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

September 21, 2020

In this supplement, we give the technical details and proofs that are omitted in the paper. In addition, we provide some material which complements 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 (3.5) 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 $T = 1, 2, \ldots$ with the following two properties: (i) $\widetilde{\Phi}_T$ has the same distribution as $\widehat{\Phi}_T$ for any T, and (ii)*

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

where Φ_T is a Gaussian statistic as defined in (3.4).

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 $\{\widetilde{\varepsilon}_t : t \in \mathbb{N}\}$ such that $[\widetilde{\varepsilon}_1, \dots, \widetilde{\varepsilon}_T] \stackrel{\mathcal{D}}{=} [\varepsilon_1, \dots, \varepsilon_T]$ for each T and

$$\max_{1 \le t \le T} \Big| \sum_{s=1}^{t} \widetilde{\varepsilon}_s - \sigma \mathbb{B}(t) \Big| = o\big(T^{1/q}\big) \quad \text{a.s.}, \tag{S.1}$$

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

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

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

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

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

$$\big| \widetilde{\Phi}_T - \Phi_T \big| \le \big| \widetilde{\Phi}_T - \Phi_T^{\diamond} \big| + \big| \Phi_T^{\diamond} - \Phi_T \big| = \big| \widetilde{\Phi}_T - \Phi_T^{\diamond} \big| + o_p\big(\rho_T \sqrt{\log T}\big), \tag{S.2}$$

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 θ and $\widetilde{\sigma}^2 = \sigma^2 + o_p(\rho_T)$. Straightforward calculations yield that

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

Using summation by parts, we further obtain that

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

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

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

Plugging (S.3) into (S.2) completes the proof. □

Auxiliary results using anti-concentration bounds

In this section, we establish some properties of the Gaussian statistic Φ_T defined in (3.4). We in particular show that Φ_T does not concentrate too strongly in small regions of the form $[x - \delta_T, x + \delta_T]$ with δ_T converging to zero.

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

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

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

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

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

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 $\mathcal{G}_T = \{x : x \in \mathcal{G}_T\} = \{x_1,\ldots,x_p\}$, where $p := |\mathcal{G}_T| \le O(T^\theta)$ for some large but fixed $\theta > 0$ by our assumptions. Moreover, for $j = 1,\ldots,p$, 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

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

where $(X_1, \ldots, X_{2p})^\top$ is a Gaussian random vector with the following properties: (i) $\mu_j := \mathbb{E}[X_j] = -\lambda(x_{j2})$ and thus $\overline{\mu} = \max_{1 \le j \le 2p} |\mu_j| \le C\sqrt{\log T}$, 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 $a_{2p} = b_{2p} \le \sqrt{2\log(2p)} \le C\sqrt{\log T}$. With this notation at hand, we can apply Proposition S.3 to obtain that

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

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 $\widetilde{\Phi}_T$ for $T = 1, 2, \ldots$ which are distributed as $\widehat{\Phi}_T$ for any $T \ge 1$ and which have the property that

$$\big|\widetilde{\Phi}_T - \Phi_T\big| = o_p\Big(\frac{T^{1/q}}{\sqrt{Th_{\min}}} + \rho_T\sqrt{\log T}\Big), \tag{S.4}$$

where Φ_T is a Gaussian statistic as defined in (3.4). The approximation result (S.4) allows us to replace the multiscale statistic $\widehat{\Phi}_T$ by an identically distributed version $\widetilde{\Phi}_T$ which is close to the Gaussian statistic Φ_T. In the next step, we show that

$$\sup_{x \in \mathbb{R}} \big|\mathbb{P}(\widetilde{\Phi}_T \le x) - \mathbb{P}(\Phi_T \le x)\big| = o(1), \tag{S.5}$$

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 $T = 1, 2, \ldots$ such that $V_T - W_T = o_p(\delta_T)$ with some $\delta_T = o(1)$. If*

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

then

$$\sup_{x \in \mathbb{R}} \big|\mathbb{P}(V_T \le x) - \mathbb{P}(W_T \le x)\big| = o(1). \tag{S.7}$$

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

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

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

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

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 $\widehat{\psi}_T(u,h) = \widehat{\psi}_T^A(u,h) + \widehat{\psi}_T^B(u,h)$ with $\widehat{\psi}_T^A(u,h) = \sum_{t=1}^T w_{t,T}(u,h)\varepsilon_t$ and $\widehat{\psi}_T^B(u,h) = \sum_{t=1}^T w_{t,T}(u,h)m_T(\frac{t}{T})$. 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,

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

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

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

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

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

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

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

$$\widehat{\psi}_T^B(u_0,h_0) = \sum_{t=1}^T w_{t,T}(u_0,h_0)\Big(\frac{\frac{t}{T} - u_0}{h_0}\Big) h_0 m_T'(\xi_{u_0,t,T}). \tag{S.12}$$

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

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

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

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

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 ℓ,

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

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 $[u-h, u+h] \subseteq [0,1]$,

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

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

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

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

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

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),

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

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

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

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

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

where the last equality holds by Theorem 3.1.

Proof of Corollary 3.1

Let $\alpha = \alpha_T \to 0$ and let $\widetilde{\Phi}_T$ be defined as in the proof of Theorem 3.1. It holds that

$$\begin{aligned}\big|\mathbb{P}(\widetilde{\Phi}_T \le q_T(\alpha_T)) - (1-\alpha_T)\big| &= \big|\mathbb{P}(\widetilde{\Phi}_T \le q_T(\alpha_T)) - \mathbb{P}(\Phi_T \le q_T(\alpha_T))\big| \\ &\le \sup_{x\in\mathbb{R}} \big|\mathbb{P}(\widetilde{\Phi}_T \le x) - \mathbb{P}(\Phi_T \le x)\big| = o(1),\end{aligned}$$

where the last equality is due to (S.5). From this, it immediately follows that $\mathbb{P}(\widetilde{\Phi}_T \le q_T(\alpha_T)) \to 1$. Moreover, since $\widetilde{\Phi}_T$ and $\widehat{\Phi}_T$ have the same distribution by construction, we obtain that

$$\mathbb{P}(\widehat{\Phi}_T \le q_T(\alpha_T)) \to 1. \tag{S.18}$$

Taking into account (S.18), Corollary 3.1 can be proven in exactly the same way as Proposition 3.3.

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 $\mathrm{Var}(W_j)=1$ for all $1\le j\le p$. Suppose that $\mathrm{Corr}(W_j,W_k)<1$ whenever $j\ne k$. Then the distribution of $\max_{1\le j\le p} W_j$ is absolutely continuous with respect to Lebesgue measure and a version of the density is given by*

$$f(x) = f_0(x) \sum_{j=1}^p e^{\mathbb{E}[W_j]x - \mathbb{E}[W_j]^2/2}\, \mathbb{P}\big(W_k \le x \text{ for all } k \ne j \,\big|\, W_j = x\big).$$

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

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

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 $\max_{1\le j\le p}\mathbb{E}[X_j^2]\le\sigma_X^2$ for some $\sigma_X^2>0$. Then for any $r>0$,*

$$\mathbb{P}\Big(\max_{1\le j\le p} X_j \ge \mathbb{E}\Big[\max_{1\le j\le p} X_j\Big] + r\Big) \le e^{-r^2/(2\sigma_X^2)}.$$

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 **?**; 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

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

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

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

Step 2. We now bound the density of Z. Without loss of generality, we assume that $\mathrm{Corr}(W_j, W_k) < 1$ for $k \neq j$. The marginal distribution of W_j is $N(\nu_j, 1)$ with $\nu_j = \mathbb{E}[W_j] = (\mu_j/\sigma_j + \overline{\mu}/\underline{\sigma}) + (x/\underline{\sigma} - x/\sigma_j) \geq 0$. Hence, by Lemmas S.2 and S.3, the random variable Z has a density of the form

$$f_p(z) = f_0(z) G_p(z), \tag{S.19}$$

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

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

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

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

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

$$G_p(z) \leq \frac{1}{1 - F_0(z)} \exp\Big(-\frac{(z - \overline{z} - a_p)_+^2}{2}\Big). \tag{S.20}$$

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

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

This together with (S.19) and (S.20) yields that

$$f_p(z) \le 2(z \vee 1) \exp\Big(-\frac{(z-\overline{z}-a_p)_+^2}{2}\Big) \quad \text{for any } z \in \mathbb{R}.$$

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

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

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 $x \ge 0$ and $\delta > 0$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le 4\delta\Big\{\frac{2\overline{\mu}}{\underline{\sigma}} + |x|\Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big) + a_p + 1\Big\}/\underline{\sigma}. \tag{S.21}$$

This inequality also holds for $x < 0$ by an analogous argument, and hence for all $x \in \mathbb{R}$. Now let $0 < \delta \le \underline{\sigma}$ and define $b_p = \mathbb{E}\max_{1\le j\le p}\{X_j - \mu_j\}$. For any $|x| \le \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$, (S.21) yields that

$$\begin{aligned}\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) &\le \frac{4\delta}{\underline{\sigma}}\Big\{\overline{\mu}\Big(\frac{3}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big) + a_p + \Big(\frac{1}{\underline{\sigma}} - \frac{1}{\overline{\sigma}}\Big)b_p \\ &\qquad + \Big(\frac{\overline{\sigma}}{\underline{\sigma}} - 1\Big)\sqrt{2\log\Big(\frac{\underline{\sigma}}{\delta}\Big)} + 2 - \frac{\underline{\sigma}}{\overline{\sigma}}\Big\} \\ &\le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\}\end{aligned} \tag{S.22}$$

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

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \frac{\delta}{\underline{\sigma}}, \tag{S.23}$$

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

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j - \mu_j\} \ge |x| - \delta - \overline{\mu}\Big). \tag{S.24}$$

If $x < -(\delta + \overline{\mu})$, then $|\max_j X_j - x| \le \delta$ implies that $\max_j\{X_j - \mu_j\} \le -|x| + \delta + \overline{\mu}$.

