Estimating Nursing Home Quality with Selection*

Andrew Olenski[1] Szymon Sacher[2]

July 17, 2023

Abstract

We use variational inference (VI), a technique from the machine learning literature, to estimate a mortality-based Bayesian model of nursing home quality accounting for selection. We demonstrate how one can use VI to quickly and flexibly estimate a high-dimensional economic model with large datasets. Using our facility quality estimates, we examine the correlates of quality and find that public report cards have near-zero correlation. We then show that in contrast to prior literature, higher quality nursing homes fared better during the pandemic: a one standard deviation increase in quality corresponds to 2.5% fewer Covid-19 cases.

*We are very grateful to Charles Angelucci, Matt Backus, Bo Cowgill, Gautam Gowrisankaran, Stephen Hansen, Suresh Naidu, Adam Sacarny, Ashley Swanson, and Pietro Tebaldi for their helpful comments and suggestions. We thank Mohan Ramanujan, Elizabeth Adams, and Maurice Dalton for their help in obtaining and managing the data. Olenski gratefully acknowledges support from the National Science Foundation Graduate Research Fellowship.

[1] Department of Economics, Lehigh University. Email: ano223@lehigh.edu

[2] Graduate School of Business, Stanford University. Email: sacher@stanford.edu

1 Introduction

In order for health care markets to function well, it is critical for patients and insurers to be able to observe provider performance. Lacking information, consumers will not be able to identify and select better providers, which in turn under-incentivizes investment in quality (Dranove and Satterthwaite 1992). Yet, quality measurement is notoriously challenging, and this complexity can diminish the returns to efforts to improve information. The nursing home industry may be the poster child of this problem. Despite significant investments in improving measurement, quality remains infamously low.[1] Meanwhile, multiple investigations have found that the public report cards that serve as the primary system of quality measurement are subject to rampant manipulation, as facilities systematically misreport information so as to maximize their ratings (Thomas 2014; Silver-Greenberg and Gebeloff 2021).

In this paper, we propose a new method to estimate the causal effect of admission to a given nursing home on 90-day survival, while controlling for patient selection. Though averting mortality is not the exclusive aim of nursing home care, this causal effect is a relevant dimension of facility quality, and may be used to evaluate the validity of existing measures. We account for endogenous patient selection using the distance between the facility and each patient's home zip code as an instrumental variable. While this is a common approach in the literature (c.f. Grabowski et al. (2013) and Einav, Finkelstein, and Mahoney (2022)), we also conduct several tests of the exclusion restriction.

To link this quasi-experimental variation in facility choice with our binary health outcome measure, we propose a structural Bayesian model of nursing home quality, in the style of Geweke, Gowrisankaran, and Town (2003) and Hull (2020). Due to the large number of parameters, typical sampling-based estimation approaches are computationally infeasible, and so we adapt a technique from the machine learning literature ('variational inference') tailored for models in which the number of parameters often stretches into the millions. This approach replaces the objective function with an approximation, thereby facilitating optimization. We show in a simulation exercise that this procedure recovers the true values and offers a substantial speedup against conventional MCMC-based approaches. In doing so, we demonstrate that VI may be used as a general purpose estimation algorithm for a wide class of computationally challenging models within economics. Despite its popularity in computer science, VI has seen very limited adoption in economics; however, now that desirable theoretical properties have been established (Wang and Blei 2019; Medina, Olea, Rush, and Velez 2021), we anticipate interest in the method to grow substantially.

1. For instance, one report found that more than one in five Medicare patients who stayed in a nursing facility for 35 days or fewer experienced harm as a result of their medical care (Office of Inspector General 2014).

Using the facility-level quality estimates, we then answer two questions. First, we evaluate the performance of the public report card system. These scores, intended to serve as easy reference points for consumers, are organized into "star ratings" and include measures of staffing, facility inspections, and patient-level outcome measures. Prior work studying similar report cards in other health care settings has found mixed results (Abaluck, Caceres Bravo, Hull, and Starc 2021; Doyle, Graves, and Gruber 2019). We find that that there is near-zero correlation between the report cards and our survival-based quality estimates. Even conditioning on narrow measures of geography, very little of the variation in nursing home quality can be explained by the variation in the star ratings.

Second, we apply the estimates to a question that has emerged in the wake of the Covid-19 pandemic's devastation of the nursing home industry: did higher quality facilities fare better? A flurry of recent articles have found no relationship between Covid-19 and star ratings, with one recent systematic review concluding that providers therefore had few levers available to combat the spread of Covid-19 (Konetzka et al. 2021). We re-visit this question using our quality estimates, and find that contrary to the consensus, higher quality nursing homes indeed fared better during the Covid-19 pandemic: facilities with one standard deviation higher quality had 2.5% fewer Covid-19 cases per bed, suggesting there may be some scope to limit adverse outcomes.

This paper primarily contributes to a considerable literature on quality estimation in the health care sector. The majority of work in this area studies hospitals: for instance, Gowrisankaran and Town (1999), Doyle, Graves, and Gruber (2019), Geweke, Gowrisankaran, and Town (2003) and Hull (2020) all estimate models of hospital quality using mortality. The latter two derive models from which we borrow heavily. For these models, we show the power of VI as a general-purpose estimation algorithm, particularly in cases where traditional methods can be computationally intractable.

Our work is closely related to contemporaneous research by Einav, Finkelstein, and Mahoney (2022) (hereafter, EFM). To study nursing home value-added, they construct a unidimensional health index that measures how fit a patient is to return to the community, and study the average change in this index over 30 days. Doing so requires estimating a model of "selection into" nursing homes – similar to ours – and "selection out," where discharge decisions are a function of the health index. This health index arguably comprises a richer, albeit less transparent, measure of value than the binary 90-day survival outcome that we use. However, this added richness comes at the cost of additional assumptions, as the authors must account for selection out of the facility. Because we observe mortality for all residents, we only have to grapple with selection in. Additionally, our estimation method allows us to consider larger markets, resulting in a small share of patients choosing a facility outside

their local market.

In terms of measuring quality and value-added (two tightly intertwined concepts[2]), our findings are complementary with those of EFM. We both find that public report cards are poor predictors of the causal effect of a given nursing home admission on health outcomes. However, our studies differ in application: EFM document considerable within- and cross-market variation in nursing home value-added, an issue we sidestep. Instead, we show that our quality estimates predict out-of-sample heterogeneity in nursing home outcomes during the Covid-19 pandemic.

2 Institutional Details and Data

2.1 Specialization and Selection

This paper considers quality estimation of CMS-certified skilled nursing facilities (SNFs), commonly referred to as nursing homes. SNFs are health care facilities that provide a wide range of services, including skilled nursing care, specialized rehabilitative services such as physical therapy, occupational therapy, and speech therapy, in addition to treatment for the mentally ill and developmentally disabled.

Given the diversity of treatment requirements across patient types, it is natural that SNFs specialize in their care provision (Mor, Banaszak-Holl, and Zinn 1995). Facilities invest in specialty care units, such as Alzheimer's wards, which may affect the types of patients who select these facilities. In addition to patient-side selection, a pair of recent papers present evidence that nursing homes themselves may worsen this selection problem by strategically admitting and discharging less profitable patients as the opportunity cost of maintaining those patients increases (Gandhi 2020; Hackmann, Pohl, and Ziebarth 2021). These papers raise an additional concern. To the extent that some nursing homes are better able to select their patient censuses on unobserved risk, quality estimation based purely on risk-adjustment would be biased. Specifically, one would systematically overestimate quality for more selective facilities, and underestimate quality for less selective facilities.

2.2 Nursing Home Compare

The most prominent public reform to raise quality is the introduction of the Nursing Home Compare website by CMS. NHC provides public report cards on each SNF, aggregating in-

2. EFM model within-patient changes in health status, and as such refer to their facility-level parameters as facility "value-added." In contrast, our set-up is closer to the traditional health care context, in which we identify institutional quality using (selection-adjusted) mean outcomes across facilities.

formation on patient outcomes and process measures, results from recent facility inspections, and data on levels of staffing. Facilities are assigned a star rating out of five for each of the components, in addition to an overall composite rating. The logic of NHC is to reduce information asymmetry; by giving consumers more information, they can choose higher quality providers.

While the NHC star ratings are widely used measures of quality by both consumers and researchers, there are substantial concerns regarding their accuracy. Two investigations found that even 5-Star facilities can have horrific living conditions, and consumers often feel misled by the system (Thomas 2014; Silver-Greenberg and Gebeloff 2021). In a comprehensive overview of the literature on the NHC system, Konetzka, Yan, and Werner (2021) provide two potential explanations for the system's shortcomings. The first is a standard multitasking moral hazard problem: because NHC includes relatively narrow quality measures, facilities can target these measures while shirking on excluded measures (Holmstrom and Milgrom 1991). For example, NHC initially reported the use of physical restraints, but not antipsychotic use – a class of drugs commonly misused as sedatives for the nursing home population, against clinical guidance. Konetzka, Brauner, Shega, and Werner (2014) find that public reporting of only physical restraints led to an increased use of antipsychotics among residents with severe cognitive impairments.

The second issue arises from the fact that two components of the NHC star ratings (the staffing and quality measures) are partially derived from records the facilities themselves submit, and suggesting substantial bandwidth for manipulation. By nature, research on this front must be indirect, but nonetheless the findings support claims of manipulation. For example, despite the large increases reported in staffing ratios under NHC, there was little evidence of increases in staffing costs (Sharma, Konetzka, and Smieliauskas 2019). In light of these shortcomings, one goal of this paper is to evaluate the correlation between NHC star ratings and survival-based quality estimates.

2.3 Data

We combine a variety of administrative and public data sources from the Centers for Medicare & Medicaid Services. The base of our analytic file is resident-level assessment data from the Minimum Data Set (MDS) spanning 2000-2017. All CMS-certified SNFs are required to complete regular assessments of each resident, beginning at admission. These assessments include a high-dimensional array of clinical and demographic information that are reported to CMS to develop quality metrics and determine reimbursement rates. As we are interested

in mortality regardless of where the beneficiary was at the time of death,[3] we merge the MDS data with the annual Medicare enrollment files, which contain each beneficiary's home zip code as well as their date of death, through 2017.

Patient data is supplemented by several public datasets containing information on the SNFs themselves. These include the LTCFocus files, which collect facility-level characteristics (such as location, size, and ownership status) from administrative sources, as well as the quarterly NHC 5-Star Ratings.[4] To examine the relationship between SNF quality and Covid-19 outcomes, we collect the cumulative number of confirmed cases and deaths through December 19th, 2021 in addition to the resident and staff vaccination rates, available from CMS.[5]

2.4 Sample Construction

As we study survival to 90 days after admissions, when constructing our analytic sample we make use only of information available on the initial admission MDS assessment. We identify all new nursing home admissions, defined as no prior admission assessment at any SNF in the prior 365 days. Further, we restrict the assessments to patients who are enrolled in the Medicare program, which is required to track mortality and residence prior to admission. We define each patient's choice set as the 100 nearest within-state SNFs to their home zip code centroid.[6] We exclude any admissions to nursing homes outside the choice set, including travel to any out-of-state SNFs. This restriction drops 4.5% of the sample. To increase statistical power while allowing time-varying quality estimates, we estimate the model over 4-year bins. This procedure leaves us with facility-level estimates for the year-bins 2001-2004, 2005-2008, 2009-2012, and 2013-2016, across each state separately.

All together, these restrictions leave us with a sample of 20,538,183 new nursing home admissions across all state-year-bins. Summary statistics are given in Appendix Table D.1. Reflecting the overall demographics of the nursing home population, our sample is disproportionately white (86.1%), older (average age 79.3), traveled 7.8 miles to their admitting SNF, and have an average 90 day survival rate of 85.8%.

3. For instance, a resident may be discharged to the hospital where their death may not be recorded in the MDS assessments.

4. The LTCFocus data are provided by the Shaping Long Term Care in America Project at Brown University and funded in part by the National Institute on Aging (1P01AG027296).

5. The latest available data at the time of download from https://data.cms.gov/covid-19/covid-19-nursing-home-data.

6. Our quality estimates are robust to alternative choice sets: we find a 0.984 correlation between our main estimates and those recovered when imposing $J = 50$ (Appendix Figure D.1).

3 Model of Nursing Home Quality

To accommodate the potential for endogenous selection, we specify the following two-index model that is similar in spirit to Geweke, Gowrisankaran, and Town (2003) and Hull (2020).

3.1 Model Set-Up

A patient i who chooses SNF j has a latent health index, h_{ij}, that depends on j's quality β_j, patient demographics and comorbidities X_i and a health shock ε_{ij}. The potential outcome (90-day survival) of patient i at SNF j is Y_{ij}. These are expressed as:

$$h_{ij} = \beta_j + \gamma X_i^T + \varepsilon_{ij} \tag{1}$$

$$Y_{ij} = \mathbb{1}\left[h_{ij} > 0\right] \tag{2}$$

Patients do not choose SNFs at random. Instead, they choose the SNF that yields the highest utility, u_{ij}, defined as:

$$u_{ij} = \underbrace{\xi_j + \pi Z_{ij}^T}_{\delta_{ij}} + \eta_{ij} \tag{3}$$

Where ξ_j parametrizes constant SNF popularity and Z_{ij} is a vector of utility shifters. Most important among these is the geodetic distance between patient i and SNF j.[7] Maximization of this latent utility index over choice of j yields the patient's choice indicator, equal to one if patient i chooses SNF j, zero otherwise:

$$D_{ij} = \mathbb{1}\left[u_{ij} > u_{ik}, \forall k \neq j\right] \tag{4}$$

We model selection as a facility-specific correlation between the preference shock η_{ij} and the health shock ε_{ij}.[8] We assume that the shocks are jointly normally distributed with mean zero, unit variance and covariance α_j. Specifically, define the health shock as $\varepsilon_{ij} = \alpha_j \eta_{ij} + \sqrt{1 - \alpha_j^2}\tilde{\varepsilon}_{ij}$ where η_{ij} and $\tilde{\varepsilon}_{ij}$ are independent standard normal random variables.[9] Notice that because the probability of selecting SNF j is increasing in $\eta_{i,j}$, facilities with higher draws of α_j will have more favorable 'unobserved'[10] selection.

