Electronic Companion

Appendix EC.1: More Details about the Extensions in §6

EC.1.1. Food-Quality Degradation

Let $t \geq 0$ be the time elapsed after an order is complete. If a customer arrives at the service facility and picks up her order at time t, her reward from the order is $Ve^{-d \cdot t}$, where $d > 0$ is a parameter that measures the magnitude of quality degradation. The pure-order-onsite scheme is not affected by d since, by definition, orders will be complete only after customers arrive. Next, we re-derive the equilibria for various order-ahead schemes. To begin with, local customers still follow the Naor threshold. For remote customers, the expected utility of ordering given initial queue length n is $U_d(n) = V\mathbb{P}(N_n > 0) + \mathbb{E}[Ve^{-d(T-X)}|T > X]\mathbb{P}(N_n = 0) - c\sum_{i=0}^{(n+1)} \frac{i}{\mu} \cdot p_n(i) - \frac{c}{\beta}$, where T is a random variable denoting travel time and X is a random variable denoting the steady-state sojourn time in the order queue. Note that in computing the expected service reward, quality degradation only arises in the event that the order is ready for pickup upon arrival at the service facility, i.e., $\{N_n = 0\}$, which is equivalent to saying that travel time exceeds sojourn time, i.e., $\{T > X\}$. Since T follows an exponential distribution with rate β, due to its memoryless property, $\mathbb{E}[Ve^{-d(T-X)}|T > X] = \mathbb{E}[Ve^{-dT}] = \frac{\beta}{\beta+d}V$. Therefore, $U_d(n) = \left(1 - \sigma^{n+1}\frac{d}{\beta+d}\right)V - \frac{c}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu}\right)$.

EC.1.1.1. Hybrid Order-Ahead

(1) Queue-length information not shared remotely. Suppose all remote customers place an order with probability $q \in [0,1]$ and local customers follow the Naor-threshold. The steady-state probability of the number of outstanding orders being i, $\pi_i^u(q)$, is the same as (2). The unconditional expected utility for a remote customer who places an order is $U_{A,d}(q) = \sum_{n=0}^{\infty} U_d(n)\pi_n^u(q)$, which is decreasing in q. Thus, $q \in (0,1)$ is an equilibrium only if $U_{A,d}(q) = 0$, $q = 1$ is an equilibrium if $U_{A,d}(1) > 0$ and $q = 0$ is an equilibrium if $U_{A,d}(0) < 0$. Hence, remote customers' equilibrium order-placing probability is: $q_{A,d} = 0, \text{if } \Lambda \geq \bar{\lambda}_{A,d}; q_{A,d} = \hat{q}_{A,d} \in (0,1), \text{if } \underline{\lambda}_{A,d} < \Lambda < \bar{\lambda}_{A,d}; q_{A,d} = 1, \text{if } \Lambda \leq \underline{\lambda}_{A,d}$, where $\hat{q}_{A,d}$ uniquely solves the equation $U_{A,d}(\hat{q}_{A,d}) = 0$. The resulting system throughput $TH_{A,d}^u = \mu(1 - \pi_0^u(\hat{q}_{A,d}))$. **(2) Queue-length information shared remotely.** Remote customers follow a threshold joining strategy. The joining threshold of remote customers $\hat{n}_e \leq n_e^* \leq n_e$ is uniquely solved by $\hat{n}_e^* \equiv \min\{n \in \mathbb{N} : U_d(n) < 0\}$. Accordingly, the steady-state probability of the number of outstanding orders being i is $\pi_i^o = \rho^i \pi_0^o,\ i = 0, 1, \cdots, \hat{n}_e; \pi_i^o = ((1-\gamma)\rho)^{i-\hat{n}_e}\rho^{\hat{n}_e}\pi_0^o,\ i = \hat{n}_e + 1, \cdots, n_e$, where $\hat{\pi}_0^o = \left(\frac{1-\rho^{\hat{n}_e}}{1-\rho} + \frac{\rho^{\hat{n}_e}(1-((1-\gamma)\rho)^{n_e - \hat{n}_e + 1})}{1-(1-\gamma)\rho}\right)^{-1}$. with $\rho \equiv \Lambda/\mu$. The resulting system throughput is $TH_{A,d}^o = \mu(1 - \hat{\pi}_0^o)$.

EC.1.1.2. Hybrid Order-Ahead with (Optimal) Rejection

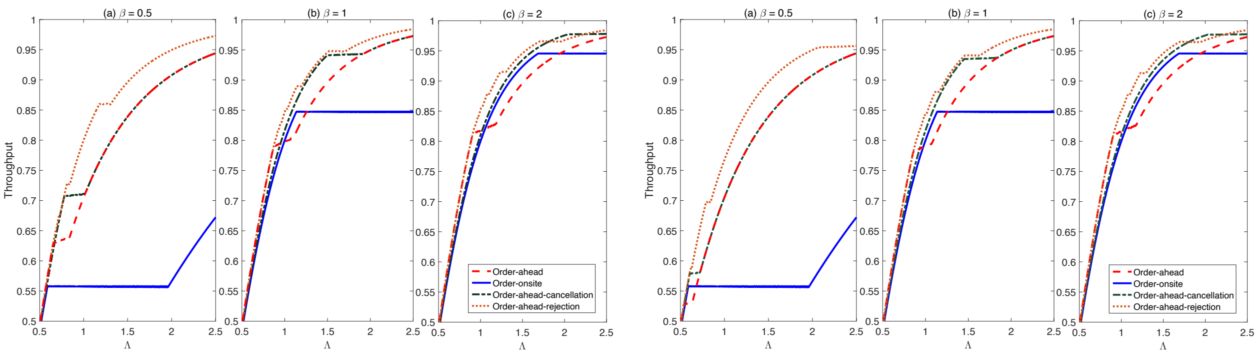
(1) Queue-length information not shared remotely. Suppose that the service provider accepts new remote orders if the number of outstanding orders is strictly less than threshold N and rejects any new remote orders otherwise. The unconditional expected utility from ordering is $U_{R,d}^N(q) = \sum_{n=0}^{N-1} U_d(n)\pi_{n,R}(q)$, The steady-state probability of the number of outstanding orders being i is given in (6) and (7). The remote customers' equilibrium order-placing probability $q_{R,d}^N$ is as follows: if $U_d(N-1) \geq 0$, $q_{R,d}^N = 1$; otherwise, $q_{R,d}^N = 0, \text{if } \Lambda \geq \bar{\lambda}_{R,d}^N; q_{R,d}^N = \hat{q}_{R,d}^N \in (0,1), \text{if } \underline{\lambda}_{R,d}^N < \Lambda < \bar{\lambda}_{R,d}^N; q_{R,d} = 1, \text{if } \Lambda \leq \underline{\lambda}_{R,d}^N$, where $\hat{q}_{R,d}^N$ uniquely solves the equation $U_{R,d}^N(\hat{q}) = 0$. The resulting system throughput is $TH_{R,d}^N = \mu[1 - \pi_{0,R}^u(q_{R,d}^N)]$.

(2) Queue-length information shared remotely. Suppose that the rejection threshold is N. Thus, a remote order effectively joins the queue according to a threshold $\min\{N, \hat{n}_e\}$. The steady-state probability of the number of outstanding orders being i is $\widetilde{\pi}_i^o = \rho^i \widetilde{\pi}_0^o,\ i = 0, 1, \cdots, x; \widetilde{\pi}_i^o = ((1-\gamma)\rho)^{i-x}\rho^x \widetilde{\pi}_0^o,\ i = x+1, \cdots, n_e$, where $x = \min\{N, \hat{n}_e\}$ and $\widetilde{\pi}_0^o = \left(\frac{1-\rho^x}{1-\rho} + \frac{\rho^x(1-((1-\gamma)\rho)^{n_e - x + 1})}{1-(1-\gamma)\rho}\right)^{-1}$. The resulting system throughput is $TH_{R,d}^o = \mu(1 - \widetilde{\pi}_0^o)$.

EC.1.1.3. Hybrid Order-Ahead with Cancellation **(1) Queue-length information not shared remotely.** If a remote customer places an order when the queue length is n, then her expected utility is given by $U_{C,d}(n) = V\mathbb{P}(0 < N_n^C \le n_e) + \mathbb{E}[Ve^{-d(T-X)}|T > X]\mathbb{P}(N_n^C = 0) - cw_d^C(n) - \frac{c}{\beta}$, where $\mathbb{P}(0 < N_n^C \le n_e) = \begin{cases} 1 - p_n^C(0) = 1 - \prod_{k=0}^{n} \frac{\mu_k}{\mu_k+\beta}, & \text{if} \quad n < n_e, \\ \sum_{i=1}^{n_e} p_n^C(i) = \sum_{i=1}^{n_e} \frac{\beta}{\mu_{i-1}+\beta} \prod_{k=i}^{n} \frac{\mu_k}{\mu_k+\beta}, & \text{otherwise,} \end{cases}$ where $\mu_n = \mu + (n - n_e)^+\beta$. The expected onsite delay is $w_{C,d}(n) \equiv \mathbb{E}[W(n)] = \sum_{i=0}^{(n+1)\wedge n_e} \mathbb{E}[W(n)|N_n = i] \cdot p_n^C(i) = \sum_{i=0}^{(n+1)\wedge n_e} \frac{i}{\mu} \cdot p_n^C(i) = \begin{cases} \frac{1}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu} - 1\right), & \text{if} \quad n < n_e; \\ \frac{\beta}{\mu+\beta} \sum_{i=1}^{n_e} \frac{i}{\mu} \prod_{k=i}^{n} \frac{\mu_k}{\mu_k+\beta}, & \text{otherwise.} \end{cases}$ Therefore, the expected utility of a remote customer who places an order is $U_{C,d}(n) = \bar{U}_{C,d}(n)\mathbf{1}_{\{n<n_e\}} + \widetilde{U}_{C,d}(n)\mathbf{1}_{\{n \ge n_e\}}$, where $\mathbf{1}_A$ is the indicator of event A, and the two functions $\bar{U}_{C,d}(n)$ and $\widetilde{U}_{C,d}(n)$ are given by $\bar{U}_{C,d}(n) \equiv V\left(1 - \frac{d}{d+\beta}\prod_{k=0}^{n} \frac{\mu_k}{\mu_k+\beta}\right) - \frac{c}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu}\right), \widetilde{U}_{C,d}(n) \equiv V\frac{\beta}{\mu+\beta}\sum_{j=1}^{n_e}\prod_{k=j}^{n} \frac{\mu_k}{\mu_k+\beta} + V\frac{\beta}{\beta+d}\prod_{k=0}^{n} \frac{\mu_k}{\mu_k+\beta} - c\frac{\beta}{\mu+\beta}\sum_{j=1}^{n_e}\frac{j}{\mu}\prod_{k=i}^{n} \frac{\mu_k}{\mu_k+\beta} - \frac{c}{\beta}$. Given that all other remote customers place orders with probability q, the expected utility of a tagged remote customer who places an order is $U_{C,d}(q) = \sum_{i=0}^{n_e-1} \bar{U}_{C,d}(i)\pi_{i,C}^u(q) + \sum_{i=n_e}^{\infty} \widetilde{U}_{C,d}(i)\pi_{i,C}^u(q) = \sum_{i=0}^{n_e-1}\left[V\left(1 - \frac{d}{d+\beta}\prod_{k=0}^{i}\frac{\mu_k}{\mu_k+\beta}\right) - \frac{c}{\beta}\left(\sigma^{i+1} + \frac{(i+1)\beta}{\mu} - 1\right)\right]\pi_{i,C}^u(q) - \frac{c}{\beta} \quad + \sum_{i=n_e}^{\infty}\left[V\left(\frac{\beta}{\beta+d}\prod_{k=0}^{i}\frac{\mu_k}{\mu_k+\beta} + \frac{\beta}{\mu+\beta}\sum_{j=1}^{n_e}\prod_{k=j}^{i}\frac{\mu_k}{\mu_k+\beta}\right) - c\frac{\beta}{\mu+\beta}\sum_{j=1}^{n_e}\frac{j}{\mu}\prod_{k=j}^{i}\frac{\mu_k}{\mu_k+\beta}\right]\pi_{i,C}^u(q)$, where t$\pi_{i,C}^u(q)$ is given by (4). Similar to what we show in the proof of Proposition 4, we can show that $U_{C,d}(q)$ is decreasing in q. Hence, remote customers' equilibrium order-placing probability is $q_{C,d} = \begin{cases} 0, & \text{if} \quad \Lambda \ge \bar{\lambda}_{C,d}, \\ \hat{q}_{C,d} \in (0,1), & \text{if} \quad \underline{\lambda}_{C,d} < \Lambda < \bar{\lambda}_{C,d}, \\ 1, & \text{if} \quad \Lambda \le \underline{\lambda}_{C,d}, \end{cases}$ where $\hat{q}_{C,d}$ uniquely solves the equation $U_{C,d}(\hat{q}_{C,d}) = 0$. The resulting system throughput is $TH_{C,d} = \mu[1 - \pi_{0,C}^u(q_{C,d})]$.

(2) Queue-length information shared remotely. Customers follow the threshold $\hat{n}_e \le n_e^*$. The resulting system throughput is the same as in the case without cancellation.

Figure EC.1 **Throughput Comparison with Food Quality Degradation**



(a) $d = 0.3$ (b) $d = 0.5$

Note. $\mu = 1$, $V = 2$, $c = 0.5$, $\gamma = 0.7$.

EC.1.1.4. Throughput Comparison We replicate the four-way throughput comparison in Figure 9 by incorporating the effect of food-quality degradation. The results are presented in Figure EC.1. We observe that when travel is fast, the system throughput of any of three order-ahead schemes is hardly affected by incorporating quality degradation. When travel is fast, quality degradation rarely occurs because customers are unlikely to arrive at the service facility after their order is complete. However, when travel is slow, quality degradation is more likely to arise and quality degradation clearly takes its toll on all three order-ahead schemes when travel is slow. In particular, when quality degradation becomes more intense, as shown in Figure EC.1-(b)-(a), the hybrid order-ahead scheme can fall short of the pure order-onsite scheme. This issue can be addressed by strategic idleness as suggested by Farahani et al. (2022) and geo-location technology that alerts the kitchen when customers are getting close so it does not start preparing orders too early. It is worth emphasizing that these supply-side interventions to mitigate quality degradation cannot address the throughput shortfall of the hybrid order-ahead scheme when travel is fast because the throughput shortfall arises even in the absence of quality degradation.

