

# Estimating Production Functions with Latent Team Structures: An Analysis of Nursing Homes

Nihal Mehta<sup>1</sup>

Last updated: November 8, 2024

[Click here for the latest version](#)

## Abstract

I consider robust specification and estimation of production functions when the researcher observes a disaggregated vector of endogenous labor inputs. Drawing on personnel and organizational economics, I develop a latent model of matching teams of worker types with bundles of tasks under time constraints and costly team formation. I adapt its implications into a penalized and shape-constrained GMM estimator and establish its consistency. Applying it to the US nursing home industry, I estimate revenue generation and predict counterfactual labor demand and health outcomes under a proposed targeted minimum staffing mandate. I find that the policy improves care quality for long-stay patients but has mixed effects for short-stay patients: it narrows disparities and raises bottom-decile quality, but reduces mean and top-decile quality.

**JEL Codes:** C13, C51, I18, L23

**Keywords:** shrinkage, machine learning, shape restrictions, GMM, organizational design, personnel economics, healthcare, nursing homes, LASSO

---

<sup>1</sup>The Pennsylvania State University. Email: [nzm5430@psu.edu](mailto:nzm5430@psu.edu).

I am indebted to Keisuke Hirano and Andres Aradillas-Lopez for all the encouragement and guidance. I am grateful to Patrik Guggenberger, Paul Grieco, and Conor Ryan for helpful comments and discussions. I thank all participants in the Econometrics Seminar, Applied Micro Brown-bag, Econometrics Workshop, and the Student Reading Groups held at Penn State.

# 1 INTRODUCTION

Production functions are among the primary building blocks of economics. They are of independent interest to researchers and also show up as a key component when estimating many other objects of interest, such as markups, counterfactual responses, firm conduct, and productivity dispersion, among others. Fundamental to the production of even the simplest good or service is the concept of division of labor—the idea that the production process can be decomposed into bundles of tasks which are assigned to different types of workers who need to collaborate and coordinate with one another (Becker and Murphy (1992)). This leads to the emergence of a complex pattern of substitution between these worker types, posing significant challenges when specifying and estimating production functions. These challenges are further exacerbated by the fact that firms’ hiring decisions are endogenous.

Given this motivation, we consider robust specification and estimation of production functions when the econometrician observes a disaggregated vector of endogenously chosen labor inputs. We develop a latent model of production informed by the theories of personnel and organizational economics. It has four key features: (1) worker types are allocated across teams, (2) teams are matched to bundles of tasks, (3) team formation is costly, and (4) each type’s time budget is constrained. The tasks and team structure are not observed by the researcher. However, as we will show later, they have implications we exploit in the form of a production function specification with constraints and sparsities over its parameters.

To clarify the link between our latent model and the econometric model, start by noting that a team is simply a group of worker types. A key insight from the latent model is that worker types are complementary within teams while team outputs are substitutable in contributing to the firm’s overall output. We incorporate this information by specifying a production function that accounts for all possible teams that can be formed from the available worker types, without requiring any knowledge of the firm’s exact team structure. However, firms’ optimizing behavior suggests that they will employ only a limited number of teams, balancing the benefits from having more teams (like efficient task completion) against the costs associated with too many or overly large teams (such as principal-agent issues and scheduling conflicts). To capture this, we introduce a novel "team sparsity" penalty. This penalty encourages our estimator to automatically disregard certain teams, while allowing the econometrician to remain agnostic about many details of the firm’s

underlying optimization problem. Additionally, our latent model imposes restrictions on the model's primitive elements— such as the "no free lunch" condition and diminishing returns to scale. These restrictions ensure that, despite its flexibility, the overall production function adheres to meaningful shape-constraints, such as non-negativity, monotonicity, concavity, and supermodularity. Given that the input choices are likely to be correlated with the firm and time varying productivity which is unobserved by the econometrician, we allow for standard ways discussed in the literature of dealing with endogeneity, such as dynamic panel and proxy variable approaches. Combining all of this, we propose a penalized and shape-constrained GMM estimator and establish its consistency. This paper demonstrates a concrete way of integrating domain specific institutional knowledge with machine learning for robust estimation of structural models.

In empirical work, it is common to see applied researchers put forth an ad hoc specification, and, in the process, drop variables or lump them into indices. This might yield parsimony and interpretability at the cost of introducing omitted variable bias and specification errors. For example, one may simply add up the hours worked by all types or split them based on heuristics, such as high-skilled versus low-skilled worker types. Alternatively, one can opt for an off-the-shelf statistical technique like LASSO with a flexible polynomial specification. This penalizes the absolute magnitude of coefficients to automatically select a subset of predictors. In practice, we might end up eliminating entire inputs or discarding higher-order and interaction terms solely based on a statistical criteria. Similarly, Principal Components Analysis (PCA) generates linear indices to maximize the variance explained by those indices in the original input matrix. These lack economic rationale and as a consequence, often display poor performance in finite samples, such as negative elasticities. We corroborate this intuition through Monte-Carlo experiments and demonstrate the superior performance of our estimator in terms of the RMSE when estimating the production function and certain functionals of it such as direct partial elasticities.

