Appendix D (Correlated Measurement Errors), Appendix E (Non-Myopic Behavior), Appendix F (Self-Updating Inference Models), Appendix G (Proofs of Results from Main Paper), Appendix H (Proofs of Results from Appendices), and Appendix I (Numerical Studies Detailed Description and Discussion) are not part of the basic e-companion for reasons of space. They are available in the unabridged version of the paper posted on SSRN.

Appendix A: General Channel Capacity

In our base model we assumed that each sensor can only target one sensor ($\lfloor k \rfloor = 1$) in each period although its target may vary across periods. We focused on this case because $\lfloor k \rfloor = 1$ represents the setting in which choosing the right target is the most crucial. We now extend our results by allowing a general value for $\lfloor k \rfloor$. As we will show, there is a *rank-ordering structure*: each sensor in each period selects the $\lfloor k \rfloor$ other sensors that provide it with the lowest variance of signal (about the state) in that period and targets them. When the channel capacity weakly exceeds the number of all other sensors in the environment ($\lfloor k \rfloor \geq |\mathcal{N}| - 1$), the results become trivial because each sensor will target all other sensors. Thus, without loss of generality, in what follows we assume ($\lfloor k \rfloor < |\mathcal{N}| - 1$). The rank-ordering structure implies (proven below) that each sensor's set of long run contacts have exactly $\lfloor k \rfloor$ members assuming that sensor qualities are asymmetric (see Definition 3 in the main body). We focus on extending some of our main results (Propositions 1, Propositions 2, Lemma 2, and Proposition 4) with the understanding that the extension of other results become straightforward given these new extensions.

We start by providing a generalization for Proposition 1 presented in the main body.

PROPOSITION EC.1 **(Target Selection and Variance Reduction)**. *Suppose in each period sensor i is allowed to communicate to a general number of other sensors denoted by $\lfloor k \rfloor$. Then $a^*_{ijt} = \mathbb{1}_{\{j \in \mathcal{J}^*_{it}\}}$, where $\mathbb{1}_{\{\cdot\}}$ is the indicator function and $\mathcal{J}^*_{it} \subseteq \mathcal{N} \setminus \{i\}$ is the set of all sensors j that have the $\lfloor k \rfloor$ lowest value of*

$$\sigma^2_t(i,j,s_{t-1}) = \frac{q_j^2 + 1/\psi_{ijt}(s_{t-1})}{q_j^4}. \tag{EC.1}$$

Proof of Proposition EC.1: Proofs of all results in the Appendices are in Appendix H. ☐

We next present the following result, which shows that Proposition 2 still holds when the number of targets in each period is allowed to be any general number.

PROPOSITION EC.2 **(Familiarity Dynamics)**. *Suppose in each period sensor i is allowed to communicate to a general number of other sensors denoted by $\lfloor k \rfloor$. Then for any $s \in \mathbb{R}$:*
(i) $\psi_{ij,t+1}(s) = \psi_{ijt}(s) + \delta(q_i, q_j, a_{ijt})$, where $\delta(q_i, q_j, a_{ijt}) \triangleq f(q_i, q_j) a_{ijt}$ and

$$f(q_i, q_j) \triangleq \frac{(1+q_i^2)}{q_j^2(1+q_i^2+q_j^2)}. \tag{EC.2}$$

(ii) For all $t = 1, 2, 3, \cdots$, we have

$$\psi_{ij,t+1}(s) = \frac{v^2_{ij0} w^2_{ij0}}{w^2_{ij0} + v^2_{ij0} s^2} + f(q_i, q_j) \sum_{l=1}^{t} a_{ijl}. \tag{EC.3}$$

Next, we extend Lemma 2. The following result shows that if there are $\lfloor k \rfloor$ other sensors that (a) have a higher quality than a given sensor j', and (b) the familiarity of sensor i with them is above a threshold in a period t, then sensor i will not target sensor j in that period. It then follows from the results established in the paper that all such $\lfloor k \rfloor$ sensors dominate sensor j' from the perspective of sensor i in period t. Due to the dominance preservation results established in the paper, this means that sensor i will not target sensor j' in any future period either. This, in turn, allows us to establish that the long-run set of contacts of sensor i will have exactly $\lfloor k \rfloor$ members. Notably, this means that each sensor will eventually settle on $\lfloor k \rfloor$ other sensors, and will not change the sensors it targets in future periods. This long-run set of contacts, however, depends on the sample path, and might differ across different realizations of state over time.

