Appendix A. Regularity conditions, proofs, and theoretical derivations

Proof of Theorem 1. Denote

$$H_{ni} = \prod_{t=T_0+1}^{T} f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})$$

and

$$B_{ni} = \sum\nolimits_{j=1}^{J} \pi_{i,j} \prod_{t=T_0+1}^{T} \widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}).$$

For any $j \in \{1, \ldots, J\}$,

$$\begin{aligned}
\log\left(\frac{H_{ni}}{B_{ni}}\right) &\leq \log\left\{\frac{\prod_{t=T_0+1}^{T} f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\pi_{i,j}\prod_{t=T_0+1}^{T} \widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\} \\
&= \log\left(\frac{1}{\pi_{i,j}}\right) + \log\left\{\frac{\prod_{t=T_0+1}^{T} f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\prod_{t=T_0+1}^{T} \widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\}. \quad \text{(A.1)}
\end{aligned}$$

Moreover, by the construction of B_{ni}, it is readily seen that

$$\begin{aligned}
B_{ni} &= \left\{\sum\nolimits_{j=1}^{J} \pi_{i,j}\widehat{g}_{i,T_0+1,j}(y_{i,T_0+1} \mid \boldsymbol{y}_i^{(T_0)})\right\} \\
&\quad \times \left\{\frac{\sum_{j=1}^{J} \pi_{i,j}\widehat{g}_{i,T_0+1,j}(y_{i,T_0+1} \mid \boldsymbol{y}_i^{(T_0)})\widehat{g}_{i,T_0+2,j}(y_{i,T_0+2} \mid \boldsymbol{y}_i^{(T_0+1)})}{\sum_{j=1}^{J} \pi_{i,j}\widehat{g}_{i,T_0+1,j}(y_{i,T_0+1} \mid \boldsymbol{y}_i^{(T_0)})}\right\} \times \\
&\quad \ldots \times \left\{\frac{\sum_{j=1}^{J} \pi_{i,j}\prod_{t=T_0+1}^{T} \widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\sum_{j=1}^{J} \pi_{i,j}\prod_{t=T_0+1}^{T-1} \widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\} \\
&= \prod_{t=T_0+1}^{T}\left\{\sum\nolimits_{j=1}^{J} w_{i,t,j}\widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})\right\} \\
&= \prod_{t=T_0+1}^{T} \widehat{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}). \qquad \text{(A.2)}
\end{aligned}$$

Combining (A.1) and (A.2), it follows that for any $j \in \{1, \ldots, J\}$,

$$\begin{aligned}
&\frac{1}{n(T-T_0)}\sum\nolimits_{i=1}^{n} \log\left(\frac{H_{ni}}{B_{ni}}\right) \\
&= \frac{1}{n(T-T_0)}\sum\nolimits_{i=1}^{n}\sum\nolimits_{t=T_0+1}^{T}\left[\log\left\{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})\right\} - \log\left\{\widehat{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})\right\}\right] \\
&\leq \frac{\sum_{i=1}^{n}\log\left(\frac{1}{\pi_{i,j}}\right)}{n(T-T_0)}
\end{aligned}$$

$$+\frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\left[\log\left\{f_{i,t}(y_{i,t}\mid \boldsymbol{y}_i^{(t-1)})\right\}-\log\left\{\widehat{g}_{i,t,j}(y_{i,t}\mid \boldsymbol{y}_i^{(t-1)})\right\}\right]}{n(T-T_0)}, \tag{A.3}$$

and the proof is concluded by taking the expectation on both sides of (A.3). $\square$

Proof of Corollary 1. Recall that $\widehat{\theta}_{i,t,j} = (b^{(1)})^{-1}(\widehat{\mu}_{i,t,j})$. It is seen that under Condition 1

$$\left.\frac{d(b^{(1)})^{-1}(\mu)}{d\mu}\right|_{\mu=\widehat{\mu}_{i,t,j}} = \frac{1}{b^{(2)}(\widehat{\theta}_{i,t,j})} \le 1/c_0 < \infty,$$

almost surely. This indicates that under Condition 1, $\widehat{\theta}_{i,t,j}(\widehat{\mu}_{i,t,j}) = (b^{(1)})^{-1}(\widehat{\mu}_{i,t,j})$ is a continuous function of $\widehat{\mu}_{i,t,j}$. Due to the fact that $(b^{(1)})^{-1}(\cdot)$ is a monotonically increasing function, we have $\widehat{\theta}_{i,t,j}(\widehat{\mu}_{i,t,j}) \in I_1 = [(b^{(1)})^{-1}(-c_1), (b^{(1)})^{-1}(c_1)]$. A similar argument also indicates that the continuous function $|b(\widehat{\theta}_{i,t,j})|$ attains its maximum in the closed interval I_1 almost surely. Then, we immediately obtain that $\bar{c}_1 < \infty$ by definition. Moreover, define $\widetilde{B}_{ni} = \sum_{j\in\widehat{\Gamma}}\widetilde{\pi}_{i,j}\prod_{t=T_0+1}^{T}\widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})$. Under the same framework that proves (A.3), we can verify that for any $j' \in \widehat{\Gamma}$

$$\begin{aligned}
&\frac{\sum_{i=1}^{n}\log\left(\frac{H_{ni}}{\widetilde{B}_{ni}}\right)}{n(T-T_0)}\\
&=\frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\left[\log\left\{\frac{f(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widetilde{g}_{\text{AFTER},i,t}(y_{i,t}|\boldsymbol{y}_i^{(t-1)},\widehat{\Gamma})}\right\}\right]}{n(T-T_0)}\\
&\le \frac{\log(\widehat{J}_1)}{T-T_0}+\frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\left[\log\left\{\frac{f_{i,t}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j'}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}\right]}{n(T-T_0)}.
\end{aligned} \tag{A.4}$$

Denote $1(j \in \widehat{\Gamma})$ as the indicator function of $\widehat{\Gamma}$. By the concavity of the logarithm function, we have

$$\begin{aligned}
\log\left\{\widetilde{g}_{\text{AFTER},i,t}(y_{i,t}\mid \boldsymbol{y}_i^{(t-1)},\widehat{\Gamma})\right\} &= \log\left\{\sum_{j\in\widehat{\Gamma}}\widetilde{w}_{i,t,j}\widehat{g}_{i,t,j}(y_{i,t}\mid \boldsymbol{y}_i^{(t-1)})\right\}\\
&\ge \sum_{j\in\widehat{\Gamma}}\widetilde{w}_{i,t,j}\log\left\{\widehat{g}_{i,t,j}(y_{i,t}\mid \boldsymbol{y}_i^{(t-1)})\right\}.
\end{aligned}$$

Then, by (A.4), for every $j \in \{1,\ldots,J\}$

$$\begin{aligned}
&\frac{1}{n(T-T_0)}\sum_{i=1}^{n}\log\left(\frac{H_{ni}}{\widetilde{B}_{ni}}\right)\\
&\le 1(j\in\widehat{\Gamma})\left[\frac{\log(\widehat{J}_1)}{T-T_0}+\frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\log\left\{\frac{f_{i,t}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}}{n(T-T_0)}\right]
\end{aligned}$$

$$
\begin{aligned}
&+1(j \notin \widehat{\Gamma})\frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\left[\log\left\{f(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})\right\} - \log\left\{\widetilde{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \widehat{\Gamma})\right\}\right]}{n(T-T_0)} \\
&\leq 1(j \in \widehat{\Gamma})\left[\frac{\log(\widehat{J}_1)}{T-T_0} + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\log\left\{\frac{f_{i,t}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}}{n(T-T_0)}\right] \\
&\quad +1(j \notin \widehat{\Gamma})\frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\left[\log\left\{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})\right\} - \sum_{j' \in \widehat{\Gamma}}\widetilde{w}_{i,t,j'}\log\left\{\widehat{g}_{i,t,j'}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})\right\}\right]}{n(T-T_0)} \\
&= \frac{\log(\widehat{J}_1)}{T-T_0} + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\log\left\{\frac{f_{i,t}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}}{n(T-T_0)} \\
&\quad +1(j \notin \widehat{\Gamma})\left\{-\frac{\log(\widehat{J}_1)}{T-T_0} + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\sum_{j' \in \widehat{\Gamma}}\widetilde{w}_{i,t,j'}\left[\log\left\{\frac{\widehat{g}_{i,t,j}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j'}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}\right]}{n(T-T_0)}\right\} \\
&\leq \frac{\log(\widehat{J}_1)}{T-T_0} + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\log\left\{\frac{f_{i,t}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}}{n(T-T_0)} \\
&\quad + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\sum_{j' \in \widehat{\Gamma}}\widetilde{w}_{i,t,j'}1(j \notin \widehat{\Gamma})\log\left\{\frac{\widehat{g}_{i,t,j}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j'}(y_{i,t}|\boldsymbol{y}_i^{(t-1)})}\right\}}{n(T-T_0)},
\end{aligned}
$$

