Electronic Companion to “Structural Estimation of Intertemporal Externalities with Application in ICU Admissions”

Yiwen Shen
School of Business and Management, The Hong Kong University of Science and Technology, yiwenshen@ust.hk

Carri W. Chan
Decision, Risk, and Operations, Columbia Business School, cwchan@columbia.edu

Fanyin Zheng
Imperial Business School, Imperial College London, f.zheng@imperial.ac.uk

Gabriel Escobar
Division of Research, Kaiser Permanente, gabriel.escobar@kp.org

EC.1. Formulae and Modeling Assumptions

EC.1.1. Explicit expressions for state transition probability $g(s'|s)$

We provide explicit expressions for the function $g(s'|s)$ used in Proposition 1, which is the transition probability from the intermediate state $s = (\tilde{n}^G_{l,t}, \tilde{n}^G_{h,t}, \tilde{n}^F_t)$ (after action is taken) to the state $s' = (n^G_{l,t+1}, n^G_{h,t+1}, n^F_{t+1})$ at the start of the next period (before action). As the GK arrivals are independent of the FSU external arrivals and departures, we have

$$g(s'|s) = g_{Q,l}\left(n^G_{l,t+1}|\tilde{n}^G_{l,t}\right) g_{Q,h}\left(n^G_{h,t+1}|\tilde{n}^G_{h,t}\right) g_F\left(n^F_{t+1}|\tilde{n}^F_t\right), \tag{EC.1}$$

where $g_{Q,l}$ and $g_{Q,h}$ denote the transition probabilities for the numbers of GK customers from low and high classes respectively, and g_F denotes the transition probability for the number of FSU customers.

For the GK transition probabilities $g_{Q,l}$ and $g_{Q,h}$, we only need to consider the new arrivals for class $i \in \{l, h\}$, which follow truncated Poisson distributions with rate $\lambda_{Q,i}$ and truncation by $M_{A,i}$ from above. Additionally accounting for the GK capacity constraint, the number of GK arrivals is capped by $\bar{A}_i = \min\{M_{A,i}, Q_i - \tilde{n}^G_{i,t}\}$. Thus, the transition probability can be computed as

$$g_{Q,i}(m|n) = \begin{cases} (\lambda_{Q,i})^{m-n}\exp(-\lambda_{Q,i})/(m-n)! & \text{if } n \le m < n + \min\{M_{A,i}, Q_i - n\} \\ \sum_{j=\bar{A}_i}^{+\infty} (\lambda_{Q,i})^j \exp(-\lambda_{Q,i})/j! & \text{if } m = n + \min\{M_{A,i}, Q_i - n\} \end{cases} \tag{EC.2}$$

and $g_{Q,i}(m|n) = 0$ elsewhere.

For the FSU transition probability g_F, we need to consider both external arrivals and departures. The number of external arrivals E_t follows a Poisson distribution with rate λ_E, and is capped by the remaining

FSU capacity $B-\tilde{n}_t^F$. With E_t external arrivals in the period, the FSU would have total $\tilde{n}_t^F+E_t$ customers. Then, the number of departures, D_t, follows a Binomial-$(\tilde{n}_t^F+E_t,\mu_I)$ distribution. Given the number of external arrivals, E_t, the number of departures follows by $D_t=\tilde{n}_t^F+E_t-n_{t+1}^F$. We note that E_t can be greater than $\max\{n_{t+1}^F-\tilde{n}_t^F,0\}$. Summing up the probability of all possible choices of the Poisson-distributed E_t (and Binomial-distributed D_t accordingly), the transition probability $g_F(m|n)$ is given by

$$
\begin{aligned}
g_F(m|n) = &\sum_{j=\max\{m-n,0\}}^{B-n-1} \lambda_E^j \frac{\exp(-\lambda_E)}{j!}\frac{(n+j)!}{(n+j-m)!m!}\mu_I^{n+j-m}(1-\mu_I)^m \\
&+\left(\sum_{j=B-n}^{+\infty}\lambda_E^k\frac{\exp(-\lambda_E)}{j!}\right)\frac{B!}{(B-m)!m!}\mu_I^{B-m}(1-\mu_I)^m, \text{ for } 0\le m\le B. \qquad \text{(EC.3)}
\end{aligned}
$$

Using (EC.2) and (EC.3), we obtain the explicit expression for state transition probability $g(s'|s)$ by (EC.1).

EC.1.2. Assumptions for Identification

This section documents the assumptions in Komarova et al. (2018), which are also satisfied by our model. In their paper, x and a denote the system state and action, respectively; ε is the random perturbation in utility.

Assumption 1 *(i) (Additive Separability) For all a, x, ε, the per-period utility follows:*

$$u(a,x,\varepsilon)=\pi(a,x)+\varepsilon(a).$$

(ii) (Conditional Independence) The transition distribution of the states has the following factorization for all x', ε', x, ε, a:

$$P(x',\varepsilon'|x,\varepsilon,a)=Q(\varepsilon')G(x'|x,a),$$

where Q is the cumulative distribution function of ε and G is the transition law of x_{t+1} conditioning on x_t and a_t. Furthermore, ε_t has finite first moments, and a positive, continuous, and bounded density.

(iii) (Finite Observed State) $X=\{1,\ldots,K\}$.

Assumption 2 (Linear-in-Parameter) *For all a, x:*

$$\pi(a,x;\theta)=\pi_0(a,x)+\theta^\top\pi_1(a,x),$$

where π_0 is a known real value function, π_1 is a known p-dimensional vector value function and θ is the p-dimensional unknown parameter.

In our setting, we have action utility given by (2) as

$$u(s_t,d_t)=u_{a,l}a_{l,t}+u_{r,l}r_{l,t}+u_{w,l}\left(n_{l,t}^G-a_{l,t}-r_{l,t}\right)+u_{a,h}a_{h,t}+u_{r,h}r_{h,t}+u_{w,h}\left(n_{h,t}^G-a_{h,t}-r_{h,t}\right).$$

This is the deterministic part of the per-period utility. Thus, it is indeed linear in parameters $\theta=\{u_{a,l},u_{r,l},u_{w,l},u_{a,h},u_{r,h},u_{w,h}\}$. In our model, the functions π_0 and π_1 specify to $\pi_0(d_t,s_t)\equiv 0$ and

$$\pi_1(d_t,s_t)=[a_{l,t},r_{l,t},n_{l,t}^G-a_{l,t}-r_{l,t},a_{h,t},r_{h,t},n_{h,t}^G-a_{h,t}-r_{h,t}]^\top.$$

EC.2. Data Selection and Estimation of System Parameters for ICU Admission

EC.2.1. Data Selection Process

We start from a total of 312,306 hospitalizations over two years. We restrict our study to the hospitalizations admitted to a medical service via the ED, which comprises the largest proportion of admitted patients. For the ED patients, they appear in our data set as soon as the admission decision has been made; as such, we do not have information about patients discharged home from the ED nor patients for whom a disposition decision has not yet been made. We drop 12 hospitalizations with unknown gender and 9,128 (4.8%) hospitalizations for patients who experience hospital transfers or transports outside of the hospital network. Our study focuses on three possible decisions for each patient in each decision epoch: keep the patient waiting in the ED, admit the patient to the ICU, or admit the patient to a non-ICU unit (e.g. the ward or TCU). We drop 3,066 (1.7%) hospitalizations where the patient was admitted to other units – e.g., OR or PAR, from the ED. Finally, we drop 1,675 (1%) hospitalizations with ED waiting time longer than 12 hours as these episodes can be considered outliers (the average waiting time is shorter than two hours).

We restrict our study cohort to the periods of each hospital with stable ICU capacity and occupancy. First, we discard the first and last month of data for all hospitals. Second, for several hospitals, we drop the period at either end of the sample where the ICU occupancy dramatically fluctuates or significantly differs from the more stable period in the middle. Finally, for hospital 21, we find that its ICU capacity experienced a substantial increase during the sample period (from 13 to 16). As a result, we split it into two parts, i.e., before and after the capacity change, and treat them as two hospitals in the estimation. We refer to 22 hospitals in our study cohort. The patient characteristics of the final study cohort and the ICU admission subset are summarized in Table OS.5 of the Online Supplement. The number of days and hospitalizations for each hospital in the final study cohort are summarized in Table OS.6 of the Online Supplement. In total, we drop 11,268 (6.4%) hospitalizations that are outside the stable periods.

EC.2.2. Estimation of System Parameters

As described in Section 2.3, the arrival and departure rates, as well as the unit capacities, are estimated directly from data – outside of the structural model. We now describe how we do this for the ICU admission problem. Recall we define each period as a two hour interval in the model. We estimate the ED arrival rates $\lambda_{Q,i}$ and maximum arrival number $M_{A,i}$ for $i \in \{l, h\}$ using the average and maximum number of arrivals to the ED for the two classes in each period. We estimate the ICU external arrival rate λ_E using the average number of patients admitted to ICU in each time slot who are not included in our low and high severity ED groups. The departure probability μ_I is estimated as the ratio of total number of departures to the total periods of ICU stay across all ICU patients. The ICU capacity B is set to be the maximum number of *all* patients (medical and surgical, emergency and elective) in the ICU observed from data. This is a reasonable assumption as ICU often operates under high congestion. We discuss the validity of these estimates in Section EC.8.

Our data captures the number of patients admitted to the hospital from the ED, but does not include any patients who are discharged from the ED. Thus, it is difficult to accurately determine the ED capacity in our model. However, we note that the number of ED boarding patients is relatively low (e.g., average 1.21 patients). Thus, we set ED capacities Q_i using the following heuristic: $Q_i = M_{Q,i} + \lfloor \sqrt{M_{A,i}} \rfloor$, where $M_{Q,i}$ is the maximum number of ED patients observed in the data; $M_{A,i}$ is the maximum number of arrivals in each period; and $\lfloor \cdot \rfloor$ denotes the floor function.The square root term $\sqrt{M_{A,i}}$ is introduced as a heuristic "safety buffer" to ensure we have ample ED capacity to avoid balking upon arrival to the ED, which rarely happens in reality. We verify by simulation that the ED rarely reaches its full capacity Q_i in our structural model. In addition, the choice probabilities are very robust to alternative choices of the ED capacity. The estimated system statistics for each hospital are reported in Table OS.6 of Online Supplement.

EC.3. Reduced-form Evidence for Discounting Behavior in ICU Admissions

In this section, we conduct reduced-form regressions to analyze the main determinants of the system's ICU admission decisions. This provides complementary evidence for the discounting behaviors of hospitals and serves as a benchmark for evaluating the explanatory power of our structural model. We apply a multinomial logistic model to estimate the ICU admission decisions. In each two-hour period, the hospital chooses one of the three options for each patient: admit to the ICU; admit to non-ICU units; or make the patient wait in the ED. At the start of each period, we construct system "snapshots" which include detailed information for each patient in the ED as well as the state of the ED and ICU. We include patient characteristics, system state variables, and seasonality effects as potential determinants of the admission decisions.

We use the non-ICU admission decision as the base case, and estimate the probabilities of the ICU admission ($d =$ ICUAdm) and waiting ($d =$ Wait) decisions relative to the non-ICU admission decision respectively. For patient i who is in the ED at the start of period t, we set up the multinomial logistic model:

$$\ln\left[\frac{\Pr(d_{it}|\mathbf{X}_i,\mathbf{S}_t)}{\Pr(\text{nonICU}_{it}|\mathbf{X}_i,\mathbf{S}_t)}\right] = \gamma_{0,d} + \gamma_{L,d}\text{LAPS2}_i + \gamma_{ICU,d}\text{ICUOccu}_t + \gamma_{ED,d}\text{EDNum}_t + \boldsymbol{\gamma}_{Z,d}^{\top}\mathbf{Z}_{i,t} + \epsilon_{it}, \tag{EC.4}$$

where $\Pr(d_{it})$ and $\Pr(\text{nonICU}_{it}|\mathbf{X}_i\mathbf{S}_t)$ denote the conditional probability of action d_{it} and admitting patient i to non-ICU units in period t, respectively. In (EC.4), LAPS2 is the patient's main severity score. ICUOccu_t denotes the current ICU occupancy level. As the ICU sizes vary dramatically across the hospitals, we use the ICU percentile rank to measure occupancy. EDNum_t denotes the number of current ED patients. $\mathbf{Z}_{i,t}$ denotes other control covariates, given by

$$\mathbf{Z}_{i,t} = \{\text{Gender}_i, \text{Age}_i, \text{COPS2}_i, \text{CHMR}_i, \text{Hosp}_i, \text{DepPre}_t, \text{AvgLAPS2}_t, \text{DayOfWeek}_t, \text{HourOfDay}_t, \text{Month}_t\}.$$

It includes patient i's gender, age, identifier for hospital admitted, as well as other two severity scores COPS2 and CHMR. In addition, it includes several system state variables: the number of patients who left the ICU in the previous period (DepPre_t), the average severity level measured by the LAPS2 score of the current ICU patients in period t (AvgLAPS2_t), and several categorical variables to capture the potential seasonality and time trend in the decisions. Here DayOfWeek_t and HourOfDay_t denote the day of week and hour of day respectively; Month_t is the dummy variable representing the month in the sample (total 23 months). To account for the heteroskedasticity, we cluster standard errors by hospitals in the regression.

We first estimate the model by combining the patient data from all hospitals. We use the McFadden's pseudo R-squared to measure the goodness of fit of the model. It is defined as $R^2 = 1 - \ln l^{mod}/\ln l^{null}$, where l^{mod} is the likelihood from the estimated model and l^{null} is the likelihood from the "null" model that only includes the intercept and a categorical variable for each hospital in (EC.4).

Table EC.1 Estimation results for Multinomial-Logistic Regression (EC.4), N = 183,691, R-squared = 0.16

	LAPS2_i	ICUOccu_t	EDNum_t
	γ_L	γ_{ICU}	γ_{ED}
Waiting	0.008	1.178	0.241
	(0.000)	(0.031)	(0.006)
ICUAdm	0.028	−0.396	−0.014
	(0.000)	(0.030)	(0.006)

Standard error is reported in parenthesis.

Table EC.1 reports the estimated coefficients for three main variables in the multinomial logistic model: LAPS2_i, ICUOccu_t, and EDNum_t. All the coefficients are statistically significant and have the expected sign. In particular, higher severity score (LAPS2) increases the probability of admission to ICU relative to other units. In addition, the estimates of γ_{ICU} and γ_{ED} suggest that, even after controlling for patient characteristics and fixed effects, a busier system state (more congested ICU or ED) decreases the probability of ICU admission and increases the probability of waiting. Such evidence suggests that the hospitals indeed internalizes the intertemporal externalities on the ICU admission decisions by adjusting their behaviors according to the current system state. The McFadden's pseudo R^2 for the combined estimation is 0.16, consistent with the magnitude seen for models of operational decisions in healthcare systems.