LEMMA EC.1. *Suppose in each period sensor i is allowed to communicate to a general number of other sensors denoted by $\lfloor k \rfloor$. Consider a subset of sensors $\mathcal{J} \subseteq \mathcal{N} \setminus \{i\}$ satisfying $|\mathcal{J}| = \lfloor k \rfloor$. For any $\epsilon > 0$, there exists a fixed threshold $\bar{\psi}_\epsilon \in \mathbb{R}$ such that if for all $j \in \mathcal{J}$ (a) $\psi_{ijt}(s_{t-1}) > \bar{\psi}_\epsilon$, and (b) $\frac{q_j}{q_{j'}} > 1 + \epsilon$, then $a^*_{ij't} = 0$.*

PROPOSITION EC.3 **(Set of Long Run Contacts)**. *Suppose in each period sensor i is allowed to communicate to a general number of other sensors denoted by $\lfloor k \rfloor$. If sensor qualities are asymmetric, then given any sample path $\mathcal{S}_\infty$, $|\mathcal{T}_i(\mathcal{S}_\infty)| = \lfloor k \rfloor$ for all $i \in \mathcal{N}$.*

Appendix B: Erroneous Initial Beliefs about Means

In our base model we assumed that at time $t = 0$ sensor i believes that sensor j's inference model parameters come from normal distributions that have correctly calibrated means ($N(\hat{\alpha}_j, 1/v^2_{ij0})$ and $N(\hat{\beta}_j, 1/w^2_{ij0})$). Since in practice, inference models are carefully built based on training data sets, it is not a strong assumption to assume that means of these distributions are relatively accurate. However, we now extend our analyses and consider the setting in which these means are erroneous as well.

In particular, suppose that at time $t = 0$, sensor i believes that sensor j's inference model parameters come from normal distributions $N(\tilde{\alpha}_j, 1/v^2_{ij0})$ and $N(\tilde{\beta}_j, 1/w^2_{ij0})$, where $(\tilde{\alpha}_j, \tilde{\beta}_j) \neq (\hat{\alpha}_j, \hat{\beta}_j)$. Let $\tilde{\boldsymbol{\mu}} = (\tilde{\alpha}_j, \tilde{\beta}_j)$, $\boldsymbol{\mu} = (\hat{\alpha}_j, \hat{\beta}_j)$, and denote the sensor i's belief about sensor j's inference model parameters at time t by $\mathbf{X} = (\alpha_{ijt}, \beta_{ijt})^T$ (a column vector). Suppose sensor i communicates to sensor j at time $t = 1$. It follows from the general results provided under Method 2 of proof of Proposition 2 that upon receiving the signal

$$\mathbf{Y} = \alpha_{ij1} + \beta_{ij1} s_0 + q_j^2 (S_1 + \epsilon_{j1})$$

from sensor j, sensor i's posterior joint distribution on sensor i's parameters is

$$\mathbf{X}|\mathbf{Y} \sim N(\boldsymbol{\Sigma}[\mathbf{A}^T \mathbf{L}(\mathbf{Y} - \mathbf{b}) + \boldsymbol{\Lambda}\tilde{\boldsymbol{\mu}}], \boldsymbol{\Sigma}), \tag{EC.4}$$

where $\mathbf{A} = (1, s_0)$, $\mathbf{b} = \mathbb{E}\left[q_j^2 (S_1 + \epsilon_{j1})\right]$, $\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^{-1}_{ij0}$, $\mathbf{L} = f(q_i, q_j)$, and

$$\boldsymbol{\Sigma} = \left[\boldsymbol{\Lambda} + \mathbf{A}^T \mathbf{L} \mathbf{A}\right]^{-1}. \tag{EC.5}$$

The difference between the posterior joint distribution obtained in (EC.4) and what one would get under the assumption made in the main body (i.e., correct means at times zero) is that in the latter case $\tilde{\boldsymbol{\mu}}$ would be replaced with $\boldsymbol{\mu}$. However, it follows from the proof of Proposition 2, that the results in Proposition 2 (and hence, other results in the paper) hold. Furthermore, under some conditions, one can also show that the initial difference between starting with mean beliefs $\tilde{\boldsymbol{\mu}}$ instead of $\boldsymbol{\mu}$ disappears as $t \to \infty$ assuming i targets j infinitely often).