where the last inequality results from both $\log(\widehat{J}_1) \geq 0$ and $\sum_{j' \in \widehat{\Gamma}}\widetilde{w}_{i,t,j'} = 1$. In the light of this, given the fact that $\widetilde{w}_{i,t,j}$, $\widehat{\theta}_{i,t,j}$ and $\widehat{\Gamma}$ are evaluated based on $\mathcal{G}_{n,t-1}$, we have, for every $j \in \{1,\ldots,J\}$

$$
\begin{aligned}
&\widetilde{R}_{nT}^{\text{AFTER}} \\
&= \frac{1}{n(T-T_0)}\sum_{i=1}^{n}\mathbb{E}\left\{\log\left(\frac{H_{ni}}{\widetilde{B}_{ni}}\right)\right\} \\
&\leq \frac{\mathbb{E}\{\log(\widehat{J}_1)\}}{T-T_0} + R_{nT}(j) \\
&\quad + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\sum_{j'=1}^{J}\mathbb{E}\left[1(j' \in \widehat{\Gamma})\widetilde{w}_{i,t,j'}1(j \notin \widehat{\Gamma})\left\{\begin{array}{c} y_{i,t}\widehat{\theta}_{i,t,j} - b(\widehat{\theta}_{i,t,j}) \\ -y_{i,t}\widehat{\theta}_{i,t,j'} + b(\widehat{\theta}_{i,t,j'}) \end{array}\right\}\right]}{\phi n(T-T_0)} \\
&= \frac{\mathbb{E}\{\log(\widehat{J}_1)\}}{T-T_0} + R_{nT}(j) \\
&\quad + \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\sum_{j'=1}^{J}\mathbb{E}\left[\widetilde{w}_{i,t,j'}1(j' \in \widehat{\Gamma}, j \notin \widehat{\Gamma})\left\{\begin{array}{c} \mathbb{E}(y_{i,t} \mid \mathcal{G}_{n,t-1})(\widehat{\theta}_{i,t,j} - \widehat{\theta}_{i,t,j'}) \\ +b(\widehat{\theta}_{i,t,j'}) - b(\widehat{\theta}_{i,t,j}) \end{array}\right\}\right]}{\phi n(T-T_0)}
\end{aligned}
$$

$$
\begin{aligned}
&= \frac{\mathbb{E}\{\log(\widehat{J}_1)\}}{T-T_0} + R_{nT}(j) \\
&+ \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\sum_{j'=1}^{J}\mathbb{E}\left[\widetilde{w}_{i,t,j'}1(j' \in \widehat{\Gamma}, j \notin \widehat{\Gamma})\left\{\begin{array}{c}\mu_{i,t}(\boldsymbol{y}_i^{(t-1)})(\widehat{\theta}_{i,t,j}-\widehat{\theta}_{i,t,j'}) \\ +b(\widehat{\theta}_{i,t,j'})-b(\widehat{\theta}_{i,t,j})\end{array}\right\}\right]}{\phi n(T-T_0)} \\
&\le \frac{\mathbb{E}\{\log(\widehat{J}_1)\}}{T-T_0} + R_{nT}(j) \\
&+ \frac{\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\sum_{j'=1}^{J}\mathbb{E}\left[\widetilde{w}_{i,t,j'}1(j' \in \widehat{\Gamma}, j \notin \widehat{\Gamma})\left\{\begin{array}{l}c_1(|\widehat{\theta}_{i,t,j}|+|\widehat{\theta}_{i,t,j'}|) \\ \quad +|b(\widehat{\theta}_{i,t,j'})|+|b(\widehat{\theta}_{i,t,j})|\end{array}\right\}\right]}{\phi n(T-T_0)} \\
&\le \frac{\mathbb{E}\{\log(\widehat{J}_1)\}}{T-T_0} + R_{nT}(j) + \frac{\bar{c}_1\sum_{i=1}^{n}\sum_{t=T_0+1}^{T}\mathbb{E}\left[1(j \notin \widehat{\Gamma})\sum_{j'=1}^{J}1(j' \in \widehat{\Gamma})\widetilde{w}_{i,t,j'}\right]}{\phi n(T-T_0)} \\
&\le \mathbb{E}\left\{\frac{\log(\widehat{J}_1)}{T-T_0}\right\} + R_{nT}(j) + \frac{\bar{c}_1}{\phi}\mathrm{P}(j \notin \widehat{\Gamma}),
\end{aligned}
$$

which concludes the proof. □

Proof of Theorem 2. Since $\widehat{g}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})$ is evaluated based on $\mathcal{G}_{n,t-1}$ and is irrelevant to $y_{i,t}$, we have that

$$
\begin{aligned}
&\mathbb{E}_{y_{i,t}|\mathcal{G}_{n,t-1}}\left[\log\left\{\frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\}\right] \\
&= \mathbb{E}_{y_{i,t}|\boldsymbol{y}_i^{(t-1)}}\left[\log\left\{\frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\}\right] \\
&= \int f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\log\left\{\frac{f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}\right\}dv(y) \\
&= -2\int f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\log\left\{\frac{\widehat{g}^{1/2}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}{f^{1/2}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}\right\}dv(y) \\
&\ge 2\int f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\left\{1-\frac{\widehat{g}^{1/2}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}{f^{1/2}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}\right\}dv(y) \\
&= 2\int\left\{f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - f^{1/2}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\widehat{g}^{1/2}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}dv(y) \\
&= \int\left\{f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) + \widehat{g}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - 2f^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\widehat{g}^{1/2}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}dv(y) \\
&= \int\left\{f^{1/2}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - \widehat{g}^{1/2}_{\text{AFTER},i,t}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}^2 dv(y)
\end{aligned}
$$

almost surely. Moreover, similar to the proof of Lemma 1 of Yang (2004), under

Condition 1,

$$
\begin{aligned}
&\left\{\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t}^{\text{AFTER}}\right\}^2 \\
&= \left[\int y \left\{f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - \widehat{g}_{\text{AFTER}}(y \mid \boldsymbol{y}_i^{(t-1)})\right\} dv(y)\right]^2 \\
&\leq \left\{\int |y| \left|f_{i,t}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)}) - \widehat{g}_{\text{AFTER}}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\right| \left|f_{i,t}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)}) + \widehat{g}_{\text{AFTER}}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\right| dv(y)\right\}^2 \\
&\leq \int |y|^2 \left\{f_{i,t}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)}) + \widehat{g}_{\text{AFTER}}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}^2 dv(y) \int \left\{f_{i,t}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)}) - \widehat{g}_{\text{AFTER}}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}^2 dv(y) \\
&\leq 2\int |y|^2 \left\{f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) + \widehat{g}_{\text{AFTER}}(y \mid \boldsymbol{y}_i^{(t-1)})\right\} dv(y) \int \left\{f_{i,t}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)}) - \widehat{g}_{\text{AFTER}}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}^2 dv(y) \\
&\leq 4c_1(c_1+\phi) \int \left\{f_{i,t}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)}) - \widehat{g}_{\text{AFTER}}^{1/2}(y \mid \boldsymbol{y}_i^{(t-1)})\right\}^2 dv(y),
\end{aligned}
$$

almost surely. Then, we have

$$
\mathbb{E}_{y_{i,t}\mid\mathcal{G}_{n,t-1}}\left[\log\left\{\frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\}\right] \geq \frac{\left\{\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t}^{\text{AFTER}}\right\}^2}{4c_1(c_1+\phi)}, \tag{A.5}
$$

almost surely. In addition, for each $j \in \{1, \ldots, J\}$, since $\widehat{\theta}_{i,t,j}$ has nothing to do with $y_{i,t}$, it follows from Condition 1 that