Considering the heterogeneity across hospitals, we further estimate model (EC.4) for each hospital separately after dropping the hospital dummy Hosp_i. As a robustness check, we also estimate a multinomial logistic model with only three variables plus an intercept term: the percentile rank of ICU occupancy, number of current ED patients, and a dummy for whether the patient is in low or high severity class. This simplified model uses the same information as in our dynamic discrete choice model. Table OS.9 reports the estimated coefficient γ_{ICU} of ICU occupancy for the ICU admission decision ($d_{it} = \text{ICUAdm}$). The results for the full (resp. simplified) multinomial logistic model are given in the second (resp. third) columns respectively. We see that the results are qualitatively similar to that for all hospitals combined in Table EC.1. Specifically, higher ICU occupancy is still associated with decreased likelihood of ICU admission decisions for most hospitals, although some coefficients are not statistically significant due to a much smaller sample size. Moreover, the identified discount factors $\hat{\beta}$ and the coefficients of ICU occupancy γ_{ICU} are highly negatively correlated: the correlation is -0.768 and -0.774 for the full and simplified model respectively. This means that the dynamic discrete choice model and the multinomial logistic model provide consistent evidence for hospitals' discounting behaviors. In the full multinomial logistic model (EC.4), the McFad-dean's R^2 for individual hospital regressions ranges from 0.14 to 0.24, with an average of 0.17. This level is also similar to our dynamic discrete choice model.

EC.4. Individual Hospital Results

Table EC.2 provides the full estimate results by individual hospital for our main model.

EC.5. Discussion on Model Identification

Action utility and customer classes: As described in Section 2, the parameters $\{u_{w,\iota}, u_{a,\iota}, u_{r,\iota}\}$ represent the average action utility within each customer class. Then, the term $\varepsilon(d_t)$ in (3) captures the idiosyncratic utility shock related to specific customer or action. Such set-up is common in dynamic discrete choice models (see, e.g., Gordon 2009, Komarova et al. 2018).

Due to the curse of dimensionality, it is prohibitive in our model to let the action utility depend on each customer's characteristics (e.g., patient's severity). The curse of dimensionality is a common challenge for the estimation of dynamic discrete choice models, as one usually needs to solve the dynamic programming equations for every candidate parameter (Rust 1987). This issue is notably significant in our context due to the fact that all the customers use the same FSU with limited capacity. Thus, our state space must include the information of *all* customers in the GK. With L customer classes, the dimension of state space is

$$\text{Dim of state space} = \prod_{i=1}^{L}(Q_i+1)\times(B+1),$$

where Q_i is the GK capacity for ith customer class and B is the FSU capacity. Clearly, the dimension of state space increases exponentially fast with the number of customer classes. For these reasons, in our main setting, we include two customer classes to capture the heterogeneity in their decisions and use utility parameters to represent the average utilities for each class. In our empirical study on ICU admission, the dimension of state space in our model is already large: the average dimension is 1318 for the 22 hospitals, with a maximum of 3584. This presents a significant challenge to the model estimation, especially since we are identifying the discount factor jointly with the utility parameters.

From the modeling perspective, our set-up is similar to the consumer purchasing/replacement problems studied in Gordon (2009) and Gowrisankaran and Rysman (2012), in which the dynamic decision is made at the system-level. Accordingly, the model state has to include the characteristics (e.g., price, quality) of *all* available products in the market. In both papers, important simplifications and assumptions have to be made

Table EC.2 Estimation results of structural model by individual hospital

Hosp	Num. of periods	Discount Factor $\hat{\beta}$	Low Severity $\hat{u}_{w,l}$	Low Severity $\hat{u}_{r,l}$	High Severity $\hat{u}_{w,h}$	High Severity $\hat{u}_{r,h}$	R^2
1	8016	0.3	−0.015	1.490	−0.749	0.301	0.19
		(0.025)	(0.016)	(0.031)	(0.060)	(0.041)	
2	6012	0.5	−0.001	1.465	−0.659	0.519	0.23
		(0.026)	(0.038)	(0.028)	(0.054)	(0.055)	
3	8016	0.1	−0.124	2.558	−1.560	1.185	0.08
		(0.052)	(0.157)	(0.052)	(0.155)	(0.067)	
4	8016	0.4	−0.065	2.307	−0.940	0.798	0.15
		(0.011)	(0.034)	(0.040)	(0.053)	(0.044)	
5	6924	0.1	−0.179	1.786	−1.147	0.336	0.10
		(0.036)	(0.096)	(0.053)	(0.117)	(0.076)	
6	8016	0.2	−0.077	1.904	−0.970	0.531	0.14
		(0.041)	(0.039)	(0.045)	(0.080)	(0.048)	
7	6948	0.9	−1.755	2.687	−1.876	1.453	0.10
		(0.177)	(0.515)	(0.083)	(0.328)	(0.096)	
8	7848	0.1	−0.249	1.446	−0.764	0.380	0.15
		(0.062)	(0.111)	(0.034)	(0.089)	(0.051)	
9	6180	0.5	0.000	1.745	−0.789	0.707	0.18
		(0.035)	(0.036)	(0.040)	(0.062)	(0.050)	
10	7320	0.1	−0.629	2.339	−1.630	0.929	0.07
		(0.325)	(0.800)	(0.053)	(0.466)	(0.086)	
11	8016	0.9	−1.127	1.774	−1.366	0.104	0.08
		(0.248)	(0.538)	(0.064)	(0.274)	(0.108)	
12	4668	0.6	−0.099	2.121	−0.558	0.637	0.22
		(0.014)	(0.033)	(0.051)	(0.058)	(0.071)	
13	8016	0.1	−0.044	2.229	−1.260	0.785	0.13
		(0.016)	(0.051)	(0.045)	(0.110)	(0.055)	
14	8016	0.3	−0.147	1.951	−0.956	0.662	0.19
		(0.050)	(0.063)	(0.036)	(0.082)	(0.044)	
15	6912	0.1	−1.082	2.155	−2.275	0.762	0.10
		(0.071)	(0.180)	(0.045)	(0.219)	(0.076)	
16	8016	0.1	−0.199	2.319	−1.233	1.005	0.11
		(0.000)	(0.072)	(0.050)	(0.123)	(0.066)	
17	6588	0.9	−1.669	2.065	−1.530	0.660	0.09
		(0.082)	(0.202)	(0.069)	(0.134)	(0.081)	
18	8016	0.4	−0.154	2.358	−1.391	1.005	0.04
		(0.046)	(0.065)	(0.060)	(0.114)	(0.069)	
19	6576	0.1	−0.886	1.809	−1.885	0.517	0.12
		(0.000)	(0.069)	(0.039)	(0.138)	(0.056)	
20	8004	0.3	−0.116	2.260	−1.335	0.862	0.09
		(0.043)	(0.054)	(0.063)	(0.131)	(0.084)	
21	4008	0.7	−0.135	2.002	−0.612	0.240	0.12
		(0.047)	(0.058)	(0.094)	(0.096)	(0.120)	
22	4008	0.3	0.000	1.513	−0.979	0.175	0.09
		(0.042)	(0.030)	(0.068)	(0.130)	(0.111)	

Note: The second column reports the number of two-hour periods in each hospital, and the last column provides the McFadden's pseudo R^2. Standard errors are reported in parenthesis. They are estimated by bootstrapping with 500 trials. We find that the standard errors of the estimated $\hat{\beta}$'s are smaller than 0.05 (resp. 0.1) for 15 (resp. 19) out of the 22 hospitals. Although not reported here, in the bootstrapping, we find that the estimated $\hat{\beta}$ remains the same or only changes to the adjacent levels (plus or minus 0.1) in more than 90% of the trials for 20 of the 22 hospitals (except Hospitals 14 and 15).

to mitigate the curse of dimension of state space. For example, Gordon (2009) only considers two product types for each of the two firms in their model, by aggregating the product characteristics within each type. This is similar to our setting with two customer classes. In contrast, our set-up is different from the structural models in, e.g., Emadi and Staats (2020) and Yu et al. (2016), in which the dynamic decision is modeled at the *individual* level. Thus, the state space only needs to contain the information for each individual. This substantially reduces the computational burden and facilitates the use of parametric or random utility coefficients to account for agent heterogeneity. Moreover, we identify the discount factor and action utilities jointly in the dynamic discrete choice model, which brings additional technical challenge.

Counterexample for non-identification: The key intuition for model identification is that there is enough variation in the observed states and actions relative to the parameter space. We now provide a simple counter-example to show that the model cannot be identified when there is not sufficient variation in the states and actions.

We consider an FSU with only one bed. We assume there is always one patient in the GK in every period. The system can take three actions for the patient: keep her waiting in GK, admit her to SSU (with unlimited capacity), or admit her to FSU. The last action is only feasible when FSU is empty. After one action is taken in each period, the FSU state may transit due to external arrival or patient departure. The transition probabilities are given by $P = [p_{00}, p_{01}; p_{10}, p_{11}]$, where p_{ij} is the probability that the FSU state changes from i to j. The action utilities are given by u_w for GK waiting, u_r for SSU admission, and u_a for FSU admission. We assume an additive idiosyncratic term in the utility as in (3). The discount factor is $\beta \in (0,1)$. Denote the action probabilities by $f(\iota_1|n_t^I = 0)$ for $\iota_1 \in \{w, r, a\}$ and $f(\iota_2|n_t^I = 1)$ for $\iota_2 \in \{w, r\}$. These action probabilities can be computed similarly to eq. (9) in our main setting.

We perform the following numerical study. First, we compute the correct choice probabilities $f(\iota_1|n_t^I = 0)$ and $f(\iota_2|n_t^I = 1)$ under a given set of parameters (β, u_a, u_r, u_w). Then, we assume an alternative, incorrect discount factor β' and search for the corresponding utility parameters (u'_a, u'_r, u'_w) to fit the correct choice probabilities. The FSU admission utility is normalized to zero ($u_a = 0$).

We find that the discount factor cannot be identified in this model. In particular, for an incorrect discount factor β', we can always find the utility parameters (u'_w, u'_r) such that the choice probabilities coincide with the true values for every state and action. Thus, the model is unidentified due to observational equivalence. In particular, the choice probabilities can be equivalently rationalized by a small discount factor combined with high action utilities or a large discount factor combined with low action utilities (see Figure OS.3 of the Online Supplement). This counter-example shows that we need enough variation in the observed states and actions to identify the discount factor and utility parameters jointly.

State-dependent action set: A novelty of our identification is that our model has state-dependent action set, which differs from Komarova et al. (2018). We now conduct a simulation study based on the ICU admission problem to show that using state-dependent action sets is essential to correctly identify the model.

We first simulate a series of state and action pairs using the correct model with state-dependent action sets for a given set of discount factor and utility parameters. We define the action in period t as $d_t = (a_{l,t}, r_{l,t}, a_{h,t}, r_{h,t})$, which includes the numbers of ICU and non-ICU admissions for the low and high severity classes, respectively. We then select a state $\bar{s} = (n_l^G, n_h^G, n^I)$ as $n_\iota^G = \max\{n_{\iota,t}^G\}$ for $\iota \in \{l, h\}$ and $n^I = B - n_l^G - n_h^G$. That is, $\bar{s}$ has the maximum numbers of ED patients for both severity classes in the simulation and enough empty beds in the ICU for admitting all ED patients. It is easy to verify that with the state $\bar{s}$, all simulated actions $d_t = (a_{l,t}, r_{l,t}, a_{h,t}, r_{h,t})$ are feasible.

We then estimate the model parameters using the simulated data from the misspecified model without state-dependent action sets. For a candidate set of parameters $\theta = (\beta, u_{w,l}, u_{r,l}, u_{w,h}, u_{r,h})$, we solve the value function using the correct model with state-dependent action sets. However, in the estimation, we incorrectly assume that the system state is given by $\bar{s}$ in every period t. That is, we use the choice probability $f(d_t|\bar{s})$ and compute the likelihood function as $l^f = \Pi_{t=1}^T f(d_t|\bar{s})$ for the simulated data. We search for the discount factor and utility parameters which maximize the likelihood function. We use the same normalization as in main specification and restrict the discount factor to the grid $\{0.1, 0.2, ..., 0.9\}$.

We select three representative hospitals, which have small, medium, and large estimated discount factors under our main specification, to conduct the simulation study. For each hospital, we generate 100 paths of state-action pairs using the estimated parameters (see Table EC.2). Each path includes 20,000 two-hour intervals with a one-month warm-up period. For each path, we estimate the discount factor and utility parameters using the correct state-dependent action set $\Pi(s_t)$ as in our main specification, or the incorrect state-independent action set $\Pi(\bar{s})$ as described above. For each estimation approach, we take the average over the 100 paths to obtain the final estimates. The results are reported in Table OS.10.

We find that when the state-independent action set $\Pi(\bar{s})$ is used in identification and estimation, the estimated discount factor is always 0.9 and the utility parameters are largely biased from their true values (see the "State-independent action set $\Pi(\bar{s})$" columns). In particular, the estimated waiting utilities $\hat{u}_{w,l}$ and $\hat{u}_{w,h}$ are close to zero with an upward bias in all cases, and the non-ICU admission utilities $\hat{u}_{r,l}$ and $\hat{u}_{r,h}$ are largely underestimated with a substantial downward bias. In contrast, using the correct state-dependent action set $\Pi(s_t)$ produces estimated parameters (both discount factor and and utility parameters) that are close to the true values (see the "Correct action set" columns).

The estimation bias can be interpreted as follows. In the sample, the hospitals slow down ICU admissions when the ICU becomes full, as hospitals would attempt save ICU capacity for future patients. However, under the state $\bar{s}$, it is always feasible to admit all ED patients to the ICU. Thus, the hospitals' capacity saving behaviors of keeping patients in the ED can only be (partially) rationalized by a large discount factor, a high (non-negative) waiting utility, and/or a low non-ICU admission utility. This leads to the observed bias in the estimation results. The above comparison shows that it is essential to account for the state-dependent action set in the identification and estimation of dynamic discrete choice models.

EC.6. Capacity Intervention of Adding One ICU Bed

We first evaluate the counterfactual impact of adding one more bed in ICU, while keeping the arrival and service rates fixed. For the hospitals in our study, adding one ICU bed has limited impact on the average ICU occupancy level. We focus on the impact on high ICU congestion measures introduced in (14) – (17).

Table EC.3 reports the reductions in ICU congestion measures for six selected hospitals. We see that adding one bed in the ICU indeed leads to substantial reductions in ICU congestion. For some hospitals, such reduction translates to 18 fewer days and 50 fewer patients exposed to high ICU congestion ("ΔDays HighCgstn" and "ΔPats HighCgstn" columns), as well as 10 more external arrivals admitted ("ΔPats Balk" column) in each year. The relative drops in $\Pr(\text{HighCgstn})$ and $\Pr(\text{Balk})$ are greater than 30% in most cases. As patients who require ICU care are usually very unstable, the observed effects have the potential to cause great improvements in their medical outcomes. Although not reported here, the standard deviations computed from the simulation confirm that all differences are statistically significant at the 1% level.