Appendix C: Delayed, Imperfect or No State Observations

In our base model we assumed that each sensor has a slower-but-more-accurate estimation approach in place (e.g., technician inspection or remote offloading) that has a delay of one period and is perfectly accurate. We analyze various relaxations in which the slower-but-more-accurate approach has a general delay but is perfectly accurate (Case I) or has a general delay with imperfect observations (Case II). We also examine settings in which the only state realization is at time 0 (see Cases III and IV).

Case I: Delayed-but-Perfect State Observations

Recall that each sensor i has its own model of state dynamics given by (3):

$$S_{it} = \hat{\alpha}_i + \hat{\beta}_i S_{t-1} + \tilde{\epsilon}_t, \tag{EC.6}$$

where $\hat{\alpha}_i$ and $\hat{\beta}_i$ are sensor i's estimates of the process parameters α and β. In the main body, we assumed that right before period t starts, the realized value of the previous period state (s_{t-1}) becomes publicly known (i.e., is revealed to each sensor), and the system moves to the next period. Here, we relax this assumption and assume that there is a delay of n periods in observations made about the sate. That is, we now assume that right before period t starts, the value of S_{t-n} (for some positive integer $n \leq t$) is observed perfectly by all sensors. Note that this is a general extension and incorporates various scenarios (including the one analyzed in the main body) as a special case. For example, as we will discuss later in this appendix, setting $n = t$ means that no state observation can be made beyond that made at the very beginning of the horizon (i.e., s_0).

Observe that replacing $S_{i,t-1} = \hat{\alpha}_i + \hat{\beta}_i S_{t-2} + \tilde{\epsilon}_{t-1}$ in (3) and repeating this process yields:

$$S_{it} = \hat{\alpha}_i \sum_{k=0}^{n-1} \hat{\beta}_i^k + \hat{\beta}_i^n S_{t-n} + \sum_{k=0}^{n-1} \hat{\beta}_i^k \tilde{\epsilon}_{t-k}. \tag{EC.7}$$

Hence, we have

$$\begin{aligned} \mathbb{E}[S_{it} | S_{i,t-n} = s_{i,t-n}] &= \hat{\alpha}_i \sum_{k=0}^{n-1} \hat{\beta}_i^k + \hat{\beta}_i^n s_{t-n} \\ &= \hat{\alpha}_i \frac{1-\hat{\beta}_i^n}{1-\hat{\beta}_i} + \hat{\beta}_i^n s_{t-n} \end{aligned} \tag{EC.8}$$

and

$$\begin{aligned} Var[S_{it} | S_{i,t-n} = s_{i,t-n}] &= \sum_{k=0}^{n-1} \hat{\beta}_i^{2k} Var[\tilde{\epsilon}_{t-k}] \\ &= \frac{1-\hat{\beta}_i^{2n}}{1-\hat{\beta}_i^2}, \end{aligned} \tag{EC.9}$$

where geometric sums are obtained by assuming (without loss of generality) that $\beta_i \neq 1$.