7. Here we draw upon the vast literature in health economics documenting a strong preference for proximity when selecting providers.

8. Notice that the scale of the latent indices h_{ij} and u_{ij} is unidentified. Therefore, without loss of generality we normalize the scale of the shocks to one: $\sigma_\varepsilon = \sigma_\eta = 1$, to mirror a conventional probit model.

9. This common independence assumption is nonetheless a substantial modeling restriction, as it may imply unrealistic substitution patterns. However, it is unclear how violations of this restriction would impact our subsequent estimates of β_j.

10. Specifically, unobserved in patient characteristics X_i.

3.2 Econometric Framework

Estimation requires evaluating the log-probability of both the observed choices and survival outcomes. Typically, estimation of discrete choice models relies on the existence of closed-form probabilities, which our model does not produce. Instead, our approach starts by observing that, conditional on the structural parameters and, crucially, the realized preference shock for the selected SNF $\eta_{i,j(i)}$, the log-probabilites take a simple form:

$$\log P(D_{ij} = 1 \mid \theta_i, X_i, Z_i) = \Sigma_{j' \neq j(i)} \log \Phi(\delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)}) \tag{5}$$

$$\log P(Y_{i,j(i)} = 1 \mid \theta_i, X_i, Z_i) = \log \left(1 - \Phi \left(\frac{-(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)} \eta_{i,j(i)})}{\sqrt{1 - \alpha_{j(i)}^2}} \right) \right) \tag{6}$$

where $\theta_i := (\beta, \alpha, \gamma, \xi, \pi, \eta_{i,j(i)})$ is a vector that stacks all of the parameters we estimate, δ_{ij} is the observed component of the utility defined in Equation (3), and Φ is the standard normal cumulative distribution function. Appendix B contains the derivations and further details.

This approach bears resemblance to a Heckman correction, as we model the selection problem instead as an omitted variable problem (Heckman 1979). Moreover, we reduce the high-dimensional selection problem (a preference shock for each SNF in the choice set) to a single, sufficient, estimable parameter for each individual (the value of the preference shock for only the chosen SNF). The quantities in equations (5) and (6) are easy to compute. By conditioning on the value of $\eta_{i,j(i)}$ we only need to evaluate the CDF of the univariate standard normal, for which highly efficient, differentiable approximations exist. While this comes at the cost of estimating $\eta_{i,j(i)}$ for each individual, with VI we are able to easily estimate a model in which the number of parameters grows linearly with the sample size.

When estimating quality, one commonly needs to account for few observations in some facilities, which may lead to noisy quality estimates. To accommodate this, we use Bayesian inference. We specify the priors in Appendix Section B.2. Importantly, we use hierarchical priors for the quality parameters β_j, which implicitly regularize the quality estimates towards the mean, increasing reliability especially for small SNFs. Empirical Bayes methods commonly used in quality estimation (e.g. Chetty, Friedman, and Rockoff 2014; Chandra, Finkelstein, Sacarny, and Syverson 2016) can be viewed as first-order approximation to the hierarchical model and it can be shown that the difference between estimates is of the order $O(J^{-1})$ (Kass and Steffey 1989). The advantage of the hierarchical model is that it ensures that the prior variance on the quality parameters is estimated and guaranteed to be consistent with the across-facility posterior variance, which is not the case for empirical Bayes

methods (Gelman 2014, Chapter 5).

4 Estimation with Variational Inference

Estimating a choice model such as ours over a large number of admissions (more than 20 million) is computationally challenging. One common solution to relieve the computational burden is to estimate separately across narrowly defined geographic markets. This solution imposes strict assumptions on patients' choice sets, resulting in limited substitution patterns between nursing homes. In addition to the statistical efficiency concerns from losing observations, in our setting it is possible that patients who travel farther into other markets may not be representative of all patients who select a given facility, giving rise to potentially biased estimates. We relax this assumption by defining large person-specific choice sets, as we do not need to impose narrow geographic markets to ensure tractability.

Defining such rich choice sets allows us to include many more observations in our estimation; however, doing so is computationally taxing, as each new observation brings with it another parameter (the shock $\eta_{ij(i)}$ for the chosen SNF). Given the size of our data, conventional approaches, such as an MCMC sampler, are computationally infeasible. As such, we use variational inference (VI), a technique from the machine learning literature commonly employed for efficiently estimating complex, high-dimensional Bayesian models in which the number of parameters often stretches into the millions. The fundamental idea of VI is to reframe the estimation problem as one of optimization rather than sampling, instead using an approximation of the posterior distribution $p(\theta|x)$ for each parameter θ given the data x.

4.1 Overview of Variational Inference

Estimating an economic model via variational inference is very rare, and so many readers may be unfamiliar with the approach. As such, we provide a quick overview of the method, with a fuller description available in Appendix Section A.

To implement VI, the researcher starts by specifying a family of distributions Q_Ψ (e.g. the family of normal distributions) indexed by a parameter $\psi \in \Psi$ (e.g. $\psi = (\mu, \sigma^2)$, the mean and variance). She then searches over possible values of the parameters ψ so as to minimize the 'distance' between the posterior $p(\theta|x)$ and the approximating distribution $q(\theta|\psi) \in Q_\Psi$. The distance between these two distributions is measured by the Kullback-Leibler divergence,

which is defined as:

$$\begin{aligned} \mathrm{KL}\big(q(\theta|\psi)||p(\theta|x)\big) &= \int q(\theta|\psi) \log \frac{q(\theta|\psi)}{p(\theta|x)} d\theta \\ &= \mathbb{E}_q[\log q(\theta|\psi) - \log p(\theta|x)]. \end{aligned} \tag{7}$$

The resulting optimization problem is:

$$\begin{aligned} \psi^* = \arg\min_{\psi \in \Psi} \quad & \mathbb{E}_q[\log q(\theta|\psi) - \log p(\theta|x)] \\ \text{s.t.} \quad & q(\theta|\psi) \in Q_\Psi \end{aligned} \tag{8}$$

The 'trick' which enables this optimization follows from Bayes' rule. While it is straightforward to compute $\log q(\theta|\psi)$, the researcher still needs to compute the true posterior $\log p(\theta|x)$. Notice that $\log p(\theta|x) = \log p(\theta, x) - \log p(x)$. The first term, the joint density of data and parameters, is easily computed. The second term, $\log p(x)$, is computationally intractable, but because it is invariant to ψ may be safely ignored in optimization.

VI offers several potential benefits for estimating economic models. Primarily, as an optimization-based approach it typically achieves faster convergence than traditional sampling-based approaches (Blei, Kucukelbir, and McAuliffe 2017). This improvement in computation time is largely owed to recent advancements in optimization, such as automatic differentiation and massive parallelization. Additionally, VI is a very flexible inference algorithm. Minimal model-specific derivations are necessary, enabling researchers to easily estimate variations of their model. For example, they can add parameters or modify distributions without encountering substantial difficulties, allowing one to assess result sensitivity to functional form assumptions.

There are applications for which VI may be less appropriate than sampling-based approaches. While MCMC methods tend to be slower, they are asymptotically exact. VI is an approximation method, the quality of the which depends on several factors, including the choice of the approximating distribution Q_Ψ, which introduces some limitations. For instance, while the point estimates are guaranteed to be consistent, VI may underestimate the posterior variances (more on this below). This suggests that VI is well-suited for large datasets, applications for which the point estimates are of primary interest, and cases in which MCMC methods are computationally infeasible. Of course, if one finds that exact methods do not pose a computational problem, using such approaches is advisable.

4.2 Implementation Details and Simulation Results

We employ a common variant known as mean-field variational inference. In this approach, the approximating distribution is taken to be fully factorized, i.e. $q(\theta|\psi) = \prod_{i=1}^{N} q(\theta_l|\psi_l)$, where θ_l represents the lth element of the θ vector. For all parameters that are defined on the real line, we select the normal distribution as the approximating distribution. As for the parameters α_j, which are constrained to the interval $[-1, 1]$, we utilize a beta distribution after applying an appropriate transformation. Lastly, for the variance of β_j, a hierarchical parameter, we opt for a gamma distribution. It is worth pointing out that the prior distributions we use are members of the same family as the approximating distributions. This choice is not necessary, but it simplifies the optimization problem due to the fact that the KL divergence between two distributions from the same family can usually be computed analytically.[11]

The mean-field assumption greatly simplifies the optimization problem but is not without tradeoffs. Specifically, while the point estimates we obtain are guaranteed to be consistent, the posterior variances may not be (Wang and Blei 2019). We explore these theoretical issues in a simulation exercise.

In the simulations, we generate data in a manner that ensures certain parameters exhibit high correlation in the posterior, thereby pressuring the assumption of fully factorizable approximating distributions. Nonetheless, we find that the point estimates obtained using VI are close to both their true values and to estimates obtained using an MCMC algorithm more common to economics (NUTS, Hoffman and Gelman 2014). The R^2 from a regression of the true β_j on the estimated values is equal to 0.86 while the R^2 from regressing those VI estimates on their MCMC-based equivalents is 0.99. However, for parameters with strong posterior correlations, the VI-based posterior variance is smaller than the asymptotically exact sampling-based posterior variance.

Underestimating the posterior variance, while important, is not a major concern in our application as we are primarily interested in point estimates of the quality parameters β_j and we do not conduct hypothesis tests on the intermediary parameters. In addition, the size of our data implies that the posterior distribution on those parameters is likely to be highly concentrated anyway, so any approximation error would be unlikely to meaningfully affect our conclusions. In Appendix Section A.6 we provide further simulation details, and demonstrate that partially relaxing the mean-field assumption can significantly improve estimates of the posterior variances for cases in which the size of the credible intervals may be of greater importance.

11. If the analytic form of the KL divergence is not available, it can be approximated numerically, but this slows down the optimization process.

5 Identification

Estimating our model presents two identification challenges. First, patients may endogenously sort across facilities based on their unobserved health status. We follow the literature and employ an instrumental variables approach. In particular, we use the geodetic distance between each patient's home zip code centroid and coordinates for each nursing home in the patient's choice set, exploiting the well-established patient preference for nursing facilities close to their home (Grabowski et al. 2013).[12] The identification assumption here is standard: conditional on the risk-adjustors X_i, the distance instruments Z_i are uncorrelated with the idiosyncratic health shock ε_i, and act only to shift preferences towards specific facilities. These instruments help to combat endogenous sorting of patients to particular facilities based on unobserved health status, for instance due to the specialization detailed in Section 2.

Second, because the geographic distribution of health is far from uniform, we also utilize the rich assessment data to incorporate a battery of risk-adjustors X_i in equation (1), so as to allow patients to vary in observed health status. Furthermore, to capture residual geographic variation in unobserved health status, in all our downstream exercises examining variation in facility quality, we include hospital referral region (HRR) fixed effects. This multi-pronged approach accounts for both mean differences in health across regions, as well as endogenous sorting of patients within an area.

While the exclusion restriction is untestable (we cannot examine the relationship between the instruments and the unobserved patient risk ε_i), we can informally assess it by examining how the instrument covaries with the risk-adjustor, X_i. A flat relationship would provide suggestive evidence in support of the exclusion restriction, under the assumption that observed and unobserved patient risks are correlated.

The high-dimensional nature of the instruments (one for each facility in the choice set) complicates efforts to visualize the relationship between the instruments and the risk-adjustors. To address this, we construct a unidimensional differential distance measure, defined as the difference between the distances to each patient's nearest high- and low-quality facilities, respectively. High-quality facilities are those whose $\hat{\beta}_{jt}$ estimate is above the median for that state-year bin, and low-quality facilities are those whose estimate is below the median.

Figure 1 demonstrates the results of this exercise. The top panel examines instrument relevance: we find that patients who are relatively closer to high-quality facilities are significantly more likely to enter such a facility. In the bottom panel, we plot ex ante patient

12. The mean distance traveled to a facility is 12.6 km.

risk (aggregated into a single predicted 90-day survival using the risk-adjustors X_i) alongside the differential distance measure. The figure reports a flat relationship, consistent with prior studies (e.g. Gupta, Howell, Yannelis, and Gupta 2021). To benchmark this result, a 10km increase in differential distance decreases predicted 90-day survival by less than 0.009 percentage points. In contrast, the same change would decrease the probability of entering a high-quality nursing home by 20.8 percentage points.

We also summarize these results in Appendix Table D.2, utilizing a binary measure of differential distance, defined as being relatively closer to high- or low-quality facilities. While we find that some patient characteristics do vary across the two groups, the magnitudes are quite small, and importantly the predicted 90-day survival rates between the two groups are indistinguishable. Despite this similarity in reported health status, we do find that patients who are relatively closer to high-quality facilities observe significantly higher survival rates, suggesting this result is not owed to a lack of statistical power.

6 Correlates of Quality

6.1 Parameter Estimates

Because we estimate the model separately across state-year-bins, for all subsequent exercises we standardize the quality estimates within-market and denote the standardized estimates as $\hat{\beta}_{jt}$, such that a SNF with $\hat{\beta}_{jt} = 1$ has quality that is one standard deviation above the mean for that state-year-bin. For comparison, we also construct quality estimates using a conventional risk-adjustment approach, which we recover by imposing no selection-on-unobservables ($\alpha = 0$) in the model estimation.[13] We denote these non-selection-adjusted estimates as $\hat{\beta}_{jt}^{NA}$.