$$
\begin{aligned}
&\mathbb{E}_{y_{i,t}\mid\mathcal{G}_{n,t-1}}\left[\log\left\{\frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\}\right] \\
&= \mathbb{E}_{y_{i,t}\mid\boldsymbol{y}_i^{(t-1)}}\left[\log\left\{\frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{i,t,j}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}\right\}\right] \\
&= \frac{\mu_{i,t}(\boldsymbol{y}_i^{(t-1)})\left\{\theta_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\theta}_{i,t,j}\right\} + b(\widehat{\theta}_{i,t,j}) - b\left\{\theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right\}}{\phi} \\
&= \frac{b^{(2)}(\widetilde{\theta}_{i,t,j})\left\{\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right\}^2}{2\phi} \\
&\leq \frac{c_1\left\{\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right\}^2}{2\phi},
\end{aligned} \tag{A.6}
$$

almost surely, where $\widetilde{\theta}_{i,t,j}$ lies between $\widehat{\theta}_{i,t,j}$ and $\theta_{i,t}(\boldsymbol{y}_i^{(t-1)})$. Furthermore, under Condition 1

$$
\begin{aligned}
\left|\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})\right| &= \left|b^{(1)}(\widehat{\theta}_{i,t,j}) - b^{(1)}\left\{\theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right\}\right| \\
&\geq c_0\left|\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right|,
\end{aligned} \tag{A.7}
$$

almost surely. Combining (A.3), (A.5), (A.6), and (A.7), we obtain that for every $j \in \{1, \ldots, J\}$,

$$
\begin{aligned}
&\frac{\sum_{i=1}^{n} \sum_{t=T_0+1}^{T} \mathbb{E}|\widehat{\mu}_{i,t}^{\text{AFTER}} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2}{n(T-T_0)} \\
&\leq \frac{4c_1(c_1+\phi)}{n(T-T_0)} \sum_{i=1}^{n} \sum_{t=T_0+1}^{T} \mathbb{E}_{\mathcal{G}_{n,t-1}} \left(\mathbb{E}_{y_{i,t}|\mathcal{G}_{n,t-1}} \left[\log \left\{ \frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widehat{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})} \right\} \right] \right) \\
&\leq \frac{4c_1(c_1+\phi)}{n(T-T_0)} \sum_{i=1}^{n} \log \left(\frac{1}{\pi_{i,j}} \right) + \frac{2c_1^2(c_1+\phi)}{c_0\phi} \frac{\sum_{i=1}^{n} \sum_{t=T_0+1}^{T} \mathbb{E}|\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2}{n(T-T_0)},
\end{aligned}
$$

and this concludes the proof. □

Proof of Corollary 2. Similar to the derivation of (A.5), note that $\widehat{\Gamma}$ is determined based on the initial information set $\mathcal{G}_{n,T_0}$. It can be verified that for each $i \geq 1$ and $t \geq T_0 + 1$,

$$
\mathbb{E}\left[\log \left\{ \frac{f_{i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)})}{\widetilde{g}_{\text{AFTER},i,t}(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \widehat{\Gamma})} \right\} \right] \geq \frac{\mathbb{E}\left| \widetilde{\mu}_{i,t}^{\text{AFTER}} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) \right|^2}{4c_1(c_1+\phi)}. \tag{A.8}
$$

Then, under Condition 1, it follows from (A.8) that

$$
\sum_{i=1}^{n} \sum_{t=T_0+1}^{T} \frac{\mathbb{E}\left| \widetilde{\mu}_{i,t}^{\text{AFTER}} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) \right|^2}{n(T-T_0)} \leq 4c_1(c_1+\phi) \frac{\sum_{i=1}^{n} \mathbb{E}\left\{ \log\left(\frac{H_{ni}}{\widetilde{B}_{ni}} \right) \right\}}{n(T-T_0)} \tag{A.9}
$$

In addition, under Condition 1, with (A.6), (A.7), and the proof of Corollary 1, we have that for each $j \in \{1, \ldots, J\}$

$$
\begin{aligned}
&\frac{\sum_{i=1}^{n} \mathbb{E}\left\{ \log\left(\frac{H_{ni}}{\widetilde{B}_{ni}} \right) \right\}}{n(T-T_0)} \\
&\leq \frac{\mathbb{E}\{\log(\widehat{J}_1)\}}{T-T_0} + \frac{c_1 \sum_{i=1}^{n} \sum_{t=T_0+1}^{T} \mathbb{E}\left| \widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) \right|^2}{2n(T-T_0)\phi c_0} + \frac{\bar{c}_1}{\phi} \mathrm{P}(j \notin \widehat{\Gamma}).
\end{aligned} \tag{A.10}
$$

The proof is concluded by combining inequalities (A.9)–(A.10). □

Proof of Theorem 3. Denote $L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda) = \exp\left(-\lambda \left|y_{i,t} - \hat{\mu}_{i,t,j}\right|^2\right)$ and

$$
D_{ni} = \sum_{j=1}^{J} \pi_{i,j} \prod_{t=T_0+1}^{T} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda).
$$

It is clear that

$$
\begin{aligned}
D_{ni} &= \sum\nolimits_{j=1}^{J} \pi_{i,j} L_j(y_{i,T_0+1} \mid \boldsymbol{y}_i^{(T_0)}, \lambda) \times \frac{\sum_{j=1}^{J} \pi_{i,j} L_j(y_{i,T_0+1} \mid \boldsymbol{y}_i^{(T_0)}, \lambda) L_j(y_{i,T_0+2} \mid \boldsymbol{y}_i^{(T_0+1)}, \lambda)}{\sum_{j=1}^{J} \pi_{i,j} L_j(y_{i,T_0+1} \mid \boldsymbol{y}_i^{(T_0)}, \lambda)} \\
&\quad \times \ldots \times \frac{\sum_{j=1}^{J} \pi_{i,j} \prod_{t=T_0+1}^{T} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda)}{\sum_{j=1}^{J} \pi_{i,j} \prod_{t=T_0+1}^{T-1} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda)} \\
&= \prod_{t=T_0+1}^{T} \sum\nolimits_{j=1}^{J} \lambda_{i,t,j} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda).
\end{aligned}
$$

Define $h_{i,t}(j) = -(y_{i,t} - \widehat{\mu}_{i,t,j})^2$. Moreover, let $\xi_{i,t}$ and $\xi_{i,t,\gamma}$ be two random variables that take values from the set of sub-indices $\{1, \ldots, J\}$ and satisfy

$$
\mathrm{P}\left(\xi_{i,t} = j\right) = \lambda_{i,t,j},
$$

and

$$
\mathrm{P}\left(\xi_{i,t,\gamma} = j\right) = \frac{\lambda_{i,t,j} \exp\left\{\gamma h_{i,t}(j)\right\}}{\sum_{j=1}^{J} \lambda_{i,t,j} \exp\left\{\gamma h_{i,t}(j)\right\}},
$$

respectively, where $\gamma \in [0, \lambda]$ is a non-stochastic constant. It follows from Lemma 3.6.1 of Catoni (2004) that

$$
\begin{aligned}
&\log(D_{ni}) \\
&= \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}_{\xi_{i,t}} \left[\exp\left\{\lambda h_{i,t}(\xi_{i,t})\right\}\right] \\
&\leq \sum\nolimits_{t=T_0+1}^{T} \lambda \mathbb{E}_{\xi_{i,t}} \left\{h_{i,t}(\xi_{i,t})\right\} \\
&+ \frac{\lambda^2}{2} \sum\nolimits_{t=T_0+1}^{T} M_{2,\xi_{i,t}} \left\{h_{i,t}(\xi_{i,t})\right\} \exp\left(\lambda \max\left[0, \sup_{\gamma \in [0,\lambda]} \frac{M_{3,\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\}}{M_{2,\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\}}\right]\right)
\end{aligned}
\tag{A.11}
$$

where

$$
M_{r,\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\} = \mathbb{E}_{\xi_{i,t,\gamma}} \left[h_{i,t}(\xi_{i,t,\gamma}) - \mathbb{E}_{\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\}\right]^r, r = 2, 3, \ldots,
$$

