Supplemental Material for "Optimal Model Averaging for Joint Value-at-Risk and Expected Shortfall Regression"*

Shoukun Jiao [a 1], Wenchao Xu [b 1], Wuyi Ye [a], and Xinyu Zhang [c a]

[a]School of Management, University of Science and Technology of China

[b]School of Statistics and Information, Shanghai University of International Business and Economics

[c]Academy of Mathematics and Systems Science, Chinese Academy of Sciences

This appendix contains four sections. Section S.1 presents the details on the proof of Theorem 1-3 presented in the main paper. Section S.2 presents a detailed verification that the assumptions made in the theorems of our paper hold for the homoskedasticity case of DGP1. Section S.3 and S.4 contain additional results in simulation experiments and empirical analysis.

S.1 Detailed proofs

This section contains the proofs of Theorem 3.1–3.3.

S.1.1 Proof of Theorem 3.1

Let $\alpha_n = \sqrt{k_m/n}$, $\boldsymbol{b}_m = (\boldsymbol{b}_m^v, \boldsymbol{b}_m^e) \in \mathbb{R}^{2k_m}$ such that $\|\boldsymbol{b}_m\| = C$ for some large enough constant C, where $\boldsymbol{b}_m^v \in \mathbb{R}^{k_m}$ and $\boldsymbol{b}_m^e \in \mathbb{R}^{k_m}$. Using the same arguments in Lu and Su (2015), in order to prove this theorem, we only need to demonstrate that for any given $\varepsilon > 0$ there is a large enough constant C such that, for large

*The authors would like to thank the Co-Editor, Viktor Todorov, and three anonymous referees for their guidance and insightful comments which substantially improved the paper. Jiao and Ye thank support by the National Natural Science Foundation of China (Grant No.72371230) and the Anhui Provincial Natural Science Foundation (Grant No. 2208085J41). Xu thanks support by the National Natural Science Foundation of China (Grant No.12101591). Zhang thanks support by the National Natural Science Foundation of China (Grant Nos. 71925007, 72091212 and 71988101).

Address correspondence to Wuyi Ye, School of Management, University of Science and Technology of China, Hefei 230026, China; e-mail: wyye@ustc.edu.cn.

[1]These authors contributed equally to this work and should be considered co-first authors.

n we have

$$\mathbb{P}\left\{\inf_{\|\boldsymbol{b}_m\|=C}\sum_{i=1}^{n}L\left(y_i,v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right),e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)>\sum_{i=1}^{n}L\left(y_i,v_{i,m}\left(\Theta_m^*\right),e_{i,m}\left(\Theta_m^*\right)\right)\right\}\geq 1-\varepsilon. \tag{S.1}$$

This implies that with probability approaching 1 (w.p.a.1) there is a local minimum $\widehat{\Theta}_m$ in the ball $\{\Theta_m^*+\alpha_n\boldsymbol{b}_m:\|\boldsymbol{b}_m\|\leq C\}$ such that $\|\widehat{\Theta}_m-\Theta_m^*\|=O_p\left(\alpha_n\right)$. Below we decompose $\sum_{i=1}^{n}\left\{L\left(y_i,v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right),e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)-L\left(y_i,v_{i,m}\left(\Theta_m^*\right),e_{i,m}\left(\Theta_m^*\right)\right)\right\}$ into several terms and showing that one term greater than $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$ and each other terms are dominated by $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$.

By Taylor expansion and the fact $|1/e_{i,m}\left(\Theta\right)|<K_2<\infty$, we obtain

$$\begin{aligned}-\frac{1}{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)}=&-\frac{1}{e_{i,m}\left(\Theta_m^*\right)}+\frac{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\\&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^3}+o\left(\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\right)\end{aligned}$$

and

$$\begin{aligned}&\log\left(-e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)-\log\left(-e_{i,m}\left(\Theta_m^*\right)\right)\\&=\frac{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{2e_{i,m}\left(\Theta_m^*\right)^2}+o\left(\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\right).\end{aligned}$$

Since $|1/e_{i,m}\left(\Theta\right)|<K_2$ and $e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)=o_p\left(1\right)$, we move the remainder $o\left(\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\right)$ into the square term, i.e., $\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{2e_{i,m}\left(\Theta_m^*\right)^2}\left\{1+o_p\left(1\right)\right\}$. Then, we can write

$$\begin{aligned}&\sum_{i=1}^{n}L\left(y_i,v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right),e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)-\sum_{i=1}^{n}L\left(y_i,v_{i,m}\left(\Theta_m^*\right),e_{i,m}\left(\Theta_m^*\right)\right)\\
&=\sum_{i=1}^{n}\left[\frac{1}{\tau e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right\}\left\{y_i-v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right\}+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)}{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)}\right.\\
&\quad\left.+\log\left(-e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)-\frac{1}{\tau e_{i,m}\left(\Theta_m^*\right)}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}\left\{y_i-v_{i,m}\left(\Theta_m^*\right)\right\}-\frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}-\log\left(-e_{i,m}\left(\Theta_m^*\right)\right)\right]\\
&=\sum_{i=1}^{n}\left[Z_{1,i}+Z_{2,i}+Z_{3,i}\right.\\
&\quad-\frac{\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)}-\frac{\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)}\\
&\quad+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2}\\
&\quad+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2}\\
&\quad-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3}\\
&\quad-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3}+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)}{e_{i,m}\left(\Theta_m^*\right)}\\
&\quad-\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}}{e_{i,m}^2\left(\Theta_m^*\right)}+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}^3\left(\Theta_m^*\right)}\end{aligned}$$

$$+\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{\tau e_{i,m}\left(\Theta_m^*\right)}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*\right)\right\}-\frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}$$

$$\left.+\frac{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{2e_{i,m}^2\left(\Theta_m^*\right)}\left\{1+o_p\left(1\right)\right\}\right],$$

where

$$\begin{aligned}
Z_{1,i}=&-\frac{\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)}+\frac{\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)},\\
Z_{2,i}=&\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2}\\
&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2},\\
\text{and}\quad Z_{3,i}=&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3}\\
&+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3}.
\end{aligned}$$

For $Z_{1,i}$, by the fact that

$$\int_0^{-v}\left\{\mathbb{I}\left(u\le s\right)-\mathbb{I}\left(u\le 0\right)\right\}ds=-\left(u+v\right)\mathbb{I}\left(u+v\le 0\right)+\left(u+v\right)\mathbb{I}\left(u\le 0\right),$$

we have

$$\begin{aligned}
\tau\sum_{i=1}^n Z_{1,i}=&\sum_{i=1}^n\left[-\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le 0\right\}\right.\\
&\left.+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le 0\right\}\right]\\
=&\sum_{i=1}^n\int_0^{\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}-\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}}\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le s\right\}-\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le 0\right\}ds\\
=&\sum_{i=1}^n\left(Z_{1,1,i}+Z_{1,2,i}\right),
\end{aligned}$$

where

$$\begin{aligned}
Z_{1,1,i}=&\mathbb{E}\left(\mathbb{E}\left[\int_0^{\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}-\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}}\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le s\right\}\right.\right.\\
&\left.\left.-\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le 0\right\}ds\right|\boldsymbol{x}_i\right]\right),
\end{aligned}$$

and

$$\begin{aligned}
Z_{1,2,i}=&\int_0^{\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}-\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}}\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le s\right\}-\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le 0\right\}ds\\
&-\mathbb{E}\left(\mathbb{E}\left[\int_0^{\frac{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}-\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}}\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le s\right\}\right.\right.\\
&\left.\left.-\mathbb{I}\left\{\frac{v_{i,m}\left(\Theta_m^*\right)-y_i}{e_{i,m}\left(\Theta_m^*\right)}\le 0\right\}ds\right|\boldsymbol{x}_i\right]\right).
\end{aligned}$$

For $Z_{1,1,i}$, by the law of iterated expectations and Taylor expansion,

$$\begin{aligned}
\sum_{i=1}^{n} Z_{1,1,i} = \mathbb{E}\Bigg(\mathbb{E}\Bigg[\sum_{i=1}^{n}\int_{0}^{\frac{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-y_i}{e_{i,m}(\Theta_m^*)}-\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}} & \mathbb{I}\left\{\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}\leq s\right\} \\
& -\mathbb{I}\left\{\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}\leq 0\right\} ds \Bigg| \boldsymbol{x}_i\Bigg]\Bigg) \\
= \mathbb{E}\left\{\sum_{i=1}^{n}\int_{0}^{\frac{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)}{e_{i,m}(\Theta_m^*)}-\frac{v_{i,m}(\Theta_m^*)}{e_{i,m}(\Theta_m^*)}} F_{y|\boldsymbol{x}}\left(v_{i,m}(\Theta_m^*)-e_{i,m}(\Theta_m^*)s\right)-F_{y|\boldsymbol{x}}\left(v_{i,m}(\Theta_m^*)\right) ds\right\} \\
= \mathbb{E}\left[\sum_{i=1}^{n}\int_{0}^{\frac{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)}{e_{i,m}(\Theta_m^*)}-\frac{v_{i,m}(\Theta_m^*)}{e_{i,m}(\Theta_m^*)}} f_{y|\boldsymbol{x}}\left(v_{i,m}(\Theta_m^*)\right)\left|e_{i,m}(\Theta_m^*)\right| s ds \left\{1+o_p(1)\right\}\right] \\
= \mathbb{E}\left[\sum_{i=1}^{n}\frac{1}{2} f_{y|\boldsymbol{x}}\left(v_{i,m}(\Theta_m^*)\right)\left|e_{i,m}(\Theta_m^*)\right|\left\{\frac{v_{i,m}(\Theta_m^*)-v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)}{e_{i,m}(\Theta_m^*)}\right\}^2\left\{1+o_p(1)\right\}\right] \\
= \mathbb{E}\left[\sum_{i=1}^{n} f_{y|\boldsymbol{x}}\left(v_{i,m}(\Theta_m^*)\right)\frac{\left\{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-v_{i,m}(\Theta_m^*)\right\}^2}{-2e_{i,m}(\Theta_m^*)}\left\{1+o_p(1)\right\}\right]. \qquad \text{(S.2)}
\end{aligned}$$

Noting that $\mathbb{E}(Z_{1,2,i})=0$ and by Assumption 3.1(iv) and (vi),

$$\begin{aligned}
& \operatorname{Var}\left(\sum_{i=1}^{n} Z_{1,2,i}\right) \\
=& \operatorname{Var}\left[\sum_{i=1}^{n}\int_{0}^{\frac{v_{i,m}(\Theta_m^*)}{e_{i,m}(\Theta_m^*)}-\frac{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)}{e_{i,m}(\Theta_m^*)}} \mathbb{I}\left\{\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}\leq s\right\}-\mathbb{I}\left\{\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}\leq 0\right\} ds\right] \\
\leq& \sum_{i=1}^{n}\mathbb{E}\left\{\mathbb{E}\left(\left[\int_{0}^{\frac{v_{i,m}(\Theta_m^*)}{e_{i,m}(\Theta_m^*)}-\frac{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)}{e_{i,m}(\Theta_m^*)}} \mathbb{I}\left\{\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}\leq s\right\}-\mathbb{I}\left\{\frac{v_{i,m}(\Theta_m^*)-y_i}{e_{i,m}(\Theta_m^*)}\leq 0\right\} ds\right]^2\middle|\boldsymbol{x}_i\right)\right\} \\
\leq& \sum_{i=1}^{n}\mathbb{E}\left[\left\{\frac{v_{i,m}(\Theta_m^*)}{e_{i,m}(\Theta_m^*)}-\frac{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)}{e_{i,m}(\Theta_m^*)}\right\}^2\right] \\
\leq& n\alpha_n^2 K_2^2\left(\boldsymbol{b}_m^v\right)^\top \mathbb{E}\left(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top\right)\boldsymbol{b}_m^v \\
=& K_2^2 n\alpha_n^2 \bar{c}_{X(m)}\left\|\boldsymbol{b}_m^v\right\|^2 \leq K_2^2 n\alpha_n^2 \bar{c}_{X(m)}\left\|\boldsymbol{b}_m\right\|^2.
\end{aligned}$$

Therefore, we have $\sum_{i=1}^{n} Z_{1,2,i} = O_p\left(\alpha_n\sqrt{n}\bar{c}_{X(m)}^{1/2}\right)\|\boldsymbol{b}_m\| = O_p\left(\alpha_n^2 n/\sqrt{k_m}\bar{c}_{X(m)}^{1/2}\right)\|\boldsymbol{b}_m\|$. Similarly, let $Z_{2,1,i}=\mathbb{E}\left(\tau Z_{2,i}\right), Z_{2,2,i}=\tau Z_{2_i}-Z_{2,1,i}$. Then, by Assumption 3.1(iv) and (v),

$$\begin{aligned}
\left|\sum_{i=1}^{n} Z_{2,1,i}\right| =& \left|\mathbb{E}\left(\tau\sum_{i=1}^{n} Z_{2,i}\right)\right| \\
\leq& \Bigg|\mathbb{E}\Bigg(\mathbb{E}\Bigg[\sum_{i=1}^{n}-\frac{\left(e_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-e_{i,m}(\Theta_m^*)\right)\left\{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-y_i\right\}}{e_{i,m}(\Theta_m^*)^2}\mathbb{I}\left\{y_i\leq v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)\right\} \\
& +\frac{\left(e_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-e_{i,m}(\Theta_m^*)\right)\left\{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-y_i\right\}}{e_{i,m}(\Theta_m^*)^2}\mathbb{I}\left\{y_i\leq v_{i,m}(\Theta_m^*)\right\}|\boldsymbol{x}_i\Bigg]\Bigg)\Bigg| \\
=& \Bigg|\mathbb{E}\Bigg\{\mathbb{E}\Bigg(\sum_{i=1}^{n}-\frac{\left|e_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-e_{i,m}(\Theta_m^*)\right|\left\{v_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-y_i\right\}}{\operatorname{sgn}\left(e_{i,m}(\Theta_m^*+\alpha_n \boldsymbol{b}_m)-e_{i,m}(\Theta_m^*)\right)e_{i,m}(\Theta_m^*)^2}
\end{aligned}$$

$$
\begin{aligned}
&\times \mathbb{I}\left[\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-y_i\right\}}{e_{i,m}\left(\Theta_m^*\right)^2} \leq 0\right] \\
&+\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-y_i\right\}}{\operatorname{sgn}\left(e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right) e_{i,m}\left(\Theta_m^*\right)^2} \\
&\left.\left.\left.\times \mathbb{I}\left[\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}}{e_{i,m}\left(\Theta_m^*\right)^2} \leq 0\right]\right| \boldsymbol{x}_i\right)\right\}\Bigg| \\
\leq & \mathbb{E}\left[\sum_{i=1}^n\left|\int_0^{\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-y_i\right\}}{e_{i,m}\left(\Theta_m^*\right)^2}-\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}}{e_{i,m}\left(\Theta_m^*\right)^2}}\right.\right. \\
& \left.\left. F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)-\frac{e_{i,m}\left(\Theta_m^*\right)^2}{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|} s\right)-F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right) d s\right|\right] \\
= & \mathbb{E}\left[\sum_{i=1}^n\left|\int_0^{\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-y_i\right\}}{e_{i,m}\left(\Theta_m^*\right)^2}-\frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}}{e_{i,m}\left(\Theta_m^*\right)^2}}\right.\right. \\
& \left.\left. f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right) \frac{e_{i,m}\left(\Theta_m^*\right)^2}{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|} s d s\right|\right] \\
= & \mathbb{E}\left[\sum_{i=1}^n \frac{1}{2} f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right) \frac{e_{i,m}\left(\Theta_m^*\right)^2}{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|}\right. \\
& \left.\times \frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^4}\right] \\
= & \mathbb{E}\left[\sum_{i=1}^n \frac{1}{2} f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right) \frac{\left|e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right|\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^2}\right] \\
\leq & \frac{1}{2} c_f n K_2^2 \alpha_n^3\left\|\boldsymbol{b}_m\right\|^3 \bar{c}_{X(m)}^{3/2} .
\end{aligned}
$$

Since $\mathbb{E}(Z_{2,2,i})=0$, similar to the calculation of $\operatorname{Var}(\sum_{i=1}^n Z_{1,2,i})$, by Assumption 3.1(iv) and (v),

$$
\begin{aligned}
& \operatorname{Var}(\sum_{i=1}^n Z_{2,2,i}) \\
\leq & \sum_{i=1}^n \mathbb{E}\left[\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^4}\right] \\
\leq & n K_2^4 \alpha_n^3\left\|\boldsymbol{b}_m\right\|^3 \bar{c}_{X(m)}^{3/2} .
\end{aligned}
$$

Therefore, we have $\sum_{i=1}^n Z_{2,2,i}=O_p\left(\alpha_n^{3/2}\sqrt{n}\bar{c}_{X(m)}^{3/4}\right)\left\|\boldsymbol{b}_m\right\|^{3/2}$.

For $Z_{3,i}$, using the steps similar to $Z_{2,i}$, we have $\mathbb{E}(\tau\sum_{i=1}^n Z_{3,i}) \leq \frac{1}{2}c_f n K_2^3\alpha_n^4\left\|\boldsymbol{b}_m\right\|^4\bar{c}_{X(m)}^2$ and $\sum_{i=1}^n\{Z_{3,i}-\mathbb{E}(Z_{3,i})\}=O_p\left(\alpha_n^2\sqrt{n}\bar{c}_{X(m)}\right)\left\|\boldsymbol{b}_m\right\|^2$. Recall that $\alpha_n\bar{c}_{X(m)}^{1/2}\left\|\boldsymbol{b}_m\right\|^{1/2}=o_p(1)$ which ensure that $\mathbb{E}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-\mathbb{E}\left(\Theta_m^*\right)=o_p(1)$ and $v\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v\left(\Theta_m^*\right)=o_p(1)$. By Assumption 3.1(v) and allowing $\left\|\boldsymbol{b}_m\right\|$ to be sufficiently large, both $Z_{1,2}$, Z_2 and Z_3 are dominated by $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$, which is positive w.p.a.1.

Note that $\Theta_m^* = \arg\min_{\Theta_m} \mathbb{E}\left\{L\left(y_i, v_{i,m}\left(\Theta_m\right), e_{i,m}\left(\Theta_m\right)\right)\right\}$, we have

$$\begin{aligned}
0 &= \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m\right)\right\} \\
&= \mathbb{E}\left\{\frac{\partial L_m\left(y_i, v_{i,m}\left(\Theta_m\right), e_{i,m}\left(\Theta_m\right)\right)}{\partial \Theta_m}\right\} \\
&= \mathbb{E}\left(\frac{\left(\boldsymbol{x}_{i,m}; \mathbf{0}_{k_m\times 1}\right)}{-e_{i,m}\left(\Theta_m\right)}\left[\frac{1}{\tau}\mathbb{I}\left\{y_i \le v_{i,m}\left(\Theta_m\right)\right\} - 1\right]\right. \\
&\qquad \left. + \frac{\left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)}{e_{i,m}\left(\Theta_m\right)^2}\left[\frac{1}{\tau}\mathbb{I}\left\{y_i \le v_{i,m}\left(\Theta_m\right)\right\}\left\{v_{i,m}\left(\Theta_m\right) - y_i\right\} - v_{i,m}\left(\Theta_m\right) + e_{i,m}\left(\Theta_m\right)\right]\right)
\end{aligned}$$

and

$$\begin{aligned}
\boldsymbol{\Lambda}_m\left(\Theta_m^*\right) &= \left.\frac{\partial \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m\right)\right\}}{\partial \Theta_m}\right|_{\Theta_m = \Theta_m^*} \\
&= \mathbb{E}\left(\left\{\frac{\left(\boldsymbol{x}_{i,m}; \mathbf{0}_{k_m\times 1}\right)\left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)^\top + \left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)\left(\boldsymbol{x}_{i,m}; \mathbf{0}_{k_m\times 1}\right)^\top}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\left\{\frac{F_{y|\boldsymbol{x}}\left(v\left(\Theta_m^*\right)\right)}{\tau} - 1\right\}\right. \\
&\qquad - \left\{\frac{2\left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)\left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)^\top}{e_{i,m}\left(\Theta_m^*\right)^3}\right\} \\
&\qquad \times \left[\frac{F_{y|\boldsymbol{x}}\left(v\left(\Theta_m^*\right)\right) v\left(\Theta_m^*\right)}{\tau} - \frac{\mathbb{E}\left\{y | y_i \le v\left(\Theta_m^*\right), \boldsymbol{x}_i\right\}}{\tau} - v\left(\Theta_m^*\right) + \mathbb{E}\left(\Theta_m^*\right)\right] \\
&\qquad \left. + \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right) | \boldsymbol{x}\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)}\left(\boldsymbol{x}_{i,m}; \mathbf{0}_{k_m\times 1}\right)\left(\boldsymbol{x}_{i,m}; \mathbf{0}_{k_m\times 1}\right)^\top + \frac{1}{e_{i,m}\left(\Theta_m^*\right)^2}\left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)\left(\mathbf{0}_{k_m\times 1}; \boldsymbol{x}_{i,m}\right)^\top\right)
\end{aligned}$$

is a positive definite matrix. Then,

$$\mathbb{E}\left[-\frac{\left\{v_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)}\right] = \mathbb{E}\left\{-\frac{v_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}\right\}, \tag{S.3}$$

$$\begin{aligned}
&\mathbb{E}\left[\frac{\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*\right) - y_i\right\}\mathbb{I}\left\{y_i \le v_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2}\right] \\
&= \mathbb{E}\left[\frac{\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*\right) - e_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2}\right]
\end{aligned} \tag{S.4}$$

and

$$\begin{aligned}
\mathbb{E}&\left[\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right) | \boldsymbol{x}\right)\left\{v_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - v_{i,m}\left(\Theta_m^*\right)\right\}^2}{-2\tau e_{i,m}\left(\Theta_m^*\right)}\right. \\
&+ \frac{\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2} \\
&- \frac{\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*\right) - y_i\right\}\mathbb{I}\left\{y_i \le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3} \\
&+ \frac{\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*\right) - v\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right)\right\}}{e_{i,m}\left(\Theta_m^*\right)^2} \\
&\left. + \frac{\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}^2}{2 e_{i,m}\left(\Theta_m^*\right)^2} + \frac{\left\{v_{i,m}\left(\Theta_m^*\right) - e_{i,m}\left(\Theta_m^*\right)\right\}\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_0\right)^3}\right]
\end{aligned}$$

$$= \alpha_n^2 \left(\boldsymbol{b}_m\right)^\top \boldsymbol{\Lambda}_m \left(\Theta_m^*\right) \boldsymbol{b}_m \geq \alpha_n^2 \left\|\boldsymbol{b}_m\right\|^2 \lambda_{\min}\left(\boldsymbol{\Lambda}_m \left(\Theta_m^*\right)\right) = \alpha_n^2 \left\|\boldsymbol{b}_m\right\|^2 \underline{c}_{D(m)}. \tag{S.5}$$

By (S.2)-(S.5), we have

$$\begin{aligned}
&\mathbb{E}\left\{\sum_{i=1}^n L\left(y_i, v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right), e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)\right) - \sum_{i=1}^n L\left(y_i, v_{i,m}\left(\Theta_m^*\right), e_{i,m}\left(\Theta_m^*\right)\right) - Z_{2,i} - Z_{3,i}\right\} \\
=&\sum_{i=1}^n \mathbb{E}\left[\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)|\boldsymbol{x}\right)\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}^2}{-2\tau e_{i,m}\left(\Theta_m^*\right)}\left\{1+o_p(1)\right\}\right. \\
&-\frac{\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)} \\
&+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2} \\
&+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2} \\
&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3} \\
&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3} \\
&+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)}{e_{i,m}\left(\Theta_m^*\right)}-\frac{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}}{e_{i,m}\left(\Theta_m^*\right)^2} \\
&+\frac{v\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^3} \\
&\left.-\frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}+\frac{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{2e_{i,m}\left(\Theta_m^*\right)^2}\left\{1+o_p(1)\right\}\right] \\
=&\sum_{i=1}^n \mathbb{E}\left[\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)|\boldsymbol{x}\right)\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}^2}{-2\tau e_{i,m}\left(\Theta_m^*\right)}\left\{1+o_p(1)\right\}\right. \\
&-\frac{\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}}{e_{i,m}\left(\Theta_m^*\right)} \\
&+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2} \\
&+\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}\left\{v_{i,m}\left(\Theta_m^*\right)-e_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^2} \\
&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3} \\
&-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*\right)-y_i\right\}\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3} \\
&+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)}{e_{i,m}\left(\Theta_m^*\right)}-\frac{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}}{e_{i,m}\left(\Theta_m^*\right)^2} \\
&+\frac{v_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^3} \\
&\left.-\frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}+\frac{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n \boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{2e_{i,m}\left(\Theta_m^*\right)^2}\left\{1+o_p(1)\right\}\right]
\end{aligned}$$

$$=n\alpha_n^2\left(\boldsymbol{b}_m\right)^\top \boldsymbol{\Lambda}_m\left(\Theta_m^*\right)\boldsymbol{b}_m\left\{1+o_p\left(1\right)\right\}+\sum_{i=1}^n\left(Z_{4,i}+Z_{5,i}\right),$$

where

$$Z_{4,i}=-\frac{\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\mathbb{I}\left\{y_i\le v_{i,m}\left(\Theta_m^*\right)\right\}}{\tau e_{i,m}\left(\Theta_m^*\right)^3}$$

and

$$Z_{5,i}=\frac{\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2}{e_{i,m}\left(\Theta_m^*\right)^3}.$$

Then, we will prove that $n\alpha_n^2\left(\boldsymbol{b}_m\right)^\top \boldsymbol{\Lambda}_m\left(\Theta_m^*\right)\boldsymbol{b}_m\left\{1+o_p\left(1\right)\right\}$ greater than $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$ and $\sum_{i=1}^n\mathbb{E}\left(Z_{4,i}\right)+\sum_{i=1}^n\mathbb{E}\left(Z_{5,i}\right)$ are dominated by $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$ w.p.a.1.

