

# Supplement to “Multiscale Inference and Long-Run Variance Estimation in Nonparametric Regression with Time Series Errors”

Marina Khismatullina  
University of Bonn

Michael Vogt  
University of Bonn

October 22, 2018

In this supplement, we provide the technical details and proofs that are omitted in the paper. In addition, we report the results of some robustness checks which complement the simulation exercises in Section 5 of the paper.

## S.1 Proofs of the results from Section 3

In this section, we prove the theoretical results from Section 3. We use the following notation: The symbol  $C$  denotes a universal real constant which may take a different value on each occurrence. For  $a, b \in \mathbb{R}$ , we write  $a_+ = \max\{0, a\}$  and  $a \vee b = \max\{a, b\}$ . For any set  $A$ , the symbol  $|A|$  denotes the cardinality of  $A$ . The notation  $X \stackrel{\mathcal{D}}{=} Y$  means that the two random variables  $X$  and  $Y$  have the same distribution. Finally,  $f_0(\cdot)$  and  $F_0(\cdot)$  denote the density and distribution function of the standard normal distribution, respectively.

### Auxiliary results using strong approximation theory

The main purpose of this section is to prove that there is a version of the multiscale statistic  $\widehat{\Phi}_T$  defined in (3.4) which is close to a Gaussian statistic whose distribution is known. More specifically, we prove the following result.

**Proposition S.1.** *Under the conditions of Theorem 3.1, there exist statistics  $\widetilde{\Phi}_T$  for  $T = 1, 2, \dots$  with the following two properties: (i)  $\widetilde{\Phi}_T$  has the same distribution as  $\widehat{\Phi}_T$  for any  $T$ , and (ii)*

$$|\widetilde{\Phi}_T - \Phi_T| = o_p\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T \sqrt{\log T}\right),$$

where  $\Phi_T$  is a Gaussian statistic as defined in (3.3).

**Proof of Proposition S.1.** For the proof, we draw on strong approximation theory for stationary processes  $\{\varepsilon_t\}$  that fulfill the conditions (C1)–(C3). By Theorem 2.1 and Corollary 2.1 in Berkes et al. (2014), the following strong approximation result holds true: On a richer probability space, there exist a standard Brownian motion  $\mathbb{B}$  and a sequence  $\{\tilde{\varepsilon}_t : t \in \mathbb{N}\}$  such that  $[\tilde{\varepsilon}_1, \dots, \tilde{\varepsilon}_T] \stackrel{\mathcal{D}}{=} [\varepsilon_1, \dots, \varepsilon_T]$  for each  $T$  and

$$\max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_s - \sigma \mathbb{B}(t) \right| = o(T^{1/q}) \quad \text{a.s.}, \quad (\text{S.1})$$

where  $\sigma^2 = \sum_{k \in \mathbb{Z}} \text{Cov}(\varepsilon_0, \varepsilon_k)$  denotes the long-run error variance. To apply this result, we define

$$\tilde{\Phi}_T = \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\tilde{\phi}_T(u, h)}{\tilde{\sigma}} \right| - \lambda(h) \right\},$$

where  $\tilde{\phi}_T(u, h) = \sum_{t=1}^T w_{t,T}(u, h) \tilde{\varepsilon}_t$  and  $\tilde{\sigma}^2$  is the same estimator as  $\hat{\sigma}^2$  with  $Y_{t,T} = m(t/T) + \varepsilon_t$  replaced by  $\tilde{Y}_{t,T} = m(t/T) + \tilde{\varepsilon}_t$  for  $1 \leq t \leq T$ . In addition, we let

$$\begin{aligned} \Phi_T &= \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_T(u, h)}{\sigma} \right| - \lambda(h) \right\} \\ \Phi_T^\diamond &= \max_{(u,h) \in \mathcal{G}_T} \left\{ \left| \frac{\phi_T(u, h)}{\tilde{\sigma}} \right| - \lambda(h) \right\} \end{aligned}$$

with  $\phi_T(u, h) = \sum_{t=1}^T w_{t,T}(u, h) \sigma Z_t$  and  $Z_t = \mathbb{B}(t) - \mathbb{B}(t-1)$ . With this notation, we can write

$$|\tilde{\Phi}_T - \Phi_T| \leq |\tilde{\Phi}_T - \Phi_T^\diamond| + |\Phi_T^\diamond - \Phi_T| = |\tilde{\Phi}_T - \Phi_T^\diamond| + o_p(\rho_T \sqrt{\log T}), \quad (\text{S.2})$$

where the last equality follows by taking into account that  $\phi_T(u, h) \sim N(0, \sigma^2)$  for all  $(u, h) \in \mathcal{G}_T$ ,  $|\mathcal{G}_T| = O(T^\theta)$  for some large but fixed constant  $\theta$  and  $\tilde{\sigma}^2 = \sigma^2 + o_p(\rho_T)$ . Straightforward calculations yield that

$$|\tilde{\Phi}_T - \Phi_T^\diamond| \leq \tilde{\sigma}^{-1} \max_{(u,h) \in \mathcal{G}_T} |\tilde{\phi}_T(u, h) - \phi_T(u, h)|.$$

Using summation by parts, we further obtain that

$$\begin{aligned} |\tilde{\phi}_T(u, h) - \phi_T(u, h)| &\leq W_T(u, h) \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_s - \sigma \sum_{s=1}^t \{\mathbb{B}(s) - \mathbb{B}(s-1)\} \right| \\ &= W_T(u, h) \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_s - \sigma \mathbb{B}(t) \right|, \end{aligned}$$

where

$$W_T(u, h) = \sum_{t=1}^{T-1} |w_{t+1,T}(u, h) - w_{t,T}(u, h)| + |w_{T,T}(u, h)|.$$

Standard arguments show that  $\max_{(u,h) \in \mathcal{G}_T} W_T(u,h) = O(1/\sqrt{Th_{\min}})$ . Applying the strong approximation result (S.1), we can thus infer that

$$\begin{aligned} |\tilde{\Phi}_T - \Phi_T^\diamond| &\leq \tilde{\sigma}^{-1} \max_{(u,h) \in \mathcal{G}_T} |\tilde{\phi}_T(u,h) - \phi_T(u,h)| \\ &\leq \tilde{\sigma}^{-1} \max_{(u,h) \in \mathcal{G}_T} W_T(u,h) \max_{1 \leq t \leq T} \left| \sum_{s=1}^t \tilde{\varepsilon}_s - \sigma \mathbb{B}(t) \right| = o_p\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}}\right). \end{aligned} \quad (\text{S.3})$$

Plugging (S.3) into (S.2) completes the proof.  $\square$

## Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic  $\Phi_T$  defined in (3.3). We in particular show that  $\Phi_T$  does not concentrate too strongly in small regions of the form  $[x - \delta_T, x + \delta_T]$  with  $\delta_T$  converging to zero.

**Proposition S.2.** *Under the conditions of Theorem 3.1, it holds that*

$$\sup_{x \in \mathbb{R}} \mathbb{P}\left(|\Phi_T - x| \leq \delta_T\right) = o(1),$$

where  $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T \sqrt{\log T}$ .

**Proof of Proposition S.2.** The main technical tool for proving Proposition S.2 are anti-concentration bounds for Gaussian random vectors. The following proposition slightly generalizes anti-concentration results derived in Chernozhukov et al. (2015), in particular Theorem 3 therein.

**Proposition S.3.** *Let  $(X_1, \dots, X_p)^\top$  be a Gaussian random vector in  $\mathbb{R}^p$  with  $\mathbb{E}[X_j] = \mu_j$  and  $\text{Var}(X_j) = \sigma_j^2 > 0$  for  $1 \leq j \leq p$ . Define  $\bar{\mu} = \max_{1 \leq j \leq p} |\mu_j|$  together with  $\underline{\sigma} = \min_{1 \leq j \leq p} \sigma_j$  and  $\bar{\sigma} = \max_{1 \leq j \leq p} \sigma_j$ . Moreover, set  $a_p = \mathbb{E}[\max_{1 \leq j \leq p} (X_j - \mu_j)/\sigma_j]$  and  $b_p = \mathbb{E}[\max_{1 \leq j \leq p} (X_j - \mu_j)]$ . For every  $\delta > 0$ , it holds that*

$$\sup_{x \in \mathbb{R}} \mathbb{P}\left(\left| \max_{1 \leq j \leq p} X_j - x \right| \leq \delta\right) \leq C\delta\{\bar{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\},$$

where  $C > 0$  depends only on  $\underline{\sigma}$  and  $\bar{\sigma}$ .