$M_{2,\xi_{i,t}} \left\{h_{i,t}(\xi_{i,t})\right\} = M_{2,\xi_{i,t,0}} \left\{h_{i,t}(\xi_{i,t,0})\right\}$, and $\mathbb{E}_{\xi_{i,t}}(\cdot)$ (or $\mathbb{E}_{\xi_{i,t,\gamma}}(\cdot)$) means taking expectation with respect to $\xi_{i,t}$ (or $\xi_{i,t,\gamma}$). Let $z_{i,t} = y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})$. Under Condition 1, we have

$$
\begin{aligned}
&\sup_{\gamma \in [0,\lambda]} \frac{M_{3,\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\}}{M_{2,\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\}} \\
&\leq \sup_{\gamma \in [0,\lambda]} \sup_{1 \leq j \leq J} \left|h_{i,t}(j) - \mathbb{E}_{\xi_{i,t,\gamma}} \left\{h_{i,t}(\xi_{i,t,\gamma})\right\}\right| \\
&= \sup_{\gamma \in [0,\lambda]} \sup_{1 \leq j \leq J} \left|\mathbb{E}_{\xi_{i,t,\gamma}} \left\{h_{i,t}(j) - h_{i,t}(\xi_{i,t,\gamma})\right\}\right|
\end{aligned}
$$

$$
\begin{aligned}
&\le \sup_{\gamma\in[0,\lambda]} \sup_{1\le j\le J} \mathbb{E}_{\xi_{i,t,\gamma}} \left|h_{i,t}(j) - h_{i,t}(\xi_{i,t,\gamma})\right| \\
&\le \sup_{1\le j,j'\le J} \left|h_{i,t}(j) - h_{i,t}(j')\right| \\
&= \sup_{1\le j,j'\le J} \left|(y_{i,t} - \widehat{\mu}_{i,t,j})^2 - (y_{i,t} - \widehat{\mu}_{i,t,j'})^2\right| \\
&\le 4\left|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})\right| \sup_{1\le j\le J}\left|\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t,j}\right| + 2\sup_{1\le j\le J}\left|\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t,j}\right|^2 \\
&\le 4|z_{i,t}|c_2 + 2c_2^2. \qquad\qquad (A.12)
\end{aligned}
$$

In addition, by Jensen's inequality and Condition 1, it is seen that

$$
\begin{aligned}
&M_{2,\xi_{i,t}}\left\{h_{i,t}(\xi_{i,t})\right\} \\
&\le \mathbb{E}_{\xi_{i,t}}\left\{|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 - |y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2\right\}^2 \\
&\le \mathbb{E}_{\xi_{i,t}}\left[\left\{2|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}| + |\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|\right\}\left|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})\right|\right]^2 \\
&\le \mathbb{E}_{\xi_{i,t}}\left[\left\{2|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| + 3|\widehat{\mu}_{i,t,\xi_{i,t}} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| + |\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|\right\}\right. \\
&\qquad \left.\times\left|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})\right|\right]^2 \\
&\le \left\{2|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| + 4\sup_{1\le j\le J}|\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t,j}|\right\}^2 \mathbb{E}_{\xi_{i,t}}\left|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})\right|^2 \\
&\le (2|z_{i,t}| + 4c_2)^2\, \mathbb{E}_{\xi_{i,t}}\left|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})\right|^2. \qquad\qquad (A.13)
\end{aligned}
$$

Now, it follows from (A.11)–(A.13) that

$$
\begin{aligned}
&\log(D_{ni}) \\
&\le -\lambda \sum_{t=T_0+1}^{T} \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 \\
&\quad + \frac{\lambda^2\exp(2\lambda c_2^2)}{2}\sum_{t=T_0+1}^{T}(2|z_{i,t}| + 4c_2)^2\exp(4\lambda c_2|z_{i,t}|)\mathbb{E}_{\xi_{i,t}}|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2, \\
&\qquad\qquad (A.14)
\end{aligned}
$$

almost surely. Note that under Condition 2 and (14), for each $t = T_0+1,\dots,T$, we have that

$$
\mathbb{E}^{1/2}_{y_{i,t}|\mathcal{G}_{n,t-1}}(2|z_{i,t}| + 4c_2)^4 \le 8\mathbb{E}^{1/2}_{y_{i,t}|\mathcal{G}_{n,t-1}}|z_{i,t}|^4 + 32c_2^2 \le 32(c_2^2 + 4c_3^2) \qquad (A.15)
$$

and

$$
\begin{aligned}
\mathbb{E}_{y_{i,t}|\mathcal{G}_{n,t-1}}\left\{\exp(8\lambda c_2|z_{i,t}|)\right\} &= 1 + \sum_{p=1}^{\infty}\frac{(8c_2)^p\lambda^p\mathbb{E}_{y_{i,t}|\mathcal{G}_{n,t-1}}(|z_{i,t}|^p)}{p!} \\
&\le 1 + \sum_{p=1}^{\infty}(8c_2c_3\lambda e)^p \\
&\le 2 \qquad\qquad (A.16)
\end{aligned}
$$

hold almost surely, where we have used the fact that $p! \geq (p/e)^p$ (Vershynin, 2010). Take $\mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}}(\cdot)$ on both sides of (A.14), with (A.15) and (A.16). Under Conditions 1–2, we have

$$
\begin{aligned}
&\mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}}\{\log(D_{ni})\} \\
&\leq -\lambda \sum_{t=T_0+1}^{T-1} \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 \\
&\quad + \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \sum_{t=T_0+1}^{T-1} (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|) \mathbb{E}_{\xi_{i,t}}|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \\
&\quad -\lambda \mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}} \left(\mathbb{E}_{\xi_{i,T}}|y_{i,T} - \widehat{\mu}_{i,T,\xi_{i,T}}|^2\right) \\
&\quad + \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \mathbb{E}_{\xi_{i,T}}|\widehat{\mu}_{i,T,\xi_{i,T}} - \mathbb{E}_{\xi_{i,T}}(\widehat{\mu}_{i,T,\xi_{i,T}})|^2 \mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}} \left\{(2|z_{it}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,T}|)\right\} \\
&\leq -\lambda \sum_{t=T_0+1}^{T-1} \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 \\
&\quad + \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \sum_{t=T_0+1}^{T-1} (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|) \mathbb{E}_{\xi_{i,t}}|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \\
&-\lambda \mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}} \left\{\mathbb{E}_{\xi_{i,T}}|y_{i,T} - \widehat{\mu}_{i,T,\xi_{i,T}}|^2\right\} \\
&+ \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \mathbb{E}_{\xi_{i,T}}|\widehat{\mu}_{i,T,\xi_{i,T}} - \mathbb{E}_{\xi_{i,T}}(\widehat{\mu}_{i,T,\xi_{i,T}})|^2 \mathbb{E}^{1/2}_{y_{i,T}|\mathcal{G}_{n,T-1}} \left\{(2|z_{i,T}| + 4c_2)^4\right\} \\
&\times \mathbb{E}^{1/2}_{y_{i,T}|\mathcal{G}_{n,T-1}} \left\{\exp(8\lambda c_2 |z_{i,T}|)\right\} \\
&\leq -\lambda \sum_{t=T_0+1}^{T-1} \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 \\
&\quad + \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \sum_{t=T_0+1}^{T-1} (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|) \mathbb{E}_{\xi_{i,t}}|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \\
&\quad -\lambda \mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}} \left\{\mathbb{E}_{\xi_{i,T}}|y_{i,T} - \widehat{\mu}_{i,T,\xi_{i,T}}|^2\right\} \\
&\quad + 16\sqrt{2}\lambda^2 \exp(2\lambda c_2^2)(c_2^2 + 4c_3^2) \mathbb{E}_{\xi_{i,T}}|\widehat{\mu}_{i,T,\xi_{i,T}} - \mathbb{E}_{\xi_{i,T}}(\widehat{\mu}_{i,T,\xi_{i,T}})|^2,
\end{aligned}
$$

almost surely. In addition, in view of that,

$$
\begin{aligned}
&\mathbb{E}_{\xi_{i,t}}|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 - |y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \\
&= \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 - \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \\
&= \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}}) + \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}}) - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 - \mathbb{E}_{\xi_{i,t}}|y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \\
&= \mathbb{E}_{\xi_{i,t}}|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 + 2\mathbb{E}_{\xi_{i,t}} \left[\{y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})\}\{\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})\}\right] \\
&= \mathbb{E}_{\xi_{i,t}}|\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2,
\end{aligned}
$$

for each $t \geq T_0 + 1$. It follows from (14) that

$$
\mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}}\{\log(D_{ni})\}
$$

$$\leq \sum_{t=T_0+1}^{T-1} \Big\{ -\lambda \mathbb{E}_{\xi_{i,t}} |y_{i,t} - \widehat{\mu}_{i,t,\xi_{i,t}}|^2 + \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|)$$
$$\times \mathbb{E}_{\xi_{i,t}} |\widehat{\mu}_{i,t,\xi_{i,t}} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2 \Big\} - \lambda \mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}} |y_{i,T} - \mathbb{E}_{\xi_{i,T}}(\widehat{\mu}_{i,T,\xi_{i,T}})|^2,$$