We find a moderate correlation of $\rho = 0.481$ between the model estimates $\hat{\beta}_{jt}$ and the non-adjusted estimates $\hat{\beta}_{jt}^{NA}$ (Appendix Figure D.2, top panel). This imperfect correlation suggests that unobserved selection is non-trivial – i.e., that there is a role for the distance instruments. Indeed, assessing the correlation between SNF quality $\hat{\beta}_{jt}$ and unobserved selection $\hat{\alpha}_{jt}$, we find a sharp negative relationship ($\rho = -0.32$, Appendix Figure D.2, bottom panel). This negative relationship indicates a traditional adverse selection mechanism: the sickest patients (lowest h_{ij}) have stronger preferences for higher quality SNFs. Because we model travel distances, rather than prices, this result implies that sicker patients travel farther for high-quality SNFs.

Appendix Table D.3 reports univariate regressions of $\hat{\beta}_{jt}$ and $\hat{\beta}_{jt}^{NA}$ on several facility

13. This approach maintains the Bayesian shrinkage of our estimation routine.

characteristics, adjusted for HRR-year fixed effects. Reassuringly, we find that measures of staffing are positively correlated with $\hat{\beta}_{jt}$. Ignoring unobserved selection ($\hat{\beta}_{jt}^{NA}$), we find that for-profit facilities have lower quality, and facilities with specialized Alzheimer's units have higher quality. However, when adjusting for selection ($\hat{\beta}_{jt}$), these relationships disappear, suggesting that unobserved patient composition underlies the results. Overall, we find that very little of the variation in $\hat{\beta}_{jt}$ can be explained by the characteristics of patients or facilities. A regression of $\hat{\beta}_{jt}$ on all of the facility and patient characteristics in Table D.3 has an R^2 of only 0.10.

6.2 Performance of Nursing Home Compare

Table 1 summarizes the findings on the relationship between SNF quality and the Nursing Home Compare star ratings. We conduct the analysis separately for the selection-adjusted estimates $\hat{\beta}_{jt}$ (columns 1-4) as well as the non-adjusted estimates $\hat{\beta}_{jt}^{NA}$ (columns 5-8). Each column represents the coefficients of a regression of quality $\hat{\beta}_{jt}$ on the overall rating, as well as the component ratings for facility inspections ('Survey'), patient-based measures ('Quality'), and overall levels of nurse and nurse-aid staffing ('Staffing'). Partial correlations between the component ratings and $\hat{\beta}_{jt}$, adjusted for geographic fixed effects, are also included. All specifications include HRR-by-year-bin fixed effects.[14] We assign the median star rating for each period. The star ratings were introduced in 2009, and so we include estimates only for the last two year-bins (2009-2012, 2013-2016).

There is startlingly little relationship between the estimates of quality $\hat{\beta}_{jt}$ and the NHC measures. We find a within-HRR correlation of only 0.005. The estimates in Table 1 suggest that moving from a 1-star to 5-star facility corresponds to an increase in quality of only 0.06 standard deviations. The only NHC rating component across which $\hat{\beta}_{jt}$ is monotonically increasing is the 'Quality' rating, which aggregates patient-based quality metrics (derived from the MDS and Medicare claims), but even in this best-performing component, the partial correlation is only 0.084. In contrast, the non-adjusted quality estimates are much more strongly correlated with the NHC ratings. We find a partial correlation of 0.166 between the Overall Rating and $\hat{\beta}_{jt}^{NA}$, suggesting that our results are not driven by a negligible correlation between facility mortality and the NHC ratings.

To examine the magnitude of these results, we conduct a counterfactual exercise in which we replace the 'Quality' component of the NHC Overall rating with our quality estimates.[15]

14. More precisely, because some HRRs cross state boundaries and $\hat{\beta}_{jt}$ is standardized within-state, we instead construct HRR-by-state indicators.

15. We construct five bins of $\hat{\beta}_{jt}$ whose size equal that of the Quality ratings by state, which are then used to compute the Overall Rating in lieu of the Quality component.

We find that doing so would downgrade approximately 27% of 5-star facilities to 3-4 stars, and upgrade only 20% of 1-star facilities to 2-3 stars (Appendix Table D.4).

7 Quality and Covid-19

The Covid-19 pandemic has been particularly devastating for nursing home residents, as SNFs have been centers of outbreak and excess mortality since the start of the crisis. In the wake of the pandemic, a flurry of recent research has sought to examine the determinants of Covid-19 spread in nursing homes, with a particular focus on whether higher quality SNFs performed better in preventing adverse outcomes. Konetzka et al. (2021) provide a systematic review of the literature, and find that there is no relationship between the NHC star ratings and various Covid-19 outcomes. The authors conclude that there is little that SNFs could have done to avert severe outcomes.

Given the lack of correlation between SNF quality $\hat{\beta}_{jt}$ and the star ratings documented in Section 6.2, it is possible that the relatively poor performance of higher rated SNFs is instead due to shortcomings in conventional metrics. We ask whether higher quality SNFs, as measured by $\hat{\beta}_{jt}$, fared relatively better during the first 21 months of the pandemic. Considering the 3-year lag between last of our year-bins (ending in 2016) and the beginning of the Covid-19 pandemic, we first verify that the estimates of $\hat{\beta}_{jt}$ are highly stable across time: the AR(1) regression coefficient of $\hat{\beta}_{jt}$ across year-bins is 0.804 (Appendix Table D.5).

Figure 2 presents binned scatter plots of Covid-19 resident cases and deaths by SNF quality $\hat{\beta}_{jt}$ from the latest available year-bin, 2013-2016. Due to the geographic variation in Covid-19 spread, we adjust the figures for HRR fixed effects.[16] Given the mechanical relationship between size and counts of Covid-19 cases, we follow McGarry, Barnett, Grabowski, and Gandhi (2022) and calculate the number of cases and deaths per bed. The figure suggests that higher quality SNFs experienced lower rates of both Covid-19 cases and deaths.

Given the negative relationship between $\hat{\beta}_{jt}$ and Covid-19 outcomes observed in Figure 2, we ask whether any omitted variables may explain these results. To assess the extent of this possibility, we run a series of regressions of the following form:

$$y_j = \lambda_1 \hat{\beta}_{jt} + \lambda X_j^T + \mu_{m(j)} + \epsilon_j \tag{9}$$

where y_j is one of several Covid-19 outcomes, X_j^T is a vector of facility characteristics, $\mu_{m(j)}$ denotes geographic (HRR or county) fixed effects, and $\hat{\beta}_{jt}$ is the quality estimate. We

16. In a robustness exercise, we also consider narrower geographic fixed effects (county-level) and find similar results (Appendix Figure D.3). A significant number of SNFs operate as the only facility in the county, however, and are therefore dropped from this regression.

include the NHC Overall Rating as it appeared on the website in December 2019. We also include measures of size, the mean household income for the zip code, for-profit ownership and chain membership, the presence of an Alzheimer's unit, and whether the SNF is located in a hospital. We consider several Covid-19 outcomes, including the cumulative number of confirmed resident cases/deaths per bed, as well as staff and resident vaccination rates.

Table 2 presents the results from these regressions for several Covid-19 outcomes. We consider four specifications. In column (1), we replicate the regressions underlying the binscatters in Figure 2, and adjust only for HRR fixed effects. In column (2), we include the NHC Overall Rating. Column (3) includes each of the additional controls contained in X_j. Column (4) replaces the HRR fixed effects with county fixed effects.

Consistent with prior literature, we find that the 2019 NHC Overall Rating is not predictive of resident Covid cases or deaths, though vaccination rates increase with facilities' ratings. In contrast, we find that $\hat{\beta}_{jt}$ significantly predicts resident cases and deaths, and is stable across specifications. Facilities with a one standard deviation higher value of $\hat{\beta}_j$ had 2.5% fewer cases and 3.4% fewer deaths due to Covid-19. We find evidence that higher quality SNFs also vaccinated their staff and residents at higher rates, though these estimates are more sensitive to the choice of controls. While these results explain relatively little of the overall variation in Covid-19 outcomes, they suggest that the reported non-correlation between quality and Covid-19 outcomes is partially due to inadequate measures of quality.

Appendix Table D.7 replicates Table 2, replacing the model estimates $\hat{\beta}_{jt}$ with their non-selection-adjusted counterparts $\hat{\beta}_{jt}^{NA}$. While $\hat{\beta}_{jt}^{NA}$ continue to predict both resident and staff vaccination rates, there is strikingly little correlation with resident cases or deaths. Indeed, in contrast to the main table, we find that facilities with higher naive quality estimates actually fare *worse* along the dimension of staff cases. Similarly, in Appendix Figure D.4 we overlay the relationship between $\hat{\beta}_{jt}^{NA}$ and Covid-19 outcomes on the binscatters presented in Figure 2, and find a flatter relationship.

8 Conclusion

The low quality of care provided at nursing homes has long posed a challenge to policymakers. In this project, we propose a new method of estimating SNF quality. Applying the method to the universe of SNF admissions, we find that conventional quality measures have near-zero correlation with our survival-based approach, and that facility. We also find that higher quality SNFs have fared better during the Covid-19 pandemic, in contrast to the medical consensus. Our results are informative for assessing the performance of conventional quality measures in addition to understanding heterogeneity in the impact of the pandemic.

Finally, our novel methodology provides a blueprint for estimating otherwise intractable high-dimensional economic models.

References

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, no. 3 (June): 1557–1610.

Bingham, Eli, Jonathan P. Chen, Martin Jankowiak, Fritz Obermeyer, Neeraj Pradhan, Theofanis Karaletsos, Rohit Singh, Paul Szerlip, Paul Horsfall, and Noah D. Goodman. 2018. "Pyro: Deep Universal Probabilistic Programming." October 18, 2018.

Blei, David M. 2003. "Latent Dirichlet Allocation." *Journal of Machine Learning Research* 3:993–1022.

Blei, David M., Alp Kucukelbir, and Jon D. McAuliffe. 2017. "Variational Inference: A Review for Statisticians." *Journal of the American Statistical Association* 112, no. 518 (April 3, 2017): 859–877.

Bonhomme, Stéphane. 2021. "Teams: Heterogeneity, Sorting, and Complementarity." *SSRN Electronic Journal.*

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, no. 8 (August): 2110–2144.

Chetty, Raj, John N. Friedman, and Jonah E. Rockoff. 2014. "Measuring the Impacts of Teachers I: Evaluating Bias in Teacher Value-Added Estimates." *American Economic Review* 104, no. 9 (September): 2593–2632.

Doyle, Joseph, John Graves, and Jonathan Gruber. 2019. "Evaluating Measures of Hospital Quality: Evidence from Ambulance Referral Patterns." *The Review of Economics and Statistics* 101, no. 5 (December): 841–852.

Dranove, David, and Mark A. Satterthwaite. 1992. "Monopolistic Competition When Price and Quality Are Imperfectly Observable." *The RAND Journal of Economics* 23 (4): 518–534.

Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2022. *Producing Health: Measuring Value Added of Nursing Homes.* Working Paper. Unpublished.

Gandhi, Ashvin. 2020. *Picking Your Patients: Selective Admissions in the Nursing Home Industry.* SSRN Scholarly Paper ID 3613950. Rochester, NY: Social Science Research Network, May 29, 2020.

Gelman, Andrew. 2014. *Bayesian Data Analysis.* Third edition. Chapman & Hall/CRC Texts in Statistical Science. Boca Raton: CRC Press.

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

Gowrisankaran, G., and R. J. Town. 1999. "Estimating the Quality of Care in Hospitals Using Instrumental Variables." *Journal of Health Economics* 18, no. 6 (December): 747–767.

Grabowski, David C., Zhanlian Feng, Richard Hirth, Momotazur Rahman, and Vincent Mor. 2013. "Effect of Nursing Home Ownership on the Quality of Post-Acute Care: An Instrumental Variables Approach." *Journal of health economics* 32, no. 1 (January): 12–21.

Guarino, Cassandra M., Michelle Maxfield, Mark D. Reckase, Paul N. Thompson, and Jeffrey M. Wooldridge. 2015. "An Evaluation of Empirical Bayes's Estimation of Value-Added Teacher Performance Measures." *Journal of Educational and Behavioral Statistics* 40, no. 2 (April 1, 2015): 190–222.

Gupta, Atul, Sabrina T. Howell, Constantine Yannelis, and Abhinav Gupta. 2021. *Does Private Equity Investment in Healthcare Benefit Patients? Evidence from Nursing Homes.* w28474. National Bureau of Economic Research, February 22, 2021.

Hackmann, Martin B, R Vincent Pohl, and Nicolas R Ziebarth. 2021. "Patient Versus Provider Incentives in Long-Term Care," 87.

Heckman, James J. 1979. "Sample Selection Bias as a Specification Error." *Econometrica* 47 (1): 153–161.

Hoffman, Matthew D., and Andrew Gelman. 2014. "The No-U-turn Sampler: Adaptively Setting Path Lengths in Hamiltonian Monte Carlo." *The Journal of Machine Learning Research* 15, no. 1 (January 1, 2014): 1593–1623.

Holmstrom, Bengt, and Paul Milgrom. 1991. "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design." *Journal of Law, Economics, & Organization* 7:24–52.

Hull, Peter. 2020. *Estimating Hospital Quality with Quasi-Experimental Data.* Working Paper. Unpublished.

Jordan, Michael I., Zoubin Ghahramani, Tommi S. Jaakkola, and Lawrence K. Saul. 1998. "An Introduction to Variational Methods for Graphical Models." In *Learning in Graphical Models,* edited by Michael I. Jordan, 105–161. Dordrecht: Springer Netherlands.

Kass, Robert E., and Duane Steffey. 1989. "Approximate Bayesian Inference in Conditionally Independent Hierarchical Models (Parametric Empirical Bayes Models)." *Journal of the American Statistical Association* 84, no. 407 (September): 717–726.

Kingma, Diederik P., and Jimmy Ba. 2017. "Adam: A Method for Stochastic Optimization." January 29, 2017.

Kingma, Diederik P., and Max Welling. 2014. "Auto-Encoding Variational Bayes." May 1, 2014.