Thus, from the perspective of sensor i, the current state S_t follows a normal distribution with mean and variance given by (EC.8) and (EC.9), respectively. Next, upon realizing the current signal $\Gamma_{it} = \gamma_{it}$, sensor i updates its prior belief about the current state according to Bayes' rule. Since both the signal received about the state and the prior on the state have a normal distribution, it follows that sensor i's posterior belief about the state is also normally distributed but with a mean and variance given by

$$\begin{aligned}\mathbb{E}[S_{it}|\Gamma_{it}=\gamma_{it}] &= \frac{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2n}}[\hat{\alpha}_i\frac{1-\hat{\beta}_i^n}{1-\hat{\beta}_i}+\hat{\beta}_i^n s_{t-n}]+q_i^2\gamma_{it}}{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2n}}+q_i^2}\\ &= \frac{g_n(\hat{\beta}_i)[\hat{\alpha}_i\frac{1-\hat{\beta}_i^n}{1-\hat{\beta}_i}+\hat{\beta}_i^n s_{t-n}]+q_i^2\gamma_{it}}{g_n(\hat{\beta}_i)+q_i^2} \qquad \text{(EC.10)}\end{aligned}$$

and

$$\begin{aligned}Var[S_{it}|\Gamma_{it}=\gamma_{it}] &= \frac{1}{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2n}}+q_i^2}\\ &= \frac{1}{g_n(\hat{\beta}_i)+q_i^2}, \qquad \text{(EC.11)}\end{aligned}$$

where

$$g_n(\hat{\beta}_i) \triangleq \frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2n}}.$$

Thus, it can be seen that the higher the quality of sensor i, the more weight it places on its signal when updating its mean belief, and the larger the associated variance reduction. In addition, assuming that the process is stationary ($|\beta_i|<1$), the smaller the n (i.e., the more recent the state observation) the smaller both the variance of the prior and the variance of posterior (see (EC.9) and (EC.11)). The special case with $n=1$ represents the case where such variances are the smallest and can be obtained by noting that $g_1(\hat{\beta}_i)=1$. Replacing $g_1(\hat{\beta}_i)=1$, yields the results obtained in the main body. However, using (EC.10) and (EC.11) with a general $g_n(\hat{\beta}_i) \triangleq \frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2n}}$ instead of their counterparts in the main body ((4) and (5), respectively) generalizes the results obtained.

Specifically, next we show that the main results obtained in the main body hold under this extension. Thus, our results reveal that our main results are *robust* to the assumption (made in the main body) that at the beginning of each period sensors have access to the realized value of the previous period's state.

Robustness of the Main Results. To study the target selection problem of sensor i, suppose it communicates with some sensor j in period t. Similar to the main body, given its privately generated signal Γ_{jt} in period t, sensor j provides sensor i with its best current estimate of state, which is $\mathbb{E}[S_{jt}|\Gamma_{jt}]$, i.e., its updated/latest expected belief about the current state S_t. Now, from sensor i's perspective, $\mathbb{E}[S_{jt}|\Gamma_{jt}]$ is formed according to:

$$\mathbb{E}[S_{ijt}|\Gamma_{jt}] = \frac{g_n(\hat{\beta}_{ijt})[\hat{\alpha}_{ijt}\frac{1-\hat{\beta}_{ijt}^n}{1-\hat{\beta}_{ijt}}+\hat{\beta}_{ijt}^n s_{t-n}]+q_j^2\Gamma_{jt}}{g_n(\hat{\beta}_{ijt})+q_j^2}. \qquad \text{(EC.12)}$$

Equation (EC.12) is based on (EC.10). However, it uses parameters $\hat{\alpha}_{ijt}$ and $\hat{\beta}_{ijt}$, which are sensor i's beliefs about sensor j's inference parameters $\hat{\alpha}_j$ and $\hat{\beta}_j$. (Also, note that (EC.12) reduces to (7) of the main body if we set $n=1$.) Since $\Gamma_{jt}=S_t+\epsilon_{jt}$, it can be seen from (EC.12) that $\mathbb{E}[S_{ijt}|\Gamma_{jt}]$ provides sensor i with the following noisy signal regarding the state S_t:

$$\frac{g_n(\hat{\beta}_{ijt})+q_j^2}{q_j^2}\mathbb{E}[S_{ijt}|\Gamma_{jt}] = S_t+\epsilon_{jt}+\frac{g_n(\hat{\beta}_{ijt})[\hat{\alpha}_{ijt}\frac{1-\hat{\beta}_{ijt}^n}{1-\hat{\beta}_{ijt}}+\hat{\beta}_{ijt}^n s_{t-n}]}{q_j^2} \qquad \text{(EC.13)}$$