almost surely. Similarly, by successively taking $\mathbb{E}_{y_{i,T-1}|\mathcal{G}_{n,T-2}}(\cdot),\ldots,\mathbb{E}_{y_{i,T_0+1}|\mathcal{G}_{n,T_0}}(\cdot)$ on both sides of (A.14), it is readily seen that

$$\mathbb{E}_{y_{i,T_0+1}|\mathcal{G}_{n,T_0}} \left(\mathbb{E}_{y_{i,T_0+2}|\mathcal{G}_{n,T_0+1}} \left[\ldots \mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}} \{\log(D_{ni})\} \ldots \right] \right)$$
$$\leq -\lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}_{y_{i,t}|\mathcal{G}_{n,t-1}} |y_{i,t} - \mathbb{E}_{\xi_{i,t}}(\widehat{\mu}_{i,t,\xi_{i,t}})|^2,$$

almost surely, which yields that

$$-\mathbb{E}\{\log(D_{ni})\} \geq \lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}|y_{i,t} - \widehat{\mu}_{i,t}^{\text{AFTQR}}|^2. \tag{A.17}$$

Inequality in (A.17) implies that

$$\mathbb{E}\left(\log \left[\frac{\prod_{t=T_0+1}^{T} \exp\left\{ -\lambda |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 \right\}}{D_{ni}} \right] \right)$$
$$\geq -\lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 + \lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}|y_{i,t} - \widehat{\mu}_{i,t}^{\text{AFTQR}}|^2$$
$$= \lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}|\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t}^{\text{AFTQR}}|^2. \tag{A.18}$$

Now, in light of (A.18), by the definition of D_{ni}, it is seen that for each $j \in \{1,\ldots,J\}$,

$$\lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}|\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t}^{\text{AFTQR}}|^2$$
$$\leq \mathbb{E}\left(\log \left[\frac{\prod_{t=T_0+1}^{T} \exp\left\{ -\lambda |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 \right\}}{D_{ni}} \right] \right)$$
$$= \mathbb{E}\left(\log \left[\frac{\prod_{t=T_0+1}^{T} \exp\left\{ -\lambda |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 \right\}}{\sum_{j'=1}^{J} \pi_{i,j'} \prod_{t=T_0+1}^{T} \exp\left(-\lambda |y_{i,t} - \widehat{\mu}_{i,t,j'}|^2 \right)} \right] \right)$$
$$\leq \mathbb{E}\left[\log \left\{ \frac{\prod_{t=T_0+1}^{T} \exp\left(-\lambda |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 \right)}{\pi_{i,j} \prod_{t=T_0+1}^{T} \exp\left(-\lambda |y_{i,t} - \widehat{\mu}_{i,t,j}|^2 \right)} \right\} \right]$$
$$= \log\left(\frac{1}{\pi_{i,j}} \right) + \lambda \sum\nolimits_{t=T_0+1}^{T} \mathbb{E}|\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t,j}|^2,$$

which concludes the proof. $\square$

Proof of Corollary 3. Denote $\widetilde{D}_{ni} = \sum_{j\in\widehat{\Gamma}} \widetilde{\pi}_{i,j} \prod_{t=T_0+1}^{T} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda)$. It is clear that

$$\widetilde{D}_{ni} = \prod_{t=T_0+1}^{T} \sum_{j\in\widehat{\Gamma}} \widetilde{\lambda}_{i,t,j} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda).$$

Let $\widetilde{\xi}_{i,t}$ and $\widetilde{\xi}_{i,t,\gamma}$ be two random variables such that for each $j \in \widehat{\Gamma}$,

$$\mathrm{P}\left(\widetilde{\xi}_{i,t} = j\right) = \widetilde{\lambda}_{i,t,j}$$

and

$$\mathrm{P}\left(\widetilde{\xi}_{i,t,\gamma} = j\right) = \frac{\lambda_{i,t,j} \exp\{\gamma h_{i,t}(j)\}}{\sum_{j\in\widehat{\Gamma}} \lambda_{i,t,j} \exp\{\gamma h_{i,t}(j)\}},$$

respectively. Then, $\widetilde{\mu}_{i,t}^{\text{AFTQR}} = \mathbb{E}_{\widetilde{\xi}_{i,t}}(\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}})$. By the same reasoning used in the proof of (A.14), it can be verified that under Conditions 1–2,

$$\begin{aligned}
&\log(\widetilde{D}_{ni}) \\
&\leq \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \sum_{t=T_0+1}^{T} (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|) \mathbb{E}_{\widetilde{\xi}_{i,t}} |\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}} - \mathbb{E}_{\widetilde{\xi}_{i,t}}(\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}})|^2 \\
&\quad -\lambda \sum_{t=T_0+1}^{T} \mathbb{E}_{\widetilde{\xi}_{i,t}} |y_{i,t} - \widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}}|^2, \qquad \text{(A.19)}
\end{aligned}$$

almost surely. In addition, by the definition of $\widetilde{D}_{ni}$ and the concavity of logarithm function,

$$\begin{aligned}
&\log\left[\frac{\prod_{t=T_0+1}^{T} \exp\{-\lambda |y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2\}}{\widetilde{D}_{ni}}\right] \\
&= -\lambda \sum_{t=T_0+1}^{T} |y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2 - \sum_{t=T_0+1}^{T} \log\left\{\sum_{j\in\widehat{\Gamma}} \widetilde{\lambda}_{i,t,j} L_j(y_{i,t} \mid \boldsymbol{y}_i^{(t-1)}, \lambda)\right\} \\
&\leq -\lambda \sum_{t=T_0+1}^{T} |y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2 + \lambda \sum_{t=T_0+1}^{T} \sum_{j\in\widehat{\Gamma}} \widetilde{\lambda}_{i,t,j} |y_{i,t} - \widehat{\mu}_{i,t,j}|^2. \qquad \text{(A.20)}
\end{aligned}$$

Moreover, by the definition of $\widetilde{D}_{ni}$, for each $j \in \widehat{\Gamma}$, we have

$$\begin{aligned}
&\log\left[\frac{\prod_{t=T_0+1}^{T} \exp\{-\lambda |y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2\}}{\widetilde{D}_{ni}}\right] \\
&\leq \log\left(\frac{1}{\widetilde{\pi}_{i,j}}\right) - \lambda \sum_{t=T_0+1}^{T} |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 \\
&\quad +\lambda \sum_{t=T_0+1}^{T} |y_{i,t} - \widehat{\mu}_{i,t,j}|^2. \qquad \text{(A.21)}
\end{aligned}$$

Now, by taking $\widetilde{\pi}_{i,j} = 1/\widehat{J}_1$, it follows from (A.19), (A.20), and (A.21) that for each $j \in \{1, \ldots, J\}$,

$$
\begin{aligned}
&\lambda \sum\nolimits_{t=T_0+1}^{T} \left\{ -|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 + \mathbb{E}_{\widetilde{\xi}_{i,t}} |y_{i,t} - \widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}}|^2 \right\} \\
&- \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \sum\nolimits_{t=T_0+1}^{T} \left\{ (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|) \mathbb{E}_{\widetilde{\xi}_{i,t}} |\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}} - \mathbb{E}_{\widetilde{\xi}_{i,t}} (\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}})|^2 \right\} \\
&\leq 1(j \in \widehat{\Gamma}) \left\{ \log\left(\widehat{J}_1\right) - \lambda \sum\nolimits_{t=T_0+1}^{T} |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 + \lambda \sum\nolimits_{t=T_0+1}^{T} |y_{i,t} - \widehat{\mu}_{i,t,j}|^2 \right\} \\
&\quad + 1(j \notin \widehat{\Gamma}) \lambda \sum\nolimits_{t=T_0+1}^{T} \left\{ -|y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2 + \sum_{s \in \widehat{\Gamma}} \widetilde{\lambda}_{i,t,(s)} |y_{i,t} - \widehat{\mu}_{i,t,(s)}|^2 \right\}. \qquad \text{(A.22)}
\end{aligned}
$$