Table EC.3 Counterfactual Estimates of Impact when Adding One Bed in ICU (Select Hospitals)

Hosp	$\hat{\beta}$	B	ΔDays HighCgstn (in # days)	ΔPats HighCgstn (in # patients)	Relative drop $\Pr(\text{HighCgstn})$	ΔPats Balk (in # patients)	Relative drop $\Pr(\text{Balk})$
1	0.3	21	6.27	52.26	35.3%	6.73	38.8%
2	0.5	26	6.78	52.78	24.9%	8.43	27.5%
5	0.1	11	18.74	68.04	38.2%	8.84	39.0%
8	0.1	16	10.90	64.88	37.5%	7.92	41.5%
9	0.5	22	8.60	78.61	31.7%	13.57	35.3%
14	0.3	36	2.36	25.90	24.0%	5.26	27.6%

Next, we look into how the benefit of adding one bed in ICU varies across hospitals. Table EC.4 reports the correlations of the reductions in the three congestion measures ("ΔDays HighCgstn", "ΔPats High-Cgstn", and "ΔPats Balk" columns in Table EC.3) with the average ICU occupancy levels and the estimated discount factors across the 22 hospitals in our sample. By the first two columns, all three effects are positively correlated with both ICU occupancy levels and estimated discount factors. This relationship still holds when we perform the following linear regression for the three effects using the estimates of the 22 hospitals:

$$\Delta\text{ICUCgstn}_i = \lambda_0 + \lambda_1 \hat{\beta}_i + \lambda_2 \text{ICUOccu}_i + \varepsilon_i.$$

By the last two columns of Table EC.4, the benefit of adding an ICU bed is positively associated with both ICU occupancy and the estimated discount factor, though the relative importance of the two varies across ICU congestion metrics. These results suggest that the benefit of adding one bed in the ICU tends to be greater for ICUs with higher occupancy levels and/or hospitals that account more for the system-level impact of their admission decisions. Although we do not interpret these results as prescriptive, they illustrate that the balancing behavior of hospitals is an important factor in capacity planning decisions. Notably, the discount factor reflects the how the hospital balances the current patients and future system capacity, while the ICU occupancy measures the average system utilization. Thus, they capture different aspects in the ICU admission decisions and are complementary in capacity planning. This reveals the benefit of understanding the hospitals' discounting behaviors from the data.

Table EC.4 Correlations between Impact of Adding One ICU Bed and ICU Occupancy or $\hat{\beta}$

	Correlation		Regression coefficient	
	$\hat{\beta}$	ICUOccu	$\hat{\beta}$	ICUOccu
ΔDays HighCgstn	0.69^{***}	0.36	18.58^{**}	3.98
ΔPats HighCgstn	0.53^{*}	0.77^{**}	19.64	140.90^{***}
ΔPats Balk	0.60^{**}	0.75^{**}	5.72^{*}	26.55^{**}

Note: $^{*}p<0.05$, $^{**}p<0.01$, and $^{***}p<0.001$.

EC.7. Robustness Checks

In this section, we conduct several robustness checks for our empirical results in Section 4. Our main findings are largely robust in these settings.

Choice of Normalization Levels: We perform two robustness checks for the normalization choice of $(u_{a,l}, u_{a,h})$, which can affect the identification results (see Section 2.2). First, we estimate our model using the periods with same or similar ED states, which mitigates the impact of normalization choice by Proposition 4. We select the state to be $(n_l^G, n_h^G) = (2,1)$, i.e., with two low severity patients and one high severity patient. This state includes multiple ED patients with both severity classes. Thus, the action space is large enough to reflect the balancing behavior of hospitals. In addition, the occurrence of the state is not too rare, so the model can be effectively estimated by combining all hospitals. In the estimation, we now calculate the log-likelihood l^f in (13) only using the actions under the ED state $(n_l^G, n_h^G) = (2,1)$.

The results are reported in Table EC.5 for all hospitals combined. The first row repeats Table 1 for our original model with all ED states. The second row reports the estimation results with only the ED state $(n_l^G, n_h^G) = (2,1)$. As an additional check, we report in the third row the estimated parameters when we use the ED states $(n_l^G, n_h^G) = (2,1)$ or $(1,1)$. This further increases the sample size. The ICU admission utilities are assumed to be zero in the estimations. We see that $\hat{\beta}$ equals to 0.3 in all the three cases, suggesting the identified discount factor is not affected when we limit to the select ED states. Figure OS.4 further shows that the likelihood decreases monotonically as we deviate from the optimal $\hat{\beta} = 0.3$ in all three cases. In addition, the waiting and non-ICU admission utilities are also similar to the original model for both classes. These findings support the robustness of our identification results.

As discussed in Section 3.3, we expect the utilities of ICU admission to be close to zero in our empirical setting. In the second robustness check, we further estimate our model for each hospital separately under

Table EC.5 **Estimation Results with Different ED States (All Hospitals Combined)**

(n_l^E, n_h^E)	Size	$\hat{\beta}$	$\hat{u}_{w,l}$	$\hat{u}_{r,l}$	$\hat{u}_{w,h}$	$\hat{u}_{r,h}$
All	154,140	0.3	-0.071	1.950	-0.932	0.671
		(0.005)	(0.001)	(0.010)	(0.018)	(0.013)
(2,1)	4,230	0.3	-0.013	2.001	-0.966	0.775
		(0.019)	(0.005)	(0.042)	(0.051)	(0.036)
(2,1) or (1,1)	11,706	0.3	-0.071	2.215	-1.059	0.727
		(0.023)	(0.044)	(0.031)	(0.040)	(0.022)

non-zero ICU admission utilities $u_{a,h}$ and $u_{a,l}$. Here all ED states are used in the estimation. In particular, we estimate the model with different normalization levels for the ICU admission utilities. We consider the cases with both $u_{a,h} = u_{a,l}$ and $u_{a,l} < u_{a,h}$ (we assume the ICU admission utility is higher for the high-severity class). The values of $u_{a,h}$ and $u_{a,l}$ are varied from -0.3 to 0.3. As in our main setting, we still assume that the ED admission utility is lower than the utility of ICU and non-ICU admission.

The estimated discount factors for individual hospitals are shown in Table OS.11. Each column in the table corresponds to a normalization choice $(u_{a,l}, u_{a,h})$. We find that the estimated discount factors are very robust under alternative normalization choices. The estimated discount factor $\hat{\beta}$ remains unchanged or only changes to adjacent levels (increase or decrease by 0.1) in all cases. We also compute the correlation (across hospitals) between $\hat{\beta}$ in the original model and each of the alternative settings. The correlation is higher than 0.9 in all cases. Thus, our results are robust to the choice of normalization levels.

Stratified Sample and Time Trend: As a robustness check, we also estimate the structural model with stratified samples based on flu vs non-flu seasons, as well as day and night periods. The flu season ranges from November to March, and the non-flu season includes the rest. The day periods include the twelve hours from 7AM to 7PM. We re-estimate the ED arrival rates $\lambda_{Q,l}$ and $\lambda_{Q,h}$, external ICU arrival rate λ_E, and ICU departure rate μ_I for each stratified sample. The ED and ICU capacities Q_l, Q_h, and B are set as those for the full sample in Table OS.6. We estimate the structural model with the stratified samples for all hospitals combined and for individual hospitals separately.

Table EC.6 reports the results from the combined estimation. We can see the estimated parameters are close to that in our main specification (first two rows). The estimated discount factor is 0.4 for the flu season, 0.2 for the non-flu season, and 0.4 for the day sample, and 0.1 for the night sample. The lower discounting level during the night can be explained as there may be more medical and operational constraints that hinder the hospital from responding to the system state during the night (e.g., shortage in nurses). For individual hospitals, we find the heterogeneity in discount factors estimated from the stratified data is largely consistent with that from the full sample given in Table EC.2. In particular, the across-hospital correlations between the estimated $\hat{\beta}$ from stratified and full sample are significantly positive: 0.87 for the flu season, 0.67 for the non-flu season and day sample, and 0.46 for the night sample. Thus, our main findings from the structural model are robust to accounting for these potential temporal variations.

In Section OS.4 of the Online Supplement, we further develop and estimate a structural model that explicitly accounts for time trend in the system transition during a day. The model includes a time indicator for each two-hour interval in a day in as a state variable, and allow the arrival and departure rates to vary with time. Our main findings remain very robust after the time trend is included in the model.

Table EC.6 Estimation of Structural Model with Stratified Data: All Hospitals Combined

	Size N	Discount factor $\hat{\beta}$	Low Severity $\hat{u}_{w,l}$	Low Severity $\hat{u}_{r,l}$	High Severity $\hat{u}_{w,h}$	High Severity $\hat{u}_{r,h}$	R^2
All	154,140	0.3	−0.071	1.950	−0.932	0.671	0.14
		(0.005)	(0.001)	(0.010)	(0.018)	(0.013)	
Flu	60,031	0.4	−0.028	1.936	−0.755	0.674	0.14
		(0.004)	(0.010)	(0.012)	(0.021)	(0.019)	
Non-flu	108,737	0.2	−0.058	1.960	−1.086	0.670	0.13
		(0.005)	(0.010)	(0.009)	(0.023)	(0.016)	
Day	84,372	0.4	−0.001	2.006	−0.806	0.655	0.13
		(0.004)	(0.009)	(0.011)	(0.020)	(0.018)	
Night	84,396	0.1	−0.045	1.875	−0.971	0.684	0.14
		(0.005)	(0.011)	(0.009)	(0.024)	(0.016)	

Standard error is reported in parenthesis.

Finer Grid Search, Three Patient Class, and Change in Unit Capacities: For computational purposes, we conduct a grid search over the discount factor with a step of 0.1 in our main specification. This facilitates the numerical implementation of our estimation as the state space in our model is relatively large. As a robustness check, we have also done a finer grid search with 0.01 increment for select Hospitals 4 and 9. The identified β's from the finer grid are very close to those from the original grid (see Figure OS.5 of the Online Supplement).

In Section OS.5 of the Online Supplement, we develop and estimate a dynamic structure model with three patient classes for all hospitals combined and each individual hospital. The estimation results (discount factor and utility parameters) are very similar under the two-class and three-class models (see Tables OS.4 and OS.13). Finally, we estimate our model under alternative ICU and ED capacity levels, which are set according to the data (see Section EC.2.2). In robustness checks, we increase or decrease the ICU capacity B by one bed, as well as drop the buffer term $\lfloor M_{A,i} \rfloor$ in the ED capacities Q_i for $i \in \{l, h\}$. The identified discount factors are reported in Tables OS.14 and OS.15, which are very similar to our main specification. These results further demonstrate the robustness of our findings.

EC.8. Summary of Model Assumptions

In this section, we present the major assumptions used in our empirical study as well as their economic justifications and robustness checks. These discussions are summarized in Table OS.16.

Linear utility function: As discussed in the paper, we achieve joint identification of discount factor and utility parameters by leveraging the results in Komarova et al. (2018), which require per-period payoff to be linear in utility parameters. This assumption is critical for our identification. It cannot be tested from data, but we believe it is a reasonable assumption to capture the first-order effect in our model. In particular, the hospitals' per-period (perceived) utility should be the sum of action utility for each patient affected by the decision, which is reflected by the linear utility function in (2). In addition, as shown in Table OS.18, the number of ED patients in each period is relatively small, mitigating the non-linearity concern regarding. Besides, we also find that the ED waiting time is relatively short for most patients (see Table OS.17), suggesting that patient deterioration in ED to have little impact.

Constant utility parameters:We assume the utility parameters are constant and independent of the state, which is standard in dynamic discrete choice model literature. This assumption is crucial for our identification and interpretation of the discount factor in our model. As discussed in Section 4.2, we expect the benefit from ICU care to remain relatively stable even when ICU is congested, since ICU typically allocates dedicated resource to each bed and maintains a fixed nurse-to-bed ratio. In addition, we find no significant impact of ICU occupancy on patient outcomes (see Table OS.2 in Section OS.3). These results provide suggestive evidence that our main findings are not driven by congestion-dependent ICU admission utility.

Normalization level: In our main model, we normalize the ICU admission utilities of both classes to be zero. The normalization assumption is necessary to achieve identification, and the specific normalization choice can affect our estimates (e.g., Aguirregabiria and Mira 2010). The normalization choice cannot be tested from the data, but we justify it based on our empirical context and provide multiple robustness checks.

The normalization assumption is needed for identification in most dynamic discrete choice models, as we cannot identify the utilities of all actions from data (e.g., Rust 1994, Magnac and Thesmar 2002). In addition, setting zero utility for a reference action is the most common practice (e.g., Magnac and Thesmar 2002, Ryan 2012, Aguirregabiria and Mira 2010). In general dynamic discrete choice models, the normalization choice can affect the identification results and counterfactual estimates (see e.g., Magnac and Thesmar 2002, Norets and Tang 2014, Aguirregabiria and Suzuki 2014, and Abbring and Daljord 2020).

As discussed in Section 3.3, the assumption of zero ICU admission utility is based on our empirical context. First, setting ICU admission utility to zero provides a proper reference point for perceived utility across all ED states, as ICU admission is the best action hospitals can do in terms of patient's clinical outcomes. Second, the marginal utility of admitting a patient to ICU should be small due to its very high fixed and operating cost. Due to data limitation, we assume the ICU admission utility is the same for the two severity classes. Note that if a patient is eventually admitted to ICU, it is likely because his unobserved factors or latest conditions imply significant risk, regardless of the pre-assigned severity class. From the data, we find that although the high severity class is four times more likely to be admitted to ICU than the low severity class (34% vs. 8%), the difference in their ICU length-of-stay (LOS) is much smaller. As shown in Table OS.19, the average ICU LOS of the low severity class is roughly 70% of that of the high severity class. In addition, the ICU LOS exhibits substantial variation for both severity classes.

Given the normalization choice impacts our identification results, we have performed multiple robustness checks using stratified samples (periods with same ED state) and alternative normalization levels. They are discussed in Section EC.7. We show that our main findings remain robust in these tests. Moreover, under current normalization assumption, the estimated dynamic discrete choice model fits our observed data well (Section OS.2) and provide evidence for hospitals' discounting behaviors that is consistent with the multinomial logistic model (Section EC.3).

Lower ED waiting utility: We assume the utility of keeping patients waiting in ED is lower than that of admitting them to the ICU and non-ICU units. This assumption is important for our empirical study, as it rules out unrealistic model estimates that are improper in our setting.

The assumption cannot be directly tested from data, but it is justified based on our empirical context. In our setting, the ED waiting time is defined as the time difference between the time the order to admit is made and the time the patient is eventually admitted to an inpatient unit. Thus, ED waiting is generally detrimental to our patient cohort, given they must be admitted to the hospital. It represents an additional cost in the patient rerouting process and is only a temporary state due to frictions or constraints.