The proof of Proposition S.3 is provided at the end of this section for completeness. To apply Proposition S.3 to our setting at hand, we introduce the following notation: We write  $x = (u, h)$  along with  $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1, \dots, x_p\}$ , where  $p := |\mathcal{G}_T| \leq O(T^\theta)$  for some large but fixed  $\theta > 0$  by our assumptions. Moreover, for  $j = 1, \dots, p$ , we set

$$\begin{aligned} X_{2j-1} &= \frac{\phi_T(x_{j1}, x_{j2})}{\sigma} - \lambda(x_{j2}) \\ X_{2j} &= -\frac{\phi_T(x_{j1}, x_{j2})}{\sigma} - \lambda(x_{j2}) \end{aligned}$$

with  $x_j = (x_{j1}, x_{j2})$ . This notation allows us to write

$$\Phi_T = \max_{1 \leq j \leq 2p} X_j,$$

where  $(X_1, \dots, X_{2p})^\top$  is a Gaussian random vector with the following properties: (i)  $\mu_j := \mathbb{E}[X_j] = -\lambda(x_{j2})$  and thus  $\bar{\mu} = \max_{1 \leq j \leq 2p} |\mu_j| \leq C\sqrt{\log T}$ , and (ii)  $\sigma_j^2 := \text{Var}(X_j) = 1$  for all  $j$ . Since  $\sigma_j = 1$  for all  $j$ , it holds that  $a_{2p} = b_{2p}$ . Moreover, as the variables  $(X_j - \mu_j)/\sigma_j$  are standard normal, we have that  $a_{2p} = b_{2p} \leq \sqrt{2\log(2p)} \leq C\sqrt{\log T}$ . With this notation at hand, we can apply Proposition S.3 to obtain that

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|\Phi_T - x| \leq \delta_T) \leq C\delta_T \left[ \sqrt{\log T} + \sqrt{\log(1/\delta_T)} \right] = o(1)$$

with  $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$ , which is the statement of Proposition S.2.  $\square$

### Proof of Theorem 3.1

To prove Theorem 3.1, we make use of the two auxiliary results derived above. By Proposition S.1, there exist statistics  $\tilde{\Phi}_T$  for  $T = 1, 2, \dots$  which are distributed as  $\hat{\Phi}_T$  for any  $T \geq 1$  and which have the property that

$$|\tilde{\Phi}_T - \Phi_T| = o_p\left(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T}\right), \quad (\text{S.4})$$

where  $\Phi_T$  is a Gaussian statistic as defined in (3.3). The approximation result (S.4) allows us to replace the multiscale statistic  $\hat{\Phi}_T$  by an identically distributed version  $\tilde{\Phi}_T$  which is close to the Gaussian statistic  $\Phi_T$ . In the next step, we show that

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(\tilde{\Phi}_T \leq x) - \mathbb{P}(\Phi_T \leq x)| = o(1), \quad (\text{S.5})$$

which immediately implies the statement of Theorem 3.1. For the proof of (S.5), we use the following simple lemma:

**Lemma S.1.** *Let  $V_T$  and  $W_T$  be real-valued random variables for  $T = 1, 2, \dots$  such that  $V_T - W_T = o_p(\delta_T)$  with some  $\delta_T = o(1)$ . If*

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|V_T - x| \leq \delta_T) = o(1), \quad (\text{S.6})$$

then

$$\sup_{x \in \mathbb{R}} |\mathbb{P}(V_T \leq x) - \mathbb{P}(W_T \leq x)| = o(1). \quad (\text{S.7})$$

The statement of Lemma S.1 can be summarized as follows: If  $W_T$  can be approximated by  $V_T$  in the sense that  $V_T - W_T = o_p(\delta_T)$  and if  $V_T$  does not concentrate too strongly

in small regions of the form  $[x - \delta_T, x + \delta_T]$  as assumed in (S.6), then the distribution of  $W_T$  can be approximated by that of  $V_T$  in the sense of (S.7).

**Proof of Lemma S.1.** It holds that

$$\begin{aligned}
& |\mathbb{P}(V_T \leq x) - \mathbb{P}(W_T \leq x)| \\
&= |\mathbb{E}[1(V_T \leq x) - 1(W_T \leq x)]| \\
&\leq |\mathbb{E}[\{1(V_T \leq x) - 1(W_T \leq x)\}1(|V_T - W_T| \leq \delta_T)]| + |\mathbb{E}[1(|V_T - W_T| > \delta_T)]| \\
&\leq \mathbb{E}[1(|V_T - x| \leq \delta_T, |V_T - W_T| \leq \delta_T)] + o(1) \\
&\leq \mathbb{P}(|V_T - x| \leq \delta_T) + o(1).
\end{aligned}
\tag*{$\square$}$$

We now apply this lemma with  $V_T = \Phi_T$ ,  $W_T = \tilde{\Phi}_T$  and  $\delta_T = T^{1/q}/\sqrt{Th_{\min}} + \rho_T\sqrt{\log T}$ : From (S.4), we already know that  $\tilde{\Phi}_T - \Phi_T = o_p(\delta_T)$ . Moreover, by Proposition S.2, it holds that

$$\sup_{x \in \mathbb{R}} \mathbb{P}(|\Phi_T - x| \leq \delta_T) = o(1). \tag{S.8}$$

Hence, the conditions of Lemma S.1 are satisfied. Applying the lemma, we obtain (S.5), which completes the proof of Theorem 3.1.

## Proof of Proposition 3.2

To start with, we introduce the notation  $\hat{\psi}_T(u, h) = \hat{\psi}_T^A(u, h) + \hat{\psi}_T^B(u, h)$  with  $\hat{\psi}_T^A(u, h) = \sum_{t=1}^T w_{t,T}(u, h)\varepsilon_t$  and  $\hat{\psi}_T^B(u, h) = \sum_{t=1}^T w_{t,T}(u, h)m_T(\frac{t}{T})$ . By assumption, there exists  $(u_0, h_0) \in \mathcal{G}_T$  with  $[u_0 - h_0, u_0 + h_0] \subseteq [0, 1]$  such that  $m'_T(w) \geq c_T\sqrt{\log T/(Th_0^3)}$  for all  $w \in [u_0 - h_0, u_0 + h_0]$ . (The case that  $-m'_T(w) \geq c_T\sqrt{\log T/(Th_0^3)}$  for all  $w$  can be treated analogously.) Below, we prove that under this assumption,

$$\hat{\psi}_T^B(u_0, h_0) \geq \frac{\kappa c_T \sqrt{\log T}}{2} \tag{S.9}$$

for sufficiently large  $T$ , where  $\kappa = (\int K(\varphi)\varphi^2 d\varphi)/(\int K^2(\varphi)\varphi^2 d\varphi)^{1/2}$ . Moreover, by arguments very similar to those for the proof of Proposition S.1, it follows that

$$\max_{(u, h) \in \mathcal{G}_T} |\hat{\psi}_T^A(u, h)| = O_p(\sqrt{\log T}). \tag{S.10}$$

With the help of (S.9), (S.10) and the fact that  $\lambda(h) \leq \lambda(h_{\min}) \leq C\sqrt{\log T}$ , we can infer that

$$\begin{aligned}
\hat{\Psi}_T &\geq \max_{(u, h) \in \mathcal{G}_T} \frac{|\hat{\psi}_T^B(u, h)|}{\hat{\sigma}} - \max_{(u, h) \in \mathcal{G}_T} \left\{ \frac{|\hat{\psi}_T^A(u, h)|}{\hat{\sigma}} + \lambda(h) \right\} \\
&= \max_{(u, h) \in \mathcal{G}_T} \frac{|\hat{\psi}_T^B(u, h)|}{\hat{\sigma}} + O_p(\sqrt{\log T})
\end{aligned}$$

$$\geq \frac{\kappa c_T \sqrt{\log T}}{2\widehat{\sigma}} + O_p(\sqrt{\log T}) \quad (\text{S.11})$$

for sufficiently large  $T$ . Since  $q_T(\alpha) = O(\sqrt{\log T})$  for any fixed  $\alpha \in (0, 1)$ , (S.11) immediately yields that  $\mathbb{P}(\widehat{\Psi}_T \leq q_T(\alpha)) = o(1)$ , which is the statement of Proposition 3.2.

**Proof of (S.9).** Write  $m_T(\frac{t}{T}) = m_T(u_0) + m'_T(\xi_{u_0,t,T})(\frac{t}{T} - u_0)$ , where  $\xi_{u_0,t,T}$  is an intermediate point between  $u_0$  and  $t/T$ . The local linear weights  $w_{t,T}(u_0, h_0)$  are constructed such that  $\sum_{t=1}^T w_{t,T}(u_0, h_0) = 0$ . We thus obtain that

$$\widehat{\psi}_T^B(u_0, h_0) = \sum_{t=1}^T w_{t,T}(u_0, h_0) \left( \frac{\frac{t}{T} - u_0}{h_0} \right) h_0 m'_T(\xi_{u_0,t,T}). \quad (\text{S.12})$$

Moreover, since the kernel  $K$  is symmetric and  $u_0 = t/T$  for some  $t$ , it holds that  $S_{T,1}(u_0, h_0) = 0$ , which in turn implies that

$$\begin{aligned} w_{t,T}(u_0, h_0) \left( \frac{\frac{t}{T} - u_0}{h_0} \right) \\ = K \left( \frac{\frac{t}{T} - u_0}{h_0} \right) \left( \frac{\frac{t}{T} - u_0}{h_0} \right)^2 / \left\{ \sum_{t=1}^T K^2 \left( \frac{\frac{t}{T} - u_0}{h_0} \right) \left( \frac{\frac{t}{T} - u_0}{h_0} \right)^2 \right\}^{1/2} \geq 0. \end{aligned} \quad (\text{S.13})$$

From (S.12), (S.13) and the assumption that  $m'_T(w) \geq c_T \sqrt{\log T / (Th_0^3)}$  for all  $w \in [u_0 - h_0, u_0 + h_0]$ , we get that

$$\widehat{\psi}_T^B(u_0, h_0) \geq c_T \sqrt{\frac{\log T}{Th_0}} \sum_{t=1}^T w_{t,T}(u_0, h_0) \left( \frac{\frac{t}{T} - u_0}{h_0} \right). \quad (\text{S.14})$$