We apply our methodology to estimate the value-added production function in the US nursing home industry. This is an ideal environment to feature our methodology for multiple reasons. First, it has been well documented that healthcare make intense use of interdisciplinary teams of workers (see [Institute of Medicine \(2001\)](#)). Second, this is a well regulated industry with high-quality data. Specifically, we are able to observe the number of hours employed of each type of worker in each facility. Third, this industry is at the forefront of critical public health concerns, particularly regarding the inadequate provision of care in

nursing homes. This problem has been further exacerbated by the Covid-19 pandemic (see [Chidambaram and Burns \(2022\)](#)) and private equity takeover (see [Gupta et al. \(2023\)](#)). As a result, many policies are currently being debated, and evaluating them would hinge on how well we can estimate this production function.

We treat net revenue as the scalar output of interest and consider, along with capital assets, the number of hours worked by five different types of workers: administrative staff, nursing staff, therapists, wellness staff, and specialists. This leads to a total of 31 teams that can potentially be formed. Our method select 8 out of these 31 teams. Large teams with 4 or more types and small singleton teams are both unlikely to occur. This suggests the presence of significant barriers to team formation. We find that specialists are concentrated in a few large, high-value teams, while the administrative staff are agnostic to team-size and the most diffuse. For each worker, we see signification variation in their within-team contribution, as well as in the contribution of that team to the overall revenue. This suggests heterogeneity in the roles these worker types play depending on who they are collaborating with.

We use these estimates to evaluate whether a minimum staffing mandate targeting nursing staff can improve patient care quality. This mandate, recently proposed by the Centers for Medicare & Medicaid Services (the US federal regulator for nursing homes), is a response to the ongoing concerns about care deficiencies in these facilities. We see that an increase in the staffing of nurses, which is expected given the incentives of the policy, is combined with a reduction in the hiring of the other worker types. However, there is considerable heterogeneity in the responses of individual firms. For example, one firm might increase hiring of the administrative staff along with the nurses to leverage skill complementarities while significantly reducing the hiring of specialists such as physicians to minimize costs.

Next, we consider the impact of this policy on the care quality for long and short stay patients, which we measure by the risk-adjusted<sup>1</sup> successful discharge rate and the percentage of patients with ADL (activities of daily living) decline respectively. We find that at the post policy staffing mix, the distribution of care quality for long-stay patients first order stochastically dominates the pre-policy distribution. This makes sense since long-stay patients primarily require health services that make use of the nursing staff (like feeding and bathing). On the other hand, the policy's impact on the distribution of care quality for short-stay patients is mixed— it (1) lowers quality disparities between facilities, (2)

---

<sup>1</sup>controlling for various facility level pre-policy patient attributes and co-morbidities.

increases quality among the bottom decile firms, (3) lowers mean quality, and (4) reduces quality among top decile facilities. Short-stay patients are admitted to nursing home after being released from a hospital following a major procedure like a surgery. They often avail services that use specialists and therapists. Reducing their staffing can thus worsen health outcomes despite an increase in nursing personnel. These trade-offs in the care quality distribution that we are able to quantify highlight the benefit of using our methodology.

Though we focus on a healthcare setting, the challenges and solutions we present are relevant to other structural models of interest. For example, in labor economics, we study how the educational outcomes of children depend on various kinds of inputs. Here, the theory of skill formation tells us of a latent grouping structure based on the nature of those inputs (time or money based) and their source (home or school based). Robust estimation of this function is important when assessing policies such as whether providing financial support to parents can narrow skill gaps across children. We shall discuss this more later.

**Related Literature:** Our paper is primarily related to the econometrics of latent group memberships. [Bonhomme and Manresa \(2015\)](#) propose grouping objects based on the K-means clustering algorithm while [Su, Shi, and Phillips \(2016\)](#) propose a variant of the LASSO for shrinking individual parameters to an unknown group-level parameter vector. Both focus primarily on grouping the fixed effects in a panel data setting so as to bypass the incidental parameter problem. So, their target is different from ours (we are interested in grouping a collection of disaggregated inputs), though their techniques can be extended to other settings. For example, [Almagro and Manresa \(2021\)](#) apply it in a BLP framework to estimate the latent structure of a nested logit model. Like [Kasahara, Schrimpf, and Suzuki \(2023\)](#), we also allow for the same input to be in multiple groups. But their focus is on latent firm types. They use a Gaussian mixture model and interpret the multiple memberships as probabilities. Penalties have been used for imposing various kinds of sparsities, such as when there are only a few valid moments in [Caner, Han, and Lee \(2018\)](#), valid instruments in [Liao \(2013\)](#), and relevant product attributes in a BLP setting in [Gillen et al. \(2019\)](#). We use penalties to impose sparsity in number of teams employed in the production process. [Bonhomme \(2021\)](#) also studies team production but their goal is to identify and estimate the individual ability of workers when we observe a collection of teams of different sizes. There is a large literature on estimation and inference under shape constraints, which we will not discuss in detail. Some important papers here are [Menzel \(2022\)](#) [Matzkin \(2013\)](#), [Blundell, Chen, and Kristensen \(2007\)](#), [Wang \(2023\)](#), [Haag, Hoderlein, and Pendakur](#)

