Congestion-Quality Trade-off: Evidence from Nursing Facilities*

Hiroki Saruya† Masaki Takahashi‡

August 29, 2025

Abstract

Directing consumers to higher-quality service providers has been considered an effective policy to improve service outcomes and consumer welfare in various contexts. However, higher-quality providers may tend to be more congested, and congestion may be detrimental to outcomes and welfare. We study this *congestion-quality trade-off* and its policy implications in the context of Japanese nursing facilities. We find that (1) *within nursing facilities*, higher occupancy leads to poorer care outcomes but (2) *between nursing facilities*, occupancy and outcome-based quality measures are positively correlated. To evaluate the welfare impact of patient reallocation policy, we then develop and estimate a model of demand for nursing facility care where choice set is potentially constrained in an unobserved manner by providers' rationing behavior. We find that congested nursing facilities are less likely to admit patients as occupancy increases but no evidence that patients dislike congestion. A simulation of a reallocation policy suggests a potential gain from occupancy smoothing even though the policy sends patients to lower-quality providers on average.

Keywords: congestion, value added, long-term care, demand estimation, choice constraint, consideration set

---

*Hiroki Saruya thanks his advisors, Jason Abaluck, Steven Berry, and Katja Seim, for their guidance and support. We are also grateful to Claudia Allende, Phil Haile, Yoko Ibuka, Jonas Lieber, Victoria Marone, Chris Neilson, Haruko Noguchi, Yuta Takahashi, Yuta Toyama, Yasutora Watanabe, and seminar participants at Asia-Pacific Industrial Organisation Conference 2024, Hitotsubashi University, Keio University, Summer Workshop on Economic Theory 2025, Waseda University, and Yale University for helpful feedback. Masaki Takahashi gratefully acknowledges the support of JSPS Grant-in-Aid for Scientific Research (20H01514).

†Economic and Social Research Institute, Cabinet Office, Government of Japan, and Sophia University. hiroki.saruya@aya.yale.edu

‡Sophia University. msk.tkhs@gmail.com

1 Introduction

Researchers and policymakers take great interest in the quality of service providers as a key determinant of various outcomes. Researchers have documented substantial heterogeneity in provider quality in various settings such as schools, teachers, hospitals, doctors, and nursing facilities,[1] and have suggested that moving service consumers (e.g., students, patients) from a low-quality provider to a high-quality provider will improve service outcomes and welfare.[2] Informational interventions have been discussed as a possible policy tool to achieve such reallocation of consumers and to improve outcomes and welfare.

However, in many industries with capacity constraints, there is another important and possibly conflicting determinant of service outcomes and welfare: congestion. A higher degree of congestion can increase the likelihood of adverse events such as hospital admission and mortality (Hoe, 2022; Gutierrez and Rubli, 2021), and can lower consumer welfare via low service intensity or long wait times.[3] Moreover, congestion and quality considerations may be in conflict. Suppose, for example, that higher-quality service providers are more congested. Then, a policy to reallocate consumers from a low-quality provider to a high-quality provider increases the average congestion for consumers. Thus, such a policy need not be beneficial, and we must understand the trade-off between improved quality and exacerbated congestion to design beneficial policies.

This paper evaluates such a trade-off, which we refer to as *a congestion-quality trade-off*, in the context of Japanese nursing facilities.[4] As in the US skilled nursing facilities, Japanese

[1] See, for example, Kane and Staiger (2008); Chetty et al. (2014a,b) for quality measures for teachers, Deming (2014); Angrist et al. (2017) for schools, Geweke et al. (2003); Chandra et al. (2016); Hull (2020); Chandra et al. (2023) for hospitals, Einav et al. (2025) for nursing facilities, and Abaluck et al. (2021) for health insurance plans. See also Chetty and Hendren (2018) for various effects of neighborhoods.

[2] For example, Einav et al. (2025) document that variation in nursing home value added within geographic markets in the US is comparable to the nationwide variation, and conclude that their finding "points to the potential for substantial gains from policies that encourage reallocation of patients to higher-quality nursing homes within their market." (p1261). Chetty et al. (2014b) illustrate the magnitude of their estimates of teacher quality by evaluating the policy counterfactual of replacing teachers in the bottom 5 percent of the distribution of teacher quality with average teachers.

[3] Since wait times are disliked by consumers, they can serve as a non-price mechanism to allocate services, in industries such as health care (Moscelli et al., 2021; Yee et al., 2022; Russo, 2024) and transportation (Fréchette et al., 2019; Buchholz et al., 2025; Castillo, 2022). See Leshno (2022) for a conceptual framework for allocation mechanisms via waiting lists.

[4] There are two components of quality: (i) congestion-dependent quality (e.g., service intensity) and (ii) quality that is determined prior to congestion (e.g., staff skills or technologies). We focus on the trade-off

nursing facilities are typically highly crowded due to regulations on capital investments along with growing demand from aging population. The resulting high occupancy rates create a congestion problem, due to inflexible adjustments in staffing. The industry also features universal long-term care insurance where provider reimbursement mostly consists of per-diem fixed payments adjusted for patient[5] care needs, which enables us to focus on congestion-quality trade-off and omit other concerns such as the choice of fee-for-service treatments.

Our research proceeds in three steps. In the first step, we document the effects of congestion on care outcomes. A key challenge is that bed occupancy (our measure of congestion) may be endogenous, because higher-quality facilities that can produce more desirable outcomes may attract more patients. To address this problem, we exploit variation in the volume of short-stay patients as an exogenous shifter of occupancy faced by long-stay patients (our research subject). Providers in our setting (i.e., nursing facilities) use their capacity to serve both short-stay patients, who receive community-based care but occasionally use facilities for temporary services, and long-stay patients, who seek in-facility services to restore their physical conditions. Therefore, variation in the volume of the former shifts congestion faced by the latter. Also, short stays typically begin with temporary unavailability of family caregivers or other reasons plausibly unrelated to outcomes of long-stay patients, and end within around two weeks, so their fluctuations are unlikely to be directly related to outcomes of long-stay patients.

Our baseline results suggest that a 1pp increase in the average daily occupancy during an episode of stay leads to a 1.3pp (3.8%) decrease in the probability of discharge to home, a desirable outcome of nursing facility care (Einav et al., 2025), and a 3.1pp (8.3%) increase in the probability of hospitalization. The large effects may arise from rehabilitation and procedural delays, as well as potential deterioration of health conditions due to insufficient treatment.

In the second step, we examine the measure of covariate-adjusted provider quality and its correlation to occupancy. We show that our quality measures unbiasedly predict the outcomes of patients who are admitted to a high-quality vs. low-quality providers due

between the latter quality and congestion, while we regard the former as a mechanism behind the negative effect of congestion.

[5]In this article, we refer to the user of nursing facilities as "patients" rather than "residents."

to arguably random geographic proximity, which implies that our quality measure is not systematically biased by nonrandom patient sorting. Furthermore, we show that our quality measure is positively correlated to occupancy, which suggests the existence of the congestion-quality trade-off.

In the final step, to evaluate the welfare impacts of congestion and counterfactual policies of patient reallocation, we estimate patients' preferences for nursing care providers and providers' admission rule. There are two challenges. First, availability of providers for an admission may be limited due to congestion, but is unobserved. Ignoring such choice restrictions can lead to biased estimates, because availability can be correlated to provider characteristics. For example, we will under-estimate patients' valuation of provider quality, if we ignore a positive correlation between provider quality and availability, because patients end up in a facility with lower-than-desired quality on average. To deal with this difficulty, we assume that short-run fluctuations in occupancy affect a provider's admission rule but do not affect patient preferences, conditional on the provider's average occupancy over a longer time period and other provider characteristics. We then formalize identification of our demand model with such (one-way) exclusion restrictions. The second challenge is, again, the endogeneity of occupancy. We address the endogeneity by adapting an approximate version of the Berry et al. (1995, henceforth BLP) approach, as proposed by Lee and Seo (2015), to our microdata setting. The approximate BLP method relaxes the market-share constraints while updating parameters, which is convenient for estimating demand models with choice constraints.

Our estimates suggest that a nursing care provider is more likely to offer an admission if the applicant lives in the same city as the provider, is female, or is of low care-need level, or if a congested provider becomes less congested. Patients prefer providers which are close to their home and which have higher quality for facilitating home discharge. However, we do not find evidence that they have distaste for congestion. A simulation of patient reallocation to smooth occupancy between the most congested and least congested providers within each market (defined by secondary medical area-quarter) suggests that occupancy smoothing has a net positive effect on home discharge for the median market, even though such reallocation sends patients to lower-quality providers on average. For patient welfare, our result suggests that such reallocation makes reallocated patients worse off.

This paper relates to the literature on the effects of congestion on healthcare outcomes and provider behavior. Hoe (2022) and Gutierrez and Rubli (2021) use admission shocks to find that hospital crowding increases unplanned readmission and in-hospital mortality, respectively. We study longer-term outcomes (home discharge, hospitalization, and mortality outcomes of an episode of length up to 365 days) in a different context (nursing facilities).[6] Another strand of literature has exploited short-run occupancy fluctuations to study providers' admission/discharge decisions in the context of long-term care (Gandhi, 2023; He and Konetzka, 2015; Hackmann et al., 2024; Saruya and Takahashi, 2025) and hospital care (Freedman, 2016; Sharma et al., 2008; Bachner et al., 2024). An advantage of the setting of Japanese nursing facilities is that there is a universal insurance system where reimbursement rate is adjusted to patient care needs, which mitigates selection incentives and facilitates the study of the causal effects of congestion on outcomes. Also, unlike many previous studies, we use an instrumental variable for occupancy and show that it can make a difference even conditional on fixed effects.

This paper also relates to the vast literature on outcome-based institutional quality, often referred to as value added; see footnote 1 for examples. Researchers have estimated institutional quality and further investigated its implications on choice of and competition among schools (Neilson, 2013; Allende, 2019; Abdulkadiroğlu et al., 2020; Beuermann et al., 2022; Ainsworth et al., 2023) and hospitals (Gaynor et al., 2016; Chandra et al., 2016). In contrast, value-added measures for nursing facilities have not been studied extensively, with a few exceptions (Einav et al., 2025; Olenski and Sacher, 2024; Bär et al., 2022; Cheng, 2023). These papers examine relatively short-run outcomes such as 30-day mortality and home discharge, whereas we study patient outcomes on a longer horizon, up to one year.[7]

Finally, this paper relates to the literature on demand estimation under choice constraints. Choice sets of goods and services from which consumers can choose are often restricted,

[6] In the context of India, Andrew and Vera-Hernández (2022) study a large cash-transfer program which incentivizes women to give birth in a health facility and find evidence that congestion induced by the program led to higher perinatal mortality in low-capacity districts, and find suggestive evidence that such effects may have persisted up to five to ten years.

[7] Olenski (2023) uses quality-of-care violations, rather than an outcome-based quality measure, to study the long-run impacts on nursing home patients through provider exits and patient reallocation.

for various reasons such as limited attention (Goeree, 2008; Ho et al., 2017; Abaluck and Adams-Prassl, 2021; Heiss et al., 2021), search frictions (De Los Santos et al., 2012; Honka, 2014), stockouts (Conlon and Mortimer, 2013; Kawaguchi et al., 2021), institutional reasons (Gaynor et al., 2016) and supply-side behavior (Dubois and Sæthre, 2020; Gandhi, 2023; Agarwal and Somaini, 2025). Agarwal and Somaini (2025) provide general identification results for many of these models using two-sided exclusion restrictions, i.e., a variable affecting consumer preferences but not choice sets and another variable affecting choice sets but not consumer preferences. We contribute to this literature by formalizing identification of a (commonly used) class of choice-constraint models using one-sided exclusion restrictions: we assume that some variable affects choice constraints without shifting preferences, but not that another variable shifts preferences without affecting choice constraints.[8] Relaxing two-sided exclusion restrictions is important in our setting, as well as in other settings where choice constraints are generated by the decisions of economic agents. In our empirical setting, providers may prefer admission of patients from their own city, so the commonly used distance instrument as a preference shifter without affecting choice constraints may be invalid. Our estimate confirms this conjecture.

The rest of this paper proceeds as follows. Section 2 describes institutional background and data. In Section 3, we present our empirical framework for and results of the effect of congestion on patient outcomes. Section 4 presents analysis of the quality measures of nursing facilities. Section 5 presents our demand model, and Section 6 provides an identification result and estimation approach. Estimation and simulation results of our demand model are presented in Section 7. Section 8 concludes. Additional results and proofs are found in the Appendix.