Standard calculations exploiting the Lipschitz continuity of the kernel  $K$  show that for any  $(u, h) \in \mathcal{G}_T$  and any given natural number  $\ell$ ,

$$\left| \frac{1}{Th} \sum_{t=1}^T K \left( \frac{\frac{t}{T} - u}{h} \right) \left( \frac{\frac{t}{T} - u}{h} \right)^\ell - \int_0^1 \frac{1}{h} K \left( \frac{w - u}{h} \right) \left( \frac{w - u}{h} \right)^\ell dw \right| \leq \frac{C}{Th}, \quad (\text{S.15})$$

where the constant  $C$  does not depend on  $u$ ,  $h$  and  $T$ . With the help of (S.13) and (S.15), we obtain that for any  $(u, h) \in \mathcal{G}_T$  with  $[u - h, u + h] \subseteq [0, 1]$ ,

$$\left| \sum_{t=1}^T w_{t,T}(u, h) \left( \frac{\frac{t}{T} - u}{h} \right) - \kappa \sqrt{Th} \right| \leq \frac{C}{\sqrt{Th}}, \quad (\text{S.16})$$

where the constant  $C$  does once again not depend on  $u$ ,  $h$  and  $T$ . (S.16) implies that  $\sum_{t=1}^T w_{t,T}(u, h) (\frac{t}{T} - u)/h \geq \kappa \sqrt{Th}/2$  for sufficiently large  $T$  and any  $(u, h) \in \mathcal{G}_T$  with  $[u - h, u + h] \subseteq [0, 1]$ . Using this together with (S.14), we immediately obtain (S.9).  $\square$

### Proof of Proposition 3.3

In what follows, we show that

$$\mathbb{P}(E_T^+) \geq (1 - \alpha) + o(1). \quad (\text{S.17})$$

The other statements of Proposition 3.3 can be verified by analogous arguments. (S.17) is a consequence of the following two observations:

(i) For all  $(u, h) \in \mathcal{G}_T$  with

$$\left| \frac{\widehat{\psi}_T(u, h) - \mathbb{E}\widehat{\psi}_T(u, h)}{\widehat{\sigma}} \right| - \lambda(h) \leq q_T(\alpha) \quad \text{and} \quad \frac{\widehat{\psi}_T(u, h)}{\widehat{\sigma}} - \lambda(h) > q_T(\alpha),$$

it holds that  $\mathbb{E}[\widehat{\psi}_T(u, h)] > 0$ .

(ii) For all  $(u, h) \in \mathcal{G}_T$  with  $[u - h, u + h] \subseteq [0, 1]$ ,  $\mathbb{E}[\widehat{\psi}_T(u, h)] > 0$  implies that  $m'(v) > 0$  for some  $v \in [u - h, u + h]$ .

Observation (i) is trivial, (ii) can be seen as follows: Let  $(u, h)$  be any point with  $(u, h) \in \mathcal{G}_T$  and  $[u - h, u + h] \subseteq [0, 1]$ . It holds that  $\mathbb{E}[\widehat{\psi}_T(u, h)] = \widehat{\psi}_T^B(u, h)$ , where  $\widehat{\psi}_T^B(u, h)$  has been defined in the proof of Proposition 3.2. As already shown in (S.12),

$$\widehat{\psi}_T^B(u, h) = \sum_{t=1}^T w_{t,T}(u, h) \left( \frac{\frac{t}{T} - u}{h} \right) h m'(\xi_{u,t,T}),$$

where  $\xi_{u,t,T}$  is some intermediate point between  $u$  and  $t/T$ . Moreover, by (S.13), it holds that  $w_{t,T}(u, h) (\frac{t}{T} - u)/h \geq 0$  for any  $t$ . Hence,  $\mathbb{E}[\widehat{\psi}_T(u, h)] = \widehat{\psi}_T^B(u, h)$  can only take a positive value if  $m'(v) > 0$  for some  $v \in [u - h, u + h]$ .

From observations (i) and (ii), we can draw the following conclusions: On the event

$$\{\widehat{\Phi}_T \leq q_T(\alpha)\} = \left\{ \max_{(u,h) \in \mathcal{G}_T} \left( \left| \frac{\widehat{\psi}_T(u, h) - \mathbb{E}\widehat{\psi}_T(u, h)}{\widehat{\sigma}} \right| - \lambda(h) \right) \leq q_T(\alpha) \right\},$$

it holds that for all  $(u, h) \in \mathcal{A}_T^+$  with  $[u - h, u + h] \subseteq [0, 1]$ ,  $m'(v) > 0$  for some  $v \in I_{u,h} = [u - h, u + h]$ . We thus obtain that  $\{\widehat{\Phi}_T \leq q_T(\alpha)\} \subseteq E_T^+$ . This in turn implies that

$$\mathbb{P}(E_T^+) \geq \mathbb{P}(\widehat{\Phi}_T \leq q_T(\alpha)) = (1 - \alpha) + o(1),$$

where the last equality holds by Theorem 3.1.

## Proof of Proposition S.3

The proof makes use of the following three lemmas, which correspond to Lemmas 5–7 in Chernozhukov et al. (2015).

**Lemma S.2.** *Let  $(W_1, \dots, W_p)^\top$  be a (not necessarily centred) Gaussian random vector in  $\mathbb{R}^p$  with  $\text{Var}(W_j) = 1$  for all  $1 \leq j \leq p$ . Suppose that  $\text{Corr}(W_j, W_k) < 1$  whenever  $j \neq k$ . Then the distribution of  $\max_{1 \leq j \leq p} W_j$  is absolutely continuous with respect to Lebesgue measure and a version of the density is given by*

$$f(x) = f_0(x) \sum_{j=1}^p e^{\mathbb{E}[W_j]x - \mathbb{E}[W_j]^2/2} \mathbb{P}(W_k \leq x \text{ for all } k \neq j \mid W_j = x).$$

**Lemma S.3.** *Let  $(W_0, W_1, \dots, W_p)^\top$  be a (not necessarily centred) Gaussian random vector with  $\text{Var}(W_j) = 1$  for all  $0 \leq j \leq p$ . Suppose that  $\mathbb{E}[W_0] \geq 0$ . Then the map*

$$x \mapsto e^{\mathbb{E}[W_0]x - \mathbb{E}[W_0]^2/2} \mathbb{P}(W_j \leq x \text{ for } 1 \leq j \leq p \mid W_0 = x)$$

*is non-decreasing on  $\mathbb{R}$ .*

**Lemma S.4.** *Let  $(X_1, \dots, X_p)^\top$  be a centred Gaussian random vector in  $\mathbb{R}^p$  with  $\max_{1 \leq j \leq p} \mathbb{E}[X_j^2] \leq \sigma_X^2$  for some  $\sigma_X^2 > 0$ . Then for any  $r > 0$ ,*

$$\mathbb{P}\left(\max_{1 \leq j \leq p} X_j \geq \mathbb{E}\left[\max_{1 \leq j \leq p} X_j\right] + r\right) \leq e^{-r^2/(2\sigma_X^2)}.$$

The proof of Lemmas S.2 and S.3 can be found in Chernozhukov et al. (2015). Lemma S.4 is a standard result on Gaussian concentration whose proof is given e.g. in Ledoux (2001); see Theorem 7.1 therein. We now closely follow the arguments for the proof of Theorem 3 in Chernozhukov et al. (2015). The proof splits up into three steps.

*Step 1.* Pick any  $x \geq 0$  and set

$$W_j = \frac{X_j - x}{\sigma_j} + \frac{\bar{\mu} + x}{\underline{\sigma}}.$$

By construction,  $\mathbb{E}[W_j] \geq 0$  and  $\text{Var}(W_j) = 1$ . Defining  $Z = \max_{1 \leq j \leq p} W_j$ , it holds that

$$\begin{aligned} \mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) &\leq \mathbb{P}\left(\left|\max_{1 \leq j \leq p} \frac{X_j - x}{\sigma_j}\right| \leq \frac{\delta}{\underline{\sigma}}\right) \\ &\leq \sup_{y \in \mathbb{R}} \mathbb{P}\left(\left|\max_{1 \leq j \leq p} \frac{X_j - x}{\sigma_j} + \frac{\bar{\mu} + x}{\underline{\sigma}} - y\right| \leq \frac{\delta}{\underline{\sigma}}\right) \\ &= \sup_{y \in \mathbb{R}} \mathbb{P}\left(|Z - y| \leq \frac{\delta}{\underline{\sigma}}\right). \end{aligned}$$

*Step 2.* We now bound the density of  $Z$ . Without loss of generality, we assume that



$\text{Corr}(W_j, W_k) < 1$  for  $k \neq j$ . The marginal distribution of  $W_j$  is  $N(\nu_j, 1)$  with  $\nu_j = \mathbb{E}[W_j] = (\mu_j/\sigma_j + \bar{\mu}/\underline{\sigma}) + (x/\underline{\sigma} - x/\sigma_j) \geq 0$ . Hence, by Lemmas S.2 and S.3, the random variable  $Z$  has a density of the form

$$f_p(z) = f_0(z)G_p(z), \quad (\text{S.18})$$

where the map  $z \mapsto G_p(z)$  is non-decreasing. Define  $\bar{Z} = \max_{1 \leq j \leq p} (W_j - \mathbb{E}[W_j])$  and set  $\bar{z} = 2\bar{\mu}/\underline{\sigma} + x(1/\underline{\sigma} - 1/\bar{\sigma})$  such that  $\mathbb{E}[W_j] \leq \bar{z}$  for any  $1 \leq j \leq p$ . With these definitions at hand, we obtain that

$$\begin{aligned} \int_z^\infty f_0(u)du G_p(z) &\leq \int_z^\infty f_0(u)G_p(u)du = \mathbb{P}(Z > z) \\ &\leq P(\bar{Z} > z - \bar{z}) \leq \exp\left(-\frac{(z - \bar{z} - \mathbb{E}[\bar{Z}])_+^2}{2}\right), \end{aligned}$$

where the last inequality follows from Lemma S.4. Since  $W_j - \mathbb{E}[W_j] = (X_j - \mu_j)/\sigma_j$ , it holds that

$$\mathbb{E}[\bar{Z}] = \mathbb{E}\left[\max_{1 \leq j \leq p} \left\{\frac{X_j - \mu_j}{\sigma_j}\right\}\right] =: a_p.$$