Then, take $\mathbb{E}_{y_{i,T}|\mathcal{G}_{n,T-1}}(\cdot)$ on both sides of (A.22). With Condition 2 and (14), we have

$$
\begin{aligned}
&\lambda \sum_{t=T_0+1}^{T-1} \left\{ -|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 + \mathbb{E}_{\widetilde{\xi}_{i,t}} |y_{i,t} - \widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}}|^2 \right\} \\
&- \frac{\lambda^2 \exp(2\lambda c_2^2)}{2} \sum_{t=T_0+1}^{T-1} \left\{ (2|z_{i,t}| + 4c_2)^2 \exp(4\lambda c_2 |z_{i,t}|) \mathbb{E}_{\widetilde{\xi}_{i,t}} |\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}} - \mathbb{E}_{\widetilde{\xi}_{i,t}} (\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}})|^2 \right\} \\
&+ \lambda |\mu_{i,T}(\boldsymbol{y}_i^{(T-1)}) - \mathbb{E}_{\widetilde{\xi}_{i,T}} (\widehat{\mu}_{i,T,\widetilde{\xi}_{i,T}})|^2 \\
&\leq 1(j \in \widehat{\Gamma}) \left\{ \log\left(\widehat{J}_1\right) - \lambda \sum_{t=T_0+1}^{T-1} |y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 + \lambda \sum_{t=T_0+1}^{T-1} |y_{i,t} - \widehat{\mu}_{i,t,j}|^2 \right\} \\
&\quad + 1(j \notin \widehat{\Gamma}) \lambda \sum_{t=T_0+1}^{T-1} \left\{ -|y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2 + \sum_{s \in \widehat{\Gamma}} \widetilde{\lambda}_{i,t,(s)} |y_{i,t} - \widehat{\mu}_{i,t,(s)}|^2 \right\} \\
&\quad + 1(j \in \widehat{\Gamma}) \left\{ \lambda |\mu_{i,T}(\boldsymbol{y}_i^{(T-1)}) - \widehat{\mu}_{i,T,j}|^2 \right\} + 1(j \notin \widehat{\Gamma}) \lambda \sum_{s \in \widehat{\Gamma}} \widetilde{\lambda}_{i,T,s} |\mu_{i,T}(\boldsymbol{y}_i^{(T-1)}) - \widehat{\mu}_{i,T,s}|^2 \\
&\leq \log\left(\widehat{J}_1\right) + 1(j \in \widehat{\Gamma}) \lambda \sum_{t=T_0+1}^{T-1} \left\{ -|y_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|^2 + |y_{i,t} - \widehat{\mu}_{i,t,j}|^2 \right\} \\
&\quad + 1(j \notin \widehat{\Gamma}) \lambda \sum_{t=T_0+1}^{T-1} \left\{ -|y_{i,t} - \mu_{i,t}(y_i^{(t-1)})|^2 + \sum_{s \in \widehat{\Gamma}} \widetilde{\lambda}_{i,t,(s)} |y_{i,t} - \widehat{\mu}_{i,t,(s)}|^2 \right\} \\
&\quad + \lambda |\mu_{i,T}(\boldsymbol{y}_i^{(T-1)}) - \widehat{\mu}_{i,T,j}|^2 + c_2^2 \lambda 1(j \notin \widehat{\Gamma}),
\end{aligned}
$$

almost surely. Similarly, by successively taking $\mathbb{E}_{y_{i,T-1}|\mathcal{G}_{n,T-2}}(\cdot)$,$\ldots$,$\mathbb{E}_{y_{i,T_0+1}|\mathcal{G}_{n,T_0}}(\cdot)$ on both sides of (A.22), it is readily seen that for each $j \in \{1, \ldots, J\}$

$$
\begin{aligned}
&\sum\nolimits_{t=T_0+1}^{T} \mathbb{E} \left\{ |\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \mathbb{E}_{\widetilde{\xi}_{i,t}} (\widehat{\mu}_{i,t,\widetilde{\xi}_{i,t}})|^2 \right\} \\
&\leq \frac{1}{\lambda} \mathbb{E} \left\{ \log\left(\widehat{J}_1\right) \right\} + \sum\nolimits_{t=T_0+1}^{T} \mathbb{E} |\mu_{i,t}(\boldsymbol{y}_i^{(t-1)}) - \widehat{\mu}_{i,t,j}|^2 + c_2^2 (T - T_0) \mathrm{P}(j \notin \widehat{\Gamma}),
\end{aligned}
$$

and this concludes the proof. $\square$

Proof of Lemma 1. In fact, for $t_0 \le t \le T$, $1 \le i \le n$, and $1 \le j \le J_0 - 1$, under Condition 3, with uniform prior assumption, it is readily seen that

$$\begin{aligned} a_{i,t,j} &= \frac{\exp\left(-l_{i,j}^{(t-1)}\right)}{\sum_{j'=1}^{J} \exp\left(-l_{i,j'}^{(t-1)}\right)} \\ &= \frac{1}{\sum_{j'=1}^{J} \exp\left(l_{i,j}^{(t-1)} - l_{i,j'}^{(t-1)}\right)} \\ &\le \frac{1}{\sum_{j'=J_0}^{J} \exp\left\{(t-T_0-1)\frac{l_{i,j}^{(t-1)} - l_{i,j'}^{(t-1)}}{t-T_0-1}\right\}} \\ &\le \frac{\exp\{-c_0(t-T_0-1)\}}{J-J_0+1}, \end{aligned}$$

and this concludes the proof. $\square$

Proof of Theorem 4. Under Conditions 1 and 3, it follows immediately from Lemma 1 that for each $t_0 + 1 \le t \le T$,

$$\begin{aligned} & |\widehat{m}_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| \\ &= \left|\sum_{j=1}^{J} a_{i,t,j}\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})\right| \\ &\le \sum_{j=1}^{J_0-1} a_{i,t,j}|\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| + \sum_{j=J_0}^{J} a_{i,t,j}|\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| \\ &\le \frac{2c_1 J_0 \exp\{-c_0(t-T_0-1)\}}{J-J_0+1} + \max_{1\le i\le n, t_0+1\le t\le T, J_0\le j\le J} |\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| \left(1 - \sum_{j=1}^{J_0-1} a_{i,t,j}\right) \\ &\le \frac{2c_1 J_0 \exp\{-c_0(t-T_0-1)\}}{J-J_0+1} + \max_{1\le i\le n, t_0+1\le t\le T, J_0\le j\le J} |\widehat{\mu}_{i,t,j} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})| \\ &\le \frac{2c_1 J_0 \exp\{-c_0(t-T_0-1)\}}{J-J_0+1} + c_1\alpha_n, \end{aligned}$$

where the last step comes from the definition of pseudo-true aggregator and this yields the first inequality. Moreover, under Conditions 1 and 3,

$$\begin{aligned} & \frac{\sum_{t=T_0+1}^{T} |\widehat{m}_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|}{T-T_0} \\ &\le \frac{\sum_{t=T_0+1}^{t_0-1} |\widehat{m}_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|}{T-T_0} + \frac{\sum_{t=t_0}^{T} |\widehat{m}_{i,t} - \mu_{i,t}(\boldsymbol{y}_i^{(t-1)})|}{T-T_0} \\ &\le \frac{2c_1(t_0-T_0-1)}{T-T_0} + \frac{2c_1 J_0 \sum_{t=t_0}^{T} \exp\{-c_0(t-T_0-1)\}}{(J-J_0+1)(T-T_0)} + \frac{c_1\alpha_n(T-t_0+1)}{T-T_0} \end{aligned}$$

$$\leq \frac{2c_1(t_0 - T_0 - 1)}{T - T_0} + \frac{2c_1 J_0 e^{-c_0(t_0 - T_0 - 1)} \left\{1 - e^{-c_0(T - t_0 + 1)}\right\}}{(J - J_0 + 1)(1 - e^{-c_0})(T - T_0)} + c_1\alpha_n,$$

and when $J/J_0 > 1 + \delta$, the quantity on the right-hand side of this inequality goes to 0 almost surely, as $n, T \to \infty$. This confirms the second inequality. Next, we verify the third inequality. To this end, note that

$$\begin{aligned}
&\left| \frac{\sum_{t=T_0+1}^{T} \widehat{G}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}{T - T_0} - \frac{\sum_{t=T_0+1}^{T} F_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})}{T - T_0} \right| \\
&\leq \frac{\sum_{t=T_0+1}^{t_0-1} \left| \widehat{G}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - F_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) \right|}{T - T_0} \\
&\quad + \frac{\sum_{t=t_0}^{T} \left| \widehat{G}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - F_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) \right|}{T - T_0} \\
&\leq \frac{2(t_0 - T_0 - 1)}{T - T_0} + \frac{\sum_{t=t_0}^{T} \left| \widehat{G}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - F_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) \right|}{T - T_0}. \qquad \text{(A.23)}
\end{aligned}$$