(2009), Chernozhukov, Newey, and Santos (2023), and Chetverikov and Wilhelm (2017).

This paper also speaks to the economics of nursing homes. Many papers study the impact of Medicaid and Medicare, such as Hackmann (2019) who measures the impact of Medicaid reimbursement rates on staffing levels. Another set of papers investigate the potential misalignment of private incentives, such as Gandhi (2023) who focuses on selective admission practices, and Gupta et al. (2023) who estimate the effect of private equity infusions. A third theme studies the impact of the market structure in this industry like firm conduct and oligopolistic competition in Lin (2015), limited supply in Ching, Hayashi, and Wang (2015), and firm exits in Olenski (2023). Grieco and McDevitt (2016) identify the trade-off between the quantity (patient load served) and quality of care and Einav, Finkelstein, and Mahoney (2022) also look at a multi-output setting. We add to this literature by studying the impact of minimum staffing policies on labor demand and health outcomes.

The rest of this paper is organized as follows: Section 3 presents a latent model of firm behavior based on personnel and organizational economics. Section 4 formalizes the econometric model and introduces the shape-constrained and penalized GMM estimator. Its finite sample performance is evaluated in a Monte-Carlo study in Section 5. Section 6 focus on the empirical application to the US nursing home industry—it discusses relevant institutional background, estimates net revenue, and predicts counterfactual labor demand and health outcome distribution under the proposed minimum staffing mandate. Section 7 concludes and lists future extensions.

## 2 TEAM BASED ORGANIZATION DESIGN AND FIRM BEHAVIOR

### 2.1 General Setup

In this subsection we will present a simple econometric framework which will help us focus on the key contribution of this paper of dealing with multiple labor inputs.

Consider a balanced panel of firms  $i = 1, \dots, n$  over periods  $t = 1, \dots, T$ . A firm’s output, capital, and labor are denoted by  $Y_{it}, K_{it}, L_{it}$  respectively, with their log values denoted by the corresponding lowercase letters.  $Y_{it}$  and  $K_{it}$  are scalars. The key departure from standard production settings is that instead of also treating labor as a scalar, we allow it to be a

$d$ -dimensional vector  $L_{it} \equiv (L_{1it}, \dots, L_{dit})'$ , with  $L_{jit}$  being the number of hours employed of the  $j$ th worker type. Let  $\omega_{it}$  and  $\varepsilon_{it}$  be scalar firm and time specific productivity and idiosyncratic shocks respectively. We further assume productivity to be Hicks-neutral and the idiosyncratic shock to be log additively separable. The latter can also be interpreted as an exogenous measurement error in output. For each firm  $i$  and time  $t$ , the researcher observes  $Y_{it}$ ,  $K_{it}$  and  $L_{it}$  but not  $\omega_{it}$  and  $\varepsilon_{it}$ . Firms are sampled from an underlying population with the asymptotics of letting the number of firms  $n \rightarrow \infty$  for a fixed  $T$ . So, the data takes the form of a short panel from which we can recover the joint distribution of  $\{(y_{it}, k_{it}, \ell_{it})\}_{t=1}^T$ . We assume that the different worker types enter into the production function through the labor technology  $H(\cdot) : \mathbb{R}^d \rightarrow \mathbb{R}$ . Let  $\beta_k$  be the output elasticity of capital. Assuming Cobb-Douglas technology between the labor technology and capital, the value-added production function becomes:

$$y_{it} = \delta_0 + h(L_{it}; \theta) + \beta_k k_{it} + \omega_{it} + \varepsilon_{it}, \quad \text{where } h(L_{it}; \theta) \equiv \log H(L_{it}; \theta). \quad (1)$$

Note than we have not yet made any assumptions about the objectives of the firms when they make their investment and hiring decisions. All that we need to assume till now is that the firm's objective is increasing in its net revenue, which means that it will never produce below its production capacity i.e. firms always operate on the production possibility frontier. Also, note than the choice of both labor and capital may be correlated with the productivity shock  $\omega$ . We will discuss in more detail the strategies of dealing with this endogeneity problem in [Section 3](#). Our focus now is in understanding this labor technology  $H$ . We do so in the next subsection.