From the data, we indeed find that the ED waiting time is relatively short (see Table OS.17) for most patients, suggesting that ED waiting is on average an undesirable state in patient flow. Therefore, we assume the ED waiting should bring lower perceived utility than ICU and non-ICU admission. Note that the utility parameters represent the *average* perceived utility associated with each action for the two patient classes. The patient-specific deviations from the average utilities are captured by the idiosyncratic term, as standard in dynamic discrete choice models (Rust 1987, Komarova et al. 2018).

Patient classes: We divide patients to two severity classes in our model and use six deterministic parameters to represent the average perceived utilities for their actions. As discussed in Section EC.5, this is necessary for our identification as our dimension of state space grows fast with the number of patient classes. We believe this assumption is innocuous as two severity classes are sufficient to capture the heterogeneity in the patients' need for service. As a robustness check on this assumption, we have shown that our main findings remain very similar when we define three patient classes with low, medium, and high severity levels. This is described in Section OS.5 (Tables OS.4 and OS.13).

Other model primitives: We assume constant model primitives including ICU and ED capacities, arrival rates to ED and ICU, and ICU departure rate. These set-ups are in line with the dynamic discrete choice model literature (e.g., Rust 1994). The model parameters are directly estimated from the observed data using standard methods. We conduct multiple robustness checks on these model primitives. For ICU and ED capacities, we estimate the model after increasing or decreasing the ICU capacity by one bed from the current level, as well as after dropping the buffer term in the ED capacity. The estimation results remain very similar to our main specification for both all hospitals combined and individual hospitals (see Tables OS.14 and OS.15). For arrival and departure rates, we estimate the model after incorporating the time trend in the state variable (see Section OS.4). That is, we allow the arrival and departure rates to vary for each two-hour interval. In addition, we estimate the model separately for daytime and nighttime, as well as for flu and non-flu seasons. Our main findings remain robust in the above settings.

In line with the literature, we assume the arrivals per period follow Poisson distributions. To validate this, we compute the ratio between the variance of arrivals per period to the mean of arrivals for each hospital. As reported in Table OS.20, the ratio is close to one for most hospitals, consistent with the Poisson distribution. We also assume the number of ICU departures follows a binomial distribution in each period with a constant rate. Table OS.21 reports the estimated departure rates for ICU occupancy levels in each decile bin (all hospitals combined). We see that the departure rates are relatively stable for different decile bins.

EC.9. Impact of Misspecifying Discount Factor in Identification

In this section, we use counterfactual simulations to show how pre-specifying an “incorrect” discount factor leads to deviations in evaluating the high ICU congestion periods. We consider four choices of discount factors for each hospital: $\beta = 0$ for a pure myopic model, $\beta = 0.3$ as the average level from the combined estimation, $\beta = \hat{\beta}$ at the estimated level, and $\beta = 0.99$ as is commonly used in the literature. For a given discount factor, the utility parameters are estimated accordingly for each hospital. By our estimation procedure, the estimated discount factor leads to the highest likelihood of hospitals’ observed actions (see Figures OS.4 and OS.5), i.e., it fits the data better than alternative discount factors. Thus, we use it as the benchmark for comparison. We also perform the likelihood-ratio test to compare different models. For all hospitals combined, the models with $\beta = 0$ and $\beta = 0.99$ are rejected compared to the identified model at the 0.1% significance level. For individual hospitals, the model with $\beta = 0$ (resp. with $\beta = 0.99$) is rejected compared to the identified model at the 1% level for 16 (resp. 15) out of the 22 hospitals.

We use counterfactual simulations to estimate the ICU congestion under three scenarios. The first is the base case with the estimated system set-up. We then consider two system changes. In the first one, we increase the ED arrival rates $\lambda_{Q,l}$ and $\lambda_{Q,h}$ by 50%. In the second one, we keep the ED arrivals at the original level but reduce the ICU capacity B by one bed. The two system changes reflect scenarios when ICU capacity gets constrained (e.g., during a pandemic). Correctly evaluating the impact of such shocks is crucial to hospitals for managing their ICUs.

Table EC.7 provides the estimates of total numbers of patients exposed to high ICU congestion and external arrivals who balk under different choices of discount factor (with corresponding utility estimates) and the three scenarios of system set-up. The numbers in the table represent the sum of the 22 hospitals in our sample. The top row “$\beta = \hat{\beta}$” shows the estimates when we jointly identify discount factor and utility parameters for each hospital. Thus, we use it as the benchmark for comparison. For other choices of β, their relative deviations compared with $\beta = \hat{\beta}$ are reported in the parenthesis.

Table EC.7 Estimated ICU Congestion Metrics under Difference Choices of Discount Factors and Corresponding Utility Estimates

Discount factor	Base case		Increasing ED arrival by 50%		Reducing one ICU bed	
	Pats HighCgstn	Pats Balk	Pats HighCgstn	Pats Balk	Pats HighCgstn	Pats Balk
$\beta = \hat{\beta}$	2,356	352	4,408	678	3308	546
$\beta = 0$	4,319 (83.4%)	718 (103.8%)	7,384 (67.5%)	1,282 (89.0%)	5,445 (64.6%)	995 (82.5%)
$\beta = 0.3$	2,961 (25.7%)	461 (30.9%)	5,335 (21.0%)	866 (27.6%)	3,993 (20.7%)	692 (26.8%)
$\beta = 0.99$	2,223 (−5.6%)	324 (−8.1%)	3,718 (−15.6%)	554 (−18.3%)	3,065 (−7.3%)	500 (−8.3%)

We have the following observations. First, setting $\beta = 0$ or 0.3 in the structural model substantially overestimates the ICU congestion compared with using the identified $\hat{\beta}$. In the base case, setting $\beta = 0$ (resp. $\beta = 0.3$) overestimates the number of patients under high ICU congestion by 83.4% (resp. 25.7%) and the external arrivals who balk by 103.8% (resp. 30.9%). In contrast, assuming $\beta = 0.99$ would underestimate the ICU congestion. In the base case, it underestimates the patients under high ICU congestion (resp. externals who balk) by 5.6% (resp. 8.1%). Such deviation becomes substantially larger when the ICU gets more constrained in adverse scenarios. After we increase ED arrivals by 50% (resp. reduce one ICU bed), the relative underestimation with $\beta = 0.99$ increases to 15.6% (resp. 7.3%) for high ICU congestion patients and 18.3% (resp. 8.3%) for balking external arrivals. The above findings are in line with our interpretation of the discount factor: Hospitals with larger discount factors respond to ICU state more proactively when making admission decisions, i.e., reducing ICU admission likelihood when ICU occupancy is high. This mitigates the high ICU congestion periods.

We also compute other key system statistics under different discount factors, including overall ICU admission probabilities, ICU occupancy, and ED waiting times. We find that using $\beta = 0.99$ and $\beta = \hat{\beta}$ with their respective estimated utility parameters in the model produce close estimates for these system statistics. This reflects the potential "observational equivalence" issue that hinders the identification (e.g., Komarova et al. 2018). However, as shown in Table EC.7, the estimates of high ICU congestion clearly hinge on the choice of discount factor. This is because the impact of hospital's balancing behavior is more visible during high ICU congestion periods. The above finding highlights the importance of identifying the discount factor using observed data. Similar analysis is conducted in Ching and Osborne (2020) under a different setting. It shows that using an "incorrect" discount factor of 0.995 would largely overestimate the increase in sales from price promotions.

References

Abbring, Jaap H, Øystein Daljord. 2020. Identifying the discount factor in dynamic discrete choice models. *Quantitative Economics* **11**(2) 471–501.

Aguirregabiria, Victor, Pedro Mira. 2010. Dynamic discrete choice structural models: A survey. *Journal of Econometrics* **156**(1) 38–67.

Aguirregabiria, Victor, Junichi Suzuki. 2014. Identification and counterfactuals in dynamic models of market entry and exit. *Quantitative Marketing and Economics* **12**(3) 267–304.

Ching, Andrew T, Matthew Osborne. 2020. Identification and estimation of forward-looking behavior: The case of consumer stockpiling. *Marketing Science* .

Emadi, Seyed Morteza, Bradley R Staats. 2020. A structural estimation approach to study agent attrition. *Management Science* **66**(9) 4071–4095.

Gordon, Brett R. 2009. A dynamic model of consumer replacement cycles in the pc processor industry. *Marketing Science* **28**(5) 846–867.

Gowrisankaran, Gautam, Marc Rysman. 2012. Dynamics of consumer demand for new durable goods. *Journal of Political Economy* **120**(6) 1173–1219.

Komarova, Tatiana, Fabio Sanches, Daniel Silva Junior, Sorawoot Srisuma. 2018. Joint analysis of the discount factor and payoff parameters in dynamic discrete choice models. *Quantitative Economics* **9**(3) 1153–1194.

Magnac, Thierry, David Thesmar. 2002. Identifying dynamic discrete decision processes. *Econometrica* **70**(2) 801–816.

Norets, Andriy, Xun Tang. 2014. Semiparametric inference in dynamic binary choice models. *Review of Economic Studies* **81**(3) 1229–1262.

Rust, John. 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* 999–1033.

Rust, John. 1994. Structural estimation of markov decision processes. *Handbook of Econometrics* **4** 3081–3143.

Ryan, Stephen P. 2012. The costs of environmental regulation in a concentrated industry. *Econometrica* **80**(3) 1019–1061.

Yu, Qiuping, Gad Allon, Achal Bassamboo. 2016. How do delay announcements shape customer behavior? an empirical study. *Management Science* **63**(1) 1–20.

Online Supplement to "Structural Estimation of Intertemporal Externalities with Application in ICU Admissions"

Yiwen Shen
School of Business and Management, The Hong Kong University of Science and Technology, yiwenshen@ust.hk

Carri W. Chan
Decision, Risk, and Operations, Columbia Business School, cwchan@columbia.edu

Fanyin Zheng
Imperial Business School, Imperial College London, f.zheng@imperial.ac.uk

Gabriel Escobar
Division of Research, Kaiser Permanente, gabriel.escobar@kp.org

OS.1. Proofs

OS.1.1. Proof for Proposition 2

By Rust (1987), the value function $\tilde{V}(s)$ can be expressed as:

$$\begin{aligned}\tilde{V}(s) &= \sum_{s'} \int_{\varepsilon'} V(s', \varepsilon')\, g(s'|s) q(\varepsilon'|s') d\varepsilon' \\ &= \sum_{s_t} \int_{\varepsilon_t} \sup_{d_t \in \Pi(s_t)} \mathsf{E}\left\{ \sum_{j=t}^{\infty} \beta^{j-t} U(s_j, d_j, \varepsilon_j) | s_t, \varepsilon_t \right\} g(s_t|s) q(\varepsilon_t|s_t) d\varepsilon_t, \end{aligned} \quad \text{(OS.1)}$$

where the second equality follows from the definition of $V(s_t, \varepsilon_t)$ in (6). The expectation above is taken over the transition of (s_j, ε_j) starting from (s_t, ε_t) based on the transition probability $g(s'|s)$ and probability density $q(\varepsilon'|s')$. Thus, the function $\tilde{V}(s)$ represents the expected future utilities starting from intermediate state s and assuming the system always takes the optimal action.

We introduce following notations. We consider two systems s and s'. For every period t, we use s_t and s'_t to denote the system states at the start of period t, which are $s_t = \left(n^G_{l,t}, n^G_{h,t}, n^F_t\right)$ and $s'_t = \left(\left(n^G_{l,t}\right)', \left(n^G_{h,t}\right)', \left(n^F_t\right)'\right)$. Additionally, we use $\tilde{s}_t = \left(\tilde{n}^G_{l,t}, \tilde{n}^G_{h,t}, \tilde{n}^F_t\right)$ and $\tilde{s}'_t = \left(\left(\tilde{n}^G_{l,t}\right)', \left(\tilde{n}^G_{h,t}\right)', \left(\tilde{n}^F_t\right)'\right)$ to denote the intermediate states after actions d_t and d'_t are taken in period t, which are given by $\tilde{s}_t = \varphi(s_t, d_t)$ and $\tilde{s}'_t = \varphi(s'_t, d'_t)$ according to (8). Assume the two systems start from intermediate states $\tilde{s}_0$ and $\tilde{s}'_0$ with

$$\tilde{n}^G_{i,0} = \left(\tilde{n}^G_{i,0}\right)' \text{ for } i \in \{l, h\}, \text{ and } \tilde{n}^F_0 \le \left(\tilde{n}^F_0\right)', \quad \text{(OS.2)}$$

then Proposition 2 translates to $\tilde{V}(\tilde{s}_0) \ge \tilde{V}(\tilde{s}'_0)$.

Coupling: Our proof is based on a coupling argument and induction in time. We first introduce the coupling of the two systems s and s' as follows: First, the two systems witness identical arrivals to GK and external arrivals to the FSU in every period, i.e., $A_{l,t} = A'_{l,t}$, $A_{h,t} = A'_{h,t}$, and $E_t = E'_t$ for every t. Next, we couple the FSU departures from the two systems as follows. Denote the numbers of FSU customers before departures by $\bar{n}_t^F$ and $(\bar{n}_t^F)'$ respectively and assume $\bar{n}_t^F \leq (\bar{n}_t^F)'$. Then, the departures D_t and D'_t are coupled as $D'_t = D_t + Z_t$, where D_t is a Binomial-$(\bar{n}_t^I, \mu_I)$ variable and Z_t is a Binomial-$((\bar{n}_t^F)' - \bar{n}_t^F, \mu_I)$ variable. That is, the number of departures in the s' system is always at least as many as the number in the s system. Finally, if an identical action d is taken in each system, the random utility components $\varepsilon_t(d)$ and $\varepsilon'_t(d)$ associated with that action d coincides for the two systems in every period t.

Under the coupling described above, we first prove following lemma.

Lemma 1 *Under the coupling, if the intermediate states in each system in period $t-1$ satisfy*

$$\tilde{n}_{i,t-1}^G = \left(\tilde{n}_{i,t-1}^G\right)' \textit{ for } i \in \{l,h\}, \textit{ and } \tilde{n}_{t-1}^F \leq \left(\tilde{n}_{t-1}^F\right)', \tag{OS.3}$$

then the states at the start of period t satisfy

$$n_{i,t}^G = \left(n_{i,t}^G\right)' \textit{ for } i \in \{l,h\} \textit{ and } n_t^F \leq \left(n_t^F\right)'. \tag{OS.4}$$

PROOF: The result follow directly from the coupled arrivals and departures in the two systems. Since we start from the intermediate state, the system evolution to period t is only dictated by the stochastic arrivals to the GK, external arrivals to the FSU, and departures from the FSU during period $t-1$.