[8] In the context of choice under risk, Barseghyan et al. (2021) propose identification of a consideration set model using preference shifters which are excluded from consideration, together with an "identification-at-infinity" type assumption. Abaluck and Adams-Prassl (2021) do not use exclusion restrictions and instead exploit symmetry of demand analogous to the Slutsky symmetry to identify consideration set models.

2 Institutional Background and Data

2.1 Nursing Home Industry in Japan

Japanese nursing facility industry features its public, universal long-term care insurance (LTCI). People over the age of 65 who are certified as in need of long-term care (LTC) services are eligible for LTCI benefits. The care-needs certification is based on an in-person health examination, which evaluates the applicant's physical and mental disability and yields a measure of the degree of their care needs, called a health score. Applicants are eligible for LTCI benefits if their health score exceeds a threshold. Eligible beneficiaries can use both home care and institutional care, typically with a 10% coinsurance rate. Due to the rapid population aging, Japanese public expenditures on LTCI have been increasing. The annual cost of LTCI was 11.5 trillion JPY in the 2023 fiscal year, accounting for about 2% of Japanese GDP (Ministry of Health, Labor and Welfare, 2023b). About one-third of the total costs consist of the cost of institutional care.

We study a type of nursing facilities called Geriatric Health Services Facilities (GHSFs), which we simply refer to as "providers" or "facilities" below.[9] GHSFs are non-profit organizations aimed at providing high-quality inpatient rehabilitation services and restoring the physical capabilities of users so that they can live back in their home or community. In this sense, they are similar to the US Skilled Nursing Facilities (SNFs). As of April 2022, there were 4,230 GHSFs, serving approximately 355,900 patients nationwide (Ministry of Health, Labor and Welfare, 2023a). The entry and changes to the bed capacity of a GHSF require an approval by the prefectural governor.

Patients stay in a GHSF for various objectives. Our analysis focuses on *long-stay* patients, who are institutionalized to receive rehabilitation services with the goal of returning to their community. Other patients, called *short-stay* patients, live in their community but occasionally use the facilities for temporary services. These short-stays services are typically used for a respite or temporary unavailability of family caregivers. The LTCI covers these services only for a short period of time, and over 70% of the short stays end within two weeks of admission (Ministry of Health, Labor and Welfare, 2017). Moreover, GHSFs serve

[9] GHSFs are also known as "Kaigo Roujin Hoken Shisetsu" or "Roken" for short in Japanese.

both short-stay patients and long-stay patients using the same capacity. Therefore, while fluctuations in short stays are likely exogenous to the health conditions of long-stay patients, they create variation in the occupancy faced by long-stay patients. These features motivate us to study the effects of occupancy on long-stay outcomes and exploit short-stay fluctuations to construct an instrument for occupancy.

Various healthcare workers, such as physicians, nurses, and social workers, stay at GHSFs to provide patients with high-quality rehabilitation and nursing care. Due to the rapidly aging population, inflexible wage setting, and harsh working conditions, GHSFs suffer from chronic labor shortages. The survey by Care Work Foundation (2016) indicates that 62.6% of facilities were understaffed and that 73.1% of such facilities responded that the primary cause of labor shortage was hiring difficulties, due to factors such as low wages (57.3%) and demanding jobs (49.6%). Because facilities cannot adjust staffing flexibly in response to an increase in patient volume, a higher occupancy rate causes greater congestion in the facility (see also Saruya and Takahashi, 2025).

2.2 Admission and Discharge Processes

Admission process is initiated by LTCI beneficiaries' application for admission to the facility. Beneficiaries consult with physicians and social workers in the application process. After the application is received by a facility, it conducts an interview with the applicant to evaluate their physical conditions and service necessity. The admission decision is based on the interview and supplementary documents. Once they are admitted in a GHSF, patients receive rehabilitation services following their care plan, which is made in advance by care managers.

Once a patient is on track to be discharged, the facility initiates the discharge process together with the patient and their family, to secure post-discharge LTC and living arrangement. Patients may be discharged to their home if they restore their health and physical conditions sufficiently. In contrast, if a patient's health condition deteriorates and requires acute care, they will need to be transferred to a hospital. Some patients move to another GHSF to continue rehabilitation, or they may move to another type of nursing homes, with an intention to live there forever.

2.3 Reimbursement

GHSF reimbursement consists of per-diem fixed payment and fee-for-service (FFS) payment. The per-diem payment is adjusted to patient care needs, based on a measure called care levels. Care levels classify LTCI beneficiaries into seven groups: support levels 1–2 and care levels 1–5, in ascending order of care needs.[10] The classification is based on the health score mentioned in Section 2.1. In principle, only patients with care level 1 or above are eligible for institutional care in GHSFs. The nursing care burden is reflected in reimbursement by setting higher rates of fixed payment for higher care levels. By contrast, FFS payment is paid to providers for certain medical procedures, such as short-term intensive rehabilitation, dementia care, and terminal care, regardless of care levels.

Table 1 shows per-diem and FFS payment in our analysis sample (defined below), by care levels. As Table 1 suggests, over 90% of reimbursement consists of per-diem payment. In the demand estimation below, we focus on patients' and providers' decisions about admissions, omitting the choices of FFS treatments.

Table 1: Per-diem Reimbursement

	Fixed payment (USD) (1)	FFS payment (USD) (2)	Fraction of fixed payment (3)
Care level 1	78.2	6.4	92.5%
Care level 2	83.4	6.6	92.6%
Care level 3	89.7	7.1	92.7%
Care level 4	95.3	7.1	93.1%
Care level 5	101.3	6.9	93.6%

Notes: The table presents daily averages of fixed and fee-for-service (FFS) payments in our analysis sample, separately by care levels. The averages are computed through the following steps. (1) Compute the daily averages of fixed and FFS payments within each patient-year-month bin, by computing the total fixed and FFS payments within each bin and then dividing them by the total days of stay within the bin. (2) Aggregate the averages to the care level.

[10] Table A1 in Appendix A describes general functional status for each support level and care level.

2.4 Data Sources and Sample Selection

Our main data source is the Survey of Long-Term Care Benefit Expenditures, which contains administrative claims data for LTC service utilization. It provides information on each LTCI recipient's service utilization at the monthly level. For each episode of a GHSF stay, the data contains information on the admission date, discharge date, and discharge destination, if admission and discharge occur within our sample period. It also contains some patient characteristics such as age, sex, care level, and coinsurance rate. We obtain claims for 2011-2017 Japanese fiscal years, i.e., from April 2011 to March 2018. We combine the claims data with the Survey of Institutions and Establishments for Long-Term Care, which provides information on each GHSF's characteristics, such as the number of beds, at the annual level. The combined data sources allow us to compute daily bed occupancy for each facility.

We use these data sources to construct a sample of episodes of GHSF stays that begin and end within our sample period. For each episode, we have information on the anonymized identifier and municipality of the provider, the admission and discharge dates, and the discharge destination. We also compute the daily averages of occupancy and peer patients' characteristics for each episode.

We present summary statistics at the episode level in Table 2. The average age at admission is 85, and 69% of patients are female. The average length of stay (LOS) is 339 days. The LOS varied widely among patients, with the 25th percentile being 60 days but the 75th percentile being 390 days. In the analysis below, we focus on episodes of admissions in a facility with LOS between 14 and 365 days.

Table 2: Summary Statistics (Patients)

	Mean (1)	SD (2)	p25 (3)	p50 (4)	p75 (5)	Obs. (6)
Age	85.07	8.13	81	86	91	1,105,046
Female	0.69	0.46	0	1	1	1,105,046
Length of stay (days)	339	519	60	139	390	1,105,046
Care level 1	323	524	58	126	358	122,751
Care level 2	344	520	61	136	394	207,679
Care level 3	353	525	64	146	411	264,722
Care level 4	332	499	60	142	385	301,078
Care level 5	338	536	52	136	382	208,807

Notes: The table presents summary statistics at the episode level, before restricting the sample for patient outcome analysis. Columns (3), (4) and (5) present 25th, 50th and 75th percentile, respectively.

3 Effects of Occupancy on Patient Outcomes

This section discusses our approach to estimating the effects of occupancy, our congestion measure, on patient outcomes, and then presents the results. Section 3.1 introduces our econometric model. Section 3.2 discusses our identification strategy. We then present our empirical specificaion in Section 3.3 and present the results in Section 3.4 .

3.1 Econometric Model

We model the care outcome of patient i if she is admitted to provider j in period $\tau (= \tau_i)$ as

$$Y_{ij\tau} = \mu_j + \beta n_{j\tau} + x'_{ij\tau}\gamma + \varepsilon_{ij\tau} \tag{1}$$

where $Y_{ij\tau}$ denotes an outcome, μ_j is the provider fixed effect (FE), $n_{j\tau}$ denotes the average occupancy of provider j in period τ, and $x_{ij\tau}$ denotes controls, including the length of stay and other FEs. As outcomes, we use indicators of whether a given episode of stay in the provider ends with (i) discharge to home, (ii) hospitalization, or (iii) death. Given the goal of providers noted in Section 2, home discharge is considered a good outcome, whereas hospitalization and death represent bad outcomes.[11] So far, μ_j only represents the average tendency of provider j to produce a specific outcome, which may include differences in case mix as well as the causal effect of the provider; we discuss its causal interpretation in Sections 4.1 and 4.2.

Model (1) yields the following regression model:

$$Y_{i\tau} = \mu_i + \beta n_{i\tau} + x'_{i\tau}\gamma + \varepsilon_{i\tau} \tag{2}$$

where $Y_{i\tau} = \sum_j Y_{ij\tau} I(j_i = j)$ denotes the realized outcome (with j_i denoting the provider to which i is admitted), and other variables are defined analogously.

[11] Similarly, the literature on nursing facility quality has examined community discharge (Einav et al., 2025), hospitalization (Rahman et al., 2016), and death (Cheng, 2023).

3.2 Identification of the Effects of Occupancy

An identification concern for Eq.(2) is that occupancy rates may be correlated to unobserved determinants of patient outcomes, $\varepsilon_{i\tau}$, in which case the ordinary least squares estimate (OLSE) of β will be a biased estimate of the causal effect of occupancy on patient outcomes. Two major sources of such endogeneity are (i) patient composition which varies with occupancy and (ii) unobserved provider heterogeneity. Endogeneity via patient composition arises if providers facing high occupancy admit/discharge patients that are systematically different from those admitted/discharged when the providers face low occupancy. Endogeneity via unobserved provider heterogeneity occurs if, e.g., providers with higher quality attract more patients.

We mitigate the concern about patient composition by including rich controls. We control for the patient's age, sex, indicator of high coinsurance rate (a proxy for the income of the patient), indicator of terminal care utilization, and care level (measure of care needs) dummies. In addition, we control for the averages of these variables among patients within the same provider and same time period. We also control for the length of stay of the focal patient. Moreover, we include provider FEs, discharge-date FEs and discharge year-month by medical area[12] FEs to control for provider- and time-specific shocks and regional trends in care outcomes, which will capture some portion of provider quality in addition to patient composition. Finally, We control for local hospital capacity, which may affect the choice of discharge destination.

We address the concern about residual unobserved heterogeneity by instrumenting occupancy with the number of short-stay discharges, to study the outcome of long-stay patients. The idea is that demand for short-stay care is unrelated to the quality of long-stay facility services but affects congestion faced by long-stay patients. For exogeneity, we note that demand for short-stay services typically arises from temporary unavailability of family caregivers. These stays typically end within two weeks of admission, as noted in Section 2. Thus, short-stay admissions and discharges are likely arranged independently of the shocks

[12] More specifically, we use the secondary medical area. The market concept segments Japan into about 350 areas and is defined so that general inpatient treatments are expected to be completed within the area. In this sense, our medical area is smaller than the hospital referral region in the US, which segments the US into about 300 regions and represents regional markets for tertiary medical are.

to long-stay patients' outcomes. Although one may worry that some shocks may affect both the number/share of short stays and the composition among long-stay patients, we show in Section 3.4 that characteristics of long-stay patients do not vary much with the number of short-stay (admissions or) discharges.