## 2.2 Team Based Production Technology

We base our latent model of matching task bundles to teams of worker types with costly team formation on the personnel and organizational economics literature. A seminal paper here is [Acemoglu and Autor \(2011\)](#), who set up this model under a scalar skill dimension. Skills are made multi-dimensional in [Lindenlaub \(2017\)](#), while [Ocampo \(2022\)](#) allows for a continuum of tasks. We extend this framework to accommodate teams.

Consider a firm that at  $t = 0$  has beliefs about future shocks and anticipates labor choice  $L$ . We will not have firm and time subscripts in this section as it is to be understood that

everything discussed will be from the point of view of a representative firm at  $t = 0$ .

Suppose there are  $d_s$  number of skills, with the intensity of each being normalized to the unit interval. This yields the skill space which is a hypercube,  $\mathcal{S} \equiv [0, 1]^{d_s}$ . Suppose  $d_s = 2$ , capturing two important skill dimensions in the nursing home context: mental health skill and physical health health. Next, suppose there are two worker types: nurse and wellness staff, represented by  $j = 1, 2$  respectively. Generalization to an arbitrary number of workers is direct. Each type  $j$  supplies  $L_j$  hours inelastically to the firm and is fully characterized by their skill level  $w_j \in \mathcal{S}$ .

Next we define the team as a non-empty collection of worker types. With just two workers, the set of all teams is given by  $\mathcal{G}^* = \{\{1\}, \{2\}, \{1, 2\}\}$ . Here,  $g = \{1\}$  and  $g = \{2\}$  denote the singleton teams of the nurse and wellness staff working alone respectively, while  $g = \{1, 2\}$  represents the team of both working together. The firm can split the labor endowment of the workers between the different teams, with  $a_{gj}$  denoting the proportion of hours of worker  $j$  that are allocated to team  $g$ .

In this skills based framework, we conceptualize a team as follows: consider the team  $g = \{1, 2\}$ , with the nurse  $j = 1$  contributing  $a_1 L_1$  hours and the wellness staff  $j = 2$  contributing  $a_2 L_2$  hours. The team pools together the resources of its constituent members, namely the skill levels and the time allocation. Suppose the pooled skill  $w_g$  lies between the skill levels of the constituent worker types i.e.  $\min_{j \in g} \{w_{js}\} \leq w_{gs} \leq \max_{j \in g} \{w_{js}\}$  for all  $s$  and  $w_g \neq w_{g'}$ . If forming a team is beneficial, then it must be that when the worker types coordinate, they exhibit some degree of complementarity. This can be reflected by their time resources being pooled in a Cobb-Douglas fashion, denoted by  $\phi_g(L; a, \gamma) = (a_{g1} L_1)^{\gamma_{g1}} (a_{g2} L_2)^{\gamma_{g2}}$ , where  $\gamma_{gj}$  is the team-time elasticity of worker  $j$  in team  $g$ . We can allow for more general ways of pooling time as well, such as using a CES specification, as long as we subject it to the constraint of the elasticity of substitution being less than or equal to 1. Some details on these generalities are provided in [Appendix B](#).

We now move onto defining tasks. A task  $t$  is a specific activity that needs to be done to realize the output of the firm. It can be fully characterized by its skill requirements, so  $t \in \mathcal{S}$ . We assume that there is a continuum of such tasks spread over the skill space with some distribution  $F$ . Given teams and tasks, we need to find a way to match them. Define the match surplus  $v_{gt}$  as the revenue realized when team  $g$  performs task  $t$  for 1 unit of



time. We can be fairly flexible about what specification it can have. What we care about are high-level conditions regarding single-crossing that ensure that this matching exists, is unique and stable. An example of such a match surplus specification is as follows:

$$v_{gt} = \exp \left( \underbrace{\left[ w_1 t_1 + w_2 t_2 \right]}_{\text{skill complementarity}} - \underbrace{\left[ (t_1 - w_1)^2 + (t_2 - w_2)^2 \right]}_{\text{skill mismatch}} \right) \quad (2)$$

Skill complementarity captures the fact that the surplus increases with the skill requirements of the task and the skill levels of the team, but gets magnified when the task and team skills are aligned along the same dimension. Skill mismatch penalizes any deviation between the task's requirements and the team's skill level.