Moreover, for each $t \geq t_0$, it is seen that

$$\begin{aligned}
&\left| \widehat{G}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - F_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) \right| \\
&\leq \sum_{j=1}^{J} a_{i,t,j} \int 1(z \leq y) \left| \widehat{g}_{i,t,j}(z \mid \boldsymbol{y}_i^{(t-1)}) - f_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) \right| dv(z) \\
&\leq \sum_{j=1}^{J_0-1} a_{i,t,j} \int 1(z \leq y) \left| \widehat{g}_{i,t,j}(z \mid \boldsymbol{y}_i^{(t-1)}) - f_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) \right| dv(z) \\
&+ \sum_{j=J_0}^{J} a_{i,t,j} \int 1(z \leq y) \left| \widehat{g}_{i,t,j}(z \mid \boldsymbol{y}_i^{(t-1)}) - f_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) \right| dv(z). \qquad \text{(A.24)}
\end{aligned}$$

Under the setup of (2) and (6), one has

$$\begin{aligned}
&\int 1(z \leq y) \left| \widehat{g}_{i,t,j}(z \mid \boldsymbol{y}_i^{(t-1)}) - f_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) \right| dv(z) \\
&= \frac{1}{\phi} \int 1(z \leq y) \left| \widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)}) \right| \left| z - b^{(1)}(\bar{\theta}_{i,t,j}) \right| \bar{g}_{i,t,j}(z \mid \bar{\theta}_{i,t,j}) dv(z),
\end{aligned}$$

where $\bar{\theta}_{i,t,j}$ lies between $\widehat{\theta}_{i,t,j}$ and $\theta_{i,t}(\boldsymbol{y}_i^{(t-1)})$, and

$$\bar{g}_{i,t,j}(z \mid \bar{\theta}_{i,t,j}) = \exp\left\{ \frac{z\bar{\theta}_{i,t,j} - b(\bar{\theta}_{i,t,j})}{\phi} + c(z, \phi) \right\}.$$

Then, under Condition 1,

$$\int 1(z \leq y) \left| \widehat{g}_{i,t,j}(z \mid \boldsymbol{y}_i^{(t-1)}) - f_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) \right| dv(z)$$

$$
\begin{aligned}
&\leq \frac{\left|\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right|}{\phi} \int \left| z - b^{(1)}(\bar{\theta}_{i,t,j}) \right| \bar{g}_{i,t,j}(z \mid \bar{\theta}_{i,t,j}) dv(z) \\
&\leq \frac{\left|\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right|}{\phi} \left\{ \int \left| z - b^{(1)}(\bar{\theta}_{i,t,j}) \right|^2 \bar{g}_{i,t,j}(z \mid \bar{\theta}_{i,t,j}) dv(z) \right\}^{1/2} \\
&= \frac{\left\{ b^{(2)}(\bar{\theta}_{i,t,j}) \right\}^{1/2} \left|\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right|}{\phi} \\
&\leq \frac{c_1^{1/2} \left|\widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)})\right|}{\phi},
\end{aligned} \tag{A.25}
$$

almost surely. Combine (A.25) with (A.24). It follows from Lemma 1 that for each $1 \leq i \leq n$,

$$
\begin{aligned}
&\frac{\sum_{t=t_0}^{T} \left| \widehat{G}_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) - F_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) \right|}{T - T_0} \\
&\leq \frac{\sum_{t=t_0}^{T} \sum_{j=1}^{J_0 - 1} a_{i,t,j} \int 1(z \leq y) \left\{ \widehat{g}_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) + f_{i,t}(z \mid \boldsymbol{y}_i^{(t-1)}) \right\} dv(z)}{T - T_0} \\
&\quad + c_1^{1/2} \max_{t_0 \leq t \leq T, J_0 \leq j \leq J} \left| \widehat{\theta}_{i,t,j} - \theta_{i,t}(\boldsymbol{y}_i^{(t-1)}) \right| \frac{\sum_{t=t_0}^{T} \sum_{j=J_0}^{J} a_{i,t,j}}{\phi(T - T_0)} \\
&\leq \frac{2J_0 e^{-c_0(t_0 - T_0 - 1)} \left\{ 1 - e^{-c_0(T - t_0 + 1)} \right\}}{(T - T_0)(J - J_0 + 1)(1 - e^{-c_0})} + \frac{c_1^{1/2} \alpha_n}{\phi},
\end{aligned}
$$

almost surely. The proof is concluded by combining this with (A.23). □

Proof of Corollary 4. In the presence of pre-screening, for $t_0 \leq t \leq T$, $1 \leq i \leq n$, and $1 \leq j \leq J_0 - 1$, under Condition 3, with uniform prior assumption, it is readily seen that

$$
\begin{aligned}
\widetilde{a}_{i,t,j} &= 1(j \in \widehat{\Gamma}) \frac{\exp\left(-l_{i,j}^{(t-1)}\right)}{\sum_{j'=1}^{J} 1(j' \in \widehat{\Gamma}) \exp\left(-l_{i,j'}^{(t-1)}\right)} \\
&= \frac{1}{\sum_{j'=1}^{J} 1(j' \in \widehat{\Gamma}) \exp\left(l_{i,j}^{(t-1)} - l_{i,j'}^{(t-1)}\right)} \\
&\leq \frac{1}{\sum_{j'=J_0}^{J} 1(j' \in \widehat{\Gamma}) \exp\left\{(t - T_0 - 1) \frac{l_{i,j}^{(t-1)} - l_{i,j'}^{(t-1)}}{t - T_0 - 1}\right\}} \\
&\leq \frac{1}{\exp\left\{c_0(t - T_0 - 1) \sum_{j'=J_0}^{J} 1(j' \in \widehat{\Gamma})\right\}} \\
&= \frac{1}{\exp\left\{c_0(t - T_0 - 1) \widehat{J}_2\right\}}
\end{aligned}
$$

$$\leq \quad e^{-c_0(t-T_0-1)},$$

almost surely. With this result, the proof is straightforward under the same framework that proves Theorem 4, and this concludes the proof. $\square$

Appendix B. Implementation

In Section 2, the parameter estimation is conducted sequentially based on the observations available up to time point $t-1$ ($t \geq T_0+1$). Moreover, exponential family panel data models are typically more complicated than linear models or generalized linear models. Therefore, the parameter estimation process of our proposed method is more computationally demanding than its counterparts in linear regression models, and we need to employ a more efficient method to implement our proposed adaptive forecast combination procedures. Liang & Zeger (1986) proposed a generalized estimating equation to model non-normal and/or non-independent data in their seminal paper. The generalized estimating equation framework is computationally attractive and can be conveniently implemented. Now we forecast $y_{i,t}$ given $\boldsymbol{y}_i^{(t-1)}$, within the framework of generalized estimating equations.

Denote the vector of available covariates as $\boldsymbol{x}_{i,t}$ ($i=1,\ldots,n, t=1,\ldots,T$), which contains observable characteristics of subjects. We consider J candidate generalized estimating equations corresponding to J candidate forecasting procedures to fit the data. In the j-th candidate generalized estimating equation, we assume that

$$\mathbb{E}(y_{i,t}) = m(\boldsymbol{x}_{i,t,(j)}^{\mathrm{T}}\boldsymbol{\beta}_{(j)}), \quad \mathrm{var}(y_{i,t}) = \phi v\{m(\boldsymbol{x}_{i,t,(j)}^{\mathrm{T}}\boldsymbol{\beta}_{(j)})\},$$

where the $p_j \times 1$-dimensional vector $\boldsymbol{x}_{i,t,(j)}$ composes some elements in $\boldsymbol{x}_{i,t}$, $\boldsymbol{\beta}_{(j)}$ is the corresponding conformable regression parameter vector, with $m(x)$ and $v(x)$ being known mean and variance functions, respectively. For the j-th candidate generalized estimating equation, denote

$$m_{i,t-1}(\boldsymbol{\beta}_{(j)}) = \{m(\boldsymbol{x}_{i,1,(j)}^{\mathrm{T}}\boldsymbol{\beta}_{(j)}), \ldots, m(\boldsymbol{x}_{i,t-1,(j)}^{\mathrm{T}}\boldsymbol{\beta}_{(j)})\}^{\mathrm{T}},$$