For relevance, GHSFs use the same bed and staff capacity to serve both short-stay and long-stay patients, and adjusting the capacity in response to such short-run fluctuations is difficult due to capital investment regulation and inflexible labor market. Consequently, fluctuations in short stays affect occupancy faced by long-stay patients. One caveat, however, is that the sign of the effect of increased short-stay admissions on the average occupancy faced by long-stay patients may be heterogeneous. Specifically, in congested facilities, short stays may *decrease* occupancy, because short-stay patients will substitute long-stay patients, the latter of whom would occupy beds for longer periods. By contrast, in less congested facilities, an increase in short-stay admissions may increase occupancy, because short-stay patients will not crowd out long stays. Such heterogeneity may lead to overestimates of the effect of congestion, by making the first-stage coefficient on the short-stay admissions small. In the empirical analysis below, we use short-stay discharges as a main instrument.[13]

Previous studies on the effects of health facilities' occupancy on care outcomes or provider behavior (Gandhi, 2023; Hackmann et al., 2024; Freedman, 2016; Hoe, 2022) similarly motivate the exogeneity assumption on their occupancy measures by noting that short-run fluctuations are perceived exogenous to the outcome of interest. One difference of our work from theirs is that we only exploit variation in short-stay admissions and discharges for identification, whereas previous studies use occupancy variation which arises from both short-stay and long-stay patients, although both of these studies control for rich fixed effects to mitigate endogeneity concerns. In Section 3.4, we show that our two-stage least squares estimates are quite different from OLSEs, which suggests that instrumenting occupancy may be important even conditional on rich controls.

[13]We are currently investigating the heterogeneity in the first stage.

3.3 Empirical Specification

The above model (2) will not fit all types of patients. On one hand, a short period of an exposure to congestion will not affect patient outcomes. On the other hand, some patients utilize a provider with intention to stay there for life. To focus on patients to whom congestion is most relevant, our analysis focuses on patients who stay with a provider for 14 to 365 days with a service code identified with a long stay. We also exclude episodes which occur at a provider with a dementia care unit; for such providers, we cannot observe whether a patient is admitted to the regular unit or a dementia unit, making it hard to identify relevant congestion level. Finally, we exclude providers which are in the bottom 20 percentile or top 5 percentile in the distribution of provider's maximum occupancy during the sample period. The former restriction is intended to exclude providers which are always empty, and to mitigate concerns about missing data.[14] The latter restriction aims to eliminate outliers of occupancy, which may reflect measurement errors.[15]

We estimate Eq.(2) using two specifications of instruments which are constructed using short-stay discharges. As an instrument for the average occupancy during episode, our benchmark analysis uses (i) the average number of discharges of short-stay patients from the provider in which the focal patient stays, where average is taken over days during the episode. To mitigate the concern that short-stay fluctuations directly correlate to long-stay outcomes, we also show regression results which use (ii) the average number of discharges of short-stay patients from the provider, during the 14 days *prior to the start date of the episode, controlling for average short-stay discharges during episode.* We also define and use analogues using short-stay admissions in robustness checks.

We normalize provider fixed effects to have mean zero. Further, we apply the empirical Bayes shrinkage (Morris, 1983; Chandra et al., 2016) to reduce the noise in their estimates.

[14] A concern is that claims from some municipalities are missing because they do not agree to provide the information for research use.

[15] We have verified that relaxing the trimming criteria does not alter qualitative results.

3.4 Results of the Effect of Occupancy

We begin by checking covariate balance across different values of our baseline instrument (episode mean of short-stay discharges) by regressing covariates on the instrument and other controls. We examine patients' age, sex, an indicator of high cost sharing (an income measure), and care level at admission, as well as an indicator of receiving terminal care during the episode and the length of stay. Table 3 shows the results of the regressions. It shows that our instrument is not systematically related to patients' cost sharing, care level, terminal-care status, or length of stay. Although the instrument is correlated to age and sex, the correlation is weak. Table A2 similarly shows that correlation between short-stay admissions and patient characteristics are weak, if any.

Table 3: Covariate Balance

	(1) Age	(2) Female	(3) High cost sharing	(4) Care level	(5) Receiving terminal care	(6) Length of stay
Short-stay discharge in pp	0.1135*	0.00928**	0.00154	-0.00920	0.000976	-0.1556
	(0.06617)	(0.00376)	(0.00146)	(0.01059)	(0.000761)	(0.8773)
Mean outcome	85.11	0.6752	0.0342	3.21	0.0109	121.61
N	599,946	599,946	599,946	599,946	599,946	599,946

*** p<0.01, **p<0.05, *p<0.1

Note: This table presents the result of regression of patient characteristic (indicated by each column) on our instrument (average number of short-stay discharges during episode, expressed as a percentage of capacity) and controls other than the dependent variable. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

Table 4 presents the results of the instrumental variable regression of patient outcomes on occupancy. The first-stage result reported in Column (1) shows that a 1pp increase in the number of short-stay discharges leads to a 0.76pp decrease in average occupancy. The pp change in occupancy is less than the pp change in short-stay discharges, likely because providers increase admissions in response to an occupancy reduction induced by short-stay discharges.

Table 4: Instrumental Variable Regression of Patient Outcomes on Occupancy

	(1) First Occupancy (pp)	(2) Second Home Discharge (pp)	(3) Second Hospitalization (pp)	(4) Second Death (pp)
Occupancy (pp)		-1.309**	3.072***	0.146
		(0.618)	(0.742)	(0.223)
Short-stay discharge (pp)	-0.760***			
	(0.121)			
Mean outcome	89.13	34.52	37.18	6.54
N	599,946	599,946	599,946	599,946

*** p<0.01, **p<0.05, *p<0.1

Note: This table presents the results of instrumental variable regressions of patient outcomes on occupancy and other controls, using average short-stay discharges during the episode as an instrument. Column (1) displays the first-stage coefficient on the instrument and Columns (2)-(4) display the second-stage coefficients on occupancy. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

Columns (2)-(4) report the IV estimates of the coefficient on occupancy. The results imply that a 1pp increase in the average daily occupancy during the episode leads to a 1.3pp decrease in the probability of home discharge (3.8% of baseline) and a 3.1pp increase in the probability of hospitalization (8.3% of baseline). These estimates may reflect rehabilitation and procedural delays, in addition to potential health deterioration due to insufficient treatment. Although occupancy is also estimated to increase the likelihood of dead discharge, the estimate is not statistically significant.

Table 5 shows regression results using as an instrument average short-stay discharges in the 14 days preceding admission, while controlling for average short-stay discharges during episode. The effects of occupancy on hospitalization and death probability are close to the estimates reported in Table 4. Although the effect of occupancy on home discharge is again negative and statistically significant, the magnitude is three times larger than our estimates with baseline instrument (episode mean of short-stay discharges).

Table 5: Instrumental Variable Regression of Patient Outcomes on Occupancy (Before-Episode Discharge IV)

	(1) First Occupancy (pp)	(2) Second Home Discharge (pp)	(3) Second Hospitalization (pp)	(4) Second Death (pp)
Occupancy (pp)		-3.722***	3.017***	0.203
		(0.601)	(0.535)	(0.218)
Short-stay discharge (pp) (before admission)	-0.456***			
	(0.0323)			
Mean outcome	89.13	34.52	37.18	6.54
N	593,518	593,518	593,518	593,518

*** p<0.01, **p<0.05, *p<0.1

Note: This table presents the results of instrumental variable regressions of patient outcomes on occupancy and other controls, using average short-stay discharges in the 14 days preceding admission as an instrument, controlling for average short-stay discharges during the episode. Column (1) displays the first-stage coefficient on the instrument and Columns (2)-(4) display the second-stage coefficients on occupancy. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

Tables A3 and A4 in Appendix A present the results of the regressions of patient outcomes using short-stay admissions as instruments. The estimates for home discharge and hospitalization using the episode mean of short-stay admissions are remarkably similar to those using the episode mean of short-stay discharges. Results using short-stay admissions before admission as an instrument are also qualitatively similar to those using short-stay discharges before admission as an instrument.

Finally, to compare with our main results, Table A5 in Appendix A presents the results of OLSE of Eq.(2). In contrast to the IV estimates, we do not find economically or statistically significant relationships between occupancy and patient outcomes. This is consistent with a hypothesis that the negative causal effect of occupancy on patient outcomes is offset by the between-provider positive correlation of occupancy and unobserved quality. The upward bias in the OLSE is reminiscent of the upward bias in the regression of quantity on price.

4 Outcome-Based Quality Measures of Facilities

This section presents our framework for studying provider quality. In Section 4.1, we introduce a method to test whether the provider fixed effects obtained from regression (2) can be interpreted as causal effects of the provider on patient outcomes. We implement the test and report the results in Section 4.2. Using validated quality measures, we then provide evidence for the congestion-quality trade-off in Section 4.3.

4.1 Validating Value Added as Causal Quality

One concern about using value-added measures as institutional quality is that the measures may be systematically biased due to consumer sorting. For example, patients admitted to provider A may be systematically better in their health status than those admitted to provider B, even conditional on observables. In such a case, the value-added measure overstates the quality of provider A relative to that of provider B, because it conflates the true causal effect with differences in patient mix. Although instrumenting provider choice may be an effective solution in theory, finding strong instruments which induce sufficient variation in the choice among a large number of providers will prove difficult in practice. Instead of using J instruments to estimate J value added measures, we follow the validation method adopted in the literature (Chetty et al., 2014a; Abaluck et al., 2021). The method only requires a single instrument to estimate a single parameter, called a forecast coefficient, which informs us of the degree of systematic bias in the value-added measures.

The idea behind such forecast regressions is simple: if the value-added measures are unbiased estimates of true causal effects, then they should provide unbiased predictions of the outcome of consumers who are (quasi-)randomly assigned to institutions. Following Abaluck et al. (2021), we can evaluate the bias of our value-added estimates $(\hat{\mu}_j)_j$ by the regression

$$Y_{i\tau} = \lambda\hat{\mu}_i + \beta n_{i\tau} + x'_{i\tau}\gamma + \varepsilon_{i\tau} \tag{3}$$

using an instrument for $\hat{\mu}_i$. An unbiased measure $(\hat{\mu}_j)_j$ of provider quality will yield $\lambda = 1$.

$1-\lambda$ can be interpreted as the degree of bias of the value added as a causal quality measure.

Because we cannot include provider FEs to Eq.(3) (as they are collinear with μ_j), estimating (3) may not yield a consistent estimate of λ. Instead, we regress $Y_{i\tau} - \hat{\beta} n_{i\tau}$ on $\hat{\mu}_i$ and $x_{i\tau}$, instrumenting the former by an exogenous variable $W_{i\tau}$.[16] As an instrument $W_{i\tau}$ for realized value added, we use the value added of the provider which is closest to the focal patient's home. The idea is that (i) the value added of the patient's closest provider is likely correlated to realized (assigned) quality, because patients tend to choose a provider in their vicinity,[17] whereas (ii) it is uncorrelated to health shocks because patients' location is conditionally random.

4.2 Results of Quality Validation

Table 6 presents the results of forecast regressions for value-added measures from the regression (3). We take the specification with the episode mean of short-stay discharges as an instrument for occupancy: value-added estimates from other specifications of instruments are validated similarly. For all outcomes, the value added unbiasedly predicts the outcome of patients who are assigned to providers due to (arguably random) geographic proximity.

[16] Under reasonable assumptions, an IV regression of Eq.(2) yields a consistent estimate of β even if $(\hat{\mu}_j)_j$ are biased.

[17] In our sample, 35% of patients are admitted to the closest provider. Because the majority of patients enter a non-closest provider, the first-stage coefficients in Table 6 are far below 1, which suggests that our result is not due to the realized quality being almost identical to the closest provider's quality.

Table 6: Forecast Regression to Validate Value-Added Measure

	(1) First Home Discharge Value Added	(2) Second Home Discharge	(3) First Hospitalization Value Added	(4) Second Hospitalization	(5) First Death Value Added	(6) Second Death
Value Added		1.079***		1.025***		1.036***
		(0.0149)		(0.0108)		(0.0363)
Value Added of the Nearest Provider	0.295***		0.271***		0.161***	
	(0.00365)		(0.00387)		(0.00315)	
Control	Y	Y	Y	Y	Y	Y
N	419,709	419,709	419,709	419,709	419,709	419,709

*** p<0.01, **p<0.05, *p<0.1

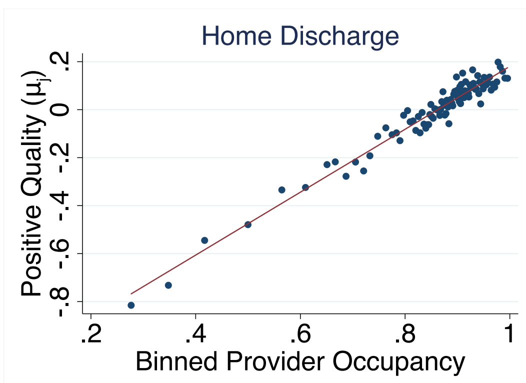
Note: This table presents the results of instrumental variable regressions of patient outcomes (residualized by the occupancy term) on the value-added measure and other controls, using the value added of the focal patient's closest provider as an excluded instrument. Columns (1), (3) and (5) display the first-stage coefficients on the excluded instrument and Columns (2), (4) and (6) display the second-stage coefficients on the value added. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects.