Hence, for every  $z \in \mathbb{R}$ ,

$$G_p(z) \leq \frac{1}{1 - F_0(z)} \exp\left(-\frac{(z - \bar{z} - a_p)_+^2}{2}\right). \quad (\text{S.19})$$

Mill's inequality states that for  $z > 0$ ,

$$z \leq \frac{f_0(z)}{1 - F_0(z)} \leq z \frac{1 + z^2}{z^2}.$$

Since  $(1 + z^2)/z^2 \leq 2$  for  $z \geq 1$  and  $f_0(z)/\{1 - F_0(z)\} \leq 1.53 \leq 2$  for  $z \in (-\infty, 1)$ , we can infer that

$$\frac{f_0(z)}{1 - F_0(z)} \leq 2(z \vee 1) \quad \text{for any } z \in \mathbb{R}.$$

This together with (S.18) and (S.19) yields that

$$f_p(z) \leq 2(z \vee 1) \exp\left(-\frac{(z - \bar{z} - a_p)_+^2}{2}\right) \quad \text{for any } z \in \mathbb{R}.$$

*Step 3.* By Step 2, we get that for any  $y \in \mathbb{R}$  and  $u > 0$ ,

$$\mathbb{P}(|Z - y| \leq u) = \int_{y-u}^{y+u} f_p(z)dz \leq 2u \max_{z \in [y-u, y+u]} f_p(z) \leq 4u(\bar{z} + a_p + 1),$$

where the last inequality follows from the fact that the map  $z \mapsto ze^{-(z-a)^2/2}$  (with  $a > 0$ ) is non-increasing on  $[a + 1, \infty)$ . Combining this bound with Step 1, we further

obtain that for any  $x \geq 0$  and  $\delta > 0$ ,

$$\mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) \leq 4\delta \left\{ \frac{2\bar{\mu}}{\underline{\sigma}} + |x| \left( \frac{1}{\underline{\sigma}} - \frac{1}{\bar{\sigma}} \right) + a_p + 1 \right\} / \underline{\sigma}. \quad (\text{S.20})$$

This inequality also holds for  $x < 0$  by an analogous argument, and hence for all  $x \in \mathbb{R}$ . Now let  $0 < \delta \leq \underline{\sigma}$  and define  $b_p = \mathbb{E} \max_{1 \leq j \leq p} \{X_j - \mu_j\}$ . For any  $|x| \leq \delta + \bar{\mu} + b_p + \bar{\sigma} \sqrt{2 \log(\underline{\sigma}/\delta)}$ , (S.20) yields that

$$\begin{aligned} \mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) &\leq \frac{4\delta}{\underline{\sigma}} \left\{ \bar{\mu} \left( \frac{3}{\underline{\sigma}} - \frac{1}{\bar{\sigma}} \right) + a_p + \left( \frac{1}{\underline{\sigma}} - \frac{1}{\bar{\sigma}} \right) b_p \right. \\ &\quad \left. + \left( \frac{\bar{\sigma}}{\underline{\sigma}} - 1 \right) \sqrt{2 \log \left( \frac{\bar{\sigma}}{\underline{\sigma}} \right) + 2 - \frac{\bar{\sigma}}{\underline{\sigma}}} \right\} \\ &\leq C\delta \{ \bar{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)} \} \end{aligned} \quad (\text{S.21})$$

with a sufficiently large constant  $C > 0$  that depends only on  $\underline{\sigma}$  and  $\bar{\sigma}$ . For  $|x| \geq \delta + \bar{\mu} + b_p + \bar{\sigma} \sqrt{2 \log(\underline{\sigma}/\delta)}$ , we obtain that

$$\mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) \leq \frac{\delta}{\underline{\sigma}}, \quad (\text{S.22})$$

which can be seen as follows: If  $x > \delta + \bar{\mu}$ , then  $|\max_j X_j - x| \leq \delta$  implies that  $|x| - \delta \leq \max_j X_j \leq \max_j \{X_j - \mu_j\} + \bar{\mu}$  and thus  $\max_j \{X_j - \mu_j\} \geq |x| - \delta - \bar{\mu}$ . Hence, it holds that

$$\mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) \leq \mathbb{P}\left(\max_{1 \leq j \leq p} \{X_j - \mu_j\} \geq |x| - \delta - \bar{\mu}\right). \quad (\text{S.23})$$

If  $x < -(\delta + \bar{\mu})$ , then  $|\max_j X_j - x| \leq \delta$  implies that  $\max_j \{X_j - \mu_j\} \leq -|x| + \delta + \bar{\mu}$ . Hence, in this case,

$$\begin{aligned} \mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) &\leq \mathbb{P}\left(\max_{1 \leq j \leq p} \{X_j - \mu_j\} \leq -|x| + \delta + \bar{\mu}\right) \\ &\leq \mathbb{P}\left(\max_{1 \leq j \leq p} \{X_j - \mu_j\} \geq |x| - \delta - \bar{\mu}\right), \end{aligned} \quad (\text{S.24})$$

where the last inequality follows from the fact that for centred Gaussian random variables  $V_j$  and  $v > 0$ ,  $\mathbb{P}(\max_j V_j \leq -v) \leq \mathbb{P}(V_1 \leq -v) = P(V_1 \geq v) \leq \mathbb{P}(\max_j V_j \geq v)$ . With (S.23) and (S.24), we obtain that for any  $|x| \geq \delta + \bar{\mu} + b_p + \bar{\sigma} \sqrt{2 \log(\underline{\sigma}/\delta)}$ ,

$$\begin{aligned} \mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) &\leq \mathbb{P}\left(\max_{1 \leq j \leq p} \{X_j - \mu_j\} \geq |x| - \delta - \bar{\mu}\right) \\ &\leq \mathbb{P}\left(\max_{1 \leq j \leq p} \{X_j - \mu_j\} \geq \mathbb{E}\left[\max_{1 \leq j \leq p} \{X_j - \mu_j\}\right] + \bar{\sigma} \sqrt{2 \log(\underline{\sigma}/\delta)}\right) \leq \frac{\delta}{\underline{\sigma}}, \end{aligned}$$

the last inequality following from Lemma S.4. To sum up, we have established that for

any  $0 < \delta \leq \underline{\sigma}$  and any  $x \in \mathbb{R}$ ,

$$\mathbb{P}\left(\left|\max_{1 \leq j \leq p} X_j - x\right| \leq \delta\right) \leq C\delta\{\bar{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\} \quad (\text{S.25})$$

with some constant  $C > 0$  that does only depend on  $\underline{\sigma}$  and  $\bar{\sigma}$ . For  $\delta > \underline{\sigma}$ , (S.25) trivially follows upon setting  $C \geq 1/\underline{\sigma}$ . This completes the proof.

## S.2 Proofs of the results from Section 4

In what follows, we prove Proposition 4.1 from Section 4. The notation is the same as in the previous section. In particular, we use the symbol  $C$  to denote a generic constant which may take a different value on each occurrence.

### Auxiliary results

To start with, we derive some auxiliary results needed for the proof of Proposition 4.1. The first lemma analyses the term

$$\xi(\ell_1, \ell_2, L) = \frac{1}{T-L} \sum_{t=L+1}^T \varepsilon_{t-\ell_1} \varepsilon_{t-\ell_2},$$

where  $\ell_1, \ell_2$  and  $L$  are natural numbers with  $0 \leq \ell_1, \ell_2 \leq L$  that may depend on the sample size  $T$ , that is,  $L = L_T$  as well as  $\ell_1 = \ell_{1,T}$  and  $\ell_2 = \ell_{2,T}$ .

**Lemma S.5.** *For any  $L = L_T$  with  $L_T/T \rightarrow 0$ , it holds that*

$$\mathbb{E}\left[\left\{\xi(\ell_1, \ell_2, L) - \gamma_\varepsilon(\ell_2 - \ell_1)\right\}^2\right] = O(T^{-1}),$$

where  $\gamma_\varepsilon(\ell) = \text{Cov}(\varepsilon_t, \varepsilon_{t-\ell})$ .