Hence, in this case,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j-\mu_j\} \le -|x| + \delta + \overline{\mu}\Big)$$
$$\le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j-\mu_j\} \ge |x| - \delta - \overline{\mu}\Big), \tag{S.25}$$

where the last inequality follows from the fact that for centred Gaussian random variables V_j and $v > 0$, $\mathbb{P}(\max_j V_j \le -v) \le \mathbb{P}(V_1 \le -v) = P(V_1 \ge v) \le \mathbb{P}(\max_j V_j \ge v)$. With (S.24) and (S.25), we obtain that for any $|x| \ge \delta + \overline{\mu} + b_p + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j-\mu_j\} \ge |x| - \delta - \overline{\mu}\Big)$$
$$\le \mathbb{P}\Big(\max_{1\le j\le p}\{X_j-\mu_j\} \ge \mathbb{E}\Big[\max_{1\le j\le p}\{X_j-\mu_j\}\Big] + \overline{\sigma}\sqrt{2\log(\underline{\sigma}/\delta)}\Big) \le \frac{\delta}{\underline{\sigma}},$$

the last inequality following from Lemma S.4. To sum up, we have established that for any $0 < \delta \le \underline{\sigma}$ and any $x \in \mathbb{R}$,

$$\mathbb{P}\Big(\Big|\max_{1\le j\le p} X_j - x\Big| \le \delta\Big) \le C\delta\big\{\overline{\mu} + a_p + b_p + \sqrt{1 \vee \log(\underline{\sigma}/\delta)}\big\} \tag{S.26}$$

with some constant $C > 0$ that does only depend on $\underline{\sigma}$ and $\overline{\sigma}$. For $\delta > \underline{\sigma}$, (S.26) trivially follows upon setting $C \ge 1/\underline{\sigma}$. This completes the proof.

S.2 Proofs of the results from Section 4

In what follows, we prove Proposition 4.1 from Section 4. Throughout the section, we assume that m is Lipschitz and that $\{\varepsilon_t\}$ is an AR(p^*) process of the form (4.3) with the following properties: $A(z) \ne 0$ for all $|z| \le 1+\delta$ with some small $\delta > 0$ and the innovations η_t have a finite fourth moment. As in the previous section, the symbol C denotes a generic real constant which may take a different value on each occurrence. Moreover, the notation $v_T \ll w_T$ means that $v_T/w_T \to 0$ as $T \to \infty$ and the symbol $\|\cdot\|_2$ denotes the usual Euclidean/spectral norm for vectors/matrices.

Auxiliary results

To start with, we derive some auxiliary results on the sample autocovariances

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

The first result bounds the L_2-distance between the ℓ-th sample autocovariance $\widehat{\gamma}_q(\ell)$ and its true counterpart $\gamma_q(\ell) = \mathrm{Cov}(\Delta_q\varepsilon_t, \Delta_q\varepsilon_{t-\ell})$ uniformly over ℓ.

Lemma S.5. *Let $1 \leq p \ll \sqrt{T}$ and $1 \leq q \ll \sqrt{T}$. For any $1 \leq \ell \leq p$,*

$$\mathbb{E}\Big[\big(\widehat{\gamma}_q(\ell) - \gamma_q(\ell)\big)^2\Big] \leq C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\},$$

where the constant C is independent of ℓ, p, q and T.

Proof of Lemma S.5. The expression $\widehat{\gamma}_q(\ell) - \gamma_q(\ell)$ can be decomposed as

$$\widehat{\gamma}_q(\ell) - \gamma_q(\ell) = \big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big) + R_{q,A}(\ell) + R_{q,B}(\ell) + R_{q,C}(\ell),$$

where

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

and

$$\begin{aligned} R_{q,A}(\ell) &= \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q m_t \Delta_q \varepsilon_{t-\ell} \\ R_{q,B}(\ell) &= \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q \varepsilon_t \Delta_q m_{t-\ell} \\ R_{q,C}(\ell) &= \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \Delta_q m_t \Delta_q m_{t-\ell} \end{aligned}$$

with $\Delta_q m_t = m(\frac{t}{T}) - m(\frac{t-q}{T})$. In what follows, we prove that for any $1 \leq \ell \leq p$,

$$\mathbb{E}\Big[\big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big)^2\Big] \leq C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2\Big\}, \tag{S.27}$$

where the constant C does not depend on ℓ, p, q and T. Applying the Cauchy-Schwarz inequality and exploiting the Lipschitz continuity of m, we further obtain that

$$\mathbb{E}R_{q,A}^2(\ell) \leq \Big(\frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \{\Delta_q m_t\}^2\Big)\Big(\frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \{\Delta_q \varepsilon_{t-\ell}\}^2\Big) \leq C\Big(\frac{q}{T}\Big)^2,$$

where C is independent of ℓ, p, q and T. Analogously, we get that $\mathbb{E}R_{q,k}^2(\ell) \leq C(q/T)^2$ for $k = B, C$. From this and (S.27), it immediately follows that

$$\begin{aligned} \mathbb{E}\Big[\big(\widehat{\gamma}_q(\ell) - \gamma_q(\ell)\big)^2\Big] &\leq 16\Big\{\mathbb{E}\Big[\big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big)^2\Big] + \mathbb{E}R_{q,A}^2(\ell) + \mathbb{E}R_{q,B}^2(\ell) + \mathbb{E}R_{q,C}^2(\ell)\Big\} \\ &\leq C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\}, \end{aligned}$$

which completes the proof.

It remains to verify (S.27). To do so, we decompose $\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)$ as

$$\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell) = \Sigma_{q,1}(\ell) - \Sigma_{q,2}(\ell) - \Sigma_{q,3}(\ell) + \Sigma_{q,4}(\ell) - \Big(1 - \frac{T-q-\ell}{T-q}\Big)\gamma_q(\ell)$$

with

$$\Sigma_{q,1}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \big\{\varepsilon_t \varepsilon_{t-\ell} - \mathbb{E}\varepsilon_t \varepsilon_{t-\ell}\big\}$$
$$\Sigma_{q,2}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \big\{\varepsilon_t \varepsilon_{t-\ell-q} - \mathbb{E}\varepsilon_t \varepsilon_{t-\ell-q}\big\}$$
$$\Sigma_{q,3}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \big\{\varepsilon_{t-q} \varepsilon_{t-\ell} - \mathbb{E}\varepsilon_{t-q} \varepsilon_{t-\ell}\big\}$$
$$\Sigma_{q,4}(\ell) = \frac{1}{T-q} \sum_{t=q+\ell+1}^{T} \big\{\varepsilon_{t-q} \varepsilon_{t-\ell-q} - \mathbb{E}\varepsilon_{t-q} \varepsilon_{t-\ell-q}\big\}$$

and prove that

$$\mathbb{E}\Sigma_{q,k}^2(\ell) \le \frac{C}{T-q} \tag{S.28}$$

for $1 \le k \le 4$, where the constant C only depends on the coefficients $c_0, c_1, c_2, \dots$ of the MA(∞) representation of $\{\varepsilon_t\}$ and the innovation variance ν^2. From this, we get that

$$\begin{aligned}
\mathbb{E}\Big[\big(\widehat{\gamma}_q^*(\ell) - \gamma_q(\ell)\big)^2\Big] &\le 25\Big\{\sum_{k=1}^{4} \mathbb{E}\Sigma_{q,k}^2(\ell) + \Big(1 - \frac{T-q-\ell}{T-q}\Big)^2 \gamma_q^2(\ell)\Big\} \\
&\le C\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2\Big\}
\end{aligned}$$

with C independent of ℓ, p, q and T, which shows (S.27).
We now turn to the proof of (S.28). As the proof is completely analogous for $k \in \{1, \dots, 4\}$, we restrict attention to $k = 1$. Since the variables ε_t have the MA(∞) expansion $\varepsilon_t = \sum_{k=0}^{\infty} c_k \eta_{t-k}$ and the autocovariances $\gamma_\varepsilon(\ell) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-\ell})$ can be written as $\gamma_\varepsilon(\ell) = (\sum_{k=0}^{\infty} c_k c_{k+\ell})\nu^2$, it holds that

$$\begin{aligned}
&\mathbb{E}\big[\varepsilon_t \varepsilon_{t-\ell} \varepsilon_{t'} \varepsilon_{t'-\ell}\big] \\
&= \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\, c_{k+\ell}\, c_{k+\ell+t'-t}\Big)\kappa + \Big(\sum_{k=0}^{\infty} c_k\, c_{k+\ell}\Big)^2 \nu^4 \\
&\quad + \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\Big)^2 \nu^4 + \Big(\sum_{k=0}^{\infty} c_k\, c_{k+\ell+t'-t}\Big)\Big(\sum_{k=0}^{\infty} c_k\, c_{k+\ell+t-t'}\Big)\nu^4 \\
&= \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\, c_{k+\ell}\, c_{k+\ell+t'-t}\Big)\kappa + \gamma_\varepsilon^2(\ell) + \gamma_\varepsilon^2(t'-t) + \gamma_\varepsilon(t'-t+\ell)\gamma_\varepsilon(t'-t-\ell)
\end{aligned}$$

with $\kappa = \mathbb{E}[\eta_0^4] - 3\nu^4$ and $c_k = 0$ for $k < 0$. From this, we obtain that $\mathbb{E}\Sigma_{q,1}^2(\ell) = \sum_{k=1}^3 \mathbb{E}\Sigma_{q,1,k}^2(\ell)$ with

$$\begin{aligned}
\mathbb{E}\Sigma_{q,1,1}^2(\ell) &= \frac{\kappa}{(T-q)^2} \sum_{t,t'=q+\ell+1}^{T} \Big(\sum_{k=0}^{\infty} c_k\, c_{k+t'-t}\, c_{k+\ell}\, c_{k+\ell+t'-t} \Big) \\
\mathbb{E}\Sigma_{q,1,2}^2(\ell) &= \frac{1}{(T-q)^2} \sum_{t,t'=q+\ell+1}^{T} \gamma_\varepsilon^2(t'-t) \\
\mathbb{E}\Sigma_{q,1,3}^2(\ell) &= \frac{1}{(T-q)^2} \sum_{t,t'=q+\ell+1}^{T} \gamma_\varepsilon(t'-t+\ell)\gamma_\varepsilon(t'-t-\ell).
\end{aligned}$$