By Assumption 3.1(iv) and (v), we have

$$n\alpha_n^2\left(\boldsymbol{b}_m\right)^\top \boldsymbol{\Lambda}_m\left(\Theta_m^*\right)\boldsymbol{b}_m\left\{1+o_p\left(1\right)\right\}\ge n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2\qquad\text{w.p.a.1,}$$

$$\begin{aligned}\sum_{i=1}^n\left|\mathbb{E}\left(Z_{4,i}\right)\right|&\le\frac{nK_2^3}{\tau}\mathbb{E}\left[\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left|v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right|\right]\\&\le\frac{nK_2^3\alpha_n^3}{\tau}\left\|\boldsymbol{b}_m\right\|^3\bar{c}_{X(m)}^{3/2}\end{aligned}$$

and

$$\begin{aligned}\sum_{i=1}^n\left|\mathbb{E}\left(Z_{5,i}\right)\right|&\le\frac{nK_2^3}{\tau}\mathbb{E}\left[\left\{e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-e_{i,m}\left(\Theta_m^*\right)\right\}^2\left|v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right|\right]\\&\le\frac{nK_2^3\alpha_n^3}{\tau}\left\|\boldsymbol{b}_m\right\|^3\bar{c}_{X(m)}^{3/2}.\end{aligned}$$

By Assumption 3.1(v) and $\alpha_n\bar{c}_{X(m)}^{1/2}\left\|\boldsymbol{b}_m\right\|^{1/2}=o_p\left(1\right)$, $\sum_{i=1}^n\mathbb{E}\left(Z_{4,i}\right)+\sum_{i=1}^n\mathbb{E}\left(Z_{5,i}\right)$ are dominated by $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$ w.p.a.1.

Finally, we need to show that

$$\begin{aligned}&\sum_{i=1}^n L\left(y_i,v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right),e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)-\sum_{i=1}^n L\left(y_i,v_{i,m}\left(\Theta_m^*\right),e_{i,m}\left(\Theta_m^*\right)\right)-Z_{2,i}-Z_{3,i}\\&\quad-\mathbb{E}\left\{\sum_{i=1}^n L\left(y_i,v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right),e_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)\right)-\sum_{i=1}^n L\left(y_i,v_{i,m}\left(\Theta_m^*\right),e_{i,m}\left(\Theta_m^*\right)\right)-Z_{2,i}-Z_{3,i}\right\}\end{aligned}\tag{S.6}$$

is dominated by $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$ w.p.a.1. As before, we can decompose (S.6) into several terms and prove that each term is dominated by $n\underline{c}_{D(m)}\alpha_n^2\left\|\boldsymbol{b}_m\right\|^2$. This work is easy and the detailed proofs are omitted for saving space, just giving an example. Let $Z_{6,i}=v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)-\mathbb{E}\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}$ be a term in the decomposition of (S.6), by Assumption 3.1(v) and the fact that $E(Z_6)=0$, we have

$$\begin{aligned}\sum_{i=1}^n\mathrm{Var}\left(Z_{6,i}\right)&\le n\,\mathbb{E}\left[\left\{v_{i,m}\left(\Theta_m^*+\alpha_n\boldsymbol{b}_m\right)-v_{i,m}\left(\Theta_m^*\right)\right\}^2\right]\\&=n\alpha_n^2\left(\boldsymbol{b}_m^v\right)^\top\mathbb{E}\left(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top\right)\boldsymbol{b}_m^v\le n\alpha_n^2\left\|\boldsymbol{b}_m^v\right\|\bar{c}_{X(m)}.\end{aligned}$$

Therefore, $\sum_{i=1}^{n} Z_{6,i} = O_p\left(\alpha_n \sqrt{n}\bar{c}_{X(m)}^{1/2}\right)\|\boldsymbol{b}_m\| = O_p\left(\alpha_n^2 n/\sqrt{k_m}\bar{c}_{X(m)}^{1/2}\right)\|\boldsymbol{b}_m\|$. Similarly, it can be proved that

$$\sum_{i=1}^{n} e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right) - \mathbb{E}\left\{e_{i,m}\left(\Theta_m^* + \alpha_n \boldsymbol{b}_m\right) - e_{i,m}\left(\Theta_m^*\right)\right\} = O_p\left(\alpha_n^2 n/\sqrt{k_m}\bar{c}_{X(m)}^{1/2}\right)\|\boldsymbol{b}_m\|.$$

By Assumption 3.1(iv) and (v), $\alpha_n \bar{c}_{X(m)}^{1/2}\|\boldsymbol{b}_m\|^{1/2} = o_p(1)$ and Hölder inequality, (S.6) is dominated by $n\underline{c}_{D(m)}\alpha_n^2\|\boldsymbol{b}_m\|^2$ w.p.a.1. ■

S.1.2 Proof of Theorem 3.2

Let

$$\widehat{\Delta}_m \equiv \sqrt{n}\left(\widehat{\Theta}_m - \Theta_m^*\right) \quad \text{and} \quad \Delta_m \equiv \sqrt{n}\left(\Theta_m - \Theta_m^*\right).$$

It follows that $\widehat{\Delta}_m = \operatorname*{argmin}_{\Delta_m} \sum_{i=1}^{n} L\left(y_i, v_{i,m}\left(\Theta_m^* + n^{-1/2}\Delta_m\right), e_{i,m}\left(\Theta_m^* + n^{-1/2}\Delta_m\right)\right)$. Let

$$V_m(\Delta) \equiv n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}\left(\Theta_m^* + n^{-1/2}\Delta\right) \quad \text{and} \quad \bar{V}_m(\Delta) \equiv \mathbb{E}\left\{V_m(\Delta)\right\}.$$

Define $\|A\|_{c_m} = \|c_m A\|$, where c_m is an arbitrary $l_m \times 2k_m$ matrix with $\|c_m\| \le L_c \underline{c}_{A(m)}^{-1/2}$ for a large constant $L_c < \infty$. We want to show that for any large constant $L < \infty$,

$$\sup_{\|\Delta\| \le \sqrt{k_m}L} \left\|V_m(\Delta) - V_m(0) - \bar{V}_m(\Delta) + \bar{V}_m(0)\right\|_{c_m} = o_p(1), \tag{S.7}$$

$$\sup_{\|\Delta\| \le \sqrt{k_m}L} \left\|\bar{V}_m(\Delta) - \bar{V}_m(0) + \mathbf{D}_m \Delta\right\|_{c_m} = o_p(1), \tag{S.8}$$

$$\text{and} \quad \left\|V_m\left(\widehat{\Delta}_m\right)\right\|_{c_m} = o_p(1). \tag{S.9}$$

Next, we are going to verify (S.7). Let $D = \{\Delta \in \mathbb{R}^{2k_m} : \|\Delta\| \le \sqrt{k_m}L\}$ for some $L < \infty$, and $|t|_\infty$ denote the maximum of the absolute values of the coordinates of t. By selecting $N_1 = (2n)^{2k_m}$ grid point, we can cover D by cubes $D_s = \{\Delta \in \mathbb{R}^{2k_m} : |\Delta - \Delta_s|_\infty \le \delta_n\}$, with sides of length $\delta_n = Lk_m^{1/2}/n$. Then,

$$\begin{aligned}
&\sup_{\|\Delta\| \le \sqrt{k_m}L} \left\|V_m(\Delta) - V_m(0) - \bar{V}_m(\Delta) + \bar{V}_m(0)\right\|_{c_m} \\
&= \sup_{\|\Theta - \Theta_m^*\| \le \sqrt{k_m}L/\sqrt{n}} \Big\| n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta) - n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}\left(\Theta_m^*\right) \\
&\qquad - \mathbb{E}\left\{n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta)\right\} + \mathbb{E}\left\{n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\Big\|_{c_m} \\
&\le \max_{1 \le s \le N_1} \Big\| n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta_s) - n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}\left(\Theta_m^*\right) \\
&\qquad - \mathbb{E}\left\{n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta_s)\right\} + \mathbb{E}\left\{n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\Big\|_{c_m} \\
&+ \max_{1 \le s \le N_1} \sup_{\Theta \in D_s} \Big\| n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta) - n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta_s) \\
&\qquad - \mathbb{E}\left\{n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta)\right\} + \mathbb{E}\left\{n^{-1/2}\sum_{i=1}^{n} \boldsymbol{g}_{i,m}(\Theta_s)\right\}\Big\|_{c_m}.
\end{aligned} \tag{S.10}$$

Since l_m is fixed, without loss of generality we assume that $l_m = 1$. Next, we will verify that the two terms of the right side of inequality (S.10) are $o_p(1)$.

Let $\eta_{i,m,s} = \sqrt{n} c_m \left[\boldsymbol{g}_{i,m}(\Theta_s) - \boldsymbol{g}_{i,m}(\Theta_m^*) - \mathbb{E}\{\boldsymbol{g}_{i,m}(\Theta_s)\} + \mathbb{E}\{\boldsymbol{g}_{i,m}(\Theta_m^*)\}\right]$. Recall that

$$\begin{aligned}\boldsymbol{g}_{i,m}(\Theta_m) =& \nabla^\top v_{i,m}(\Theta_m) \frac{\mathbb{I}\{y_i \le v_{i,m}(\Theta_m)\}}{-\tau e_{i,m}(\Theta_m)} - \frac{\nabla^\top v_{i,m}(\Theta_m)}{e_{i,m}(\Theta_m)} + \frac{v_{i,m}(\Theta_m)\nabla^\top e_{i,m}(\Theta_m)}{\tau e_{i,m}(\Theta_m)^2}\mathbb{I}\{y_i \le v_{i,m}(\Theta_m)\} \\ & - \frac{y_i \nabla^\top e_{i,m}(\Theta_m)}{\tau e_{i,m}(\Theta_m)^2}\mathbb{I}\{y_i \le v_{i,m}(\Theta_m)\} - \frac{v_i(\Theta_m)\nabla^\top e_{i,m}(\Theta_m)}{e_{i,m}(\Theta_m)^2} - \frac{\nabla^\top e_{i,m}(\Theta_m)}{e_{i,m}(\Theta_m)}.\end{aligned}$$

Since

$$\mathrm{Var}(\eta_{i,m}) = n\mathrm{Var}\{c_m \boldsymbol{g}_{i,m}(\Theta_s) - c_m \boldsymbol{g}_{i,m}(\Theta_m^*)\} \le n\,\mathbb{E}\left[\{c_m \boldsymbol{g}_{i,m}(\Theta_s) - c_m \boldsymbol{g}_{i,m}(\Theta_m^*)\}^2\right],$$

in the following, we will show that $\mathbb{E}\left[\{c_m \boldsymbol{g}_{i,m}(\Theta_s) - c_m \boldsymbol{g}_{i,m}(\Theta_m^*)\}^2\right] \le C\underline{c}_{A(m)}^{-1} k_m^{3/2} \|\Theta_s - \Theta_m^*\|$ by decomposing $\{c_m \boldsymbol{g}_{i,m}(\Theta_s) - c_m \boldsymbol{g}_{i,m}(\Theta_m^*)\}$ into six terms and showing that the expectation of the square of each term is bounded by $\underline{c}_{A(m)}^{-1} k_m^{3/2} \|\Theta_s - \Theta_m^*\|$.

To emphasize that the derivative of $v_{i,m}$ and $e_{i,m}$ are only related to $\boldsymbol{x}_{i,m}$, we denote $H_1(\boldsymbol{x}_{i,m}) \equiv \nabla v_{i,m}(\Theta_m)$ and $H_2(\boldsymbol{x}_{i,m}) \equiv \nabla e_{i,m}(\Theta_m)$, respectively. Note that $\|H_1(\boldsymbol{x}_{i,m})\| = \|\boldsymbol{x}_{i,m}\|$ and $\|H_2(\boldsymbol{x}_{i,m})\| = \|\boldsymbol{x}_{i,m}\|$.

First term:

$$\begin{aligned}|c_m \mu_{1,m}| \equiv& \left| c_m \left\{ \frac{\nabla^\top v_i(\Theta_s)}{e_i(\Theta_s)} - \frac{\nabla^\top v_i(\Theta_m^*)}{e_i(\Theta_m^*)} \right\} \right| = \left| c_m H_1^\top(\boldsymbol{x}_{i,m}) \left\{ \frac{1}{e_i(\Theta_s)} - \frac{1}{e_i(\Theta_m^*)} \right\} \right| \\ =& \left| c_m H_1^\top(\boldsymbol{x}_{i,m}) \left\{ \frac{e_i(\Theta_m^*) - e_i(\Theta_s)}{e_i(\Theta_s) e_i(\Theta_m^*)} \right\} \right| \le \left| c_m H_1^\top(\boldsymbol{x}_{i,m}) K_2^2 \{e_i(\Theta_m^*) - e_i(\Theta_s)\} \right| \\ =& \left| c_m H_1^\top(\boldsymbol{x}_{i,m}) \right| K_2^2 \left| \nabla^\top e_i\left(\tilde{\Theta}\right)(\Theta_s - \Theta_m^*) \right| \le \left| c_m H_1^\top(\boldsymbol{x}_{i,m}) \right| K_2^2 \|H_1(\boldsymbol{x}_{i,m})\| \|\Theta_s - \Theta_m^*\|\end{aligned}$$

where $\tilde{\Theta}$ lies between Θ_s and Θ_m^*.

On the other hand,

$$|c_m \mu_{1,m}| = \left| c_m \left\{ \frac{\nabla^\top v_i(\Theta_s)}{e_i(\Theta_s)} - \frac{\nabla^\top v_i(\Theta_m^*)}{e_i(\Theta_m^*)} \right\} \right| \le 2 \left| c_m H_1^\top(\boldsymbol{x}_{i,m}) \right| K_2.$$

Therefore,

$$\begin{aligned}\mathbb{E}\left\{ (c_m \mu_{1,m})^2 \right\} &\le \mathbb{E}\left\{ 2 c_m H_1^\top(\boldsymbol{x}_{i,m}) K_2 c_m H_1^\top(\boldsymbol{x}_{i,m}) K_2^2 \|H_1(\boldsymbol{x}_{i,m})\| \|\Theta_s - \Theta_m^*\| \right\} \\ &\le \mathbb{E}\left\{ 2\|c_m\|^2 \|H_1(\boldsymbol{x}_{i,m})\|^3 2K_2^3 \right\} \|\Theta_s - \Theta_m^*\| \\ &\le 4K_2^3\, \mathbb{E}\left(\|c_m\|^8 \right)^{1/4} \mathbb{E}\left\{ \|H_1(\boldsymbol{x}_{i,m})\|^4 \right\}^{3/4} \|\Theta_s - \Theta_m^*\| \le C\underline{c}_{A(m)}^{-1} k_m^{3/2} \|\Theta_s - \Theta_m^*\|,\end{aligned}$$

where C is a strictly positive constant.

Second term: Using a mean-value expansion around Θ_m^*, Hölder inequality and Assumption 3.1 we obtain:

$$\begin{aligned}|c_m \mu_{2,m}| \equiv& \left| \frac{c_m \nabla^\top v_{i,m}(\Theta_s)\mathbb{I}\{y_i \le v_{i,m}(\Theta_s)\}}{-\tau e_{i,m}(\Theta_s)} - \frac{c_m \nabla^\top v_{i,m}(\Theta_m^*)\mathbb{I}\{y_i \le v_{i,m}(\Theta_m^*)\}}{-\tau e_{i,m}(\Theta_m^*)} \right| \\ \le& \frac{1}{\tau} \left| c_m \frac{\nabla^\top v_{i,m}(\Theta_m^*)}{\tau e_{i,m}(\Theta_m^*)} \left| \mathbb{I}\{y_i \le v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \le v_{i,m}(\Theta_m^*)\} \right| \right| \\ & + \frac{1}{\tau} \left| c_m \left\{ \frac{\nabla^\top v_{i,m}(\Theta_s)}{e_{i,m}(\Theta_s)} - \frac{\nabla^\top v_{i,m}(\Theta_m^*)}{e_{i,m}(\Theta_m^*)} \right\} \right|\end{aligned}$$

$$
\begin{aligned}
&\leq \frac{1}{\tau}\left|c_m H_1^\top\left(\boldsymbol{x}_{i,m}\right)\right| K_2\left|\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\right| \\
&\qquad+\frac{1}{\tau}\left|c_m\left\{\frac{\nabla^\top v_{i,m}\left(\Theta_s\right)}{e_{i,m}\left(\Theta_s\right)}-\frac{\nabla^\top v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}\right\}\right|.
\end{aligned}
$$

Therefore

$$
\begin{aligned}
\mathbb{E}\left\{\left(c_m\mu_{2,m}\right)^2\right\} &\leq \frac{2}{\tau^2}\,\mathbb{E}\left[\left\|c_m\right\|^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|^2 K_2^2\left|\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\right|\right] \\
&\qquad+C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\| \\
&=\frac{2}{\tau^2}\,\mathbb{E}\left(\left\|c_m\right\|^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|^2 K_2^2\left|\mathbb{E}\left[\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\mid \boldsymbol{x}_i\right]\right|\right) \\
&\qquad+C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\| \\
&=\frac{2}{\tau^2}\,\mathbb{E}\left\{\left\|c_m\right\|^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|^2 K_2^2\left|\int_{v_{i,m}\left(\Theta_m^*\right)}^{v_{i,m}\left(\Theta_s\right)} f_{y|\boldsymbol{x}}(y)\,dy\right|\right\}+C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\| \\
&\leq \frac{2}{\tau^2}\,\mathbb{E}\left\{\left\|c_m\right\|^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|^2 K_2^2 c_f\left|v_{i,m}\left(\Theta_s\right)-v_{i,m}\left(\Theta_m^*\right)\right|\right\}+C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\| \\
&=\frac{2}{\tau^2}\,\mathbb{E}\left\{\left\|c_m\right\|^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|^2 K_2^2 c_f\left\|\nabla v_{i,m}\left(\Theta_1\right)\right\|\left\|\Theta_s-\Theta_m^*\right\|\right\}+C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\| \\
&\leq \frac{2}{\tau^2}c_f K_2^2\,\mathbb{E}\left(\left\|c_m\right\|^8\right)^{1/4}\mathbb{E}\left[\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|^4\right]^{3/4}\left\|\Theta_s-\Theta_m^*\right\|+C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\| \\
&\leq C_1\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\|,
\end{aligned}
$$

where Θ_1 lies between Θ_s and Θ_m^*, C and C_1 is a strictly positive constant. The penultimate inequality holds by the Hölder inequality.

Third term:

$$
\begin{aligned}
\left|c_m\mu_{3,m}\right| &\equiv \left|c_m\left\{\frac{v_{i,m}\left(\Theta_s\right)\nabla^\top e_{i,m}(\Theta)}{e_{i,m}\left(\Theta_s\right)^2}-\frac{v_{i,m}\left(\Theta_m^*\right)\nabla^\top e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\right| \\
&=\left|c_m H_2^\top\left(\boldsymbol{x}_{i,m}\right)\left\{\frac{v_{i,m}\left(\Theta_s\right)}{e_{i,m}\left(\Theta_s\right)^2}-\frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\right| \\
&\leq \left|2c_m H_2^\top\left(\boldsymbol{x}_{i,m}\right)\left\{\frac{1}{e_{i,m}\left(\Theta_s\right)}-\frac{1}{e_{i,m}\left(\Theta_m^*\right)}\right\}\right| \\
&=\left|2c_m H_2^\top\left(\boldsymbol{x}_{i,m}\right)\frac{\nabla e_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^2}\left(\Theta_s-\Theta_m^*\right)\right| \\
&\leq 2\left|c_m H_2^\top\left(\boldsymbol{x}_{i,m}\right)\right| K_2^2\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|\Theta_s-\Theta_m^*\right\|
\end{aligned}
$$

where $\tilde{\Theta}$ lies between Θ_s and Θ_m^*.

On the other hand,

$$
\begin{aligned}
\left|c_m\mu_{3,m}\right| &= c_m\left|\frac{v_{i,m}\left(\Theta_s\right)\nabla^\top e_{i,m}(\Theta)}{e_{i,m}\left(\Theta_s\right)^2}-\frac{v_{i,m}\left(\Theta_m^*\right)\nabla^\top e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right| \\
&=\left|c_m H_2^\top\left(\boldsymbol{x}_{i,m}\right)\left\{\frac{v_{i,m}\left(\Theta_s\right)}{e_{i,m}\left(\Theta_s\right)^2}-\frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\right|
\end{aligned}
$$

$$\leq \left|2c_m H_2^\top(\boldsymbol{x}_{i,m})\left\{\frac{1}{e_{i,m}(\Theta_s)} - \frac{1}{e_{i,m}(\Theta_m^*)}\right\}\right| \leq \left|4c_m H_2^\top(\boldsymbol{x}_{i,m}) K_2\right|.$$

Then, we have

$$\begin{aligned}
\mathbb{E}\left\{(c_m\mu_{3,m})^2\right\} &\leq 4\,\mathbb{E}\left\{\left|c_m H_2^\top(\boldsymbol{x}_{i,m}) K_2 c_m 2H_2^\top(\boldsymbol{x}_{i,m}) K_2^2 \left\|H_2(\boldsymbol{x}_{i,m})\right\| \left\|\Theta_s - \Theta_m^*\right\|\right|\right\} \\
&\leq 8K_2^3\,\mathbb{E}\left\{\|c_m\|^2 \left\|H_2(\boldsymbol{x}_{i,m})\right\|^3\right\} \left\|\Theta_s - \Theta_m^*\right\| \\
&\leq C\underline{c}_{A(m)}^{-1} k_m^{3/2} \left\|\Theta_s - \Theta_m^*\right\|,
\end{aligned}$$

where C is a strictly positive constant.

Fourth term:

$$\begin{aligned}
&|c_m\mu_{4,m}| \equiv \left|c_m\left[\frac{v_{i,m}(\Theta_s)\nabla^\top e_{i,m}(\Theta_s)}{\tau e_{i,m}(\Theta_s)^2}\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \frac{v_{i,m}(\Theta_m^*)\nabla^\top e_{i,m}(\Theta_m^*)}{\tau e_{i,m}(\Theta_m^*)^2}\mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}\right]\right| \\
&\leq \frac{1}{\tau}\left|c_m\frac{v_{i,m}(\Theta_m^*)\nabla^\top e_{i,m}(\Theta_m^*)}{\tau e_{i,m}(\Theta_m^*)^2}|\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}|\right| + \frac{|c_m\mu_{3,m}|}{\tau} \\
&\leq \frac{1}{\tau}\left|c_m\frac{\nabla^\top e_{i,m}(\Theta_m^*)}{\tau e_{i,m}(\Theta_m^*)}|\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}|\right| + \frac{|c_m\mu_{3,m}|}{\tau} \\
&\leq \frac{1}{\tau}\left|c_m H_2^\top(\boldsymbol{x}_{i,m})\right| K_2|\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}| + \frac{|c_m\mu_{3,m}|}{\tau}.
\end{aligned}$$

Then, we have

$$\begin{aligned}
&\mathbb{E}\left\{(c_m\mu_{4,m})^2\right\} \\
&\leq \frac{2}{\tau^2}\,\mathbb{E}\left\{\|c_m\|^2\left\|H_2(\boldsymbol{x}_{i,m})\right\|^2 K_2^2 c_f \left\|H_1(\boldsymbol{x}_{i,m})\right\| \left\|\Theta_s - \Theta_m^*\right\|\right\} + \frac{2}{\tau^2}\,\mathbb{E}\left[(c_m\mu_{3,m})^2\right] \\
&\leq \frac{1}{\tau^2}c_f K_2^2\,\mathbb{E}\left(\|c_m\|^8\right)^{1/4}\mathbb{E}\left(\|x_{i,m}\|^4\right)^{3/4}\left\|\Theta_s - \Theta_m^*\right\| + C\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s - \Theta_m^*\right\| \\
&\leq C_1\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s - \Theta_m^*\right\|,
\end{aligned}$$

where C and C_1 are strictly positive constants.