**Proof of Lemma S.5.** Since the variables  $\varepsilon_t$  have the expansion  $\varepsilon_t = \sum_{k=0}^{\infty} c_k \eta_{t-k}$  and  $\gamma_\varepsilon(\ell) = (\sum_{k=0}^{\infty} c_k c_{k+\ell}) \nu^2$ , it holds that

$$\mathbb{E}[\xi^2(\ell_1, \ell_2, L)] = \frac{1}{(T-L)^2} \sum_{t, t'=L+1}^T \mathbb{E}[\varepsilon_{t-\ell_1} \varepsilon_{t-\ell_2} \varepsilon_{t'-\ell_1} \varepsilon_{t'-\ell_2}],$$

where

$$\begin{aligned} & \mathbb{E}[\varepsilon_{t-\ell_1} \varepsilon_{t-\ell_2} \varepsilon_{t'-\ell_1} \varepsilon_{t'-\ell_2}] \\ &= \left( \sum_{k=0}^{\infty} c_k c_{k+\ell_1-\ell_2} c_{k+t'-t} c_{k+t'-t+\ell_1-\ell_2} \right) \kappa + \left( \sum_{k=0}^{\infty} c_k c_{k+\ell_1-\ell_2} \right)^2 \nu^4 \\ &+ \left( \sum_{k=0}^{\infty} c_k c_{k+t'-t} \right)^2 \nu^4 + \left( \sum_{k=0}^{\infty} c_k c_{k+t'-t-\ell_1+\ell_2} \right) \left( \sum_{k=0}^{\infty} c_k c_{k+t'-t+\ell_1-\ell_2} \right) \nu^4 \end{aligned}$$

$$\begin{aligned}
&= \left( \sum_{k=0}^{\infty} c_k c_{k+\ell_1-\ell_2} c_{k+t'-t} c_{k+t'-t+\ell_1-\ell_2} \right) \kappa + \gamma_{\varepsilon}^2(\ell_1 - \ell_2) \\
&\quad + \gamma_{\varepsilon}^2(t' - t) + \gamma_{\varepsilon}(t' - t - \ell_1 + \ell_2) \gamma_{\varepsilon}(t' - t + \ell_1 - \ell_2)
\end{aligned}$$

with  $\kappa = \mathbb{E}[\eta_0^4] - 3\nu^4$  and  $c_k = 0$  for  $k < 0$ . Noting that

$$\mathbb{E}\left[\left\{\xi(\ell_1, \ell_2, L) - \gamma_{\varepsilon}(\ell_1 - \ell_2)\right\}^2\right] = \mathbb{E}\left[\xi^2(\ell_1, \ell_2, L)\right] - \gamma_{\varepsilon}^2(\ell_1 - \ell_2),$$

we can infer that

$$\begin{aligned}
&\mathbb{E}\left[\left\{\xi(\ell_1, \ell_2, L) - \gamma_{\varepsilon}(\ell_1 - \ell_2)\right\}^2\right] \\
&= \frac{1}{(T-L)^2} \sum_{t,t'=L+1}^T \left( \sum_{k=0}^{\infty} c_k c_{k+\ell_1-\ell_2} c_{k+t'-t} c_{k+t'-t+\ell_1-\ell_2} \right) \kappa + \frac{1}{(T-L)^2} \sum_{t,t'=L+1}^T \gamma_{\varepsilon}^2(t' - t) \\
&\quad + \frac{1}{(T-L)^2} \sum_{t,t'=L+1}^T \gamma_{\varepsilon}(t' - t - \ell_1 + \ell_2) \gamma_{\varepsilon}(t' - t + \ell_1 - \ell_2) \\
&= O(T^{-1}),
\end{aligned}$$

the last equality following from the fact that the autocovariances  $\gamma_{\varepsilon}(\ell)$  are absolutely summable and the coefficients  $c_k$  decay exponentially fast to zero.  $\square$

We next show that the empirical autocovariances

$$\hat{\gamma}_q(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^T \Delta_q Y_{t,T} \Delta_q Y_{t-\ell,T}$$

of the process  $\{\Delta_q Y_{t,T}\}$  have the following property.

**Lemma S.6.** *For any  $q = q_T$  with  $q_T/\sqrt{T} \rightarrow 0$  and any  $1 \leq \ell \leq p+1$ , it holds that*

$$\hat{\gamma}_q(\ell) - \gamma_q(\ell) = O_p(T^{-1/2}),$$

where  $\gamma_q(\ell) = \text{Cov}(\Delta_q \varepsilon_t, \Delta_q \varepsilon_{t-\ell})$ .

**Proof of Lemma S.6.** To analyse the term  $\hat{\gamma}_q(\ell)$ , we decompose it as follows:

$$\hat{\gamma}_q(\ell) = \hat{\gamma}_q^*(\ell) + R_A + R_B + R_C,$$

where

$$\hat{\gamma}_q^*(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^T \Delta_q \varepsilon_t \Delta_q \varepsilon_{t-\ell}$$

as well as  $R_A = (T-q)^{-1} \sum_{t=q+\ell+1}^T \Delta_q m_t \Delta_q \varepsilon_{t-\ell}$ ,  $R_B = (T-q)^{-1} \sum_{t=q+\ell+1}^T \Delta_q \varepsilon_t \Delta_q m_{t-\ell}$  and  $R_C = (T-q)^{-1} \sum_{t=q+\ell+1}^T \Delta_q m_t \Delta_q m_{t-\ell}$  with  $\Delta_q m_t = m(\frac{t}{T}) - m(\frac{t-q}{T})$ . With the

help of Lemma S.5, it is straightforward to show that

$$\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell) = O_p(T^{-1/2}).$$

Moreover, the Cauchy-Schwarz inequality yields that

$$\mathbb{E}[R_A^2] \leq \left\{ \frac{1}{T-q} \sum_{t=q+\ell+1}^T (\Delta_q m_t)^2 \right\} \mathbb{E} \left[ \frac{1}{T-q} \sum_{t=q+\ell+1}^T (\Delta_q \varepsilon_{t-\ell})^2 \right].$$

Since  $m$  is Lipschitz by assumption, we get that  $(T-q)^{-1} \sum_{t=q+\ell+1}^T (\Delta_q m_t)^2 \leq C(q/T)^2$ . In addition, it obviously holds that  $\mathbb{E}[(T-q)^{-1} \sum_{t=q+\ell+1}^T (\Delta_q \varepsilon_{t-\ell})^2] = O(1)$ . Hence, we can infer that

$$\mathbb{E}[R_A^2] = O\left(\left\{\frac{q}{T}\right\}^2\right),$$

which implies that  $R_A = o_p(T^{-1/2})$ . Similar arguments yield that  $R_j = o_p(T^{-1/2})$  for  $j = B, C$  as well. Putting everything together, we arrive at the statement of Lemma S.6.  $\square$

## Proof of Proposition 4.1

We first show that the pilot estimator  $\widetilde{\mathbf{a}}_q$  converges to  $\mathbf{a}$ . In particular, we verify that  $\widetilde{\mathbf{a}}_q - \mathbf{a} = O_p(T^{-1/2})$ . By Lemma S.6, it holds that  $\widehat{\mathbf{\Gamma}}_q = \mathbf{\Gamma}_q + O_p(T^{-1/2})$  and  $\widehat{\boldsymbol{\gamma}}_q = \boldsymbol{\gamma}_q + O_p(T^{-1/2})$ . Since  $\mathbf{\Gamma}_q$  is invertible, this implies that

$$\widetilde{\mathbf{a}}_q = \mathbf{\Gamma}_q^{-1} \boldsymbol{\gamma}_q + O_p(T^{-1/2}).$$

With the help of equation (4.10), we can further infer that

$$\widetilde{\mathbf{a}}_q - \mathbf{a} = -\nu^2 \mathbf{\Gamma}_q^{-1} \mathbf{c}_q + O_p(T^{-1/2}).$$

As already noted in Section 4.2, the entries of the vector  $\mathbf{c}_q = (c_{q-1}, \dots, c_{c-p})^\top$  decay exponentially fast to zero, that is,  $|c_k| \leq C\rho^k$  for some  $0 < \rho < 1$ . Moreover, it holds that  $\gamma_q(\ell) \rightarrow 2\gamma_\varepsilon(\ell)$  for any fixed  $\ell$  as  $q \rightarrow \infty$ . Consequently,  $\|\nu^2 \mathbf{\Gamma}_q^{-1} \mathbf{c}_q\|_\infty = o(T^{-1/2})$ , where  $\|\cdot\|_\infty$  denotes the usual supremum norm for vectors. As a result, we obtain that  $\widetilde{\mathbf{a}}_q - \mathbf{a} = O_p(T^{-1/2})$ .

We next show that  $\widehat{\mathbf{a}}_r - \mathbf{a} = O_p(T^{-1/2})$ , where  $r \geq 1$  is any fixed integer that does not grow with the sample size  $T$ . By definition, it holds that  $\widehat{\mathbf{a}}_r = \widehat{\mathbf{\Gamma}}_r^{-1}(\widehat{\boldsymbol{\gamma}}_r + \widetilde{\nu}^2 \widetilde{\mathbf{c}}_r)$ . From Lemma S.6, it follows that  $\widehat{\mathbf{\Gamma}}_r^{-1} = \mathbf{\Gamma}_r^{-1} + O_p(T^{-1/2})$  and  $\widehat{\boldsymbol{\gamma}}_r = \boldsymbol{\gamma}_r + O_p(T^{-1/2})$ . Moreover, with the help of the fact that  $\widetilde{\mathbf{a}}_q - \mathbf{a} = O_p(T^{-1/2})$ , it is straightforward to verify that  $\widetilde{\nu}^2 - \nu^2 = O_p(T^{-1/2})$  and  $\widetilde{\mathbf{c}}_r - \mathbf{c}_r = O_p(T^{-1/2})$ . Hence, we arrive at

$$\widehat{\mathbf{a}}_r = \mathbf{\Gamma}_r^{-1}(\boldsymbol{\gamma}_r + \nu^2 \mathbf{c}_r) + O_p(T^{-1/2}) = \mathbf{a} + O_p(T^{-1/2}), \quad (\text{S.26})$$

where the last equality is due to equation (4.10).