Let $\#\{t'-t=r\}$ be the number of pairs (t,t') with $q+1 \le t,t' \le T$ such that $t'-t=r$ and note that $\#\{t'-t=r\} \le T-q$ for any r. It holds that

$$\begin{aligned}
|\mathbb{E}\Sigma_{q,1,1}^2(\ell)| &\le \frac{\kappa}{(T-q)^2} \sum_{r=-T}^{T} \#\big\{t'-t=r\big\} \sum_{k=0}^{\infty} |c_k\, c_{k+r}\, c_{k+\ell}\, c_{k+\ell+r}| \\
&\le \frac{\kappa}{T-q} \sum_{k=0}^{\infty} |c_k\, c_{k+\ell}| \sum_{r=-\infty}^{\infty} |c_{k+r}\, c_{k+\ell+r}| \\
&\le \frac{\kappa\{\max_j |c_j|\}^2 \{\sum_{k=0}^{\infty} |c_k|\}^2}{T-q} \le \frac{C}{T-q},
\end{aligned} \tag{S.29}$$

where C only depends on the parameters $c_0, c_1, c_2, \dots$ of the MA(∞) representation of $\{\varepsilon_t\}$. Moreover,

$$\mathbb{E}\Sigma_{q,1,2}^2(\ell) \le \frac{1}{(T-q)^2} \sum_{r=-T}^{T} \#\big\{t'-t=r\big\}\gamma_\varepsilon^2(r) \le \frac{\{\sum_{r=-\infty}^{\infty} \gamma_\varepsilon^2(r)\}}{T-q} \le \frac{C}{T-q} \tag{S.30}$$

and analogously $|\mathbb{E}\Sigma_{q,1,3}^2(\ell)| \le C/(T-q)$, where C only depends on the MA parameters $c_0, c_1, c_2, \dots$ and the innovation variance ν^2 (noting that $\gamma_\varepsilon(k) = \sum_{j=0}^{\infty} c_j c_{j+k} \nu^2$). With (S.29) and (S.30), we immediately arrive at (S.28). $\square$

With the help of Lemma S.5, we can derive the following bounds on $\mathbb{E}\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2$ and $\mathbb{E}\|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2$, where $\widehat{\boldsymbol{\gamma}}_q = (\widehat{\gamma}_q(1), \dots, \widehat{\gamma}_q(p))^\top$ and $\boldsymbol{\gamma}_q = (\gamma_q(1), \dots, \gamma_q(p))^\top$ as well as $\widehat{\boldsymbol{\Gamma}}_q = (\widehat{\gamma}_q(i-j) : 1 \le i,j \le p)$ and $\boldsymbol{\Gamma}_q = (\gamma_q(i-j) : 1 \le i,j \le p)$.

Lemma S.6. *Let $1 \le p \ll \sqrt{T}$ and $1 \le q \ll \sqrt{T}$. It holds that*

$$\begin{aligned}
\mathbb{E}\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2 &\le C\sqrt{p}\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\}^{1/2} \\
\mathbb{E}\|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2 &\le Cp\Big\{\frac{1}{T-q} + \Big(\frac{p}{T-q}\Big)^2 + \Big(\frac{q}{T}\Big)^2\Big\}^{1/2}
\end{aligned}$$

with some constant C independent of p, q and T.

Proof of Lemma S.6. The first statement immediately follows from Lemma S.5. The second statement is obtained by using Lemma S.5 and the bound

$$\|\widehat{\mathbf{\Gamma}}_q - \mathbf{\Gamma}_q\|_2 \leq \max_{1 \leq i \leq p} \Big(\sum_{j=1}^{p} \big|\widehat{\gamma}_q(i-j) - \gamma_q(i-j)\big| \Big) \leq \sum_{\ell=-p}^{p} \big|\widehat{\gamma}_q(\ell) - \gamma_q(\ell)\big|,$$

which follows from Gershgorin's theorem. □

The final auxiliary result summarizes some properties of the inverse autocovariance matrices $\widehat{\mathbf{\Gamma}}_q^{-1}$ and $\mathbf{\Gamma}_q^{-1}$.

Lemma S.7. *Let $p \to \infty$ and $(1+\delta)p \leq q \ll \sqrt{T}$ for some small $\delta > 0$. Then*

(i) $\|\mathbf{\Gamma}_q^{-1}\|_2 \leq C$ *for sufficiently large T with C independent of p, q and T*

(ii) $\|\widehat{\mathbf{\Gamma}}_q^{-1} - \mathbf{\Gamma}_q^{-1}\|_2 = O_p(p/\sqrt{T})$

(iii) $\|\widehat{\mathbf{\Gamma}}_q^{-1}\|_2 = O_p(1)$.

Proof of Lemma S.7. We first prove (i). As $\gamma_q(\ell) = 2\gamma_\varepsilon(\ell) - \gamma_\varepsilon(q-\ell) - \gamma_\varepsilon(q+\ell)$ with $\gamma_\varepsilon(\ell) = \mathrm{Cov}(\varepsilon_t, \varepsilon_{t-\ell})$, it holds that

$$\mathbf{\Gamma}_q = 2\mathbf{\Gamma} - \boldsymbol{R},$$

where $\mathbf{\Gamma} = (\gamma_\varepsilon(i-j) : 1 \leq i,j \leq p)$ and $\boldsymbol{R} = (\gamma_\varepsilon(q+i-j) + \gamma_\varepsilon(q-i+j) : 1 \leq i,j \leq p)$. Since the spectral density f_ε of the AR(p^*) process $\{\varepsilon_t\}$ is bounded away from zero and infinity, we can use Proposition 4.5.3 in **?** to obtain that the eigenvalues of the autocovariance matrix $\mathbf{\Gamma}$ lie in some interval $[c_\Gamma, C_\Gamma]$ with constants $0 < c_\Gamma \leq C_\Gamma < \infty$ that are independent of p. From this, we can infer that

$$\|\mathbf{\Gamma}^{-1}\|_2 \leq C \text{ with some constant } C \text{ independent of } p. \tag{S.31}$$

Moreover, it holds that

$$\|\mathbf{\Gamma}_q/2 - \mathbf{\Gamma}\|_2 = \|\boldsymbol{R}/2\|_2 \leq \sum_{\ell=-p}^{p} |\gamma_\varepsilon(q-\ell) + \gamma_\varepsilon(q+\ell)|/2 \leq C p \xi^{q-p} = o(1), \tag{S.32}$$

where we have used Gershgorin's theorem and the fact that $|\gamma_\varepsilon(\ell)| \leq C\xi^\ell$ with some $\xi \in (0,1)$. We now make use of the inequality

$$\|\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1}\|_2 \leq \frac{\|\boldsymbol{B}^{-1}\|_2^2 \, \|\boldsymbol{B} - \boldsymbol{A}\|_2}{1 - \|\boldsymbol{B} - \boldsymbol{A}\|_2 \, \|\boldsymbol{B}^{-1}\|_2}, \tag{S.33}$$

which holds for general invertible matrices $\boldsymbol{A}$ and $\boldsymbol{B}$ and which is a simple consequence of the fact that $\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} = (\boldsymbol{A}^{-1} - \boldsymbol{B}^{-1} + \boldsymbol{B}^{-1})(\boldsymbol{B} - \boldsymbol{A})\boldsymbol{B}^{-1}$. Setting $\boldsymbol{A} = \mathbf{\Gamma}_q/2$ and

$\boldsymbol{B} = \boldsymbol{\Gamma}$ in (S.33) and applying (S.31) together with (S.32), we get that

$$\|(\boldsymbol{\Gamma}_q/2)^{-1} - \boldsymbol{\Gamma}^{-1}\|_2 \leq \frac{\|\boldsymbol{\Gamma}^{-1}\|_2^2 \, \|\boldsymbol{\Gamma}_q/2 - \boldsymbol{\Gamma}\|_2}{1 - \|\boldsymbol{\Gamma}_q/2 - \boldsymbol{\Gamma}\|_2 \, \|\boldsymbol{\Gamma}^{-1}\|_2} \leq C \tag{S.34}$$

for sufficiently large T with C independent of p, q and T. Statement (i) easily follows upon combining (S.31) and (S.34).
We next turn to the proof of (ii). With (S.33), we obtain that

$$\|\widehat{\boldsymbol{\Gamma}}_q^{-1} - \boldsymbol{\Gamma}_q^{-1}\|_2 \leq \frac{\|\boldsymbol{\Gamma}_q^{-1}\|_2^2 \, \|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2}{1 - \|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2 \, \|\boldsymbol{\Gamma}_q^{-1}\|_2} = O_p\Big(\frac{p}{\sqrt{T}}\Big),$$

since $\|\boldsymbol{\Gamma}_q^{-1}\|_2 \leq C$ by (i) and $\|\widehat{\boldsymbol{\Gamma}}_q - \boldsymbol{\Gamma}_q\|_2 = O_p(p/\sqrt{T})$ by Lemma S.6. Finally, (iii) is an immediately consequence of (i) and (ii). $\square$

Proof of Proposition 4.1

We have to prove the following three statements:

$$\|\widetilde{\boldsymbol{a}}_q - \boldsymbol{a}\|_2 = O_p\Big(\sqrt{\frac{p}{T}}\Big) \tag{S.35}$$

$$\|\widehat{\boldsymbol{a}} - \boldsymbol{a}\|_2 = O_p\Big(\sqrt{\frac{p^3}{T}}\Big) \tag{S.36}$$

$$\widehat{\sigma}^2 - \sigma^2 = O_p\Big(\sqrt{\frac{p^4}{T}}\Big). \tag{S.37}$$

We carry out the proof for case (B), that is, we impose the following conditions on the parameters p, q, $\underline{r}$ and $\overline{r}$: $q \ll \sqrt{T}$, $C \log T \leq p \ll \min\{T^{1/5}, q\}$ for some sufficiently large C, $\underline{r} = (1+\delta)p$ for some small $\delta > 0$ and $\overline{r} - \underline{r}$ remains bounded as $T \to \infty$. The proof for case (A) is a simplified version of that for case (B).