4.3 Congestion-Quality Trade-off

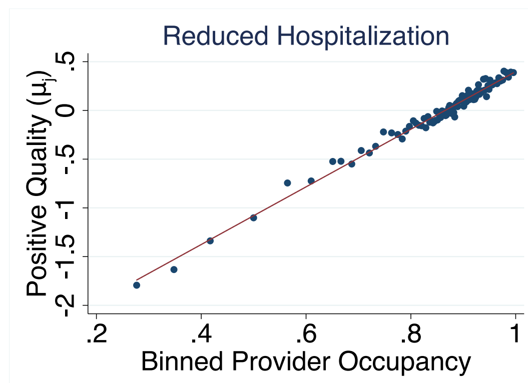
Figure 1 displays the scatter plots of provider-level value-added measures (fixed effects from Eq.(2)) against provider-specific average occupancy rate. Value-added measures for home discharge (Panel (a)) and reduced hospitalization (Panel(b)) exhibit strong positive correlation with provider occupancy. This suggests a *congestion-quality trade-off*: moving patients to a higher-quality provider tends to exacerbate the congestion faced by the patients, which might substantially offset the benefit of better quality.

Figure 1: Congestion and Quality

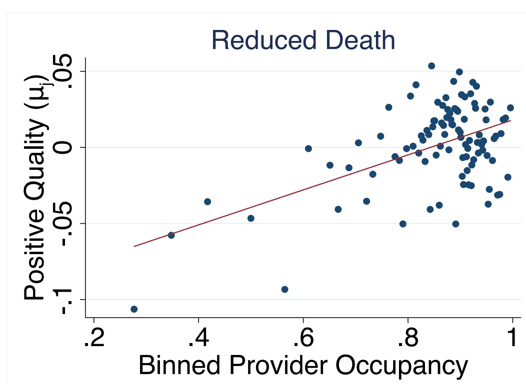
(a) Congestion and Home Discharge



(b) Congestion and Reduced hospitalization



(c) Congestion and Reduced Death



Note: This figure plots provider-level positive quality measure (fixed effects from Eq.(2) signed so that larger values imply the higher likelihood of a desirable outcome) against provider-level average occupancy rate. We use estimates of regressions with episode averages of discharges as instruments.

A caveat about the quality measure μ_j is that it measures *occupancy-adjusted productivity*:

it measures productivity of each provider holding observed production factors constant. While covariate-adjusted productivity is a common measure of value added used in the literature, it may overstate quality differences between providers which take substantially different values of covariates. Thus, our quality measure is likely more reliable to compare the quality of providers which are similar in occupancy.[18]

5 Demand Model with Choice Constraints

We now introduce a model of demand for nursing facility care. Estimating patient preferences is crucial for policy evaluation for two reasons. First, patients may value non-quality provider characteristics, in which case reallocation to smooth occupancy may be harmful to their welfare. For example, if they dislike distance from their previous home, then sending them to a less congested but more distant provider may be welfare-reducing. Second, if patients dislike congestion, then they will (partially) internalize the congestion externalities, which may reduce the need for policy interventions. To address these issues, we build and estimate a model of patients' demand for nursing facility admission and providers' admission decisions.

Timing. Patients (denoted by i) arrive sequentially to market $t = t_i$. We define markets by medical area-year-quarter combinations. Upon patient i's arrival, admission is realized via the following decisions.

- Each provider $j \in \mathcal{J}_t$ decides whether to offer i an admission.
- Patient i chooses which offer to accept.

An advantage of this (seemingly simplistic) assumption is that it leads to a familiar demand model with choice set constraints, commonly referred to as a consideration set model (Goeree, 2008). In our model, the consideration sets (i.e., restricted choice sets) are induced by providers' acceptance/rejection decisions.

[18]Note, however, that the congestion-quality trade-off is not a mechanical result. If providers at different occupancy levels are not systematically different in their quality, then we will observe a negative correlation between occupancy and patient outcomes. In actual data, this is not the case: recall that the OLSE of the coefficient on occupancy is statistically insignificant. Therefore, these patterns inform us that congested providers are better in unobserved terms.

Moreover, under additional assumptions, the above timing assumption yields the same choice outcome as the following, possibly more realistic, alternative:

- Upon arrival, i applies for facilities sequentially, in order of preference.
- Upon receiving an application, provider j decides whether to accept the application.

In either case, patient i is matched to her most preferred provider which accepts her application.

Payoffs. Provider j offers patient i an admission according to the following admission rule:[19]

$$O_{ijt} = I\left(\tilde{v}_{ijt} > 0\right) \tag{4}$$

$$\tilde{v}_{ijt} = v_j\left(x_{ijt}, n_{ijt}\right) - \eta_{ijt} \tag{5}$$

where $O_{ijt} = 1$ indicates that provider j offers an admission to patient i who arrives at market t. The "admission desirability" of patient i for provider j, $\tilde{v}_{ijt}$ (which we simply call provider's "utility"), depends on observed patient-provider characteristics $x_{ijt} \in \mathcal{X}_j$, episode-level occupancy $n_{ijt} \in \mathcal{N}_j = \left(\underline{n}_j, \bar{n}_j\right)$ and an unobserved variable $\eta_{ijt} \sim F_{\eta_{jt}|x_{ijt}}$ which admits a density function.

Patient i's utility from admission to provider j is $u_{ijt} = u\left(x_{ijt}, \xi_{jt}\right) + \varepsilon_{ijt}$, where ξ_{jt} denotes unobserved demand shocks to provider j and ε_{ijt} denotes an idiosyncratic shock with a distribution function G. As discussed below, we allow x_{ijt} to include the average occupancy at the provider-market level.

Other modelling issues. We omit discharge decisions. We also omit dynamic considerations of occupancy management, which is less of a concern because patients in our empirical setting are likely relatively homogeneous in their profitability compared to the context of US SNFs. Finally, an outside option contains nursing facilities of other types (private or public). We will normalize the systematic component of the utility of the outside option to zero, instead of accounting for their heterogeneous quality and occupancy.

[19] For simplicity, in the empirical analysis below, we assume v is common across j, and de-mean n_{ijt} by j-specific averages to account for baseline heterogeneity in occupancy.

Equilibrium. Let $C_i \subseteq \mathcal{J}_t$ be the subset of "inside" facilities that offer i an admission. The outside option is always available. The "equilibrium" of this model is the strategy profile which satisfies the following conditions: for each i and j,

1. patient i chooses provider j iff $j \in C_i \cup \{0\}$ and $u_{ijt} \geq u_{ij't}$ for all $j' \in C_i \cup \{0\}$.

2. provider $j \neq 0$ offers patient i an admission iff $v_{ijt} \geq 0$. Thus, $C_i = \{j \in \mathcal{J}_t : v_{ijt} \geq 0\}$.

6 Identification and Estimation of the Demand Model

6.1 Identification

In our empirical analysis below, we assume that the distribution of $(\eta_{ijt}, \varepsilon_{ijt})$ is known and estimate the parameterized version of u and v using observations on $x_{it} = (x_{ijt})_{j \in \mathcal{J}_t}$, $n_{it} = (n_{ijt})_{j \in \mathcal{J}_t}$ and choice j_i. There are two challenges with our empiricala analysis:

1. Offers C_i are unobserved.

2. Occupancy n_{ijt} may be correlated to current and past demand unobservables $\xi^t \equiv (\xi_{j\tau})_{j \in \mathcal{J}_\tau, \tau \leq t}$.

Failing to address the first problem leads to biased parameter estimates, hence biased welfare conclusions. To illustrate this point, suppose that patients positively value provider quality and that providers' acceptance probability is decreasing in occupancy. Then higher-quality providers are more likely to be excluded from the choice set C_i, as they attract more patients and face higher occupancy. If we estimate patient preferences assuming that all options are available, then we will underestimate patients' valuation of the quality, because they will be admitted to providers which are on average of lower quality than they would choose were it not for choice constraints. The second problem implies that occupancy is not excluded from patients' choice probabilities conditional on other observables.

In this section, we explain how we address the above problems, in two steps. First, we show that consideration probabilities and choice probabilities are separately identified from observed data, if we tentatively assume that occupancy rate n_{ijt} is independent of $(\xi_{jt})_{j\in\mathcal{J}_t}$. We then discuss how to restore the independence assumption using a control function. We omit market subscript t for notational simplicity.

6.1.1 Identification of Choice and Consideration Probabilities

We show that offer probabilities $\Pr(j \text{ offers } i \text{ an admission}|x_{ij}, n_{ij}) = F_{\eta_j|x_{ij}}(v_j(x_{ij}, n_{ij}))$ and the choice probabilities $\Pr(i \text{ chooses } j|C_i = C, x_i)$ are separately identified, using an exclusion restriction that n_{ij} shifts offer probability but not choice probability.

More specifically, we impose Assumptions 1 and 2.

Assumption 1

(i) η_{ij} *is independent across* j *conditional on* $x_i = (x_{ij})_{j\in\mathcal{J}}$ *and* $n_i = (n_{ij})_{j\in\mathcal{J}}$.

(ii) $\varepsilon_i = (\varepsilon_{ij})_{j\in\mathcal{J}}$, $\xi = (\xi_j)_{j\in\mathcal{J}}$, $\eta_i = (\eta_{ij})_{j\in\mathcal{J}}$ *and* n_i *are mutually independent conditional on* x_i.

(iii) For each j *and* x_{ij}, $\lim_{n_{ij}\to\underline{n}_j} F_{\eta_j|x_{ij}}(v_j(x_{ij}, n_{ij})) = 1$.[20]

Assumption 1(i) yields an "alternative specific consideration model" (Abaluck and Adams-Prassl, 2021), where the probabilities of a consideration set can be written as a product of independent consideration probabilities of each alternative. Assumption 1(ii) excludes n_i from choice probabilities conditional on x_i and C_i. Assumption 1(iii) is an "identification-at-infinity" assumption which has been imposed in most previous studies to fix the location parameter.

The above assumptions are different from those proposed by Agarwal and Somaini (2025) for identification of general demand models with choice set constraints. On the one hand, their identification result is applicable to a wider class of models, as they do not impose independent consideration (Assumption 1(i)). On the other hand, they impose two-way exclusion restrictions, i.e., that some variable affects choice set without affecting preference and another variable affects preference without affecting choice set.

We also need a rank condition:

[20] We can slightly relax this assumption to accommodate, e.g., $v(x, n) = \alpha + \beta x \cdot \left(\frac{1}{n} - 2\right)$, $n \in (0, 1)$.

Assumption 2 *Let*

$$
\begin{aligned}
F\left(C; x_i, n_i\right) &= \Pr\left(C_i = C \mid x_i, n_i\right) \\
&= \prod_{k \in C} F_{\eta_k | x_{ik}}\left(v_k(x_{ik}, n_{ik})\right) \prod_{l \in J \setminus C} \left[1 - F_{\eta_l | x_{il}}\left(v_l(x_{il}, n_{il})\right)\right]
\end{aligned}
$$

denote the conditional probability of consideration set C*. For each* j*, let* $\mathcal{C}(j) = \{C \subseteq \mathcal{J} : j \in C\} \equiv \left\{C_1^j, \ldots, C_K^j\right\}$ *denote the set of all consideration sets which contain* j*. Then, for each* j *and* x_i*, there exist* $n^{1,j}, \ldots, n^{R,j}$ *such that the matrix*

$$
F^j(x_i, \mathbf{n^j}) = \begin{bmatrix} F\left(C_1^j; x_i, n^{1,j}\right) & \cdots & F\left(C_K^j; x_i, n^{1,j}\right) \\ \vdots & \ddots & \vdots \\ F\left(C_1^j; x_i, n^{R,j}\right) & \cdots & F\left(C_K^j; x_i, n^{R,j}\right) \end{bmatrix}
$$

has rank K*.*

We then obtain an identification result for choice and consideration probabilities.

Proposition 1 *Suppose that Assumptions 1 and 2 hold. Then the offer probability*

$$
\Pr\left(j \text{ offers } i \text{ an admission} | x_{ij}, n_{ij}\right) = F_{\eta_j | x_{ij}}\left(v_j(x_{ij}, n_{ij})\right)
$$

and the choice probability

$$
\Pr\left(i \text{ chooses } j | C_i = C, x_i\right)
$$

are identified for all (j, x_i, n_i, C)*.*

Proof. See Appendix B.