It is trivial to see the relationship for the customers in GK holds as we assume same GK arrival processes by coupling. For the FSU customers, we consider both external arrivals and departures. We have

$$\begin{aligned}\left(n_t^F\right)' - n_t^F &= \min\left\{(\tilde{n}_{t-1}^F)' + E'_{t-1}, B\right\} - \min\left\{\tilde{n}_{t-1}^F + E_{t-1}, B\right\} - \left(D'_{t-1} - D_{t-1}\right) \\ &= \min\left\{(\tilde{n}_{t-1}^F)' + E_{t-1}, B\right\} - \min\left\{\tilde{n}_{t-1}^F + E_{t-1}, B\right\} - Z_{t-1} \geq 0.\end{aligned}$$

The last equality follows from the coupling $E_t = E'_t$ for external arrivals and $D_{t-1} + Z_{t-1} = D'_{t-1}$ for departures. The last inequality holds as Z_{t-1} follows a Binomial-(L_{t-1}, μ_I) distribution with $L_{t-1} = \min\left\{(\tilde{n}_{t-1}^F)' + E_{t-1}, B\right\} - \min\left\{\tilde{n}_{t-1}^F + E_{t-1}, B\right\}$. This completes the proof for Lemma 1. □

Mimicking Policy: We now define the policies used in each system. We assume the system s' always takes its optimal action which achieve the supremum in (OS.1). For system s, we define a mimicking policy π which mimics the action taken in the s' system whenever possible; if it is not possible, it takes its own optimal action. We denote the value function associated with this policy by $V^\pi(s)$, which is defined by (OS.1) with optimal action d_t replaced by the one under policy π. By definition, we have

$$\tilde{V}(\tilde{s}_0) \geq V^\pi(\tilde{s}_0). \tag{OS.5}$$

To prove the proposition, we will establish following two properties under our coupling and the policy π. First, two systems always have same number of customers in the GK, but system s has no more customers in the FSU:

$$n_{i,t}^G = \left(n_{i,t}^G\right)' \text{ for } i \in \{l,h\} \text{ and } n_t^F \leq \left(n_t^F\right)', \quad \forall t. \tag{OS.6}$$

Second, the action taken in the s' system is always admissible for system s; thus system s always mimics the action of s' under π:

$$d_t = d_t' \in \Pi(s_t), \quad \forall t. \tag{OS.7}$$

Note that (OS.6) directly implies (OS.7), as it follows from (1) that given the same number of customers in GK, the system with fewer FSU customers has a larger admissible action set, leading to $d_t' \in \Pi(s_t') \subseteq \Pi(s_t)$.

Induction: We establish (OS.6) for every t by induction.

Base Case: The base case follows directly from the relationship of the initial intermediate states $\tilde{s}_0$ and $\tilde{s}_0'$, which satisfy (OS.2), and from Lemma 1. Thus we have $n_{i,1}^G = \left(n_{i,1}^G\right)'$ for $i \in \{l,h\}$ and $n_1^F \leq \left(n_1^F\right)'$.

Inductive Step: We assume (OS.6) holds for period j and show this implies it holds for period $j+1$. In period j, under policy π, system s takes the same action of s' since by the inductive hypothesis the action is admissible, i.e., $d_j = d_j'$. Given the same action is taken in each system, it is easy to verify the intermediate states after action also coincide, i.e., $\tilde{n}_{i,j}^G = \left(\tilde{n}_{i,j}^G\right)'$ for $i \in \{l,h\}$ and $\tilde{n}_j^F = \left(\tilde{n}_j^F\right)'$. Finally, we can apply Lemma 1 to prove the relationship (OS.6) holds for period $j+1$. This completes the inductive step.

Per-Period Utilities: We have shown that under our coupling and the policy π for the s system, at the start of each period, the two systems always have same numbers of customers in the GK, and system s always has fewer customers in the FSU than that in system s'. Thus, the system s always mimics the action by s' under the policy π. We next prove the per-period utilities always coincide for the two systems, which follows by:

$$\begin{aligned} U(s_t, d_t, \varepsilon_t) &= u(s_t, d_t) + \varepsilon_t(d_t) = \sum_{i \in \{l,h\}} u_{a,i} a_{i,t} + \sum_{i \in \{l,h\}} u_{r,i} r_{i,t} + \sum_{i \in \{l,h\}} u_{w,i} \left(n_{i,t}^G - a_{i,t} - r_{i,t}\right) + \varepsilon_t(d_t) \\ &= \sum_{i \in \{l,h\}} u_{a,i} a_{i,t}' + \sum_{i \in \{l,h\}} u_{r,i} r_{i,t}' + \sum_{i \in \{l,h\}} u_{w,i} \left(\left(n_{i,t}^G\right)' - a_{i,t}' - r_{i,t}'\right) + \varepsilon_t'(d_t') = U(s_t', d_t', \varepsilon_t'). \end{aligned}$$

This is because: (i) Both systems take the same action, thus they send same numbers of customers to the FSU and SSU, leading to same routing utilities; (ii) As both systems have same numbers of customers in the GK, the number of customers remaining in the GK after actions are also the same, leading to the same waiting utilities; (iii) By our coupling, the random utility components coincide for the same action $d_t = d_t'$, i.e., $\varepsilon_t(d_t) = \varepsilon_t'(d_t')$.

As the per-period utilities coincide for every period given system s takes policy π and system s' takes its own optimal policy, we have $V^\pi(\tilde{s}_0) = \tilde{V}(\tilde{s}_0')$. Then it follows by (OS.5)

$$\tilde{V}(\tilde{s}_0) \geq V^\pi(\tilde{s}_0) = \tilde{V}(\tilde{s}_0').$$

This proves the proposition.

OS.1.2. Proof for Proposition 3

We prove equation (11) in below. The proof of equation (12) follow in a similar way. For the two states $s_1 = (1,0,0)$ and $s_2 = (0,1,0)$, denote the probabilities of admitting the patient to FSU and SSU under the them by $\Pr(a|s_1)$, $\Pr(r|s_2)$, $\Pr(a|s_1)$, and $\Pr(r|s_2)$ respectively. By Proposition 1, we can compute the log of these choice probabilities as

$$\ln \Pr(a|s_1) = \beta \tilde{V}(s_a) - C(s_1), \quad \ln \Pr(r|s_1) = -c_{r,l} + \beta \tilde{V}(s_r) - C(s_1), \tag{OS.8}$$

and

$$\ln \Pr(a|s_2) = \beta \tilde{V}(s_a) - C(s_2), \quad \ln \Pr(r|s_2) = -c_{r,h} + \beta \tilde{V}(s_r) - \ C(s_2), \tag{OS.9}$$

where $s_a = (0,0,1)$ and $s_r = (0,0,0)$ denote the system state after the customer is admitted to the FSU and SSU respectively. Here we use the assumption $u_{a,l} = u_{a,h} = 0$. Taking the differences in the choice probabilities across the two states for FSU and SSU admission decisions respectively, we obtain

$$\ln \Pr(a|s_1) - \ln \Pr(a|s_2) = C(s_2) - C(s_1). \tag{OS.10a}$$

and

$$\ln \Pr(r|s_1) - \ln \Pr(r|s_2) = u_{r,l} - u_{r,h} + C(s_2) - C(s_1). \tag{OS.10b}$$

Here we see the terms involving the value function $\tilde{V}(\cdot)$ in (OS.8) – (OS.9) are canceled out in the differences. Further subtracting (OS.10a) from (OS.10b), we get rid of the terms related to state-dependent function $C(\cdot)$ and identify the SSU admission utility difference as

$$u_{r,l} - u_{r,h} = \ln\left(\frac{\Pr(r|s_1)}{\Pr(a|s_1)}\right) - \ln\left(\frac{\Pr(r|s_2)}{\Pr(a|s_2)}\right).$$

This proves (11). The above steps show how well-constructed state and action pairs can be used to disentangle the complex structure in the choice probability expression (9). The proof of (12) follows similarly.

OS.1.3. Proof for Proposition 4

With non-zero utility $u_{a,h}$ for sending the high-priority customer to the FSU, the per-period utility of action is given by

$$\begin{aligned} u'(s_t, d_t) &= u_{a,l}a_{l,t} + u_{r,l}r_{l,t} + u_{w,l}\left(n^G_{l,t} - a_{l,t} - r_{l,t}\right) + u_{a,h}a_{h,t} + u_{r,h}r_{h,t} + u_{w,h}\left(n^G_{h,t} - a_{h,t} - r_{h,t}\right) \\ &= \sum_{\iota \in \{l,h\}} \Delta u_{w,\iota}\left(n^G_{\iota,t} - a_{\iota,t} - r_{\iota,t}\right) + \sum_{\iota \in \{l,h\}} \Delta u_{r,\iota} r_{\iota,t} + \Delta u_{a,l}a_{l,t} + u_{a,h}n^G_t, \end{aligned} \tag{OS.11}$$

where $n^G_t = n^G_{l,t} + n^G_{h,t}$ denotes the total number of customers in GK in period t. In addition, we denote $\Delta u_{w,\iota} = u_{w,\iota} - u_{a,h}$, and $\Delta u_{r,\iota} = u_{r,\iota} - u_{a,h}$ for $\iota \in \{l,h\}$, as well as $\Delta u_{a,l} = u_{a,l} - u_{a,h}$. They represent the relative actions utilities compared with sending the high-priority customer to the FSU.

We see that, with non-zero utility for sending high-priority customer to the FSU, the last term $u_{a,h}n_t^G$ is involved in the per-period utility. Denote the per-period utility and value function with non-zero FSU routing utility $u_{a,h}$ by $U'(s_t,d_t,\varepsilon_t)$ and $V'(s_t,\varepsilon_t)$, respectively. By (6), we have

$$V'(s_t,\varepsilon_t)=\sup_{d_t\in\Pi(s_t)}\mathsf{E}\left\{\sum_{j=t}^{\infty}\beta^{j-t}U'(s_j,d_j,\varepsilon_j)|s_t,\varepsilon_t\right\}=\sup_{d_t\in\Pi(s_t)}\mathsf{E}\left\{\sum_{j=t}^{\infty}\beta^{j-t}\left[U(s_j,d_j,\varepsilon_j)+u_{a,h}n_j^G\right]|s_t,\varepsilon_t\right\}, \tag{OS.12}$$

where $U(s_j,d_j,\varepsilon_j)$ denotes the per-period utility with zero FSU routing utility $u_{a,h}$ and same relative utilities of other actions, i.e., $\Delta u_{w,\iota}$ and $\Delta u_{r,\iota}$ for $\iota\in\{l,h\}$ as well as $\Delta u_{a,l}$.

With same total number of customers in GK for each period, we can drop the subscript j in n_j^G and write it as n^G. Then the term $u_{a,h}n^G$ in (OS.12) is independent of action d_t and can be extracted out of the conditional expectation. We can write (OS.12) as

$$V'(s_t,\varepsilon_t)=\sup_{d_t\in\Pi(s_t)}\mathsf{E}\left\{\sum_{j=t}^{\infty}\beta^{j-t}U(s_j,d_j,\varepsilon_j)|s_t,\varepsilon_t\right\}+\frac{1}{1-\beta}u_{a,h}n^G, \tag{OS.13}$$

where the last term is a constant number and independent of the optimal actions. We note that by (6), the first term in (OS.13) is exactly the value function under the model with zero FSU routing utility for high-priority customers and same relative utilities for other actions. Thus, we can establish the following relationship between the value functions with different normalization levels of $u_{a,h}$:

$$V'(s_t,\varepsilon_t)=V(s_t,\varepsilon_t)+\frac{1}{1-\beta}u_{a,h}n^G.$$

Similarly, by (OS.1), we can prove the function $\tilde V'(s)$ under non-zero utility $u_{a,h}$ satisfies

$$\tilde V'(s)=\tilde V(s)+\frac{1}{1-\beta}u_{a,h}n^G. \tag{OS.14}$$

By (OS.11) and (OS.14), we have

$$u'(s,d)+\beta\tilde V'(\varphi(s,d))=u(s,d)+\beta\tilde V(\varphi(s,d))+\frac{1}{1-\beta}u_{a,h}n^G.$$

Plugging this in to (10), it is easy to verify the new function $\tilde V'(s)$ satisfies the functional equation. Moreover, by (9), we can show that the new choice probability $f'(d_t|s_t)$ coincides with the original ones under $u_{a,h}=0$. Thus, model identification is not affected by the normalization level of $u_{a,h}$.

OS.2. Goodness of Fit of Estimated Structural Model

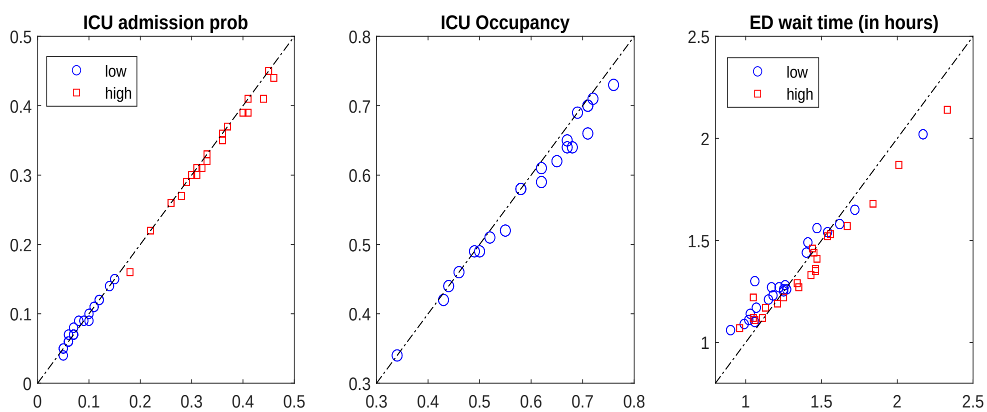
We show our estimated structural model provides good fits for both the hospitals' decisions and the overall system statistics. First, we compare the explanatory power for the hospitals' decisions, as measured by the McFadden's pseudo R^2, from our structural model with that from the reduced-form multinomial logistic regression model in Section EC.3 of the Online Supplement. As shown in Tables 1 and EC.1, the pseudo

R^2 for the combined estimation of all hospitals is similar from the structural model and logistic regression (0.14 versus 0.16). The average pseudo R^2 for individual hospital estimations is also comparable (0.13 vs 0.16). We note that the multinomial model contains a comprehensive set of variables that might influence admission decisions, including patient's characteristics, system states, and multiple dummies for seasonality fixed effects. Thus, the similar R^2 suggests that our structural model has reasonable explanatory power in capturing the hospitals' decisions.

We also find the structural model and reduced-form regression provide consistent evidences for the heterogeneity in hospitals' discounting behaviors. In the multinomial logistic model, we measure the impact of ICU occupancy on ICU admission likelihood by the coefficient γ_{ICU}. As discussed in Section EC.3, γ_{ICU} is significantly negative for most hospitals, suggesting the ICU admission probability drops as ICU occupancy increases. We find that the estimated $\hat{\beta}$ and γ_{ICU} are negatively correlated: the correlation is -0.770 across the 22 hospitals, statistically significant at the 0.1% level. As such, hospitals with larger estimated discount factors indeed tend to be more responsive to ICU congestion when admitting patients.