There are two independent sources of noise in this signal: (a) the inherent white noise ϵ_{jt} in sensor j's measurement Γ_{jt} (which has a variance of $1/q_j^2$), and (b) the noise caused by sensor i's lack of familiarity with sensor j's inference model (the third term in the right hand side of (EC.13)). For notational convenience, we define the random variable

$$\tilde{\Xi}_{ijt}(s_{t-n}) = g_n(\hat{\beta}_{ijt})\left[\hat{\alpha}_{ijt}\frac{1-\hat{\beta}_{ijt}^n}{1-\hat{\beta}_{ijt}} + \hat{\beta}_{ijt}^n s_{t-n}\right],$$

and let its precision be denoted by $\tilde{\psi}_{ijt}(s_{t-n}) \triangleq 1/Var\left[\tilde{\Xi}_{ijt}(s_{t-n})\right]$. With these, based on (EC.13), the variance of the signal's noise (if i targets j in period t) is

$$\tilde{\sigma}_t^2(i,j,s_{t-n}) = \frac{q_j^2 + 1/\tilde{\psi}_{ijt}(s_{t-n})}{q_j^4},$$

which provides an extension for Equation (10) of the main body.

Similar to the main body, the target selection is then as follows. Sensor i targets the sensor that has the minimum $\tilde{\sigma}_t^2(i,j,s_{t-n})$. That is,

$$\begin{aligned} j_{it}^* &\triangleq \arg\min_{j\in\mathcal{N}\setminus\{i\}} \tilde{\sigma}_t^2(i,j,s_{t-n}) \\ &= \arg\min_{j\in\mathcal{N}\setminus\{i\}} \left\{\frac{q_j^2 + 1/\tilde{\psi}_{ijt}(s_{t-n})}{q_j^4}\right\}. \end{aligned} \tag{EC.14}$$

Now that the target selection rule is defined, we are ready to show that the main results presented in the main body hold. In particular, we show that when sensor qualities differ across sensors, for any given sensor i and along any fixed sample path $\mathcal{S}_\infty$: (a) the set of long run contacts $\mathcal{T}_i(\mathcal{S}_\infty)$ is a singleton, i.e., $|\mathcal{T}_i(\mathcal{S}_\infty)| = 1$, and (b) the unique long run contact in $\mathcal{T}_i(\mathcal{S}_\infty)$ can be identified in the almost sure sense in finite time, i.e., $\mathcal{T}_i(\mathcal{S}_\infty) = \mathcal{T}_i(\mathcal{S}_\infty'^{t^*})$ *a.s.* for some $t^* < \infty$. These two results, in turn, mean that one can fully define the long run communication network as a *random directed graph*, i.e., a directed graph with given probabilities assigned to each link ij that indicate the probability that j will be the long run target for i. Furthermore, there exists a finite time after which the graph can be defined as a deterministic directed graph, i.e., with all probabilities being zero or one, that fully specifies the long run target for each sensor.

To establish these results for the extension under consideration, we start by establishing the following result which extends Lemma 2 of the main body.

LEMMA EC.2. *For any $\epsilon > 0$, there exists a fixed threshold $\bar{\psi}_\epsilon \in \mathbb{R}$ such that if $\tilde{\psi}_{ijt}(s_{t-n}) > \bar{\psi}_\epsilon$ and $\frac{q_j}{q_{j'}} > 1+\epsilon$ then $j_{it}^* \neq j'$.*

Lemma (EC.2) allows us to show that the set of long run contacts of any sensor is a singleton, and that this result holds regardless of the duration of delay in state realization (i.e., parameter n). That is, Proposition 4 of the main body holds not only when $n = 1$, but also for any general n.

PROPOSITION EC.4 **(Unique Long Run Contact)**. *If sensor qualities are asymmetric, then given any sample path $\mathcal{S}_\infty$, $|\mathcal{T}_i(\mathcal{S}_\infty)| = 1$ for all $i \in \mathcal{N}$.*

Finally, we can show that another main result presented in the main body holds for the extension under consideration: when the sensor qualities are asymmetric, the long run set of contacts of each sensor can be determined in finite time. This means that transient analysis is sufficient for characterizing the communication network that will be formed in the long run, even when state realization has a general delay. Intuitively, this is because the role of state in target selection eventually vanishes: the effect of familiarity built via past communications eventually outweighs the role of state.