As discussed above, an advantage of the above result relative to that of Agarwal and Somaini (2025) is that our result only requires an exclusion of some choice set shifter from patient preference. Also, beyond independence and exclusion, our result imposes few properties on choice probabilities, so it can be used as a basis for testing some properties.[21]

[21] For example, we can test whether the choice probabilities satisfy Slutsky-like symmetry (Abaluck and

6.1.2 Identification of Admission Rule and Preference

Identification of choice and consideration probabilities does not imply identification of preference u and admission parameter v. Identification of the primitives is required, for example, to examine the convexity of the admission cost captured by $-v$: if v is identified only up to some monotonic transformation, then we cannot tell whether the true $-v$ is convex in occupancy.

Conditions for the identification of v are available in the literature (e.g., Matzkin 1992). These conditions typically assume the existence of some large-support variable which enters the function linearly. However, in our case, we may be interested in investigating whether occupancy (the large-support variable) enters the negative provider utility in a convex manner. Therefore, in Appendix B, we develop an identification result which requires that some variable, not necessarily with a large support, enters the cost function linearly. Intuitively, such a variable allows us to identify the derivatives of v, which together with a location normalization yields v.[22]

For identification of patient preferences, we invoke existing results (Berry and Haile, 2016), conditional on identification of choice probabilities.

6.1.3 Addressing Endogeneity of Occupancy

Proposition 1 is based on the conditional independence of n_i and ξ. This may fail, however, because current occupancy is partly the result of past demand shocks and demand shocks are serially correlated. Specifically, we will have $n_{ijt} = n_{ijt}\left(\xi^{t-1}, x^{t-1}\right)$ where $\xi^t = \left(\xi_{j\tau}\right)_{j\in\mathcal{J}_\tau,\tau\leq t}$, etc.

To address this problem, we make some assumptions. Recall that the occupancy rate varies even within market, defined by medical area-quarter in the empirical analysis below, because occupancy fluctuates at the daily level. Let n_{ijt} denote the occupancy rate at the beginning of episode i at provider j in market t.[23] Also, let n^e_{jt} denote the provider-quarter

Adams-Prassl, 2021) or the random utility axioms (Falmagne, 1978).

[22]The assumption requires that the provider utility be continuously differentiable with respect to occupancy. In the empirical analysis below, we specify v as a piecewise linear function of occupancy, so the identification assumption does not hold exactly.

[23]We assume this is what the provider cares about. Alternative assumptions are possible.

average of occupancy. Define the residualized occupancy as $\tilde{n}_{ijt} = n_{ijt} - n^e_{jt}$.

Assumption 3

(i) Patients only care about the provider-quarter average occupancy n^e_{jt}.

(ii) $\tilde{n}_{ijt}$ is independent of $(\xi_{j\tau})_{j,\tau \leq t}$ conditional on n^e_{jt} and other observables.

Assumption 3(i) means that patients who arrive on different days in market t use the same expected occupancy to make an application decision.[24] Assumption 3(ii) suggests that the occupancy fluctuations around its market-provider-specific average is independent of demand shocks and is excluded from choice probability, conditional on n^e_{jt}. Therefore, n^e_{jt} serves as a control function to restore independence between (residualized) occupancy and demand. Specifically, by controlling for n^e_{jt} both in choice and consideration probabilities, the residual variation in occupancy becomes excluded from demand, which enables us to apply Proposition 1 for identification.[25] This assumption is violated if, for example, there is a serially correlated demand shock which varies at the daily level.

Conditional on the average occupancy n^e_{jt}, residual fluctuations in occupancy come from unexpected patient inflows/outflows such as emergency admissions and patient deaths. Although these factors may be correlated to provider quality (hence ξ_{jt}) in the long run, the precise timing of emergencies to family caregivers and deaths is likely random.

Because n^e_{jt} is endogenous to demand, we need to instrument for it. Motivated by the analysis in Section 3, we use the number of short-stay discharges in the previous quarter as an IV. Note that our approach yields patients' preference over average occupancy n^e_{jt}, which will be of interest for evaluating welfare effects of policies.

[24] With additional complexity, we can allow the occupancy in patients' information set to be patient-specific.

[25] A caveat is that controlling for n^e_{jt} can affect the validity of Assumption 1(iii). In the empirical analysis below, we only assume that the identification assumptions hold approximately, partially relying on the parametric assumptions.

6.2 Empirical Specification

We adopt simple specifications for provider and patient utility functions as follows:[26]

$$
\begin{aligned}
v_{ijt} &= \alpha_{n1} n_{ijt} + \alpha_{n2} n_{ijt} I\left(n_{ijt} \geq K\right) + \alpha_{e1} n^e_{jt} + y^{f\prime}_{ijt} \alpha_y - \eta_{ijt} \\
u_{ijt} &= \underbrace{w'_{jt}\beta_w + \beta_{e1} n^e_{jt} + \xi_{jt}}_{\equiv \delta_{jt}} + y^{p\prime}_{ijt}\gamma_y + \varepsilon_{ijt} \\
u_{i0t} &= \varepsilon_{i0t}
\end{aligned}
$$

where y^f_{ijt} denotes a vector of patient characteristics which affect facility j's decision to accept or reject an application from patient i (hence a superscript f), y^p_{ijt} is a vector of patient-provider characteristics which affect patients' preferences, and w_{jt} is a vector of facility characteristics. K denotes a threshold value of daily occupancy, which we set to the average occupancy in sample. The portion of patient utility which varies at the provider-market level is denoted by δ_{jt}. We normalize $\delta_{0t} + y^{p\prime}_{ijt}\gamma_y = 0$. η_{ijt} follows i.i.d. logistic distribution, and ε_{ijt} follows i.i.d. Type-I Extreme Value distribution.

In our empirical analysis, y^f_{ijt} includes a constant, an indicator of at least 75 years of age, a female indicator, an indicator of care level being 3 or higher (on a scale of 1-5), and an indicator of whether the patient is from the same city of the provider. As y^p_{ijt}, we use the distance between patient i's (former) residence and provider j. w_{jt} includes a constant and the value added for home discharge.

Similarly to the analysis of patient outcomes in Section 3, we prefer to assume that our instrument is exogenous only after eliminating systematic differences across providers. To do so, we apply within-provider transformation to our occupancy variables.

Finally, we restrict the sample for the structural analysis. We focus on patients who are at age 65 or older at admission; those who access long-term care below this age may have special care needs and choose providers differently. Also, due to computational costs, we focus on observations in the Tokyo prefecture.

[26] Although not ideal, we assume that controlling for the linear term of n^e_{jt} eliminates the relevant portion of correlation between occupancy and demand shocks.

6.3 Estimation

We aim to estimate $\tilde{\theta} = (\theta, \beta)$ where $\theta \equiv \left(\alpha_{n1}, \alpha_{n2}, \alpha_{e1}, \alpha_y', \gamma_y'\right)'$ collects parameters commonly called *nonlinear parameters* in the IO literature, and $\beta = (\beta_w', \beta_{e1})'$ denotes so-called *linear parameters*.

An estimation challenge is to deal with the correlation between average occupancy n_{jt}^e and unobserved demand shocks ξ_{jt} in the nonlinear setting. A common approach to this problem is the BLP method (Berry et al., 1995). In this approach, estimation proceeds in a loop structure where in the inner loop given a trial parameter value, they find a "mean utility" vector δ which equates the predicted market share to the empirical market share. They provide a contraction mapping algorithm to solve for δ. Besides its potentially slow convergence, applying their approach to models with choice set constraints involves an additional difficulty: for some parameter value θ, there may not exist $\delta = \delta(\theta)$ that solves the market-share equation. If, for example, the parameter value is such that the acceptance probability of option j is around 0.2, then no vector δ can rationalize its empirical market share of 0.3.[27] Therefore, an estimation approach which relaxes the market share constraint is desired. Although there are alternative estimation approaches which do not (always) impose market share equations during parameter search (Dubé et al., 2012; Grieco et al., 2025), they involve optimization over a large number of parameters, which may be difficult in complex models with choice set constraints.

We therefore adapt an approximate BLP method (Lee and Seo, 2015) to our setting with choice set constraints and microdata. With "just identification" (meaning that the number of endogenous variables is the same as that of excluded instruments), our approach can further be simplified, yielding the following iterative procedure:

1. Given an initial value δ^0, we update parameter in the h-th iteration by

$$\theta^h = \arg\max_{\theta \in \Theta_{NL}} \ln L\left(\theta; \delta^{h-1}\right)$$

where $\ln L\left(\theta; \delta\right)$ denotes the log-likelihood of the episode-level sample and Θ_{NL} denotes

[27] Strictly speaking, this example contradicts the identification-at-infinity assumption above. However, the intuition will remain valid if the acceptance probability approaches one only infrequently.

the support of the nonlinear parameters θ.

2. Update δ by

$$\begin{aligned}\delta^h &= \delta\left(\theta^h, \delta^{h-1}\right)\\ &\equiv \delta^{h-1} + \left[\nabla_{\delta'} \ln s\left(\delta^{h-1}; \theta^h\right)\right]^{-1} \left[\ln S - \ln s\left(\delta^{h-1}; \theta^h\right)\right]\end{aligned}$$

where S denotes the vector of empirical market share and $s\left(\delta;\theta\right)$ is the vector of predicted market share.

3. Upon convergence, we obtain estimates $(\hat{\theta}, \hat{\delta})$. We then estimate β by

$$\hat{\beta} = \left[Z'X\right]^{-1} Z'\hat{\delta}$$

where X and Z are a matrix of provider characteristics and that of instruments, respectively.

This approach avoids imposing market share equations at implausible parameter values. Instead of obtaining $\delta(\theta)$ which solves $\ln S = \ln s(\delta;\theta)$, we update δ based on the Taylor approximation of this log market share equation. Although we have not investigated formal properties of convergence, the update of δ alone follows a Newton-Raphson step, so we conjecture that similar convergence property may hold if we also update θ. The convergent point $\hat{\delta}$ solves the market share equation.[28]

Note also that this approach exploits the separability of the micro and macro objective functions conditional on δ (Grieco et al., 2025). Specifically, given δ, the micro likelihood function is devoted to pinning down parameters governing patient heterogeneity, whereas standard macro moments are devoted to addressing the provider-level endogeneity problem. The estimation approach becomes slightly more complicated if the number of excluded instruments is strictly larger than that of endogenous variables, because the choice of θ will have to account for macro moments in addition to micro likelihood.

[28]We are also working on an alternative estimator suggested by Grieco et al. (2025), which does not involve the concern about convergence.

The likelihood for patient i admitted to provider j is

$$\begin{aligned}
l_{ij}\left(x_{it}, n_{it}\right) &= \sum_{C \subseteq \mathcal{J}} \prod_{k \in C} \frac{\exp\left(\bar{v}_{ikt}\right)}{1+\exp\left(\bar{v}_{ikt}\right)} \prod_{l \in \mathcal{J} \setminus C} \frac{1}{1+\exp\left(\bar{v}_{ilt}\right)} \frac{I\left(j \in C\right) \exp\left(\bar{u}_{ijt}\right)}{1+\sum_{j' \in C} \exp\left(\bar{u}_{ij't}\right)} \\
&\equiv \sum_{C \subseteq \mathcal{J}} \prod_{k \in C} F_{ikt}(\theta^f) \prod_{l \in \mathcal{J} \setminus C} \left(1 - F_{ilt}(\theta^f)\right) P_{ijt}(C; \theta^p) \\
&\equiv \sum_{C \subseteq \mathcal{J}} Q_{it}(C; \theta^f) P_{ijt}(C; \theta^p).
\end{aligned}$$

Instead of evaluating this, we simulate the likelihood by drawing consideration sets, with importance sampling to smooth the objective function. Details are described in Appendix C.

7 Estimation and Simulation Results

7.1 Estimation Results

Table 7 shows the estimates of provider parameters. Providers are more likely to accept an application if the applicant lives in the same city as the provider, is female, is of lower care-need level, or if the occupancy rate becomes lower relative to high baseline occupancy. To facilitate interpretation, Table 7 also displays the marginal effect of each characteristic for the acceptance probability of a single provider, evaluated at the mean value of characteristics. For example, acceptance probability is 4.4 pp higher if the applicant is from the same city.

Table 8 similarly presents estimates of patient parameters and implied marginal effects. We isolate the marginal effect on conditional choice probability (from effect on consideration) by presenting the marginal effect of a characteristic on the probability that an alternative is chosen over an outside alternative, from the binary choice set which contains these two options. Patients prefer providers with higher value added for home discharge and dislike distant providers. The coefficient on provider-quarter-specific average occupancy is imprecisely estimated, which may indicate that patients do not take congestion into account when choosing a provider.