Fifth term:

$$\begin{aligned}
&|c_m\mu_{5,m}| \equiv \left|c_m\left[\frac{y_i\nabla^\top e_{i,m}(\Theta_s)}{\tau e_{i,m}(\Theta_s)^2}\mathbb{I}\{y_i \leq v_i(\Theta_s)\} - \frac{y_i\nabla^\top e_{i,m}(\Theta_m^*)}{\tau e_{i,m}(\Theta_m^*)^2}\mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}\right]\right| \\
&\leq \frac{1}{\tau}\left|c_m\frac{\nabla^\top e_{i,m}(\Theta_m^*)}{\tau e_{i,m}(\Theta_m^*)^2}\;|y_i[\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}]|\right| + \frac{|c_m\mu_{3,m}|}{\tau}|y_i\mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}| \\
&\leq \frac{1}{\tau}\left|c_m H_2^\top(\boldsymbol{x}_{i,m})\right| K_2^2\,|y_i[\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}]| + \frac{|c_m\mu_{3,m}|}{\tau}|y_i\mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}|.
\end{aligned}$$

Therefore,

$$\begin{aligned}
&\mathbb{E}\left\{(c_m\mu_{5,m})^2\right\} \\
\leq&\frac{2}{\tau^2}\,\mathbb{E}\left\{\|c_m\|^2\left\|H_2(\boldsymbol{x}_{i,m})\right\|^2 K_2^4\,\mathbb{E}\left(y_i^2\left[\mathbb{I}\{y_i \leq v_{i,m}(\Theta_s)\} - \mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}\right]|\boldsymbol{x}_i\right)\right\} \\
&\qquad + \frac{2}{\tau^2}\sqrt{\mathbb{E}\left[|y_i\mathbb{I}\{y_i \leq v_{i,m}(\Theta_m^*)\}|^4\right]\mathbb{E}\left[(c_m\mu_{3,m})^4\right]},
\end{aligned}$$

where

$$
\begin{aligned}
&\mathbb{E}\left(y_i^2\left[\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\right] \mid \boldsymbol{x}_i\right) \\
\leq &\mathbb{E}\left\{y_i^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|\Theta_s-\Theta_m^*\right\| c_f \mid \boldsymbol{x}_i\right\} \\
= &\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|\Theta_s-\Theta_m^*\right\| c_f \,\mathbb{E}\left(y_i^2 \mid \boldsymbol{x}_{i,m}\right)=C_1\left\|H_1\left(\boldsymbol{x}_i\right)\right\|\left\|\Theta_s-\Theta_m^*\right\| .
\end{aligned}
$$

Then, using $\mathbb{E}\left[\left|y_i \mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\right|^4\right]<\mathbb{E}\left(y_i^4\right)<C$ and $\mathbb{E}\left[\left(c_m \mu_{3,m}\right)^4\right] \leq C \underline{c}_{A(m)}^{-2} k_m^3\left\|\Theta_s-\Theta_m^*\right\|^2$, we have

$$
\begin{aligned}
\mathbb{E}\left\{\left(c_m \mu_{5,m}\right)^2\right\} \leq & C_1 \,\mathbb{E}\left\{\left\|c_m\right\|^2\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|^2\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|\right\}\left\|\Theta_s-\Theta_m^*\right\| \\
& +C_2 \underline{c}_{A(m)}^{-1} k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\|=C_3 \underline{c}_{A(m)}^{-1} k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\|,
\end{aligned}
$$

where C_2 and C_3 are strictly positive constants.

Sixth term:

Since

$$
\begin{aligned}
\left|c_m \mu_{6,m}\right| \equiv & \left|c_m\left\{\frac{\nabla^{\top} e_{i,m}(\Theta)}{e_{i,m}(\Theta)}-\frac{\nabla^{\top} e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)}\right\}\right| \\
= & \left|c_m H_2^{\top}\left(\boldsymbol{x}_{i,m}\right)\left\{\frac{1}{e_{i,m}(\Theta)}-\frac{1}{e_{i,m}\left(\Theta_m^*\right)}\right\}\right| \\
\leq & \left|c_m H_2^{\top}\left(\boldsymbol{x}_{i,m}\right) K_2^2\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|\Theta-\Theta_m^*\right\|\right|
\end{aligned}
$$

and

$$
\left|c_m \mu_{6,m}\right|=\left|c_m H_2^{\top}\left(\boldsymbol{x}_{i,m}\right)\left\{\frac{1}{e_{i,m}(\Theta)}-\frac{1}{e_{i,m}\left(\Theta_m^*\right)}\right\}\right| \leq\left|c_m H_2^{\top}\left(\boldsymbol{x}_{i,m}\right) 2 K_2\right| .
$$

We have

$$
\begin{aligned}
\mathbb{E}\left\{\left(c_m \mu_{6,m}\right)^2\right\} & \leq \mathbb{E}\left\{\left|c_m H_2^{\top}\left(\boldsymbol{x}_{i,m}\right) 2 K_2 c_m H_2^{\top}\left(\boldsymbol{x}_{i,m}\right) K_2^2\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|\Theta-\Theta_m^*\right\|\right|\right\} \\
& \leq \mathbb{E}\left\{\left\|c_m\right\|^2\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|^3 2 K_2^3\right\}\left\|\Theta_s-\Theta_m^*\right\| \\
& \leq C \underline{c}_{A(m)}^{-1} k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\|,
\end{aligned}
$$

where C is a strictly positive constant.

Thus, we have

$$
\mathbb{E}\left[\left\{c_m \boldsymbol{g}_{i,m}\left(\Theta_s\right)-c_m \boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}^2\right] \leq 6 \sum_{i=1}^6 \mathbb{E}\left\{\left(c_m \mu_{i,m}\right)^2\right\} \leq C \underline{c}_{A(m)}^{-1} k_m^{3/2}\left\|\Theta_s-\Theta_m^*\right\|
$$

and

$$
\begin{aligned}
& \left\|c_m \boldsymbol{g}_{i,m}\left(\Theta_s\right)-c_m \boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\| \\
\leqslant & C_1 \underline{c}_{A(m)}^{-1/2}\left\|\boldsymbol{x}_{i,m}\right\|^2\left\|\Theta_s-\Theta_m^*\right\|+C_2 \underline{c}_{A(m)}^{-1/2}\left\|\boldsymbol{x}_{i,m}\right\|\left|\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\right| \\
& +C_3 \underline{c}_{A(m)}^{-1/2}\left\|\boldsymbol{x}_{i,m}\right\|\left|y_i\left[\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i \leq v_{i,m}\left(\Theta_m^*\right)\right\}\right]\right| \qquad \text{(S.11)}
\end{aligned}
$$

for some constant C, C_1, C_2 and C_3, which imply that

$$
\operatorname{Var}\left(\eta_{i,m,s}\right)=n^{-1} \operatorname{Var}\left\{c_m \boldsymbol{g}_{i,m}\left(\Theta_s\right)-c_m \boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}
$$

$$\leq n^{-1}\,\mathbb{E}\left[\left\{c_m \boldsymbol{g}_{i,m}\left(\Theta_s\right) - c_m \boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}^2\right]$$

$$\leq 6n^{-1}\sum_{i=1}^{6}\mathbb{E}\left\{\left(c_m\mu_{i,m}\right)^2\right\} \leq Cn^{-1}\underline{c}_{A(m)}^{-1}k_m^{3/2}\left\|\Theta_s - \Theta_m^*\right\| \leq C_1\underline{c}_{A(m)}^{-1}k_m^2 n^{-3/2},$$

and

$$\left\|\mathbb{E}\left\{n^{-1/2}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\} - \mathbb{E}\left\{n^{-1/2}\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\right\|_{c_m}$$

$$\leqslant n^{-1/2}Cc_{A(m)}^{-1/2}\sqrt{k_m^{3/2}\left\|\Theta_s - \Theta_m^*\right\|} = n^{-1/2}Cc_{A(m)}^{-1/2}k_m^{3/4}k_m^{1/4}n^{1/4} = n^{-1/4}Cc_{A(m)}^{-1/2}k_m,$$

for some constant C and C_1. Let $e_{1n} = n^{1/8}$, then there are constants C_3 and C_4 such that

$$\left\|\sqrt{n}\left[\boldsymbol{g}_{i,m}\left(\Theta_s\right) - \boldsymbol{g}_{i,m}\left(\Theta_m^*\right) - \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\} + \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\right]\right.$$

$$\left.\times\,\mathbb{I}\left(\left\|\boldsymbol{x}_{i,m}\right\| \leqslant e_{1n}, \left|y_i\right| \leqslant e_{1n}\right)\right\|_{c_m}$$

$$\leqslant C\sqrt{n}\underline{c}_{A(m)}^{-1/2}\left(e_{1n}^2 k_m^{1/2}n^{-1/2} + e_{1n} + e_{1n}^2\right).$$

By Bernstein's inequality (see, e.g., Serfling, 1980, page 95), we have

$$\mathbb{P}\left[\max_{1\leq s\leq N_1}\left\|\frac{1}{n}\sum\nolimits_{i=1}^{n}\sqrt{n}\left[\boldsymbol{g}_{i,m}\left(\Theta_s\right) - \boldsymbol{g}_{i,m}\left(\Theta_m^*\right) - \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\} + \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\right]\right.\right.$$

$$\left.\left.\times\,\mathbb{I}\left(\left\|\boldsymbol{x}_{i,m}\right\| \leqslant e_{1n}, \left|y_i\right| \leqslant e_{1n}\right)\right\|_{c_m} > \varepsilon\right]$$

$$= \mathbb{P}\left(\max_{1\leq s\leq N_1}\left|\frac{1}{n}\sum_{i=1}^{n}\eta_{i,m,s}\mathbb{I}\left(\left\|\boldsymbol{x}_{i,m}\right\| \leqslant e_{1n}, \left|y_i\right| \leqslant e_{1n}\right)\right| > \varepsilon\right)$$

$$\leq N_1\max_{1\leq s\leq N_1}\mathbb{P}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\eta_{i,m,s}\mathbb{I}\left(\left\|\boldsymbol{x}_{i,m}\right\| \leqslant e_{1n}, \left|y_i\right| \leqslant e_{1n}\right)\right| > \varepsilon\right)$$

$$\leq 2N_1\exp\left(-\frac{n^2\varepsilon^2}{2nC_1\underline{c}_{A(m)}^{-1}k_m^2 n^{-3/2} + 2C_2 n^{3/2}\underline{c}_{A(m)}^{-1/2}\left(n^{1/4}k_m^{1/2}n^{-1/2} + n^{1/8} + n^{1/4}\right)/3}\right)$$

$$= 2\exp\left(2k_m\log 2n\right)\exp\left(-\frac{n^2\varepsilon^2}{2n^{-1/2}C_1\underline{c}_{A(m)}^{-1}k_m^2 + 2C_2\underline{c}_{A(m)}^{-1/2}\left(n^{5/4}k_m^{1/2} + n^{13/8} + n^{7/4}\right)/3}\right).$$

Assumption 3.2(iii) $\log\left(n\right)k_m^3/n^{5/2}\underline{c}_{A(m)} \to 0$ implies $\log\left(n\right)^{1/2}k_m^{3/2}/n^{5/4}\underline{c}_{A(m)}^{1/2} \to 0$. Then $\frac{n^{-5/2}\underline{c}_{A(m)}^{-1}k_m^2}{\underline{c}_{A(m)}^{-1/2}n^{-3/4}k_m^{1/2}} = \frac{k_m^{3/2}}{\underline{c}_{A(m)}^{1/2}n^{7/4}} = \frac{\log(n)^{1/2}k_m^{3/2}}{n^{5/4}\underline{c}_{A(m)}^{1/2}}\frac{1}{\log(n)^{1/2}n^{1/2}} \to 0$. Combining $k_m^4 n^{-3} \to 0$, we have

$$2\exp\left(2k_m\log 2n\right)\exp\left(-\frac{n^2\varepsilon^2}{2n^{-1/2}C_1\underline{c}_{A(m)}^{-1}k_m^2 + 2C_2\underline{c}_{A(m)}^{-1/2}\left(n^{5/4}k_m^{1/2} + n^{13/8} + n^{7/4}\right)/3}\right)$$

$$\leq 2C_3\exp\left(2k_m\log 2n\right)\exp\left(-\frac{n^{5/2}}{\underline{c}_{A(m)}^{-1}k_m^2}\right).$$

Thus, combining $\log\left(n\right)k_m^3/n^{5/2}\underline{c}_{A(m)} \to 0$, we have

$$\mathbb{P}\left[\max_{1\leq s\leq N_1}\left\|\frac{1}{n}\sum\nolimits_{i=1}^{n}\sqrt{n}\left[\boldsymbol{g}_{i,m}\left(\Theta_s\right) - \boldsymbol{g}_{i,m}\left(\Theta_m^*\right) - \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\} + \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\right]\right.\right.$$

$$\times \mathbb{I}\left(\|\boldsymbol{x}_{i,m}\| \leqslant e_{1n}, |y_i| \leqslant e_{1n}\right)\|_{c_m} > \varepsilon\Big] \to 0.$$

Furthermore, by the dominated convergence theorem,

$$\begin{aligned}
&\mathbb{P}\left[\max_{1\leq s\leq N_1}\left\|\frac{1}{n}\sum\nolimits_{i=1}^{n}\sqrt{n}\left[\boldsymbol{g}_{i,m}\left(\Theta_s\right)-\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)-\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\}+\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\right]\right.\right.\\
&\qquad \left.\left.\times\left[1-\mathbb{I}\left(\|\boldsymbol{x}_{i,m}\|\leqslant e_{1n},|y_i|\leqslant e_{1n}\right)\right]\right\|_{c_m}>\varepsilon\right]\\
&\leqslant \mathbb{P}\left(\max_{1\leq i\leq m_1}\|\boldsymbol{x}_{i,m}\|>e_{1n}\right)+\mathbb{P}\left(\max_{1\leq i\leq m_1}\|y_i\|>e_{1n}\right)\\
&\leqslant n\,\mathbb{P}\left(\|\boldsymbol{x}_{i,m}\|>e_{1n}\right)+n\,\mathbb{P}\left(\|y_i\|>e_{1n}\right)\\
&\leqslant \frac{n\,\mathbb{E}\left[\|\boldsymbol{x}_{i,m}\|^8\,\mathbb{I}\left(\|\boldsymbol{x}_{i,m}\|>e_{1n}\right)\right]}{e_{1n}^8}+\frac{n\,\mathbb{E}\left[\|y_i\|^8\,\mathbb{I}\left(\|y_i\|>e_{1n}\right)\right]}{e_{1n}^8}\\
&\to 0.
\end{aligned}$$

This means that the first term of (S.10) is $o_p(1)$.

Next, we verify the second term of (S.10). According to the definition of D_s, $\|\Theta-\Theta_s\| \le L\frac{k_m}{n^{3/2}}$ for all $\Theta \in D_s$. Therefore,

$$\begin{aligned}
&\max_{1\leq s\leq N_1}\sup_{\Theta\in D_s}\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right.\\
&\qquad\qquad \left.-\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)\right\}+\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\}\right\|_{c_m}\\
\leq&\max_{1\leq s\leq N_1}\sup_{\Theta\in D_s}\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\|_{c_m}\\
&\qquad\qquad +\max_{1\leq s\leq N_1}\sup_{\Theta\in D_s}\left\|\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\}\right\|_{c_m}.
\end{aligned}$$

Then, using the arguments as (S.11), we have

$$\begin{aligned}
&\|c_m\boldsymbol{g}_{i,m}\left(\Theta_s\right)-c_m\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\|\\
&\leqslant C_1\underline{c}_{A(m)}^{-1/2}\|\boldsymbol{x}_{i,m}\|^2\|\Theta_s-\Theta_m^*\|+C_2\underline{c}_{A(m)}^{-1/2}\|\boldsymbol{x}_{i,m}\|\left|\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}\right|\\
&\qquad +C_3\underline{c}_{A(m)}^{-1/2}\|\boldsymbol{x}_{i,m}\|\left|y_i\left[\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_s\right)\right\}-\mathbb{I}\left\{y_i\leq v_{i,m}\left(\Theta_m^*\right)\right\}\right]\right|,\\
&\|\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\}-\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}\|_{c_m}\leqslant Cc_{A(m)}^{-1/2}\sqrt{k_m^{3/2}\|\Theta_s-\Theta_m^*\|}=Cc_{A(m)}^{-1/2}k_m^{5/4}n^{-3/4},
\end{aligned}$$

and

$$\mathbb{E}\left\{\sup_{\Theta\in D_s}\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\|_{c_m}\right\}\leqslant Cc_{A(m)}^{-1/2}k_m^{5/4}n^{-3/4}.$$

Thus, using the same steps as Bernstein's inequality on the previous page, we can prove that

$$\begin{aligned}
&\max_{1\leq s\leq N_1}\sup_{\Theta\in D_s}\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right.\\
&\qquad\qquad \left.-\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)\right\}+\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\}\right\|_{c_m}=o_p(1).
\end{aligned}$$

Furthermore,

$$\mathbb{P}\left[\max_{1\le s\le N_1}\sup_{\Theta\in D_s}\left\|\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\}\right\|_{c_m}>\varepsilon\right]$$
$$\le\mathbb{P}\left(Cn^{1/2}\underline{c}_{A(m)}^{-1}k_m\left\|\Theta-\Theta_s\right\|>\varepsilon\right)\to 0.$$
$$\mathbb{P}\left[\max_{1\le s\le N_1}\mathbb{E}\left\{\sup_{\Theta\in D_s}\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta\right)-n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\Theta_s\right)\right\|_{c_m}\right\}>\varepsilon\right]$$
$$\le\mathbb{P}\left(Cn^{1/2}\underline{c}_{A(m)}^{-1}k_m\left\|\Theta-\Theta_s\right\|>\varepsilon\right)\to 0.$$

This implies that the second term of (S.10) is $o_p(1)$. Consequently (S.7) holds.

Next, we will show (S.8):

$$\sup_{\|\Delta\|\le\sqrt{k_m L}}\left\|\bar{V}_m\left(\Delta\right)-\bar{V}_m\left(0\right)+\mathbf{D}_m\Delta\right\|_{c_m}=o_p\left(1\right).$$

Consider a mean-value expansion of $E\{\boldsymbol{g}_{i,m}(\Delta)\}$ around Δ_m^*:

$$\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta\right)\right\}-\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}=\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)\left(\Theta-\Theta_m^*\right),$$

where $\tilde{\Theta}_i$ lies between Θ and Θ_m^*. Then, we will prove $\left\|c_m\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)-c_m\mathbf{D}_m\right\|=\underline{c}_{A(m)}^{-1/2}O\left(\left\|\tilde{\Theta}_i-\Theta_m^*\right\|\right)$ by decomposing $\left\|c_m\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)-c_m\mathbf{D}_m\right\|$ into five terms and showing that each is bounded by a $\underline{c}_{A(m)}^{-1/2}O\left(\left\|\tilde{\Theta}_i-\Theta_m^*\right\|\right)$ term.

First term:

$$\Bigg\|c_m\,\mathbb{E}\left[\left\{\frac{\nabla^2 v_{i,m}\left(\tilde{\Theta}\right)}{-e_{i,m}\left(\tilde{\Theta}\right)}+\frac{\nabla v_{i,m}\left(\tilde{\Theta}\right)^{\top}\nabla e_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^2}+\frac{\nabla e_{i,m}\left(\tilde{\Theta}\right)^{\top}\nabla v_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^2}\right\}\left\{\frac{F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)\right)}{\tau}-1\right\}\right]$$
$$-c_m\,\mathbb{E}\left[\left\{\frac{\nabla^2 v_{i,m}\left(\Theta_m^*\right)}{-e_{i,m}\left(\Theta_m^*\right)}+\frac{\nabla v_{i,m}\left(\Theta_m^*\right)^{\top}\nabla e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}+\frac{\nabla e_{i,m}\left(\Theta_m^*\right)^{\top}\nabla v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\left\{\frac{F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right)}{\tau}-1\right\}\right]\Bigg\|$$
$$\le\left(\frac{1}{\tau}\right)\Bigg\|c_m\,\mathbb{E}\left\{\frac{\nabla v_{i,m}\left(\tilde{\Theta}\right)^{\top}\nabla e_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^2}-\frac{\nabla v_{i,m}\left(\Theta_m^*\right)^{\top}\nabla e_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}$$
$$+c_m\,\mathbb{E}\left\{\frac{\nabla e_{i,m}\left(\tilde{\Theta}\right)^{\top}\nabla v_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^2}-\frac{\nabla e_{i,m}\left(\Theta_m^*\right)^{\top}\nabla v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\Bigg\|$$
$$=\left(\frac{1}{\tau}\right)\left\|c_m\,\mathbb{E}\left\{\frac{H_2^{\top}(\boldsymbol{x}_{i,m})H_1(\boldsymbol{x}_{i,m})}{e_{i,m}\left(\tilde{\Theta}\right)^2}-\frac{H_2^{\top}(\boldsymbol{x}_{i,m})H_1(\boldsymbol{x}_{i,m})}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\right\|$$
$$+\left(\frac{1}{\tau}\right)\left\|c_m\,\mathbb{E}\left\{\frac{H_1^{\top}(\boldsymbol{x}_{i,m})H_2(\boldsymbol{x}_{i,m})}{e_{i,m}\left(\tilde{\Theta}\right)^2}-\frac{H_1^{\top}(\boldsymbol{x}_{i,m})H_2(\boldsymbol{x}_{i,m})}{e_{i,m}\left(\Theta_m^*\right)^2}\right\}\right\|$$
$$=\left(\frac{2}{\tau}\right)\left\|\mathbb{E}\left\{\frac{c_m H_2^{\top}(\boldsymbol{x}_{i,m})H_1(\boldsymbol{x}_{i,m})}{e_{i,m}\left(\Theta^1\right)^3}H_2(x_{i,m})\left(\tilde{\Theta}-\Theta_m^*\right)\right\}\right\|$$
$$+\left(\frac{2}{\tau}\right)\left\|\mathbb{E}\left\{\frac{c_m H_1^{\top}(\boldsymbol{x}_{i,m})H_2(\boldsymbol{x}_{i,m})}{e_{i,m}\left(\Theta^2\right)^3}H_2(x_{i,m})\left(\tilde{\Theta}-\Theta_m^*\right)\right\}\right\|$$

$$
\begin{aligned}
&= \left(\frac{4}{\tau}\right) K_2^3 \, \mathbb{E}\left\{\left\|c_m H_2^\top(\boldsymbol{x}_{i,m}) H_1(\boldsymbol{x}_{i,m})\right\| \|H_2(\boldsymbol{x}_{i,m})\|\right\} \left\|\left(\tilde{\Theta} - \Theta_m^*\right)\right\| \\
&\leq \left(\frac{4}{\tau}\right) K_2^3 \, \mathbb{E}\left\{\|c_m\| \left\|H_2^\top(\boldsymbol{x}_{i,m})\right\| \|H_1(\boldsymbol{x}_{i,m})\| \|H_2(\boldsymbol{x}_{i,m})\|\right\} \left\|\left(\tilde{\Theta} - \Theta_m^*\right)\right\| \\
&= \left(\frac{4}{\tau}\right) K_2^3 \, \mathbb{E}\left\{\|c_m\| \|H_1(\boldsymbol{x}_{i,m})\|^3\right\} \left\|\left(\tilde{\Theta} - \Theta_m^*\right)\right\| \\
&\leq C \underline{c}_{A(m)}^{-1/2} k_m^{3/2} \left\|\left(\tilde{\Theta} - \Theta_m^*\right)\right\|,
\end{aligned}
$$

where Θ^1 and Θ^2 lie between $\tilde{\Theta}$ and Θ_m^*, and $\left|F_{y|\boldsymbol{x}}(x)/\tau - 1\right| < 1/\tau$ holds for all x and τ.