Next, we assume that part of this surplus will be transferred to the workers in the form of their wages. This ensures that the principle of comparative advantage holds and we see the assignment of tasks to teams so as to maximize the match surplus. The set of tasks allocated to team  $g$  is given by:

$$\mathcal{T}_g = \{t \in \mathcal{S} : v_{gt} > v_{g't} \text{ for all } g' \in \mathcal{G}^* \text{ such that } g' \neq g\}. \quad (3)$$

We do not worry about ties since the set of tasks for which that holds is of measure zero. To build intuition for all that we have discussed thus far, look at the right panel of [Figure 1](#) below. We see two worker types: nurse and wellness staff, as well as their team whose embedding in the skill space is determined by the skill pooling we discussed above. Tasks are uniformly distributed, with 3 of them illustrated by a white dot. Their assignment to the workers and the team is shown by different colors, with a black hyperplane separating them. Consider a task like wound dressing, which requires physical care skills, such as medical precision and attention to hygiene, making it well-suited to the nurse's expertise. In contrast, dementia care relies heavily on mental care skills, such as providing emotional stability and cognitive engagement, areas where the wellness staff excels. Stroke rehabilitation requires a balance of both. Thus, the team of both workers should collaborate to perform it, with the nurse helping the patient regain their strength and the wellness staff helping them manage their frustration.

If collaboration is restricted, as in the left panel, then the nurse may take the lead due to a better skill alignment, despite neither worker being a good fit. Lastly, tasks like patient ambulation (not shown in the figure) require minimal physical and mental care. The

Figure 1: Task Allocation Under Two Different Organization Designs



specification in Equation 2 captures that fact that it would generate lower revenue compared to a task like stroke rehab which requires higher levels of both skills.

We can now obtain the match surplus corresponding to team  $g$ , denoted by  $v_g$ , by simply integrating  $v_{gt}$  over all tasks  $t$  that are assigned to  $g$ . Formally,

$$v_g = \int_{\mathcal{T}_g} v_{gt} dF(t). \quad (4)$$

The output of team  $g$  is then simply  $v_g$  scaled by the effective units of time worked by that team,  $\phi_g(L; a, \gamma)$ . We are now in a position to look at the firm's choice of organization design i.e. the collection of teams that it employs as well as the allocation of workers' hours within those teams. We will only look at the intuition here, with the formal notation deferred to the next subsection. In the 2 worker case, there are 8 different organization designs to choose from, with details in Table 1.

Given organization design  $\mathcal{G} \subseteq \mathcal{G}^*$ , the revenue is the surplus generated from all the tasks done, which is simply the sum of the revenue from each of the teams:

$$\sum_{g \in \mathcal{G}} v_g \phi_g(L; a, \gamma).$$

Note that adding teams is beneficial to the firm since it allows better performance of some tasks leading to higher match surplus. But there is no free lunch. The firm has to incur

Table 1: Organisation Designs in Illustrative Model with Two Worker Types

S.No	Org Design $\mathcal{G}$	# Teams	# Teams (Nurse)	# Teams (Wellness)
1	$\{\{1\}, \{2\}\}$	2	1	1
2	$\{\{1\}, \{1,2\}\}$	2	2	1
3	$\{\{2\}, \{1,2\}\}$	2	1	2
4	$\{\{1,2\}\}$	1	1	1
5	$\{\{1\}, \{2\}, \{1,2\}\}$	3	2	2
6	$\{\{1\}\}$	1	1	0
7	$\{\{2\}\}$	1	0	1
8	$\{\{\}\}$	0	0	0

fixed costs whenever it attempts to instantiate any new team.

There are primarily two types of costs. First is the coordination cost which increases with the number of workers in any team. Larger teams may face communication barriers, increased potential for conflict, and difficulties in managing workflow, which can diminish overall productivity. Second is the management cost, which increases with the number of teams. More teams necessitate more managers, more meetings, and potentially more layers of bureaucracy, all of which contribute to higher administrative and overhead expenses. There is a literature on micro-foundations of these costs. For example, in [Caplin et al. \(2023\)](#), the costs emerge from rational inattention of the managers who decide on the worker-task assignment. On the other hand, [Dessein, Lo, and Minami \(2022\)](#) attributes the costs to principal-agent incentive misalignments within the organizational hierarchy.

For a given organization design  $\mathcal{G}$ , the firm chooses an allocation vector which maximizes the total revenue net of costs associated with that design. This optimization is subject to the allocation vector satisfying two constraints. First, it must be non-negative and second, it must respect the time budget constraint, i.e., it should sum up to exactly unity for each worker across the different teams. Designs 6, 7, and 8 in [Table 1](#) are automatically rejected since they fail to meet this constraint.

Among the five remaining designs, consider the extremes: designs 1 and 4. With high management costs, the firm would consolidate and choose design 4— a single large team with both worker types. Conversely, under high coordination costs, the firm would decentralize and opt for design 1— two singleton teams working independently. Regardless of the specific cost structure, the key takeaway is that the firm is likely to employ only a subset

of all possible team configurations.