Table 7: Estimates of Provider Parameters

Parameter	Estimate	SE	Marginal Effect
Constant	0.0115	0.0144	0.0029
Same City	0.1767***	0.0074	0.0441
Old	0.0109	0.0120	0.0027
Female	0.0144*	0.0075	0.0036
High Care Level	-0.0261***	0.0084	-0.0065
Occupancy	0.0890	0.1325	0.0222
Occupancy$\times$ I(Occupancy > Mean)	-0.3100*	0.1752	-0.0774

Note: This table presents the estimates of provider parameters. Rows indicate the coefficient on the constant term, the indicator of whether the patient's home is in the same city as the provider, indicator of whether the patient is 75 years old or older, female indicator, indicator of care level 3 or above (indicating relatively high care needs), de-meaned (by provider-specific average) daily occupancy, and de-meaned occupancy interacted by an indicator of the occupancy above mean (zero by construction). The column "Marginal Effect" displays the marginal effect of each characteristic on the offer probability of a single provider, evaluated at the mean value of characteristics. Estimate of the coefficient on the control covariate (provider-quarter-specific average occupancy) is omitted. Occupancy is denoted by the absolute number (1=100pp).

7.2 Simulation Results

To illustrate the effect of reallocation, we simulate a simple policy of smoothing occupancy. Instead of considering a complicated procedure to smooth occupancy across different episodes, we treat each year-quarter as consisting of homogeneous episodes and predict how the probability of home discharge changes with the policy, using the estimate of the coefficient on occupancy and provider quality estimates (using the specification with episode average of short-stay discharges as an instrument, as reported in Table 4).

More specifically, we assume that the occupancy rate is constant within provider-year-quarter bin (at the average within the bin). This generates a hypothetical set of homogeneous episodes, each lasting from the beginning of the year-quarter to the end of it. We then consider reallocating patients from the most congested provider to the least congested provider within each market (medical area-year-quarter) to equalize their average occupancy in the year-quarter. We ignore changes in variables other than occupancy and quality, and

Table 8: Estimates of Patient Parameters

Parameter	Estimate	SE	Marginal Effect with Binary Choice
Constant	0.8531***	0.0317	0.0981
Value Added for Home Discharge	1.2406***	0.1363	0.1426
Occupancy	5.5632	7.0981	0.6395
Distance (km)	-0.4856***	0.0018	-0.0558

Note: This table presents the estimates of patient parameters. Rows indicate the coefficient on the constant term, the value added for home discharge, provider-quarter-specific average occupancy, and distance to the provider. The column "Marginal Effect with Binary Choice" displays the marginal effect of each characteristic on the probability that an option with mean characteristics is chosen over an outside option, from the binary choice set. Occupancy is denoted by the absolute number (1=100pp).

decompose the effect of reallocation on patient i's outcome as follows:

$$\begin{aligned}\Delta_i &= E\left[Y_{ij_i^{\text{post}}} \mid n_j^{\text{post}}\right] - E\left[Y_{ij_i^{\text{pre}}} \mid n_j^{\text{pre}}\right] \\ &= \underbrace{\beta\left(n_{ij_i^{\text{post}}} - n_{ij_i^{\text{pre}}}\right)}_{\text{occupancy}-\text{smoothing effect}} + \underbrace{\mu_{j_i^{\text{post}}} - \mu_{j_i^{\text{pre}}}}_{\text{quality effect}} \equiv \Delta_i^o + \Delta_i^q \end{aligned} \tag{6}$$

where β is the coefficient on average occupancy in Eq.(2) and j_i^{post} (j_i^{pre}) is the facility to which patient i is assigned after (before) reallocation. The first term of Eq.(6) is an occupancy-smoothing effect, due to changing occupancy. Let h (l) denotes the provider with the higher (lower) occupancy. For home discharge ($\beta < 0$), we have $\Delta_i^o > 0$ if $j_i^{\text{pre}} = h$ and $\Delta_i^o < 0$ if $j_i^{\text{pre}} = l$, and the net occupancy-smoothing effect within each pair of providers, $\sum_{i:j_i^{\text{pre}} \in \{h,l\}} \Delta_i^o$, is *always* non-negative.[29] On the other hand, with congestion-quality trade-off, we expect $\sum_{i:j_i^{\text{pre}} \in \{h,l\}} \Delta_i^q$ to be negative. We report the result for aggregate home discharges at the market level. We also compute changes in the utility of patients, measured relative to the disutility from distance.

Table 9 shows simulation results. Panel (a) shows that the occupancy-smoothing effect is positive but the quality effect is typically negative, as reallocation transfers patients to a lower-quality provider on average. The total effect is negative on average; however, the

[29] Proof is omitted. Roughly, this is because there are more patients who enjoy an occupancy reduction than those who suffer an occupancy increase.

median effect is positive, so there are more winner markets than there are loser markets. Panel (b) shows that the welfare is negatively affected by reallocation, with average disutility equivalent to an increase in distance of 7.3km.

Table 9: Simulated Effect of Occupancy Smoothing on Patient Outcomes and Welfare

	Mean (1)	Std. Dev. (2)	Median (3)	Obs. (4)
(a) Change in Per-Patient Likelihood of Discharge to Home				
Occupancy-smoothing (pp)	0.273	4.117	0.287	262
Quality (pp)	-0.918	1.870	-0.074	262
Total (pp)	-0.646	4.082	0.159	262
(b) Change in Utility				
Total (relative to dist. coef)	-7.30	2.06	-8.04	1,416

Notes: This table presents changes in discharge to home and patient utility following reallocation. The row "Occupancy-smoothing (pp)" in panel (a) shows the per-patient net occupancy-smoothing effect (sum of Δ_i^o in Eq. (6) within each medical area-quarter, divided by the number of affected patients), in percentage points. Similarly, the row "Quality (pp)" shows the net quality effect (sum of Δ_i^q in Eq. (6) within each medical area-quarter) and the row "Total (pp)" shows the net total effect (sum of Δ_i in Eq. (6) within each medical area-quarter), both on per-patient basis and represented in pp. The row "Total (relative to dist. coef)" in panel (b) shows changes in patient utility, divided by the absolute value of distance coefficient.

However, this analysis is preliminary because we only consider reallocated patients, and it is partly driven by noisy estimates of utility parameters. Other limitations of the current simulation exercise are mentioned in Section 8.

8 Conclusion

Heterogeneity in outcome-based quality measures across service providers has attracted attention of many researchers and policymakers, who have then suggested possible gains from steering consumers to higher-quality providers. This paper suggests that policy debates may need to consider another factor, congestion, which may affect service outcomes and welfare negatively. Using fluctuations in short-stay patient volume as an instrument for occupancy (our congestion measure), we find that a 1pp increase in the average occupancy during an episode of nursing-home stay leads to a 1.3pp (3.8%) decrease in the probability of home discharge and a 3.1pp (8.3%) increase in the probability of hospitalization. On the other hand, providers with higher covariate-adjusted quality also tend to be more congested, which generates a *congestion-quality trade-off* in producing patient outcomes.

We then build and estimate a demand model for nursing facilities, to evaluate the welfare impacts of reallocating patients to smooth occupancy. The model accommodates choice restrictions due to providers' rejections of admission, and we show that the model is identified by using daily occupancy fluctuations as a choice set shifter that is excluded from patient preferences. In estimation, we also address the endogeneity of aggregate occupancy by adapting an approximate BLP method to our microdata setting. Our estimates suggest that patients prefer a closer, higher-quality provider, but we find no evidence that they care about congestion. On the other hand, highly occupied providers partially internalize the negative impacts of congestion by reducing the probability of admitting a patient when occupancy increases. Our simulation suggests that the net effect of occupancy smoothing on the home discharge outcome is positive for the median market, even though the reallocated patients face lower-quality providers on average. In terms of welfare, they are made worse off by the reallocation.

Our findings have important implications for policies that aim to steer consumers toward high-quality providers. Informational interventions to steer more consumers to better providers may backfire if they generate excessive congestion. This trade-off between the negative externality of congestion and improved choice is reminiscent of the finding of Handel (2013), who suggests that improving choice by eliminating inertia may have unintended negative consequences due to exacerbated adverse selection. Beyond informational interventions, our findings have implications for capacity policies, such as entry/exist regulations. While provider exits may reallocate consumers to higher-quality providers (Olenski, 2023), there may emerge an offsetting effect of congestion (Avdic et al., 2024). Thus, the policymaker should pay attention to congestion along with the quality of operating institutions.

This paper leaves some issues up to future research. First, the effects of congestion may be heterogeneous; for example, the effects may be stronger for high-occupancy episodes. We are currently investigating effect heterogeneity and how it may affect our conclusions on congestion-quality trade-off. Second, we have only considered the welfare impact of reallocation on the patients who are transferred to a different provider. Given the statistically insignificant estimate of patients' valuation of congestion, our model predicts that reallocation will have no significant impact on the welfare of patients who remain in the same facility. However, the insignificant estimate may be due to the negative valuation of congestion being offset by positive effects, such as a signaling value of congestion. Third, we have not discussed whether the current allocation of patients is efficient. Discussions of spatial misallocation will require us to take stance on the "right" preference and what causes deviations of occupancy distribution from the efficient one. Finally, we have not investigated what policy can lead to efficient outcomes (relative to the current or an alternative benchmark). We will consider these extensions in the future updates of this paper.

References

Abaluck, Jason, and Abi Adams-Prassl. 2021. "What do Consumers Consider Before They Choose? Identification from Asymmetric Demand Responses." *The Quarterly Journal of Economics* 136 (3): 1611–1663.

Abaluck, Jason, Mauricio Caceres Bravo, Peter Hull, and Amanda Starc. 2021. "Mortality Effects and Choice Across Private Health Insurance Plans." *The Quarterly Journal of Economics* 136 (3): 1557–1610.

Abdulkadiroğlu, Atila, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. 2020. "Do Parents Value School Effectiveness?" *American Economic Review* 110 (5): 1502–39.

Agarwal, Nikhil, and Paulo J Somaini. 2025. "Demand analysis under latent choice constraints."Technical report, National Bureau of Economic Research.

Ainsworth, Robert, Rajeev Dehejia, Cristian Pop-Eleches, and Miguel Urquiola. 2023. "Why Do Households Leave School Value Added on the Table? The Roles of Information and Preferences." *American Economic Review* 113 (4): 1049–82.

Allende, Claudia. 2019. "Competition Under Social Interactions and the Design of Education Policies." Working Paper.

Andrew, Alison, and Marcos Vera-Hernández. 2022. "Incentivizing Demand for

Supply-Constrained Care: Institutional Birth in India." *The Review of Economics and Statistics* 1–45.

Angrist, Joshua D., Peter D. Hull, Parag A. Pathak, and Christopher R. Walters. 2017. "Leveraging Lotteries for School Value-Added: Testing and Estimation." *The Quarterly Journal of Economics* 132 (2): 871–919.

Avdic, Daniel, Petter Lundborg, and Johan Vikström. 2024. "Does Health Care Consolidation Harm Patients? Evidence from Maternity Ward Closures." *American Economic Journal: Economic Policy* 16 (1): 160–89.

Bachner, Florian, Martin Halla, and Gerald J Pruckner. 2024. "Do Empty Beds Cause Cesarean Deliveries?".

Bär, Marlies, Pieter Bakx, Bram Wouterse, and Eddy van Doorslaer. 2022. "Estimating the health value added by nursing homes." *Journal of Economic Behavior Organization* 203 1–23.

Barseghyan, Levon, Francesca Molinari, and Matthew Thirkettle. 2021. "Discrete Choice under Risk with Limited Consideration." *American Economic Review* 111 (6): 1972–2006.

Berry, Steven, and Philip Haile. 2016. "Identification in Differentiated Products Markets." *Annual Review of Economics* 8 27–52.

Berry, Steven, James Levinsohn, and Ariel Pakes. 1995. "Automobile prices in market equilibrium." *Econometrica* 63 (4): 841–890.

Beuermann, Diether W, C Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. 2022. "What is a Good School, and Can Parents Tell? Evidence on the Multidimensionality of School Output." *The Review of Economic Studies* 90 (1): 65–101.

Buchholz, Nicholas, Laura Doval, Jakub Kastl, Filip Matejka, and Tobias Salz. 2025. "Personalized Pricing and the Value of Time: Evidence From Auctioned Cab Rides." *Econometrica* 93 (3): 929–958.

Care Work Foundation. 2016. "Nursing Care Labor Survey." https://www.kaigo-center.or.jp/content/files/report/h28_chousa_kekka.pdf (Japanese), Accessed July 31, 2025.

Castillo, Juan Camilo. 2022. "Who benefits from surge pricing?" *Available at SSRN 3245533*.

Chandra, Amitabh, Maurice Dalton, and Douglas O Staiger. 2023. "Are Hospital Quality Indicators Causal?" Working Paper 31789, National Bureau of Economic Research.