Proof of (S.35). It holds that

$$\begin{aligned} \widetilde{\boldsymbol{a}}_q - \boldsymbol{a} = \widehat{\boldsymbol{\Gamma}}_q^{-1} \widehat{\boldsymbol{\gamma}}_q - \boldsymbol{a} &= \widehat{\boldsymbol{\Gamma}}_q^{-1} \big[(\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q) + (\boldsymbol{\gamma}_q - \boldsymbol{\Gamma}_q \boldsymbol{a}) + (\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\big] \\ &= \widehat{\boldsymbol{\Gamma}}_q^{-1} \big[(\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q) - \nu^2 \boldsymbol{c}_q + \boldsymbol{\rho}_q + (\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\big] \end{aligned}$$

and thus

$$\|\widetilde{\boldsymbol{a}}_q - \boldsymbol{a}\|_2 \leq \|\widehat{\boldsymbol{\Gamma}}_q^{-1}\|_2 \Big[\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2 + \nu^2 \|\boldsymbol{c}_q\|_2 + \|\boldsymbol{\rho}_q\|_2 + \|(\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\|_2\Big], \tag{S.38}$$

where we have used that $\boldsymbol{\gamma}_q - \boldsymbol{\Gamma}_q \boldsymbol{a} = -\nu^2 \boldsymbol{c}_q + \boldsymbol{\rho}_q$. We now make use of the following facts:
(i) By Lemmas S.6 and S.7, it holds that $\|\widehat{\boldsymbol{\gamma}}_q - \boldsymbol{\gamma}_q\|_2 = O_p(\sqrt{p/T})$ and $\|\widehat{\boldsymbol{\Gamma}}_q^{-1}\|_2 = O_p(1)$.
(ii) Since $\boldsymbol{c}_q = (c_{q-1}, \ldots, c_{q-p})^\top$ and $|c_j| \leq C\xi^j$ for some $\xi \in (0,1)$, it holds that $\|\boldsymbol{c}_q\|_2 \leq$

$C\sqrt{p}\xi^{q-p}$. Similarly, as $\boldsymbol{\rho}_q = (\rho_q(1), \ldots, \rho_q(p))^\top$ with $\rho_q(\ell) = \sum_{j=p+1}^{p^*} a_j \gamma_q(\ell - j)$, we have $|\rho_q(\ell)| \leq C\sum_{j=p+1}^{p^*} \xi^j \leq C\xi^p$ and thus $\|\boldsymbol{\rho}_q\|_2 \leq C\sqrt{p}\xi^p$. (iii) It holds that $\|(\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\|_2 = O_p(\sqrt{p/T})$, which can be seen as follows:

$$
\begin{aligned}
&\|(\boldsymbol{\Gamma}_q - \widehat{\boldsymbol{\Gamma}}_q)\boldsymbol{a}\|_2^2 \\
&= \sum_{i,j,j'=1}^{p} \big\{\gamma_q(i-j) - \widehat{\gamma}_q(i-j)\big\}\big\{\gamma_q(i-j') - \widehat{\gamma}_q(i-j')\big\} a_j a_{j'} \\
&\leq \sum_{j,j'=1}^{p} |a_j a_{j'}| \Big(\sum_{i=1}^{p} \big\{\gamma_q(i-j) - \widehat{\gamma}_q(i-j)\big\}^2\Big)^{1/2} \Big(\sum_{i=1}^{p} \big\{\gamma_q(i-j') - \widehat{\gamma}_q(i-j')\big\}^2\Big)^{1/2} \\
&\leq \Big(\max_{1\leq j\leq p} \sum_{i=1}^{p} \big\{\gamma_q(i-j) - \widehat{\gamma}_q(i-j)\big\}^2\Big)\Big(\sum_{j,j'=1}^{p} |a_j a_{j'}|\Big) \\
&\leq C \sum_{\ell=-p}^{p} \big\{\gamma_q(\ell) - \widehat{\gamma}_q(\ell)\big\}^2 = O_p\Big(\frac{p}{T}\Big),
\end{aligned}
$$

where the last equality follows by Lemma S.5. Plugging (i)–(iii) into (S.38), we arrive at

$$
\|\widetilde{\boldsymbol{a}}_q - \boldsymbol{a}\|_2 = O_p\Big(\sqrt{\frac{p}{T}} + \sqrt{p}\xi^{q-p} + \sqrt{p}\xi^p\Big) = O_p\Big(\sqrt{\frac{p}{T}}\Big),
$$

which completes the proof. □

Proof of (S.36). It suffices to show that $\|\widehat{\boldsymbol{a}}_r - \boldsymbol{a}\|_2 = O_p(\sqrt{p^3/T})$ for $r = (1+\delta)p$ with some small $\delta > 0$. By the same arguments as for (S.35), we obtain the bound

$$
\begin{aligned}
\|\widehat{\boldsymbol{a}}_r - \boldsymbol{a}\|_2 &\leq \|\widehat{\boldsymbol{\Gamma}}_r^{-1}\|_2 \Big[\|\widehat{\boldsymbol{\gamma}}_r - \boldsymbol{\gamma}_r\|_2 + \|\widetilde{\nu}^2\widetilde{\boldsymbol{c}}_r - \nu^2\boldsymbol{c}_r\|_2 + \|\boldsymbol{\rho}_r\|_2 + \|(\boldsymbol{\Gamma}_r - \widehat{\boldsymbol{\Gamma}}_r)\boldsymbol{a}\|_2\Big] \\
&= \|\widehat{\boldsymbol{\Gamma}}_r^{-1}\|_2 \, \|\widetilde{\nu}^2\widetilde{\boldsymbol{c}}_r - \nu^2\boldsymbol{c}_r\|_2 + O_p\Big(\sqrt{\frac{p}{T}}\Big). \qquad \text{(S.39)}
\end{aligned}
$$

Straightforward but lengthy calculations yield that $\widetilde{\nu}^2 - \nu^2 = O_p(\sqrt{p/T})$. Moreover, as proven below,

$$
\max_{1\leq j\leq p} |\widetilde{c}_j - c_j| = O_p\Big(\frac{p}{\sqrt{T}}\Big), \qquad \text{(S.40)}
$$

which implies that $\|\widetilde{\nu}^2\widetilde{\boldsymbol{c}}_r - \nu^2\boldsymbol{c}_r\|_2 = O_p(\sqrt{p^3/T})$. Plugging this into (S.39) completes the proof of (S.36).

It remains to verify (S.40). By assumption, $A(z) \neq 0$ for all complex $|z| \leq 1+\delta$ with some small $\delta > 0$. Hence, $A^{-1}(z)$ has a (unique) series expansion of the form $A^{-1}(z) = \sum_{j=0}^{\infty} c_j z^j$ for all $|z| < 1+\delta$, where the coefficients c_j satisfy the integral representation

$$
c_j = \frac{1}{2\pi i} \oint_{|w|=1} \frac{A^{-1}(w)}{w^{j+1}} dw \qquad \text{for all } j \geq 0. \qquad \text{(S.41)}
$$

Here and in what follows, the symbol $\oint_{|w|=1}$ denotes the line integral along the unit circle.

An analogous integral representation can be derived for the estimates $\widetilde{c}_j$: By classical time series results, $\widetilde{A}(z) = 1 - \sum_{j=1}^{p} \widetilde{a}_j z^j \neq 0$ for all complex $|z| \leq 1$ whenever $\widehat{\gamma}_q(0) > 0$. Since $\widetilde{A}(z)$ has exactly p (and thus finitely many) roots, it even holds that $\widetilde{A}(z) \neq 0$ for all $|z| \leq 1 + \widetilde{\delta}$ with some small $\widetilde{\delta} > 0$ whenever $\widehat{\gamma}_q(0) > 0$. Note that $\widetilde{\delta} = \widetilde{\delta}(\widetilde{a}_1, \ldots, \widetilde{a}_p)$ is a function of the parameter estimates $\widetilde{a}_1, \ldots, \widetilde{a}_p$ and thus random. As $\widehat{\gamma}_q(0) > 0$ with probability approaching 1, we can infer that $\widetilde{A}(z) \neq 0$ for all $|z| \leq 1 + \widetilde{\delta}$ with probability tending to 1. This in turn implies that with probability tending to 1, $\widetilde{A}^{-1}(z)$ has a (unique) series expansion of the form $\widetilde{A}^{-1}(z) = \sum_{j=0}^{\infty} \widetilde{\phi}_j z^j$ for all $|z| < 1 + \widetilde{\delta}$, where the coefficients $\widetilde{\phi}_j$ satisfy the integral representation $\widetilde{\phi}_j = (2\pi i)^{-1} \oint_{|w|=1} \widetilde{A}^{-1}(w)/w^{j+1} dw$ for all $j \geq 0$. By definition, the estimates $\widetilde{c}_j$ are the solution to the recursive equations (4.2). Importantly, this is equivalent to them being the coefficients in the series expansion of $A^{-1}(z)$. Hence, $\widetilde{c}_j = \widetilde{\phi}_j$ for all $j \geq 0$ and thus

$$\widetilde{c}_j = \frac{1}{2\pi i} \oint_{|w|=1} \frac{\widetilde{A}^{-1}(w)}{w^{j+1}} dw \qquad \text{for all } j \geq 0 \tag{S.42}$$

with probability tending to 1. Combining (S.41) and (S.42), we finally obtain that

$$\max_{1 \leq j \leq p} |\widetilde{c}_j - c_j| = \max_{1 \leq j \leq p} \left| \frac{1}{2\pi i} \oint_{|w|=1} \frac{\widetilde{A}^{-1}(w) - A^{-1}(w)}{w^{j+1}} dw \right|$$

with probability tending to 1. The right-hand side of this formula can be bounded with the help of (S.35) and straightforward arguments to arrive at (S.40). $\square$