Next, we use simulation to show that our structural model produces system statistics close to those observed in the data. We examine several important system statistics, including the average ICU occupancy, overall proportion eventually admitted to the ICU for the two classes, as well as the ED waiting times for the two classes. The statistics estimated from our structural model are averaged from 200 simulation trials as in the counterfactual studies in Section 5.

Figure OS.1 Comparison of System Statistics from Structural Model and Real Data



Note: The figure compares the system statistics simulated from the structural model (y-coordinate) and observed from the real data (x-coordinate), including ICU admission probability (left), ICU occupancy (middle), and ED waiting times (right).

Figure OS.1 compares the simulated and observed system statistics: ICU admission probabilities (left), ICU occupancy (middle), and ED waiting times (right). In the panels, each point represents a hospital in our study; its x-coordinate (y-coordinate) corresponds to the observed (simulated) value of the system statistic.

We plot the 45-degree line in each panel, which represents a perfect fit. As we can see, most points fall very close to the 45-degree line, implying that our estimated structural model produces system statistics that are very close to the observed data. We also check the correlation and the relative mean absolute deviation (MAD) between the observed and simulated system statistics. The results are reported in Table OS.1. We see the correlations are very close to one and the relative deviation is small (2.2% to 4.5%). This further supports its effectiveness in modeling the admission process for ED patients.

Table OS.1 Correlation and Relative MAD between Simulated and Observed Statistics

	ICU Adm. Prob	ICU Avg Occu	Avg ED Wait
Correlation	0.999	0.993	0.976
Relative MAD	2.21%	2.58%	4.49%

With a small number of parameters, over-fitting is unlikely for our structural model. To further address this concern, we divide the sample to first and second halves for each hospital. We estimate the structural model with all hospitals combined using the first half sample, and do out-of-sample prediction on the second half. The McFadden's pseudo R^2 from the out-of-sample prediction (0.13) is very similar to that from the in-sample estimation (0.14) and the level from the full sample estimation in Table 1 (0.14).

OS.3. Impact of ICU Occupancy on Patient Outcomes

In this section, we use reduced-form analysis to examine if ICU congestion affects the outcomes of patients admitted to the ICU. We run an OLS regression to examine how ICU occupancy affects patient outcomes:

$$Y_i = \gamma_{0,d} + \gamma_{L,d}\text{LAPS2}_i + \gamma_{ICU,d}\text{ICUOccu}_t + \gamma_{ED,d}\text{EDNum}_t + \boldsymbol{\gamma}_{Z,d}^\top \mathbf{Z}_{i,t} + \epsilon_{it}, \tag{OS.15}$$

where Y_i is one of the following patient outcomes: 30-day mortality, hospital readmission, and hospital length-of-stay (LOS). ICUOccu_t is the ICU occupancy level at the time of patient admission, measured by the percentage occupancy relative to the ICU capacity; EDNum_t is the number of patients in the ED (including both high and low severity types). The other patient and system-level controls $\mathbf{Z}_{i,t}$ are introduced in Section EC.3. The set-up of the OLS regression (OS.15) is similar to our reduced-form analysis for the patient admission decision in Section EC.3.

As an additional test, we also examine the impact of the average ICU occupancy during a patient's ICU stay, rather than the occupancy level at the time of ICU admission. The average ICU occupancy is also measured relative to the ICU capacity.

$$Y_i = \gamma_{0,d} + \gamma_{L,d}\text{LAPS2}_i + \gamma_{ICU,d}\text{AvgICUOccu}_i + \gamma_{ED,d}\text{EDNum}_t + \boldsymbol{\gamma}_{Z,d}^\top \mathbf{Z}_{i,t} + \epsilon_{it}, \tag{OS.16}$$

The average ICU occupancy, AvgICUOccu_i, is computed using the two-hour snapshots from our structural model. For example, if a patient enters the ICU at 12:45 and leaves at 17:16, we take the average of ICU occupancy levels recorded at the 13:00, 15:00, and 17:00 snapshots. Note the average ICU occupancy level is not available to the physician at the time of a patient's ICU admission, but this analysis allows us to understand the association between congestion and outcomes. We replace ICUOccu_t by AvgICUOccu_i in regression (OS.15). The two variables are not included simultaneously as they are highly correlated (a correlation of 0.85).

Table OS.2 Estimation Results for OLS Regression (OS.15), N = 19,680

Panel A: ICU occupancy at the time of admission

	Mortality	Readmission	Hospital LOS
LAPS2	0.004^{***}	0.0006^{***}	0.0355^{***}
	(8.02×10^{-5})	(8.68×10^{-5})	(0.002)
ICUOccu_t	-0.0076	0.0022	-0.381
	(0.017)	(0.019)	(0.491)
EDNum_t	-0.0009	0.0048^{*}	0.0898
	(0.002)	(0.002)	(0.055)

Panel B: Average ICU occupancy during a patient's stay

	Mortality	Readmission	Hospital LOS
LAPS2	0.0004^{***}	0.0006^{***}	0.0354^{***}
	(8.01×10^{-5})	(8.68×10^{-5})	(0.002)
AvgICUOccu_i	-0.011	0.0153	-0.0423
	(0.020)	(0.021)	(0.556)
EDNum_t	-0.0008	0.0047^{*}	$0.0897^{\dagger}$
	(0.002)	(0.002)	(0.053)

Note: ${}^{\dagger}p < 0.1$, ${}^{*}p < 0.05$, ${}^{**}p < 0.01$, and ${}^{***}p < 0.001$.

Since we focus on the ED medical patients in our structural model, we estimate (OS.15) and (OS.16) using the sample of ED medical patients admitted to the ICU. The estimation results are reported in Table OS.2. The coefficients of ICU occupancy level are statistically insignificant for all three outcomes. This holds for both the ICU occupancy level at the time of admission (Panel A) and the average ICU occupancy level during a patient's stay (Panel B). We also note that the magnitudes of the estimated coefficients are small across all three outcomes and two specifications, except for LOS in Panel A, where the impact of ICU occupancy on hospital LOS is negative–i.e., higher ICU occupancy is possibly associated with shorter hospital LOS. These results suggest that, even if we had larger sample size, it is still unlikely that we would find higher ICU occupancy associated with worse patient outcomes. Therefore, we do not find empirical evidence that ICU congestion leads to worse patient outcomes in our study cohort.

As expected, the LAPS2 severity score is significantly positive for all three outcomes. In addition, we find that the number of patients in the ED at the time of ICU admission is associated with a higher readmission rate and a longer hospital length of stay. The latter effect is statistically significant only at the 10% level, and only when we control for the average ICU occupancy during a patient's stay (Panel B). One possible explanation for the ED occupancy effect is that higher ED occupancy may delay the initiation of care due to longer boarding time. In addition, the care quality in the ED may be more sensitive to patient occupancy than that in the ICU, as the ICU typically operates under a fixed nurse-to-bed ratio.

OS.4. Dynamic Model with Intraday Time Trend

In this section, we extend our dynamic discrete choice model for the ICU admission problem by including a time indicator in the state. This captures the time trend in the system transition during a day, e.g., the arrival and departure rates may vary substantially during day time and night time.

The model is set up as follows. Recall our data consists of state-action pairs for each two-hour interval. We now include a time indicator $m_t \in \{1, 2, ..., 12\}$ in the state variable, which denotes the specific time (two-hour interval) of the day. For example, $m_t = 1$ means the time of the day is from 12AM to 2AM. Thus, our new state variable becomes $s_t = (m_t, n_{l,t}^G, n_{h,t}^G, n_t^F)$. Naturally, the transition of m_t is given by

$$m_{t+1} = \begin{cases} m_t + 1, & \text{if } m_t < 12 \\ 1, & \text{if } m_t = 12. \end{cases}$$

To capture the time trend in the system state, we estimate the ED arrival rates ($\lambda_{Q,l}$ and $\lambda_{Q,h}$), external ICU arrival rate (λ_E), and ICU departure rate (μ_I) for each of the 12 two-hour time intervals separately. That is, we now estimate $\{\lambda_{Q,l}^{(m)}\}$, $\{\lambda_{Q,h}^{(m)}\}$, and $\{\lambda_E^{(m)}\}$ for $m = 1, 2, ..., 12$ following the procedures in Section EC.2.2. This accounts for the intra-day variation in the system state transition. In other words, we allow the state transition probability $g(s'|s)$ of $(n_{l,t}^G, n_{h,t}^G, n_t^F)$ derived in Section EC.1.1 to depend on the time interval m_t using the estimated arrival and departure rates for the specific time period. The ED and ICU capacities are still fixed as they do not vary over time.

We estimate the new model using our data. However, including the time indicator increases our state space by 12 times. To reduce computational burden, we now drop the buffer term in the ED capacity (see Section EC.2.2) and set it for each severity class as the maximum number of patients observed in the data. This has limited impact on the estimation since, in most cases, there are usually only one or two patients in the ED. Like our main specification, we consider a discrete grid $\{0.1, 0.2, ..., 0.9\}$ for potential discount factor levels.

The estimation results for the time-trend model are reported in Table OS.3 for all hospitals combined and Table OS.12 for individual hospitals. The $\hat{\beta}$ column reports the estimated discount factor, and the four subsequent columns ($\hat{u}_{w,l}$, $\hat{u}_{r,l}$, $\hat{u}_{w,h}$, $\hat{u}_{r,h}$) report the estimated utility parameters for the low and high severity classes. By Table OS.3, we see that the estimated model parameters under the time-trend model are very similar to those in our main specification. In particular, the estimated discount factor $\hat{\beta}$ moves from 0.3 to the adjacent level of 0.2. The utility estimates are also very similar.

Table OS.3 Comparison of Main Model and Time-trend Model: All Hospitals Combined

	$\hat{\beta}$	$\hat{u}_{w,l}$	$\hat{u}_{r,l}$	$\hat{u}_{w,h}$	$\hat{u}_{r,h}$
Main specification	0.3	-0.071	1.950	-0.932	0.674
Time-trend model	0.2	-0.003	1.887	-0.839	0.680

In Table OS.12 for individual hospitals, the last "$\hat{\beta}$ (main)" column additionally reports the estimated discount factor in our main setting without time trend. We see that the estimated discount factors are very similar for most hospitals between the two specifications. In particular, the estimated discount factors are exactly the same for 15 out of the 22 hospitals and changes to adjacent levels for 5 hospitals. Only for two hospitals (Hospital 7 and 11), the estimated $\hat{\beta}$ experiences a larger change after the time trend is incorporated. For these hospitals, we have a smaller sample size and thus less variation. This explains the change in the discount factor once we incorporate the time trend. The above results suggest that our main findings are robust to including the time trend in the dynamic discrete choice model.

OS.5. Dynamic Model with Three Patient Classes

To further capture the heterogeneity in patient's utility coefficients, we estimate a dynamic structure model with three patient classes. The three severity classes include patients with severity scores (LAPS2 score) below the 60th percentile, between the 60th and 85th percentile, and those above the 85th percentile, respectively. We normalize the ICU admission utility to zero for all three classes and allow them to have different utility parameters for ED waiting and non-ICU admission. The arrival rates and ED capacity of the three classes are estimated directly from data as in our main set-up. Thus, we have in total seven parameters to estimate in the identification, including the discount factor, and the waiting and non-ICU admission utilities for the three classes. The discount factor is still estimated on the grid $\{0.1, 0.2, ..., 0.9\}$. To control the state space dimension, we drop the buffer term in the ED capacity and set it for each class as the maximum number of patients in the data.

The estimation results for three-class model are reported in Table OS.4 for all hospitals combined and Table OS.13 for individual hospitals. By Table OS.4, we see that our estimation results remain very robust when using the three-class model. The identified discount factor is still 0.3. The estimated utility parameters are also similar. Recall that the high severity class is defined for the same patient cohort in the main model and the three-class model (the patients with LAPS2 above the 85th percentile). We see that their utility estimates $u_{w,h}$ and $u_{r,h}$ are indeed very close in the two models (-0.972 vs. -0.950 for $u_{w,l}$ and 0.674 vs 0.684 for $u_{r,h}$).

Table OS.4 Comparison of Main Model and Three-class Model: All Hospitals Combined

Main specification					Three-class model						
$\hat{\beta}$	$\hat{u}_{w,l}$	$\hat{u}_{r,l}$	$\hat{u}_{w,h}$	$\hat{u}_{r,h}$	$\hat{\beta}$	$\hat{u}_{w,l}$	$\hat{u}_{r,l}$	$\hat{u}_{w,m}$	$\hat{u}_{r,m}$	$\hat{u}_{w,h}$	$\hat{u}_{r,h}$
0.3	-0.071	1.95	-0.932	0.674	0.3	-0.001	2.299	-0.418	1.794	-0.95	0.684

Table OS.13 reports the estimation results with three patient classes for individual hospitals. The second column reports the dimensions of state space in our three-class model, which are indeed relatively large.

The third column ($\hat{\beta}_{3c}$) reports the estimated discount factor in the three-class model. The subsequent three columns ($\hat{u}_{w,1}$, $\hat{u}_{w,2}$, and $\hat{u}_{w,3}$) report the estimated ED waiting utilities for the low-, medium-, and high-severity patients. The next three columns ($\hat{u}_{r,1}$, $\hat{u}_{r,2}$, and $\hat{u}_{r,3}$) report the estimated non-ICU admission utilities for the three classes. In the last column, we report the estimated discount factor from the two-class model as in our main manuscript.

We have the following observations about the results. First, the estimated discount factors remain similar to the main specification with two patient classes for most of the individual hospitals. The absolute difference between $\hat{\beta}_{2c}$ and $\hat{\beta}_{3c}$ is zero for 13 hospitals, and 0.1 for 8 hospitals This suggests that our main identification and estimation results are robust under the three-class model. Second, we find that for most hospitals, the estimated ED waiting and non-ICU admission utilities are also similar to those in the main specification and are decreasing in patient severity level. This also makes sense as more severe patients on average benefit more from the ICU admission (recall the ICU admission utilities are normalized to zero for all three classes). It also suggests that our main model with two classes of patients captures most of the patient heterogeneity in the sample.