Chandra, Amitabh, Amy Finkelstein, Adam Sacarny, and Chad Syverson. 2016. "Health Care Exceptionalism? Performance and Allocation in the US Health Care Sector." *American Economic Review* 106 (8): 2110–44.

Cheng, Alden. 2023. "Demand for Quality in the Presence of Information Frictions: Evidence from the Nursing Home Market."

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014a. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104 (9): 2593–2632.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014b. "Measuring the Impacts of Teachers II: Teacher Value-Added and Student Outcomes in Adulthood." *American Economic Review* 104 (9): 2633–79.

Chetty, Raj, and Nathaniel Hendren. 2018. "The Impacts of Neighborhoods on Intergenerational Mobility II: County-Level Estimates." *The Quarterly Journal of Economics* 133 (3): 1163–1228.

Conlon, Christopher T., and Julie Holland Mortimer. 2013. "Demand Estimation under Incomplete Product Availability." *American Economic Journal: Microeconomics* 5 (4): 1–30.

De Los Santos, Babur, Ali Hortaçsu, and Matthijs R. Wildenbeest. 2012. "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior." *American Economic Review* 102 (6): 2955–80.

Deming, David J. 2014. "Using School Choice Lotteries to Test Measures of School Effectiveness." *American Economic Review* 104 (5): 406–11.

Dubé, Jean-Pierre, Jeremy T. Fox, and Che-Lin Su. 2012. "Improving the Numerical Performance of Static and Dynamic Aggregate Discrete Choice Random Coefficients Demand Estimation." *Econometrica* 80 (5): 2231–2267.

Dubois, Pierre, and Morten Sæthre. 2020. "On the Effect of Parallel Trade on Manufacturers' and Retailers' Profits in the Pharmaceutical Sector." *Econometrica* 88 (6): 2503–2545.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2025. "Producing Health: Measuring Value Added of Nursing Homes." *Econometrica* 93 (4): 1225–1264.

Falmagne, Jean-Claude. 1978. "A Representation Theorem for Finite Random Scale Systems." *Journal of Mathematical Psychology* 18 (1): 52–72.

Fréchette, Guillaume R., Alessandro Lizzeri, and Tobias Salz. 2019. "Frictions in a Competitive, Regulated Market: Evidence from Taxis." *American Economic Review* 109 (8): 2954–92.

Freedman, Seth. 2016. "Capacity and Utilization in Health Care: The Effect of Empty Beds on Neonatal Intensive Care Admission." *American Economic Journal: Economic Policy* 8 (2): 154–185. http://dx.doi.org/10.1257/pol.20120393.

Gandhi, Ashvin. 2023. "Picking your patients: Selective admissions in the nursing home industry." *Working Paper.*

Gaynor, Martin, Carol Propper, and Stephan Seiler. 2016. "Free to Choose? Reform, Choice, and Consideration Sets in the English National Health Service." *American Economic Review* 106 (11): 3521–57.

Geweke, John, Gautam Gowrisankaran, and Robert J. Town. 2003. "Bayesian Inference for Hospital Quality in a Selection Model." *Econometrica* 71 (4): 1215–1238.

Goeree, Michelle Sovinsky. 2008. "Limited Information and Advertising in the US Personal Computer Industry." *Econometrica* 76 (5): 1017–1074.

Grieco, Paul LE, Charles Murry, Joris Pinkse, and Stephan Sagl. 2025. "Optimal Estimation of Discrete Choice Demand Models with Consumer and Product Data."

Gutierrez, Emilio, and Adrian Rubli. 2021. "Shocks to Hospital Occupancy and Mortality: Evidence from the 2009 H1N1 Pandemic." *Management Science* 67 (9): 5943–5952. https://doi.org/10.1287/mnsc.2021.4010.

Hackmann, Martin B., R. Vincent Pohl, and Nicolas R. Ziebarth. 2024. "Patient versus Provider Incentives in Long-Term Care." *American Economic Journal: Applied Economics* 16 (3): 178–218.

Handel, Benjamin R. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review* 103 (7): 2643–82.

He, Daifen, and R. Tamara Konetzka. 2015. "Public Reporting and Demand Rationing: Evidence from the Nursing Home Industry." *Health Economics* 24 (11): 1437–1451. https://doi.org/10.1002/hec.3097.

Heiss, Florian, Daniel McFadden, Joachim Winter, Amelie Wuppermann, and Bo Zhou. 2021. "Inattention and Switching Costs as Sources of Inertia in Medicare Part D." *American Economic Review* 111 (9): 2737–81.

Ho, Kate, Joseph Hogan, and Fiona Scott Morton. 2017. "The impact of consumer inattention on insurer pricing in the Medicare Part D program." *The RAND Journal of Economics* 48 (4): 877–905.

Hoe, Thomas P. 2022. "Does Hospital Crowding Matter? Evidence from Trauma and Orthopedics in England." *American Economic Journal: Economic Policy* 14 (2): 231–262. 10.1257/pol.20180672.

Honka, Elisabeth. 2014. "Quantifying search and switching costs in the US auto insurance industry." *The RAND Journal of Economics* 45 (4): 847–884.

Hull, Peter. 2020. "Estimating Hospital Quality with Quasi-experimental Data." Working paper.

Kane, Thomas J, and Douglas O Staiger. 2008. "Estimating Teacher Impacts on Student Achievement: An Experimental Evaluation." Working Paper 14607, National Bureau of Economic Research.

Kawaguchi, Kohei, Kosuke Uetake, and Yasutora Watanabe. 2021. "Designing

Context-Based Marketing: Product Recommendations Under Time Pressure." *Management Science* 67 (9): 5642–5659.

Lee, Jinhyuk, and Kyoungwon Seo. 2015. "A computationally fast estimator for random coefficients logit demand models using aggregate data." *The RAND Journal of Economics* 46 (1): 86–102.

Leshno, Jacob D. 2022. "Dynamic Matching in Overloaded Waiting Lists." *American Economic Review* 112 (12): 3876–3910.

Matzkin, Rosa L. 1992. "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and The Binary Choice Models." *Econometrica* 60 (2): 239–270.

Ministry of Health, Labor and Welfare. 2017. "Short-Stay Residential and Medical Care." https://www.mhlw.go.jp/file/05-Shingikai-12601000-Seisakutoukatsukan-Sanjikanshitsu_Shakaihoshoutantou/0000168704.pdf (Japanese), Accessed August 4, 2025.

Ministry of Health, Labor and Welfare. 2023a. "Geriatric Health Services Facilities." https://www.mhlw.go.jp/content/12300000/001131788.pdf (Japanese), Accessed June 2, 2025.

Ministry of Health, Labor and Welfare. 2023b. "Overview of Long-Term Care Cost Statistics." https://www.mhlw.go.jp/toukei/saikin/hw/kaigo/kyufu/23/dl/03.pdf (Japanese), Accessed June 2, 2025.

Morris, Carl N. 1983. "Parametric Empirical Bayes Inference: Theory and Applications." *Journal of the American Statistical Association* 78 (381): 47–55.

Moscelli, Giuseppe, Hugh Gravelle, and Luigi Siciliani. 2021. "Hospital competition and quality for non-emergency patients in the English NHS." *The RAND Journal of Economics* 52 (2): 382–414.

Neilson, Christopher. 2013. "Targeted vouchers, competition among schools, and the academic achievement of poor students." Working Paper.

Olenski, Andrew. 2023. "Reallocation and the (In)efficiency of Exit in the US Nursing Home Industry."

Olenski, Andrew, and Szymon Sacher. 2024. "Estimating Nursing Home Quality with Selection." *The Review of Economics and Statistics* 1–31.

Rahman, Momotazur, David C. Grabowski, Vincent Mor, and Edward C. Norton. 2016. "Is a Skilled Nursing Facility's Rehospitalization Rate a Valid Quality Measure?" *Health Services Research* 51 (6): 2158–2175.

Russo, Anna. 2024. "Waiting or Paying for Healthcare: Evidence from the Veterans Health Administration." Working paper.

Saruya, Hiroki, and Masaki Takahashi. 2025. "Improving Access or Inducing Demand? Analyzing Trade-offs in Healthcare Capacity."

Sharma, Rajiv, Miron Stano, and Renu Gehring. 2008. "Short-Term Fluctuations in Hospital Demand: Implications for Admission, Discharge, and Discriminatory Behavior." *RAND Journal of Economics* 39 (2): 586–606. https://doi.org/10.1111/j.0741-6261.2008.00029.x.

Yee, Christine A., Yevgeniy Feyman, and Steven D. Pizer. 2022. "Dually-enrolled patients choose providers with lower wait times: Budgetary implications for the VHA." *Health Services Research* 57 (4): 744–754.

A Additional Figures and Tables

Table A1: Functional Status of Each Support and Care Level

Support level 1-2	The patient is able to perform most of the basic activities of daily living on her own, but some nursing care is required for complex daily activities.
Care level 1	The patient's ability to perform complex daily activities has declined further from the state of Support level.
Care level 2	In addition to the condition of care level 1, the patient requires nursing care for basic daily activities.
Care level 3	Compared to the state of care level 2, there is a significant decline in terms of both basic and complex daily activities, and almost total nursing care is required.
Care level 4	In addition to the condition of care level 3, the patient's ability to move is further reduced and it becomes difficult for her to carry out daily living without nursing care.
Care level 5	The patient's ability to perform daily activities is even worse than the state of care level 4, and it is almost impossible for the patient to carry out daily living without nursing care.

Notes: This table, replicated from Saruya and Takahashi (2025), describes physical status of patients in each category of support levels and care levels.

Table A2: Covariate Balance (Admission Instrument)

	(1) Age	(2) Female	(3) High cost sharing	(4) Care level	(5) Receiving terminal care	(6) Length of stay
Short-stay admission in pp	0.09056	0.00937**	0.00120	-0.00226	0.000598	1.578*
	(0.06669)	(0.00380)	(0.00148)	(0.01069)	(0.000769)	(0.8861)
Mean outcome	85.11	0.6752	0.0342	3.21	0.0109	121.61
N	599,946	599,946	599,946	599,946	599,946	599,946

*** $p<0.01$, **$p<0.05$, *$p<0.1$

Note: This table presents the result of regression of patient characteristic (indicated by each column) on our instrument (average number of short-stay admissions during episode, expressed as a percentage of capacity) and controls other than the dependent variable. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

Table A3: Instrumental Variable Regression of Patient Outcomes on Occupancy (Admissions IV)

	(1) First Occupancy (pp)	(2) Second Home Discharge (pp)	(3) Second Hospitalization (pp)	(4) Second Death (pp)
Occupancy (pp)		-1.371**	3.157***	0.278
		(0.668)	(0.803)	(0.243)
Short-stay admission (pp)	-0.714***			
	(0.122)			
N	599,946	599,946	599,946	599,946

*** p<0.01, **p<0.05, *p<0.1

Note: This table presents the results of instrumental variable regressions of patient outcomes on occupancy and other controls, using average short-stay admissions during the episode as an instrument. Column (1) displays the first-stage coefficient on the instrument and Columns (2)-(4) display the second-stage coefficients on occupancy. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

Table A4: Instrumental Variable Regression of Patient Outcomes on Occupancy (Before-Episode Admission IV)

	(1) First Occupancy (pp)	(2) Second Home Discharge (pp)	(3) Second Hospitalization (pp)	(4) Second Death (pp)
Occupancy (pp)		-1.802***	2.157***	-0.299
		(0.643)	(0.606)	(0.255)
Short-stay admission (pp) (before admission)	-0.410***			
	(0.0334)			
N	593,518	593,518	593,518	593,518

*** p<0.01, **p<0.05, *p<0.1

Note: This table presents the results of instrumental variable regressions of patient outcomes on occupancy and other controls, using average short-stay admissions in the 14 days preceding admission as an instrument, controlling for average short-stay admissions during the episode. Column (1) displays the first-stage coefficient on the instrument and Columns (2)-(4) display the second-stage coefficients on occupancy. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

Table A5: OLS Estimates of the coefficient on Occupancy

	(1) Home Discharge (pp)	(2) Hospitalization (pp)	(3) Death (pp)
Occupancy (pp)	-0.00853	0.000176	-0.00197
	(0.0136)	(0.0141)	(0.00688)
N	599,946	599,946	599,946

*** p<0.01, **p<0.05, *p<0.1

Note: This table presents the results of OLS regressions of patient outcomes on occupancy and other controls. Controls included are: age, female indicator, indicator of high cost sharing, care level, indicator of receiving terminal care, average of these variables at the provider during the episode, length of stay, local hospital capacity, discharge date fixed effects and medical area by discharge year-month fixed effects, as well as provider fixed effects.