EC.1.2. Channel Choice

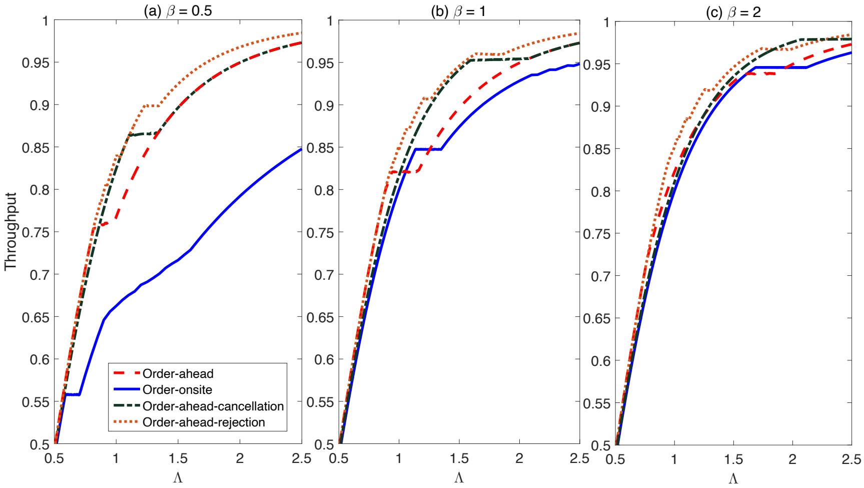
As argued in 6.2, modeling the channel choice will not affect the pure order-onsite scheme or the hybrid order-ahead-with-cancellation scheme. It will not affect the hybrid order-ahead scheme (with or without rejection) when queue-length information is shared remotely. The only cases in which modeling the channel choice will make a difference are the hybrid order-ahead scheme and that with rejection when queue-length information is not shared remotely. We rederive the equilibria below for these two cases.

EC.1.2.1. Hybrid Order-Ahead All customers who order onsite follow a threshold strategy n_e. We suppose that remote customers order ahead with probability $q_A \in [0,1]$ and order onsite with probability $q_S \in [0,1]$, and do not order with probability $1-q_A-q_S$, where $q_A+q_S \le 1$. The corresponding steady-state probability of the number of outstanding orders being i is $\pi_i^u(q_A,q_S) = \begin{cases} \rho_{T1}^i \pi_0^u(q_A,q_S), & i<n_e, \\ \rho_{R1}^{i-n_e}\rho_{T1}^{n_e}\pi_0^u(q_A,q_S), & i \ge n_e, \end{cases}$ for $\rho_{R1}<1$, and $\rho_{T1} = \frac{\gamma\Lambda(q_A+q_S)+(1-\gamma)\Lambda}{\mu}$ and $\rho_{R1} = \frac{\gamma\Lambda q_A}{\mu}$, and $\pi_0^u(q_A,q_S) = \left(\frac{1-\rho_{T1}^{n_e}}{1-\rho_{T1}} + \frac{\rho_{T1}^{n_e}}{1-\rho_{R1}}\right)^{-1}$. The expected utility for a remote customer who orders ahead is $U_A^R(q_A,q_S) = \sum_{n=0}^{\infty} \bar{U}(n)\pi_n^u(q_A,q_S)$. The expected utility for a remote customer who orders onsite is $U_S^R(q_A,q_S) = \sum_{n=0}^{n_e-1}\left(V - \frac{(n+1)c}{\mu}\right)\pi_n^u(q_A,q_S) - \frac{c}{\beta}$. Thus, remote customers' equilibrium order-placing probability is: $(q_A^e,q_S^e) = \begin{cases} (q_A^R,0), & \text{if } U_A^R(q_A^R,0)\ge 0 \text{ and } U_A^R(q_A^R,0) > U_S^R(q_A^R,0), \\ (0,q_S^R), & \text{if } U_S^R(0,q_S^R)\ge 0 \text{ and } U_S^R(0,q_S^R) > U_A^R(0,q_S^R), \\ (q_{A1}^R, 1-q_{A1}^R), & \text{if } U_A^R(q_{A1}^R,1-q_{A1}^R) = U_S^R(q_{A1}^R,1-q_{A1}^R) \ge 0, \\ (q_{A2}^R,q_{S2}^R), & \text{if } U_A^R(q_{A2}^R,q_{S2}^R) = U_S^R(q_{A2}^R,q_{S2}^R) = 0. \end{cases}$ When $q_A^R\in(0,1)$, it is solved by $U_A^R(q_A^R,0)=0$; when $q_S^R\in(0,1)$, it is solved by $U_S^R(0,q_S^R)=0$. The resulting system throughput is $TH^u = \mu[1-\pi_0^u(q_A^e,q_S^e)]$.

EC.1.2.2. Hybrid Order-Ahead with (Optimal) Rejection Given rejection threshold N and the order-placing probabilities (q_A,q_S) of remote customers, we first give the steady-state probability of the number of outstanding orders being i. If $N>n_e$ the steady-state probability of the number of outstanding orders being i is $\pi_{i,R}^u(q_A,q_S) = \begin{cases} \rho_{T1}^i\pi_{0,R}^u(q_A,q_S), & i<n_e, \\ \rho_{R1}^{i-n_e}\rho_{T1}^{n_e}\pi_{0,R}^u(q_A,q_S), & i=n_e,\cdots,N, \end{cases}$ where $\pi_{0,R}^u(q_A,q_S) = \left(\frac{1-\rho_{T1}^{n_e}}{1-\rho_{T1}} + \frac{\rho_{T1}^{n_e}(1-\rho_{R1}^{N-n_e+1})}{1-\rho_{R1}}\right)^{-1}$. If $N\le n_e$, the steady-state probability of the number of outstanding orders being i is $\pi_{i,R}^u(q_A,q_S) = \begin{cases} \rho_{T1}^i\pi_{0,R}^u(q_A,q_S), & i<N, \\ \rho_L^{i-N}\rho_{T1}^N\pi_{0,R}^u(q_A,q_S), & i=N,\cdots,n_e, \end{cases}$ where $\pi_{0,R}^u(q) = \left(\frac{1-\rho_{T1}^N}{1-\rho_{T1}} + \frac{\rho_{T1}^N(1-\rho_{L1}^{n_e-N+1})}{1-\rho_{L1}}\right)^{-1}$, and $\rho_{T1} = \frac{\gamma\Lambda(q_A+q_S)+(1-\gamma)\Lambda}{\mu}$ and $\rho_{R1} = \frac{\gamma\Lambda q_A}{\mu}$, and $\rho_{L1} = \frac{\gamma\Lambda q_S+(1-\gamma)\Lambda}{\mu}$. Thus, for a remote customer who places an order ahead, with probability $\pi_{N,R}^u(q_A,q_S)$, her order will be rejected (from which she gets zero utility); with probability $1-\pi_{N,R}^u(q_A,q_S)$, her order will be accepted (which implies the queue length at the moment is less than N). Thus, her unconditional expected utility from ordering ahead is $U_{R,N}^u(q_A,q_S) = \sum_{n=0}^{N-1}\bar{U}(n)\pi_{n,R}^u(q_A,q_S)$. For a remote customer who places an onsite order, her unconditional expected utility from ordering is $U_{S,R}^R(q_A,q_S) = \sum_{n=0}^{n_e-1}\left(V-\frac{(n+1)c}{\mu}\right)\pi_{n,R}^u(q_A,q_S) - \frac{c}{\beta}$. Thus, remote customers' equilibrium order-placing probability is: $(q_A^e,q_S^e) = \begin{cases} (q_A^R,0), & \text{if } U_{A,R}^R(q_A^R,0)\ge 0 \text{ and } U_{A,R}^R(q_A^R,0) > U_{S,R}^R(q_A^R,0), \\ (0,q_S^R), & \text{if } U_{S,R}^R(0,q_S^R)\ge 0 \text{ and } U_S^R(0,q_S^R) > U_{A,R}^R(0,q_S^R), \\ (q_A^R,1-q_A^R), & \text{if } U_{A,R}^R(q_A^R,1-q_A^R) = U_{S,R}^R(q_A^R,1-q_A^R)\ge 0, \\ (q_A^R,q_S^R), & \text{if } U_{A,R}^R(q_A^R,q_S^R) = U_{S,R}^R(q_A^R,q_S^R) = 0. \end{cases}$

When $q_A^R\in(0,1)$, it is solved by $U_{A,R}^R(q_A^R,0)=0$; when $q_S^R\in(0,1)$, it is solved by $U_{S,R}^R(0,q_A^R)=0$. The resulting system throughput is $TH^u = \mu[1-\pi_0^u(q_A^e,q_S^e)]$.

Figure EC.2 Throughput Comparison with Channel Choice



Note. $\mu = 1$, $V = 2$, $c = 0.5$, $\gamma = 0.7$.

EC.1.2.3. Throughput Comparison We replicate the four-way throughput comparison in Figure 9 by incorporating the channel choice. The result is presented in Figure EC.2. We observe that Figure EC.2 is qualitatively similar to Figure 9, suggesting that the insights from our base model carry over. When travel is slow, incorporating the channel choice barely affects the equilibrium outcome, as judged by the similarly between Figure EC.2-(a) and Figure 9-(a). In such a case, ordering ahead has such a substantial advantage that remote customers will not forgo it. When travel is fast, some remote customers switch to ordering onsite, making the hybrid order-ahead scheme more similar to the pure-order-onsite scheme, as illustrated by Figure EC.2-(c). Still, we find that providing the order-ahead option can result in lower throughput, but such shortfall can be addressed through cancellation or rejection.

EC.1.3. Heterogeneous Travel Speed of Remote Customers

Denote the cumulative distribution function of β by F. In the hybrid order-ahead scheme, when queue-length information is not shared remotely, we represent remote customers' equilibrium by (λ_A, λ_S), where λ_A and λ_S denote the arrival rates of remote customers who choose to order ahead and order onsite, respectively, By definition, $\lambda_A + \lambda_S \le \gamma\Lambda$. Further, the arrival rate of local customers is $\lambda_L = (1-\gamma)\Lambda$. Given the triplet $\boldsymbol{\lambda} \equiv (\lambda_A, \lambda_S, \lambda_L)$, let $\hat{\rho}_T = (\lambda_A + \lambda_S + \lambda_L)/\mu$ and $\hat{\rho}_R = \lambda_A/\mu$; for $\hat{\rho}_R < 1$, the corresponding steady-state probability of the number of outstanding orders being i is: $\hat{\pi}_0^u(\boldsymbol{\lambda}) = \left(\frac{1-\hat{\rho}_T^{n_e}}{1-\hat{\rho}_T} + \frac{\hat{\rho}_T^{n_e}}{1-\hat{\rho}_R}\right)^{-1}$; $\hat{\pi}_i^u(\boldsymbol{\lambda}) = \hat{\rho}_T^i \hat{\pi}_0^u(\boldsymbol{\lambda}), i < n_e$; $\hat{\pi}_i^u(\boldsymbol{\lambda}) = \hat{\rho}_R^{i-n_e}\hat{\rho}_T^{n_e}\hat{\pi}_0^u(\boldsymbol{\lambda}), i \ge n_e$. For a remote customer with travel speed β, let $U_A(\boldsymbol{\lambda},\beta)$ and $U_S(\boldsymbol{\lambda},\beta)$ be her expected utility of ordering ahead and that of ordering onsite, respectively. Thus, $U_A(\boldsymbol{\lambda},\beta) = \sum_{n=0}^{\infty} \bar{U}(n)\hat{\pi}_n^u(\boldsymbol{\lambda})$ and $U_S(\boldsymbol{\lambda},\beta) = \sum_{n=0}^{n_e-1}\left(V - \frac{(n+1)c}{\mu}\right)\hat{\pi}_n^u(\boldsymbol{\lambda}) - \frac{c}{\beta}$. Therefore, the equilibrium (λ_A, λ_S) solves the following set of fixed-point equations: $\lambda_A = \Lambda\gamma\int_a^b \mathbf{1}_{\{U_A(\boldsymbol{\lambda},\beta) > [U_S(\boldsymbol{\lambda},\beta)]^+\}} dF(\beta)$; $\lambda_S = \Lambda\gamma\int_a^b \mathbf{1}_{\{U_S(\boldsymbol{\lambda},\beta) > [U_A(\boldsymbol{\lambda},\beta)]^+\}} dF(\beta)$. The equilibrium for the hybrid order-ahead-with-rejection scheme and that for the pure order-onsite scheme can be similarly defined.

Appendix EC.2: Proofs

We first give the following two technical lemmas that will be repeatedly used in the subsequent proofs.

Lemma EC.2.1. *For a strictly decreasing function $f(n)$ and two probability distributions $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$, supported over $\{0, 1, \cdots, \bar{n}\}$, where $\bar{n}$ can possibly be ∞, if $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ cross each other only once (there exists an n^* such that $\pi_n^1 \ge \pi_n^2$ when $n \le n^*$ and $\pi_n^1 < \pi_n^2$ when $n > n^*$), then for two random variables X_1 and X_2 following $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$, we have $\mathbb{E}[f(X_1)] = \sum_{n=0}^{\bar{n}} f(n)\pi_n^1 > \sum_{n=0}^{\bar{n}} f(n)\pi_n^2 = \mathbb{E}[f(X_2)]$.*

Proof of Lemma EC.2.1 We write $\sum_{n=0}^{\bar{n}} f(n)\pi_n^1 - \sum_{n=0}^{\bar{n}} f(n)\pi_n^2 = \sum_{n=0}^{n^*} f(n)(\pi_n^1 - \pi_n^2) + \sum_{n=n^*+1}^{\bar{n}} f(n)(\pi_n^1 - \pi_n^2) > f(n^*)\left(\sum_{n=0}^{n^*}(\pi_n^1 - \pi_n^2) + \sum_{n=n^*+1}^{\bar{n}}(\pi_n^1 - \pi_n^2)\right) = 0$, where the inequality holds due to the single-crossing property, and the last equality holds because both $\boldsymbol{\pi}^1$ and $\boldsymbol{\pi}^2$ are well defined probability distributions over $\{0, 1, \cdots, \bar{n}\}$, i.e., $0 = 1 - 1 = \sum_{n=0}^{\bar{n}}(\pi_n^1 - \pi_n^2) = \sum_{n=0}^{n^*}(\pi_n^1 - \pi_n^2) + \sum_{n=n^*+1}^{\bar{n}}(\pi_n^1 - \pi_n^2)$. □

Lemma EC.2.2. *Consider a birth-and-death process with birth rate λ_{i-1} and death rate μ_i for state i and denote $\rho_i = \lambda_{i-1}/\mu_i$. We consider two systems indexed by (1) and (2) with stationary distributions $\boldsymbol{\pi}^{(1)}$ and $\boldsymbol{\pi}^{(2)}$, respectively. If $\rho_i^{(1)} \geq \rho_i^{(2)}$ for all $i = 1, \cdots$, we must have $\pi_0^{(1)} \leq \pi_0^{(2)}$, and the queue length of the first system is stochastically larger than that of the second system, i.e., $Q^{(1)} \geq_{st} Q^{(2)}$.*

Proof of Lemma EC.2.2 In a birth-and-death system, $\pi_0 = \frac{1}{1+\rho_1+\rho_1\rho_2+\cdots+\prod_{i=1}^{n}\rho_i+\cdots}$. Therefore, a larger ρ_i induces a smaller π_0. Hence, $\pi_0^{(1)} \leq \pi_0^{(2)}$. The steady-state probability of the number of customers being i in the two systems are $\pi_i^{(1)} = \rho_1^{(1)} \cdots \rho_i^{(1)} \pi_0^{(1)}$, and $\pi_i^{(2)} = \rho_1^{(2)} \cdots \rho_i^{(2)} \pi_0^{(2)}$. Thus, $\pi_i^{(1)}/\pi_i^{(2)}$ is increasing in i. Hence, $Q^{(1)} \geq_{st} Q^{(2)}$ in the likelihood ratio order. □