Proof of (S.37). With the help of (S.36) and straightforward but lengthy calculations, it can be shown that $\widehat{\nu}^2 = \nu^2 + O_p(\sqrt{p^3/T})$ and $(1 - \sum_{j=1}^{p} \widehat{a}_j)^2 - (1 - \sum_{j=1}^{p^*} a_j)^2 = O_p(\sqrt{p^4/T})$. The statement (S.37) is a simple consequence of these two facts. $\square$

S.3 Supplementary material for the simulation study of Section 5

Implementation of SiZer in Section 5.1

(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

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

Since the autocovariance function $\gamma_\varepsilon(\cdot)$ of the error process $\{\varepsilon_t\}$ is assumed to be known, we can calculate the value of $\mathrm{Var}(\bar{Y})$ 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 $(u,h) \in \mathcal{G}_T$, we finally calculate the effective sample size for dependent data

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

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 estimator and its standard deviation:

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

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

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

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

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 critical values:

For a given significance level α and for each bandwidth h with $(u,h) \in \mathcal{G}_T^*$, we compute the term

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

where Φ is the distribution function of a standard normal random variable, g is the number of locations u with $(u,h) \in \mathcal{G}_T^*$, and the cluster index θ is defined on p.1519 in Park et al. (2009). The SiZer test is carried out as follows: For each $(u,h) \in \mathcal{G}_T^*$, the test rejects the null hypothesis $H_0(u,h)$ that m is constant on $[u-h, u+h]$ if $|\widehat{m}_h'(u)/\text{sd}(\widehat{m}_h'(u))| > q(h)$. More specifically, the test indicates an increase in m on the interval $[u-h, u+h]$ if $\widehat{m}_h'(u)/\text{sd}(\widehat{m}_h'(u)) > q(h)$ and a decrease if $-\widehat{m}_h'(u)/\text{sd}(\widehat{m}_h'(u)) > q(h)$.

Power simulations additional to Section 5.1.2

In the following simulation exercises, we compare the performance of the tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ (i) when m is the blocks signal of **?** that was investigated in detail by Hannig and Marron (2006) in the SiZer context and (ii) when m is the sine curve $m(u) = \sin(6\pi u)$ that was considered in Park et al. (2009). We define the blocks signal exactly as **?** and Hannig and Marron (2006). Specifically, we set

$$\begin{aligned} m(x) = (0.6/9.2)\big\{&4\text{ssgn}(x-0.1) - 5\text{ssgn}(x-0.13) + 3\text{ssgn}(x-0.15) \\ &- 4\text{ssgn}(x-0.23) + 5\text{ssgn}(x-0.25) - 4.2\text{ssgn}(x-0.4) \\ &+ 2.1\text{ssgn}(x-0.44) + 4.3\text{ssgn}(x-0.65) - 3.1\text{ssgn}(x-0.76) \\ &+ 2.1\text{ssgn}(x-0.78) - 4.2\text{ssgn}(x-0.81) + 2\big\} + 0.2, \end{aligned}$$

where $\text{ssgn}(x) = (1 + \text{sgn}(x))/2$ and $\text{sgn}(x)$ is the standard sign function. In both the blocks and the sine case, we model the error terms as an AR(1) process $\varepsilon_t = a_1\varepsilon_{t-1} + \eta_t$, where $a_1 \in \{-0.5, 0.5\}$ and η_t are i.i.d. normal with $\mathbb{E}[\eta_t] = 0$ and $\mathbb{E}[\eta_t^2] = \nu^2$. In the blocks example, we set $\nu^2 = (1 - a_1^2)/100$. This implies that $\text{Var}(\varepsilon_t) = (0.1)^2$, which matches the variance of the i.i.d. errors in the blocks example of Hannig and Marron (2006). In the sine example, we choose $\nu^2 = (1 - a_1^2)$, which implies that $\text{Var}(\varepsilon_t) = 1$. A plot of the blocks signal is given in the two top panels of Figure S.1. As can be seen, the signal is a piecewise constant function with several jumps. We could replace this jump function by a slightly smoothed and thus differentiable version with very steep increases and decreases. However, as this would leave the simulation results essentially unchanged, we stick to the original blocks signal.

For both the blocks and the sine example, we simulate a representative data sample of length $T = 1000$ and carry out the four tests on the simulated sample for the significance level $\alpha = 0.05$. The results are presented by SiZer maps in Figures S.1 and S.2 which are to be read as follows: Each pixel of the SiZer map corresponds to a location-scale point (u, h), or put differently, to a time interval $[u - h, u + h]$. The pixel (u, h) is coloured blue if the test indicates an increase in the trend m on the interval $[u - h, u + h]$, red if the test indicates a decrease and purple if the test does not reject the null hypothesis that m is constant on $[u - h, u + h]$. Moreover, a pixel (u, h) is coloured grey if the effective sample size $\text{ESS}^*(u, h)$ is smaller than 5, in which case the pixel (u, h) is not included in the location-scale grid $\mathcal{G}_T^*$.

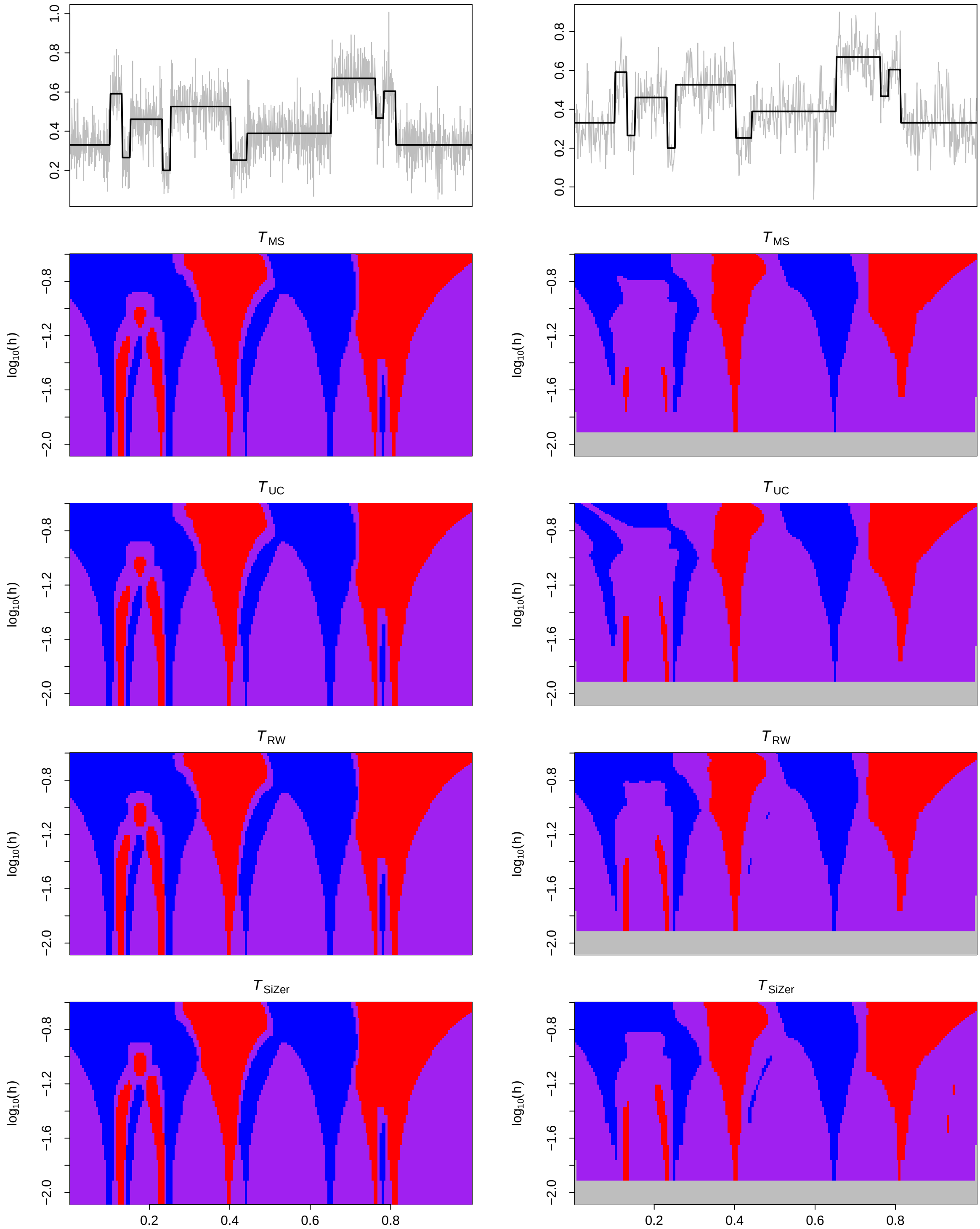
The results for the blocks example are reported in Figure S.1, the left-hand panels of subfigure (a) corresponding to the case with $a_1 = -0.5$ and the right-hand panels of subfigure (b) to the case with $a_1 = 0.5$. Let us first have a closer look at subfigure (a). The top panel depicts the blocks signal with the simulated data sample in the background. The other panels show the SiZer maps produced by the four tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$. As can be seen, the SiZer maps are fairly similar. In particular, all four tests pick up the increases and decreases (that is, the upward and downward jumps) in the signal m quite accurately. The situation is a bit different in subfigure (b), that is, in the case with $a_1 = 0.5$. Overall, the colour patterns in the four SiZer maps look fairly similar. However, on closer inspection, the following differences become apparent:

(i) In the SiZer maps of the two row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$, there is a small stripe of blue pixels around the jump location $u = 0.44$. Hence, the row-wise methods detect the small upward jump in the blocks signal at $u = 0.44$, whereas the global methods do not pick up this jump.