Konetzka, R. Tamara, Daniel J. Brauner, Joseph Shega, and Rachel M. Werner. 2014. "The Effects of Public Reporting on Physical Restraints and Antipsychotic Use in Nursing Home Residents with Severe Cognitive Impairment." *Journal of the American Geriatrics Society* 62, no. 3 (March): 454–461.

Konetzka, R. Tamara, Elizabeth M. White, Alexander Pralea, David C. Grabowski, and Vincent Mor. 2021. "A Systematic Review of Long-Term Care Facility Characteristics Associated with COVID-19 Outcomes." *Journal of the American Geriatrics Society* 69 (10): 2766–2777.

Konetzka, Tamara, Kevin Yan, and Rachel M. Werner. 2021. "Two Decades of Nursing Home Compare: What Have We Learned?" *Medical Care Research and Review* 78, no. 4 (August 1, 2021): 295–310.

McGarry, Brian E., Michael L. Barnett, David C. Grabowski, and Ashvin D. Gandhi. 2022. "Nursing Home Staff Vaccination and Covid-19 Outcomes." *New England Journal of Medicine* 386, no. 4 (January 27, 2022): 397–398.

Medina, Marco Avella, José Luis Montiel Olea, Cynthia Rush, and Amilcar Velez. 2021. "On the Robustness to Misspecification of Alpha-Posteriors and Their Variational Approximations." April 16, 2021.

Mele, Angelo, and Lingjiong Zhu. 2021. "Approximate Variational Estimation for a Model of Network Formation." *The Review of Economics and Statistics* (Forthcoming 2021).

Mor, Vincent, Jane Banaszak-Holl, and Jacqueline S. Zinn. 1995. "The Trend Towward Specialization in Nursing Care Facilities." *Generations: Journal of the American Society on Aging* 19 (4): 24–29.

Office of Inspector General. 2014. *Adverse Events in Skilled Nursing Facilities: National Incidence Among Medicare Beneficiaries.* Department of Health and Human Services, February.

Phan, Du, Neeraj Pradhan, and Martin Jankowiak. 2019. "Composable Effects for Flexible and Accelerated Probabilistic Programming in NumPyro." December 24, 2019.

Robbins, Herbert, and Sutton Monro. 1951. "A Stochastic Approximation Method." *The Annals of Mathematical Statistics* 22, no. 3 (September): 400–407.

Ruiz, Francisco J. R., Susan Athey, and David M. Blei. 2020. "SHOPPER: A Probabilistic Model of Consumer Choice with Substitutes and Complements." *The Annals of Applied Statistics* 14, no. 1 (March 1, 2020): 1–27.

Sharma, Hari, R. Tamara Konetzka, and Fabrice Smieliauskas. 2019. "The Relationship Between Reported Staffing and Expenditures in Nursing Homes." *Medical Care Research and Review* 76, no. 6 (December 1, 2019): 758–783.

Silver-Greenberg, Jessica, and Robert Gebeloff. 2021. "How U.S. Ratings of Nursing Homes Mislead the Public." The New York Times, March 13, 2021. Accessed 1/11/2022. https://www.nytimes.com/2021/03/13/business/nursing-homes-ratings-medicare-covid.html.

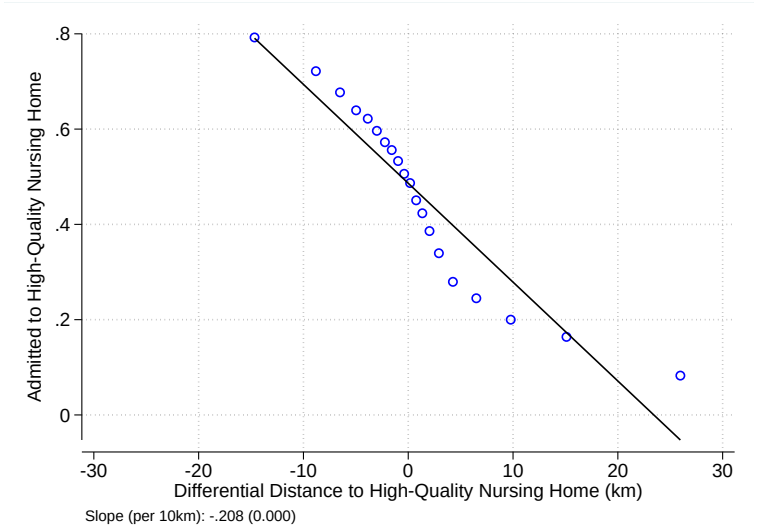
Thomas, Katie. 2014. "Medicare Star Ratings Allow Nursing Homes to Game the System." The New York Times, August 24, 2014. Accessed 11/22/2019. https://www.nytimes.com/2014/08/25/business/medicare-star-ratings-allow-nursing-homes-to-game-the-system.html.

Vafa, Keyon, Suresh Naidu, and David M. Blei. 2020. "Text-Based Ideal Points." July 21, 2020.

Wang, Yixin, and David M. Blei. 2019. "Frequentist Consistency of Variational Bayes." *Journal of the American Statistical Association* 114, no. 527 (July 3, 2019): 1147–1161.

9 Tables and Figures

(a) Probability of entering high-quality nursing home



(b) Predicted 90-day survival

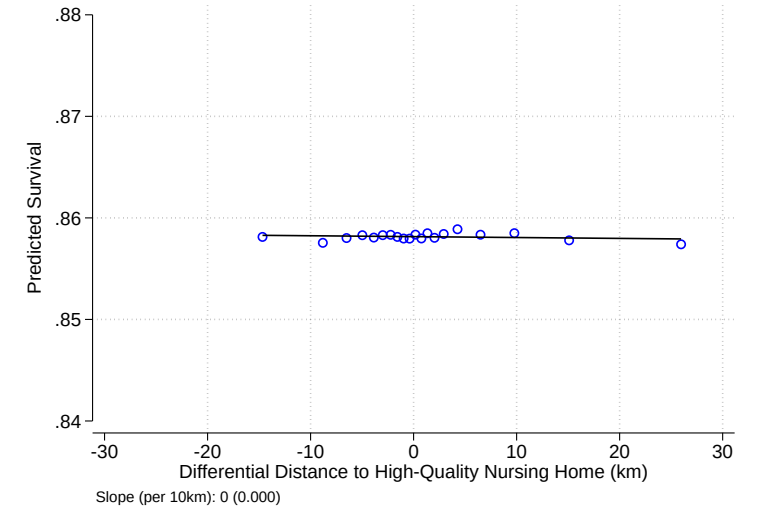
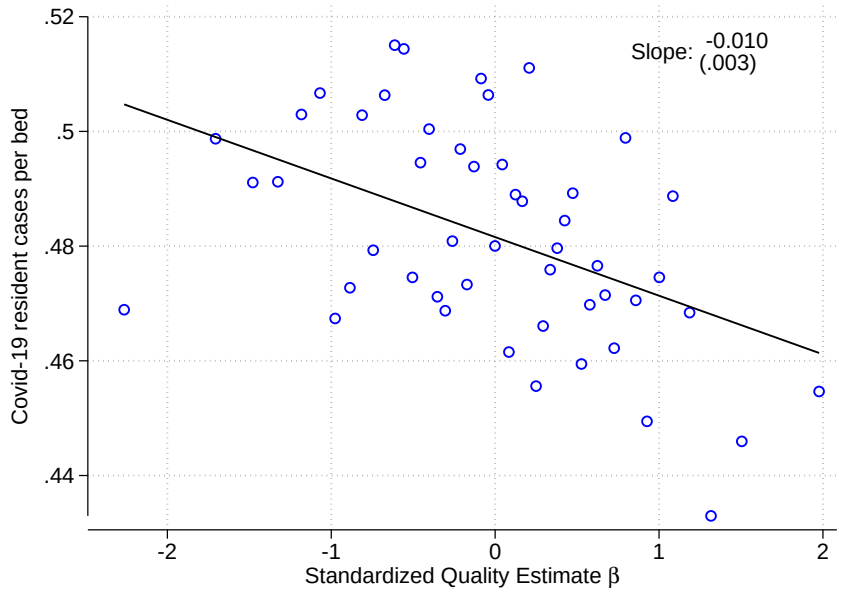


Figure 1: Patient Characteristics by Differential Distance to High Quality Nursing Home

Notes: Figures explore the exogeneity and relevance assumptions of the distance instruments. 'High-quality' nursing homes are defined as those whose $\hat{\beta}_j$ estimate is above the median for each state-year bin. Top panel presents a binned scatterplot of predicted 90-day survival by differential distance to the nearest high quality nursing home and nearest low-quality nursing home. The flat relationship indicates that patients who are relatively closer to high-quality nursing homes appear similar on ex ante mortality risk, based on observables. Bottom panel presents the probability of entering a nursing home with above-median quality by differential distance. Both are adjusted for hospital referral region fixed effects.

(a) Cumulative Covid-19 resident cases per bed



(b) Cumulative Covid-19 resident deaths per bed

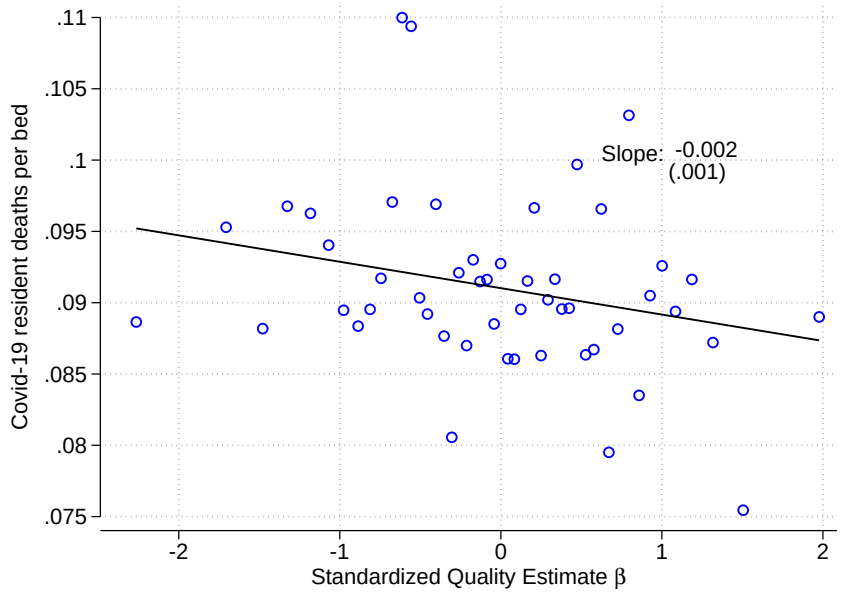


Figure 2: Covid-19 Outcomes by Nursing Home Quality Estimates β

Notes: Figures establish the negative relationship between nursing home quality in 2013-2016 and Covid-19 outcomes through 2021. Top panel presents a binned scatterplot of cumulative confirmed Covid-19 cases per bed through December 19, 2021 by nursing home quality $\hat{\beta}_j$ estimated in the last available year-bin, 2013-2016. Bottom panel presents Covid-19 deaths per bed by nursing home quality. Both are adjusted for hospital referral region fixed effects.

	(a) Selected-Adjusted Quality Estimates $\hat{\beta}_{jt}$				(b) Non-Adjusted Quality Estimates $\hat{\beta}_{jt}^{NA}$			
	(1) Overall	(2) Survey	(3) Quality	(4) Staffing	(5) Overall	(6) Survey	(7) Quality	(8) Staffing
2-Star	-0.0263	-0.0856***	0.0824**	-0.0205	0.0910***	0.0963***	0.0762**	0.0398
	(0.0195)	(0.0164)	(0.0255)	(0.0216)	(0.0188)	(0.0167)	(0.0294)	(0.0212)
3-Star	-0.0921***	-0.136***	0.128***	-0.0295	0.193***	0.189***	0.149***	0.0770***
	(0.0199)	(0.0174)	(0.0251)	(0.0218)	(0.0197)	(0.0176)	(0.0289)	(0.0212)
4-Star	-0.0738***	-0.109***	0.208***	0.000206	0.315***	0.323***	0.238***	0.178***
	(0.0204)	(0.0186)	(0.0253)	(0.0225)	(0.0207)	(0.0194)	(0.0288)	(0.0227)
5-Star	0.0609**	-0.0249	0.288***	0.134***	0.563***	0.451***	0.433***	0.320***
	(0.0235)	(0.0257)	(0.0278)	(0.0310)	(0.0250)	(0.0293)	(0.0314)	(0.0376)
Observations	28738	28738	28738	28668	28738	28738	28738	28668
Partial R^2	.0035	.0032	.0071	.0021	.0292	.0184	.0143	.0067
Partial ρ	.005	-.026	.084	.024	.166	.135	.114	.077

Table 1: Regressions of Quality Estimates on Nursing Home Compare 5-Star Ratings

Notes: Table establishes the weak correlation between nursing home quality and public report cards. Panel (a) reports regressions of standardized SNF quality estimates $\hat{\beta}_{jt}$ on each component of the Nursing Home Compare 5-Star ratings with HRR by year-bin fixed effects. Panel (b) reports the same regressions using the non-selection-adjusted quality estimates (i.e., those recovered when imposing $\alpha = 0$, analogous to a conventional risk-adjusted mortality approach). The mean of $\hat{\beta}_{jt}$ is zero. Partial R^2 reports the fraction of the variance in $\hat{\beta}_{jt}$ explained by the star ratings, conditional on the fixed effects. Partial ρ is the partial correlation between the star ratings (as continuous variables) and quality $\hat{\beta}_{jt}$, conditional on the fixed effects. We include estimates only for the 2009-2012 and 2013-2016 year-bins, due to availability of the star ratings. Standard errors in parentheses are all clustered at the facility-level. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