Proposition EC.5 **(Transient Analysis)**. *If sensor qualities are asymmetric, then along any sample path $\mathcal{S}_\infty$ there exists a finite period t^* such that for all $i \in \mathcal{N}$*

$$\mathcal{T}_i(\mathcal{S}_\infty) = \mathcal{T}_i(\mathcal{S}'^{t^*}_\infty) \qquad a.s.$$

Proposition EC.6 **(Deterministic Random Directed Graph)**. *If sensor qualities are asymmetric, then there exists a finite time t^* such that given the sample path up to t^* (i.e., $\mathcal{S}_{t^*}$), the long run communication network can be defined as a graph $\vec{G}(\mathcal{N}, \mathcal{E}, \mathcal{P})$ with the additional property that $p_{ij} \in \{0, 1\}$ for all $i, j \in \mathcal{N}$.*

Case II: Delayed-and-Imperfect State Observations

We next consider the case in which each sensor's slower-but-more-accurate estimation approach has a general delay of n periods but is imperfect. In other words, the delayed state observations are subject to errors. To extend Case I, we assume that right before period t starts, the value of state at $t-n$ (for some positive integer $n \le t$) is revealed with some random error that is independent across sensors. Specifically, we model this revelation for sensor i as $S'_{i,t-n} = S_{t-n} + \upsilon_i$ with a realized value denoted by $S'_{i,t-n} = s'_{i,t-n}$, where S_{t-n} is the true value of state at period $t-n$, and υ_i represents a normally distributed random noise with mean 0 and variance $1/q_{i0}^2$. In this setting, q_{i0}^2 denotes the precision (i.e., quality) of sensor i's slower-but-more-accurate approach in measuring the state. For expositional ease, we will assume this quality is the same across sensors, that is, $q_{i0} = q_0 \ \forall i$. Because the slower approach is more accurate than the regular sensors, it follows that $q_0 >> q_i \ \forall i$. Of note, if $q_0 = \infty$ this setting is equivalent to Case I studied earlier. However, we now generalize our previous analyses by allowing $q_0 < \infty$. We assume the υ_i noise terms are independent across sensors. However, the case in which there is one "universal" slower-but-more accurate approach that gives the same (imperfect) measurement to all sensors is easily handled by simply setting $\upsilon_i = \upsilon \ \forall i$ in what follows.

Recall from (EC.7) that

$$S_{it} = \hat{\alpha}_i \sum_{k=0}^{n-1} \hat{\beta}_i^k + \hat{\beta}_i^n S_{t-n} + \sum_{k=0}^{n-1} \hat{\beta}_i^k \tilde{\epsilon}_{t-k}. \tag{EC.15}$$

Hence, we have

$$\begin{aligned}\mathbb{E}[S_{it} | S'_{i,t-n} = s'_{i,t-n}, S'_{i,t-n} = S_{t-n} + \upsilon_i] &= \hat{\alpha}_i \sum_{k=0}^{n-1} \hat{\beta}_i^k + \hat{\beta}_i^n s'_{i,t-n} \\ &= \hat{\alpha}_i \frac{1 - \hat{\beta}_i^n}{1 - \hat{\beta}_i} + \hat{\beta}_i^n s'_{i,t-n} \end{aligned} \tag{EC.16}$$

and

$$\begin{aligned} Var[S_{it} | S'_{i,t-n} = s'_{i,t-n}, S'_{i,t-n} = S_{t-n} + \upsilon_i] &= \frac{\hat{\beta}_i^{2n}}{q_0^2} + \sum_{k=0}^{n-1} \hat{\beta}_i^{2k} Var[\tilde{\epsilon}_{t-k}] \\ &= \frac{\hat{\beta}_i^{2n}}{q_0^2} + \frac{1 - \hat{\beta}_i^{2n}}{1 - \hat{\beta}_i^2}, \end{aligned} \tag{EC.17}$$

where geometric sums are obtained by assuming (without loss of generality) that $\beta_i \neq 1$.