OS.6. Supplementary Tables and Figures

Figure OS.2 Probability of ICU admission by ICU occupancy rates (selected hospitals)

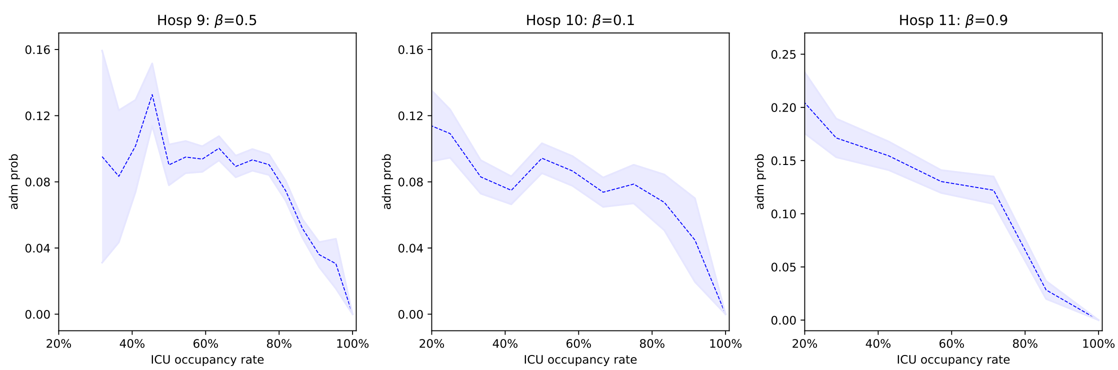


Table OS.5 Summary Statistics of ED Medical Patient Characteristics

Final study cohort: N=164,167

	Min	Max	Mean	Median	SD
Male			0.53		
Age (years)	18.00	113.00	67.27	70.00	17.59
LAPS2	0.00	294.00	74.11	70.00	37.47
EDWait (hours)	0.02	12.00	1.30	0.88	1.41

ICU admission cohort: N=19,683

	Min	Max	Mean	Median	SD
Male			0.48		
Age	18.00	111.00	64.52	67.00	17.48
LAPS2	0.00	294.00	105.03	102.00	45.98
EDWait	0.02	11.98	1.36	0.90	1.45

Note: LAPS2 is severity of illness score. EDWait corresponds to the ED boarding time

Figure OS.3 Utility Estimates for Each β' in the Counterexample

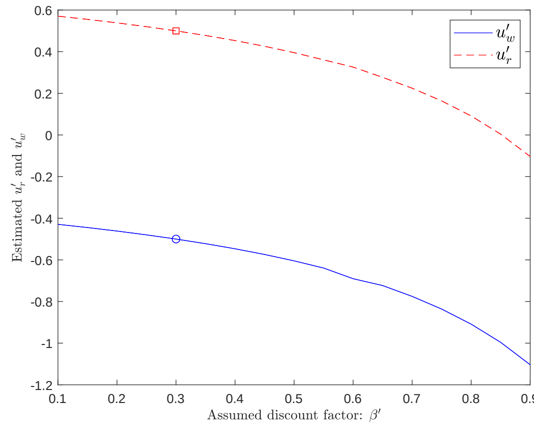


Figure OS.4 Log Likelihood at Different β for Select ED States (All Hospitals Combined)

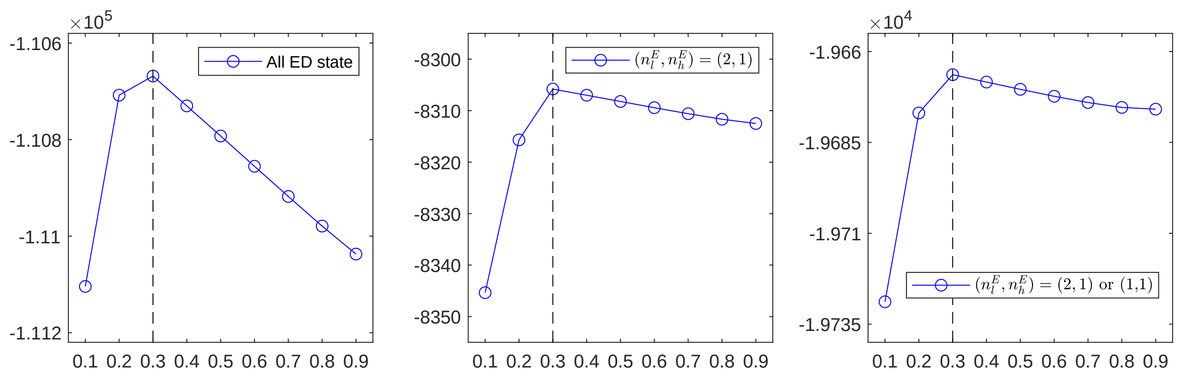
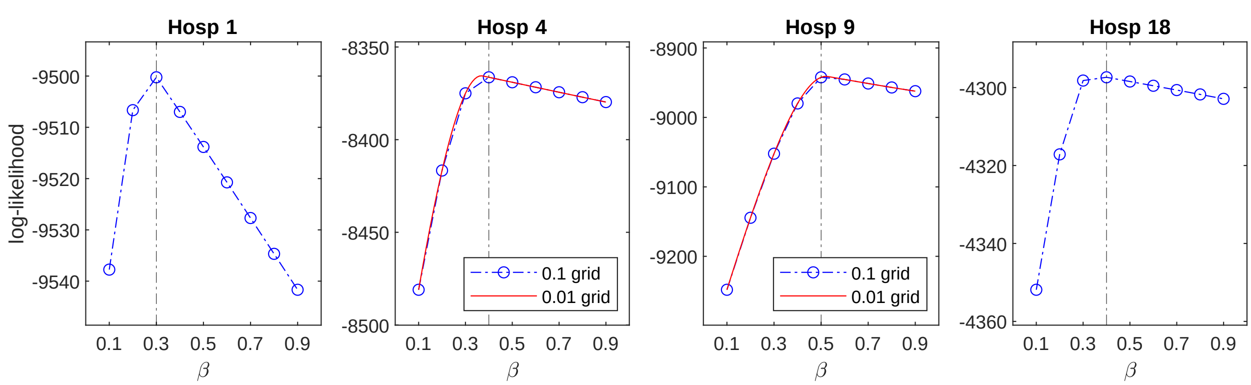


Figure OS.5 Examples of Log-likelihood versus Discount Factor for a Subset of Hospitals



Note: The estimated log-likelihood at $\beta = \{0.1, 0.2, \ldots, 0.9\}$ for Hospitals 1, 4, 9, and 18. The β with the best likelihood is plotted by the black vertical line. The red solid line shows the likelihood with a 0.01 grid for Hospitals 4 and 9.

Table OS.6 System Summary Statistics by Hospital

Hosp	Num. of days	Num. of hospitalizations	Q_l	Q_h	B	ICUOccu	$\lambda_{Q,l}$	$\lambda_{Q,h}$	λ_E	μ_I	$\Pr(a_l)$	$\Pr(a_h)$
1	667	11,676	11	5	21	0.67	1.230	0.221	0.252	0.035	0.12	0.41
2	500	9,902	14	7	26	0.76	1.429	0.227	0.268	0.026	0.11	0.36
3	667	8,039	8	4	12	0.49	0.859	0.146	0.101	0.030	0.05	0.22
4	667	14,595	13	7	31	0.71	1.529	0.327	0.519	0.031	0.05	0.29
5	576	5,082	7	3	11	0.65	0.633	0.098	0.108	0.030	0.11	0.41
6	667	10,577	9	5	21	0.58	1.123	0.208	0.278	0.036	0.08	0.36
7	578	4,915	7	4	11	0.71	0.583	0.111	0.188	0.030	0.05	0.18
8	653	8,400	10	6	16	0.67	0.868	0.185	0.158	0.033	0.14	0.40
9	514	12,355	15	5	22	0.71	1.807	0.264	0.354	0.036	0.07	0.31
10	609	5,978	8	4	12	0.52	0.714	0.092	0.228	0.048	0.06	0.28
11	667	2,655	6	4	7	0.55	0.284	0.045	0.056	0.029	0.12	0.46
12	388	6,751	12	6	24	0.69	1.256	0.173	0.299	0.026	0.06	0.33
13	667	8,061	9	4	16	0.50	0.817	0.182	0.110	0.028	0.07	0.30
14	667	12,841	11	6	36	0.72	1.332	0.257	0.493	0.026	0.07	0.33
15	575	7,208	8	4	16	0.43	0.897	0.111	0.114	0.031	0.07	0.31
16	667	7,190	8	4	13	0.44	0.752	0.133	0.108	0.033	0.06	0.26
17	548	3,511	7	4	9	0.62	0.418	0.099	0.066	0.024	0.09	0.32
18	667	7,702	9	6	32	0.58	0.796	0.122	0.557	0.034	0.06	0.26
19	547	7,476	8	4	25	0.46	0.966	0.170	0.196	0.031	0.10	0.37
20	666	5,096	8	4	11	0.34	0.545	0.087	0.065	0.035	0.07	0.29
21	333	2,109	8	4	13	0.68	0.451	0.076	0.139	0.025	0.10	0.44
22	333	2,048	8	4	16	0.62	0.422	0.090	0.182	0.028	0.15	0.45

Note: System summary statistics for each hospital: Q_i for $i \in \{l,h\}$ is the ED capacity for the two classes of patients; B is the ICU capacity; ICUOccu is the average ICU occupancy level; $\lambda_{Q,i}$ for $i \in \{l,h\}$ is the ED arrival rate; λ_E is the external arrival rate to ICU; μ_I is the ICU departure rate; $\Pr(a_i)$ for $i \in \{l,h\}$ is the overall ICU admission probability for the ED patients.

Table OS.7 Hospital Characteristics Considered in the Heterogeneity Analysis

Size	ICU size, ED size for low and high severity classes
Arrival rates	ED arrival rate for low and high severity classes, external ICU arrival rate
Departure rate	ICU departure rate
ICU congestion	Average ICU occupancy (as % of capacity), median ICU occupancy (as % of capacity), % of periods with fewer than one or two empty ICU beds
Variation in system	Standard deviations of ED arrivals, ICU external arrivals, and ICU occupancy across periods
Patient characteristics	average age, average LAPS2 score, average ED waiting time for two classes, % of high severity class patients, % of ICU admissions from external arrivals
Utility parameters	waiting and non-ICU admission utilities for both classes

Table OS.8 Counterfactual Estimates of Impact when β Increases from $\hat{\beta}$ to 0.9 (All Hospitals)

Hosp	$\hat{\beta}$	ΔDays HighCgstn (in days)			ΔPats HighCgstn (in patients)			ΔPats Balk (in patients)			Increase in ED Time (in hours)	
		β to 0.9	$B+1$	Heuris	β to 0.9	$B+1$	Heuris	β to 0.9	$B+1$	Heuris	Low Class	High Class
1	0.3	4.19	6.27	3.38	34.94	52.26	28.20	4.68	6.73	4.27	2.05	0.63
2	0.5	6.04	6.78	6.44	47.03	52.78	50.12	7.53	8.43	8.73	1.84	0.63
3	0.1	1.75	6.06	2.26	6.89	23.82	8.88	0.81	2.11	0.77	1.59	0.29
4	0.4	1.60	4.22	1.17	17.89	47.08	13.11	3.95	8.63	2.81	1.33	0.42
5	0.1	6.79	18.74	5.24	24.67	68.04	19.02	4.55	8.84	3.11	1.75	0.44
6	0.2	1.15	2.73	1.36	10.07	23.77	11.83	1.27	2.42	1.52	1.70	0.50
7	0.9	-0.45	23.23	0.01	-1.62	84.04	0.04	-0.71	21.20	1.33	0.00	0.00
8	0.1	5.58	10.90	5.45	33.21	64.88	32.41	4.69	7.92	4.62	1.36	0.76
9	0.5	4.27	8.60	6.37	38.99	78.61	58.22	6.42	13.57	10.56	1.55	0.49
10	0.1	0.81	8.23	1.09	5.10	51.80	6.87	0.81	7.56	1.51	0.73	0.26
11	0.9	-0.11	26.98	2.79	-0.24	56.65	5.86	-0.04	6.61	0.54	0.00	0.00
12	0.6	2.14	6.06	2.96	15.20	43.17	21.05	3.35	7.67	4.07	1.15	0.71
13	0.1	0.96	2.34	0.79	4.79	11.65	3.92	0.50	0.95	0.32	2.08	0.38
14	0.3	1.58	2.36	1.73	17.43	25.90	19.07	3.26	5.26	3.54	1.26	0.46
15	0.1	0.22	0.85	0.11	1.25	4.77	0.64	0.13	0.30	0.00	0.37	0.11
16	0.1	0.65	2.52	0.95	3.09	12.07	4.56	0.29	0.95	0.52	1.48	0.44
17	0.9	0.35	21.54	0.65	0.82	50.58	1.52	-0.03	6.60	0.35	0.00	0.00
18	0.4	0.08	0.65	0.16	1.07	8.24	2.05	0.23	1.16	0.62	1.42	0.28
19	0.1	0.05	0.08	0.02	0.40	0.75	0.20	0.04	0.09	0.05	0.45	0.16
20	0.3	0.25	0.94	0.06	1.05	3.91	0.26	0.02	0.15	0.08	1.90	0.33
21	0.7	1.90	16.61	2.71	6.77	59.02	9.62	1.86	11.97	2.59	1.33	0.49
22	0.3	1.12	8.27	1.35	9.75	42.31	6.91	2.43	7.59	2.29	3.23	0.47

Table OS.9 **Comparison of Identified Discount Factor $\hat{\beta}$ and Coefficient of ICU Occupancy γ_{ICU}**

Hosp	Discount factor $\hat{\beta}$ Structural model	Coefficient γ_{ICU} Full logit model	Coefficient γ_{ICU} Simplified logit model
1	0.3	-0.372^{**}	$-0.162^{\dagger}$
	(0.012)	(0.114)	(0.089)
2	0.5	-0.642^{***}	-0.445^{***}
	(0.010)	(0.116)	(0.099)
3	0.1	-0.191	-0.175
	(0.019)	(0.193)	(0.153)
4	0.4	-0.344^{*}	-0.020
	(0.010)	(0.140)	(0.105)
5	0.1	-0.782^{***}	-0.488^{***}
	(0.023)	(0.183)	(0.140)
6	0.2	$-0.270^{\dagger}$	0.144
	(0.014)	(0.146)	(0.107)
7	0.9	-1.489^{***}	-1.170^{***}
	(0.016)	(0.230)	(0.208)
8	0.1	-0.591^{***}	-0.404^{***}
	(0.020)	(0.117)	(0.099)
9	0.5	-0.419^{***}	-0.258^{**}
	(0.009)	(0.126)	(0.104)
10	0.1	-0.434	$-0.275^{\dagger}$
	(0.027)	(0.201)	(0.166)
11	0.9	-0.969^{***}	-0.956^{***}
	(0.017)	(0.216)	(0.185)
12	0.6	-1.025^{***}	-0.508^{***}
	(0.010)	(0.219)	(0.149)
13	0.1	-0.214	-0.06
	(0.018)	(0.184)	(0.131)
14	0.3	-0.269^{*}	-0.127
	(0.012)	(0.132)	(0.100)
15	0.1	-0.083	-0.035
	(0.033)	(0.182)	(0.145)
16	0.1	-0.393^{*}	-0.227
	(0.020)	(0.171)	(0.147)
17	0.9	-1.444^{***}	-0.922^{***}
	(0.021)	(0.237)	(0.176)
18	0.4	0.001	0.104
	(0.013)	(0.218)	(0.159)
19	0.1	-0.341^{*}	0.014
	(0.029)	(0.172)	(0.122)
20	0.3	-0.744^{***}	-0.494^{**}
	(0.017)	(0.208)	(0.165)
21	0.7	-1.091^{***}	-1.023^{***}
	(0.012)	(0.253)	(0.215)
22	0.3	-0.751^{**}	$-0.347^{\dagger}$
	(0.025)	(0.276)	(0.195)