Proof of Lemma 1 Consider a remote customer with n existing orders upon her arrival. Let $T \sim \text{Exp}(\beta)$ be her travel time, and let I.I.D. $\text{Exp}(\mu)$ r.v.'s $S_1, \cdots, S_n$ denote the service times for the n outstanding orders, with S_i corresponding to the i^{th} order to be processed in the order queue. Let S_0 be the service time of the tagged customer. We next derive the distribution of N_n, the tagged remote customer' updated queue position (including herself) when she arrives at the service facility. Note that $N_n \in \{0, 1, \cdots, n+1\}$. We denote the probabilities by $p_n(0), \ldots, p_n(n+1)$. It is straightforward to see that the updated queue position is i if and only if there are exactly $n - i + 1$ service completions when the tagged customer arrives at the service facility; this corresponds to the event $\{S_1 + \cdots + S_{n-i+1} < T < S_1 + \cdots + S_{n-i+1} + S_{n-i+2}\}$. Let $\sigma \equiv \mu/(\mu + \beta)$. By the memoryless property of the exponential distribution, we have $p_n(n+1) \equiv \mathbb{P}(N_n = n+1) = \mathbb{P}(T < S_1) = 1 - \sigma; p_n(i) \equiv \mathbb{P}(N_n = i) = \mathbb{P}(T > S_1) \times \cdots \times \mathbb{P}(T > S_{n-i+1}) \times \mathbb{P}(T < S_{n-i+2}) = (1-\sigma)\sigma^{n-i+1}, i = 1, 2, \cdots, n; p_n(0) \equiv \mathbb{P}(N_n = 0) = \mathbb{P}(T > S_1) \times \cdots \times \mathbb{P}(T > S_n) \times \mathbb{P}(T > S_0) = \sigma^{n+1}$. □

Proof of Proposition 1 The proof of Proposition 1 will be based on Lemmas EC.2.1 and EC.2.2. Recall that when the queue-length information is not shared remotely, the expected utility for a remote customer who places an order is $U^u(q) = \sum_{n=0}^{\infty} \bar{U}(n)\pi_n^u(q)$. The next lemma establishes properties of $U_\rho^u(q)$.

Lemma EC.2.3 **(Property of U^u function)**. *The utility function U^u has the following properties: (i)$U_\infty^u(1) = -\infty$, $U_0^u(1) > 0$, $U_\infty^u(0) = \bar{U}(n_e) < 0$ and $U_0^u(0) > 0$; (ii) $U_\rho^u(q)$ is continuous and strictly decreasing in q for a fixed ρ; (iii) $U_\rho^u(1)$ and $U_\rho^u(0)$ are continuous and strictly decreasing in ρ.*

Proof of Lemma EC.2.3 To prove Part (i), first, we have $U_0^u(1) = U_0^u(0) = \bar{U}(0) > 0$ by Assumption 1. When the system load goes to infinity and if all remote customers place orders (i.e., $q = 1$), the birth-death process becomes unstable so that $U_\infty^u(1) = -\infty$. When the system load goes to infinity and if no other remote customers join (i.e., $q = 0$), the steady-state probability $\pi_{n_e}^u = 1$ because there are infinite local customers keeping the queue size at its capacity n_e. Hence, we have $U_\infty^u(0) = \bar{U}(n_e) < 0$. For Part (ii), continuity is obvious. To prove monotonicity, we pick $q_1 < q_2$ and consider the two corresponding steady state distributions $\{\pi_n^u(q_1)\}$ and $\{\pi_n^u(q_2)\}$, we have that $\pi_0^u(q_1) > \pi_0^u(q_2)$ since the system 2 is busier than system 1 according to Lemma EC.2.2. In addition, there must exist some integers n' such that $\pi_{n'}^u(q_1) < \pi_{n'}^u(q_2)$. Otherwise, we cannot have $\sum_{n=0}^{\infty} \pi_n^u(q_1) = \sum_{n=0}^{\infty} \pi_n^u(q_2) = 1$. We then claim that for any n', if $\pi_{n'}^u(q_1) < \pi_{n'}^u(q_2)$, then $\pi_n^u(q_1) < \pi_n^u(q_2)$ for $n \geq n' + 1$ due to the geometric structure of $\boldsymbol{\pi}$ distribution. Then it is straightforward to see these two distributions satisfy Condition (ii) of Lemma EC.2.1. And because $\bar{U}(n)$ is decreasing in n, it follows from Lemma EC.2.1 that $U_\rho^u(q_1) > U_\rho^u(q_2)$. Hence, $U_\rho^u(q)$ is decreasing in q. The proof of (iii) is similar to that of (ii). □

Finishing the proof of Proposition 1. From Parts (i) and (iii) of Lemma EC.2.3 and $U_0^u(1) = \bar{U}(0) > 0$ by Assumption 1, there must be a unique solution $\underline{\rho}_A^u$ to equation $U_\rho^u(1) = 0$ for $\rho \in (0,1)$. Similarly, there must be a unique solution $\bar{\rho}_A^u$ to equation $U_\rho^u(0) = 0$. Denote $\underline{\lambda}_A^u = \underline{\rho}_A^u \mu$ and $\bar{\lambda}_A^u = \bar{\rho}_A^u \mu$. By Lemma EC.2.3, $U^u(1) > 0$ when $\Lambda < \underline{\lambda}_A^u$, which implies that $q_A^u = 1$. On the other hand, when $\Lambda \geq \bar{\lambda}_A^u$, $q_A^u = 0$. Otherwise, q_A^u must satisfy $U^u(q) = 0$. □

Proof of Theorem 1 We first consider the small β case. When $\beta = \beta_0 \equiv c/(V - c/\mu)$, the cost of travel and undergoing a single service time becomes too high so no remote customer will join in the order-onsite model. Hence, the order-onsite model reduces to a standard $M/M/1/n_e$ model with arrival rate $(1-\gamma)\Lambda$ and service rate μ. On the other hand, the parallelization effect achieves some waiting time reduction so some remote customers may still join the system which makes the order-ahead model stochastically busier than the order-onsite model (to see this we again invoke Lemma EC.2.2). Hence, the order-ahead model yields higher throughput than the order-onsite model. We next consider the large β case. We define $\Delta\lambda \equiv \underline{\lambda}_S^u - \bar{\lambda}_A^u$, where $\underline{\lambda}_S^u$ is defined in the order-onsite model and $\bar{\lambda}_A^u$ is defined in (3). Our strategy is to study the asymptotic behavior of $\underline{\lambda}_S^u$ and $\bar{\lambda}_A^u$ when β is sufficiently large. We will show that, when β grows large, $\underline{\lambda}_S^u$ increases without bound whereas $\bar{\lambda}_A^u$ does not (it approaches a finite number). Then, we will have an interval $[\bar{\lambda}_A^u, \underline{\lambda}_S^u]$ such that $q_A^u(\Lambda) = 0$ and $q_S^u(\Lambda) = 1$ for all $\Lambda \in [\bar{\lambda}_A^u, \underline{\lambda}_S^u]$, as long as β is sufficiently large. This result will ensure that the order-onsite Continuous Time Markov Chain (CTMC) is stochastically busier than the order-ahead CTMC when $\Lambda \in [\bar{\lambda}_A^u, \underline{\lambda}_S^u]$. To see this, note that the two models have an equal death rate but the former has a strictly larger birth rate than the latter. Hence, invoking Lemma EC.2.2, we must have $\pi_0^S(\Lambda) < \pi_0^A(\Lambda)$, so that the order-onsite model yields strictly higher throughput for all $\Lambda \in [\bar{\lambda}_A^u, \underline{\lambda}_S^u]$. For $\bar{\lambda}_A^u$, we let $\beta \to \infty$, so that a remote customer's utility in the order-ahead model with $q = 0$ is $U^u(0) = \sum_{i=0}^{n_e} \left(V - \frac{(i+1)c}{\mu} - \frac{c}{\beta}\sigma^{i+1}\right) \pi_i^u(0) \to \sum_{i=0}^{n_e} \left(V - \frac{(i+1)c}{\mu}\right) \pi_i^u(0) = \sum_{i=0}^{n_e-1} \underbrace{\left(V - \frac{(i+1)c}{\mu}\right)}_{\geq 0} \pi_i^u(0) + \underbrace{\left(V - \frac{(n_e+1)c}{\mu}\right)}_{<0} \pi_{n_e}^u(0)$, where $\pi_i^u(0) = \frac{\rho_L^i(1-\rho_L)}{1-\rho_L^{n_e+1}}$, $i = 0, 1, \ldots, n_e$, and $\rho_L = (1-\gamma)\Lambda/\mu$. As Λ increases, $\pi_{n_e}^u(0)$ will have a bigger weight so the negative term in $U^u(0)$ will dominate the positive terms, so that $U^u(0)$ will become negative when $\Lambda > \bar{\Lambda}$ for some finite $\bar{\Lambda}$. Recall that $\bar{\lambda}_A^u$ is the root of $U^u(0) = 0$, $\bar{\lambda}_A^u$ must be a finite number as β grows large. On the other hand, recall that the utility function of remote customers in the order-onsite model is $U_S^u(q) = \sum_{i=0}^{n_e-1} \left(V - \frac{(i+1)c}{\mu}\right) \pi_{i,S}^u(q) - \frac{c}{\beta}$. When β is sufficiently large, $U_S^u(q) > 0$ for any $q \in [0,1]$ (because $V - (i+1)c/\mu > 0$ for all $i = 0, \ldots, n_e - 1$). So we must have $\underline{\lambda}_S^u = \infty$ and $\Delta\lambda = -\infty$. Hence, $\Delta\lambda$ is sufficiently negative when β is large. Because $\Delta\lambda$ is continuous in β, there exists a $\bar{\beta}$ such that $\Delta\lambda < 0$ for all $\beta > \bar{\beta}$. By the definition of $\underline{\lambda}_S^u$ and $\bar{\lambda}_A^u$, we now have $q_A^u(\Lambda) = 0$ and $q_S^u(\Lambda) = 1$ for all $\Lambda \in [\bar{\lambda}_A^u, \underline{\lambda}_S^u]$, as long as $\beta > \bar{\beta}$. Thus, Lemma EC.2.2 implies that the order-ahead scheme has lower throughput than the order-onsite model. □

Proof of Proposition 2 Recall that the expected utility of a remote customer who sees a queue length n and places an order is $\bar{U}(n) = V - \frac{c}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu}\right)$. We leverage the following properties of $\bar{U}(n)$: (i) $\lim_{n\to\infty} \bar{U}(n) = -\infty$; (ii) $\bar{U}(n)$ is strictly decreasing in n since $\bar{U}(n) - \bar{U}(n-1) = c(\sigma^{n+1} - 1)/\mu < 0$. Since $\bar{U}(0) > 0$ (Assumption 1), the equilibrium strategy must be of a threshold type, given by $n_e^* \equiv \min\{n \geq 0 : \bar{U}(n) < 0\}$, and the threshold n_e^* is at most n_e. Since $\bar{U}(n_e) = V - \frac{c}{\beta}\left(\sigma^{n_e+1} + \frac{(n_e+1)\beta}{\mu}\right) = V - \frac{c(n_e+1)}{\mu} - \frac{c}{\beta}\sigma^{n_e+1} < 0$, and $\bar{U}(n)$ is decreasing in n, the joining threshold n_e^* must be attained in $\{0, 1, \ldots, n_e\}$, i.e., $n_e^* = \min\{n \geq 0 : \bar{U}(n) < 0\}$. When $\lfloor \frac{\mu V}{c} \rfloor = \frac{\mu V}{c}$, for any finite β, $\bar{U}(n_e - 1) = -\frac{c}{\beta}\sigma^{n_e} < 0$, so that the joining threshold $n_e^* \leq n_e - 1 < n_e$. Otherwise, from Equation (1), $\bar{U}(n)$ decreases in n and increases in β. Note that $n_e^* = n_e$ if and only if $\bar{U}(n_e - 1) = V - \frac{cn_e}{\mu} - \frac{c}{\beta}\sigma^{n_e} \geq 0$, which requires that $\beta \geq \underline{\beta}$, where $\underline{\beta}$ is given by $V - \frac{cn_e}{\mu} - \frac{c}{\underline{\beta}}\sigma^{n_e} = 0$. □

Proof of Proposition 3 Given any Λ, there exists a unique order-placing probability $\tilde{q} \in (0,1)$, which induces the same throughput regardless of whether queue-length information is shared directly or not remotely, i.e.,