B Proofs and Additional Results

B.1 Proof of Proposition 1.

By Assumption 1(i)(ii), the probability that i is admitted to j can be written as

$$
\begin{aligned}
l_{ij}(x_i, n_i) &= \Pr\left(j_i = j | x_i, n_i\right) \\
&= \sum_{C \subseteq \mathcal{J}} \Pr\left(C_i = C | x_i, n_i\right) \Pr\left(j_i = j | C_i = C, x_i\right) \\
&= \sum_{C \subseteq \mathcal{C}(j)} \prod_{k \in C} F_{\eta_k | x_{ik}}\left(v_k(x_{ik}, n_{ik})\right) \prod_{l \in \mathcal{J} \setminus C} \left[1 - F_{\eta_l | x_{il}}\left(v_l(x_{il}, n_{il})\right)\right] \Pr\left(j | C, x_i\right) \\
&= F_{\eta_j | x_{ij}}\left(v_j(x_{ij}, n_{ij})\right) \\
&\quad \times \sum_{C \subseteq \mathcal{C}(j)} \prod_{k \in C \setminus \{j\}} F_{\eta_k | x_{ik}}\left(v_k(x_{ik}, n_{ik})\right) \prod_{l \in \mathcal{J} \setminus C} \left[1 - F_{\eta_l | x_{il}}\left(v_l(x_{il}, n_{il})\right)\right] \Pr\left(j | C, x_i\right)
\end{aligned}
$$

where the third equality holds because of independence between η_i and ε_i and independence of η_{ij} across j, with $\mathcal{C}(j)$ denoting the collection of subsets of $\mathcal{J}$ that contain j.

Below, we proceed by slightly relaxing Assumption 1(iii):

Assumption 1(iii-b) *For each j, (i) we know x_{ij}^0 such that* $\lim_{n_{ij} \to \underline{n}_j} F_{\eta_j | x_{ij}^0}\left(v_j(x_{ij}^0, n_{ij})\right) = 1$. *(ii) we know n_j^0 such that* $\lim_{n_{ij} \to n_j^0} F_{\eta_j | x_{ij}}\left(v_j(x_{ij}, n_{ij})\right) > 0$ *does not depend on x_{ij}.*

This assumption slightly relaxes Assumption 1(iii). For example, with $v(x, n) = \alpha + \beta x \cdot \left(\frac{1}{n} - 2\right)$, $n \in (0, 1)$, $\beta > 0$ and $F_{\eta_j | x_{ij}} = F_{\eta_j}$, Assumption 1(iii) does not hold at $x = 0$ but Assumption 1(iii-b) still holds with $x^0 = 1$ and $n^0 = 0.5$.

Fix j. Let $n = (n_1, \ldots, n_J)$ and $n' = \left(n_1, \ldots, n_{j-1}, n_j', n_{j+1}, \ldots, n_J\right)$ be any two vectors of daily occupancy which only differ in j's occupancy. We then have $\frac{l_{ij}(x_i, n)}{l_{ij}(x_i, n')} = \frac{F_{\eta_j | x_{ij}}\left(v_j(x_{ij}, n_j)\right)}{F_{\eta_j | x_{ij}}\left(v_j(x_{ij}, n_j')\right)}$. Therefore, by Assumption 1(iii-b), at $x_i = x_i^0 = (x_{i1}^0, \ldots, x_{iJ}^0)$,

$$
F_{\eta_j | x_{ij}^0}\left(v_j(x_i^0, n_j)\right) = \lim_{n_{j'} \to \underline{n}_j} \frac{F_{\eta_j | x_{ij}^0}\left(v_j(x_i^0, n_j)\right)}{F_{\eta_j | x_{ij}^0}\left(v_j(x_i^0, n_j')\right)} = \lim_{n_{j'} \to \underline{n}_j} \frac{l_{ij}(x_i^0, n)}{l_{ij}(x_i^0, n')}
$$

is identified at all n_j.

Next, at any (x_{ij}, n_j),

$$\begin{aligned} F_{\eta_j|x_{ij}}\left(v_j(x_{ij}, n_j)\right) &= \lim_{n'_j \to n^0_j} \frac{F_{\eta_j|x_{ij}}\left(v_j(x_{ij}, n_j)\right)}{F_{\eta_j|x_{ij}}\left(v_j(x_{ij}, n'_j)\right)} F_{\eta_j|x_{ij}}\left(v_j(x_{ij}, n'_j)\right) \\ &= \lim_{n'_j \to n^0_j} \frac{l_{ij}(x_i, n)}{l_{ij}(x_i, n')} \lim_{n'_j \to n^0_j} F_{\eta_j|x^0_{ij}}\left(v_j(x^0_{ij}, n'_j)\right), \end{aligned}$$

is identified. Repeating this argument, we identify $F_{\eta_{j'}|x_{ij'}}\left(v_{j'}(x_{ij'}, n_{j'})\right)$ for all j', $x_{ij'}$ and $n_{j'}$.

Now, pick any x_i and $n^{1,j}, \ldots, n^{R,j}$ that satisfies Assumption 2. Stack the above equations at this value in a vector

$$\begin{aligned} L_j(x_i, \mathbf{n^j}) &\equiv \begin{bmatrix} l_{ij}(x_i, n^{1,j}) \\ \\ l_{ij}(x_i, n^{R,j}) \end{bmatrix} \\ &= \begin{bmatrix} F\left(C^j_1; x_i, n^{1,j}\right) & \cdots & F\left(C^j_K; x_i, n^{1,j}\right) \\ \vdots & \ddots & \vdots \\ F\left(C^j_1; x_i, n^{R,j}\right) & \cdots & F\left(C^j_K; x_i, n^{R,j}\right) \end{bmatrix} \begin{bmatrix} \Pr\left(j_i = j | C_i = C^j_1, x_i\right) \\ \\ \Pr\left(j_i = j | C_i = C^j_K, x_i\right) \end{bmatrix} \\ &= F^j(x_i, \mathbf{n^j}) P_j(x_i) \end{aligned}$$

and stack the matrices further as

$$\begin{aligned} L(x_i, \mathbf{n}) \equiv \begin{bmatrix} L_1(x_i, \mathbf{n^1}) \\ \\ L_J(x_i, \mathbf{n^J}) \end{bmatrix} &= \begin{bmatrix} F_1(x_i, \mathbf{n^1}) & & 0 \\ & \ddots & \\ 0 & & F_J(x_i, \mathbf{n^J}) \end{bmatrix} \begin{bmatrix} P_1(x_i) \\ \\ P_J(x_i) \end{bmatrix} \\ &\equiv F(x_i, \mathbf{n}) P(x_i). \end{aligned}$$

We then identify the choice probabilities as $P(x_i) = \left[F(x_i, \mathbf{n})^\top F(x_i, \mathbf{n})\right]^{-1} F(x_i, \mathbf{n})^\top L(x_i)$. We can repeat this for all x_i. ■

B.2 Identification of Admission Parameters

We discuss identification of providers' utility function v and distribution of idiosyncratic cost, $F_{\eta_j|x_{ij}}$.

Assumption 4 *For each j,*

(i) the utility of not offering i an admission is normalized to zero.

(ii) $F_{\eta_j|x_{ij}}$ is strictly increasing and differentiable, and v_j is continuously differentiable in n_j and some element $x_{ij}^{(s)}$ with a finite derivative.

(iii) $\frac{\partial v_j(x_{ij},n_j)}{\partial x_{ij}^{(s)}} \overset{restrict}{=} \beta_s \overset{normalize}{=} 1$.

(iv) $F_{\eta_j|x_{ij}}$ does not depend on $x_{ij}^{(s)}$.

(v) the τ-th quantile of $\eta_j|x_{ij}$ is known. For each x_{ij}, we observe n_j such that $F_{\eta_j|x_{ij}}(v_j(x_{ij}, n_j)) = \tau$.

Assumption 4(i) is standard normalization. Assumptions 4(ii), (iii) and (iv) together allow us to identify the derivative of v_j with respect to n_j. Note that we do not assume $x_{ij}^{(s)}$ has a large support. Finally, Assumption 4(v) fixes the location. It holds if, e.g., for all x_{ij}, $F_{\eta_j|x_{ij}}(\cdot)$ and $v_j(x_{ij}, \cdot)$ are continuous, $v_j(x_{ij}, \cdot)$ is decreasing and $F_{\eta_j|x_{ij}}(v_j(x_{ij}, \bar{n}_j)) < \tau < F_{\eta_j|x_{ij}}\big(v_j(x_{ij}, \underline{n}_j)\big)$. A more specific example is that $\eta_j|x_{ij}$ has conditional median zero, and that $F_{\eta_j|x_{ij}}(v_j(x_{ij}, n_j)) = 0.5$ for some n_j.

Proposition 2 *Suppose that Assumption 4 holds and we observe $F_{\eta_j|x_{ij}}(v_j(x_i, n_j))$ for all (j, x_{ij}, n_j). Then $v_j(x_{ij}, n_j)$ is identified for all (j, x_{ij}, n_j). Moreover, $F_{\eta_j|x_{ij}}$, $x_{ij} \in \mathcal{X}_j$, is identified on $v_j(\mathcal{X}_j, \mathcal{N}_j)$, where $\mathcal{X}_j$ and $\mathcal{N}_j$ denote the support of x_{ij} and n_j, respectively.*

Proof. By Assumptions 4(ii)(iii)(iv),

$$\frac{\frac{\partial F_{\eta_j|x_{ij}}(v_j(x_{ij},n_j))}{\partial n_j}}{\frac{\partial F_{\eta_j|x_{ij}}(v_j(x_{ij},n_j))}{\partial x_{ij}^{(s)}}} = \frac{\frac{\partial v_j(x_{ij},n_j)}{\partial n_j}}{\frac{\partial v_j(x_{ij},n_j)}{\partial x_{ij}^{(s)}}} = \frac{\partial v_j(x_{ij}, n_j)}{\partial n_j}.$$

is identified for all (x_{ij}, n_j). Take any x_{ij}. We observe some $n_j(x_{ij})$ such that $v_j(x_{ij}, n_j(x_{ij})) = q_j^\tau(x_{ij})$, where $q_j^\tau(x_{ij})$ is the known τ-th quantile of $\eta_j|x_{ij}$. Therefore,

$$v_j(x_{ij}, n_j) = q_j^\tau(x_{ij}) + \int_{n_j(x_{ij})}^{n_j} \frac{\partial v_j(x_{ij}, n_j')}{\partial n_j'} dF_j(n_j'),$$

where $F_j(n_j)$ is the distribution function of n_j, is identified. Thus, at any $(x_{ij}, n_j) \in \mathcal{X}_j \times \mathcal{N}_j$, $F_{\eta_j|x_{ij}}(v_j(x_{ij}, n_j))$ is identified. ■

C Estimation Details

Note the individual likelihood can be rewritten as

$$\begin{aligned} l_{ij} &= \sum_{C \subseteq \mathcal{J}} \prod_{k \in C} F_{ikt}(\theta^f) \prod_{l \in \mathcal{J} \setminus C} \left(1 - F_{ilt}(\theta^f)\right) P_{ijt}(C; \theta^p) \\ &= \sum_{C \subseteq \mathcal{J}} Q_{it}(C; \theta^f) P_{ijt}(C; \theta^p) \\ &= \sum_{C \subseteq \mathcal{J}} \left\{ \frac{Q_{it}(C; \theta^f)}{Q_{it}(C; \theta_0^f)} P_{ijt}(C; \theta^p) \right\} Q_{it}(C; \theta_0^f) \end{aligned}$$

where θ_0 is an initial value. This suggests the following simulation algorithm:

1. Draw $U_{ijt}^r \sim U[0, 1]$, $r = 1, \ldots, R$, for each i and j.
2. Obtain $C_{it}^r = \left\{ j \in \mathcal{J}_t : F_{ijt}\left(\theta_0^f\right) \geq U_{ijt}^r \right\}$ for each i and r.
3. Calculate $P_{ijt}(C_{it}^r; \theta^p) = I(j \in C_{it}^r) \frac{\exp(\bar{u}_{ijt})}{1 + \sum_{j' \in C_{it}^r} \exp\left(\bar{u}_{ij't}\right)}$.
4. Calculate $\hat{l}_{ijt} = \frac{1}{R} \sum_{r=1}^R \frac{Q_{it}\left(C_{it}^r; \theta^f\right)}{Q_{it}\left(C_{it}^r; \theta_0^f\right)} P_{ij}(C_{it}^r; \theta^p)$.

The simulated likelihood is smooth in parameters. Moreover, we can hold simulated consideration sets fixed throughout estimation.