		(1)	(2)	(3)	(4)
(a) Resident Cases Per Bed	SNF Quality $\hat{\beta}_{jt}$	-0.0102***	-0.0106***	-0.0100***	-0.0118**
		(0.00274)	(0.00275)	(0.00277)	(0.00395)
	2019 NHC Rating		-0.00667***	0.00174	0.00367
			(0.00178)	(0.00188)	(0.00210)
	Mean	0.482	0.482	0.482	0.480
	Observations	13160	13087	13074	12245
(b) Resident Deaths Per Bed	SNF Quality $\hat{\beta}_{jt}$	-0.00185*	-0.00184*	-0.00236**	-0.00308*
		(0.000860)	(0.000863)	(0.000886)	(0.00122)
	2019 NHC Rating		0.000132	0.000851	0.00116
			(0.000565)	(0.000606)	(0.000695)
	Mean	0.0911	0.0912	0.0912	0.0910
	Observations	13160	13087	13074	12245
(c) Staff Cases Per Bed	SNF Quality $\hat{\beta}_{jt}$	-0.00885***	-0.00820***	-0.0111***	-0.0138***
		(0.00226)	(0.00220)	(0.00217)	(0.00300)
	2019 NHC Rating		0.0343***	0.0209***	0.0223***
			(0.00148)	(0.00143)	(0.00159)
	Mean	0.467	0.467	0.467	0.465
	Observations	13160	13087	13074	12245
(d) Vaccinated Residents, %	SNF Quality $\hat{\beta}_{jt}$	0.264*	0.311**	0.0621	0.445**
		(0.102)	(0.101)	(0.100)	(0.146)
	2019 NHC Rating		1.509***	1.219***	1.296***
			(0.0723)	(0.0746)	(0.0851)
	Mean	87.25	87.26	87.26	87.22
	Observations	13158	13085	13072	12247
(e) Vaccinated Staff, %	SNF Quality $\hat{\beta}_{jt}$	1.395***	1.447***	1.282***	-0.137
		(0.141)	(0.139)	(0.140)	(0.186)
	2019 NHC Rating		1.513***	1.289***	1.382***
			(0.0919)	(0.0955)	(0.103)
	Mean	78.23	78.25	78.25	78.56
	Observations	13163	13090	13077	12251
	Additional Controls	No	No	Yes	Yes
	HRR FEs	Yes	Yes	Yes	No
	County FEs	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, **$p < 0.01$, ***$p < 0.001$

Table 2: Regressions of Covid-19 Outcomes on Nursing Home Quality Estimates

Notes: Table reports regressions of several Covid-19 outcomes on several facility characteristics. SNF quality estimates $\hat{\beta}_j$ are from the last available year-bin, 2013-2016. The Nursing Home Compare Overall Rating is as it appeared on the website in December 2019. Additional controls include the log of mean household income for the SNF's zip code, log number of total beds, and indicators for the for-profit status, chain membership, the presence of an Alzheimer's unit, and whether the SNF is hospital-based. Standard errors in parentheses are all clustered at the facility-level.

A Details on Variational Inference

A.1 Overview of Variational Inference

Here we provide a more detailed overview of variational inference. Readers interested in a more in-depth treatment should refer to Blei, Kucukelbir, and McAuliffe (2017).

The goal of Bayesian computation is to estimate $p(\theta|x)$, the posterior density of the parameters of the model, θ, given the data, x. By Bayes' rule this is proportional to the joint density of data and parameters,

$$p(\theta|x) = \frac{p(\theta, x)}{p(x)} \tag{10}$$

For a given prior, $p(\theta)$ the numerator in (10) is easily computed since $p(\theta, x) = p(x|\theta)p(\theta)$, where $p(\theta)$ is the prior density. However, the denominator is generally computationally intractable.

Variational inference starts by postulating a parametrized family of distributions:

$$Q_\Psi = \{q(\theta|\psi) \quad \forall \psi \in \Psi\}$$

It then approximates the posterior density in (10) by finding a member of Q_Ψ that minimizes the distance between $p(\theta|x)$ and $q(\theta|\psi)$. Given a family of distributions the goal is then to find $\psi^\star$ that minimizes the distance, measured by the Kullback-Leibler divergence, between the posterior and the approximating distribution:

$$\psi^* = \arg\min_{\psi\in\Psi} KL(q(\theta|\psi)||p(\theta|x)), \tag{11}$$

where KL divergence is defined as

$$\begin{aligned} \mathrm{KL}(q(\theta|\psi) \;||\; p(\theta|x)) &= \int q(\theta|\psi) \log \frac{q(\theta|\psi)}{p(\theta|x)} d\theta \\ &= \mathbb{E}_q[\log q(\theta|\psi) - \log p(\theta|x)]. \end{aligned} \tag{12}$$

The quantity in (12) is not computable as it still involves the constant, $p(x)$. However, for the purpose of optimization, the constant can be ignored as its value does not depend on ψ. The resulting objective function, known as the Evidence Lower Bound (ELBO), is obtained by subtracting the log evidence, $p(x)$ from the KL divergence and switching the sign.

$$\begin{aligned} \mathrm{ELBO}(\psi) &:= \log p(x) - \mathrm{KL}(q(\theta|\psi) \;||\; p(\theta|x)) \\ &= \log p(x) + \mathbb{E}_q[\log p(\theta|x) - \log q(\theta|\psi)] \\ &= \log p(x) + \mathbb{E}_q[\log p(\theta, x) - \log p(x) - \log q(\theta|\psi)] \\ &= \mathbb{E}_q[\log p(\theta, x)] - \mathbb{E}_q[\log q(\theta|\psi)] \end{aligned} \tag{13}$$

Notice that maximizing the ELBO is equivalent to minimizing the KL divergence. That is,

$$\psi^* = \arg\min_{\psi\in\Psi} KL(q(\theta|\psi)||p(\theta|x)) \iff \psi^* = \arg\max_{\psi\in\Psi} \mathrm{ELBO}(\psi). \tag{14}$$

A typical tradeoff in Bayesian computation is that (a) the parameters with a high posterior density should be good at fitting the data, and (b) the posterior distribution of parameters should be close to the prior. This tradeoff can be easily seen by rewriting (13) as

$$\text{ELBO}(\psi) = \mathbb{E}_q[\log p(\theta, x)] - \mathbb{E}_q[\log q(\theta|\psi)] = \mathbb{E}_q[\log p(x|\theta)] - \text{KL}(q(\theta|\psi) \mid p(\theta)).$$

The first term in the equation above ensures that the parameters fit the data, while the second ensures that the approximate posterior $q(\theta|\psi)$ is as close as possible to the prior $p(\theta)$.

A.2 Use of Variational Inference in Economics

Variational inference and related approaches have long been popular in computer science (see Jordan, Ghahramani, Jaakkola, and Saul 1998). It has been used, for instance, to estimate the Latent Dirichlet Allocation model (Blei 2003) which is now popular among economists who use text data. Despite this popularity, there are few instances of the use of VI for estimating economic models. Notable exceptions include Bonhomme (2021) (team production) and Mele and Zhu (2021) (network formation). There are also several examples at the intersection of economics and computer science including Ruiz, Athey, and Blei (2020) (demand) and Vafa, Naidu, and Blei (2020) (polarization in text). One reason for this limited adoption may be that, until recently, there have been few theoretical results on the statistical properties of the approach. However, recent work (Wang and Blei 2019; Medina, Olea, Rush, and Velez 2021) established appealing properties of the approach regarding asymptotic behavior and robustness to model mis-specification. It is worth noting that VI has a great potential in terms of scalability (though massive parallelization and data-subsampling) and can benefit from modern specialized hardware (GPUs/TPUs). We anticipate that VI may see widespread adoption by economists in the near future.

A.3 Implementation

The parameters whose posterior distributions we estimate are: (a) the parameters of the choice model $((\xi_j)_{j\in\mathcal{J}},\ \pi)$, (b) the parameters of the survival model $((\beta_j)_{j\in\mathcal{J}},\ \gamma)$, (c) the selection parameters $(\alpha_j)_{j\in\mathcal{J}}$ and (d) the preference shocks for the selected SNFs $((\eta_{i,j(i)})_{i\in\mathcal{I}})$. Each of these parameters receives a variational family that we describe in the subsequent section. All the parameters of the variational families are collected into a single vector ψ. The joint log-probability of the model given the parameters is given by

$$\log p(\theta, x) = \sum_i (\log p(x_i|\theta)) + \log p(\theta) = \sum_i \left(\log P(Y_{i,j(i)}|\theta) + \log P(D_{i,j(i)}|\theta)\right) + \log p(\theta), \tag{15}$$

where $p(\theta)$ is the prior density of the parameters. We use independent priors for each parameter so $\log p(\theta) = \sum_i \log p(\eta_{i,j(i)}) + \sum_j \log p(\beta_j) + \sum_j \log p(\xi_j) + \sum_j \log p(\alpha_j) + \log p(\gamma) + \log p(\pi)$. Likewise, we use a factorizable variational family so the approximate posterior

density of the parameters is given by

$$\begin{aligned}\log q(\theta|\psi) = &\sum_i \log q(\eta_{i,j(i)}|\psi_{\eta_{i,j(i)}}) + \sum_j \log q(\beta_j|\psi_{\beta_j}) + \sum_j \log q(\xi_j|\psi_{\xi_j}) \\ &+ \sum_j \log q(\alpha_j|\psi_{\alpha_j}) + \log q(\gamma|\psi_\gamma) + \log q(\pi|\psi_\pi)\end{aligned} \tag{16}$$

We specify the ELBO using Numpyro (Phan, Pradhan, and Jankowiak 2019; Bingham et al. 2018), a Python library for probabilistic programming. Full implementation details are beyond the scope of this paper, though we emphasize a few key points. First, as Equations (15) and (16) show, the ELBO is an expectation over a sum of components related to individual observations which makes it easily parallelizable. Second, we are able to leverage automatic differentiation to compute an unbiased estimate of the gradient of the ELBO. This is achieved using the so-called 'reparametrization trick' of Kingma and Welling (2014). This trick enables expressing the gradient of the ELBO as an expectation of the gradient with respect to a noise distribution, τ where the density does not depend on the parameters ψ.

$$\nabla_\psi ELBO(\psi) = \nabla_\psi \mathbb{E}_q(\log p(\theta, x) - \log q(\theta|\psi)) = \mathbb{E}_\tau \left(\nabla_\psi(\log p(f(\psi,\tau)) - \log q(f(\psi,\tau)|\psi))\right), \tag{17}$$

where f is the reparametrizing function.[17]

Consequently, we can obtain an unbiased estimate of $\nabla_\psi \text{ELBO}(\psi)$ by taking draws from the noise distribution. This is critical given the dimension of the parameter space (which in our model scales linearly with the sample size), so gradient-free optimization methods are infeasible. On the other hand, the analytic form of $\nabla_\psi \text{ELBO}(\psi)$ may not exist. Fortunately, this unbiased estimate of the gradient is sufficient to specify an algorithm that is guaranteed to converge to a local optimum (Robbins and Monro 1951).[18]

A.4 Variational Families

Several factors contribute to our choice of the variational family. First, due to the number of parameters we estimate (including the value of the preference for the selected SNF $\eta_{i,j(i)}$), and the fact that in the subsequent analysis we focus on posterior means of the parameters (and not the variance), we selected factorizable families. Second, for each parameter we chose a family that includes the prior. For example, for each parameter whose prior is Gaussian, we specified a Gaussian variational approximation. For α_j, whose prior is Uniform, we specify a Beta variational approximation. Finally, for the posterior scale of the hierarchical prior we specify a Gamma variational approximation. Specifying the variational family in such a way is not necessary, but it facilitates analytically computing the KL divergence between

17. As an example, if $\theta \sim N(\mu, \sigma)$ then $\theta = \mu + \sigma\tau = f((\mu,\sigma),\tau)$ where τ is the standard normal noise distribution. The reparametrization trick is not available for some distributions, for instance categorical distributions.

18. This approach differs significantly from Bonhomme (2021) who instead uses a (gradient-free) variational expectation maximization algorithm to maximize an objective function closely related to the ELBO. By exploiting a gradient-based optimizer, our approach performs well in significantly larger parameter spaces.

the prior and the approximate posterior, which reduces the variance in the estimation of the gradient of the ELBO and may speed up convergence.

A.5 Optimization Details

We minimize the ELBO with the Adam optimizer (Kingma and Ba 2017) using clipped gradients (to ensure numerical stability) and geometrically decreasing step size. For a typical state-year bin, a single Adam update step takes less than a second on a single core CPU[19] and our model converged after approximately 30 minutes to one hour, depending on the state-year bin. The optimization of the VI objective function can further benefit from specialized hardware (such as a GPU), but we did not implement this due to hardware limitations on the secure server where our patient data reside.

A.6 Simulation Exercises

To evaluate the performance of variational inference in our setting, we conduct a brief simulation exercise. We generate a dataset containing $I = 25,000$ patients and $J = 50$ facilities according to the data-generating process specified in Section 3. For each facility $j \in \{1, \ldots, 50\}$, we sample a quality parameter $\beta_j \sim N(0, 0.8)$ and a selection parameter $\alpha_j \sim \text{Uniform}(-1, 1)$.

In addition, for each patient i we simulate five highly correlated risk-adjusters drawn from a multivariate normal distribution with mean 0 and variance-covariance matrix with values 0.09 and 0.08 on and off diagonal, respectively. We specify the coefficients on the risk-adjusters, γ, to be a vector of ones. The instruments Z_{ij} are drawn from a Triangular(0,30) distribution with a mode at the right end.

In addition to estimating the model with VI, we also use a state-of-the-art Markov Chain Monte Carlo approach, the No-U-Turn-Sampler (Hoffman and Gelman 2014) implemented in Numpyro. We run the chain for 2000 steps, split equally between warmup and sampling and ensure that the potential scale reduction factor is close to 1 for all parameters, indicating convergence. Given the theoretical guarantees of the NUTS algorithm, we use the posterior mean and variance of the parameters as the benchmark against which we compare the VI estimates.