$TH_A^u(\widetilde{q}) = TH_A^o$. It holds since $TH_A^u(0) < TH_A^o < TH_A^u(1)$ by Lemma EC.2.3 and the throughput when queue-length information is not shared remotely increases in the order-placing probability q given the market size. Recall from Proposition 1 that the equilibrium order-placing probability of remote customers under the hybrid order-ahead scheme decreases in the market size, and $q_A^u = 1$ when the market size $\Lambda < \underline{\lambda}_A^u$. Hence, $\widetilde{q} < q_A^u = 1$ when $\Lambda < \underline{\lambda}_A^u$. Pick two market sizes Λ_1, Λ_2 ($\underline{\lambda}_A^u < \Lambda_1 < \Lambda_2$), there exist two order-placing probabilities $\widetilde{q}_1(\Lambda_1), \widetilde{q}_2(\Lambda_2)$ that satisfy $TH_A^u(\widetilde{q}_1(\Lambda_1)) = TH_A^o(\Lambda_1)$ and $TH_A^u(\widetilde{q}_2(\Lambda_2)) = TH_A^o(\Lambda_2)$. We next prove $U^u(\widetilde{q}_1(\Lambda_1)) - U^u(\widetilde{q}_2(\Lambda_2)) = \sum_{i=0}^{\infty} \bar{U}(n)(\pi_i^u(\widetilde{q}_1(\Lambda_1)) - \pi_i^u(\widetilde{q}_2(\Lambda_2)) > 0$. Since $TH_A^u(\widetilde{q}_1(\Lambda_1)) = TH_A^o(\Lambda_1)$ and $TH_A^u(\widetilde{q}_2(\Lambda_2)) = TH_A^o(\Lambda_2)$, we also have $\pi_0^u(\widetilde{q}_1(\Lambda_1)) = \pi_0^o(\Lambda_1) > \pi_0^o(\Lambda_2) = \pi_0^u(\widetilde{q}_2(\Lambda_2))$, and there must exists some $\tilde{n}$ such that $\pi_{\tilde{n}}^u(\widetilde{q}_1(\Lambda_1)) < \pi_{\tilde{n}}^u(\widetilde{q}_2(\Lambda_2))$. Otherwise, we cannot have $\sum_{i=0}^{\infty} \pi_i^u(\widetilde{q}_1(\Lambda_1)) = \sum_{i=0}^{\infty} \pi_i^u(\widetilde{q}_2(\Lambda_2)) = 1$. Recall that $\pi_0^u(\widetilde{q}_1) = \left(\frac{1-\rho_{T1}^{n_e}}{1-\rho_{T1}} + \frac{\rho_{T1}^{n_e}}{1-\rho_{R1}}\right)^{-1}$ and $\pi_0^u(\widetilde{q}_2) = \left(\frac{1-\rho_{T2}^{n_e}}{1-\rho_{T2}} + \frac{\rho_{T2}^{n_e}}{1-\rho_{R2}}\right)^{-1}$, where $\rho_{T1} = \frac{\gamma\Lambda_1\widetilde{q}_1 + (1-\gamma)\Lambda_1}{\mu}, \rho_{T2} = \frac{\gamma\Lambda_2\widetilde{q}_2 + (1-\gamma)\Lambda_2}{\mu}, \rho_{R1} = \frac{\gamma\Lambda_1\widetilde{q}_1}{\mu}, \rho_{R2} = \frac{\gamma\Lambda_2\widetilde{q}_2}{\mu}$. First, we aim to show that $\rho_{T1} < \rho_{T2}$. To see this, assume $\rho_{T1} \geq \rho_{T2}$, then because $\Lambda_1 < \Lambda_2$, we must have $\rho_{R1} > \rho_{R2}$. The geometric structure of the steady-state probability in (2), along with $\pi_0^u(\widetilde{q}_1) > \pi_0^u(\widetilde{q}_2)$ implies that $\pi_i^u(\widetilde{q}_1) > \pi_i^u(\widetilde{q}_2)$ for all i. Hence, a contradiction. Next, we aim to show that $\rho_{R1} < \rho_{R2}$. To see this, assume that $\rho_{R1} \geq \rho_{R2}$ which is equivalent to $\Lambda_1\tilde{q}_1 \geq \Lambda_2\tilde{q}_2$. Consider the special case $\gamma = 1$, where the steady-state distribution follows an exact geometric structure with $\rho_{T1} = \rho_{R1} \geq \rho_{T2} = \rho_{R2}$. Because $\pi_0^u(\widetilde{q}_1) > \pi_0^u(\widetilde{q}_2)$, similarly, we have a contradiction. Hence, we conclude that $\rho_{T1} < \rho_{T2}$ as well as $\rho_{R1} < \rho_{R2}$. Note that the two probability distributions have the same structure (both are geometric-like, with the former having a bigger probability mass at 0 than the latter. Hence, it is straightforward to see that there must exists some $\tilde{n}$ such that $\pi_n^u(q_1(\Lambda_1)) > (\leq) \pi_n^u(q_2(\Lambda_2))$ when $n < \tilde{n}$ ($n \geq \tilde{n}$). We shall show that the probability distribution $\{\pi_n^u(\widetilde{q}_2)\}_{n=0}^{\infty}$ stochastically dominates the probability distribution $\{\pi_n^u(\widetilde{q}_1)\}_{n=0}^{\infty}$. Hence, the distribution of $\{\pi_n^u(\widetilde{q})\}_{n=0}^{\infty}$ satisfies the condition (ii) of technical Lemma EC.2.1 and $U^u(\widetilde{q}_1(\Lambda_1)) - U^u(\widetilde{q}_2(\Lambda_2)) > 0$. is proved. Further, the utility function of remote customers under equilibrium is $U^u(\Lambda) = 0$ when $\underline{\lambda}_A^u \leq \Lambda < \bar{\lambda}_A^u$. Then, $U^u(\widetilde{q}(\Lambda)) - U^u(q_A^u)$ is decreasing in Λ, and this indicates $\widetilde{q} - q_A^u$ increases in $\Lambda \in (\underline{\lambda}_A^u, \bar{\lambda}_A^u)$. Since we have know that the equilibrium order-placing probability $q_A^u = 0 < \widetilde{q} \in (0,1)$ when the market size $\Lambda \geq \bar{\lambda}_A^u$ and $q_A^u = 1 > \widetilde{q} \in (0,1)$ when the market size $\Lambda \leq \bar{\lambda}_A^u$, this reminds us that there exists a unique market size $\widetilde{\Lambda} \in (\underline{\lambda}_A^u, \bar{\lambda}_A^u)$ that enables $q_A^u > (\leq) \widetilde{q}$ when $\Lambda < (\geq) \widetilde{\Lambda}$. Therefore, the throughput under two information provision policies are equal when the market size $\Lambda = \widetilde{\Lambda}$, and when $\Lambda < \widetilde{\Lambda}$, $TH_A^u(q_A^u) > TH_A^u(\widetilde{q}) = TH_A^o$; when $\Lambda > \widetilde{\Lambda}$, $TH_A^u(q_A^u) < TH_A^u(\widetilde{q}) = TH_A^o$. □

Proof of Theorem 2 To compare the throughput of the two models under optimal information, we first consider the order-onsite model. Recall that when the market size $\Lambda \leq \underline{\lambda}_S^u$, the equilibrium order-placing probability of remote customers is $q_S^u = 1$ of the order-onsite model. We then focus on the specific market size $\Lambda = \underline{\lambda}_S^u$. In this case, $TH_S^* = TH_S^u$ by using the result in technical Lemma EC.2.2. We compare the throughput under optimal information of the two models by considering two cases: First, when $TH_A^* = TH_A^o$, the order-onsite model outperforms the order-ahead model by achieving a higher throughput ($TH_S^u > TH_A^* = TH_A^o$) by using the result in technical Lemma EC.2.2. Otherwise, when the maximum throughput of the order-ahead model is $TH_A^* = TH_A^u$, the equilibrium order-placing probability $q_S^u = 1$ while q_A^u is solved by Proposition 1. The rest of the proof is similar to the last part of the proof of Theorem 1, where we showed that for a sufficiently large β, we must have $TH_S^u > TH_A^u$ when $\Lambda = \underline{\lambda}_S^u$. □

Proof of Lemma 2 Consider an arriving customer with N outstanding orders ahead of hers. (i) If $N < n_e$, her expected utility if she keeps on waiting is no less than $V - c(N+1)/\mu \geq 0$. This lower bound $V - c(N+1)/\mu$ would be attained if no customers ahead of her were to cancel their order. Since even the lower bound is nonnegative, any customer seeing $N < n_e$ keeps on waiting. (ii) If $N \geq n_e$, the arriving customer knows (based on the preceding argument) that the first n_e outstanding orders will not be canceled. Thus, her expected utility if she keeps on waiting is no greater than $V - c(n_e+1)/\mu < 0$. This upper bound $V - c(n_e+1)/\mu$ would be attained if all customers ahead of her beyond the first n_e were to cancel their order. Since even the upper bound is negative, any customer seeing $N \geq n_e$ abandons and cancels her order. $\square$

Proof of Lemma 3 Consider a tagged remote customer who observes n existing orders upon her arrival. Let $T \sim \text{Exp}(\beta)$ denote her travel time, and $S_0 \sim \text{Exp}(\mu)$ denote her service time. Let r.v.'s $Y_1, \ldots, Y_n$ denote inter-departure time of the order for the n outstanding orders (excluding the tagged customer), either by service completion or by order cancellation, with Y_i corresponding to the i^{th} order in the order queue. The corresponding departure rate is $\mu_i = \mu + (i - n_e)^+\beta, \quad i = 0, 1, \cdots$. We next derive the distribution for N_n^C, the tagged remote customer' updated queue position (including herself) when she arrives at the service facility. Note that $N_n^C \in \{0, 1, \cdots, n+1\}$. We denote the probabilities by $p_n^C(0), \ldots, p_n^C(n+1)$. It is straightforward to see that the updated queue position is i if and only if there are exactly $n - i + 1$ orders removed (either for service completion or cancellation) from the order queue when the tagged customer arrives at the service facility; this corresponds to the event $\{Y_1 + \cdots + Y_{n-i+1} < T < Y_1 + \cdots + Y_{n-i+1} + Y_{n-i+2}\}$. (i) If $n > n_e$, the probability distribution of N_n^C is given by $p_n^C(n+1) \equiv \mathbb{P}(N_n^C = n+1) = \mathbb{P}(T < Y_{n+1}) = \frac{\beta}{\mu_n + \beta}$; $p_n^C(i) \equiv \mathbb{P}(N_n^C = i) = \mathbb{P}(T > Y_{n+1}) \times \cdots \times \mathbb{P}(T > Y_{i+1}) \times \mathbb{P}(T < Y_i) = \frac{\beta}{\mu_{i-1}+\beta} \prod_{k=i}^{n} \frac{\mu_k}{\mu_k + \beta}, 1 \leq i \leq n$; $p_n^C(0) \equiv \mathbb{P}(N_n^C = 0) = \mathbb{P}(T > Y_{n+1}) \times \cdots \times \mathbb{P}(T > Y_1) \times \mathbb{P}(T > S_0) = \prod_{k=0}^{n} \frac{\mu_k}{\mu_k+\beta}$; (ii) If $n \leq n_e$, the departure rate of outstanding orders is $\mu_n = \mu$ and then N_n^C has the same distribution as N_n given by Lemma 1. $\square$

Proof of Proposition 4 First, we characterize remote customers' expected utility function. Consider a tagged remote customer who observes n outstanding orders upon experiencing a need and places an order. The probability she will continue to wait upon arriving onsite is $\vartheta_C(n) = \mathbb{P}(N_n^C \leq (n+1) \wedge n_e) = \begin{cases} 1, & \text{if } \; n < n_e \\ \sum_{i=0}^{n_e} p_n^C(i) = \prod_{k=0}^{n} \frac{\mu_k}{\mu_k+\beta} + \frac{\beta}{\mu+\beta} \sum_{i=1}^{n_e} \prod_{k=i}^{n} \frac{\mu_k}{\mu_k+\beta}, & \text{otherwise} \end{cases}$ where the probabilities $p_n^C(i)$ are given in Lemma 3, and $x \wedge y \equiv \min\{x, y\}$. Hence, the mean remaining onsite waiting time if joining is $w_C(n) \equiv \sum_{i=0}^{(n+1)\wedge n_e} \frac{i}{\mu} \cdot p_n^C(i) = \begin{cases} \frac{1}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu} - 1\right), & \text{if } \; n < n_e; \\ \frac{\beta}{\mu+\beta} \sum_{i=1}^{n_e} \frac{i}{\mu} \prod_{k=i}^{n} \frac{\mu_k}{\mu_k+\beta}, & \text{otherwise.} \end{cases}$ Let $U_C(n)$ denote the expected utility of a remote customer who observes a queue length n and places an order to join the queue at Stage 1. Thus, $U_C(n) = V\vartheta_C(n) - cw_C(n) - \frac{c}{\beta} = \bar{U}_C(n)\mathbf{1}_{\{n<n_e\}} + \widetilde{U}_C(n)\mathbf{1}_{\{n \geq n_e\}}$, where the indicator function $\mathbf{1}_A$ is equal to 1 if condition A holds and 0 otherwise, and the two functions $\bar{U}_C(n)$ and $\widetilde{U}_C(n)$ are given by $\bar{U}_C(n) \equiv V - \frac{c}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu} - 1\right) - \frac{c}{\beta} = \bar{U}(n), \widetilde{U}_C(n) \equiv V\left(\prod_{k=0}^{n} \frac{\mu_k}{\mu_k+\beta} + \frac{\beta}{\mu+\beta} \sum_{j=1}^{n_e} \prod_{k=j}^{n} \frac{\mu_k}{\mu_k+\beta}\right) - c\frac{\beta}{\mu+\beta} \sum_{j=1}^{n_e} \frac{j}{\mu} \prod_{k=j}^{n} \frac{\mu_k}{\mu_k+\beta} - \frac{c}{\beta}$. Therefore, when the queue-length information is not shared remotely in the cancellation model, given that all other remote customers place orders with probability q, the expected utility for a tagged customer to place an order is $U_C^u(q) = \sum_{i=0}^{n_e-1} \bar{U}_C(i)\pi_{i,C}^u(q) + \sum_{i=n_e}^{\infty} \widetilde{U}_C(i)\pi_{i,C}^u(q) = \sum_{i=0}^{n_e-1}\left(V - \frac{c}{\beta}\left(\sigma^{i+1} + \frac{(i+1)\beta}{\mu} - 1\right)\right)\pi_{i,C}^u(q) - \frac{c}{\beta} + \sum_{i=n_e}^{\infty}\left[V\left(\prod_{k=0}^{i} \frac{\mu_k}{\mu_k+\beta} + \frac{\beta}{\mu+\beta} \sum_{j=1}^{n_e} \prod_{k=j}^{i} \frac{\mu_k}{\mu_k+\beta}\right) - c\frac{\beta}{\mu+\beta} \sum_{j=1}^{n_e} \frac{j}{\mu} \prod_{k=j}^{i} \frac{\mu_k}{\mu_k+\beta}\right]\pi_{i,C}^u(q)$, where the steady-state probabilities are given by $\pi_{i,C}^u(q) = \left(\frac{1-\rho_T^{n_e+1}}{1-\rho_T} + \rho_T^{n_e} \sum_{j=1}^{\infty} \prod_{k=1}^{j} \frac{\gamma\Lambda q}{\mu+k\beta}\right)^{-1} \rho_T^{(i \wedge n_e)} \prod_{k=1}^{(i-n_e)^+} \frac{\gamma\Lambda q}{\mu+k\beta}, \quad i = 0, 1, \cdots$, and $\rho_T = [\gamma\Lambda q + (1-\gamma)\Lambda]/\mu$. To show $U_C(n)$ is strictly decreasing in n, we have $\bar{U}_C(n) - \bar{U}_C(n-1) = \frac{c}{\mu}\left(\sigma^{n+1} - 1\right) < 0, \widetilde{U}_C(n) - \widetilde{U}_C(n-1) = \left(\frac{\mu_n}{\mu_n+\beta} - 1\right)\left[V\prod_{k=0}^{n-1} \frac{\mu_k}{\mu_k+\beta} + \frac{\beta}{\mu+\beta} \sum_{i=1}^{n_e}\left(V - \frac{ic}{\mu}\right)\prod_{k=i}^{n-1} \frac{\mu_k}{\mu_k+\beta}\right] < 0$. In addition, $\widetilde{U}_C(n_e) - \bar{U}_C(n_e - 1) = (\sigma - 1)\left(V - \frac{n_e c}{\mu}\right) + \frac{c}{\mu+\beta}\left(\sigma^{n_e} - 1\right) < 0$, which implies that $U_C(n)$ is decreasing in n. To give customers' equilibrium joining strategy, we first give some properties of $U_C^u(q)$ in (5) in the following Lemma. Let $U_{C,\rho}^u(q)$ denote the expected utility of a joining remote customer when $\Lambda/\mu = \rho$ and other remote customers join with probability q.