Second term: Since

$$
\begin{aligned}
&\mathbb{E}\left[\left\{\nabla^2 e_{i,m}(\Theta) \frac{1}{e_{i,m}(\Theta)^2} - \frac{2\nabla e_{i,m}(\Theta)^\top \nabla e_{i,m}(\Theta)}{e_{i,m}(\Theta)^3}\right\}\left\{-\frac{1}{\tau}\mathbb{E}\left[y_i \mathbb{I}\left\{y_i \leq v_{i,m}(\Theta)\right\} | \boldsymbol{x}_i\right]\right\}\right] \\
&= \mathbb{E}\left[\frac{2H_2^\top(\boldsymbol{x}_{i,m}) H_2(\boldsymbol{x}_{i,m})}{\tau e_{i,m}(\Theta)^3} y_i \mathbb{I}\left\{y_i \leq v_{i,m}(\Theta)\right\}\right]
\end{aligned}
$$

and $\frac{\partial \mathbb{E}(y_i \mathbb{I}(y_i \leq v_{i,m}(\Theta)) | \boldsymbol{x}_i)}{\partial \Theta} = \frac{\partial}{\partial \Theta} \int_{-\infty}^{v_{i,m}(\Theta)} y f_{y|\boldsymbol{x}}(y) \, dy = v_{i,m}(\Theta) f_{y|\boldsymbol{x}}(v_{i,m}(\Theta)) \nabla v_{i,m}(\Theta)$, we have

$$
\begin{aligned}
&\left\| c_m \mathbb{E}\left[2H_2^\top(\boldsymbol{x}_{i,m}) H_2(\boldsymbol{x}_{i,m}) \frac{\mathbb{E}\left[y_i \mathbb{I}\left\{y_i \leq v_{i,m}\left(\tilde{\Theta}\right)\right\} | \boldsymbol{x}_i\right]}{\tau e_{i,m}\left(\tilde{\Theta}\right)^3} \right.\right. \\
&\qquad \left.\left. - 2H_2^\top(\boldsymbol{x}_{i,m}) H_2(\boldsymbol{x}_{i,m}) \frac{y \, \mathbb{E}\left[y_i \mathbb{I}\left\{y_i \leq v_{i,m}(\Theta_m^*)\right\} | \boldsymbol{x}_i\right]}{\tau e_{i,m}(\Theta_m^*)^3}\right]\right\| \\
&= \left\| c_m \mathbb{E}\left[2H_2^\top(\boldsymbol{x}_{i,m}) H_2(\boldsymbol{x}_{i,m}) \frac{v_{i,m}(\Theta^1) f_{y|\boldsymbol{x}_i}\left(v_{i,m}(\Theta^1)\right) \nabla v_{i,m}(\Theta^1)}{\tau e_{i,m}(\Theta^1)^3} \right.\right. \\
&\qquad \left.\left. - \frac{3\nabla e_{i,m}(\Theta^1) \mathbb{E}\left[y_i \mathbb{I}\left\{y_i \leq v_{i,m}(\Theta^1)\right\} | \boldsymbol{x}_i\right]}{\tau e_{i,m}(\Theta^1)^4}\right\}\left(\tilde{\Theta} - \Theta_m^*\right) \right. \\
&\leq C \underline{c}_{A(m)}^{-1/2} k_m^{3/2} \left\|\tilde{\Theta} - \Theta_m^*\right\| + C \underline{c}_{A(m)}^{-1/2} k_m^{3/2} \left\|\tilde{\Theta} - \Theta_m^*\right\| \sqrt{\mathbb{E}\left[y_i^2 \mathbb{I}\left\{y_i \leq v_{i,m}(\Theta^1)\right\}\right]} \\
&= C \underline{c}_{A(m)}^{-1/2} k_m^{3/2} \left\|\tilde{\Theta} - \Theta_m^*\right\|,
\end{aligned}
$$

where Θ^1 lies between $\tilde{\Theta}$ and Θ_m^*.

Third term:

Since

$$
\begin{aligned}
&\mathbb{E}\left\{\left\{\nabla^2 e_{i,m}(\Theta) \frac{1}{e_{i,m}(\Theta)^2} - \frac{2}{e_{i,m}(\Theta)^3} \nabla e_{i,m}(\Theta)^\top \nabla e_{i,m}(\Theta)\right\} \right. \\
&\qquad \left. \times \left(\left\{\frac{F_{y|\boldsymbol{x}}(v_{i,m}(\Theta))}{\tau} - 1\right\} v_{i,m}(\Theta) + e_{i,m}(\Theta)\right)\right\} \\
&= \mathbb{E}\left[\frac{2H_2^\top(\boldsymbol{x}_{i,m}) H_2(\boldsymbol{x}_{i,m})}{e_{i,m}(\Theta)^3}\left(\left\{1 - \frac{F_{y|\boldsymbol{x}}(v_{i,m}(\Theta))}{\tau}\right\} v_{i,m}(\Theta) - e_{i,m}(\Theta)\right)\right],
\end{aligned}
$$

we have

$$
\left\| c_m \mathbb{E}\left[2H_2^\top(\boldsymbol{x}_{i,m}) H_2(\boldsymbol{x}_{i,m})\left(\left\{1 - \frac{F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)\right)}{\tau}\right\} \frac{v_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^3} - \frac{1}{e_{i,m}\left(\tilde{\Theta}\right)^2}\right]\right\|
$$

$$
- \left\{1 - \frac{F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right)}{\tau}\right\} \frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^3} + \frac{1}{e_{i,m}\left(\Theta_m^*\right)^2}\Bigg)\Bigg]\Bigg\|
$$

$$
\begin{aligned}
&\leqslant \left\| c_m \mathbb{E}\left[\frac{2H_2^\top\left(\boldsymbol{x}_{i,m}\right) H_2\left(\boldsymbol{x}_{i,m}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^3}\left(\frac{1}{e_{i,m}\left(\tilde{\Theta}\right)^2} - \frac{1}{e_{i,m}\left(\tilde{\Theta}\right)^2}\right)\right]\right\| \\
&+ \left\| c_m \mathbb{E}\left[2H_2^\top\left(\boldsymbol{x}_{i,m}\right) H_2\left(\boldsymbol{x}_{i,m}\right)\left(\left\{1 - \frac{F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)\right)}{\tau}\right\}\left\{\frac{v_{i,m}\left(\tilde{\Theta}\right)}{e_{i,m}\left(\tilde{\Theta}\right)^3} - \frac{v_{i,m}\left(\Theta_m^*\right)}{e_{i,m}\left(\Theta_m^*\right)^3}\right\}\right)\right]\right\| \\
&+ \left\| c_m \mathbb{E}\left[2H_2^\top\left(\boldsymbol{x}_{i,m}\right) H_2\left(\boldsymbol{x}_{i,m}\right)\left(\frac{v_{i,m}\left(\Theta_m^*\right)}{\tau e_{i,m}\left(\Theta_m^*\right)^3}\left\{F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)\right) - F_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)\right)\right\}\right)\right]\right\| \\
&\leq C \underline{c}_{A(m)}^{-1/2} k_m^{3/2}\left\|\tilde{\Theta} - \Theta_m^*\right\|.
\end{aligned}
$$

Fourth term:

$$
\begin{aligned}
&\left\| c_m \mathbb{E}\left\{\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\tilde{\Theta}\right)} \nabla v_{i,m}\left(\tilde{\Theta}\right)^\top \nabla v_{i,m}\left(\tilde{\Theta}\right) - \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} \nabla v_{i,m}\left(\Theta_m^*\right)^\top \nabla v_{i,m}\left(\Theta_m^*\right)\right\}\right\| \\
=& \Bigg\| c_m \mathbb{E}\Bigg\{\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\tilde{\Theta}\right)} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right) - \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right) \\
&\quad + \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right) - \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} \nabla H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right)\Bigg\}\Bigg\| \\
=& \left\| c_m \mathbb{E}\left\{\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\tilde{\Theta}\right)} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right) - \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right)\right\}\right\| \\
&\quad + \left\| c_m \mathbb{E}\left\{\frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\tilde{\Theta}\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right) - \frac{f_{y|\boldsymbol{x}}\left(v_{i,m}\left(\Theta_m^*\right)|\boldsymbol{x}_i\right)}{-\tau e_{i,m}\left(\Theta_m^*\right)} \nabla H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right)\right\}\right\| \\
\leq& 2\left\| c_m \mathbb{E}\left\{\frac{c_f K_2^2}{\tau} H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right)\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|\tilde{\Theta} - \Theta_m^*\right\|\right\}\right\| \\
&\quad + 2\left\| c_m \mathbb{E}\left\{\frac{K_2 H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right) c_f\left|v_{i,m}\left(\tilde{\Theta}\right) - v_{i,m}\left(\Theta_m^*\right)\right|}{\tau}\right\}\right\| \\
\leq& \Bigg[2\,\mathbb{E}\left\{\frac{c_f K_2^2}{\tau}\left\|c_m\right\|\left\|H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|\right\} \\
&\quad + 2\,\mathbb{E}\left\{\frac{c_f K_2\left\|c_m\right\|\left\|H_1^\top\left(\boldsymbol{x}_{i,m}\right) H_1\left(\boldsymbol{x}_{i,m}\right)\right\|\left\|H_1\left(\boldsymbol{x}_{i,m}\right)\right\|}{\tau}\right\}\Bigg]\left\|\tilde{\Theta} - \Theta_m^*\right\| \\
&\leq C \underline{c}_{A(m)}^{-1/2} k_m^{3/2}\left\|\left(\tilde{\Theta} - \Theta_m^*\right)\right\|.
\end{aligned}
$$

Fourth term:

$$\left\|c_m\,\mathbb{E}\left\{\frac{1}{e_{i,m}\left(\tilde{\Theta}\right)^2}\nabla e_{i,m}\left(\tilde{\Theta}\right)^\top\nabla e_{i,m}\left(\tilde{\Theta}\right)-\frac{1}{e_{i,m}\left(\Theta_m^*\right)^2}\nabla e_{i,m}\left(\Theta_m^*\right)^\top\nabla e_{i,m}\left(\Theta_m^*\right)\right\}\right\|$$
$$=\left\|c_m\,\mathbb{E}\left\{\frac{1}{e_{i,m}\left(\tilde{\Theta}\right)^2}H_2^\top\left(\boldsymbol{x}_{i,m}\right)H_2\left(\boldsymbol{x}_{i,m}\right)-\frac{1}{e_{i,m}\left(\Theta_m^*\right)^2}H_2^\top\left(\boldsymbol{x}_{i,m}\right)H_2\left(\boldsymbol{x}_{i,m}\right)\right\}\right\|$$
$$\leq\mathbb{E}\left\{\|c_m\|\left\|H_2^\top\left(\boldsymbol{x}_{i,m}\right)H_2\left(\boldsymbol{x}_{i,m}\right)\right\|2K_2^3\left\|H_2\left(\boldsymbol{x}_{i,m}\right)\right\|\right\}\left\|\tilde{\Theta}-\Theta_m^*\right\|$$
$$\leq C\underline{c}_{A(m)}^{-1/2}k_m^{3/2}\left\|\left(\tilde{\Theta}-\Theta_m^*\right)\right\|.$$

Thus, we have

$$\left\|c_m\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)-c_m\mathbf{D}_m\right\|\leq\sum\nolimits_{i=1}^n\text{each term}\leq C\underline{c}_{A(m)}^{-1/2}k_m^{3/2}\left\|\left(\tilde{\Theta}-\Theta_m^*\right)\right\|\leq C\underline{c}_{A(m)}^{-1/2}k_m^{3/2}\left\|\left(\Theta-\Theta_m^*\right)\right\|.$$

Therefore, (S.8) hold by

$$\left\|\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^n\boldsymbol{g}_{i,m}\left(\Theta\right)\right\}-\mathbb{E}\left\{n^{-1/2}\sum\nolimits_{i=1}^n\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}-\sqrt{n}\mathbf{D}_m\left(\Theta-\Theta_m^*\right)\right\|_{c_m}$$
$$=\left|n^{-1/2}c_m\sum\nolimits_{i=1}^n\left[\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta\right)\right\}-\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\}-\mathbf{D}_m\left(\Theta-\Theta_m^*\right)\right]\right|$$
$$=\left|n^{-1/2}c_m\sum\nolimits_{i=1}^n\left\{\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)\left(\Theta-\Theta_m^*\right)-\mathbf{D}_m\left(\Theta-\Theta_m^*\right)\right\}\right|$$
$$=\left|n^{-1/2}\sum\nolimits_{i=1}^n\left[\left\{c_m\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)-c_m\mathbf{D}_m\right\}\left(\Theta-\Theta_m^*\right)\right]\right|$$
$$\leq\left|n^{-1/2}\sum\nolimits_{i=1}^n\left\{\left\|c_m\boldsymbol{\Lambda}_m\left(\tilde{\Theta}_i\right)-c_m\mathbf{D}_m\right\|\left\|\Theta-\Theta_m^*\right\|\right\}\right|$$
$$\leq\left|n^{-1/2}\sum\nolimits_{i=1}^n\left\{C\underline{c}_{A(m)}^{-1/2}k_m^{3/2}\left\|\Theta-\Theta_m^*\right\|^2\right\}\right|$$
$$=C\underline{c}_{A(m)}^{-1/2}k_m^{5/2}n^{-1/2}=o\left(\underline{c}_{A(m)}^{-1}k_m^{5/2}n^{-1/2}\right)\to 0,$$

for all $\|\Theta-\Theta_m^*\|\leq\sqrt{k_mL/n}$.

Let $\{b_j\}_{j=1}^{2k_m}$ be the standard basis of $\mathbb{R}^{2k_m}$ and define

$$L_n^j\left(a\right)=n^{-1/2}\sum_{i=1}^n L\left(y_t,v_{i,m}\left(\widehat{\Theta}_m+ab_j\right),e_{i,m}\left(\widehat{\Theta}_m+ab_j\right)\right),$$

where a is a scalar. And let G_n^j (a scalar) be the right partial derivative of $L_n^j\left(a\right)$, that is

$$G_n^j\left(a\right)=T^{-1/2}\sum_{i=1}^n\left(\frac{\nabla_j v_{i,m}\left(\widehat{\Theta}_m+ab_j\right)}{-e_{i,m}\left(\widehat{\Theta}_m+ab_j\right)}\left[\frac{1}{\tau}\mathbb{I}\left\{y_t\leq\nabla_j v_{i,m}\left(\widehat{\Theta}_m+ab_j\right)\right\}-1\right]\right.$$
$$+\frac{\nabla_j e_{i,m}\left(\widehat{\Theta}_m+ab_j\right)}{e_{i,m}\left(\widehat{\Theta}_m+ab_j\right)^2}\left[\frac{1}{\tau}\mathbb{I}\left\{y_t\leq\nabla_j v_{i,m}\left(\widehat{\Theta}_m+ab_j\right)\right\}\right]$$
$$\left.\times\left\{v_{i,m}\left(\widehat{\Theta}_m+ab_j\right)-y_t\right\}-v_{i,m}\left(\widehat{\Theta}_m+ab_j\right)+e_{i,m}\left(\widehat{\Theta}_m+ab_j\right)\right).$$

For (S.9), following the proof of Lemma 2 in Patton et al. (2019), we have[2]

$$
\begin{aligned}
&\mathbb{P}\left\{c_m n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\widehat{\Theta}_m\right)\geq\varepsilon\right\}\\
\leq&\,\mathbb{P}\left\{\|c_m\|\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\widehat{\Theta}_m\right)\right\|\geq\varepsilon\right\}\\
\leq&\,\mathbb{P}\left\{\|c_m\|\left\|n^{-1/2}\sum\nolimits_{i=1}^{n}\boldsymbol{g}_{i,m}\left(\widehat{\Theta}_m\right)\right\|\geq\varepsilon\right\}\\
\leq&\,\mathbb{P}\left\{\underline{c}_{A(m)}^{-1/2}L_c\sqrt{\sum_{j=1}^{k_m}|G_n^j(0)|^2}\geq\varepsilon\right\}\\
=&\,\mathbb{P}\left\{\sum_{j=1}^{k_m}|G_n^j(0)|^2\geq\varepsilon^2L_c^{-2}\underline{c}_{A(m)}\right\}\\
\leq&\sum_{j=1}^{k_m}\mathbb{P}\left\{|G_n^j(0)|^2>\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}\right\}\\
\leq&\sum_{j=1}^{k_m}\mathbb{P}\left\{\frac{K_2^2K_1^2}{\tau^2n}\max_{1\leq t\leq n}\|H_1(\boldsymbol{x}_{t,m})\|^2>\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}\right\}\\
\leq&\sum_{j=1}^{k_m}\mathbb{P}\left\{\max_{1\leq t\leq n}\|H_1(\boldsymbol{x}_{t,m})\|^2>\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}n\tau^2K_2^{-2}K_1^{-2}\right\}\\
\leq&\sum_{j=1}^{k_m}\sum_{t=1}^{n}\mathbb{P}\left\{\|H_1(\boldsymbol{x}_{t,m})\|^2>\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}n\tau^2K_2^{-2}K_1^{-2}\right\}\\
\leq&\sum_{j=1}^{k_m}\sum_{t=1}^{n}\frac{\mathbb{E}\left\{\|H_1(\boldsymbol{x}_{t,m})\|^4\right\}}{\left(\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}n\tau^2K_2^{-2}K_1^{-2}\right)^2}=\frac{\mathbb{E}\left\{\|H_1(\boldsymbol{x}_{t,m})\|^4\right\}k_m^3}{\left(\varepsilon^2L_c^{-2}\underline{c}_{A(m)}\tau^2K_2^{-2}K_1^{-2}\right)^2n}=\frac{k_m^5}{\underline{c}_{A(m)}^2n}\to 0.
\end{aligned}
$$

Thus, (S.7), (S.8) and (S.9) are proved.

Combining (S.7), (S.8) and Theorem 3.1, we have $\left\|V_m\left(\widehat{\Delta}_m\right)-V_m(0)+\mathbf{D}_m\widehat{\Delta}_m\right\|_{\mathbf{C}_m}=o_p(1)$. Then, under Assumption 3.2(i), $\widehat{\Delta}_m\equiv\sqrt{n}\left(\widehat{\Theta}_m-\Theta_m^*\right)=\mathbf{D}_m^{-1}V_m(0)-\mathbf{D}_m^{-1}V_m\left(\widehat{\Delta}_m\right)+\mathbf{D}_m^{-1}R_m$ and $\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\widehat{\Delta}_m=\sqrt{n}\left(\widehat{\Theta}_m-\Theta_m^*\right)=\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}V_m(0)-\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}V_m\left(\widehat{\Delta}_m\right)+\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}R_m\equiv T_{1,m}+T_{2,m}+T_{3,m}$. where $\|R_m\|_{c_m}=o_p(1)$ for any c_m with $\|c_m\|<L_c\underline{c}_{A(m)}^{-1/2}$. Let $\eta_{ni}\equiv n^{-1/2}\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\boldsymbol{g}_{i,m}(\Theta_m^*)$, we have

[2] By Assumption 3.2, $\mathbb{P}\left(|G_n^j(0)|^2>\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}\right)\leq\mathbb{P}\left(\frac{K_2^2K_1^2}{\tau^2n}\max_{1\leq t\leq n}\|H_1(\boldsymbol{x}_{t,m})\|^2>\varepsilon^2L_c^{-2}\underline{c}_{A(m)}k_m^{-1}\right)$ is the same as Eq.9 in the supplemental appendix of Patton et al. (2019).

$T_{1,m} = \sum_{i=1}^n \eta_{ni}$. Based on the fact that $\mathrm{tr}(AB) < \lambda_{\max}(A)\mathrm{tr}(B)$,

$$
\begin{aligned}
\mathbb{E}\left\|\eta_{ni}\right\|^4 &= n^{-2}\,\mathbb{E}\left[\left\{\mathrm{tr}\left(\mathbf{C}_m \mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)^\top \mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1/2}\mathbf{C}_m^{-1}\right)\right\}^2\right] \\
&= n^{-2}\,\mathbb{E}\left[\left\{\mathrm{tr}\left(\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)^\top \mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1/2}\mathbf{C}_m^{-1}\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\right)\right\}^2\right] \\
&\leq n^{-2}\,\mathbb{E}\left[\left\{\mathrm{tr}\left(\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)^\top\right)\right\}^2\right] \times \left\{\lambda_{\max}\left(\mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1/2}\mathbf{C}_m^{-1}\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\right)\right\}^2 \\
&= n^{-2}\,\mathbb{E}\left\|\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\|^4 \times \left\{\lambda_{\max}\left(\mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1/2}\mathbf{C}_m^{-1}\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\right)\right\}^2 \\
&\leq n^{-2}\,\mathbb{E}\left\|\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\|^4 \times \left\{\lambda_{\max}\left(\mathbf{C}_m^{-1}\mathbf{C}_m\right)\right\}^2\left\{\lambda_{\max}\left(\mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1}\mathbf{D}_m^{-1}\right)\right\}^2 \\
&= n^{-2}\,\mathbb{E}\left\|\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\|^4 \times \left\{\lambda_{\max}\left(\mathbf{C}_m^{-1}\mathbf{C}_m\right)\right\}^2\left\{\lambda_{\max}\left(\mathbf{A}_m^{-1}\right)\right\}^2 \\
&= O\left(n^{-2}k_m^2\underline{c}_{A(m)}^{-2}\right).
\end{aligned}
$$

Then, given Assumption 3.2(iii), $\sum_{i=1}^n \mathbb{E}\left[\left\|\eta_{ni}\right\|^2 \mathbb{I}\left\{\left\|\eta_{ni}\right\| \geq \varepsilon\right\}\right] = n\,\mathbb{E}\left[\left\|\eta_{ni}\right\|^2 \mathbb{I}\left\{\left\|\eta_{ni}\right\| \geq \varepsilon\right\}\right] \leq n\,\mathbb{E}\left(\left\|\eta_{ni}\right\|^4\right)^{1/2} \times \mathbb{P}\left(\left\|\eta_{ni}\right\| \geq \varepsilon\right)^{1/2} \leq n\varepsilon^{-2}\,\mathbb{E}\left(\left\|\eta_{ni}\right\|^4\right) = O\left(n^{-1}k_m^2\underline{c}_{A(m)}^{-2}\right) = o(1)$. Thus, η_{ni} satisfies the conditions of the Lindeberg-Feller central limit theorem and we have

$$
T_{1,m} \xrightarrow{d} N\left(0, \mathbf{C}_0\right).
$$

For $T_{2,m}$ and $T_{3,m}$, we take $c_m = \mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}$. By the fact that $\lambda_{\max}(A^\top A) = \lambda_{\max}(AA^\top)$, we have

$$
\begin{aligned}
\left\|c_m\right\| &= \left\{\mathrm{tr}\left(\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1/2}\mathbf{C}_m^\top\mathbf{C}_m\right)\right\}^{1/2} \\
&\leq \left\|\mathbf{C}_m\right\|\left\{\lambda_{\max}\left(\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}\mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1/2}\right)\right\}^{1/2} \\
&= \left\|\mathbf{C}_m\right\|\left\{\lambda_{\max}\left(\mathbf{D}_m^{-1}\mathbf{\Sigma}_m^{-1}\mathbf{D}_m^{-1}\right)\right\}^{1/2} \\
&= \left\|\mathbf{C}_m\right\|\left\{\lambda_{\max}\left(\mathbf{A}_m^{-1}\right)\right\}^{1/2} \leq \left\|\mathbf{C}_m\right\|\underline{c}_{A(m)}^{-1/2} \leq L_c\underline{c}_{A(m)}^{-1/2}.
\end{aligned}
$$

Then by (S.9),

$$
\left\|T_{2,m}\right\| = \left\|\bar{V}_m\left(\widehat{\Delta}_m\right)\right\|_{\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}} = o_p(1)
$$

and

$$
\left\|T_{3,m}\right\| = \left\|R_m\right\|_{\mathbf{C}_m\mathbf{\Sigma}_m^{-1/2}\mathbf{D}_m^{-1}} = o_p(1).
$$

This completes the proof. ■

S.1.3 Proof of Theorem 3.3

We follow the proof strategy in Lu and Su (2015) which uses the results of Rice (1984). In other words, we replace the loss function in the proof of Theorem 3.2 of Lu and Su (2015), and the entire proof step still holds. Let $\bar{L}_m\left(\Theta_m\right) = \mathbb{E}\left\{L_m\left(y_i, v_{i,m}(\Theta), e_{i,m}(\Theta)\right)\right\}$, we have

$$
\begin{aligned}
&\bar{L}_m\left(\Theta_m\right) - \bar{L}_m\left(\Theta_m^*\right) \\
=&\,\mathbb{E}\left\{L_m\left(y_i, v_{i,m}(\Theta), e_{i,m}(\Theta)\right\} - L_m\left(y_i, v_{i,m}\left(\Theta_m^*\right), e_{i,m}\left(\Theta_m^*\right)\right)\right) \\
=&\left(\Theta_m - \Theta_m^*\right)^\top \mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\} + \frac{1}{2}\left(\Theta_m - \Theta_m^*\right)^\top \mathbf{\Lambda}_m\left(\Theta_m^*\right)\left(\Theta_m - \Theta_m^*\right) + o_p\left(\left\|\Theta_m - \Theta_m^*\right\|\right).
\end{aligned}
$$

Note that $\mathbb{E}\left\{\boldsymbol{g}_{i,m}\left(\Theta_m^*\right)\right\} = 0$, we have

$$\bar{L}_m\left(\Theta_m\right) - \bar{L}_m\left(\Theta_m^*\right) \simeq \frac{1}{2}\left(\Theta_m - \Theta_m^*\right)^\top \mathbf{D}_m\left(\Theta_m - \Theta_m^*\right).$$

Under Assumptions 3.1–3.2, the rest of the proof of Theorem 3.3 is identical to Theorem 3.2 of Lu and Su (2015), and thus we omit it here. ■

S.2 Verification of assumptions

In this part, we will verify that some high-level assumptions in the theorems of this paper hold for the homoskedasticity case of DGP1. Other data generation processes are not verified one by one.