Our model and motivation are closely related to the divisional structure explored in personnel economics as in Lazear and Gibbs (2014) and organizational design as in Burton, Obel, and Håkonsson (2021). It builds on the principle that inputs within a group are complementary in generating the group's output, while outputs across groups are substitutable in contributing to the firm's overall output. In our framework, complementarity among workers within a team is captured by the resource pooling, whereas the substitutability between teams is captured by aggregating the effective surplus generated by each team to arrive at the total revenue.

Besides the divisional structure, another popular class of organization designs is referred to as the functional structure where inputs within a group are substitutable, while outputs across groups are complementary in producing the overall output. This can be relevant for production in other contexts. For example, suppose you are a labor economist interested in skill formation. You would like to know whether a remediation policy that targets parental monetary investment can narrow the ability gaps across children. It could be that inputs such as the school's spending on learning resources and teacher's instructional time are substitutable in determining 'school quality' while the time that the parents play with their child, and the family income determine 'house quality'. The production of the final output i.e. test scores can then be modelled as being complementary in these two qualities. It could also be that the school's expenditures and family income are substitutable in determining 'monetary investment' while the instructional and parental times determine 'temporal investment'. The test scores can then be complementary in these two investments. In such a case, our methodology offers researchers the ability to incorporate their domain knowledge of this functional structure while letting the grouping choice remain data-driven.

### 2.3 Firm's Choice of Organization Design

We first extend the previous model to a finite but arbitrary number of worker types  $\{1, \dots, d\}$ . The firm anticipates demanding labor  $L = (L_1, \dots, L_d)'$ . A team is a non-empty collection of worker types, with the set of all possible teams being  $\mathcal{G}^* \equiv \mathcal{P}(\{1, \dots, d\}) \setminus \{\emptyset\}$ , where  $\mathcal{P}(\cdot)$  is the power set. Let  $a_{gj}$  represent the proportion of hours of worker  $j$  allocated to team  $g$ . Each team  $g$  performs a fixed bundle of tasks for an effective time  $\phi_g(L; a, \gamma)$ .

Then, the revenue from team  $g$  is given by:

$$v_g \phi_g(L; a, \gamma) \equiv v_g \prod_{j \in g} [a_{gj} L_j]^{\gamma_{gj}}, \quad (5)$$

where  $\gamma_{gj}$  is the team-time elasticity of worker  $j$  in team  $g$ . Let  $v_g$  be the match surplus. Next, we define an organization design to be a collection of teams  $\mathcal{G} \subseteq \mathcal{G}^*$  along with a feasible allocation matrix  $A_{\mathcal{G}}$  that dictates how much time each worker type devotes to each team. Let  $a_g = a'_g$  be the allocation vector for team  $g$  and let  $a = (a_g \text{ for } g \in \mathcal{G}^*)$  be the stacked allocation vector across all teams. Then, the set of feasible allocation vectors  $A_{\mathcal{G}}$  is:

$$A_{\mathcal{G}} = \left\{ a \in \mathbb{R}^{d2^{d-1}} \mid a_{gj} = 0 \text{ if } g \notin \mathcal{G} \text{ else } a_{gj} > 0, \|a_{\bullet j}\|_1 = 1 \right\}. \quad (6)$$

For example, in the case with 2 workers that we described in the previous subsection, consider the design  $\mathcal{G} = \{\{1\}, \{1, 2\}\}$ . Then  $a_{22}$  is equal to zero to meet the time budget constraint and the space of feasible allocations  $A_{\mathcal{G}}$  is simply given by:

$$A_{\mathcal{G}} = \left\{ a \equiv [a_{11}, 0, 1 - a_{11}, 1]' \in \mathbb{R}^4 \mid a_{11} \in (0, 1) \right\}. \quad (7)$$

The revenue generated by the firm under a given organization design  $(\mathcal{G}, a)$  is:

$$S(\mathcal{G}, a; L) = \sum_{g \in \mathcal{G}} v_g \phi_g(L; a, \gamma)$$

Next we define the costs that intuitively introduced in the previous subsection. A firm with an organizational structure  $\mathcal{G}$  incurs two types of costs: The coordination cost, which increases with the number of workers in any team  $\sum_{g \in \mathcal{G}} |g|$ , and the management cost, which increases with the number of teams  $|\mathcal{G}|$ .