Appendix Figure D.5 presents the results obtained with VI. We plot the estimated mean posterior values of β_j and α_j against their true values. In that case the R-squared values from regressions of the true β_j and true α_j on their estimated equivalents are 0.856 and 0.597, respectively. Comparing this results against asymptotically exact MCMC-based equivalents obtained using the No-U-Turn-Sampler we find that the estimates are virtually identical between the two algorithms for both β_j ($R^2 = 0.9869$) and α_j ($R^2 = 0.9865$). This should reassure readers that in our setting, despite being an approximate method, VI performs extremely well and matches MCMC approach in precision.

We do not report runtimes, but in all of our experiments we observed at least a 10-fold improvement using VI compared with NUTS, and often much larger. In our simulations

19. The biggest state-year bin, California 2013-2016, contains 500k+ admissions. A single Adam update there takes approximately 2.5 seconds and objective function converges in less than 2 hours.

both algorithms (VI and NUTS) benefited from GPU acceleration (using a single consumer-grade GPU, i.e. Nvidia Geforce GTX 1660 Ti) resulting in approximately 12-fold speedup compared with the deployment on a 16-core CPU (Intel i7-1077H). Unfortunately, hardware accelerators are not available at the secure server where our data reside. Further, despite using the GPU, to keep runtime under two hours, we had to restrict the size of our simulated data set to less than 10% of the size of the biggest market in the real data. We therefore conclude that in our application even the state-of-the-art MCMC approach is prohibitively slow, while VI converges in a reasonable time.

These simulation results are in line with the literature examining theoretical properties of variational inference – namely, that point estimates (such as means of the approximate posterior distributions) are likely to be consistent estimators (in the frequentist sense) for the parameters of interest. However, echoing Wang and Blei (2019) we warn that using too simple a family of distributions to approximate the posterior may be undesirable. For the purpose of hypotheses testing, the approximations may lead to Bayesian credible intervals being incorrect. This may be a problem when the true posterior is poorly approximated by a factorizable family, for example if it is multimodal, or the parameters are strongly correlated.

We illustrate this problem in Appendix Figure D.7, which presents the point estimates and 95% credible intervals for the parameters on the risk-adjusters γ. Because the data were simulated a way that the risk-adjusters are highly correlated, the true posterior distribution for γ also features correlation. As a result, while the point estimates are very similar, the credible intervals are too small using VI, as compared with (asymptotically exact) MCMC. This points to a potential tradoff between computational burden and the quality of inference.

This problem should theoretically be resolved if one could use sufficiently rich families of distributions. For example, Wang and Blei (2019) show that the family of full-rank multivariate normal distributions have desirable asymptotic properties in that the posterior distribution converges (in total variation) to the asymptotic distribution of an unbiased and asymptotically normal frequentist estimator. Unfortunately, this approach is only computationally feasible when the parameter vector θ is sufficiently low-dimensional. In our scenario, the presence of high-dimensional nuisance parameters, $\eta_{i,j(i)}$ (the preference shocks for the chosen SNFs) make this intractable. Nevertheless, we experimented with an intermediate approach that mixes factorizable normal approximation for some parameters and full-rank multivariate normal for others. We found that allowing for full-rank multivariate normal distributions for the 5-element parameter γ is sufficient to recover credible intervals that are virtually indistinguishable from those obtained with MCMC, while the computational time is almost identical to using a fully factorizable family. This is a promising result, but we warn that this approach is not guaranteed to work in all settings. We note that this approach is similar to Bonhomme (2021) who also uses a mean-field variational approximation for the nuisance parameters, but uses a different maximum likelihood approach for the remaining (low-dimensional) parameters.

B Further Model Details

B.1 Deriving Likelihoods

A necessary element in computing the VI objective, the ELBO, is the log-joint likelihood of the data and parameters. As mentioned in the main text, we explicitly condition on and estimate the value of the preference shock for each person for their selected facility, $\eta_{ij(i)}$. Adding in the assumption that the shocks for different facilities are independent, we can then express the probability that i selects SNF j as the product of the probabilities that the utility of SNF j exceeds that of $\tilde{j}$, for each other facility $\tilde{j}$ in i's choice set. The remaining randomness then comes the utility shocks for those non-chosen facilities $\tilde{j}$, which we do not estimate. Consequentely, the log-probability of the observed choice can be derived as:

$$\begin{aligned}
\log P(D_{ij} = 1 \mid \theta_i, X_i, Z_i) &= \log P(u_{ij(i)} > u_{ij'}, \ \forall j' \neq j(i) \mid \theta_i) \\
&= \log P(\eta_{ij'} < \delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)}, \ \forall j' \neq j(i) \mid \theta_i) \\
&= \log \Pi_{j' \neq j(i)} P\left(\eta_{ij'} < \delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)} \mid \theta_i\right) \\
&= \Sigma_{j' \neq j(i)} \log \Phi(\delta_{ij(i)} - \delta_{ij'} + \eta_{ij(i)})
\end{aligned} \tag{18}$$

where Φ is the standard normal cumulative distribution function.

Similarly, by plugging in for our definition of the health shock ε_{ij}, we can express the log-probability of the observed survival indicator as:

$$\begin{aligned}
\log P(Y_{i,j(i)} = 1 \mid \theta_i, X_i, Z_i) &= \log P\big(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)} + \sqrt{1-\alpha_{j(i)}^2}\tilde{\varepsilon}_{i,j(i)} > 0 \mid \theta_i\big) \\
&= \log P\left(\tilde{\varepsilon}_{i,j(i)} > \frac{-(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)})}{\sqrt{1-\alpha_{j(i)}^2}} \mid \theta_i\right) \\
&= \log\left(1 - \Phi\left(\frac{-(\beta_{j(i)} + \gamma X_i^T + \alpha_{j(i)}\eta_{i,j(i)})}{\sqrt{1-\alpha_{j(i)}^2}}\right)\right)
\end{aligned} \tag{19}$$

B.2 Prior Distributions

Our variational inference estimation approach does not restrict us to conjugate priors – it is only required that the joint log-likelihood is differentiable. This restriction is mild. While it excludes models with explicit discrete latent variables, these models can often be parametrized to remove these latent variables, a process known as marginalization.

Starting with the quality parameters, β_j, we specify a hierarchical prior $\beta_j \sim N(\mu_\beta, \sigma_\beta)$. The mean and scale of this prior are given the following prior: $\mu_\beta \sim N(0, 5)$, $\sigma_\beta \sim \text{Gamma}(3, \frac{1}{3})$. This specification implicitly regularizes (or shrinks) the quality estimates towards the state-year bin mean, increasing reliability especially for small SNFs that have relatively few patients. Note that the empirical Bayes methods frequently used in quality estimation to reduce noise (e.g. Chetty, Friedman, and Rockoff 2014; Guarino et al. 2015; Chandra, Finkelstein, Sacarny, and Syverson 2016) can be viewed as an approximation to

the hierarchical model.

For the remaining parameters we use uninformative or weakly informative priors. Specifically, we use a uniform prior for the selection parameter, α_j. For the coefficients on the health (γ) and preference (π) shifters, we employ dispersed normal priors with mean zero and scale 8. For the observed SNF popularity parameters ξ_j, we use a standard normal prior. It is well-known that utility parameters are identified only up to a constant. We do not normalize any of the ξ_j, instead we rely on the prior to fix the mean of these parameters.

C Sample Construction

We begin with all MDS admissions assessments for Medicare-enrolled residents during the period 2000-2017. We further restrict these admission assessments to only those assessments that occur within 30 days of entry. In the event that an admission spans multiple assessments, we collect the non-missing covariates from the last assessment conducted, and we exclude any assessments with missing information on diagnoses or the activities of daily living.

We combine the cleaned sample of admissions assessments with the Long-Term Care Focus files; a small number of assessments are excluded at this stage if their provider numbers do not match any active SNFs from the LTCFocus panel. We then impose several sample restrictions. First, we require no prior nursing home admission assessments in the prior 365 days. Second, for this reason we exclude any assessments for admissions that begin prior to 2001. Third, we exclude assessments for whom we have no home zip code or for whom the home zip code does not align with the state code[20] provided in the beneficiary summary files. Fourth, we exclude any patients residing in Alaska, as the LTCFocus files do not contain any Alaskan SNFs. Finally, after we pool admissions across 4-year bins, we exclude all assessments at SNFs with fewer than 50 patients.

To construct the patient choice sets, we calculate the distance to each nursing home in the state, for all beneficiaries in the new admissions sample. To do so we assign each beneficiary a home zip code, coming from the Medicare beneficiary summary files.[21] Distance is calculated between the beneficiary's home zip code centroid[22] and the coordinates of the nursing home, which we geocode from the address available in the LTCFocus files.

A natural concern with using zip code centroids is that our distances may not accurately represent the true distance for each beneficiary to nursing home. This may reflect both that a beneficiary's residence may not be close to the centroid, as well the distinction between geodetic and driving distance. However, the extent to which we are misattributing distance will only bias the disutility of distance parameters in our selection equation towards zero. The two-stage least squares equivalent is that measurement error will lead our instruments to be weaker than they otherwise would be.

20. We collect the zip code ranges available in each state from the IRS: https://www.irs.gov/pub/irs-utl/zip_code_and_state_abbreviations.pdf

21. For long-term stays at nursing homes, it is common for a resident to update her address on file with Social Security to that of the nursing home. To account for this, for admissions in year t, we assign the zip code in year $t-1$ if the zip code matches that of the nursing home; otherwise, we use the zip code from year t.

22. Available at https://www.nber.org/research/data/zip-code-distance-database.

D Additional Tables and Figures

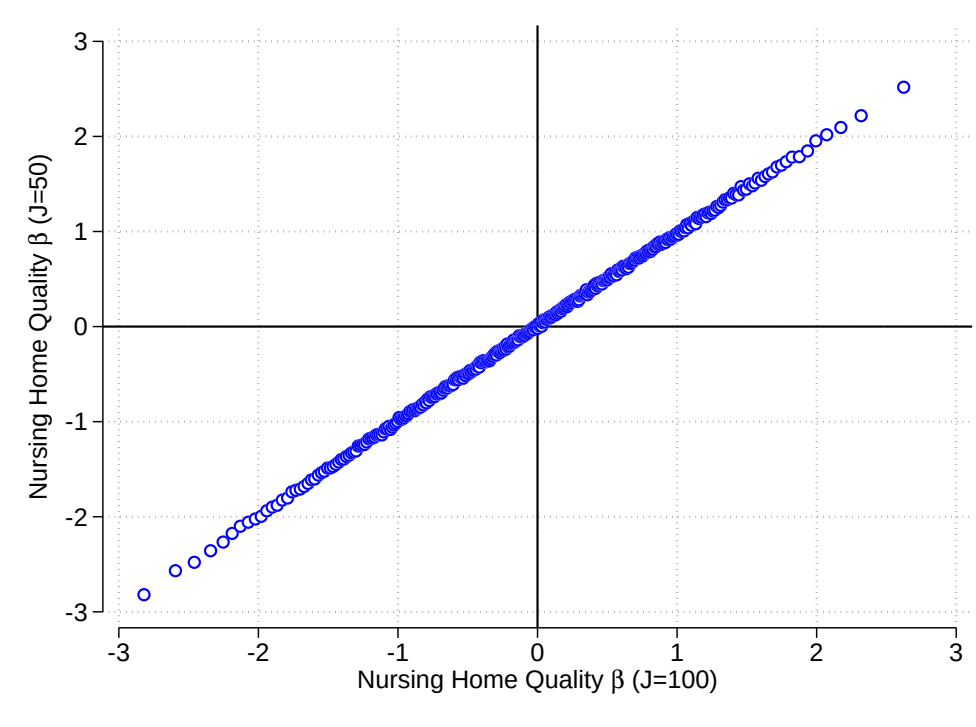
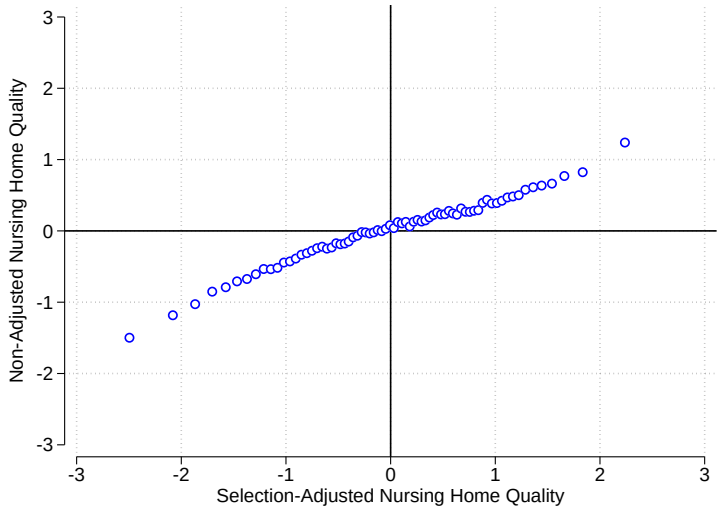


Figure D.1: Robustness to Alternative Choice Set Sizes

Notes: Figure presents a binscatter plot of the market-standardized nursing home quality parameter $\hat{\beta}_{jt}$ computed using choice set size $J = 50$ and $J = 100$. The correlation coefficient between the two sets of estimates is 0.984.