DGP1 with heteroscedastic errors:

$$y_i = \alpha \sum_{j=1}^{1000} j^{-1} x_{i,j} + \xi_i,$$

where $x_{i,1} = -1$ and $x_{i,j}, j = 2, 3, \ldots$ are each i.i.d. $N(0,1)$ and are mutually independent of each other. $\xi_i \sim N(0,1)$ and independent of $x_{i,j}$.

First, taking a candidate model with regressor $\{1, x_{i,2}\}$ to serve as an illustrative example, we will verify the existence and uniqueness of pseudo-true parameter Θ_m^*. In other words, we estimate the VaR and ES of y_i using

$$\begin{cases} \mathrm{VaR}_\tau(y_i|x_{i,2}) = \vartheta_1 + \vartheta_2 x_{i,2}, \\ \mathrm{ES}_\tau(y_i|x_{i,2}) = \theta_1 + \theta_2 x_{i,2}, \end{cases} \tag{S.12}$$

where $\vartheta_1, \vartheta_2, \theta_2$ and θ_2 are unknown parameters. Since $x_{i,2}$ is independent of $\{x_{i,3}, \ldots, x_{i,1000}\}$ and ξ_i, DGP1 can be considered as

$$y_i = -\alpha + \frac{\alpha}{2} x_{i,2} + \eta_i,$$

where $\eta_i \sim N(0, 1 + \sum_{j=3}^{1000} j^{-2})$ and independent of $x_{i,2}$. Therefore, parameter estimation of the joint model (S.12) can be considered as the model is correctly specified rather than misspecified. Thus, the pseudo-true parameter

$$\Theta_m^* = (\vartheta_1^*, \vartheta_2^*, \theta_2^*, \theta_2^*)^\top = (-\alpha + \mathrm{VaR}_\tau(\eta_i), \alpha/2, -\alpha + \mathrm{ES}_\tau(\eta_i), \alpha/2)^\top.$$

Next, we will demonstrate the validity of Assumption 1, which asserts that the sample must satisfy the conditions $0 < |1/e_{i,m}(\Theta)| < K_2 < \infty$ and $e_{i,m}(\Theta_m^*) < v_{i,m}(\Theta_m^*) < 0$ for any candidate model. To fulfill this requirement, we propose a linear transformation of y. Specifically, within our simulation design for DGP1-3, we first normalize $\{y_i\}_{i=1}^n$, and then shift the normalized $\{y_i\}_{i=1}^n$ to guarantee that the maximum true VaR of in-sample as well as out-of-sample is -2. As a rule of thumb, this treatment ensures that the in-sample estimators and the out-of-sample predictors of VaR and ES are always smaller than zero. Consequently, this ensures the validity of Assumption 4.

Assumptions 3.1 and 3.2 impose several constraints on $\underline{c}_{A(m)}$, $\bar{c}_{A(m)}$, $\underline{c}_{D(m)}$, $\bar{c}_{D(m)}$, $\underline{c}_{X(m)}$ and $\bar{c}_{X(m)}$. In order to visualize the eigenvalues, i.e., $\lambda_{\min}(\mathbf{A}_m)$, $\lambda_{\max}(\mathbf{A}_m)$, $\lambda_{\min}(\mathbf{D}_m)$, $\lambda_{\max}(\mathbf{D}_m)$, $\lambda_{\min}\{\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)\}$ and $\lambda_{\max}\{\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)\}$, we will perform numerical simulations to calculate them. For simplicity, α in DGP1 is taken to be 1. Furthermore, in consideration of the requirement for Assumption 1 to hold, we consider the following data generation process:

$$y_i = -3 + \sum_{j=2}^{1000} j^{-1} x_{i,j} + \xi_i.$$

Since the regressor dimension k_m of the mth candidate model is allowed to diverge with increasing sample size, we assume that the regressors used for the mth candidate model gradually change from $\{1, x_{i,2}\}$ to $\{1, x_{i,2}, x_{i,3}, \dots\}$. Based on 100000 samples, the numerical estimates of the maximum and minimum eigenvalues of $\mathbf{A}_m$, $\mathbf{D}_m$ and $\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)$ are presented in Figure S.1. Notice that the DGP1 implies that $\lambda_{\min}\{\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)\} = \lambda_{\max}\{\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)\} = 1$. This indicates that the numerical estimates of $\lambda_{\min}\{\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)\}$ and $\lambda_{\max}\{\mathbb{E}(\boldsymbol{x}_{i,m}\boldsymbol{x}_{i,m}^\top)\}$ in Figure S.1 are close to their actual values. The results in Figure S.1 show that $\lambda_{\min}(\mathbf{A}_m)$, $\lambda_{\max}(\mathbf{A}_m)$, $\lambda_{\min}(\mathbf{D}_m)$ and $\lambda_{\max}(\mathbf{D}_m)$ hardly vary with k_m. Therefore, Figure S.1 ensures that the assumptions concerning $\underline{c}_{A(m)}$, $\bar{c}_{A(m)}$, $\underline{c}_{D(m)}$, $\bar{c}_{D(m)}$, $\underline{c}_{X(m)}$ and $\bar{c}_{X(m)}$ hold.

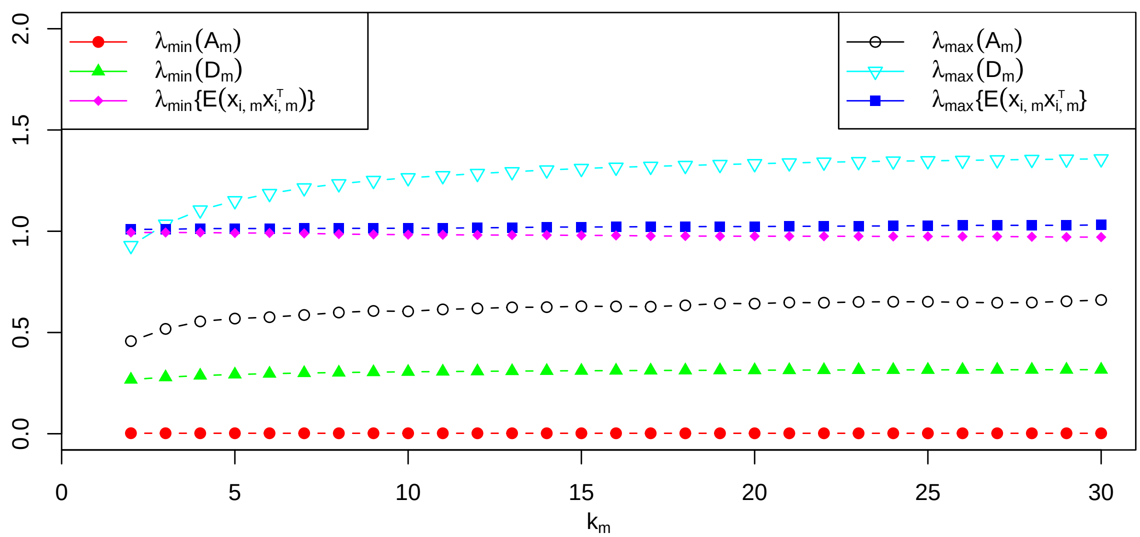

Figure S.1: Maximum and minimum eigenvalues.

S.3 Additional simulation results

As the title of this paper suggests, the main focus of our paper is the JMA method for the joint estimation of VaR and ES. Since our results involve both VaR and ES, it is important to discuss the performance of each individually. Hence, in this section and next section, we will investigate the performance of our JMA estimators separately for VaR and ES.

To evaluate the performance of the VaR estimators, we will report the results of FPE_{VaR}, after conducting simulation experiments based on DGP1-6, as presented in our main text, where

$$\text{FPE}^*_{\text{VaR}} = \frac{1}{100}\sum_{s=1}^{100} \rho_\tau \left(y_s - \widehat{\text{VaR}}_\tau \left(y_s | \boldsymbol{x}_s \right) \right),$$

with check function $\rho_\tau(u) = u\{\tau - \mathbb{I}(u \leq 0)\}$. In other words, when performing simulation studies for DGP1-6, we focus on the results of FPE_{VaR} rather than FPE in (4.1). The other settings remain unchanged. In addition, we also consider the JMA estimator of VaR (abbreviated as JMA_{VaR}) proposed by Lu and Su

(2015) to examine the performance of joint and separate estimation methods. We only add the JMA method of Lu and Su (2015) as a comparison because the dominance of JMA over other estimation methods has been confirmed in Lu's paper.

As mentioned in the introduction of our article, ES is not elicitability and there is no consensus criterion to evaluate the estimated performance. Here, we utilize the $\mathrm{FPE}_{\mathrm{ES}}$ based on mean square error (MSE) as the evaluation criterion, i.e.,

$$\mathrm{FPE}_{\mathrm{ES}} = \frac{1}{100} \sum_{s=1}^{100} \left(\widehat{\mathrm{ES}}_{\tau} \left(y_s | \boldsymbol{x}_s \right) - \mathrm{ES}_{\tau} \left(y_s | \boldsymbol{x}_s \right) \right)^2,$$

where $\mathrm{ES}_{\tau} \left(y_s | \boldsymbol{x}_s \right)$ is the true ES obtained from the data generation process.

Figures S.2-S.12 report the $\mathrm{FPE}_{\mathrm{VaR}}$ of DGP1-6. Overall, for small R^2, $\mathrm{JMA}_{\mathrm{VaR}}$ performs significantly worse than the four methods based on the joint estimation of VaR and ES. However, this gap diminishes significantly as n increases. At $n = 400$, $\mathrm{JMA}_{\mathrm{VaR}}$ exhibits a clear dominance on larger R^2. We interpret this phenomenon to suggest that methods based on the joint estimation of VaR and ES comprehensively incorporate tail information when the information validity is insufficient (small R^2). However, the VaR-focused method ($\mathrm{JMA}_{\mathrm{VaR}}$) offers a more accurate estimate of VaR when ample information is available (large R^2). For the four methods that jointly model VaR and ES, the performance of VaR for JMA is almost always optimal and SWMA approaches the performance level of JMA.

Figures S.13-S.23 present the results of $\mathrm{FPE}_{\mathrm{ES}}$. For DGP1 and 2, the $\mathrm{FPE}_{\mathrm{ES}}$ obtained from JMA outperforms the other methods for all n and R^2. In the case of DGP3, SWMA performs slightly better than JMA when R^2 is small, while the opposite holds true for large R^2. The results of DGP4, DGP5 and DGP6 presented in Figures S.19-S.23 show that the JMA method is optimal in most cases, although it is occasionally slightly weaker than SWMA. Additionally, both CVMS and Largest exhibit significantly inferior performance compared to the two model averaging methods.

Overall, regardless of whether the performance of VaR and ES is discussed individually or jointly, the JMA method proposed in this paper consistently demonstrates significant advantages. Of course, if one's primary focus is on modeling VaR, the quantile JMA method proposed by Lu and Su (2015) is also a good choice. Additionally, the unsatisfactory outcomes of the model selection method and Largest model underscore the necessity for a model averaging approach.

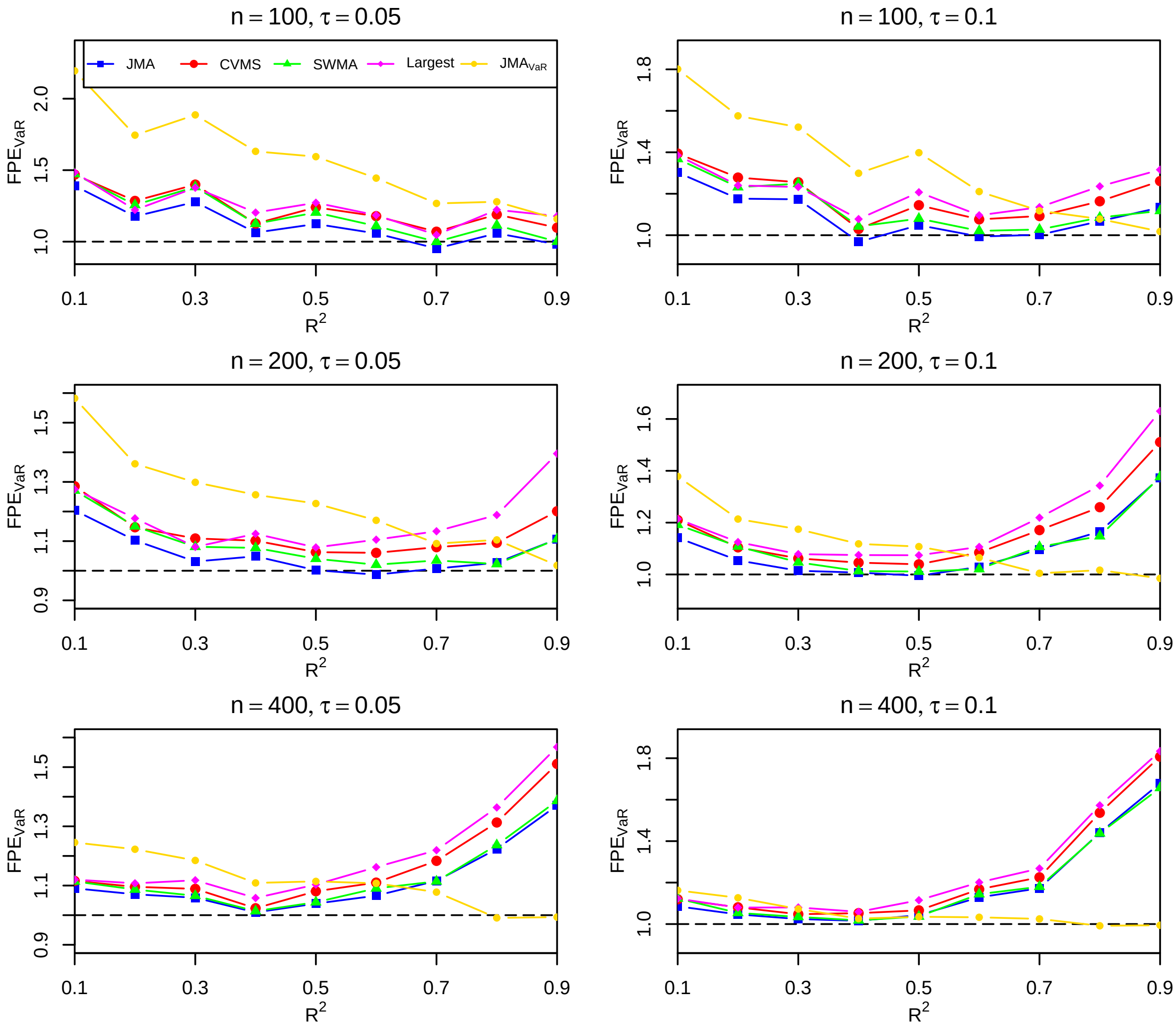

Figure S.2: Out-of-sample performance of VaR: DGP1, Homoskedasticity.

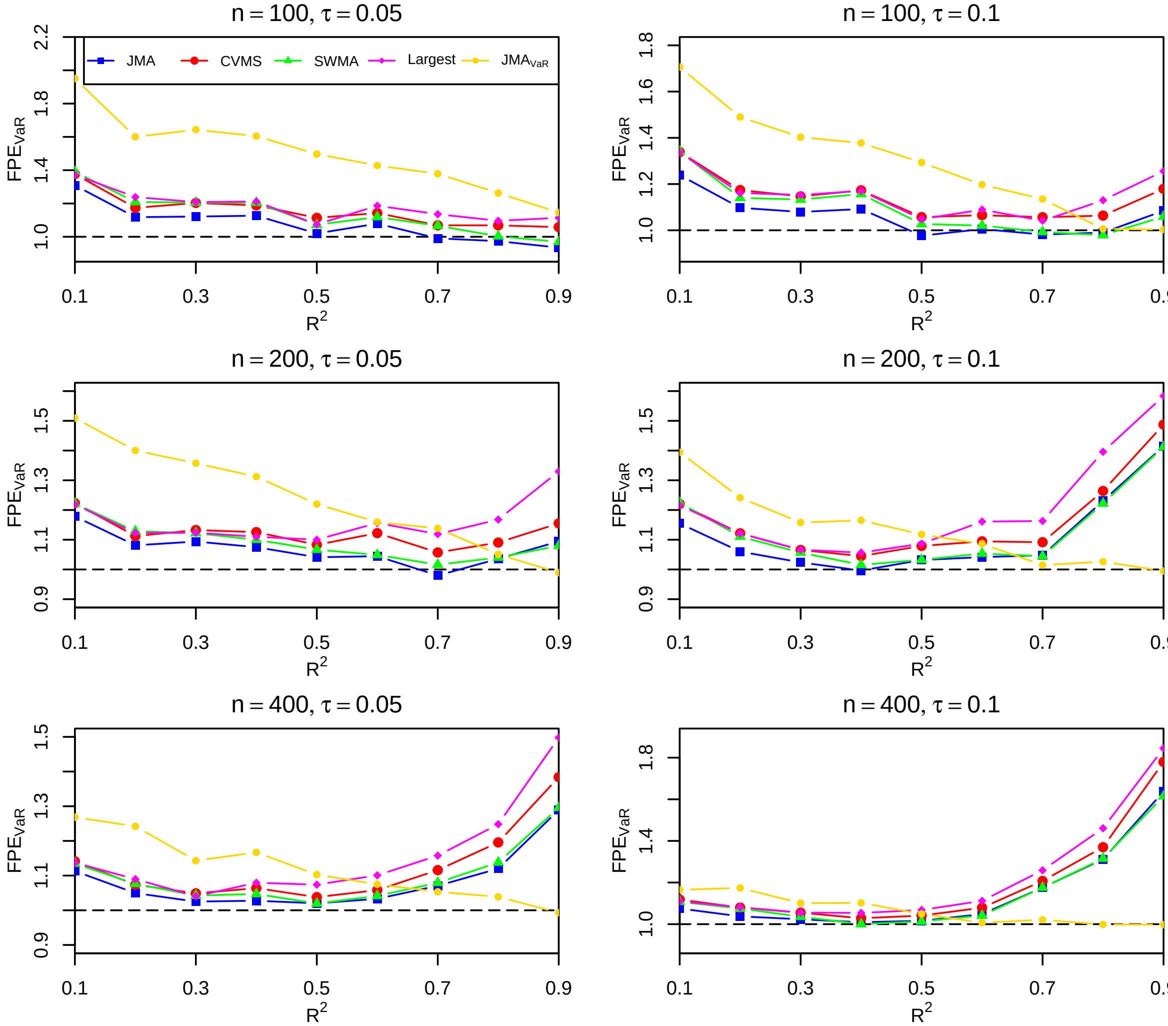

Figure S.3: Out-of-sample performance of VaR: DGP1, Heteroskedasticity.

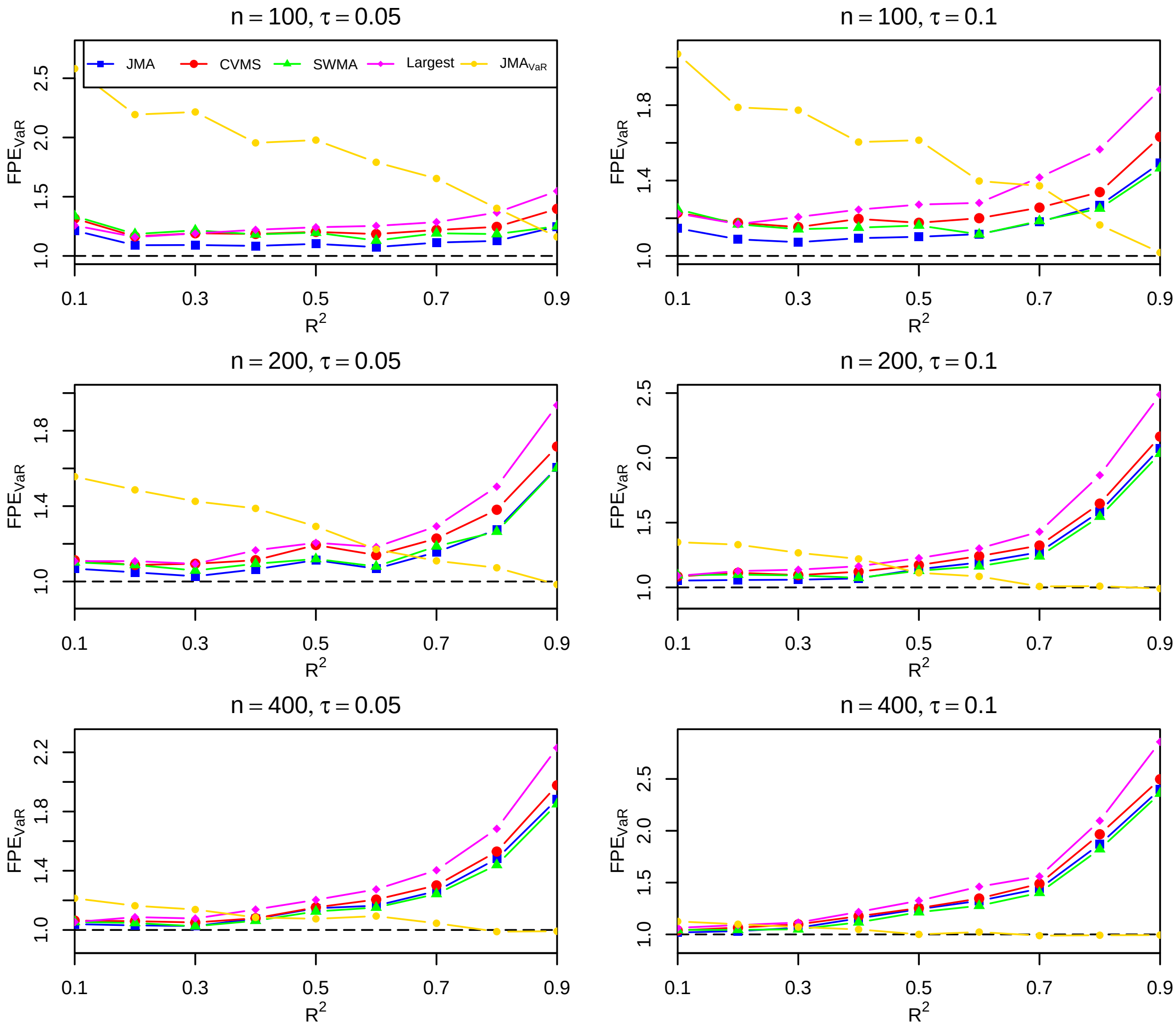

Figure S.4: Out-of-sample performance of VaR: DGP2, Homoskedasticity.