LEMMA EC.2.4 **(Property of U_C^u function)**. *The utility function U_C^u has the following properties: (i) $U_{C,\infty}^u(1) = U_{C,\infty}^u(0) = -c/\beta < 0$ and $U_{C,0}^u(1) = U_{C,0}^u(0) = \bar{U}(0) > 0$. (ii) $U_C^u(q)$ is continuous and strictly decreasing in q. (iii) $U_C^u(1)$ is continuous and strictly decreasing in ρ, where $U_C^u(1)$ is given by* $U_C^u(1) = \sum_{i=0}^{n_e-1}\left(V - \frac{c}{\beta}\left(\sigma^{i+1} + \frac{(i+1)\beta}{\mu} - 1\right)\right)\pi_{i,C}^u(1) - \frac{c}{\beta} + \sum_{i=n_e}^{\infty}\left[V\left(\prod_{k=0}^{i} \frac{\mu_k}{\mu_k+\beta} + \frac{\beta}{\mu+\beta} \sum_{j=1}^{n_e} \prod_{k=j}^{i} \frac{\mu_k}{\mu_k+\beta}\right) - c\frac{\beta}{\mu+\beta} \sum_{j=1}^{n_e} \frac{j}{\mu} \prod_{k=j}^{i} \frac{\mu_k}{\mu_k+\beta}\right]\pi_{i,C}^u(1).$

The proof of Lemma EC.2.4 is similar to that of Lemma EC.2.3 and thus omitted due to the page limit.

Finishing the proof of Proposition 4. Similar to the proof of Proposition 1, and by the property of the U_C^u function, there must be a unique solution $\underline{\rho}_C^u$ to equation $U_C^u(1)=0$. Similarly, there must be a unique solution $\bar{\rho}_C^u$ to equation $U_C^u(0)=0$. Denoting $\underline{\lambda}_C^u=\mu\underline{\rho}_C^u$ and $\bar{\lambda}_C^u=\mu\bar{\rho}_C^u$ completes the proof. □

Proof of Theorem 3 The proof proceeds in two steps. **Step 1**: We first prove $TH_C^u \geq TH_S^u$. We compare the system throughput of the order-ahead-with-cancellation (OAC) model and the order-onsite model in the following cases specified by the equilibrium order-placing probabilities of remote customers (q_C^u and q_S^u): **Case 1:** $q_C^u \geq q_S^u$. This case includes three subcases: (a) $q_C^u=1, q_S^u\in[0,1]$, (b) $q_S^u=q_C^u=0$, and (c) $q_C^u\in(0,1), q_S^u=0$. Note that $\pi_{0,C}^u(q)=\left(\frac{1-\rho_T^{n_e+1}}{1-\rho_T}+\rho_T^{n_e}\sum_{j=1}^{\infty}\prod_{k=1}^{j}\frac{\gamma\Lambda q}{\mu+k\beta}\right)^{-1}\leq\left(\frac{1-\rho_T^{n_e+1}}{1-\rho_T}\right)^{-1}=\pi_{0,S}^u(q)$. Therefore, $\pi_{0,C}^u(q_C^u)\leq\pi_{0,C}^u(q_S^u)\leq\pi_{0,S}^u(q_S^u)$, which implies that $TH_C^u\geq TH_S^u$. **Case 2:** $q_C^u<q_S^u$. In this case, the equilibrium order-placing probability of remote customers must be strictly positive with $q_S^u>0$ and we must have $q_C^u<1$. When $n_e=1$, in the order-onsite model, the remote customer's expected utility from traveling is $\pi_{0,S}^u(q_S^u)(V-c/\mu)-c/\beta\geq0$, where $\pi_{0,S}^u$ is the idle probability. We then consider the following two subcases specified by the value of q_C^u. **Case 2a:** We first consider the case $q_C^u\in(0,1)$. In the OAC model, the remote customer's expected utility from ordering is $p_0V+p_1(V-c/\mu)-c/\beta=0$, where p_0 is the probability that the order is ready when the customer arrives at the service facility, p_1 is the probability that the order is not ready when the customer arrives at the service facility, and $1-p_0-p_1$ is the cancellation probability. Let $\pi_{0,C}^u$ be the idle probability in the OAC model. Thus, $p_0+p_1>\pi_{0,C}^u$, because if the system is idle when a customer places the order, the customer will surely not cancel the order. Moreover, we have $\underbrace{\pi_{0,S}^u(V-c/\mu)-c/\beta}_{\geq0}\geq\underbrace{p_0V+p_1(V-c/\mu)-c/\beta}_{=0}>(p_0+p_1)(V-c/\mu)-c/\beta$, which implies $\pi_{0,S}^u>p_0+p_1$. Hence $\pi_{0,S}^u>\pi_{0,C}^u\Longleftrightarrow\mu(1-\pi_{0,C}^u)>\mu(1-\pi_{0,S}^u)$. Therefore, the OAC model yields higher throughput than the order-onsite model. **Case 2b:** We next consider the case $q_C^u=0$. First, we must have $p_0V+p_1(V-c/\mu)-c/\beta\leq0$. The OAC system with $q_C^u=0$ is equivalent to an $M/M/1/1$ queue. Thus, the steady-state probability that the system has exactly one outstanding order is $\pi_{1,C}^u=\frac{\Lambda(1-\gamma)}{\mu+\Lambda(1-\gamma)}$, and the steady-state probability that the system is empty is $\pi_{0,C}^u=1-\pi_{1,C}^u$. The cancellation probability is $(1-p_0-p_1)$, i.e., the probability of seeing exactly two outstanding orders (including one's own order) ahead after arriving onsite. Hence, $1-p_0-p_1=\pi_{1,C}^u\frac{\beta}{\beta+\mu}<\pi_{1,C}^u$. Thus, $p_0+p_1>\pi_{0,C}^u$. Note that in the order-onsite system, $\pi_{0,S}^u(q)$ is decreasing in q and $\pi_{0,S}^u(0)=\pi_{0,C}^u(0)$. Because $q_S^u>0$, we have $\pi_{0,C}^u(0)>\pi_{0,S}^u(q_S^u)$. Thus, $p_0+p_1>\pi_{0,C}^u(0)>\pi_{0,S}^u(q_S^u)$, which implies that $p_0V+p_1(V-c/\mu)-c/\beta>\pi_{0,S}^u(q_S^u)(V-c/\mu)-c/\beta\geq0$. That is, $p_0V+p_1(V-c/\mu)-c/\beta>0$. However, this contradicts $p_0V+p_1(V-c/\mu)-c/\beta\leq0$. Therefore, the case $q_C^u=0$ and $q_S^u>0$ does not exist. In summary, when $n_e=1$, $TH_C^u\geq TH_S^u$.

Step 2: We next prove $TH_C^o\geq TH_S^o$ for $n_e=1$. Note that $TH_C^o=TH_A^o$. Hence, we need to prove $TH_A^o\geq TH_S^o$ for $n_e=1$. This is done later in Theorem EC.1. □

To facilitate our proofs, we introduce an order-ahead model variant when queue information is shared with remote customers. In this variant, remote customers can cancel their orders upon arrival at the store, based on the updated queue position of their order at that time. We refer to this variant as "order-ahead with onsite balking" (OAOB).

Lemma EC.2.5. *In the OAOB model, a remote customer orders if and only if the queue length $n<n_e^*$.*

Proof of Lemma EC.2.5 Consider a tagger remote customer who observes n orders upon arriving, N_n^{OB} denotes her updated queue position (including herself) upon arrival at the service facility if she places an order. If she places an order and travels to the service facility, she will keep the order with probability $\vartheta(n)\equiv\mathbb{P}(N_n^{OB}\leq n_e)$. Because $N_n^{OB}\leq n+1$, we have $\vartheta(n)=\mathbb{P}(N_n^{OB}\leq(n+1)\wedge n_e)=\begin{cases}1, & \text{if } n<n_e,\\ \sum_{i=0}^{n_e}p_n(i)=\sigma^{n-n_e+1}, & \text{otherwise,}\end{cases}$ where probabilities $p_n(i)$ are given in Lemma 1, and $x\wedge y\equiv\min\{x,y\}$. Her expected onsite waiting time, $w(n)$, is $w(n)=$

$\sum_{i=0}^{(n+1)\wedge n_e} \frac{i}{\mu} \cdot p_n(i) = \begin{cases} \frac{1}{\beta}\left(\sigma^{n+1} + \frac{(n+1)\beta}{\mu} - 1\right), & \text{if} \quad n < n_e, \\ \frac{1}{\beta}\left(\sigma^{n+1} - \left(1 - \frac{n_e\beta}{\mu}\right)\sigma^{n-n_e+1}\right), & \text{otherwise.} \end{cases}$ Let $U^{OB}(n)$ denote her expected utility. Then $U^{OB}(n) = V\vartheta(n) - cw(n) - \frac{c}{\beta} = \bar{U}(n)\mathbf{1}_{\{n<n_e\}} + \widetilde{U}(n)\mathbf{1}_{\{n\geq n_e\}}$, where $\bar{U}(n)$ is given by (1) and $\widetilde{U}(n) = V\sigma^{n-n_e+1} - \frac{c}{\beta}\left(\sigma^{n+1} - \left(1 - \frac{n_e\beta}{\mu}\right)\sigma^{n-n_e+1} + 1\right)$. It remains to show that $U^{OB}(n) < 0$ for all $n \geq n_e^*$. If $n_e^* < n_e$, $U^{OB}(n_e^*) = \bar{U}(n_e^*) < 0$; otherwise, $U^{OB}(n_e^*) = \widetilde{U}(n_e)$ and $\widetilde{U}(n_e) = \sigma\left(V - \frac{n_e c}{\mu}\right) + \frac{c}{\beta}\left(\sigma\left(1 - \sigma^{n_e}\right) - 1\right) = \sigma\left[\left(V - \frac{c(n_e+1)}{\mu}\right) - \frac{c}{\beta}\sigma^{n_e}\right] < 0$. We then conclude that $U^{OB}(n_e^*) < 0$. To show the monotonicity of $U^{OB}(n)$ in n, we establish the monotonicity for both $\bar{U}(n)$ and $\widetilde{U}(n)$, namely, $\bar{U}(n) - \bar{U}(n-1) = \frac{c}{\mu}\left(\sigma^{n+1} - 1\right) < 0$ and $\widetilde{U}(n) - \widetilde{U}(n-1) = -\sigma^{n-n_e}(1 - \sigma)\left(V - \frac{n_e c}{\mu} + \frac{(1-\sigma^{n_e})c}{\beta}\right) < 0$. Further, $\widetilde{U}(n_e) - \bar{U}(n_e - 1) = -(1-\sigma)\left(V - \frac{n_e c}{\mu} + \frac{(1-\sigma^{n_e})c}{\beta}\right) < 0$. Hence, $U^{OB}(n)$ is decreasing in n. We then have $U^{OB}(n) < 0$ for all $n \geq n_e^*$. $\square$

THEOREM EC.1. *When queue-length information is shared remotely and $n_e = 1$, the hybrid order-ahead scheme has higher throughput than the pure order-onsite scheme, i.e., $TH_A^o \geq TH_S^o$.*

Proof of Theorem EC.1 In the pure order-onsite system, given that remote customers observe $n \leq n_e$ orders in the onsite queue, we will show that the maximum onsite queue length under which a remote customer is willing to travel with a positive probability must be less than n_e^*. We draw on the results of the auxiliary OAOB model given in Lemma EC.2.5. Consider a tagged remote customer who observes n outstanding orders upon arrival, let N_n^S denote her updated queue position (including herself) if she places an onsite order upon arrival at the service facility. We then have $U_S(n) = \mathbb{E}\left[V - \frac{cN_n^S}{\mu}\right]^+ - \frac{c}{\beta} \leq \mathbb{E}\left[V - \frac{cN_n^{OB}}{\mu}\right]^+ - \frac{c}{\beta} = U^{OB}(n)$, where N_n^{OB} and $U^{OB}(n)$ are the updated queue position and expected utility function defined in the OAOB model, and the inequality holds because $N_n^S \geq_{st} N_n^{OB}$. Because remote customers in the OAOB model will not place an order when $n \geq n_e^*$ (as shown in Lemma EC.2.5), neither will the aforementioned tagged customer in the present order-onsite model when $n \geq n_e^*$.

Now consider the case $n_e = 1$. Since we have proved $n_e^* \leq n_e$ in Proposition 2. Also, we have that $n_e^* \geq 1$ according to Assumption 1. Hence, $n_e^* = 1$ in this case. The throughput in the hybrid order-ahead system is $TH_A^o = \Lambda\mu/(\Lambda+\mu)$. We consider the following two cases for the pure order-onsite system: (1) No remote customers travel in the order-onsite system. Then clearly, the order-ahead throughput is higher. (2) Remote customers in the order-onsite system travel with a positive probability $p > 0$ if and only if they see an empty onsite queue. Let $\lambda_L = \Lambda(1-\gamma)$ be the arrival rate of local customers and $\lambda_R = \Lambda\gamma p$ be the travel rate of remote customers, and $\lambda_L + \lambda_R \leq \Lambda$. Let $\pi_{k,i}$ be the steady-state probability of state (k,i), where $k \in \{0,1\}$ is the number of customers in the onsite queue and $i \in \{0,1,...\}$ is the number of traveling customers. The balance equations are

$$(i\beta + \mu)\pi_{1,i} = \lambda_L\pi_{0,i} + (i+1)\beta\pi_{0,i+1} + (i+1)\beta\pi_{1,i+1}, \quad i = 0,1,... \tag{EC.1}$$

$$(\lambda_R + \lambda_L + i\beta)\pi_{0,i} = \mu\pi_{1,i} + \lambda_R\pi_{0,i-1}, \quad i = 1,2,... \tag{EC.2}$$

$$(\lambda_R + \lambda_L)\pi_{0,0} = \mu\pi_{1,0}. \tag{EC.3}$$

We prove the following: $\lambda_R\pi_{0,i} = (i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1})$, $i = 0,1,...$ We first show that it holds for $i = 0$. Equation (EC.1) gives $\mu\pi_{1,0} = \lambda_L\pi_{0,i} + \beta\pi_{0,1} + \beta\pi_{1,1}$. Combining this with (EC.3) gives $(\lambda_R + \lambda_L)\pi_{0,0} = \lambda_L\pi_{0,i} + \beta\pi_{0,1} + \beta\pi_{1,1}$. Hence $\lambda_R\pi_{0,0} = \beta(\pi_{0,1} + \pi_{1,1})$. Next, we prove that $\lambda_R\pi_{0,i} = (i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1})$, $i = 0,1,...$ holds for $i \geq 1$.