Thus, similar to Case I, from the perspective of sensor i, the current state S_t follows a normal distribution with mean and variance given by (EC.16) and (EC.17), respectively. Next, upon realizing the current signal $\Gamma_{it} = \gamma_{it}$, sensor i updates its prior belief about the current state according to

Bayes' rule. Since both the signal received about the state and the prior on the state have a normal distribution, it follows that sensor i's posterior belief about the state is also normally distributed but with a mean and variance given by

$$\begin{aligned}
\mathbb{E}[S_{it}|\Gamma_{it}=\gamma_{it}] &= \frac{\big[\frac{\hat{\beta}_i^{2n}}{q_0^2}+\frac{1-\hat{\beta}_i^{2n}}{1-\hat{\beta}_i^2}\big]^{-1}\big[\hat{\alpha}_i\frac{1-\hat{\beta}_i^n}{1-\hat{\beta}_i}+\hat{\beta}_i^n s'_{i,t-n}\big]+q_i^2\gamma_{it}}{\big[\frac{\hat{\beta}_i^{2n}}{q_0^2}+\frac{1-\hat{\beta}_i^{2n}}{1-\hat{\beta}_i^2}\big]^{-1}+q_i^2}\\
&= \frac{g_n(\hat{\beta}_i)\big[\hat{\alpha}_i\frac{1-\hat{\beta}_i^n}{1-\hat{\beta}_i}+\hat{\beta}_i^n s'_{i,t-n}\big]+q_i^2\gamma_{it}}{g_n(\hat{\beta}_i)+q_i^2} \qquad \text{(EC.18)}
\end{aligned}$$

and

$$\begin{aligned}
Var[S_{it}|\Gamma_{it}=\gamma_{it}] &= \frac{1}{\big[\frac{\hat{\beta}_i^{2n}}{q_0^2}+\frac{1-\hat{\beta}_i^{2n}}{1-\hat{\beta}_i^2}\big]^{-1}+q_i^2}\\
&= \frac{1}{g_n(\hat{\beta}_i)+q_i^2}, \qquad \text{(EC.19)}
\end{aligned}$$

where

$$g_n(\hat{\beta}_i) \triangleq \Big[\frac{\hat{\beta}_i^{2n}}{q_0^2}+\frac{1-\hat{\beta}_i^{2n}}{1-\hat{\beta}_i^2}\Big]^{-1}.$$

When $q_0=\infty$ we observe that $g_n(\hat{\beta}_i)$ given above is equivalent to that obtained under Case I. However, it provides a generalization by allowing $q_0<\infty$. Importantly, using the $g_n(\hat{\beta}_i)$ given above instead of that in Case I, it can easily been shown that all the results provided under Case I still hold. This follows directly if the slower-but-more accurate approach is common to all sensors (i.e., they all receive the same delayed-but-imperfect measurement). If the delayed-but-imperfect measurements are independent then the results also go through directly subject to a mild assumption that sensors do not try to learn about the delayed measurements received by other sensors. Any benefits of such learning (to improve estimation) would be very low (and so can be ignored) because the slower-but-more-accurate estimation approach has a much higher precision than the real-time sensor measurement, i.e., $q_0>>q_i$.

Case III: No State Observations, Transient Behavior

In the previous subsection, we considered the case where state is realized with some delay (Case I) and showed that the main results presented in the main body hold under such an extension. We now consider the case in which there is no state observation beyond the one in the very first period, s_0 (see below for another extension in which even s_0 is not known). This case can be analyzed by setting $n=t$ and using the result of Case I. Specifically, for $n=t$, we observe that

$$\begin{aligned}
\mathbb{E}[S_{it}|\Gamma_{it}=\gamma_{it}] &= \frac{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}\big[\hat{\alpha}_i\frac{1-\hat{\beta}_i^t}{1-\hat{\beta}_i}+\hat{\beta}_i^t s_0\big]+q_i^2\gamma_{it}}{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}+q_i^2}\\
&= \frac{g_t(\hat{\beta}_i)\big[\hat{\alpha}_i\frac{1-\hat{\beta}_i^t}{1-\hat{\beta}_i}+\hat{\beta}_i^t s_0\big]+q_i^2\gamma_{it}}{g_t(\hat{\beta}_i)+q_i^2} \qquad \text{(EC.20)}
\end{aligned}$$

and

$$Var[S_{it}|\Gamma_{it}=\gamma_{it}] = \frac{1}{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}+q_i^2}$$

$$= \frac{1}{g_t(\hat{\beta}_i) + q_i^2}, \tag{EC.21}$$

where

$$g_t(\hat{\beta}_i) \triangleq \frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}.$$