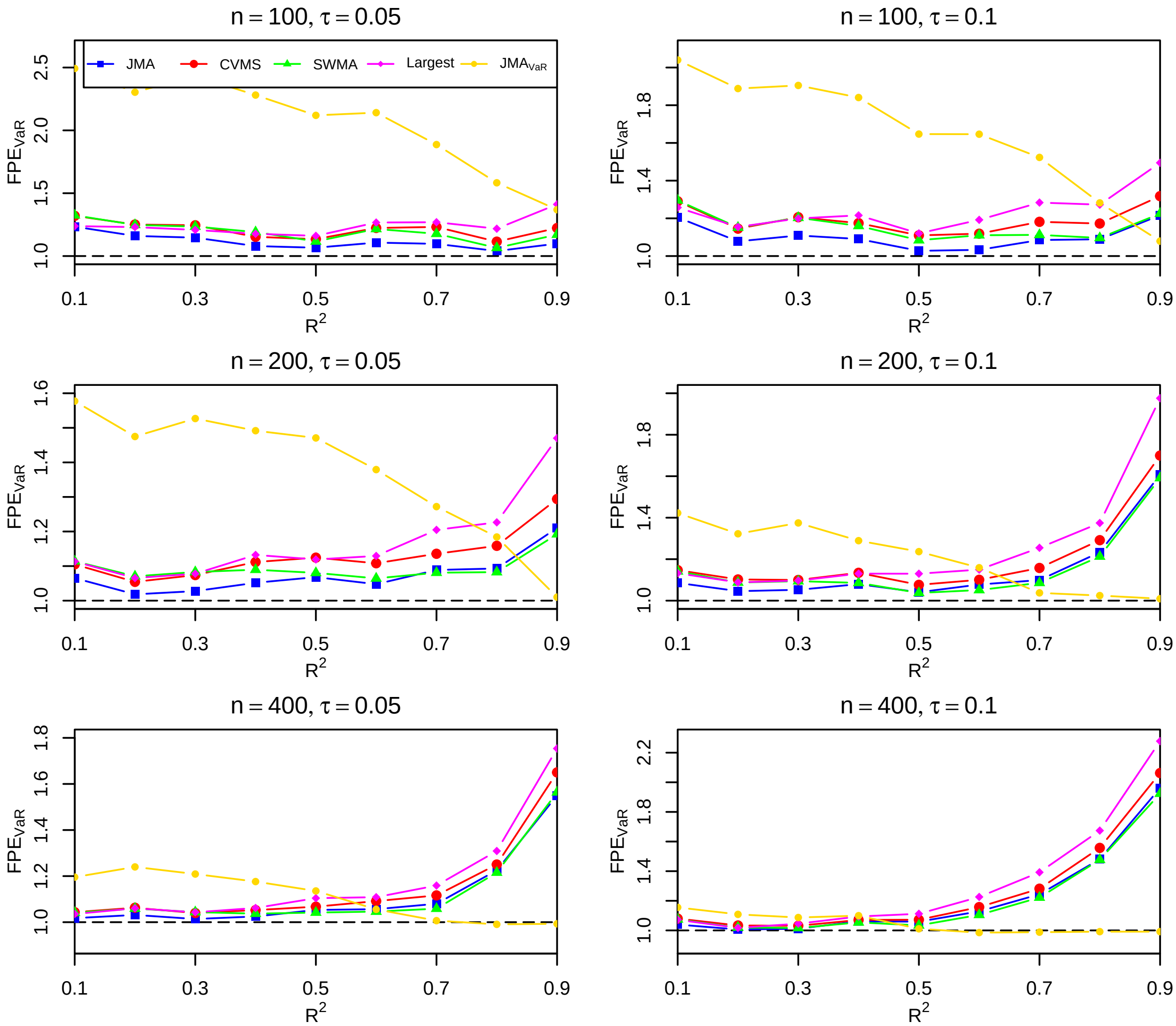

Figure S.5: Out-of-sample performance of VaR: DGP2, Heteroskedasticity.

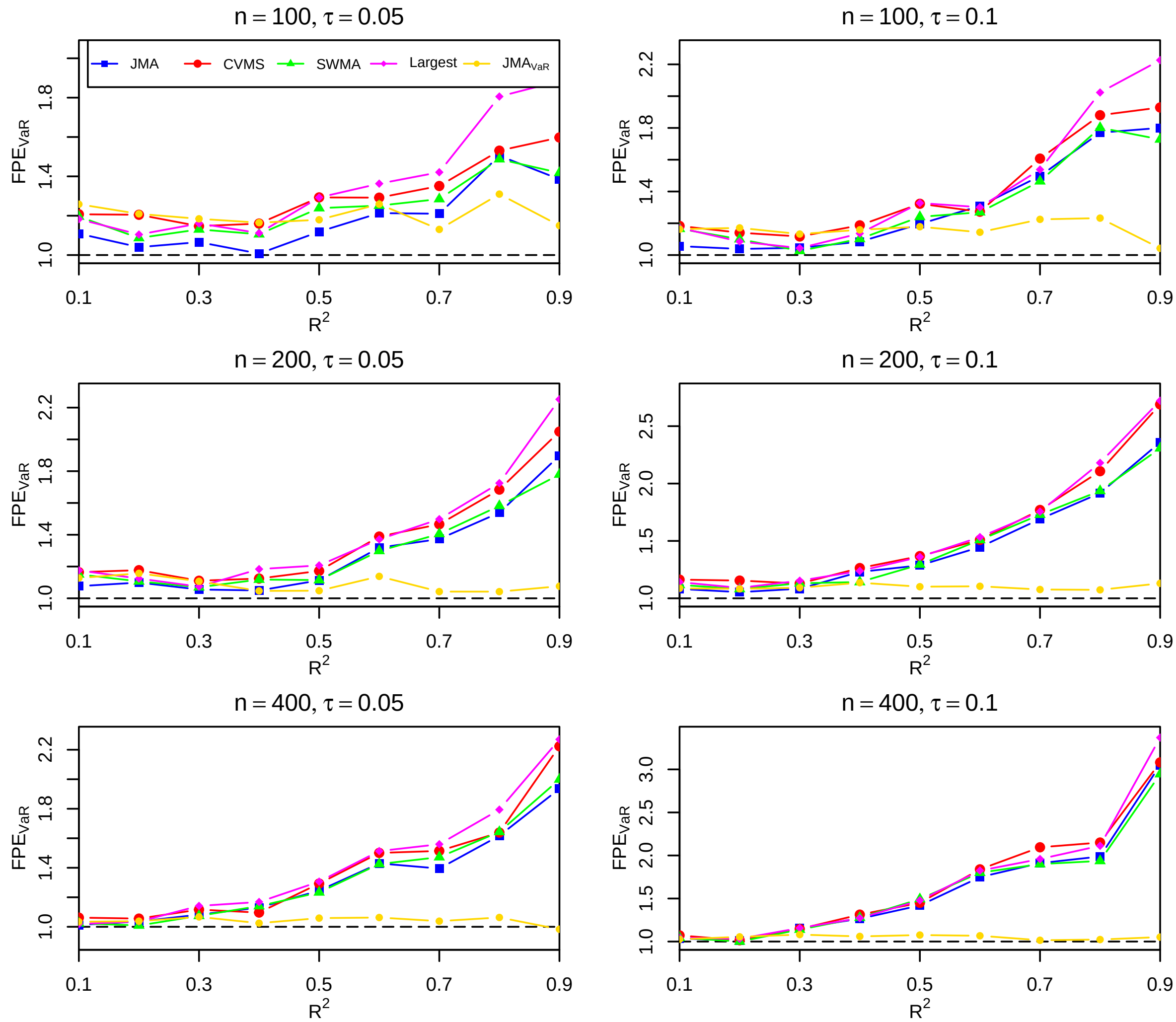

Figure S.6: Out-of-sample performance of VaR: DGP3, Homoskedasticity.

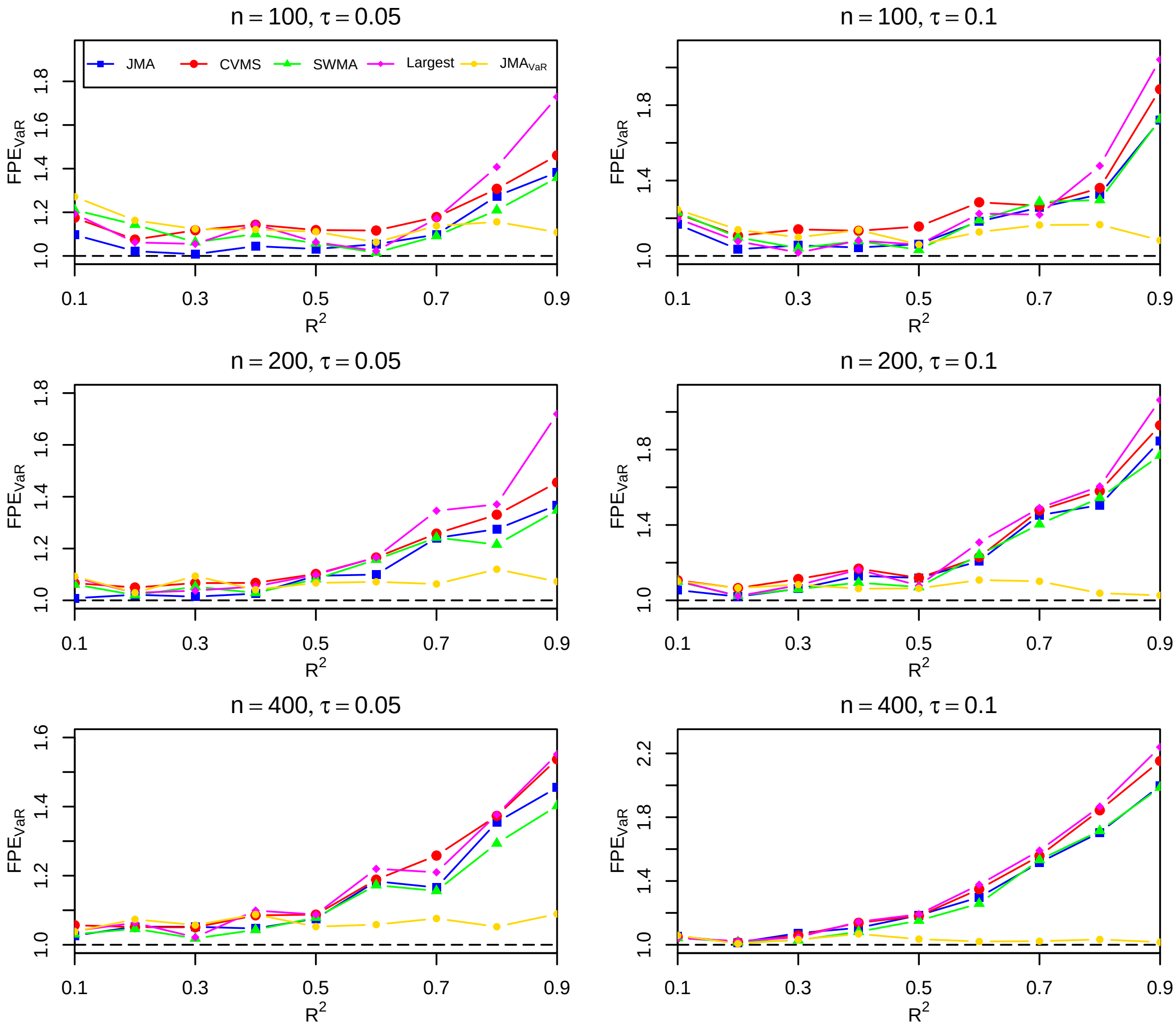

Figure S.7: Out-of-sample performance of VaR: DGP3, Heteroskedasticity.

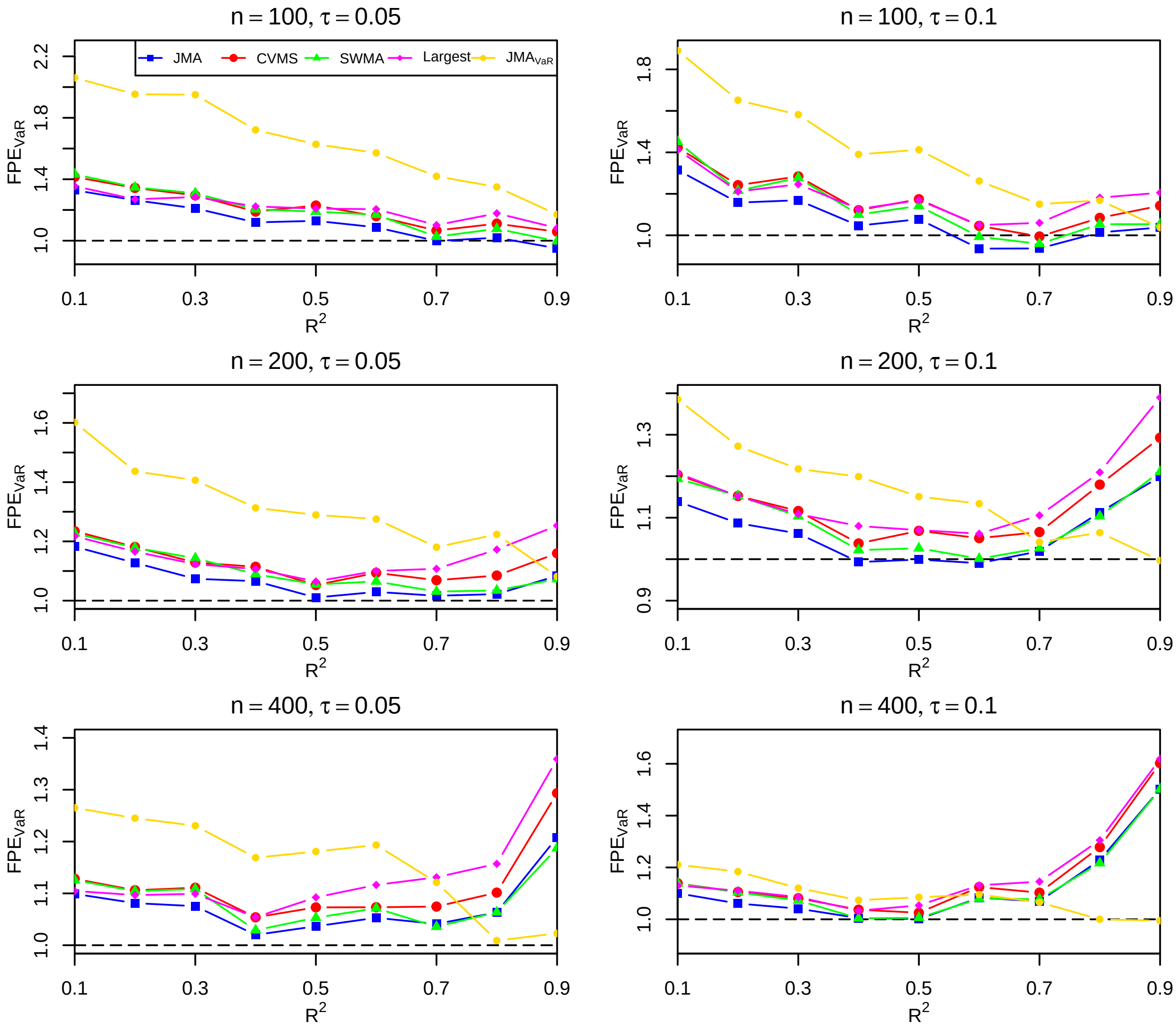

Figure S.8: Out-of-sample performance of VaR: DGP4, Homoskedasticity.

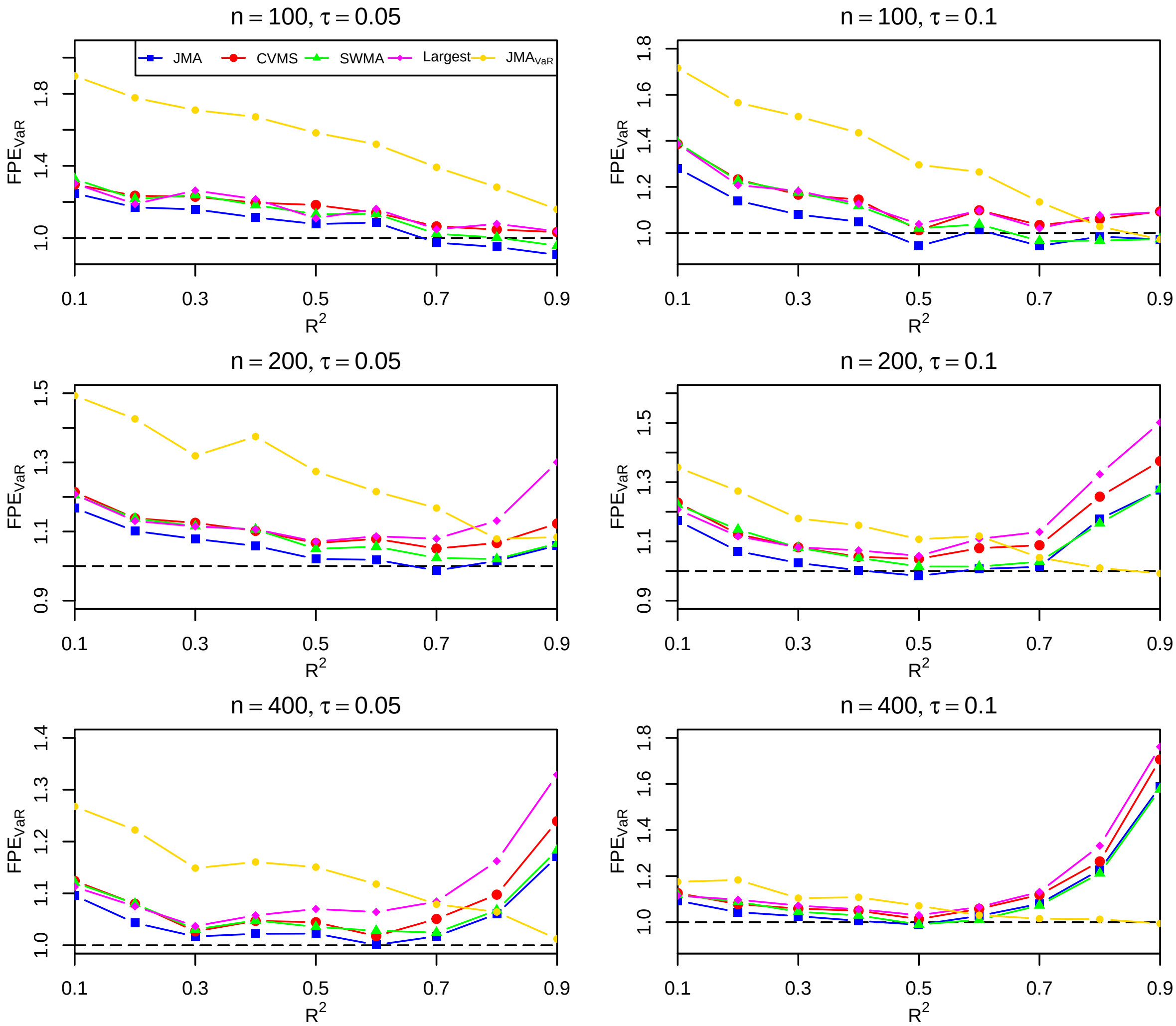

Figure S.9: Out-of-sample performance of VaR: DGP4, Heteroskedasticity.

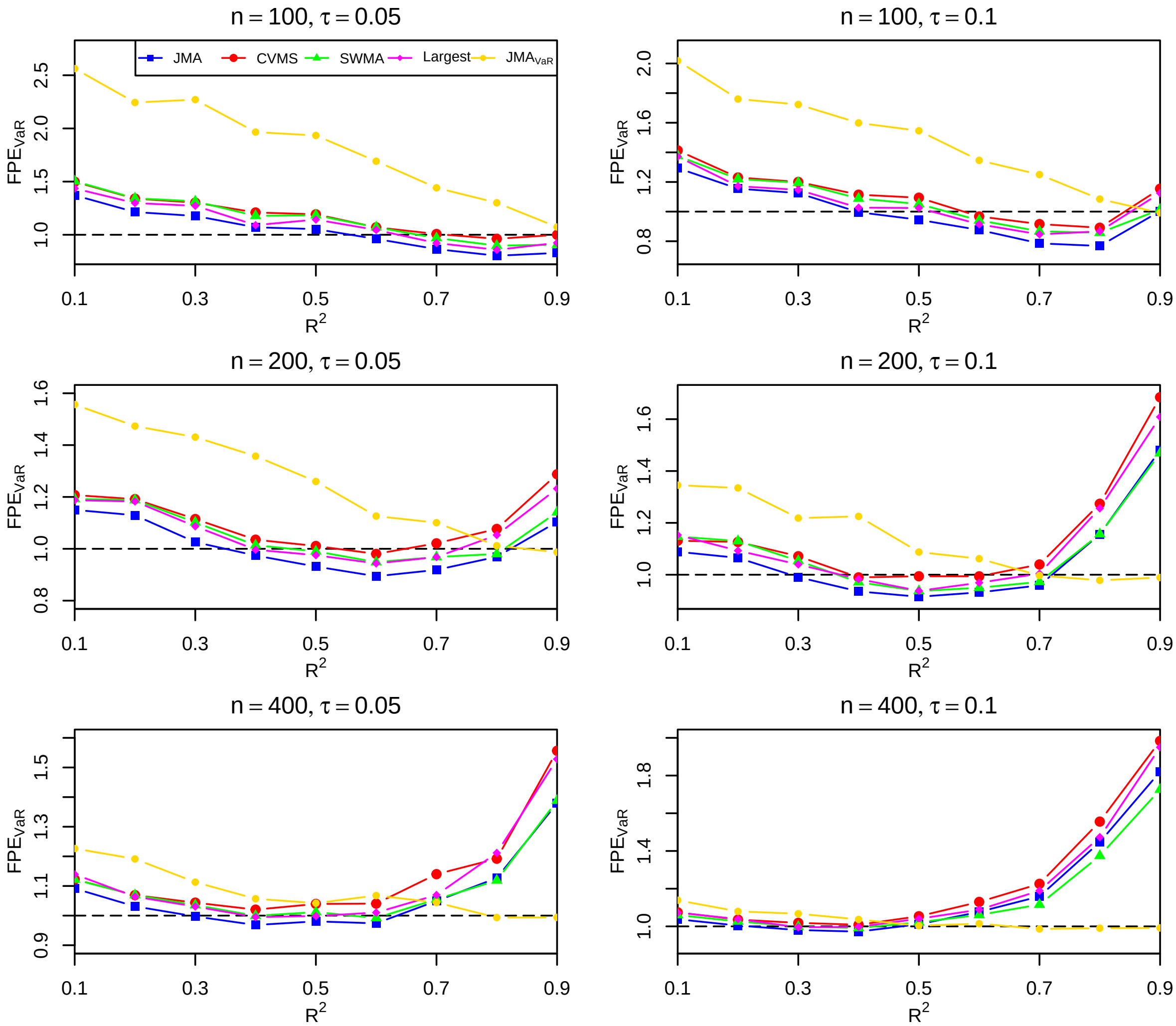

Figure S.10: Out-of-sample performance of VaR: DGP5, Homoskedasticity.