Equation (EC.1) gives $\mu\pi_{1,i} - \lambda_L\pi_{0,i} = (i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1}) - i\beta\pi_{1,i}$ for $i \geq 0$. Equation (EC.2) gives $\mu\pi_{1,i} - \lambda_L\pi_{0,i} = (\lambda_R + i\beta)\pi_{0,i} - \lambda_R\pi_{0,i-1}$ for $i \geq 1$. Hence, for $i \geq 1$, $(i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1}) - i\beta\pi_{1,i} = (\lambda_R + i\beta)\pi_{0,i} - \lambda_R\pi_{0,i-1}$, $(i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1}) - i\beta(\pi_{1,i} + \pi_{0,i}) = \lambda_R(\pi_{0,i} - \pi_{0,i-1})$, $\sum_{i=1}^{j}[(i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1}) - i\beta(\pi_{1,i} + \pi_{0,i})] = \sum_{i=1}^{j}\lambda_R(\pi_{0,i} - \pi_{0,i-1})$, $(j+1)\beta(\pi_{0,j+1} + \pi_{1,j+1}) - \beta(\pi_{1,1} + \pi_{0,1}) = \lambda_R(\pi_{0,j} - \pi_{0,0})$. Since we have proven $\lambda_R\pi_{0,0} = \beta(\pi_{0,1} + \pi_{1,1})$, it follows that $\lambda_R\pi_{0,i} = (i+1)\beta(\pi_{0,i+1} + \pi_{1,i+1})$ for $i \geq 0$. Hence, $i\beta\pi_{0,i} \leq \lambda_R\pi_{0,i-1}$ for $i \geq 1$. Combining this inequality with (EC.2) shows that $(\lambda_R + \lambda_L)\pi_{0,i} \geq \mu\pi_{1,i}$. Thus, $\mu\sum_{i=0}^{\infty}\pi_{1,i} \leq (\lambda_L + \lambda_R)\sum_{i=0}^{\infty}\pi_{0,i}$. Since $\sum_{i=0}^{\infty}\pi_{1,i} + \sum_{i=0}^{\infty}\pi_{0,i} = 1$, the throughput $TH_S^o = \mu\sum_{i=0}^{\infty}\pi_{1,i} \leq \mu(\lambda_L + \lambda_R)/(\lambda_L + \lambda_R + \mu)$. Since $\lambda_R + \lambda_L \leq \Lambda$ and $TH_A^o = \Lambda\mu/(\Lambda+\mu)$, it follows that $TH_S^o \leq TH_A^o$. $\square$

Proof of Theorem 4 When we compare the order-ahead-with-cancellation (OAC) model and the plain order-ahead model, we note that given any fixed q, the birth rates of the two models are equal and the death rate of the OAC model is higher ($\mu_i > \mu$ when $i > n_e$). Hence, invoking Lemma EC.2.2 gives $\pi^u_{0,C}(q) > \pi^u_0(q)$ and then $\pi^u_{i,C}(q) = \rho^i_T \pi^u_{0,C}(q) > \pi^u_i(q) = \rho^i_T \pi^u_0(q)$ for $i = 0, 1, \cdots, n_e$. There must exist some $n' > n_e$ such that $\pi^u_{n',C}(q) < \pi^u_{n'}(q)$. Otherwise, we cannot have $\sum_{n=0}^{\infty} \pi^u_{n,C}(q) = \sum_{n=0}^{\infty} \pi^u_n(q) = 1$. We then claim that for any $\hat{n} \geq n_e$, if $\pi^u_{\hat{n},C}(q) > \pi^u_{\hat{n}}(q)$ and $\pi^u_{\hat{n}+1,C}(q) \leq \pi^u_{\hat{n}+1}(q)$, then $\pi^u_{n,C}(q) < \pi^u_n(q)$ for all $n > \hat{n}+1$. To show this claim, note that $\pi^u_{\hat{n}+1,C}(q) = \pi^u_{\hat{n},C}(q)\frac{\gamma\Lambda q}{\mu+(\hat{n}-n_e)\beta}$, $\pi^u_{\hat{n}+1}(q) = \pi^u_{\hat{n}}(q)\frac{\gamma\Lambda q}{\mu}$. Hence, $\frac{\gamma\Lambda q}{\mu+(\hat{n}-n_e)\beta} < \frac{\gamma\Lambda q}{\mu}$. This implies $\frac{\gamma\Lambda q}{\mu+(n-n_e)\beta} < \frac{\gamma\Lambda q}{\mu}$ for any $n \geq \hat{n}+1$, which further implies $\pi^u_{n,C}(q) < \pi^u_n(q)$ for $n > \hat{n}+1$. And this satisfies Part (ii) of Lemma EC.2.1, where $\boldsymbol{\pi}^C$ and $\boldsymbol{\pi}^A$ cross each other only once: there exists an $\hat{n}$ such that $\pi^u_{n,C} \geq \pi^u_n$ when $n \leq \hat{n}$ and $\pi^u_{n,C} < \pi^u_n$ when $n > \hat{n}$. In addition, $\bar{U}_C(n) = \bar{U}(n)$, and $\widetilde{U}_C(n) > \bar{U}(n)$ and thus $U_C(n) \geq \bar{U}(n)$ for all n. Both $\bar{U}(n)$ and $U_C(n)$ are decreasing in n. Hence, $U^u_C(q) = \sum_{n=0}^{\infty} U_C(n)\pi^u_{n,C}(q) \geq \sum_{n=0}^{\infty} \bar{U}(n)\pi^u_{n,C}(q) > \sum_{n=0}^{\infty} \bar{U}(n)\pi^u_n(q) = U^u(q)$, and the second inequality holds by Lemma EC.2.1. The above ranking of the utility functions implies the ranking of the solutions for $U^u_C(q) = 0$ and $U^u(q)$ due to the properties established for the two utility functions in Lemmas EC.2.3 and EC.2.4. That is, $\underline{\lambda}^u_A < \underline{\lambda}^u_C$ and $\bar{\lambda}^u_A < \bar{\lambda}^u_C$, hence the equilibrium joining probabilities exhibit $q^u_A \leq q^u_C$. Note that in the plain order-ahead model, $q^u_A = 1$ for $\Lambda \leq \underline{\lambda}^u_A$. In the OAC model, $q^u_C = 1$ for $\Lambda \leq \underline{\lambda}^u_C$. The birth-rate of the two systems are equal while the death-rate in the OAC model is larger, which implies (based on Lemma EC.2.2) that the plain order-ahead system is busier, and thus $TH^u_A > TH^u_C$ for $\Lambda < \min\{\underline{\lambda}^u_A, \underline{\lambda}^u_C\} = \underline{\lambda}^u_A$. It remains to show that for sufficiently small Λ, $TH^*_A = TH^u_A$ (which has already been proved in Proposition 3) and $TH^*_C = TH^u_C$, which can be similarly proved. $\square$

Proof of Proposition 5 The expected utility of placing a remote order in the order-ahead-with-rejection (OAR) model is $U^u_{R,N}(q) = \sum_{n=0}^{N-1} \bar{U}(n)\pi^u_{n,R}(q)$. If $\bar{U}(N-1) \geq 0$, because $\bar{U}(n)$ decreases in n, we have $U^u_{R,N}(q) \geq 0$ for all $q \in [0,1]$, so that the equilibrium order-placing probability is $q^u_{R,N} = 1$. Next, we consider the case $\bar{U}(N-1) < 0$. Define $\tilde{U}^u_{R,N}(q) \equiv \frac{U^u_{R,N}(q)}{\sum_{j=0}^{N-1} \pi^u_{j,R}(q)} = \sum_{n=0}^{N-1} \bar{U}(n)\frac{\pi^u_{n,R}(q)}{\sum_{j=0}^{N-1} \pi^u_{j,R}(q)} \equiv \sum_{n=0}^{N-1} \bar{U}(n) f^u_{n,R}(q)$. We first consider $q = 1$ (the case $q = 0$ is similar). To establish that there exists a unique $\rho > 0$ such that $U^u_{R,N}(q) = 0$, it suffices to show that there exists a unique $\rho > 0$ such that $\tilde{U}^u_{R,N}(q) = 0$ because the latter is the former scaled by a positive term $\sum_{j=0}^{N-1} \pi^u_{j,R}$.

Lemma EC.2.6 **(Property of $\tilde{U}^u_{R,N}$ function)**. *The function $\tilde{U}^u_{R,N}$ exhibits the following properties: (i) $\tilde{U}^u_{R,N}(q)$ is continuous and strictly decreasing in q for a fixed ρ. (ii) $\tilde{U}^u_{R,N}(1)$ and $\tilde{U}^u_{R,N}(0)$ are continuous and strictly decreasing in ρ. (iii) When $\rho = 0$, $\tilde{U}^u_{R,N}(0) = \tilde{U}^u_{R,N}(1) = \bar{U}(0) > 0$; when $\rho \to \infty$, $\tilde{U}^u_{R,N}(0) < 0$ and $\tilde{U}^u_{R,N}(1) < 0$.*

Proof of Lemma EC.2.6 To prove Part (i), we consider two cases: (1) If $N > n_e$. Define $\bar{f}^u_{i,R}(\rho) \equiv f^u_{i,R}(q) = \begin{cases} \frac{\rho_T^i}{\sum_{j=0}^{n_e-1} \rho_T^j + \sum_{j=n_e}^{N-1} \rho_R^{j-n_e} \rho_T^{n_e}}, & \text{if} \quad i < n_e, \\ \frac{\rho_R^{i-n_e} \rho_T^{n_e}}{\sum_{j=0}^{n_e-1} \rho_T^j + \sum_{j=n_e}^{N-1} \rho_R^{j-n_e} \rho_T^{n_e}}, & \text{if} \quad i = n_e, \cdots, N-1, \end{cases}$ where $\rho_R = \gamma\rho q$ and $\rho_T = \gamma\rho q + (1-\gamma)\rho$. For two traffic intensities $\rho_1 < \rho_2$, we have $\bar{f}^u_{0,R}(\rho_2) < \bar{f}^u_{0,R}(\rho_1)$. We find that $\bar{f}^u_{i,R}(\rho_2)/\bar{f}^u_{i,R}(\rho_1)$ increases in i since $\gamma\rho_1 q < \gamma\rho_2 q$ and $\gamma\rho_1 q + (1-\gamma)\rho_1 < \gamma\rho_2 q + (1-\gamma)\rho_2$. This satisfies condition (ii) of Lemma EC.2.1. Because $\bar{U}(n)$ decreases in n, we conclude that $\tilde{U}^u_{R,N}(q)$ decreases in ρ. Next, when the traffic intensity ρ goes to infinity, it is evident that $\lim_{\rho\to\infty} \bar{f}^u_{N-1,R}(\rho) = 1$ when $q \in (0,1]$, and $\lim_{\rho\to\infty} \bar{f}^u_{n_e,R}(\rho) = 1$ when $q = 0$. Hence, the joining utility of a tagged remote customer approaches $\bar{U}(N-1))$ when $q > 0$ and $\bar{U}(n_e)$ when $q = 0$; in both cases, it converges to a negative value. (2) If $N \leq n_e$, from equation (7) we have $\bar{f}^u_{i,R}(\rho) \equiv f^u_{i,R}(q) = \frac{\rho_T^i}{\sum_{j=0}^{N-1} \rho_T^j}, \quad i = 0, \cdots, N-1$. For two traffic intensities $\rho_1 < \rho_2$, we have $\bar{f}^u_{0,R}(\rho_2) < \bar{f}^u_{0,R}(\rho_1)$. We find that $\bar{f}^u_{i,R}(\rho_2)/\bar{f}^u_{i,R}(\rho_1)$ increases in i for $i = 0, 1, \cdots, N-1$ since $\gamma\rho_1 q + (1-\gamma)\rho_1 < \gamma\rho_2 q + (1-\gamma)\rho_2$. Hence, $\tilde{U}^u_{R,N}(q)$ decreases in q for a fixed ρ. When $\rho \to \infty$, the joining utility of a tagged remote customer is $\tilde{U}^u_{R,N}(q)$ approaches $\bar{U}(N-1) < 0$ for all $q \in [0,1]$ because $\lim_{\rho\to\infty} \bar{f}^u_{N-1,R}(\rho) = 1$. The proof of Part (ii) is similar to Part (i). To prove Part (iii), $\tilde{U}^u_{R,N}(0) = \tilde{U}^u_{R,N}(1) = \bar{U}(0) > 0$ by Assumption 1. Combined with Part (ii), we have $\tilde{U}^u_{R,N}(0) < 0$ and $\tilde{U}^u_{R,N}(1) < 0$ when $\rho \to \infty$. $\square$

Finishing the proof of Proposition 5. Similar to the proof of Proposition 1, and by the property of the utility function $\tilde{U}^u_{R,N}$, both $U^u_{R,N}(0)=0$ and $U^u_{R,N}(1)=0$ have a unique solution, denoted by $\bar{\rho}^u_{R,N}, \underline{\rho}^u_{R,N}$, respectively. Furthermore, let $\bar{\lambda}^u_{R,N}=\mu\bar{\rho}^u_{R,N}$, $\underline{\lambda}^u_{R,N}=\mu\underline{\rho}^u_{R,N}$. □

Proof of Theorem 5 First, if the service provider uses a rejection threshold $N\le n_e^*$, then all remote customers have nonnegative utilities so that their order-placing probability is $q=1$. In addition, the system throughput under rejection threshold $N<n_e^*$ is lower than that under threshold n_e^*, because the former model rejects customers who observe $i, i=n_e^*,\cdots,N-1$ outstanding orders, who are supposed to place an order in the latter model. Therefore, the optimal rejection threshold N^* must satisfy $N^*\ge n_e^*$. Hence, it suffices to focus on the case $N\ge n_e^*$ below.

Next, we prove that $N^*=n_e^*$ for sufficiently large Λ. For any rejection threshold $N>n_e^*$, $\bar{U}(N-1)<0$, and Proposition 5 shows that the order-placing probability $q=0$ for sufficiently large Λ. By contrast, $q=1$ for $N=n_e^*$. Hence, by Lemma EC.2.2, the throughput under rejection threshold n_e^* is higher than that under any rejection threshold $N>n_e^*$ for sufficiently large Λ.