From (S.26), it immediately follows that  $\hat{\mathbf{a}} - \mathbf{a} = O_p(T^{-1/2})$ , which in turn allows us to infer that  $\hat{\nu}^2 - \nu^2 = O_p(T^{-1/2})$  and  $\hat{\sigma}^2 = \sigma^2 + O_p(T^{-1/2})$  by straightforward arguments.

### S.3 Robustness checks and implementation details for the simulations in Section 5

#### Robustness checks for Section 5.3

In what follows, we carry out some robustness checks to assess how sensitive the estimators  $\hat{a}$  and  $\hat{\sigma}^2$  are to the choice of the tuning parameters  $q$  and  $\bar{r}$ . To do so, we repeat the simulation exercises of Section 5.3 for different values of  $q$  and  $\bar{r}$ . In addition, we consider different choices of the tuning parameters  $(m_1, m_2)$  on which the estimators of Hall and Van Keilegom (2003) depend. As in Section 5.3, we choose  $m_1$  and  $m_2$  such that  $q$  lies between these values. We thus keep the parameters  $q$  and  $(m_1, m_2)$  roughly comparable.

To start with, we consider the simulation scenarios with a moderate trend ( $s_\beta = 1$ ). The MSE values of the estimators  $\hat{a}$ ,  $\hat{a}_{\text{HvK}}$ ,  $\hat{a}_{\text{oracle}}$  and  $\hat{\sigma}^2$ ,  $\hat{\sigma}_{\text{HvK}}^2$ ,  $\hat{\sigma}_{\text{oracle}}^2$  for these scenarios are presented in Figure 3 of Section 5.3. These MSEs are re-calculated in Figures S.1 and S.2 for a range of different choices of  $q$ ,  $\bar{r}$  and  $(m_1, m_2)$ . As one can see, the MSEs in the different plots of Figures S.1 and S.2 are very similar. Hence, the MSE results reported in Section 5.3 for the scenarios with a moderate trend appear to be fairly robust to different choices of the tuning parameters. In particular, our estimators  $\hat{a}$  and  $\hat{\sigma}^2$  seem to be quite insensitive to the choice of tuning parameters, at least as far as their MSEs are concerned.

We next turn to the simulation designs with a pronounced trend ( $s_\beta = 10$ ). The MSE values of the estimators in these scenarios are reported in Figure 4 of Section 5.3. Analogously as before, we re-calculate these MSEs for different tuning parameters in Figures S.3–S.5. Figure S.4 is a zoomed-in version of Figure S.3 which is added for better visibility. As can be seen, our estimators appear to be barely influenced by the choice of  $q$ . However, the MSE values become somewhat larger when  $\bar{r}$  is chosen bigger. This is of course not very suprising: The main reason why the estimator  $\hat{a}$  works well in the presence of a strong trend is that it is only based on differences of small orders. If we increase  $\bar{r}$ , we use larger differences to compute  $\hat{a}$ , which results in not eliminating the trend  $m$  appropriately any more. This becomes visible in somewhat larger MSE values. Nevertheless, our estimators appear not to be strongly influenced by the choice of tuning parameters (in terms of MSE) as long as these are chosen within reasonable bounds.

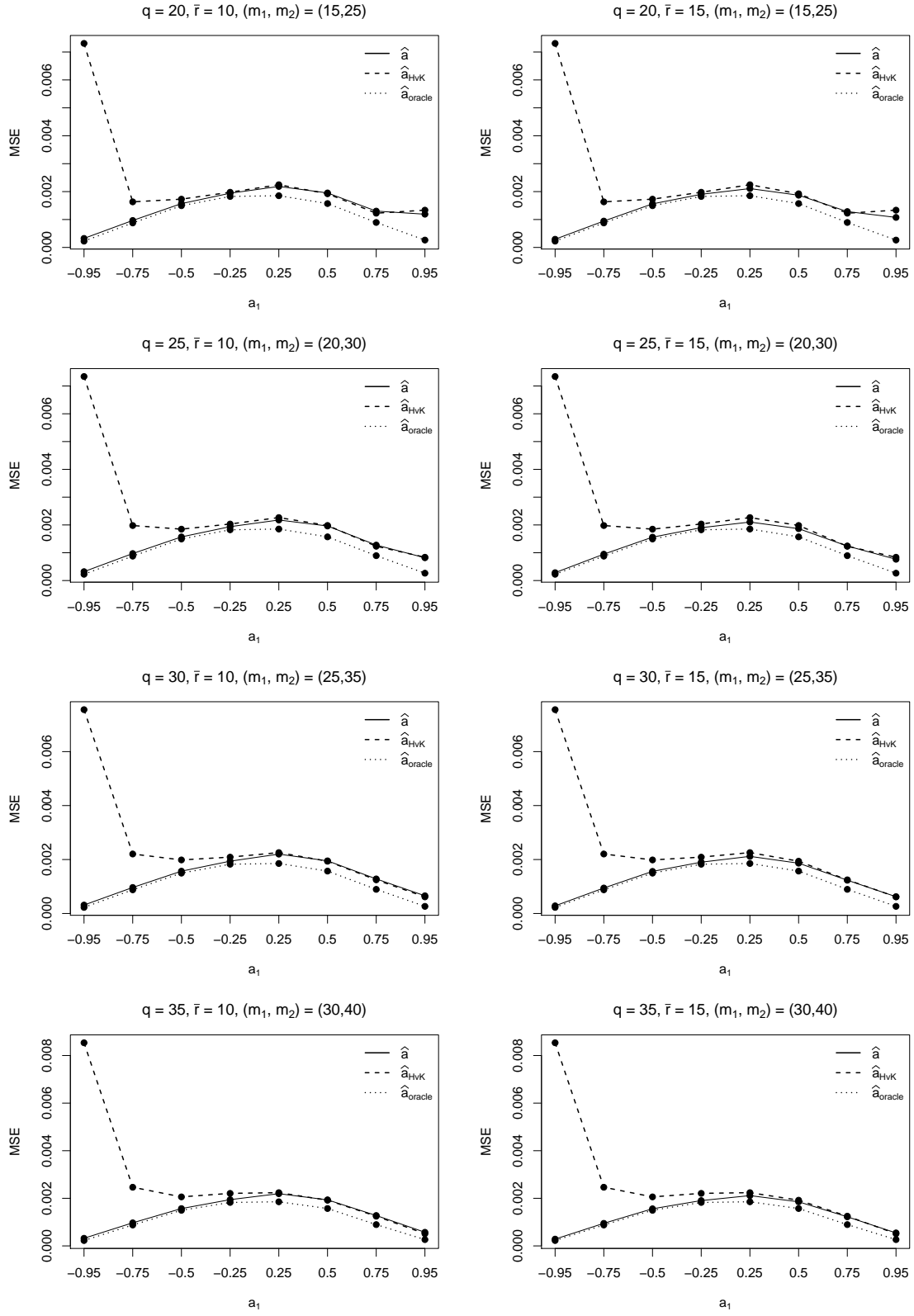


Figure S.1: MSE values for the estimators  $\hat{a}$ ,  $\hat{a}_{HvK}$  and  $\hat{a}_{oracle}$  in the scenario with a moderate trend ( $s_\beta = 1$ ).