Similar to Case I, using (EC.20) and (EC.21) with a general $g_t(\hat{\beta}_i) \triangleq \frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}$ instead of their counterparts in the main body ((4) and (5), respectively) generalizes the results obtained (after some necessary modifications).

What if even s_0 is not available? Note that the availability of s_0 in the above analyses is not restrictive. For example, suppose s_0 is not available to the sensors but they either make use of the historical mean of the state value (e.g., using their training observations) to find/estimate a proxy for s_0, or rely on a predicated value provided by an outside entity (e.g., a third-party's smart device). In either case, the s_0 value in the above analysis can be replaces with $\tilde{s}_0$, where $\tilde{s}_0$ represents the proxy available to the sensors. Replacing s_0 with its proxy $\tilde{s}_0$ (which might differ from s_0), we have

$$\begin{aligned}
\mathbb{E}[S_{it}|\Gamma_{it}=\gamma_{it}] &= \frac{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}[\hat{\alpha}_i\frac{1-\hat{\beta}_i^t}{1-\hat{\beta}_i} + \hat{\beta}_i^t\tilde{s}_0] + q_i^2\gamma_{it}}{\frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}} + q_i^2} \\
&= \frac{g_t(\hat{\beta}_i)[\hat{\alpha}_i\frac{1-\hat{\beta}_i^t}{1-\hat{\beta}_i} + \hat{\beta}_i^t\tilde{s}_0] + q_i^2\gamma_{it}}{g_t(\hat{\beta}_i) + q_i^2} \qquad \text{(EC.22)}
\end{aligned}$$

Using using (EC.22) and (EC.21) with a general $g_t(\hat{\beta}_i) \triangleq \frac{1-\hat{\beta}_i^2}{1-\hat{\beta}_i^{2t}}$ instead of their counterparts in the main body ((4) and (5), respectively), we again see that the main results are generalizable. Furthermore, in the next case (Case III), we analyze the setting in which the system dynamics has reached steady-state. We note that in that case, even a proxy for s_0 is not needed.

Case IV: No State Observations, Steady-State Behavior

We now consider a similar case to that of Case II by assuming that there is no state observation. However, unlike Case II, we assume that at the beginning of the horizon the state dynamics has reached steady-state. This case can be analyzed by setting $n \to \infty$, and nothing that the process is stationary if $|\beta_i| < 1$. To this end, we define

$$g_\infty(\hat{\beta}_i) \triangleq 1 - \hat{\beta}_i^2.$$

It can then be seen that

$$\mathbb{E}[S_{it}|\Gamma_{it}=\gamma_{it}] = \frac{g_\infty(\hat{\beta}_i)[\hat{\alpha}_i\frac{1}{1-\hat{\beta}_i}] + q_i^2\gamma_{it}}{g_\infty(\hat{\beta}_i) + q_i^2} = \frac{(1+\hat{\beta}_i)\hat{\alpha}_i + q_i^2\gamma_{it}}{g_\infty(\hat{\beta}_i) + q_i^2} \tag{EC.23}$$

and

$$Var[S_{it}|\Gamma_{it}=\gamma_{it}] = \frac{1}{g_\infty(\hat{\beta}_i) + q_i^2}. \tag{EC.24}$$

Again, similar to cases I and II, using (EC.23) and (EC.24) with $g_\infty(\hat{\beta}_i) \triangleq 1 - \hat{\beta}_i^2$ instead of their counterparts in the main body ((4) and (5), respectively) generalizes the results obtained (after some necessary modifications).