Table OS.10 Simulation Study with the Correct State Space $\Pi(s_t)$ and State-Independent State Space $\Pi(\bar{s})$

	Hospital 2			Hospital 13			Hospital 22		
	True value	Correct action set	State-independent action set $\Pi(\bar{s})$	True value	Correct action set	State-independent action set $\Pi(\bar{s})$	True value	Correct action set	State-independent action set $\Pi(\bar{s})$
$\hat{\beta}$	0.300	0.380	0.900	0.500	0.566	0.900	0.900	0.834	0.900
		(0.007)	(0.000)		(0.008)	(0.000)		(0.010)	(0.000)
$\hat{u}_{w,l}$	-0.015	-0.101	0.000	0.000	-0.124	0.000	-1.669	-1.570	0.000
		(0.012)	(0.000)		(0.014)	(0.000)		(0.022)	(0.000)
$\hat{u}_{r,l}$	1.489	1.486	0.248	1.745	1.742	0.364	2.065	2.074	0.010
		(0.011)	(0.007)		(0.001)	(0.007)		(0.003)	(0.005)
$\hat{u}_{w,h}$	-0.750	-0.798	0.000	-0.789	-0.873	0.000	-1.530	-1.475	0.000
		(0.008)	(0.000)		(0.010)	(0.000)		(0.012)	(0.000)
$\hat{u}_{r,h}$	0.301	0.299	-0.863	0.707	0.707	-0.533	0.660	0.665	-1.177
		(0.002)	(0.002)		(0.002)	(0.003)		(0.004)	(0.002)

Table OS.11 Estimated Discount Factor $\hat{\beta}$ Under Alternative Normalization Choices of $(u_{a.l}, u_{a,h})$

Hosp	(0.3,0.3)	(0.2,0.2)	(0.2,0.3)	(0.1,0.2)	(0.1,0.1)	(0,0)	(-0.1,-0.1)	(-0.2,-0.1)	(-0.2,-0.2)	(-0.3,-0.2)	(-0.3,-0.3)
1	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
2	0.4	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6
3	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
4	0.3	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4	0.4
5	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
6	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2	0.2
7	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
8	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
9	0.5	0.5	0.5	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6
10	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
11	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
12	0.5	0.5	0.5	0.6	0.6	0.6	0.6	0.6	0.6	0.7	0.7
13	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
14	0.2	0.2	0.2	0.2	0.2	0.3	0.3	0.3	0.3	0.3	0.3
15	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
16	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
17	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9	0.9
18	0.3	0.3	0.3	0.3	0.3	0.4	0.4	0.4	0.4	0.4	0.4
19	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1	0.1
20	0.2	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3
21	0.6	0.6	0.6	0.6	0.6	0.7	0.7	0.7	0.7	0.8	0.8
22	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.3	0.4	0.4

Table OS.12 Estimation Results of Dynamic Discrete Choice Model with Time Trend

Hosp	$\hat{\beta}$	$\hat{u}_{w,l}$	$\hat{u}_{r,l}$	$\hat{u}_{w,h}$	$\hat{u}_{r,h}$	$\hat{\beta}$ (main)
1	0.2	0.000	1.445	-0.692	0.304	0.3
2	0.5	-0.023	1.486	-0.682	0.522	0.5
3	0.1	-0.169	2.568	-1.562	1.188	0.1
4	0.4	-0.099	2.322	-1.024	0.839	0.4
5	0.1	-0.223	1.834	-1.209	0.290	0.1
6	0.2	-0.088	1.907	-1.011	0.528	0.2
7	0.6	-0.101	2.731	-0.752	1.605	0.9
8	0.1	-0.277	1.465	-0.794	0.387	0.1
9	0.5	-0.001	1.808	-0.843	0.716	0.5
10	0.1	-0.620	2.350	-1.726	0.932	0.1
11	0.4	-0.056	1.846	-0.908	0.129	0.9
12	0.6	-0.134	2.139	-0.559	0.640	0.6
13	0.1	-0.074	2.235	-1.292	0.810	0.1
14	0.2	0.000	1.930	-0.853	0.673	0.3
15	0.1	-1.105	2.160	-2.280	0.824	0.1
16	0.1	-0.194	2.334	-1.221	1.012	0.1
17	0.9	-1.701	2.070	-1.540	0.660	0.9
18	0.3	-0.001	2.336	-1.244	1.008	0.4
19	0.1	-0.911	1.820	-1.888	0.513	0.1
20	0.2	0.000	2.209	-1.238	0.906	0.3
21	0.6	0.000	1.980	-0.489	0.247	0.7
22	0.3	-0.033	1.546	-1.037	0.119	0.3

Table OS.13 Estimation Results for Individual Hospitals with Three Severity Classes

Hosp	Dim. of state	$\hat{\beta}_{3c}$	$\hat{u}_{w,1}$	$\hat{u}_{w,2}$	$\hat{u}_{w,3}$	$\hat{u}_{r,1}$	$\hat{u}_{r,2}$	$\hat{u}_{r,3}$	$\hat{\beta}_{2c}$ (two class)
2	11616	0.3	0.000	-0.344	-0.750	1.699	1.459	0.301	0.3
3	18144	0.5	0.000	-0.251	-0.659	1.742	1.401	0.518	0.5
5	4680	0.2	-0.229	-0.654	-1.709	3.045	2.250	1.186	0.1
7	22528	0.5	-0.180	-0.733	-1.064	2.870	2.209	0.798	0.4
8	1920	0.2	-0.161	-0.805	-1.244	2.216	1.301	0.333	0.1
9	8316	0.3	-0.130	-0.636	-1.076	2.342	1.715	0.530	0.2
10	2400	0.9	-1.794	-2.032	-1.875	2.881	2.689	1.454	0.9
11	7140	0.1	-0.191	-0.383	-0.763	1.751	1.278	0.380	0.1
13	13248	0.6	-0.068	-0.570	-0.912	2.243	1.658	0.706	0.5
14	3640	0.9	-2.563	-3.034	-2.667	2.740	1.898	0.920	0.1
15	1400	0.9	-1.134	-1.275	-1.366	2.144	1.272	0.105	0.9
16	13475	0.6	-0.061	-0.388	-0.558	2.595	1.897	0.637	0.6
17	4080	0.2	-0.131	-0.545	-1.379	2.641	2.052	0.785	0.1
19	19943	0.2	0.000	0.000	-0.840	2.181	2.006	0.662	0.3
20	3400	0.1	-0.784	-1.850	-2.276	2.531	1.881	0.761	0.1
21	3360	0.1	0.000	-0.621	-1.233	2.688	2.041	1.005	0.1
22	1750	0.9	-1.812	-1.649	-1.530	2.427	1.732	0.660	0.9
24	10395	0.4	-0.021	-0.597	-1.392	2.710	2.106	1.006	0.4
25	8190	0.1	-0.615	-1.447	-1.885	2.202	1.574	0.516	0.1
27	2400	0.3	-0.010	-0.372	-1.333	2.569	1.952	0.861	0.3
29	3780	0.7	-0.104	-0.305	-0.612	2.379	1.594	0.241	0.7
30	3060	0.4	-0.118	-0.309	-1.073	1.865	1.205	0.174	0.3

Table OS.14 Estimation Results Under Changes in Capacity Choices (All Hospitals Combined)

	$\hat{\beta}$	$\hat{u}_{w,l}$	$\hat{u}_{r,l}$	$\hat{u}_{w,h}$	$\hat{u}_{r,h}$
Main Specification	0.3	−0.071	1.950	−0.932	0.671
Increase ICU Capacity by One Bed	0.3	−0.066	1.956	−0.928	0.677
Decrease ICU Capacity by One Bed	0.3	−0.090	1.931	−0.958	0.651
Remove Buffer Term In ED Capacities	0.3	−0.070	1.949	−0.931	0.672

Table OS.15 Estimated Discount Factor $\hat{\beta}$ Under Changes in Capacity Choices

Hosp	Main Specification	Increase ICU Capacity by One Bed	Decrease ICU Capacity by One Bed	Remove Buffer Term in ED Capacity
1	0.3	0.3	0.3	0.3
2	0.5	0.5	0.5	0.5
3	0.1	0.1	0.1	0.1
4	0.4	0.4	0.4	0.4
5	0.1	0.1	0.1	0.1
6	0.2	0.2	0.2	0.2
7	0.9	0.9	0.3	0.9
8	0.1	0.1	0.1	0.1
9	0.5	0.5	0.5	0.5
10	0.1	0.1	0.1	0.1
11	0.9	0.4	0.3	0.9
12	0.6	0.6	0.6	0.6
13	0.1	0.1	0.1	0.1
14	0.3	0.3	0.2	0.3
15	0.1	0.1	0.1	0.1
16	0.1	0.1	0.1	0.1
17	0.9	0.2	0.1	0.9
18	0.4	0.4	0.4	0.4
19	0.1	0.1	0.1	0.1
20	0.3	0.3	0.3	0.3
21	0.7	0.7	0.7	0.7
22	0.3	0.3	0.3	0.3

Table OS.16 Summary of Model Assumptions and Justifications/Robustness Checks

Assumption	Summary	Justification/Robustness checks
Linear utility function	Linearity is critical to achieve joint identification of discount factor and utility parameters (Komarova et al. 2018). It cannot be directly tested from the data.	It is a reasonable assumption as per period utility should be the sum of action utilities for each patient. In addition, we find in the data that the number of patients in ED is small for most periods (Table OS.18) and the ED waiting time is short for most patients (Table OS.17). We also find that the ICU departure rates are generally similar for different ICU occupancy levels (Table OS.21).
Constant utility parameters	We assume the utility parameters are constant and independent of the state. This is essential crucial for our identification and interpretation of the discount factor in our model. This assumption can be validated empirically given sufficient variation in the data.	The ICU typically allocates dedicated resource to each bed and maintains a fixed nurse-to-bed ratio, thus the benefit of ICU care tends to remain stable even when the ICU is congested. Consistent with this, our reduced-form analysis finds no significant impact of ICU occupancy on patient outcomes (see Table OS.2 in Section OS.3).
Normalization choices	Normalization assumption is needed to identify dynamic discrete choice model (e.g., Magnac and Thesmar 2002). Different normalization choices can affect our estimates of discount factor and utility parameters. The normalization choice cannot be pinned down from the data.	We set the ICU admission utility to zero as (i) it provides a proper reference point for current period utility across ED states; (ii) the ICU has very high fixed and operating costs. We perform robustness checks with stratified sample and alternative normalization choices (see Section EC.7).
Lower ED waiting utility	This assumption is critical in our problem as it rules out unrealistic model estimates (long-term benefit from ED waiting). It cannot be directly tested from the data.	ED waiting represents an additional cost in the patient rerouting process, and is a only temporary state instead of final destination. Thus, on average, it brings lower perceived utility than ICU and non-ICU admission. We find in the data that the ED waiting time is indeed short for most patients (see Table OS.17).
Patient classes	This assumption facilitates our estimation by reducing the state space dimension, which is important as our state space grows exponentially in number of patient classes. It is innocuous as two patient classes are sufficient to capture the difference in patients' needs for care.	We perform robustness checks with three patient classes in Section OS.5. The estimated discount factors and heterogeneity across hospitals are similar to our main specification.
Constant arrival/departure rates	These set-ups are standard in dynamic discrete choice models. The arrival/departure rates are estimated from data.	We perform robustness checks using time-varying transition rates (Section OS.4) and stratified sample analysis (Section EC.7).
ED and ICU capacity	These set-ups are standard in dynamic discrete choice models. The units capacity levels are directly estimated from the data.	The ICU/ED capacity levels are directly estimated from data. We perform robustness checks by varying capacity levels of ED and ICU (Section EC.7).
Unlimited capacity for non-ICU units	This assumption facilitates our estimation by reducing the state space dimension. It can be validated from the observed data.	We observe in the data that the occupancy of non-ICU units seldom reaches their maximum observed level (see footnote 2). Thus, the capacity of non-ICU units should not be a constraint. This is also confirmed in our communication with hospitals.

Table OS.17 Statistics of Patients ED Length-of-Stay

	Full cohort (N = 180051)			ICU admission cohort (N = 20977)		
ED LOS	$\leq$ 2 hours	$\leq$ 4 hours	$\leq$ 6 hours	$\leq$ 2 hours	$\leq$ 4 hours	$\leq$ 6 hours
Proportion	85.40%	94.10%	96.60%	82.40%	93.80%	97.00%

Table OS.18 Summary Statistics of ED Occupancy

	25th percentile	Median	Mean	75th percentile	90th percentile
Low severity class	0.00	1.00	1.02	2.00	3.00
High severity class	0.00	0.00	0.19	0.00	1.00

Table OS.19 Summary Statistics of ICU Length-of-Stay (Hours)

Severity Class	Count	Mean	Std
Low	11490	45.09	48.33
High	8189	66.04	59.98

Table OS.20 Variance-to-mean ratio for arrivals in each periods

Hosp	ED low severity	ED high severity	External ICU
1	1.066	0.985	1.047
2	1.259	1.025	1.072
3	1.025	1.009	1.050
4	1.134	0.999	1.244
5	1.090	1.011	1.038
6	1.085	1.029	1.056
7	1.054	1.012	1.049
8	1.091	0.987	1.043
9	1.083	1.010	1.039
10	1.110	0.994	1.015
11	1.002	1.010	1.016
12	1.316	1.051	1.048
13	1.054	1.009	1.026
14	1.095	0.965	1.175
15	1.077	0.994	1.079
16	1.100	1.012	1.022
17	0.970	0.975	1.020
18	1.171	1.069	1.391
19	1.019	1.003	1.059
20	1.007	0.990	0.996
21	1.003	0.997	1.076
22	1.037	0.976	1.033

Table OS.21 ICU Departure Rate by ICU Occupancy Level (All Hospitals Combined)

ICU Occupancy Decile Bin	ICU Departure Rate μ_I
1	0.031
2	0.031
3	0.032
4	0.032
5	0.031
6	0.029
7	0.032
8	0.032
9	0.030
10	0.029

References

Komarova, Tatiana, Fabio Sanches, Daniel Silva Junior, Sorawoot Srisuma. 2018. Joint analysis of the discount factor and payoff parameters in dynamic discrete choice models. *Quantitative Economics* **9**(3) 1153–1194.

Magnac, Thierry, David Thesmar. 2002. Identifying dynamic discrete decision processes. *Econometrica* **70**(2) 801–816.

Rust, John. 1987. Optimal replacement of gmc bus engines: An empirical model of harold zurcher. *Econometrica: Journal of the Econometric Society* 999–1033.