(ii) In the SiZer map of $\mathcal{T}_{\text{SiZer}}$, there are two small stripes of red pixels near the location $u = 0.95$ corresponding to scales h with $\log_{10}(h)$ between -1.2 and -1.6. Hence, the row-wise SiZer test $\mathcal{T}_{\text{SiZer}}$ spuriously finds a decrease in the trend m on a short time interval around $u = 0.95$. As a specific example, the pixel $(u, h) = (0.93, 0.03)$ is coloured red, implying that $\mathcal{T}_{\text{SiZer}}$ spuriously finds a decrease on the interval $[0.90, 0.96]$. Inspecting the grey time series plot in the top panel of subfigure (b), it indeed looks as if there is a short downward trend in the time series towards the end of the sample. However, this downward movement of the time series is not due to an actual decrease in the trend function m. It is rather produced by the autocorrelation structure in the error terms.

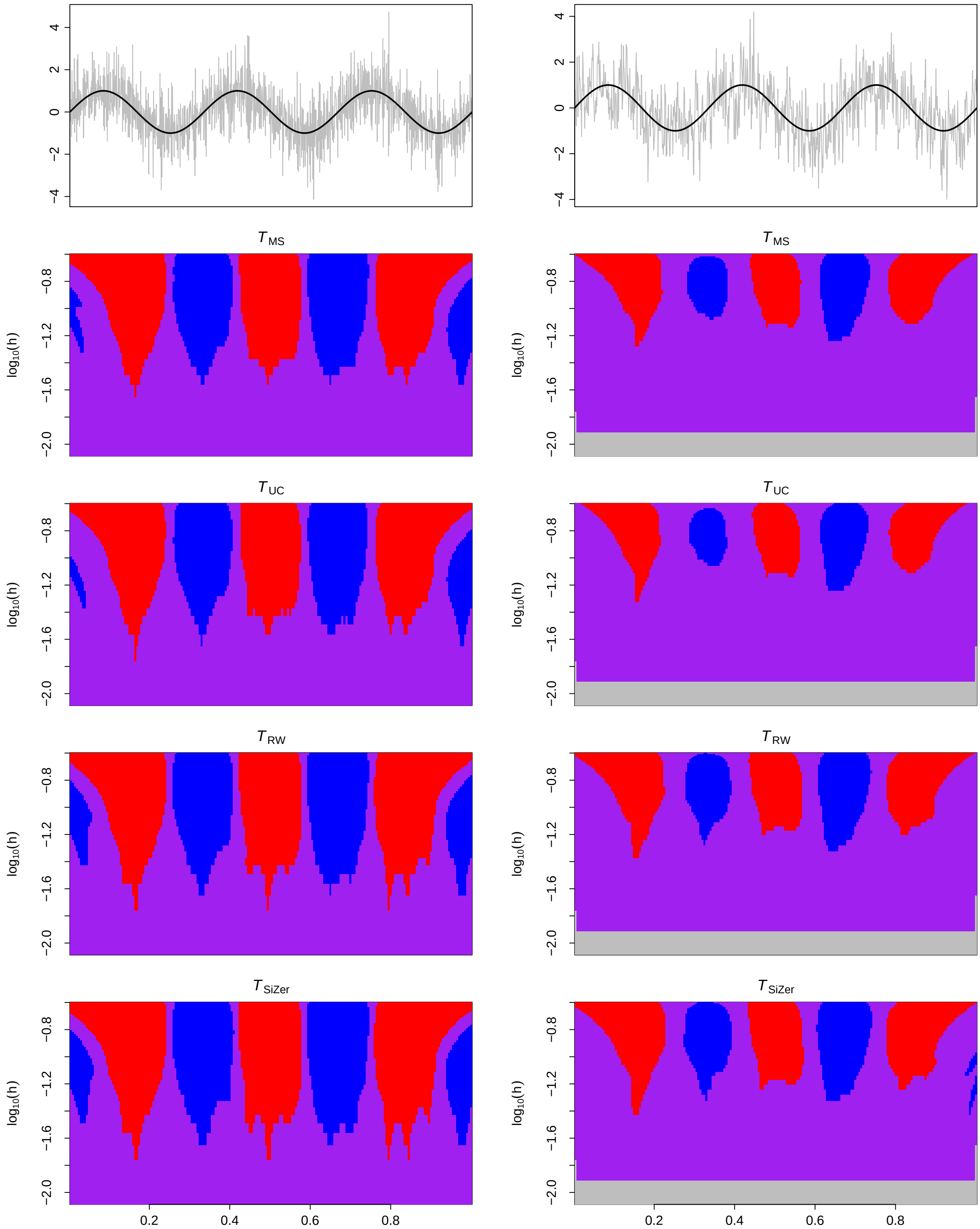
(i) indicates that the row-wise methods $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$ tend to be more powerful than the global tests $\mathcal{T}_{\text{MS}}$ and $\mathcal{T}_{\text{UC}}$. (ii) shows that this gain of power comes at a cost: The row-wise methods tend to find spurious increases/decreases more often than the global ones. Hence, the SiZer maps of subfigure (b) nicely illustrate the main findings of our power simulations in Section 5.1.2.

The SiZer maps for the sine example are depicted in Figure S.2. Overall, they convey a picture very similar to the SiZer maps of the blocks example. The SiZer maps for the case with $a_1 = -0.5$ show that the increases and decreases of the sine curve m are picked up appropriately by all four tests. In the case with $a_1 = 0.5$, in contrast, the four tests only detect the increases and decreases of m in the interior of the support $[0, 1]$. The increase of m at the left-hand boundary of the support is not picked up by any of the tests, the increase at the right-hand boundary is only detected by row-wise SiZer $\mathcal{T}_{\text{SiZer}}$, which is indicated by the small blue area at the right-hand boundary of the SiZer map. This again illustrates that the row-wise methods tend to be more powerful than the global tests.



(a) $a = -0.5$ (b) $a = 0.5$

Figure S.1: SiZer maps for the blocks example. The left-hand panels of subfigure (a) show the results for $a_1 = -0.5$, the right-hand panels of subfigure (b) those for $a_1 = 0.5$. The two upper panels depict the trend curve m with the simulated data sample in the background. The other panels show the SiZer maps produced by the four tests $\mathcal{T}_{\text{MS}}$, $\mathcal{T}_{\text{UC}}$, $\mathcal{T}_{\text{RW}}$ and $\mathcal{T}_{\text{SiZer}}$.



(a) $a = -0.5$ (b) $a = 0.5$

Figure S.2: SiZer maps for the sine example. The left-hand panels of subfigure (a) show the results for $a_1 = -0.5$, the right-hand panels of subfigure (b) those for $a_1 = 0.5$. The two upper panels depict the sine curve with the simulated data sample in the background. The other panels show the SiZer maps produced by the four tests $\mathcal{T}_{\rm MS}$, $\mathcal{T}_{\rm UC}$, $\mathcal{T}_{\rm RW}$ and $\mathcal{T}_{\rm SiZer}$.

Robustness checks for Section 5.2

In what follows, we carry out some robustness checks to assess how sensitive the estimators $\widehat{a}$ and $\widehat{\sigma}^2$ are to the choice of the tuning parameters q and $(\underline{r}, \overline{r})$. To do so, we repeat the simulation exercises of Section 5.2 for different values of q and $\overline{r}$. The parameter $\underline{r}$, in contrast, is not varied but set to $\underline{r} = 1$ throughout. The reason is as follows: Our estimators can be expected to perform well as long as the differencing orders r with $\underline{r} \le r \le \overline{r}$ are sufficiently small. Choosing $\underline{r}$ as small as possible is thus completely unproblematic (and indeed most natural). The interesting issue is rather how strongly our estimators depend on the choice of $\overline{r}$. 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 5.2, 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 $\widehat{a}$, $\widehat{a}_{\text{HvK}}$, $\widehat{a}_{\text{oracle}}$ and $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\text{HvK}}$, $\widehat{\sigma}^2_{\text{oracle}}$ for these scenarios are presented in Figure 4 of Section 5.2. These MSEs are re-calculated in Figures S.3 and S.4 for a range of different choices of q, $\overline{r}$ and (m_1, m_2). As one can see, the MSEs in the different plots of Figures S.3 and S.4 are very similar. Hence, the MSE results reported in Section 5.2 for the scenarios with a moderate trend appear to be fairly robust to different choices of the tuning parameters. In particular, our estimators $\widehat{a}$ and $\widehat{\sigma}^2$ 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 5 of Section 5.2. Analogously as before, we re-calculate these MSEs for different tuning parameters in Figures S.5–S.7. Figure S.6 is a zoomed-in version of Figure S.5 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 $\overline{r}$ is chosen bigger. This is of course not very surprising: The main reason why the estimator $\widehat{a}$ works well in the presence of a strong trend is that it is only based on differences of small orders. If we increase $\overline{r}$, we use larger differences to compute $\widehat{a}$, which results in not eliminating the trend m appropriately any more. This becomes visible in somewhat larger MSE values. Nevertheless, overall, 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.