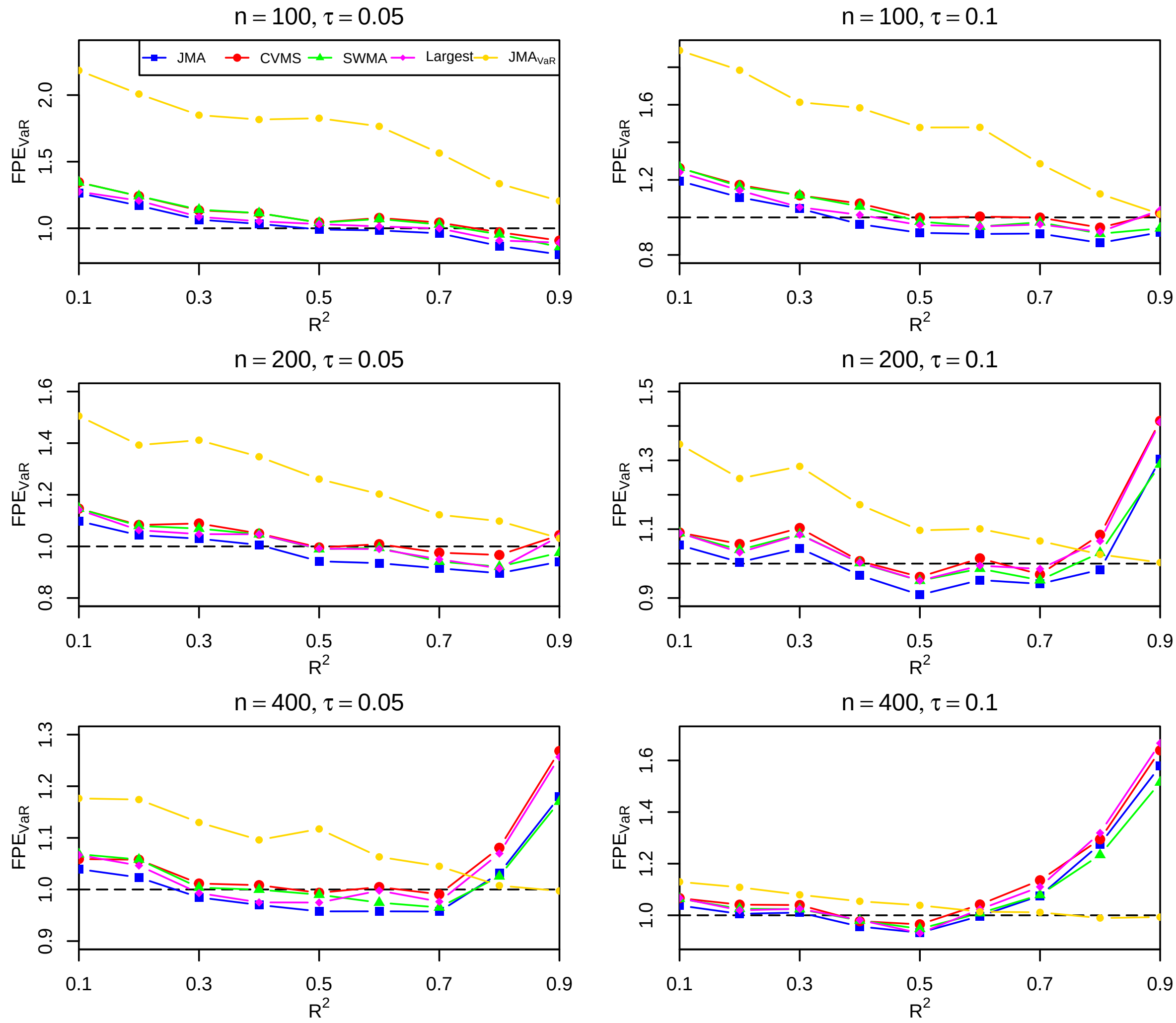

Figure S.11: Out-of-sample performance of VaR: DGP5, Heteroskedasticity.

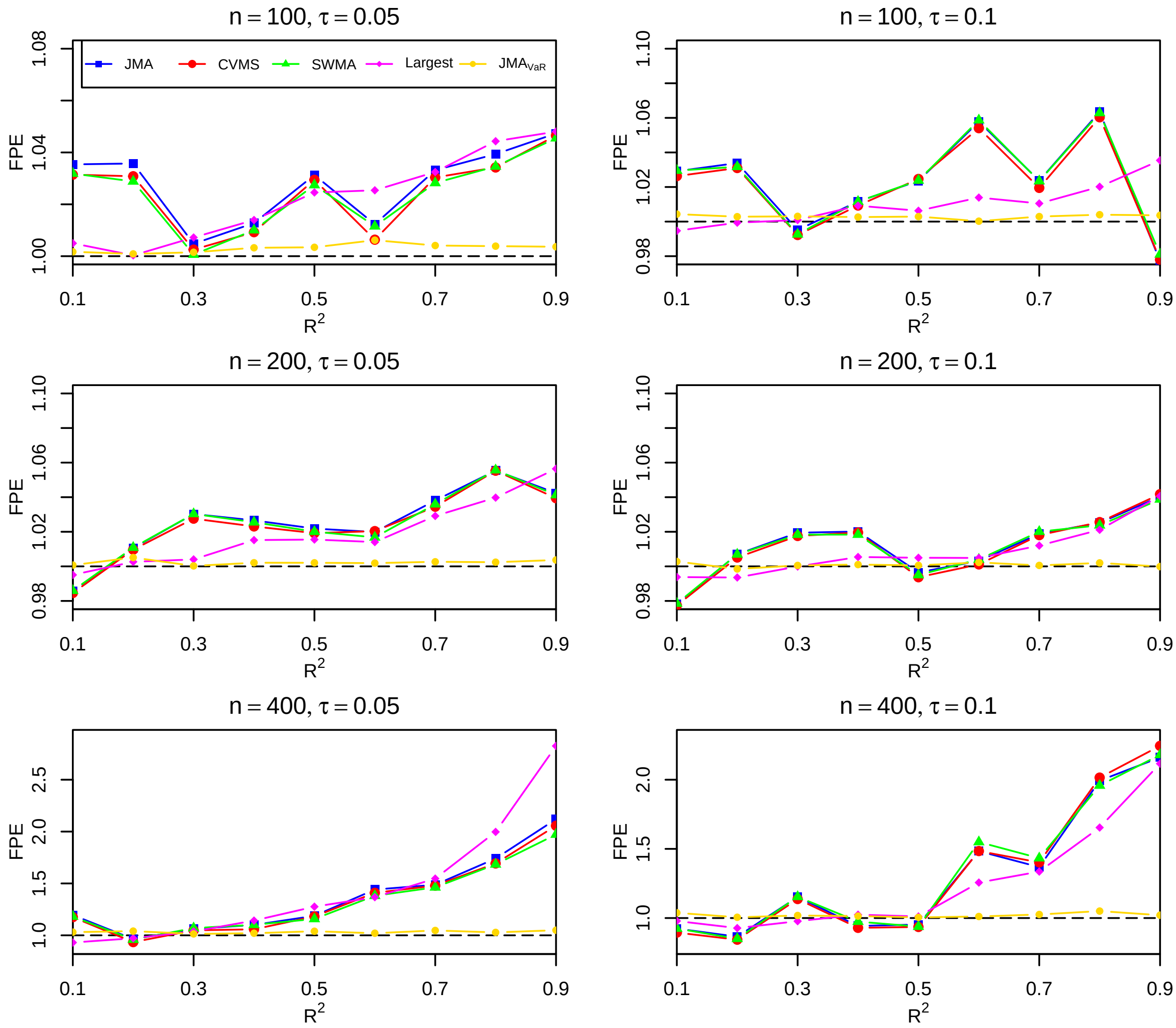

Figure S.12: Out-of-sample performance of VaR: DGP6.

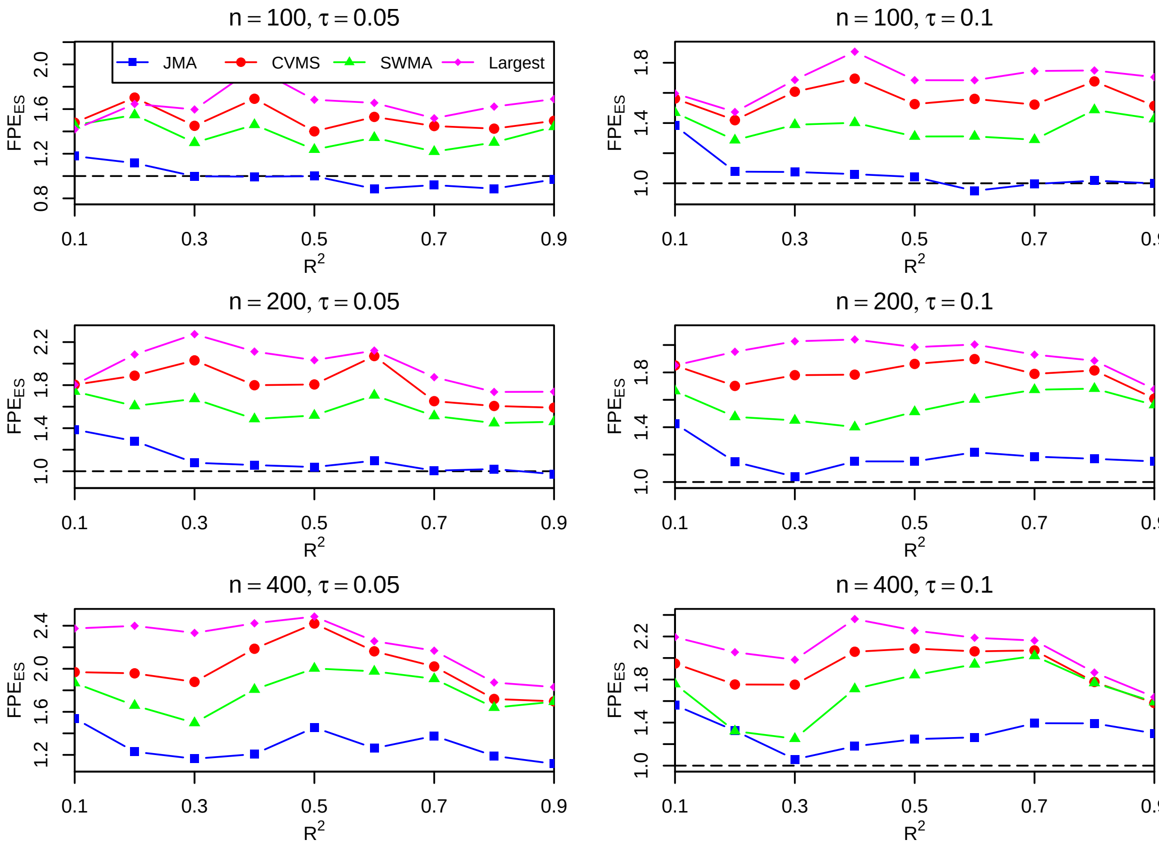

Figure S.13: Out-of-sample performance of ES: DGP1, Homoskedasticity.

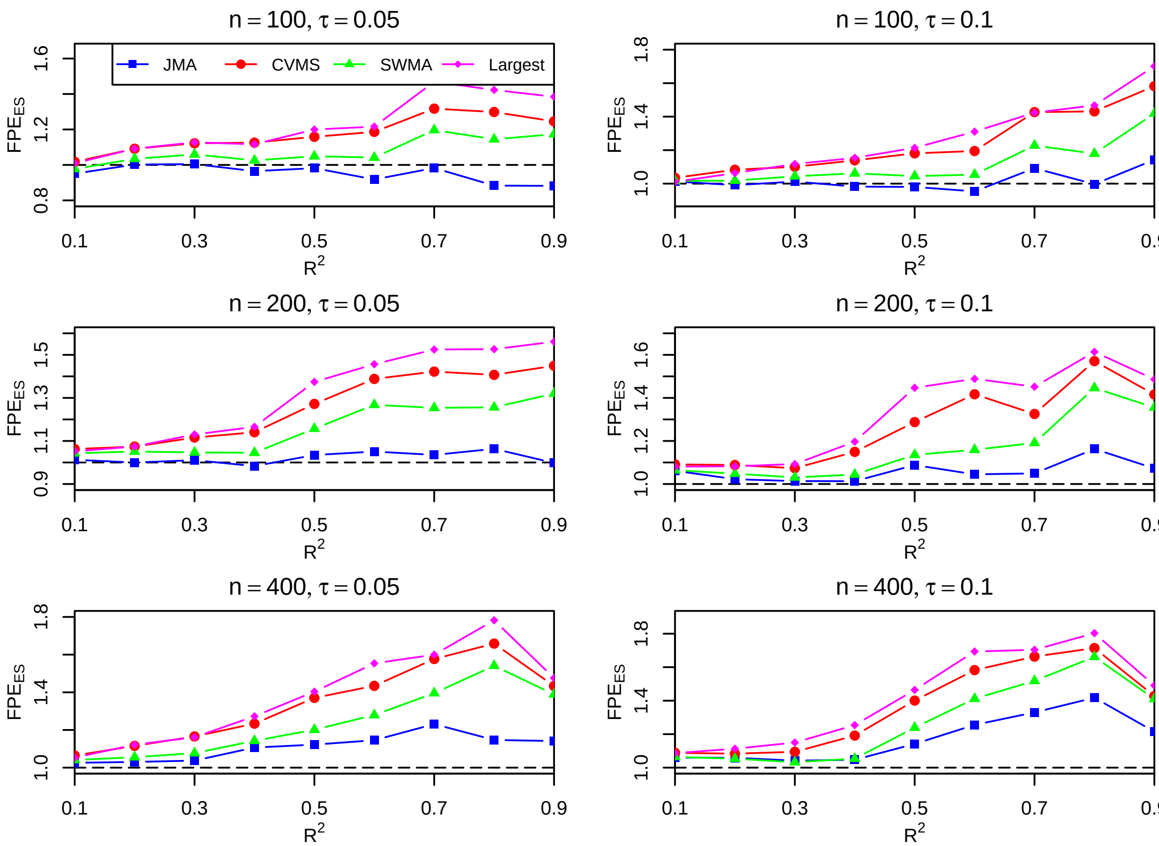

Figure S.14: Out-of-sample performance of ES: DGP1, Heteroskedasticity.

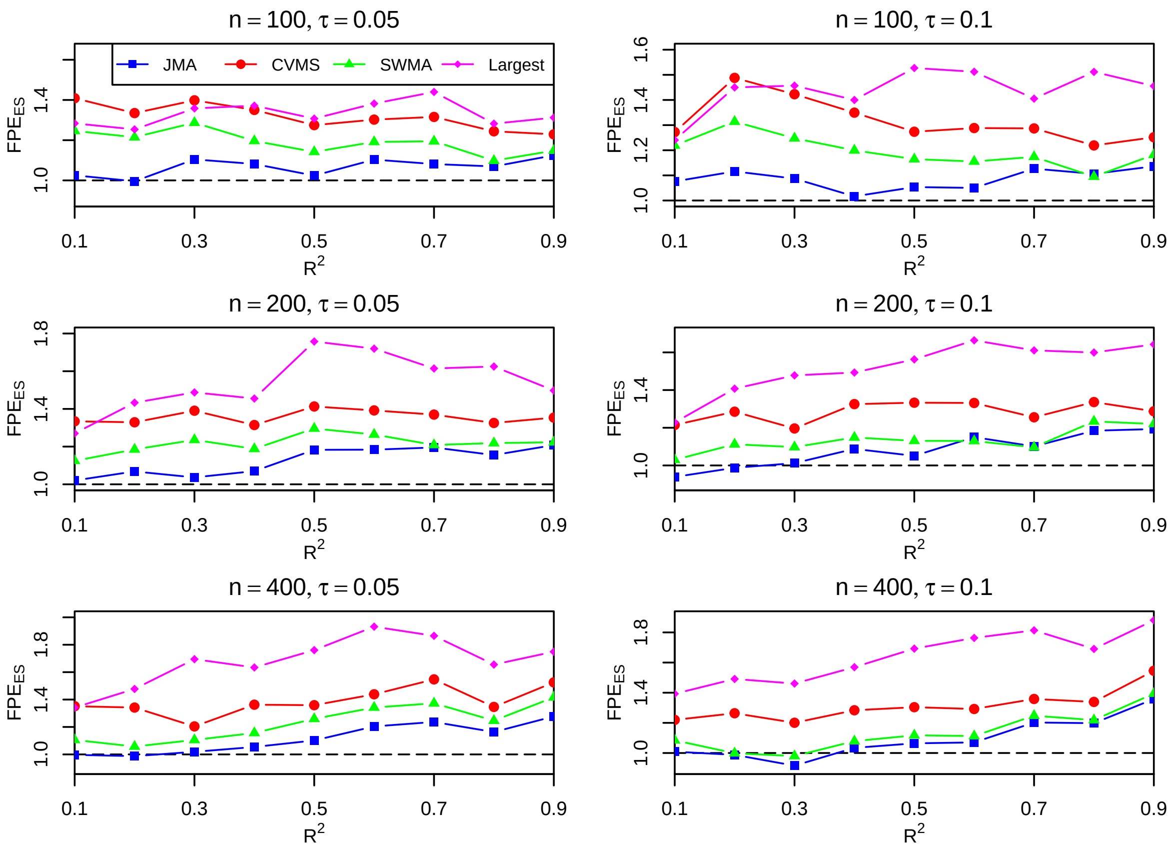

Figure S.15: Out-of-sample performance of ES: DGP2, Homoskedasticity.

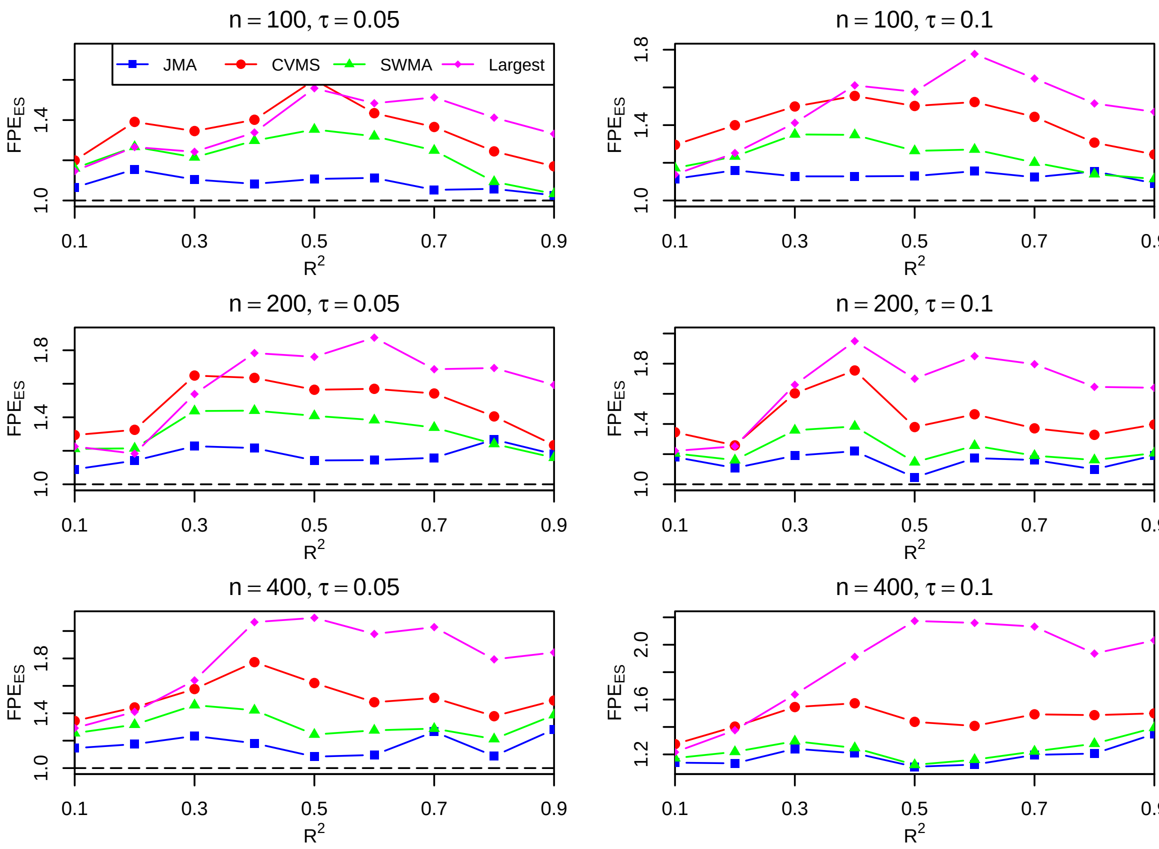

Figure S.16: Out-of-sample performance of ES: DGP2, Heteroskedasticity.

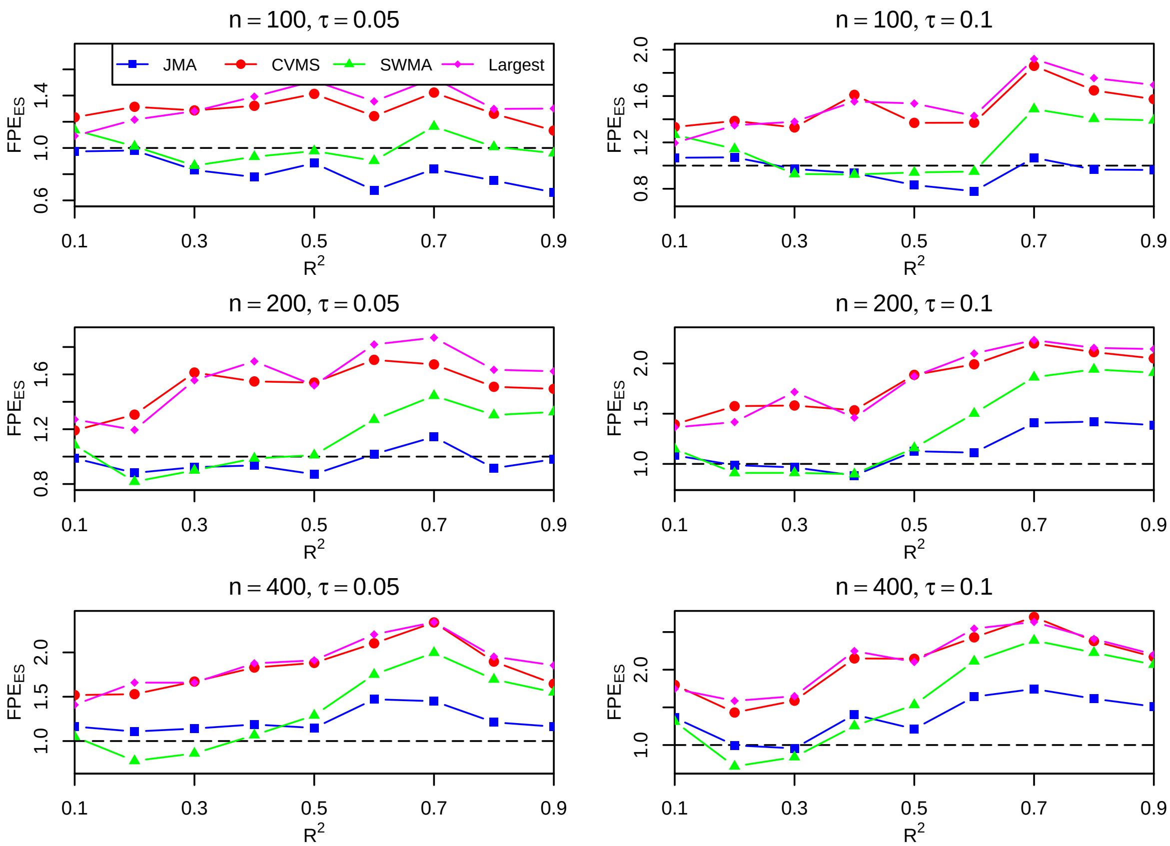

Figure S.17: Out-of-sample performance of ES: DGP3, Homoskedasticity.

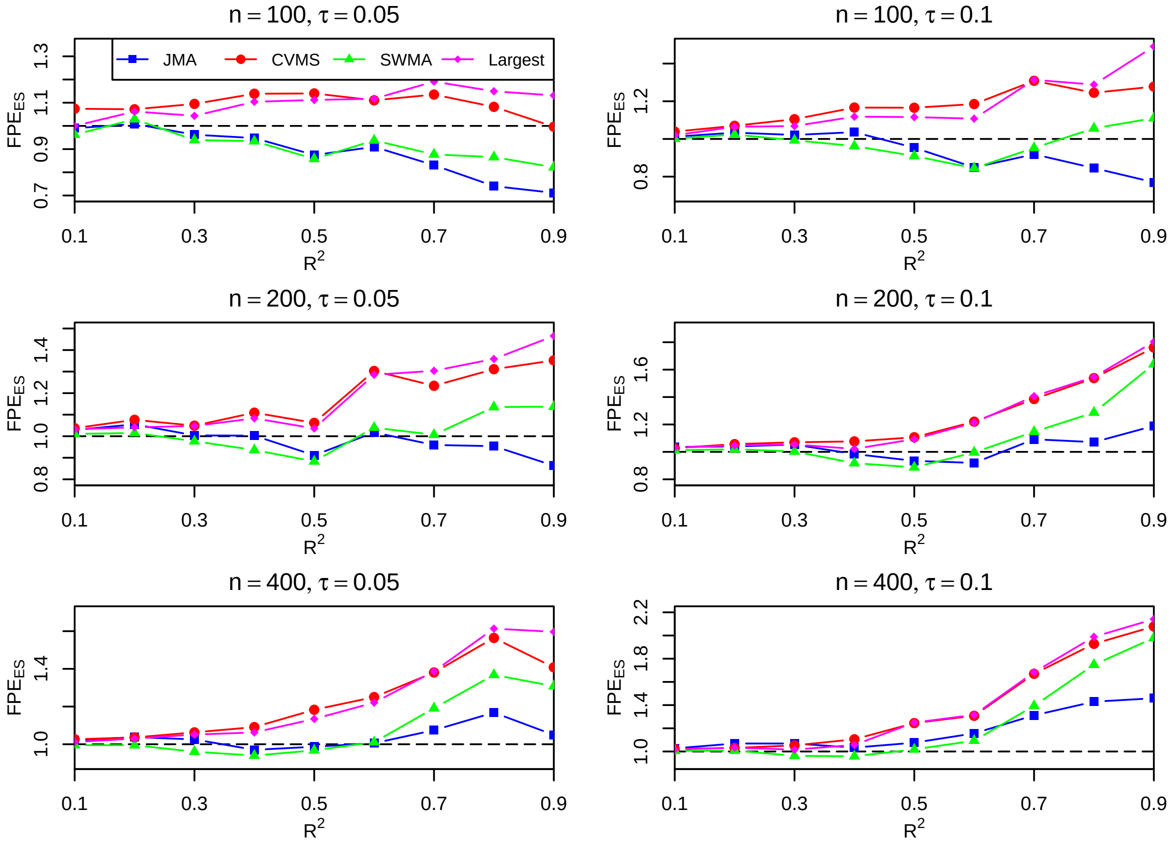

Figure S.18: Out-of-sample performance of ES: DGP3, Heteroskedasticity.