(a) Binscatter of Non-Selection-Adjusted Nursing Home Quality $\hat{\beta}_{jt}^{NA}$ by Selection-Adjusted Nursing Home Quality $\hat{\beta}_{jt}$



(b) Binscatter of Unobserved Selection $\hat{\alpha}_{jt}$ by Nursing Home Quality $\hat{\beta}_{jt}$

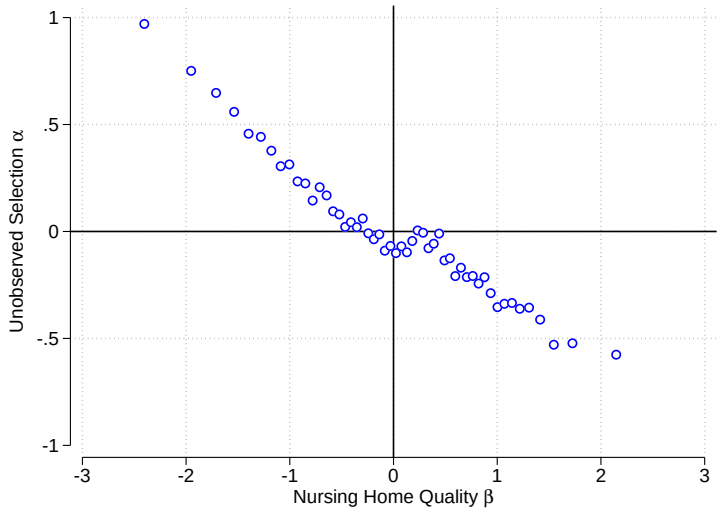
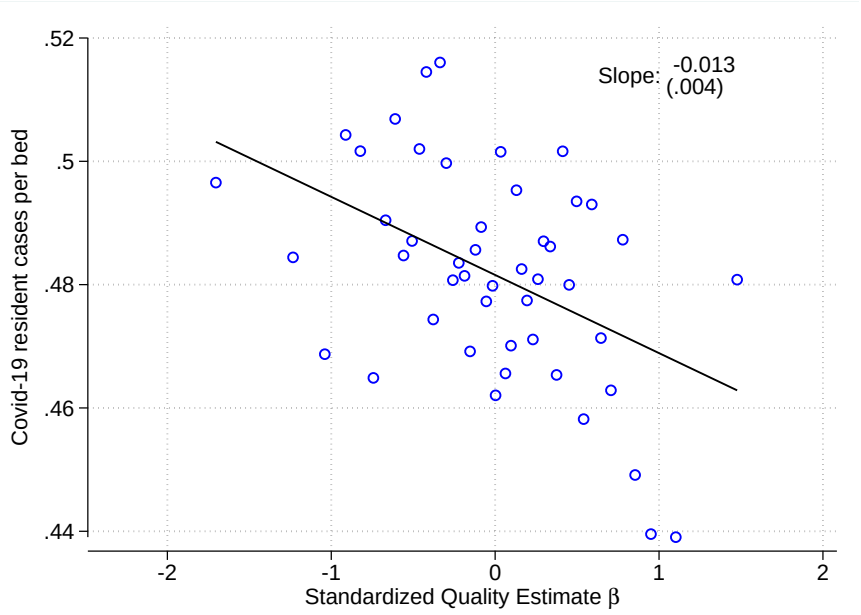
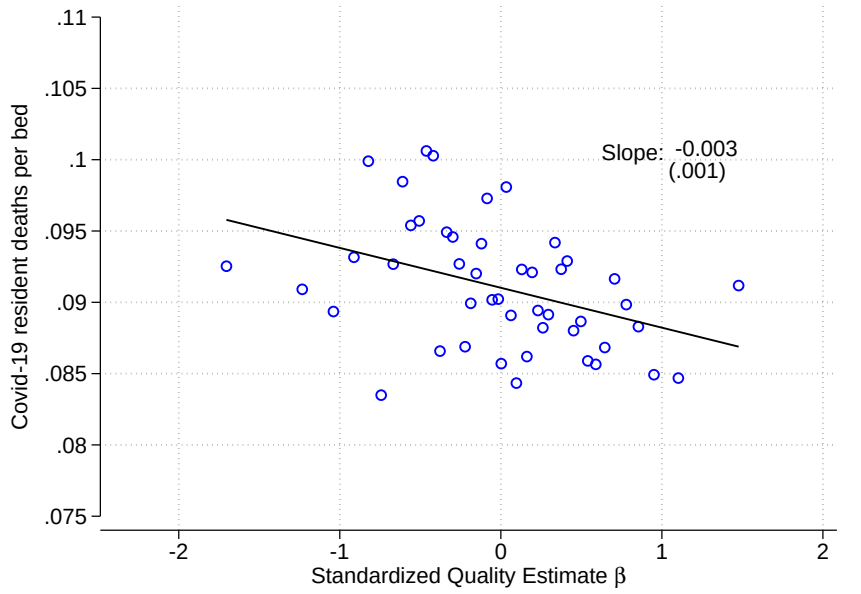


Figure D.2: Relationships Between $\hat{\beta}_{jt}$, $\hat{\beta}_{jt}^{NA}$, and $\hat{\alpha}_{jt}$

Notes: Figure documents the relationships between the selection-adjusted quality $\hat{\beta}_{jt}$, the non-selection-adjusted quality estimate $\hat{\beta}_{jt}^{NA}$, and the unobserved selection parameter $\hat{\alpha}_{jt}$. The top panel presents a binscatter showing the positive relationship between selection-adjusted quality and non-selection-adjusted quality. The bottom panel presents a binscatter plot of the selection parameter $\hat{\alpha}_{jt}$ by the quality parameter $\hat{\beta}_{jt}$. The negative relationship indicates the presence of adverse selection.



(a) Cumulative Covid-19 resident cases per bed

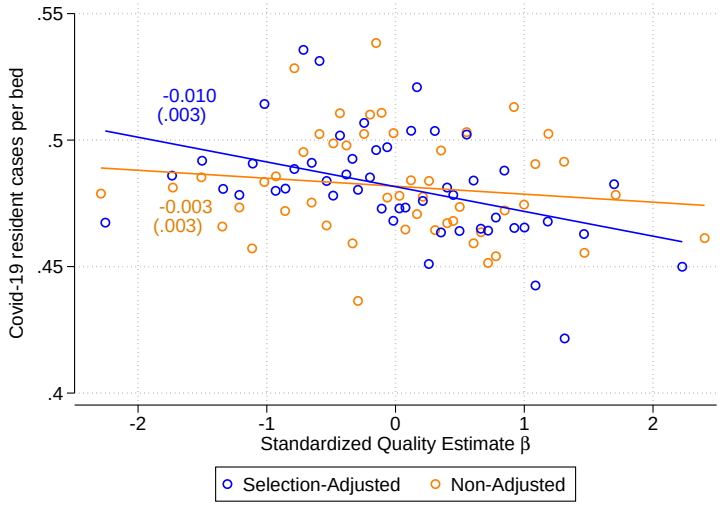


(b) Cumulative Covid-19 resident deaths per bed

Figure D.3: Binscatters of Covid-19 Outcomes by Nursing Home Quality Estimates β with County Fixed Effects

Notes: Top panel presents a binned scatterplot of cumulative confirmed Covid-19 cases per bed through December 19, 2021 by nursing home quality β estimated in the last available year-bin, 2013-2016. Bottom panel presents Covid-19 deaths per bed by nursing home quality. Both are adjusted for county fixed effects.

(a) Cumulative Covid-19 resident cases per bed



(b) Cumulative Covid-19 resident deaths per bed

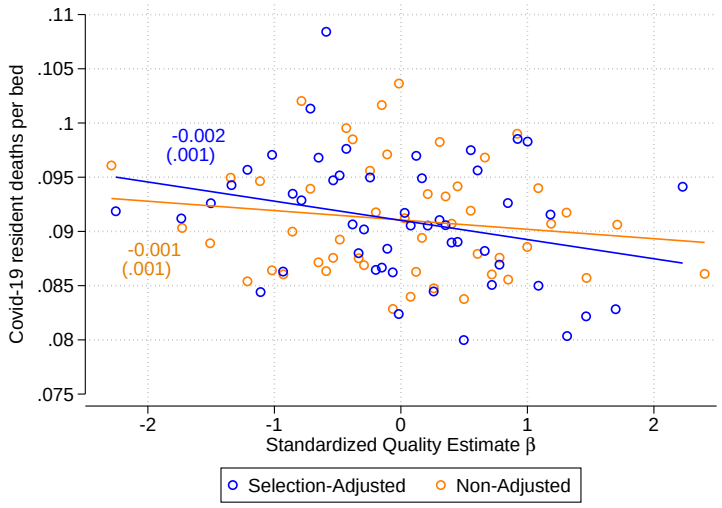
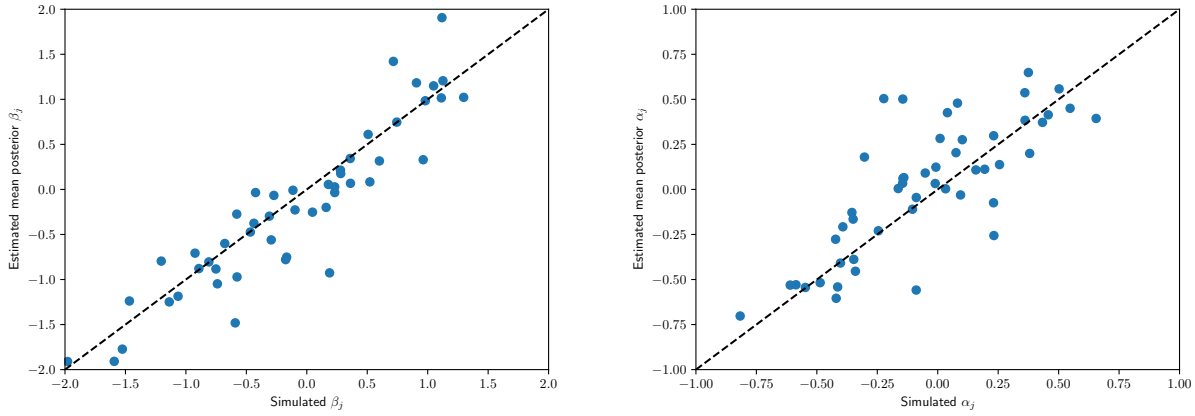


Figure D.4: Covid-19 Outcomes by Nursing Home Quality Estimates

Notes: Replication of Figure 2, with non-selection-adjusted quality estimated $\hat{\beta}_{jt}^{NA}$ overlaid. Top panel presents a binned scatterplot of cumulative confirmed Covid-19 cases per bed through December 19, 2021 by nursing home quality β estimated in the last available year-bin, 2013-2016. Bottom panel presents Covid-19 deaths per bed by nursing home quality. Both are adjusted for HRR fixed effects.

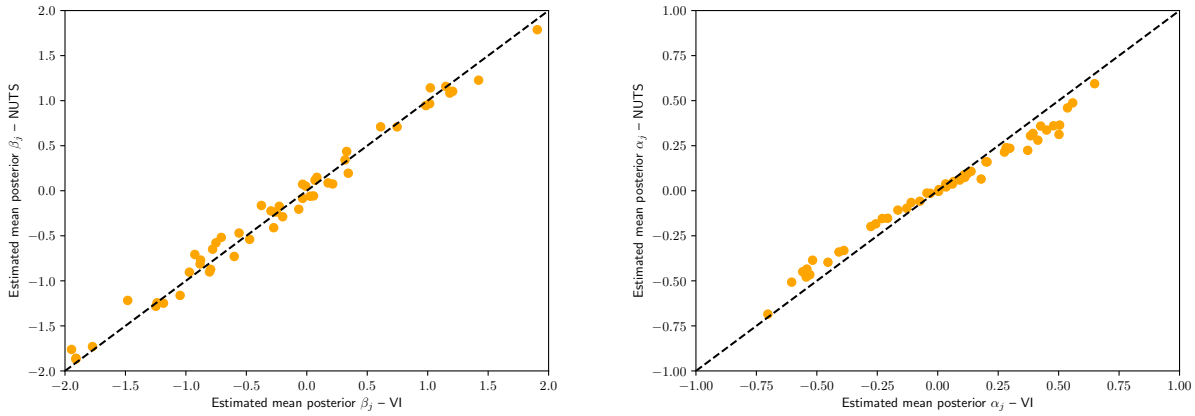
Figure D.5: Simulation Study: VI Estimates vs. True Values



(a) Quality β_j **(b)** Selection α_j

Notes: Figure presents scatter plots of the true parameters (on the horizontal axes) and the means of the respective posterior distributions estimated using variational inference. In all cases we use $I = 50,000$ observations, or approximately 100 observations per facility.

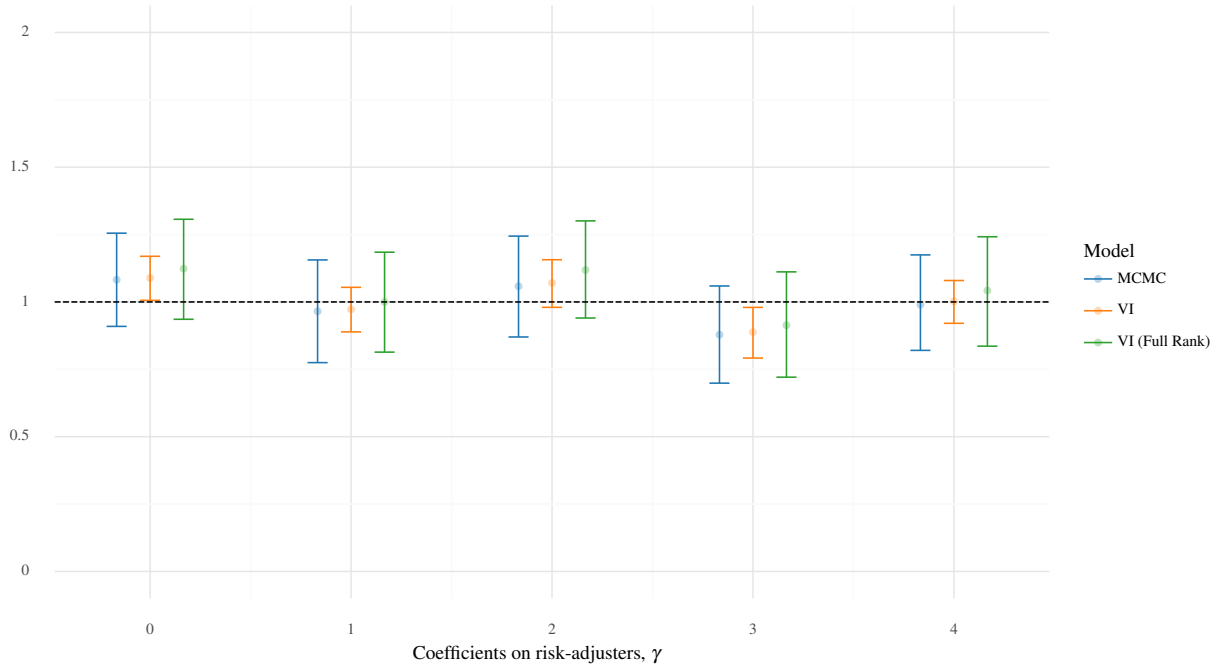
Figure D.6: Simulation Study: VI Estimates vs NUTS Estimates



(a) Quality β_j **(b)** Selection α_j

Notes: Figure presents scatter plots of the means of the posterior distributions estimated with variational inference (on the horizontal axes) against their equivalents obtained with an MCMC approach, NUTS. In all cases we use $I = 50,000$ observations, or approximately 100 observations per facility.