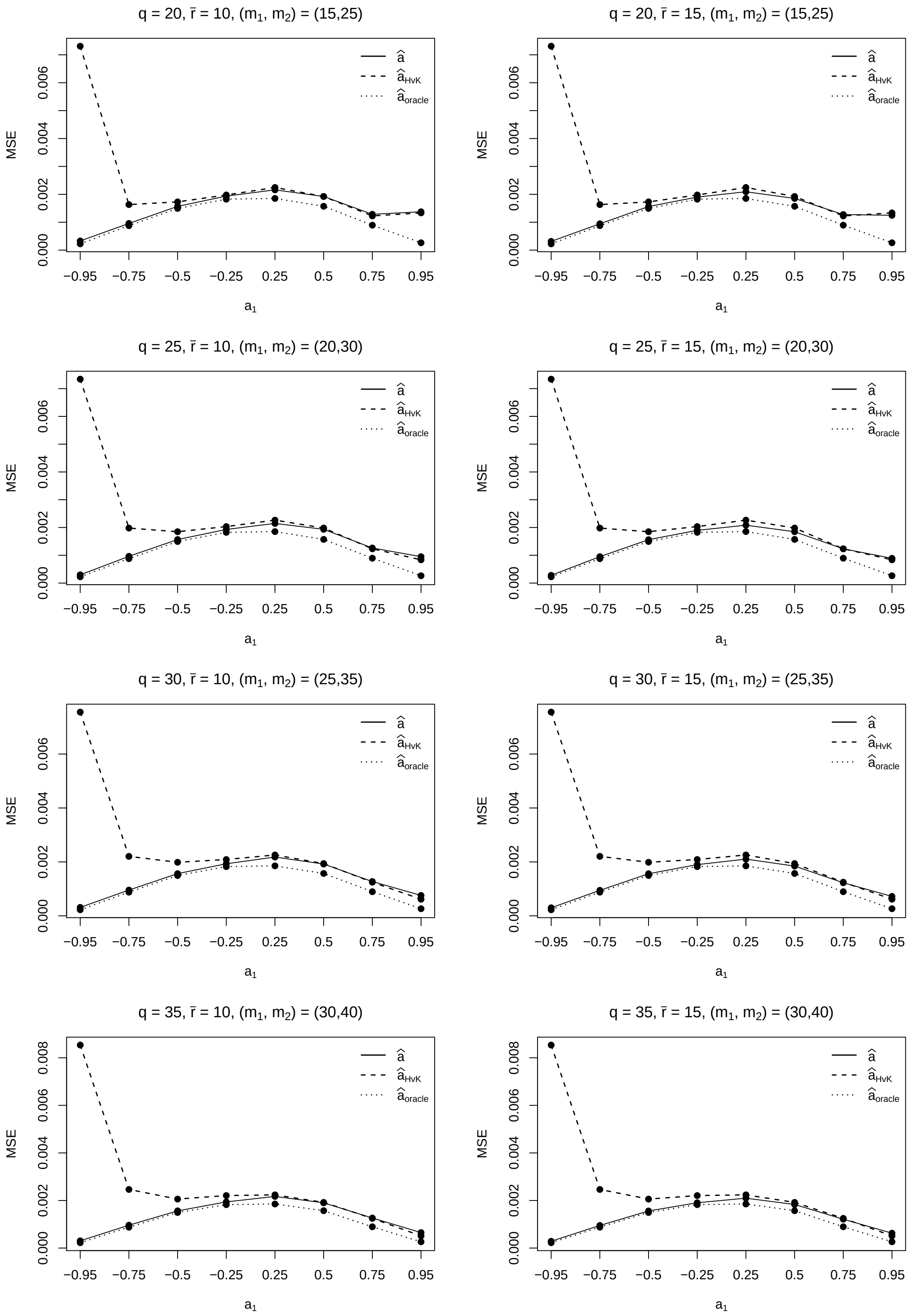


Figure S.3: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$ and $\widehat{a}_{\mathrm{oracle}}$ in the scenario with a moderate trend ($s_\beta = 1$).

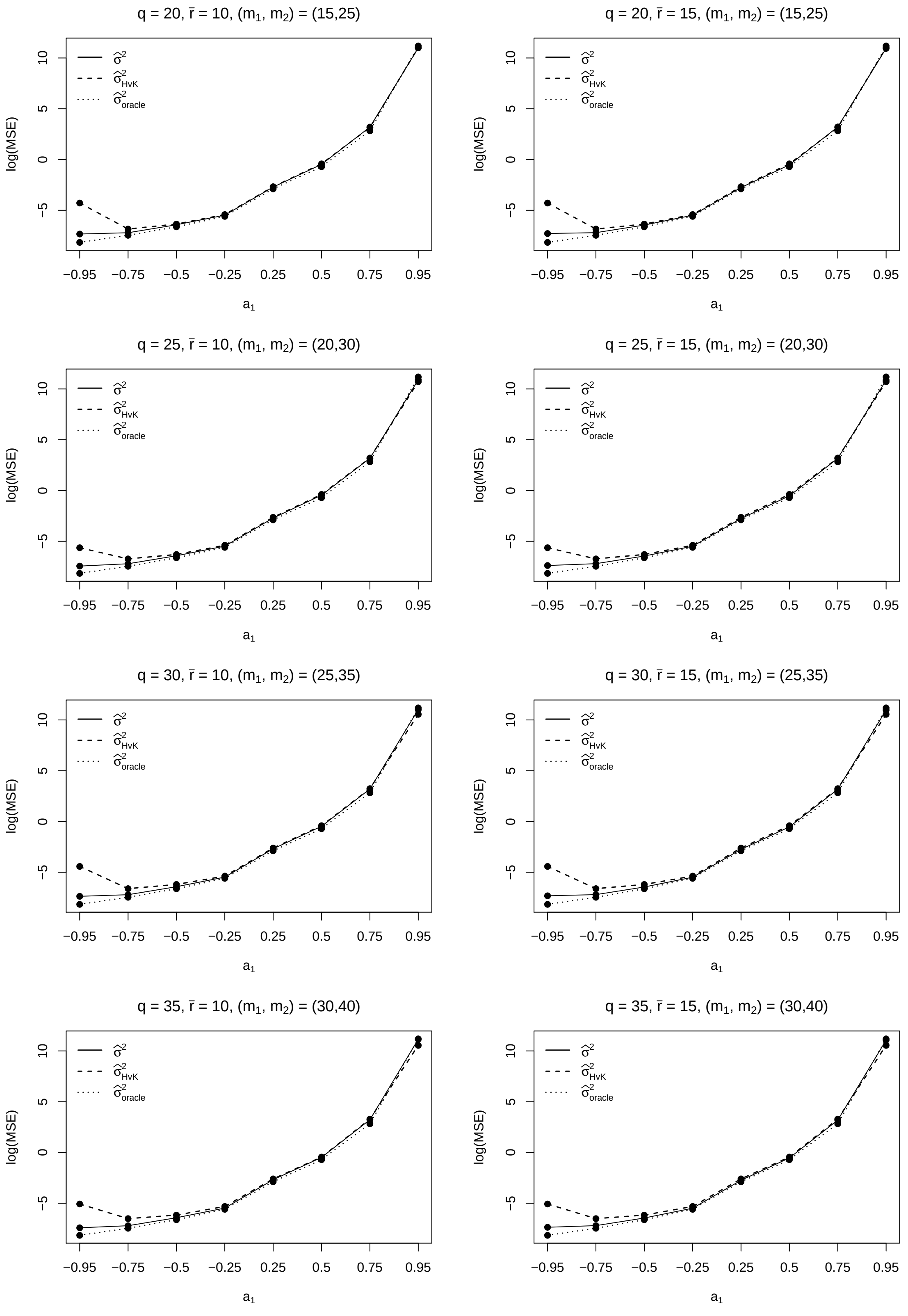


Figure S.4: Logarithmic MSE values for the estimators $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$ and $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the scenario with a moderate trend ($s_\beta = 1$).