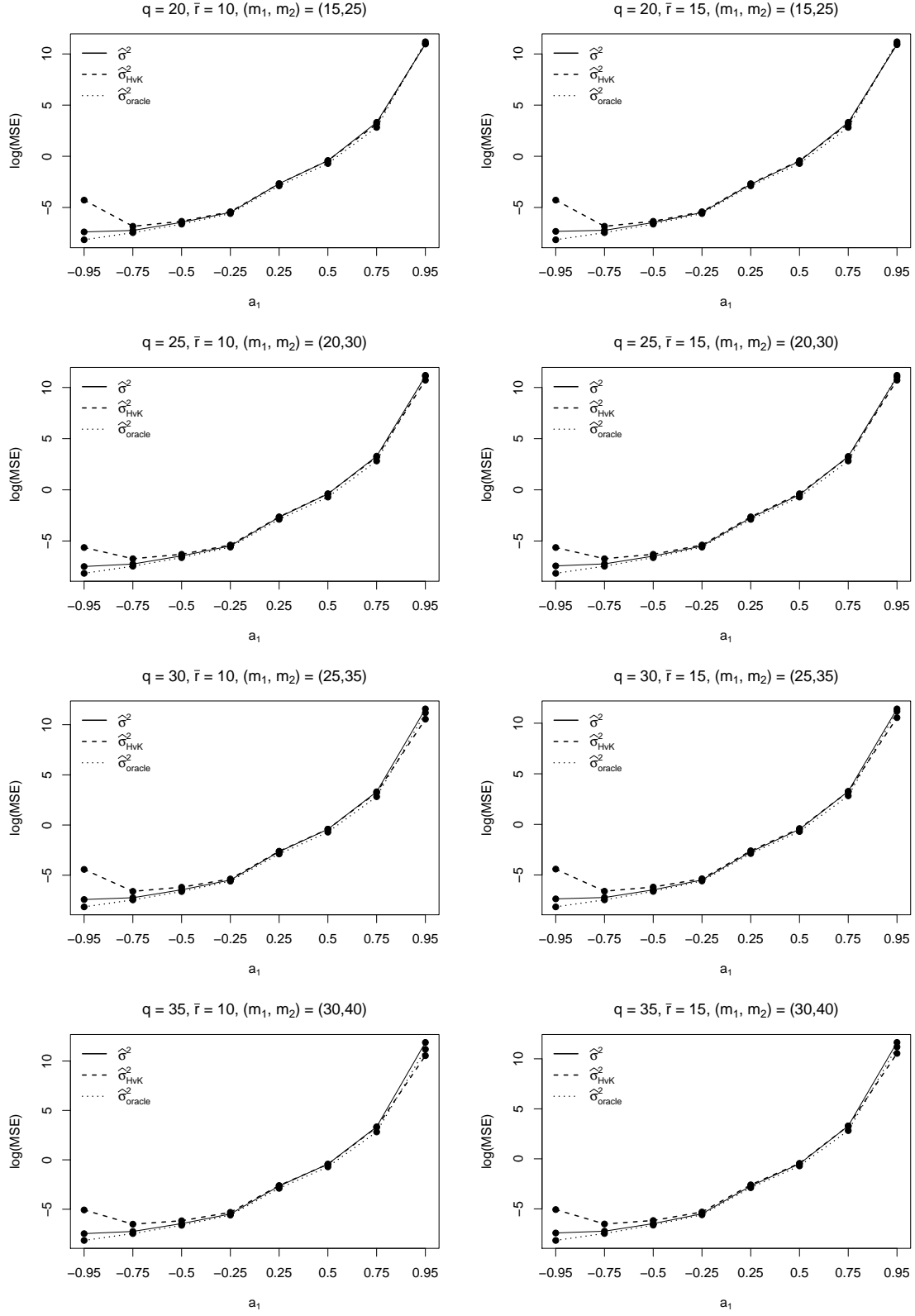


Figure S.2: Logarithmic MSE values for the estimators  $\hat{\sigma}^2$ ,  $\hat{\sigma}_{\text{HvK}}^2$  and  $\hat{\sigma}_{\text{oracle}}^2$  in the scenario with a moderate trend ( $s_\beta = 1$ ).



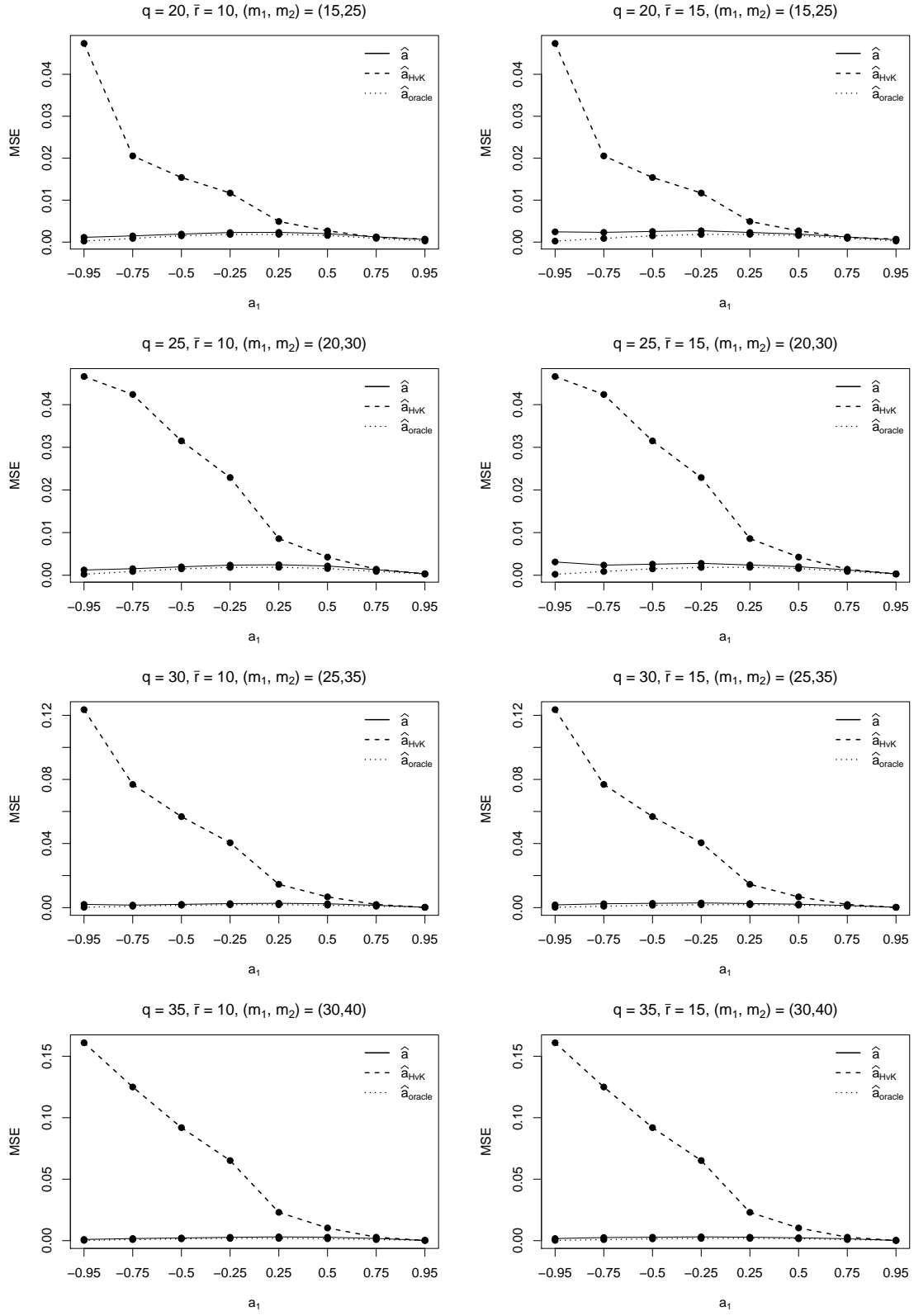


Figure S.3: MSE values for the estimators  $\hat{a}$ ,  $\hat{a}_{\text{HvK}}$  and  $\hat{a}_{\text{oracle}}$  in the scenario with a pronounced trend ( $s_\beta = 10$ ).

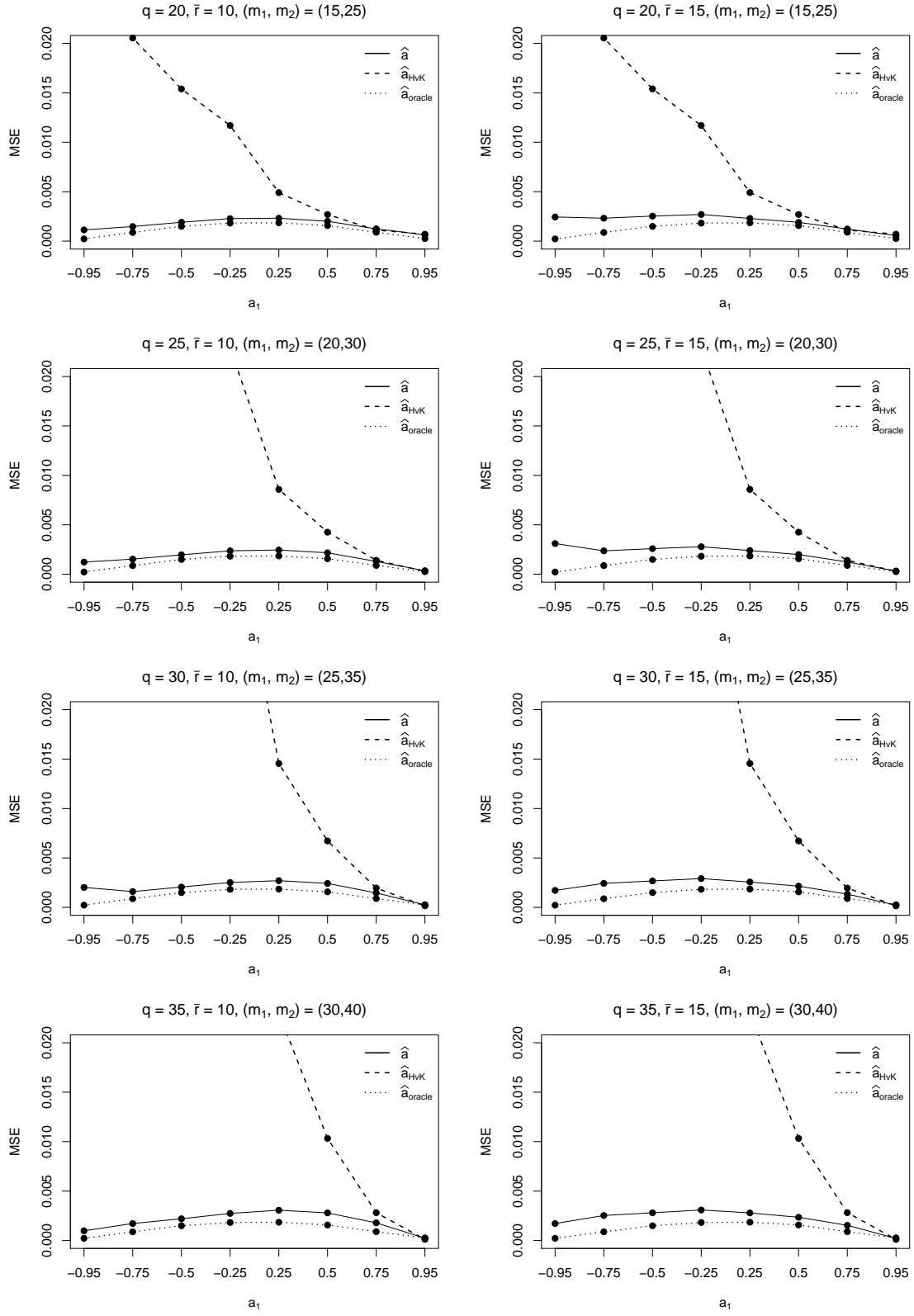


Figure S.4: MSE values for the estimators  $\hat{a}$ ,  $\hat{a}_{HvK}$  and  $\hat{a}_{oracle}$  in the scenario with a pronounced trend ( $s_\beta = 10$ ).