Figure D.7: Simulation Study: Comparison Between VI, MCMC, and full-rank VI



Notes: Figure presents the means and 95% credible interval of the posterior distributions of the coefficients on the 5 strongly correlated risk-adjusters. The true value for each is set to 1 and is indicated by the dashed line. We find that while MCMC and VI recover nearly identical point estimates, the credibility intervals for VI are under-estimated, relative to the MCMC estimates, which are theoretically known to be exact. Adopting a full-rank approximating distribution for VI successfully recovers the 'true' credibility intervals as estimated by MCMC.

	Mean	P10	P50	P90
Female	0.64	0	1	1
Age	79.33	66	81	91
White	0.86	0	1	1
Black	0.10	0	0	0
Number of ADLs	5.18	1	6	8
Number of Diagnoses	1.66	0	1	3
Disability	0.19	0	0	1
ESRD	0.01	0	0	0
BMI<18.5	0.08	0	0	0
BMI 18.5-24.9	0.39	0	0	1
BMI 25-29.9	0.26	0	0	1
BMI >30	0.26	0	0	1
Mortality, 30-day	0.05	0	0	0
Mortality, 90-day	0.14	0	0	1
Mortality, 180-day	0.21	0	0	1
Mortality, 365-day	0.30	0	0	1
Distance to NH (mi)	7.84	1	4	17
Observations	20538183			

Table D.1: Individual summary statistics

Notes: Table reports summary statistics of the estimating sample of new nursing home admissions. Demographics and comorbidities are derived from the admission assessment for each stay. Disability, ESRD status, and mortality are all derived from the Beneficiary Summary Files.

Patient Attribute	Nearest Facility's Quality Estimate		
	Above Median	Below Median	p-value
	(1)	(2)	(3)
Quality $\geq$ Median	0.638	0.249	< 0.001
Age at Entry	79.268	79.417	< 0.001
Disability	0.186	0.185	0.242
Female	0.639	0.634	< 0.001
Race: Black	0.121	0.076	< 0.001
Race: Other	0.043	0.036	< 0.001
Weight: Underweight	0.084	0.081	< 0.001
Weight: Overweight	0.261	0.265	< 0.001
Weight: Obese	0.260	0.266	< 0.001
ADL: Bath	0.928	0.930	< 0.001
ADL: Bed	0.612	0.599	< 0.001
ADL: Transferring	0.642	0.630	< 0.001
ADL: Walking	0.285	0.287	0.011
ADL: Locomotion	0.660	0.641	< 0.001
ADL: Dressing	0.680	0.667	< 0.001
ADL: Eating	0.148	0.143	< 0.001
ADL: Toileting	0.693	0.679	< 0.001
ADL: Hygiene	0.575	0.569	< 0.001
Diabetes	0.313	0.314	0.418
Alzheimer's	0.250	0.247	< 0.001
Stroke	0.135	0.133	< 0.001
Pneumonia	0.092	0.094	< 0.001
Hip Fracture	0.085	0.085	0.101
Predicted 90-Day Survival	0.858	0.858	0.285
90-Day Survival	0.863	0.853	< 0.001

Table D.2: Patient Attributes' Balance

Notes: Table reports the mean values of patients' attributes with respect to the level of nurse staffing in the facility nearest to patient's home zip code. Specifically, we group facilities based on whether the level of RN hours' per patient is above or below the state-year median. Column (1) presents the statistics for patients whose nearest facility (not necessarily the one they selected) has an above-median level of RN staffing. The first row reports the fraction of patients who selected an above-median facility. Patient characteristics include demographic variables, comorbidities (ESRD status and the presence of four common conditions), and whether the patient requires assistance with the activities of daily living. Finally, we report the predicted and realized 90-day survival.

	Quality Estimates			
	(a) Selected-Adjusted $\hat{\beta}_{jt}$		(b) Non-Adjusted $\hat{\beta}_{jt}^{NA}$	
For-Profit	-0.00662	(0.0141)	-0.0992***	(0.0137)
Chain	0.00684	(0.0119)	-0.0203	(0.0117)
Alzheimer's Unit	-0.00135	(0.0154)	0.129***	(0.0143)
Hospital-Based	-0.314***	(0.0243)	-0.283***	(0.0288)
log(Total Beds)	-0.0467***	(0.0114)	0.0672***	(0.0126)
% Occupancy	0.00204***	(0.000454)	0.00659***	(0.000470)
log(RNs/Resident)	0.0179*	(0.00785)	0.0591***	(0.00991)
log(LPNs/Resident)	0.0408**	(0.0131)	0.0334*	(0.0148)
log(CNAs/Resident)	0.202***	(0.0196)	0.244***	(0.0220)
% White	-0.0100***	(0.000373)	0.00312***	(0.000358)
% Medicare	-0.00399***	(0.000375)	0.00315***	(0.000528)
% Medicaid	0.000670*	(0.000279)	-0.00480***	(0.000330)
Mean ADL Score	0.0258***	(0.00274)	0.0703***	(0.00301)
Mean Age	-0.0192***	(0.00108)	0.0199***	(0.00116)

Table D.3: Univariate Regressions of Nursing Home Quality on Facility Characteristics

Notes: Table reports a series of univariate regressions of estimated nursing home quality on several facility characteristics, collected from the Long-Term Care Focus files. Column (a) reports regressions on our standardized selection-adjusted quality estimates $\hat{\beta}_{jt}$. Column (b) reports regressions on the non-selection-adjusted quality measure $\hat{\beta}_{jt}^{NA}$. For-profit, chain, Alzheimer's unit, and hospital-based are each indicator variables. % Occupancy, White, Medicare, and Medicaid are each scaled from 0 to 100. log(RNs/Resident), log(LPNs/Resident), and log(CNAs/Resident) measure the log number of hours worked per resident-days for registered nurses, licensed practical nurses, and certified nursing aides, respectively. Mean ADL Score is an index of patient severity (higher values indicate that patients require more help with activities of daily living). All regressions include HRR-by-year-bin fixed effects. Standard errors in parentheses are clustered at the facility-level. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

NHC Rating	Counterfactual Rating (%)				
	1-Star	2-Star	3-Star	4-Star	5-Star
	(1)	(2)	(3)	(4)	(5)
1-Star	79.9	19.6	0.5	0.0	0.0
2-Star	12.5	68.8	18.4	0.3	0.0
3-Star	0.4	17.0	66.8	15.5	0.3
4-Star	0.0	0.4	20.7	64.9	14.0
5-Star	0.0	0.0	0.7	26.5	72.8

Table D.4: Transition Matrix from Nursing Home Compare to $\hat{\beta}_{jt}$-Based Ratings

Notes: Table reports the results from a counterfactual exercise in which the 'Quality' component of the Nursing Home Compare Overall Rating is replaced by our estimates of $\hat{\beta}_{jt}$. For each state-year-bin, we compute the share of facilities with each Quality star rating (1 to 5) and create equal-sized bins of $\hat{\beta}_{jt}$ for that state-year-bin. We then construct new Overall Ratings by replacing the Quality component with these $\hat{\beta}_{jt}$ bins. Column (1) reports the overall share of facilities with each Overall Rating. Columns (2)-(6) report the shares of facilities with each new Overall Rating under the counterfactual Quality ratings.

	2001	2005	2009	2013
2001	1	0.804	0.747	0.714
2005	0.804	1	0.815	0.767
2009	0.747	0.815	1	0.82
2013	0.714	0.767	0.82	1

Table D.5: Autocorrelation of $\hat{\beta}_{jt}$

Notes: Table reports correlations of $\hat{\beta}_{jt}$ across 4-year bins. The estimates suggest that the quality is highly stable across time.

Variable	Mean	% Positive	% Negative	% Positive Significant	% Negative Significant
(a) Choice Model					
Distance	-1.66	0.00	1.00	0.00	1.00
Distance Sq	0.65	1.00	0.00	1.00	0.00
(b) Survival Model					
Female	0.11	1.00	0.00	1.00	0.00
Age at Entry	-0.08	0.04	0.96	0.02	0.78
Age at Entry Sq	-0.05	0.17	0.83	0.10	0.63
ADL: Bath	-0.01	0.36	0.64	0.15	0.29
ADL: Bed	-0.07	0.00	1.00	0.00	0.90
ADL: Dressing	-0.02	0.16	0.84	0.02	0.17
ADL: Eating	-0.15	0.00	1.00	0.00	1.00
ADL: Hygiene	-0.08	0.00	1.00	0.00	0.95
ADL: Walking	0.04	0.98	0.02	0.87	0.00
ADL: Locomotion	-0.05	0.03	0.97	0.00	0.70
ADL: Toileting	-0.03	0.05	0.95	0.00	0.31
ADL: Transfering	0.00	0.46	0.54	0.05	0.03
Disability	0.01	0.84	0.16	0.24	0.01
ESRD	-3.15	0.01	0.99	0.00	0.81
Fall: 1 month	0.01	0.81	0.19	0.44	0.04
Fall: Last 6 months	-0.01	0.32	0.68	0.07	0.30
Fracture: last 6 months	0.08	1.00	0.00	1.00	0.00
Race: Black	-0.82	0.80	0.20	0.72	0.18
Race: Other	-1.10	0.68	0.32	0.51	0.28
Weight: Underweight	-0.07	0.00	1.00	0.00	1.00
Weight: Overweight	0.06	1.00	0.00	0.99	0.00
Weight: Obese	0.08	1.00	0.00	0.97	0.00

Table D.6: Parameter Estimates of the Choice and Survival Models

Notes: Table reports the parameter estimates of the (a) choice and (b) survival models. Additionally, we report the percentage of markets in which the estimates are positive or negative, as well as the share in which they are statistically significant at the 5% level. For brevity we omit the coefficients from the 19 diagnosis codes that enter the survival equation.

		(1)	(2)	(3)	(4)
(a) Resident Cases Per Bed	Non-Selection Adjusted $\hat{\beta}_{jt}^{NA}$	-0.00327	-0.00179	-0.000517	-0.00145
		(0.00253)	(0.00256)	(0.00254)	(0.00285)
	2019 NHC Rating		-0.00639***	0.00209	0.00429*
			(0.00180)	(0.00190)	(0.00212)
	Mean	0.482	0.482	0.482	0.480
	Observations	13160	13087	13074	12245
(b) Resident Deaths Per Bed	Non-Selection Adjusted $\hat{\beta}_{jt}^{NA}$	-0.000897	-0.000993	-0.00132	-0.00189*
		(0.000791)	(0.000800)	(0.000805)	(0.000904)
	2019 NHC Rating		0.000258	0.00107	0.00151*
			(0.000569)	(0.000608)	(0.000697)
	Mean	0.0911	0.0912	0.0912	0.0910
	Observations	13160	13087	13074	12245
(c) Staff Cases Per Bed	Non-Selection Adjusted $\hat{\beta}_{jt}^{NA}$	0.0175***	0.00989***	0.0117***	0.0118***
		(0.00216)	(0.00214)	(0.00208)	(0.00229)
	2019 NHC Rating		0.0332***	0.0199***	0.0214***
			(0.00149)	(0.00144)	(0.00161)
	Mean	0.467	0.467	0.467	0.465
	Observations	13160	13087	13074	12245
(d) Vaccinated Residents, %	Non-Selection Adjusted $\hat{\beta}_{jt}^{NA}$	0.723***	0.401***	0.205*	0.246*
		(0.101)	(0.103)	(0.0998)	(0.113)
	2019 NHC Rating		1.462***	1.194***	1.249***
			(0.0739)	(0.0755)	(0.0862)
	Mean	87.25	87.26	87.26	87.22
	Observations	13158	13085	13072	12247
(e) Vaccinated Staff, %	Non-Selection Adjusted $\hat{\beta}_{jt}^{NA}$	1.346***	1.026***	0.852***	0.660***
		(0.129)	(0.131)	(0.130)	(0.140)
	2019 NHC Rating		1.386***	1.156***	1.308***
			(0.0943)	(0.0975)	(0.105)
	Mean	78.23	78.25	78.25	78.56
	Observations	13163	13090	13077	12251
	Additional Controls	No	No	Yes	Yes
	HRR FEs	Yes	Yes	Yes	No
	County FEs	No	No	No	Yes

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table D.7: Regressions of Covid-19 Outcomes on Nursing Home Quality Estimates using Non-Selection-Adjusted Quality Estimates $\hat{\beta}_{jt}^{NA}$

Notes: Table reports regressions of several Covid-19 outcomes on several facility characteristics. Instead of our selection-corrected estimates, we use estimates of β_{jt} obtained by setting facility-specific selection parameters α_{jt} to 0. This is equivalent to a simple risk-adjusted mortality-based quality estimate, which still includes the shrinkage properties of our Bayesian estimator. The Nursing Home Compare Overall Rating is as it appeared on the website in December 2019. Additional controls include the log of mean household income for the SNF's zip code, log number of total beds, and indicators for the for-profit status, chain membership, the presence of an Alzheimer's unit, and whether the SNF is hospital-based. Standard errors in parentheses are all clustered at the facility-level.