The firm's objective is to choose the optimal organizational structure and allocation of labor hours to maximize total revenue while minimizing coordination and management costs. The optimization problem can be formulated as follows:

$$\mathcal{G}^0, a^0 = \arg \max_{\mathcal{G} \subseteq \mathcal{G}^*, a \in A_{\mathcal{G}}} \Gamma \left( \underbrace{S(\mathcal{G}, a; L)}_{\text{Benefit (+)}}, \underbrace{\sum_{g \in \mathcal{G}} |g|, |\mathcal{G}|}_{\text{Cost (-)}} \right), \quad (8)$$

where the cost is assumed to be monotonically increasing in both of its arguments. Higher management cost will reduce the number of selected teams while a higher coordination

cost reduces the size of the selected teams. The allocation parameters must satisfy  $a_{gj} \geq 0$  and  $\|a_{\cdot j}\|_1 = 1$ . The latter ensures that all the hours of each worker are accounted for, reflecting the time budget constraint. Note that unless a worker is part of only a single team in the selected design, a boundary allocation will never be optimal for that worker since that means that there exists a team that the firm could drop without affecting its revenue while it would save on the costs it had to incur to set that team up in the first place.

The team-time elasticity parameters and the match surplus are bounded below by some strictly positive number, so  $\gamma_{gj} \geq \underline{\gamma} > 0$  and  $v_g \geq \underline{v} > 0$ . This is because in the economics of occupation, we assume that the surplus from any worker being matched to any task is strictly positive depending on the distance between the task and the worker in the skill space, as in [Equation 2](#). Furthermore, we expect the effective time worked by each team to exhibit diminishing returns, i.e.  $\|\gamma_{g\cdot}\|_1 \leq 1$ .

### 3 PENALIZED AND SHAPE-CONSTRAINED GMM ESTIMATOR

#### 3.1 Econometric Model

Recall that our value added production function in logs is given by:

$$y_{it} = \delta_0 + h(L_{it}; \theta) + \beta_k k_{it} + \omega_{it} + \varepsilon_{it}, \quad \text{where } h(L_{it}; \theta) \equiv \log H(L_{it}; \theta). \quad (9)$$

A priori, the researcher does not know the true organization design  $\mathcal{G}^o$ . Suppose we consider the superset of all teams  $\mathcal{G}^*$  yielding the following specification for  $H$ :

$$H(L_{it}; \theta) = \sum_{g \in \mathcal{G}^*} v_g \prod_{j \in g} [a_{gj} L_j]^{\gamma_{gj}}, \quad (10)$$

where  $\theta_g = (a'_{g\cdot}, \gamma'_{g\cdot}, v_g)'$  collect all parameters for team  $g$ , and  $\theta = (\theta_g \text{ for } g \in \mathcal{G}^*)$ . Observe that there is no  $i$  subscript in the labor index. For the latent model we presented in the last section, this would imply that we are assuming that all the nursing homes face the same cost structure, anticipated labor demand, and the matching function and the choice of the organization design exist, are stable and unique. Although allowing for latent firm types is a non-trivial extension, we can allow the production technology to differ based on known attributes of the firms. For example, if we except urban nursing homes with a

dedicated memory care unit to have a systematically different organization design compared to their rural standalone counterparts, then we can simply estimate the production functions separately for these two subsets of firms. We are currently undertaking this empirical exercise.

We now need to deal with the fact that both labor and capital are endogenous since the investment and hiring decisions are likely correlated with the productivity. Depending on the assumptions we are willing to make regarding input timings and the law of motion of productivity, there are two broad strategies of dealing with this. The first one is informally dubbed as the ‘proxy variable’ approach in the literature. The seminal paper here is by [Olley and Pakes \(1996\)](#), followed up with important contributions from [Levinsohn and Petrin \(2003\)](#) and [Akerberg, Caves, and Frazer \(2015\)](#). The other body of work is referred to as the ‘dynamic panel’ approach. The primary paper here is [Blundell and Bond \(2000\)](#). Going forward we shall stick to the latter approach, as set up in [Akerberg \(2023\)](#). The first strategy requires extending our framework to a semi-nonparametric setting, since some auxiliary functions are estimated using sieves (see [Appendix A](#))<sup>2</sup>. We make the following assumptions on input timings and latent variables:

**Assumption 3.1** (Information Set and Latent Variables).

1. The firm’s information set just before realizing the idiosyncratic shock  $\varepsilon_{it}$ :

$$\mathcal{I}_{it} \equiv \{k_{is}, \ell_{is}, \omega_{is}, y_{is-1}\}_{s=1}^t, \quad \text{for } i = 1, \dots, n, \text{ and } t = 1, \dots, T.$$

2. Given innovation  $\xi$ ,  $\omega$  follows a first order auto-regressive AR(1) process:

$$\omega_{it+1} = \delta_1 \omega_{it} + \xi_{it+1}, \quad \text{for } i = 1, \dots, n, \text{ and } t = 1, \dots, T-1.$$

3. The innovation shock  $\xi$  has mean zero conditional on past information:

$$\mathbb{E}[\xi_{it+1} \mid \mathcal{I}_{it}] = 0, \quad \text{for } i = 1, \dots, n, \text{ and } t = 1, \dots, T-1.$$

4. The idiosyncratic shock  $\varepsilon$  is mean zero conditional on current information:

$$\mathbb{E}[\varepsilon_{it} \mid \mathcal{I}_{it}] = 0, \quad \text{for } i = 1, \dots, n, \text{ and } t = 1, \dots, T.$$

---

<sup>2</sup>Papers such as [Valmari \(2023\)](#), [Demirer \(2020\)](#) and [Akerberg, Hahn, and Pan \(2022\)](#) focus on multi-product firms, multi dimensional productivity, and Non-Hicksian factor augmenting productivity. Extensions to those settings are possibly interesting directions for future work.