$\boldsymbol{A}_{i,t-1}(\boldsymbol{\beta}_{(j)}) = \mathrm{diag}\left[\phi v\{m(\boldsymbol{x}_{i,l,(j)}^{\mathrm{T}}\boldsymbol{\beta}_{(j)})\}\right]$, and $\boldsymbol{R}_{t-1,j}(\boldsymbol{\alpha}_{(j)})$ as the $(t-1) \times (t-1)$ working correlation matrix, where $\boldsymbol{\alpha}_{(j)}$ is a q_j-dimensional correlation parameter. At each time point $t-1$ ($t \geq T_0+1$), the generalized estimating equation estimator for $\boldsymbol{\beta}_{(j)}$ can be obtained by solving

$$\sum\nolimits_{i=1}^{n} \frac{\partial m_{i,t-1}^{\mathrm{T}}(\boldsymbol{\beta}_{(j)})}{\partial \boldsymbol{\beta}_{(j)}} \boldsymbol{V}_{i,t-1,j}^{-1}(\boldsymbol{\beta}_{(j)}, \boldsymbol{\alpha}_{(j)}) \left\{\boldsymbol{y}_i^{(t-1)} - m_{i,t-1}(\boldsymbol{\beta}_{(j)})\right\} = \mathbf{0}_{p_j \times 1},$$

where $\boldsymbol{V}_{i,t-1,j}(\boldsymbol{\beta}_{(j)}, \boldsymbol{\alpha}_{(j)}) = \boldsymbol{A}_{i,t-1}^{1/2}(\boldsymbol{\beta}_{(j)})\boldsymbol{R}_{t-1,j}(\boldsymbol{\alpha}_{(j)})\boldsymbol{A}_{i,t-1}^{1/2}(\boldsymbol{\beta}_{(j)})$ is the working covariance matrix. The generalized estimating equation estimator of correlation parameter $\boldsymbol{\alpha}_{(j)}$ is obtained via the method of moments (Liang & Zeger, 1986). It is worthwhile noting that all the candidate generalized estimating equations can be subject to misspecification. That is, the functional form of $m(\boldsymbol{x}_{i,t,(j)}^{\mathrm{T}}\boldsymbol{\beta}_{(j)})$ (or $\boldsymbol{R}_{t-1,j}(\boldsymbol{\alpha}_{(j)})$) can be different from that of the true but unknown $\mathbb{E}(y_{i,t})$ (or $\mathrm{corr}(\boldsymbol{y}_i^{(t-1)})$). More discussions about model misspecification can be found in White (1982), Flynn et al. (2013), Lv & Liu (2014), and Yu et al. (2018).

Under the generalized estimating equation framework, the marginal mean $p_{i,t}$ is forecasted by $\widehat{p}_{i,t,j} = m(\boldsymbol{x}_{i,t,(j)}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{t-1,(j)})$,

$$\widehat{\boldsymbol{p}}_{i,j}^{(t-1)} = \{m(\boldsymbol{x}_{i,1,(j)}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{t-1,(j)}), \ldots, m(\boldsymbol{x}_{i,t-1,(j)}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}_{t-1,(j)})\}^{\mathrm{T}},$$

$\widehat{\boldsymbol{\kappa}}_{i,t,j} = \widehat{\boldsymbol{G}}_{i,t,j}^{-1}\widehat{\boldsymbol{S}}_{i,t,j}$, where $\widehat{\boldsymbol{G}}_{i,t,j}$ and $\widehat{\boldsymbol{S}}_{i,t,j}$ compose the $(t-1)\times(t-1)$ northwest corner and the first $t-1$ elements in the t-th column of $\boldsymbol{V}_{i,t,j}(\widehat{\boldsymbol{\beta}}_{t-1,(j)}, \widehat{\boldsymbol{\alpha}}_{t-1,(j)})$, respectively. Based on these estimators, our adaptive forecasts proposed in Section 2 can be constructed accordingly. The screening strategies based on information criteria (Pan, 2001; Imori, 2015) and the corresponding weighting scheme based on information score are provided in Appendix C. It is also worthwhile noting that the risk-bound properties established in the main paper hold for any reasonable forecasting methods, and therefore the AFTER and AFTQR procedures are flexible for various forecasting methods.

Appendix C. Alternative strategies for combining forecasts under a generalized estimating equations framework

We can also consider alternative multi-model forecast strategies under a generalized estimating equations framework. Model selection has been studied extensively in the context of generalized estimating equations (Imori, 2015). Among many others, Pan (2001) proposed the quasi-likelihood under the independence model criterion (QIC), which is considered as the representative criterion in the generalized estimating equations framework (Imori, 2015). In addition, Imori (2015) proposed a modified version of QIC. Denote the QIC by Pan (2001) as $\mathrm{QIC}_{\mathrm{Pan}}$ and the modified QIC of Imori (2015) as $\mathrm{QIC}_{\mathrm{Imori}}$.

Following Buckland et al. (1997), we can define a model averaging forecast based on the scores of $\mathrm{QIC}_{\mathrm{Pan}}$ or $\mathrm{QIC}_{\mathrm{Imori}}$. Under the framework of Pan (2001), the weight corresponding to the j-th generalized estimating equation is defined as

$$\widehat{w}_{t,j}^{\mathrm{Pan}} = \frac{\exp\left(-\frac{\mathrm{QIC}_{\mathrm{Pan},t-1,j} - \min_{1\le j\le J}\mathrm{QIC}_{\mathrm{Pan},t-1,j}}{2}\right)}{\sum_{j=1}^{J}\exp\left(-\frac{\mathrm{QIC}_{\mathrm{Pan},t-1,j} - \min_{1\le j\le J}\mathrm{QIC}_{\mathrm{Pan},t-1,j}}{2}\right)},$$

and the model averaging forecast of $f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})$ is

$$\widehat{g}_{\mathrm{Pan},i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) = \sum_{j=1}^{J}\widehat{w}_{t,j}^{\mathrm{Pan}}\widehat{g}_{i,t,j}(y \mid \boldsymbol{y}_i^{(t-1)}), \tag{C.1}$$

where $\text{QIC}_{\text{Pan},t-1,j}$ is the QIC_{Pan} of j-th generalized estimating equation evaluated based on $\mathcal{G}_{n,t-1}$. Similarly, we can define

$$\widehat{w}_{t,j}^{\text{Imori}} = \frac{\exp\left(-\frac{\text{QIC}_{\text{Imori},t-1,j} - \min_{1\le j\le J}\text{QIC}_{\text{Imori},t-1,j}}{2}\right)}{\sum_{j=1}^{J}\exp\left(-\frac{\text{QIC}_{\text{Imori},t-1,j} - \min_{1\le j\le J}\text{QIC}_{\text{Imori},t-1,j}}{2}\right)},$$

and the model averaging forecast of $f_{i,t}(y \mid \boldsymbol{y}_i^{(t-1)})$ is

$$\widehat{g}_{\text{Imori},i,t}(y \mid \boldsymbol{y}_i^{(t-1)}) = \sum_{j=1}^{J} \widehat{w}_{t,j}^{\text{Imori}}\widehat{g}_{i,t,j}(y \mid \boldsymbol{y}_i^{(t-1)}), \tag{C.2}$$

where $\text{QIC}_{\text{Imori},t-1,j}$ is the $\text{QIC}_{\text{Imori}}$ of the j-th generalized estimating equation, obtained based on $\mathcal{G}_{n,t-1}$. Model averaging forecasts proposed in (C.1) and (C.2) can be viewed as the smoothed extensions of the model selection methods based on QIC_{Pan} and $\text{QIC}_{\text{Imori}}$, respectively. Unlike the weight in (7) or (13), which is adaptive for each subject, the calculation of $\widehat{w}_{t,j}^{\text{Pan}}$ or $\widehat{w}_{t,j}^{\text{Imori}}$ involves the entire dataset in $\mathcal{G}_{n,t-1}$. Thus, $\widehat{w}_{t,j}^{\text{Pan}}$ or $\widehat{w}_{t,j}^{\text{Imori}}$ remains unchanged for each subject. Moreover, we can employ QIC_{Pan} and $\text{QIC}_{\text{Imori}}$ to screen out the top K procedures among all candidate forecasting procedures based on the initial information set $\mathcal{G}_{n,T_0}$ and combine them by the weighting scheme defined in Equation (10) or (15).