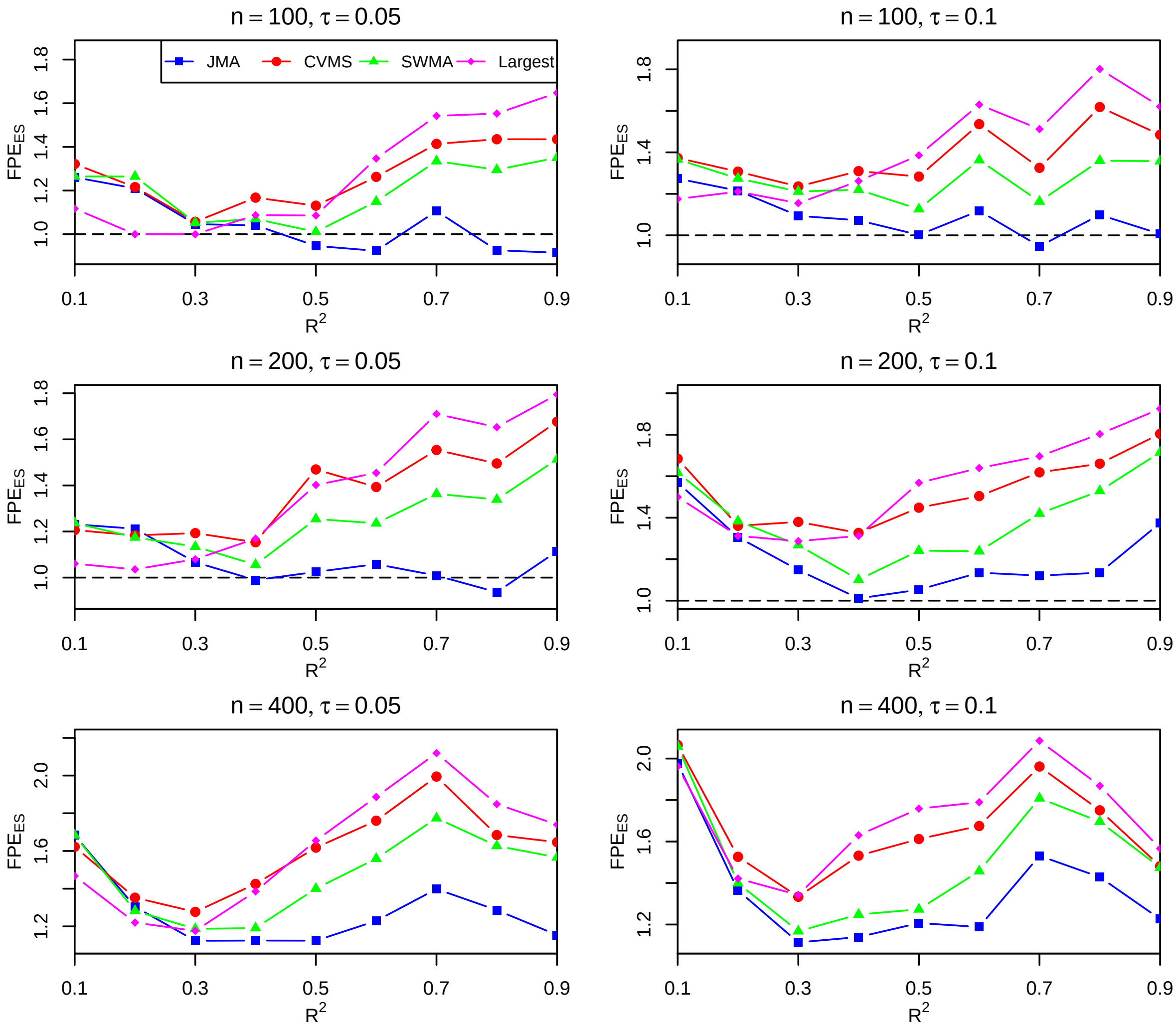

Figure S.19: Out-of-sample performance of ES: DGP4, Homoskedasticity.

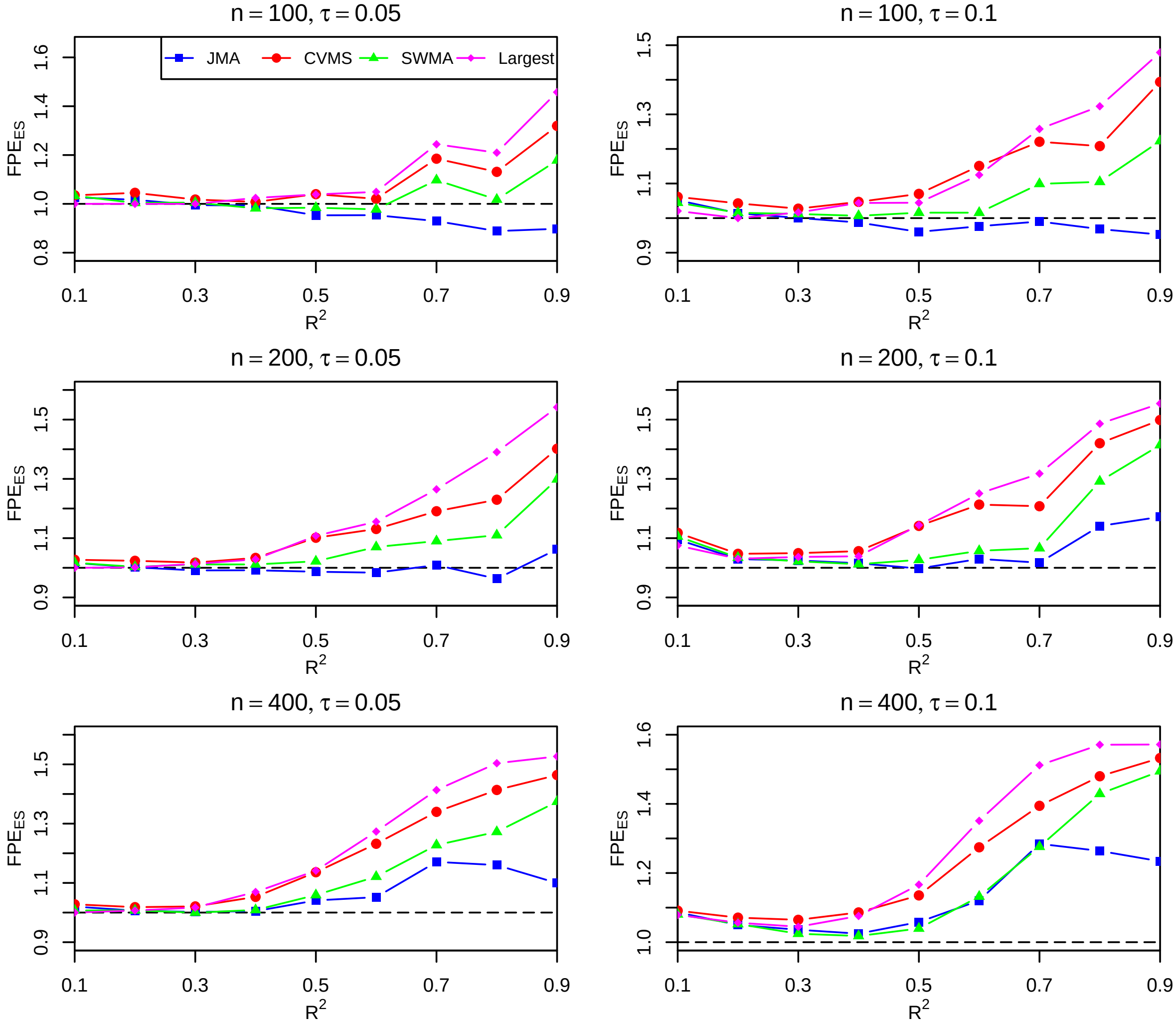

Figure S.20: Out-of-sample performance of ES: DGP4, Heteroskedasticity.

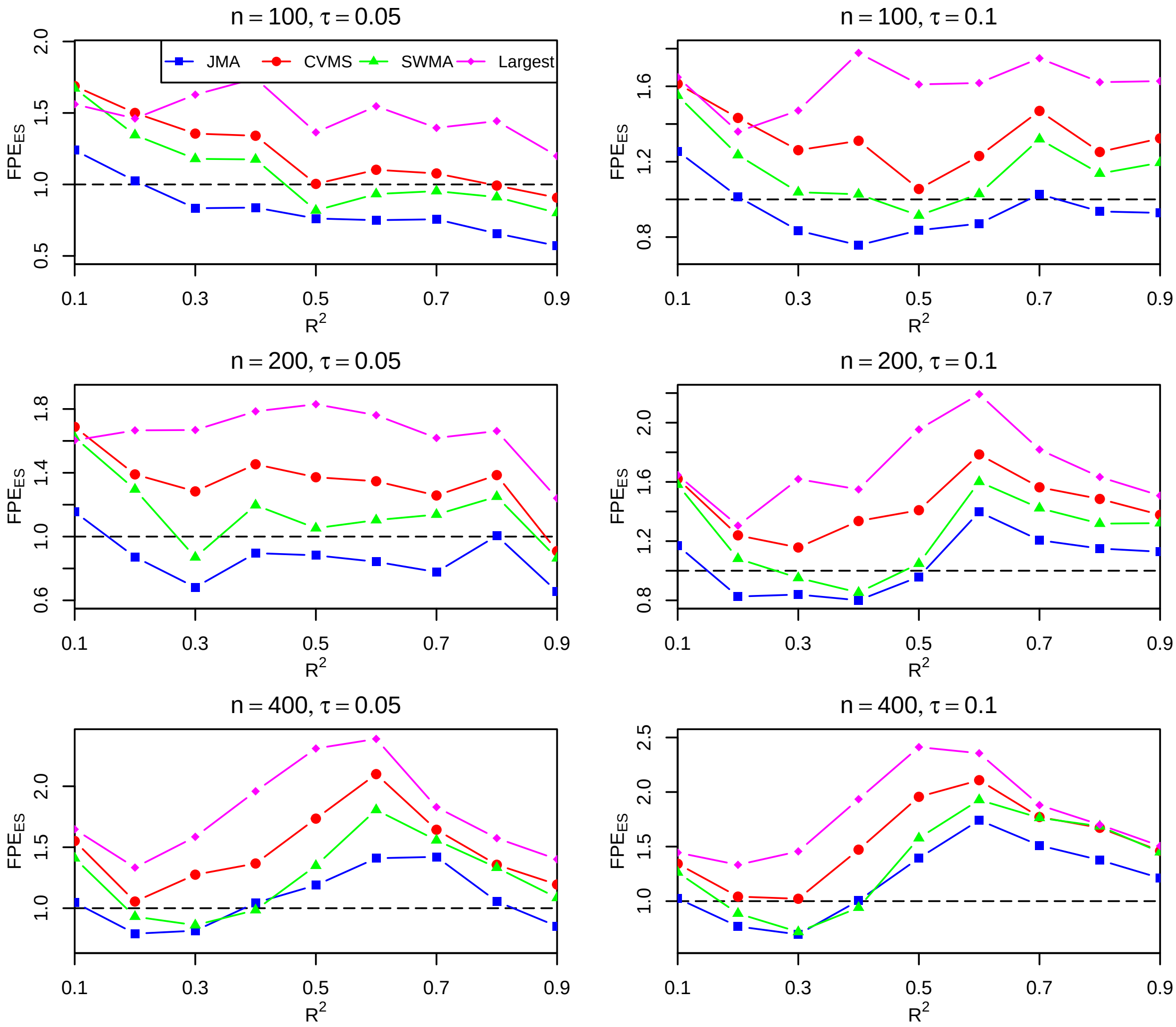

Figure S.21: Out-of-sample performance of ES: DGP5, Homoskedasticity.

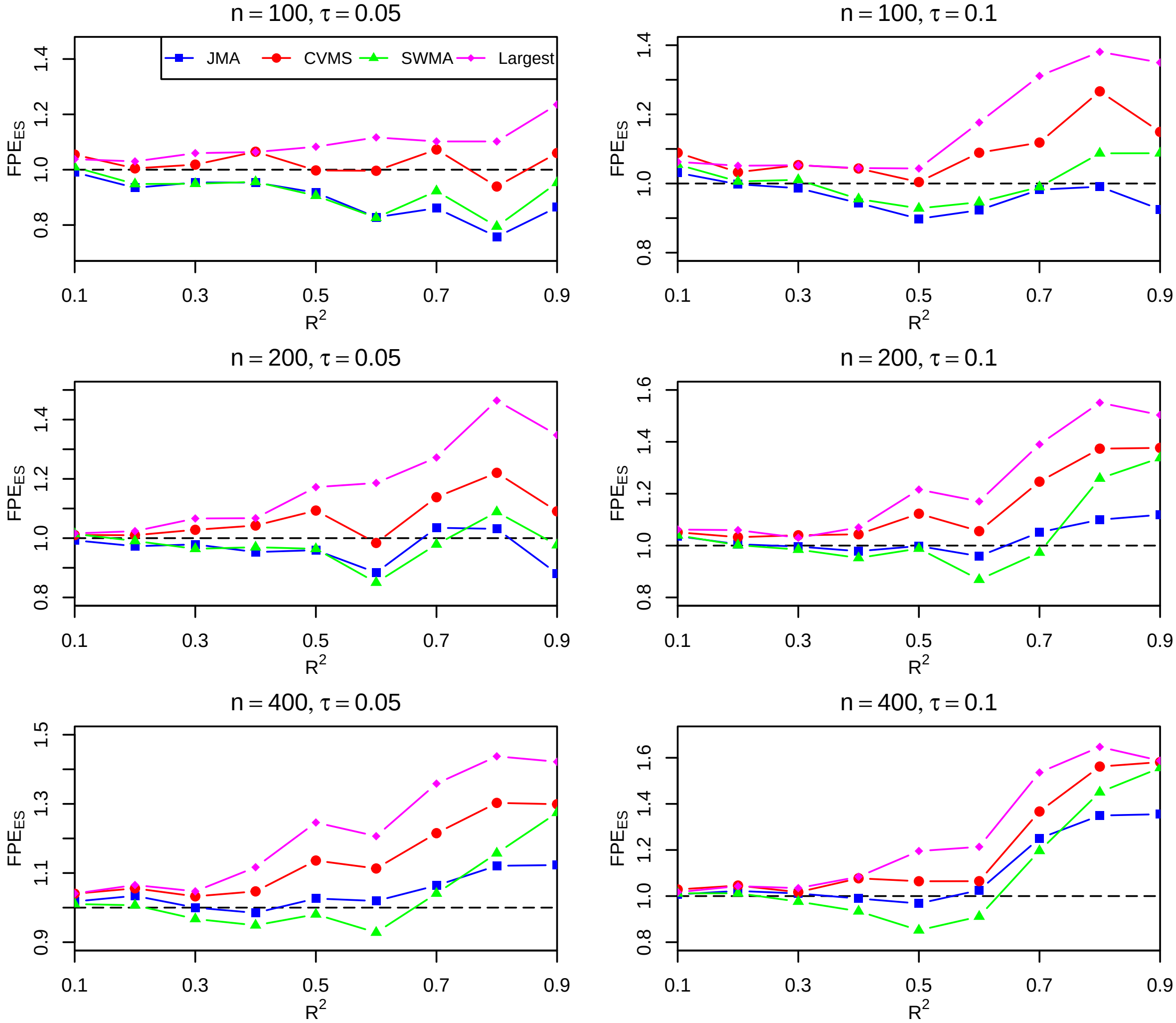

Figure S.22: Out-of-sample performance of ES: DGP5, Heteroskedasticity.

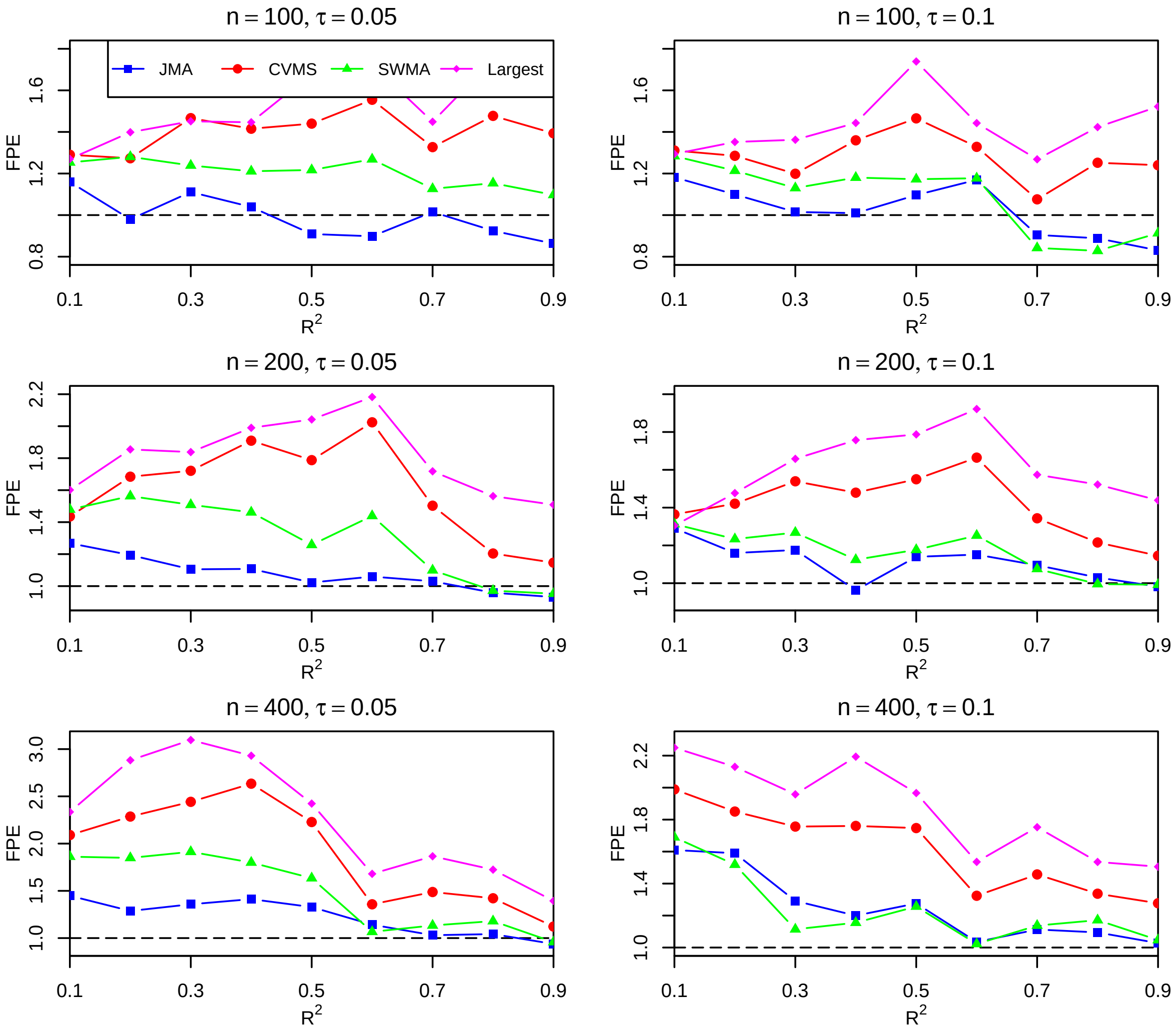

Figure S.23: Out-of-sample performance of ES: DGP6.

S.4 Additional empirical results

Our empirical analysis yields not only the EFPE results presented in the main text, but also the $\mathrm{FPE}_{\mathrm{VaR}}$ results, as shown in Tables S.1-S.3. Here, similar to the EFPE in the main text, we define $\mathrm{FPE}_{\mathrm{VaR}} = [\sum_{t=T_1+1}^{T} \{\rho_\tau(y_t, \widehat{v}_t) - \rho_{\tau,\min}(y_t, \widehat{v}_t)\}]/(T - T_1)$ or $\mathrm{FPE}_{\mathrm{VaR}} = [\sum_{t \in testing\ set} \{\rho_\tau(y_t, \widehat{v}_t) - \rho_{\tau,\min}(y_t, \widehat{v}_t)\}]/(n - n_1)$. The results in Tables S.1-S.3 indicate that there is little difference between the different methods when we focus on VaR only, and no method exhibits dominance. These findings are in line with the simulation results of Section S.3.

Due to not knowing the true value of ES and the absence of a corresponding loss function for ES, reporting the results of $\mathrm{FPE}_{\mathrm{ES}}$ in this section, as in Section S.3, is not feasible. Therefore, this section does not discuss the empirical results for ES.

Table S.1: Out-of-sample $\mathrm{FPE}_{\mathrm{VaR}}$ for the stock returns data.

τ	T_1	JMA	SWMA	CVMS	Largset	$\mathrm{JMA}_{\mathrm{VaR}}$
0.1	100	0.0356	0.0364	0.0382	0.0355	0.0483
0.1	200	0.0315	0.0332	0.0329	0.0325	0.0377
0.1	400	0.0294	0.0294	0.0294	0.0321	0.0289
0.05	100	0.0211	0.0221	0.0226	0.0230	0.0332
0.05	200	0.0212	0.0217	0.0222	0.0222	0.0304
0.05	400	0.0220	0.0226	0.0231	0.0220	0.0208

Table S.2: Out-of-sample $\mathrm{FPE}_{\mathrm{VaR}}$ for the wage data.

τ	n_1	JMA	SWMA	CVMS	Largset	$\mathrm{JMA}_{\mathrm{VaR}}$
0.1	100	0.0015	0.0017	0.0022	0.0015	0.0050
0.1	200	0.0015	0.0026	0.0018	0.0017	0.0015
0.1	400	0.0025	0.0042	0.0026	0.0024	0.0011
0.05	100	0.0012	0.0012	0.0019	0.0011	0.0064
0.05	200	0.0007	0.0019	0.0011	0.0008	0.0023
0.05	400	0.0011	0.0026	0.0012	0.0011	0.0016

Table S.3: Out-of-sample FPE_{VaR} for the WTI returns data.

τ	T_1	JMA	SWMA	CVMS	Largest	JMA_{VaR}
0.1	100	0.0192	0.0189	0.0190	0.0197	0.0195
0.1	200	0.0134	0.0129	0.0130	0.0127	0.0134
0.1	400	0.0073	0.0072	0.0072	0.0067	0.0075
0.05	100	0.0134	0.0133	0.0136	0.0129	0.0135
0.05	200	0.0086	0.0092	0.0087	0.0093	0.0106
0.05	400	0.0053	0.0053	0.0052	0.0051	0.0058

References

Lu, X. and Su, L. (2015). Jackknife model averaging for quantile regressions. *Journal of Econometrics*, 188(1):40–58.

Patton, A. J., Ziegel, J. F., and Chen, R. (2019). Dynamic semiparametric models for expected shortfall (and value-at-risk). *Journal of Econometrics*, 211(2):388–413.

Rice, J. (1984). Bandwidth choice for nonparametric regression. *The Annals of Statistics*, 12(4):1215–1230.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics.* John Wiley & Sons.