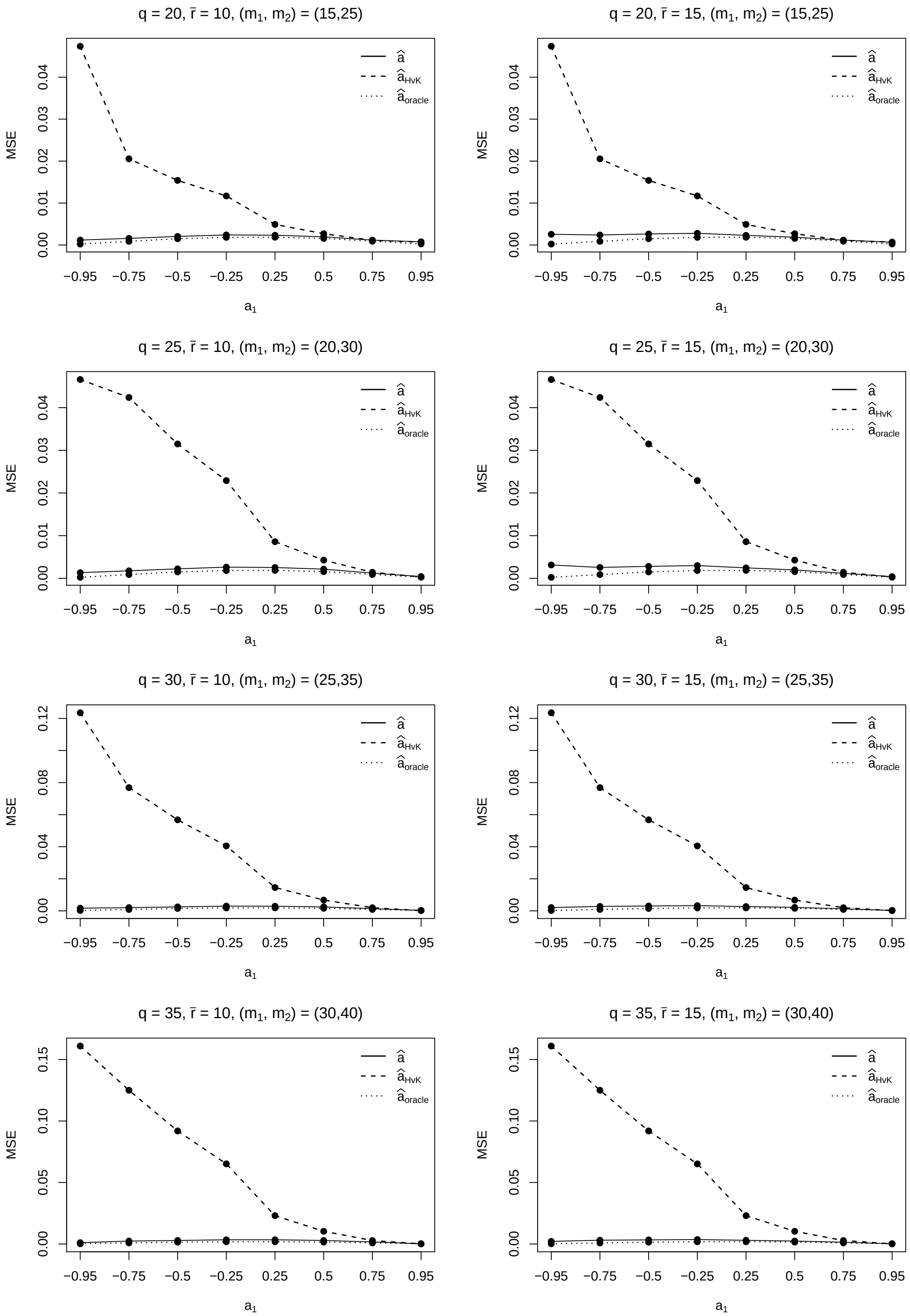


Figure S.5: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\mathrm{HvK}}$ and $\widehat{a}_{\mathrm{oracle}}$ in the scenario with a pronounced trend ($s_\beta = 10$).

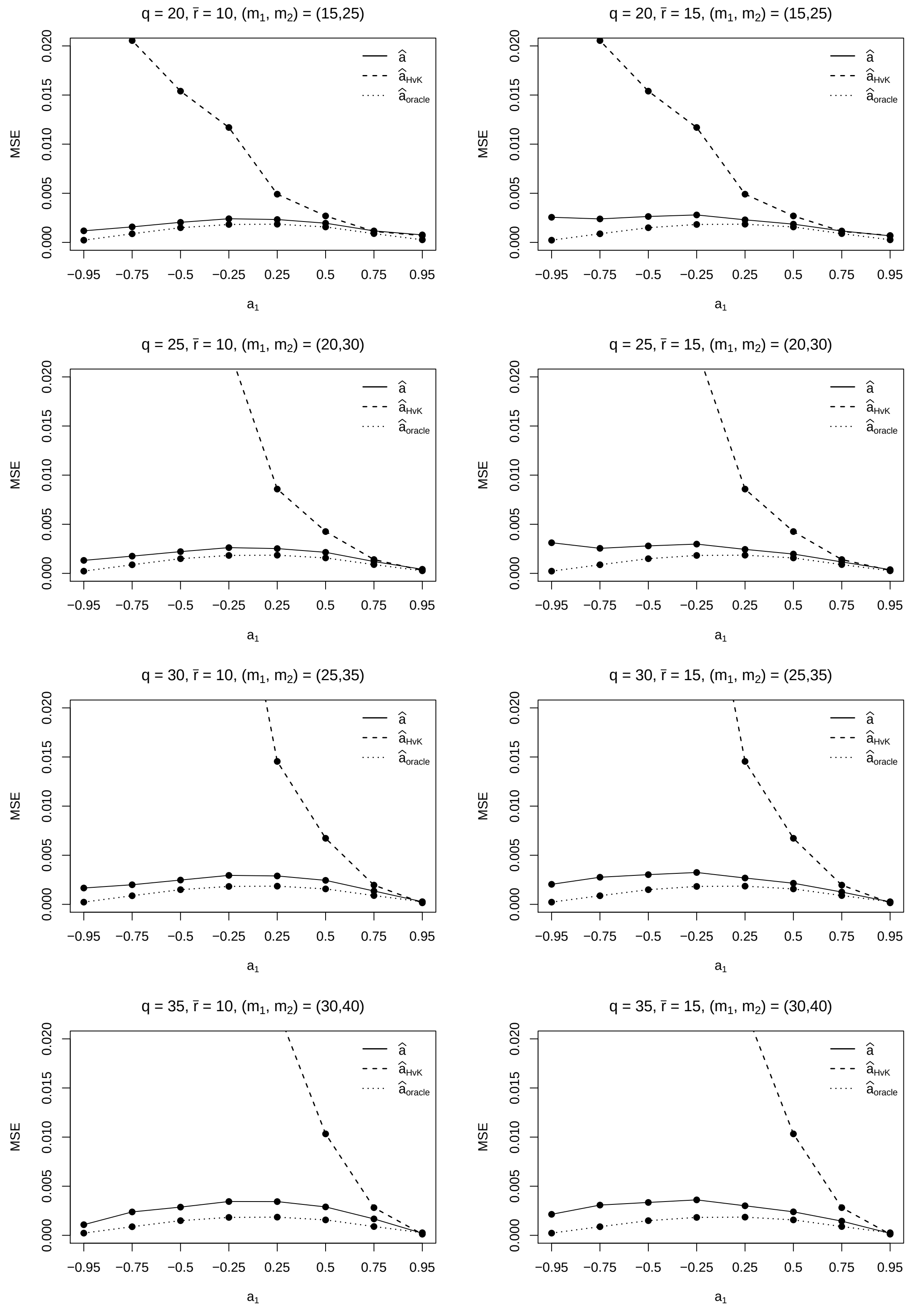


Figure S.6: MSE values for the estimators $\widehat{a}$, $\widehat{a}_{\text{HvK}}$ and $\widehat{a}_{\text{oracle}}$ in the scenario with a pronounced trend ($s_\beta = 10$). The plots are zoomed-in versions of the respective plots in Figure S.5.

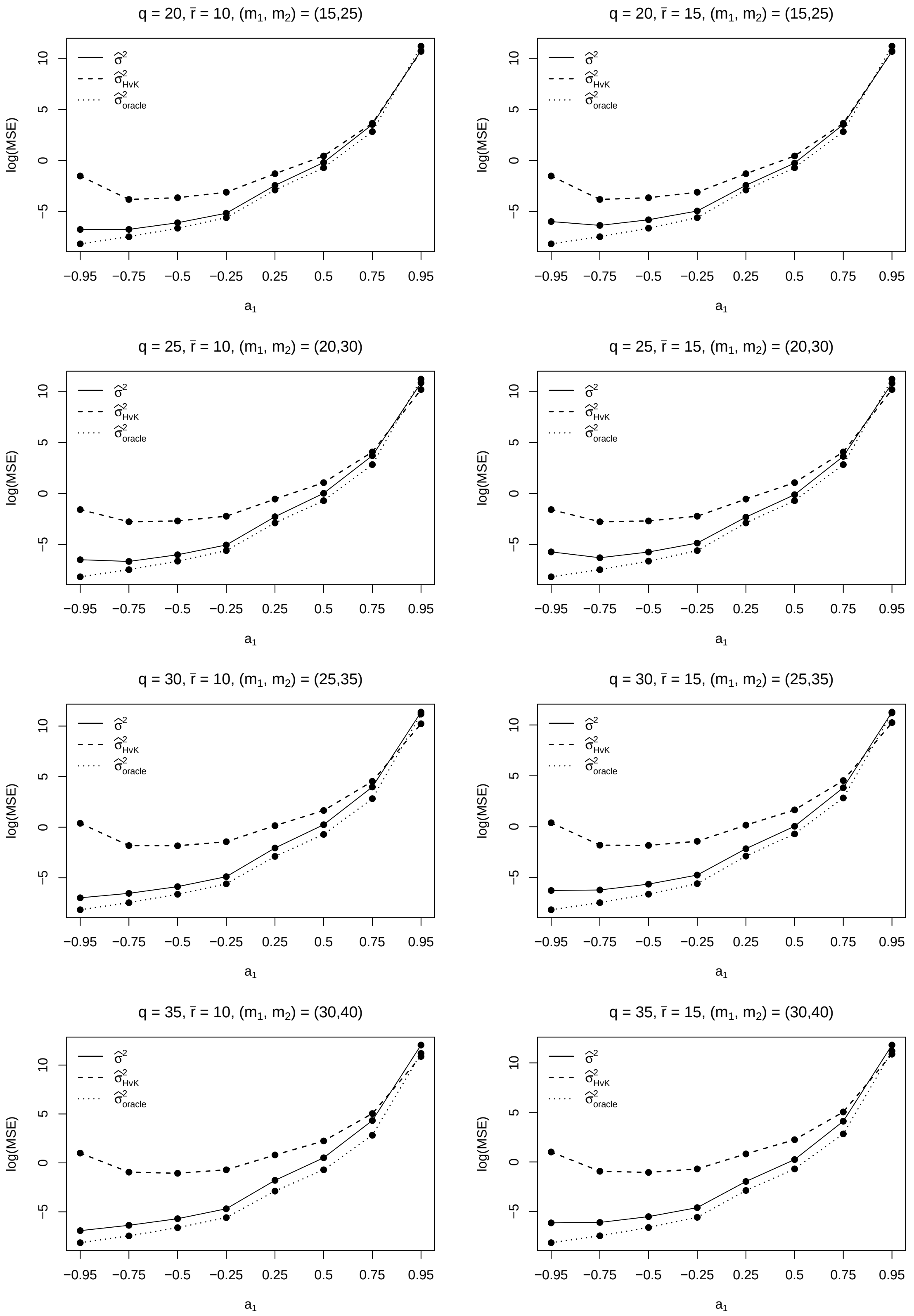


Figure S.7: Logarithmic MSE values for the estimators $\widehat{\sigma}^2$, $\widehat{\sigma}^2_{\mathrm{HvK}}$ and $\widehat{\sigma}^2_{\mathrm{oracle}}$ in the scenario with a pronounced trend ($s_\beta = 10$).