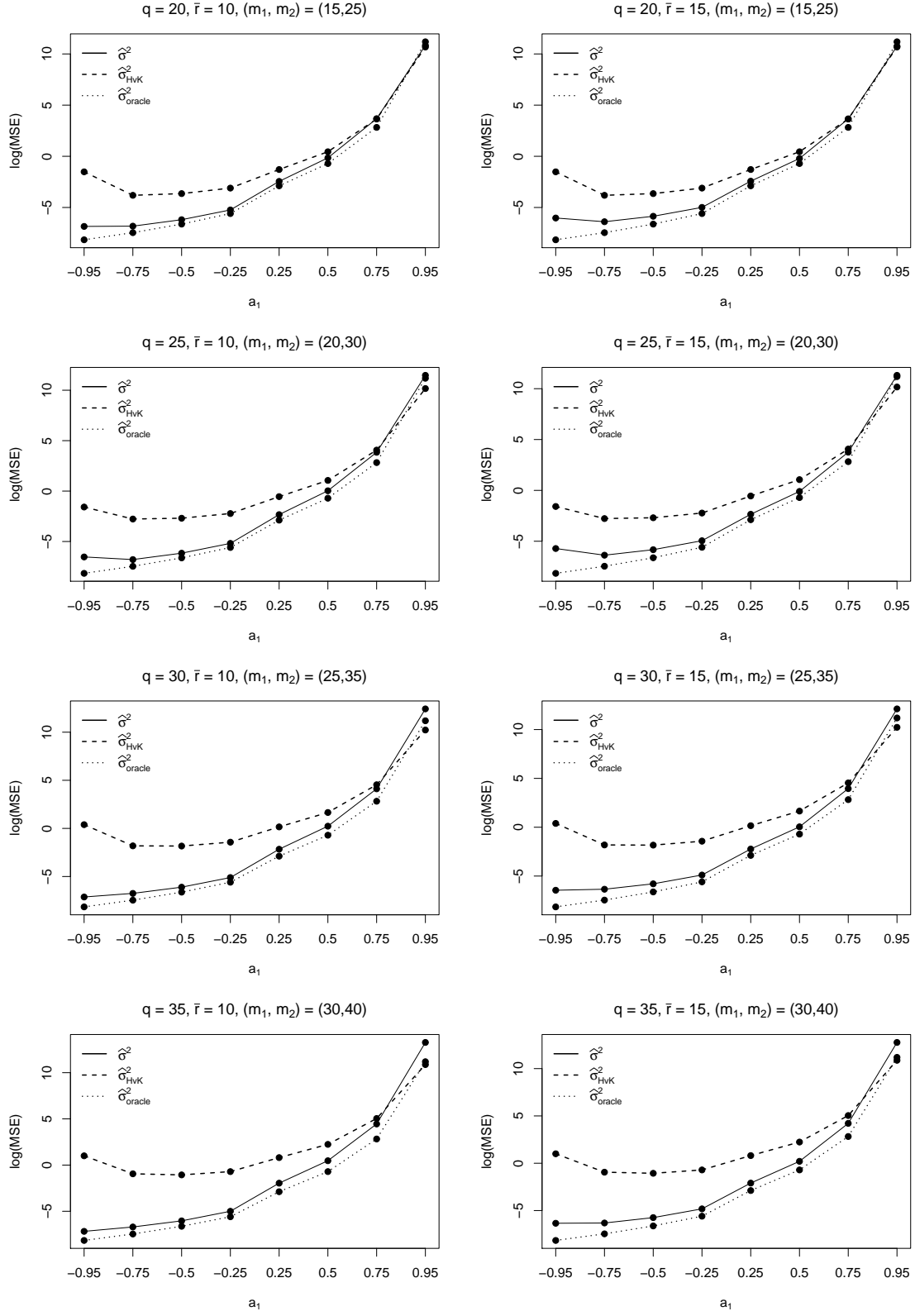


Figure S.5: Logarithmic MSE values for the estimators  $\hat{\sigma}^2$ ,  $\hat{\sigma}_{\text{HvK}}^2$  and  $\hat{\sigma}_{\text{oracle}}^2$  in the scenario with a pronounced trend ( $s_\beta = 10$ ).

## Implementation of SiZer in Section 5.2

The SiZer methods for the comparison study in Section 5.2 are implemented as follows:

(a) Computation of the grid  $\mathcal{G}_T^*$ :

To start with, we compute the variance of  $\bar{Y} = T^{-1} \sum_{t=1}^T Y_{t,T}$ , which is given by

$$\text{Var}(\bar{Y}) = \frac{\gamma_\varepsilon(0)}{T} + \frac{2}{T} \sum_{k=1}^{T-1} \left(1 - \frac{k}{T}\right) \gamma_\varepsilon(k).$$

Since the autocovariance function  $\gamma_\varepsilon(\cdot)$  is known by assumption, we obtain  $\text{Var}(\bar{Y})$  by using the formula  $\gamma_\varepsilon(k) = \nu^2 a_1^{|k|} / (1 - a_1^2)$  together with the true parameters  $a_1$  and  $\nu^2 = \mathbb{E}[\eta_t^2]$ . We next compute

$$T^* = \frac{\gamma_\varepsilon(0)}{\text{Var}(\bar{Y})},$$

which can be interpreted as a measure of information in the data. For each point  $(u, h) \in \mathcal{G}_T$  from (5.1), we finally calculate the effective sample size for dependent data

$$\text{ESS}^*(u, h) = \frac{T^* \sum_{t=1}^T K_h(t/T - u)}{K_h(0)}$$

with  $K_h(v) = h^{-1}K(v/h)$  and set  $\mathcal{G}_T^* = \{(u, h) \in \mathcal{G}_T : \text{ESS}^*(u, h) \geq 5\}$ .

(b) Computation of the local linear estimators and their standard deviations:

For each  $(u, h) \in \mathcal{G}_T^*$ , we compute a standard local linear estimator  $\hat{m}'_h(u)$  of the derivative  $m'(u)$  together with its standard deviation  $\text{sd}(\hat{m}'_h(u))$ . The latter is given by  $\text{sd}(\hat{m}'_h(u)) = \{\text{Var}(\hat{m}'_h(u))\}^{1/2}$ , where  $\text{Var}(\hat{m}'_h(u)) = e^\top V e$  with  $e = (0 \ 1)^\top$  and

$$V = (X^\top W X)^{-1} (X^\top \Sigma X) (X^\top W X)^{-1}.$$

The matrices  $X$ ,  $W$  and  $\Sigma$  are defined as follows:  $\Sigma$  is a  $T \times T$  matrix with the elements

$$\Sigma_{st} = \gamma_\varepsilon(s - t) K_h\left(\frac{s}{T} - u\right) K_h\left(\frac{t}{T} - u\right),$$

$W$  is a  $T \times T$  diagonal matrix with the diagonal entries  $K_h(t/T - u)$  and

$$X = \begin{pmatrix} 1 & (1/T - u) \\ 1 & (2/T - u) \\ \vdots & \vdots \\ 1 & (1 - u) \end{pmatrix}.$$

(c) Computation of the confidence intervals:

For a given confidence level  $\alpha$  and for each bandwidth value  $h$  with  $(u, h) \in \mathcal{G}_T^*$ , we compute the Gaussian quantile

$$q(h) = \Phi^{-1}\left(\left(1 - \frac{\alpha}{2}\right)^{1/(\theta g)}\right),$$

where  $\Phi$  is the distribution function of a standard normal random variable,  $g$  is the number of locations  $u$  in the grid  $\mathcal{G}_T^*$ , and the cluster index  $\theta$  is defined on p.1519 in Park et al. (2009). The confidence interval of  $\hat{m}'_h(u)$  is then computed as  $[\hat{m}'_h(u) - q(h) \text{sd}(\hat{m}'_h(u)), \hat{m}'_h(u) + q(h) \text{sd}(\hat{m}'_h(u))]$ .

## References

- BERKES, I., LIU, W. and WU, W. B. (2014). Komlós-Major-Tusnády approximation under dependence. *Annals of Probability*, **42** 794–817.
- CHERNOZHUKOV, V., CHETVERIKOV, D. and KATO, K. (2015). Comparison and anti-concentration bounds for maxima of Gaussian random vectors. *Probability Theory and Related Fields*, **162** 47–70.
- HALL, P. and VAN KEILEGOM, I. (2003). Using difference-based methods for inference in nonparametric regression with time series errors. *Journal of the Royal Statistical Society: Series B*, **65** 443–456.
- LEDoux, M. (2001). *Concentration of Measure Phenomenon*. American Mathematical Society.
- PARK, C., HANNIG, J. and KANG, K.-H. (2009). Improved SiZer for time series. *Statistica Sinica*, **19** 1511–1530.