Next, we prove that $N^*=\infty$ for sufficiently small Λ. Let $\rho=\Lambda/\mu$, $\rho_T=(\gamma\Lambda q+(1-\gamma)\Lambda)/\mu$, $\rho_R=\gamma\Lambda q/\mu$, $\rho_L=(1-\gamma)\Lambda/\mu$. Define steady-state probabilities: $\pi^u_{i,R}(q)=\begin{cases}\rho_T^{i\wedge n_e}\rho_R^{(i-n_e)^+}\pi^u_{0,R}(q), & \text{if } N>n_e,\\ \rho_T^{i\wedge N}\rho_L^{(i-N)^+}\pi^u_{0,R}(q), & \text{if } N\le n_e,\end{cases}$, $i=0,1,\ldots,N\vee n_e$, where $\pi^u_{0,R}(q)=\begin{cases}\left(\frac{1-\rho_T^{n_e}}{1-\rho_T}+\frac{\rho_T^{n_e}\left(1-\rho_R^{N-n_e+1}\right)}{1-\rho_R}\right)^{-1}, & \text{if } N>n_e,\\ \left(\frac{1-\rho_T^{N}}{1-\rho_T}+\frac{\rho_T^{N}\left(1-\rho_L^{n_e-N+1}\right)}{1-\rho_L}\right)^{-1}, & \text{if } N\le n_e.\end{cases}$ Let $\pi^u_{n,R}(q;N)$ be the steady-state probability of n orders in the OAR model with rejection threshold N, where remote customers place orders with probability q. Also, let $U^u_R(q;N)(q;N)$ denote the joining utility of a remote customer in the OAR model with rejection threshold N, where remote customers place orders with probability q. **(1)** We first prove that the expected utility when all customers join $U^u_R(1;N)=\sum_{n=0}^{N-1}\bar{U}(n)\pi^u_{n,R}(1;N)$ is decreasing in rejection threshold N for $N\ge n_e^*$. To prove this claim, recognize that $\pi^u_{n,R}(1;N)>\pi^u_{n,R}(1;N+1)$ for $n=0,1,\ldots,N$. Hence, distribution $\{\pi^u_{n,R}(1;N+1)\}$ stochastically dominates distribution $\{\pi^u_{n,R}(1;N)\}$. That is, we can stochastically rank the steady-state queue length $Q(N)$ and $Q(N+1)$ under the two thresholds N and $N+1$ as: $Q(N)\le_{\text{st}}Q(N+1)$. Define the function $f(x)\equiv\bar{U}(x)\mathbf{1}_{\{x\le N\}}$, we have $U^u_R(1;N)=\sum_{n=0}^{N-1}\bar{U}(n)\pi^u_{n,R}(1;N)>\sum_{n=0}^{N}\bar{U}(n)\pi^u_{n,R}(1;N)=\mathbb{E}[f(Q(N))]\ge\mathbb{E}[f(Q(N+1))]=\sum_{n=0}^{N}\bar{U}(n)\pi^u_{n,R}(1;N+1)=U^u_R(1;N+1)$, where the first inequality holds because $\bar{U}(n)<0$ for $n\ge n_e^*$, and the second inequality holds because the function $f(x)$ decreases in x. **(2)** Given that all customers join, the throughput $\mu(1-\pi^u_{0,R}(1))$ is increasing in rejection threshold N due to a larger birth-rate (see Lemma EC.2.2). **(3)** It follows from (1) and (2) that if $U^u_R(1;\infty)\ge 0$, then the optimal rejection threshold is ∞. Further $U^u_R(1;\infty)$ is decreasing in $\Lambda\in(0,\mu)$ because distribution $\{\pi^u_{n,R}(1;\infty,\Lambda)\}$ stochastically increases with Λ. Hence, $\exists\underline{\Lambda}$ such that if $\Lambda\le\underline{\Lambda}$, the optimal rejection threshold is ∞. Note that when $N=\infty$, this model reduces to the plain order-ahead model, hence $\underline{\Lambda}=\underline{\lambda}^u_A$. □

Proof of Theorem 6 When queue-length information is shared remotely, we define N_o^* to be the optimal rejection threshold in the rejection model. First, we consider the case where $N_o^*\ge n_e^*$. In this case, the joining behavior of remote customers coincides with the behavior of the plain order-ahead model when queue-length information is shared remotely (balk with threshold n_e^*), that is $TH^o_{R,N_o^*}=TH^o_A$. We next consider the case where $N_o^*<n_e^*$. In this case, customers will always join since $\bar{U}(N_o^*)>0$, which further implies that the joining probability of remote customers is $q^o_R=1$. Compared to the case where $N_o^*=n_e^*$, in which all remote customers place orders, the birth rate in the former case is smaller. Lemma EC.2.2 implies that the latter system is busier, which consequently results in a higher system throughput, i.e., $TH^o_{R,N_o^*}<TH^o_A$. This implies that system throughput in the OAR model when queue-length information is shared remotely will not exceed that of the plain order-ahead model when queue-length information is shared remotely, that is, $TH^o_R\le TH^o_A$. Recall from Theorem 5 that when the queue-length information is not shared

remotely, the optimal rejection threshold satisfies $N^* \geq n_e^*$, thus resulting in a higher system throughput than that of the plain order-ahead model when queue-length information is shared remotely. In summary, $TH_R^u \geq TH_R^o$. $\square$

Proof of Proposition 6 From Theorem 5, we have proved the optimal rejection threshold $N^* \geq n_e^*$. Consider a tagged rejected remote customer. The number of outstanding orders satisfies $n \geq n_e^*$. Let N_n^S denote her updated queue position (including herself) upon arrival at the service facility if she travels to the service facility. Let $U_S(n)$ be her expected utility if she travels, and we then have $U_S(n) = \mathbb{E}\left[V - \frac{cN_n^S}{\mu}\right]^+ - \frac{c}{\beta} \leq \mathbb{E}\left[V - \frac{cN_n^{OB}}{\mu}\right]^+ - \frac{c}{\beta} = U^{OB}(n) < 0$, where N_n^{OB} and $U^{OB}(n)$ are the updated queue size and utility function defined in the OAOB model (see proof of Lemma EC.2.5), the first inequality holds because $N_n^S \geq_{st} N_n^{OB}$, and the last inequality holds by Lemma EC.2.5. $\square$

Proof of Theorem 7 First, consider the case of queue-length information not being shared remotely. We suppose the rejection threshold $N = n_e$. Recall that the steady-state probability of the number of outstanding orders in OAR model is $\pi_{i,R}^u(q) = \frac{\rho_T^i}{\sum_{j=0}^{n_e} \rho_T^j}, i = 0, 1, \cdots, n_e$. and the steady-state probability of the number of outstanding orders in the order-onsite model is $\pi_{i,S}^u(q) = \frac{\rho_T^i}{\sum_{j=0}^{n_e} \rho_T^j} = \pi_{i,R}^u(q), i = 0, 1, \cdots, n_e$. We next compare the utilities of a tagged remote customer who decides to join the order-onsite model and the OAR model. We have $U_{R,N}^u(q) = \sum_{i=0}^{n_e-1} \bar{U}(i)\pi_{i,R}^u(q) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu}\right)\pi_{i,R}^u(q) - \frac{c}{\beta}\sum_{i=0}^{n_e-1}\sigma^i \pi_{i,R}^u(q) > \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu}\right)\pi_{i,S}^u(q) - \frac{c}{\beta} = U_S^u(q)$, for any given q, where the strict inequality holds because $\sigma < 1$ and $\pi_{0,R}^u(q) < 1$ under a rejection threshold n_e. We next use the normalized utility $\tilde{U}_{R,N}^u(q)$. According to the above inequality, we have $\tilde{U}_{R,N}^u(1) \geq U_{R,N}^u(1) > U_S^u(1)$, whenever $\tilde{U}_{R,N}^u(1) \geq 0$ because the normalization factor of $\tilde{U}_{R,N}^u(1)$ is $\sum_{j=0}^{N-1} \pi_{j,R}^u(1) \in (0,1)$. Also, since $\tilde{U}_{R,N}^u(q)$ decreases in q, the two solutions of $\tilde{U}_{R,N}^u(1) = 0$ and $U_S^u(1) = 0$ must satisfy $\underline{\lambda}_{R,N}^u > \underline{\lambda}_S^u$. Similarly, we have $\tilde{U}_{R,N}^u(0) \geq U_{R,N}^u(0) > U_S^u(0)$, whenever $\tilde{U}_{R,N}^u(0) \geq 0$, so that the two solutions of $\tilde{U}_{R,N}^u(0) = 0$ and $U_S^u(0) = 0$ satisfy that $\bar{\lambda}_{R,N}^u > \bar{\lambda}_S^u$. By Proposition 5 and the order-onsite model, the equilibrium order-placing probability in the OAR model and travel probability in the order-onsite model must satisfy the ordering $q_{R,N}^u \geq q_S^u$. Finally, the steady-state probabilities of $\pi_{i,R}^u(q)$ and $\pi_{i,S}^u(q)$ imply that $\pi_{0,R}^u \leq \pi_{0,S}^u$, showing that the OAR model achieves higher throughput under rejection threshold $N = n_e$. The OAR model under the optimal rejection threshold only achieves even higher throughput and thus, $TH_R^u \geq TH_S^u$.

Next, consider the case of queue-length information being shared remotely. We know that $TH_R^o = TH_A^o$. We further know from Theorem EC.1 that $TH_A^o \geq TH_S^o$ for $n_e = 1$. Hence, $TH_R^o \geq TH_S^o$ for $n_e = 1$. $\square$

Proof of Theorem 8 Recall from Theorem 6 that the service provider has no incentive to share queue-length information in the OAR model. Moreover, when the queue-length information is shared in the OAC model, customers behave the same as they would in the hybrid order-ahead model, which represents a special case of the OAR model with an optimal rejection threshold. Hence, under optimal information, when it is optimal for the OAC model to share information ($TH_C^* = TH_C^o$), we always have $TH_R^* = TH_R^u \geq TH_A^o = TH_C^o$; when it is optimal for the OAC model not to share information ($TH_C^* = TH_C^u$), since $TH_R^* = TH_R^u$, it suffices to prove that $TH_R^u \geq TH_C^u$ when the market size is sufficiently small or large. The rest of the proof focuses on the case where queue-length information is not shared remotely. First, according to Theorem 4, we know that the throughput of the plain order-ahead model dominates that of the OAC model when the market size is small (i.e., when $\Lambda \leq \underline{\lambda}_A^u$). In addition, under the optimal rejection threshold, the OAR model yields higher throughput than the plain order-ahead model. Hence, the OAR model dominates the OAC model when the market size is sufficiently small. When the market size is large, we consider the OAR model with rejection threshold $N = n_e$ and show that this model already yields higher throughput than the OAC model. **Case 1:** If $\bar{U}(n_e - 1) \geq 0$, the joining probability of remote customers is $q_R^u = 1$ in the OAR model (Proposition 5). On the other hand, in the OAC model, if $\Lambda \geq \bar{\lambda}_C^u$, the joining probability of remote customers is $q_C^u = 0$, so that the OAR model is stochastically more congested than the OAC model (to see this, we again invoke Lemma EC.2.2). Hence, OAR yields higher throughput than OAC under a large market size ($\Lambda \geq \bar{\lambda}_C^u$). **Case 2:** Suppose $\bar{U}(n_e-1)<0$. Assume that the order-placing probability of remote customers in the OAC model is 0, so that the joining utility for a remote customer is $U_C^u(0)=\sum_{n=0}^{n_e-1}\left(\bar{U}(n)+\frac{c}{\beta}\right)\pi_{n,C}^u(0)+\left(\widetilde{U}(n_e)+\frac{c}{\beta}\right)\pi_{n_e,C}^u(0)-\frac{c}{\beta}$, where the corresponding steady-state probabilities are given by: $\pi_{i,C}^u(0)=\frac{\rho_T^i(1-\rho_T)}{1-\rho_T^{n_e+1}}, \quad i=0,1,\cdots,n_e$, where $\rho_T=(1-\gamma)\rho$. On the other hand, the utility of a remote customer in the OAR model is $U_{R,n_e}^u(0)=\sum_{n=0}^{n_e-1}\bar{U}(n)\pi_{n,R}^u(0)$, where the steady-state probability of the number of outstanding orders being i is $\pi_{i,R}^u(0)=\frac{\rho_T^i(1-\rho_T)}{1-\rho_T^{n_e+1}}=\pi_{i,C}^u(0), \quad i=0,\cdots,n_e$. A straightforward comparison of the above the two utility functions reveals $U_{R,n_e}^u(0)-U_C^u(0)=\pi_{n_e,C}^u(0)\frac{c}{\beta}-\left(\widetilde{U}(n_e)+\frac{c}{\beta}\right)\pi_{n_e,C}^u(0)=-\pi_{n_e,C}^u(0)\widetilde{U}(n_e)>0$, where the inequality holds because $\widetilde{U}(n_e)<\bar{U}(n_e-1)<0$. Consequently, we must have that $\bar{\lambda}_C^u<\bar{\lambda}_R^u$. When the market size $\Lambda\in(\bar{\lambda}_C^u,\bar{\lambda}_R^u)$, the OAR model is stochastically more congested than the OAC model (Lemma EC.2.2), so the former yields higher throughput than the latter. When the market size $\Lambda\geq\bar{\lambda}_R^u$, the order-placing probabilities under the two models are $q_C^u=q_R^u=0$, which yields identical system throughput. Hence, the OAR model under the rejection threshold n_e already dominates the OAC model by giving higher throughput when the marker size $\Lambda>\bar{\lambda}_C^u$. We conclude that the OAR model has higher throughput than the OAC model in this case. □

Proof of Theorem 9 We first characterize the queueing system for a given rejection threshold N_1 and cancellation threshold N_2. Given remote customers' order-placing probability q, the number of outstanding orders i evolves according to a birth-death process with a state-dependent birth rate $\lambda_i(q)$ and death rate μ_i:

$$\lambda_i(q)=\gamma\Lambda q\cdot\mathbf{1}_{\{i\leq N_1-1\}}+(1-\gamma)\Lambda\cdot\mathbf{1}_{\{i\leq n_e-1\}} \quad\text{and}\quad \mu_i=\mu+\beta(i-N_2)^+,\quad i=0,1,\cdots, \tag{EC.4}$$

Given the birth and death rates, the steady-state probability of the number of the outstanding orders X being i, $\hat{\pi}_{i,C}^u(q)\equiv\mathbb{P}(X=i)$, satisfies the balance equations below:

$$[\gamma\Lambda q\cdot\mathbf{1}_{\{i\leq N_1-1\}}+(1-\gamma)\Lambda\cdot\mathbf{1}_{\{i\leq n_e-1\}}]\hat{\pi}_{i,C}^u(q)=[\mu+(i-N_2)^+\beta]\hat{\pi}_{i+1,C}^u(q),\quad i=0,1,\cdots,\max\{N_1,n_e\}.$$

Denote X_a as the steady-state number of outstanding orders when a remote customer's order is accepted. Thus, given q, $\mathbb{P}(X_a=i)=\mathbb{P}(X=i|X<N_1)=\hat{\pi}_{i,C}^u(q)/(1-\hat{\pi}_{N_1,C}^u(q))$. The expected utility of a remote customer who places an order is

$$\begin{aligned} U_{R,C}^u(q)&\equiv\mathbb{E}\left[\left(V-\frac{c\cdot N_{X_a}}{\mu}\right)\mathbf{1}_{\{N_{X_a}<N_2\}}\right]-\frac{c}{\beta} \\ &=\sum_{i=0}^{N_2-1}\bar{U}(i)\hat{\pi}_{i,C}^u(q)+\sum_{i=N_2}^{N_1-1}\left[\sum_{j=0}^{N_2}\left(V-\frac{cj}{\mu}\right)p_i^{R,C}(j)-\frac{c}{\beta}\right]\hat{\pi}_{i,C}^u(q), \end{aligned} \tag{EC.5}$$

where $p_i^{R,C}(j)$ represents the probability distribution of the updated queue position N_{X_a} in the rejection-cancellation system, which is the same as in Lemma 3, except that n_e is replaced by N_2.

We next derive the optimal (N_1,N_2) for a sufficiently small market size Λ. In the integrated scheme: the birth rate $\lambda_i(q)=\gamma\Lambda q\cdot\mathbf{1}_{\{i\leq N_1\}}+(1-\gamma)\Lambda\cdot\mathbf{1}_{\{i\leq n_e-1\}}$ is maximized when $N_1=\infty$ and $q=1$, and the death rate $\mu_i=\mu+\beta(i-N_2)^+$ is minimized when $N_2=\infty$. Hence, from Lemma EC.2.2, the throughput is indeed maximized by setting $N_1=N_2=\infty$ and $q=1$. On the other hand, According to Proposition 1, when Λ is sufficiently small, the hybrid order-ahead model (with $N_1=N_2=\infty$) induces the remote customers' order-placing probability $q_A^u=1$. Hence, when Λ is sufficiently small, $N_1=N_2=\infty$ achieves the maximum throughput.

We next derive the optimal (N_1,N_2) when the market size Λ is sufficiently large. Suppose that the service provider adopts a rejection threshold $N_1\leq n_e^*$, then cancellation will not kick in since the cancellation threshold satisfies $N_2\geq n_e$ and the queue position of each remote arrival at the service facility will never exceed n_e^*. Thus, according to Theorem 5, any rejection threshold $N_1<n_e^*$ is throughput-dominated by rejection threshold n_e^*.

Suppose that the service provider adopts a rejection threshold $N_1 > n_e^*$. We next show that if $N_1 > n_e^*$, then for any $N_2 \geq n_e$, remote customers' order-placing probability $q^u_{R,N_1} = 0$ for sufficiently large Λ. To see this, first, for a birth-death process with birth rates $\{\lambda_i, 0 \leq i \leq N_1 - 1\}$ and death rates $\{\mu_i, 1 \leq i \leq N_1\}$, we know its steady-state probability $\pi_i = \prod_{j=0}^{i-1} \rho_i / (\sum_{k=0}^{N_1} \prod_{l=0}^{k-1} \rho_l)$ for $i = 0, \ldots, N_1$, where $\bar{\rho}_i \equiv \lambda_i / \mu_{i+1}$. We prove by contradiction. Now suppose $q^u_{R,N_1} > 0$, we know from (EC.4) and (EC.5) that as $\Lambda \to \infty$, we have $\rho_i \equiv \lambda_{i-1}(q^u_{R,N_1})/\mu_i(q^u_{R,N_1}) \to \infty$, and the probability that a remote customer is accepted at state $N_1 - 1$ is

$$\frac{\hat{\pi}^u_{N_1-1,C}(q^u_{R,N_1})}{1 - \hat{\pi}^u_{N_1,C}(q^u_{R,N_1})} = \frac{\prod_{j=0}^{N_1-2} \rho_i / (\sum_{k=0}^{N_1} \prod_{l=0}^{k-1} \rho_l)}{1 - \prod_{j=0}^{N_1-1} \rho_i / (\sum_{k=0}^{N_1} \prod_{l=0}^{k-1} \rho_l)} = \frac{\prod_{j=0}^{N_1-2} \rho_i}{1 + \rho_0 + \rho_0\rho_1 + \cdots + \prod_{j=0}^{N_1-2} \rho_i} \to 1,$$

so that

$$\begin{aligned} U^u_{R,C}(q^u_{R,N_1}) &\to \mathbb{E}\left[\left(V - \frac{c \cdot N_{N_1-1}}{\mu}\right) \mathbf{1}_{\{N_{N_1-1} < N_2\}}\right] - \frac{c}{\beta} \\ &= \underbrace{\left(\mathbb{E}\left[\left(V - \frac{c \cdot N_{N_1-1}}{\mu}\right)\right] - \frac{c}{\beta}\right)}_{\bar{U}(N_1-1)} \cdot \mathbb{P}(N_{N_1-1} < N_2) + \left(-\frac{c}{\beta}\right) \cdot \mathbb{P}(N_{N_1-1} \geq N_2). \end{aligned}$$

Because $\bar{U}(N_1 - 1) \leq \bar{U}(n_e^*) < 0$, the above limit is strictly negative regardless of the cancellation threshold N_2. Thus, there must exist a sufficiently large $\bar{\Lambda}_{R,C}$ such that, for any $\Lambda > \bar{\Lambda}_{R,C}$, a remote customer's joining expected utility is negative if $q^u_{R,N_1} > 0$, which leads to a contradiction. Hence, $q^u_{R,N_1} = 0$ for sufficiently large Λ when $N_1 > n_e^*$.

Thus, for sufficiently large Λ, the model with $N_1 > n_e^*$ and the model with $N_1 = n_e^*$ have identical states $i = 0, 1, \cdots, n_e$ and death rates, while the former has smaller birth rates (i.e., $(1-\gamma)\Lambda$) than those of the latter (Λ) at states $i = 0, 1, \cdots, n_e - 1$, thus resulting in lower throughput (according to Lemma EC.2.2).

Based on the above analysis, when the market size is sufficiently large, $N_1 = n_e^*, N_2 = \infty$ is optimal. □

Proof of Proposition 7 We first establish Lemma EC.2.7, where $U_A(\boldsymbol{\lambda}, \beta)$ and $U_S(\boldsymbol{\lambda}, \beta)$ are defined in §EC.1.3.

LEMMA EC.2.7. *Given arrival rate $\boldsymbol{\lambda}$, the utility functions $U_A(\boldsymbol{\lambda}, \beta)$ and $U_S(\boldsymbol{\lambda}, \beta)$ are increasing in β. In addition, $\Delta U(\boldsymbol{\lambda}, \beta) \equiv U_A(\boldsymbol{\lambda}, \beta) - U_S(\boldsymbol{\lambda}, \beta)$ is decreasing in β, and the two utility functions have exactly one intersection.*

Proof of Lemma EC.2.7. Note that $\bar{U}(n)$ increases in β, and $\hat{\pi}^u_n(\boldsymbol{\lambda})$ is independent of β. It is straightforward to see that both $U_A(\boldsymbol{\lambda}, \beta)$ and $U_S(\boldsymbol{\lambda}, \beta)$ increase in β.

Let N be the steady-state queue length of an $M/M/1$ queue with birth rate $\lambda_i = \Lambda$ if $i < n_e$ and $\lambda_i = \lambda_A$ otherwise, and death rate $\mu_i = \mu$ for $i > 0$. The expected utility of a tagged customer with β who places a remote order is

$$U_A(\boldsymbol{\lambda}, \beta) = V - c\mathbb{E}[\max\{T(\beta), X_{N+1,\mu}\}],$$

where $T(\beta) \sim \text{Exp}(\beta)$ is the random travel time with speed β, and $X_{N+1,\mu}$ denotes an Erlang random variable with $(N+1)$ phases and rate μ. The expected utility of a tagged customer with β who places an onsite order is $U_S(\boldsymbol{\lambda}, \beta) = V\mathbb{P}(N < n_e) - c\mathbb{E}\left[T(\beta) + X_{N+1,\mu} \cdot \mathbf{1}_{\{N<n_e\}}\right]$. The difference between the two utility functions is

$$\triangle U(\boldsymbol{\lambda}, \beta) = V\mathbb{P}(N \geq n_e) - c\mathbb{E}[\max\{0, X_{N+1,\mu} - T(\beta)\} - X_{N+1,\mu} \cdot \mathbf{1}_{\{N<n_e\}}].$$

To show that $\triangle U(\boldsymbol{\lambda}, \beta)$ decreases in β, it suffices to show that the random variable $\max\{0, X_{N+1,\mu} - T(\beta)\}$ stochastically increases in β. To see this, let $\beta_1 \leq \beta_2$; it is straightforward to see that $T(\beta_1) \sim \text{Exp}(\beta_1) \geq_{st} \text{Exp}(\beta_2) \sim T(\beta_2)$.

Next, because $T(\beta) \sim \text{Exp}(\beta)$, we have $\lim_{\beta \to 0} \Delta U(\boldsymbol{\lambda}, \beta) = \infty$, and

$$\begin{aligned} \lim_{\beta \to \infty} \Delta U(\boldsymbol{\lambda}, \beta) &= V\mathbb{P}(N \geq n_e) - c\mathbb{E}[X_{N+1,\mu} - X_{N+1,\mu} \cdot \mathbf{1}_{\{N<n_e\}}] \\ &= V\mathbb{P}(N \geq n_e) - c\mathbb{E}[X_{N+1,\mu} \cdot \mathbf{1}_{\{N \geq n_e\}}] \\ &= \mathbb{E}[(V - c \cdot X_{N+1,\mu}) \cdot \mathbf{1}_{\{N \geq n_e\}}] < 0, \end{aligned}$$

where the inequality holds by the definition of n_e. Hence, the above analysis ensures that there exists only one intersection between utility functions $U_A(\boldsymbol{\lambda},\beta)$ and $U_S(\boldsymbol{\lambda},\beta)$. $\square$

Finishing the proof of Proposition 7. From Lemma EC.2.7, there exist thresholds β_1 and β_2 with $a \le \beta_1 \le \beta_2 \le b$ such that $0 > \max\{U_A(\boldsymbol{\lambda},\beta), U_S(\boldsymbol{\lambda},\beta)\}$ (and hence customers do not order) if $\beta < \beta_1$; $U_A(\boldsymbol{\lambda},\beta) > \max\{0, U_S(\boldsymbol{\lambda},\beta)\}$ (and hence customers order ahead) if $\beta \in (\beta_1,\beta_2)$; $U_S(\boldsymbol{\lambda},\beta) > \max\{0, U_A(\boldsymbol{\lambda},\beta)\}$ (and hence customers order onsite) if $\beta > \beta_2$. $\square$

Proof of Theorem 10 We prove the throughput dominance under rejection threshold n_e, which would imply throughput dominance under the optimal rejection threshold.

In the OAR model, let the arrival rate for remote customers who choose to order ahead and order onsite be λ_A and λ_S respectively, with $\lambda_A + \lambda_S \le \gamma\Lambda$; let the arrival rate for local customers be $\lambda_L = (1-\gamma)\Lambda$. The order queue evolves as a birth-death process with state-dependent birth rate $\lambda_i = (\lambda_A + \lambda_S + \lambda_L)\cdot \mathbf{1}_{\{i<n_e\}}$ and death rate $\mu_i = \mu$ for $i > 0$. The steady-state probability of the number of outstanding orders being $i \le n_e$ is $\hat{\pi}_i^u(\lambda_A+\lambda_S+\lambda_L) = (1-(\lambda_A+\lambda_S+\lambda_L)/\mu)((\lambda_A+\lambda_S+\lambda_L)/\mu)^i/(1-(\lambda_A+\lambda_S+\lambda_L)^{n_e+1})$. Let the expected utility for a remote customer who orders ahead be U_A^R and that for a remote customer who orders onsite be U_S^R:

$$U_A^R(\lambda_A,\lambda_S,\lambda_L,\beta) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu} - \frac{c}{\beta}\sigma^{i+1}\right)\hat{\pi}_i^u(\lambda_A+\lambda_S+\lambda_L),$$
$$U_S^R(\lambda_A,\lambda_S,\lambda_L,\beta) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu}\right)\hat{\pi}_i^u(\lambda_A+\lambda_S+\lambda_L) - \frac{c}{\beta},$$

where $\sigma = \mu/(\beta+\mu)$. It is straightforward that $U_A^R(\lambda_A,\lambda_S,\lambda_L,\beta) \ge U_S^R(\lambda_A,\lambda_S,\lambda_L,\beta)$. Thus, no remote customers place onsite orders ($\lambda_S = 0$) in equilibrium. Moreover, since $U_A^R(\lambda_A,\lambda_S,\lambda_L,\beta)$ increases in β, there exists a threshold for travel speed $\bar{\beta}_A$ that uniquely solves $U_A^R(\lambda_A,0,\lambda_L,\bar{\beta}_A) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu} - \frac{c}{\bar{\beta}_A}\sigma^{i+1}\right)\hat{\pi}_i^u(\lambda_A+\lambda_L) = 0$ with $\lambda_A = \gamma\Lambda(1-F(\bar{\beta}_A))$ such that customers place orders (ahead) if and only if $\beta > \bar{\beta}_A$.

In the order-onsite model, let the arrival rate for remote customers be $\lambda'_S \le \gamma\Lambda$ and the arrival rate for local customers be $\lambda_L = (1-\gamma)\Lambda$. The order queue state is a birth-death process with state-dependent birth rate $\lambda_i = (\lambda'_S + \lambda_L)\cdot\mathbf{1}_{\{i<n_e\}}$ and death rate $\mu_i = \mu$ for $i > 0$. The corresponding steady-state probability of the number of outstanding orders being $i \le n_e$ is $\hat{\pi}_i^u(\lambda'_S+\lambda_L) = (1-(\lambda'_S+\lambda_L)/\mu)((\lambda'_S+\lambda_L)/\mu)^i/(1-(\lambda'_S+\lambda_L)^{n_e+1})$. Thus, the expected utility of a joining remote customer with travel speed β is $U_S^S(\lambda'_S,\lambda_L,\beta) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu}\right)\hat{\pi}_i^u(\lambda'_S+\lambda_L) - \frac{c}{\beta}$. It is straightforward to see that $U_S^S(\lambda'_S,\lambda_L,\beta)$ increases in β. Hence, remote customers join when their travel speed β exceeds a threshold $\bar{\beta}_S$, which uniquely solves $U_S^S(\lambda'_S,\lambda_L,\bar{\beta}_S) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu}\right)\hat{\pi}_i^u(\lambda'_S+\lambda_L) - \frac{c}{\bar{\beta}_S} = 0$ with $\lambda'_S = \gamma\Lambda(1-F(\bar{\beta}_S))$.

Next, we prove $\lambda_A > \lambda'_S$, i.e., $\bar{\beta}_A < \bar{\beta}_S$. Note that $0 = U_A^R(\lambda_A,0,\lambda_L,\bar{\beta}_A) = \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu} - \frac{c}{\bar{\beta}_A}\sigma^{i+1}\right)\hat{\pi}_i^u(\lambda_A + \lambda_L) > \sum_{i=0}^{n_e-1}\left(V - \frac{(i+1)c}{\mu}\right)\hat{\pi}_i^u(\gamma\Lambda(1-F(\bar{\beta}_A))+\lambda_L) - \frac{c}{\bar{\beta}_A} = U_S^S(\gamma\Lambda(1-F(\bar{\beta}_A)),\lambda_L,\bar{\beta}_A)$. Since $U_S^S(\gamma\Lambda(1-F(\beta)),\lambda_L,\beta)$ is increasing in β and $U_S^S(\gamma\Lambda(1-F(\bar{\beta}_A)),\lambda_L,\bar{\beta}_A) < 0$ and $U_S^S(\gamma\Lambda(1-F(\bar{\beta}_S)),\lambda_L,\bar{\beta}_S) = 0$, we have $\bar{\beta}_A < \bar{\beta}_S$. Hence, $\lambda_A > \lambda'_S$, which implies that $\hat{\pi}_0(\lambda_A+\lambda_L) < \hat{\pi}_0(\lambda'_S+\lambda_L)$. Therefore, the throughput of the OAR model with rejection threshold n_e surpasses that of the order-onsite model. The throughput of the OAR model under the optimal rejection threshold will only be even higher. $\square$