Note that the parameters of interest are  $\alpha_0 \equiv (\theta_0, \beta_0) \in \Theta \times \mathcal{B} \equiv \mathcal{A}$ , where  $\theta_0$  characterizes the labor index and  $\beta_0 \equiv (\beta_k, \delta_0, \delta_1) \in \mathbb{R}^3$  is a finite dimensional vector. Let  $X_{it} \equiv (y_{it+1}, y_{it}, k_{it+1}, k_{it}, L_{it+1}, L_{it})'$  and define the  $\delta_1$ -differenced residual:

$$\begin{aligned} \rho(X_{it}, \alpha_0) &= \delta_0 + y_{it+1} - h(L_{it+1}; \theta_0) - \beta_k k_{it+1} - \delta_1 (\delta_0 + y_{it} - h(L_{it}; \theta_0) - \beta_k k_{it}) \\ &= \omega_{it+1} + \varepsilon_{it+1} - \delta_1 (\omega_{it} + \varepsilon_{it}) && \text{(Equation 9)} \\ &= \delta_1 \omega_{it} + \xi_{it+1} + \varepsilon_{it+1} - \delta_1 (\omega_{it} + \varepsilon_{it}) && \text{(Assumption 4.1.2)} \\ &= \xi_{it+1} + \varepsilon_{it+1} - \delta_1 \varepsilon_{it}. && (11) \end{aligned}$$

Under [Assumption 3.1](#), our model yields  $T - 1$  conditional moment restrictions:

$$\mathbb{E}[\rho(X_{it}, \alpha_0) | \mathcal{I}_{it}] = 0, \quad \text{for } t = 1, \dots, T - 1. \quad (12)$$

This follows from the fact that  $\mathbb{E}[\xi_{it+1} | \mathcal{I}_{it}] = 0$  from [Assumption 4.1.3](#),  $\mathbb{E}[\varepsilon_{it} | \mathcal{I}_{it}] = 0$  from [Assumption 4.1.4](#), and  $\mathbb{E}[\varepsilon_{it+1} | \mathcal{I}_{it}] = 0$  from the law of iterated expectations and [Assumption 4.1.4](#). Note that the vector  $X_{it}$  in [Equation 12](#) can be partitioned into 2 sets—the endogenous variables  $\tilde{X}_{it}$ , and the pre-determined variables  $\tilde{\mathcal{I}}_{it}$ , which are a subset of the information set  $\mathcal{I}_{it}$ . For the information vector  $I_{it} \equiv \text{vec}(\mathcal{I}_{it})$ , we consider transformations  $\{Z_{imt}\}_{m=1}^{d_{znt}}$  which we collect into a vector  $Z_{it}^{d_{znt}} \equiv (Z_{i1t}, \dots, Z_{id_{znt}t})'$ . This is a sequence of known basis functions which can approximate any real-valued square integrable function of  $I_{it}$  arbitrarily well as  $d_{znt} \rightarrow \infty$ . Setting  $I_i \equiv (I_{i1}, \dots, I_{iT-1})'$ , we further set  $d_{zn} \equiv \sum_{t=1}^{T-1} d_{znt}$ ,  $Z_i^{d_{zn}} \equiv \left( (Z_{i1}^{d_{zn1}})', \dots, (Z_{iT-1}^{d_{znT-1}})' \right)'$ , and  $\rho(X_i, \alpha) \equiv (\rho(X_{i1}, \alpha), \dots, \rho(X_{iT-1}, \alpha))'$ . We are now ready to define the following object:

$$\psi_i^{d_{zn}}(\alpha) \equiv \begin{matrix} \rho(X_i, \alpha) & * & Z_i^{d_{zn}} \\ ((T-1) \times 1) & (d_{zn} \times 1) \end{matrix} \equiv \begin{matrix} \left[ \begin{array}{c} \rho(X_{i1}, \alpha) \otimes Z_{i1}^{d_{zn1}} \\ \vdots \\ \rho(X_{iT-1}, \alpha) \otimes Z_{iT-1}^{d_{znT-1}} \end{array} \right] \\ (d_{zn} \times 1) \end{matrix}. \quad (13)$$

For each  $\alpha$ , we take the product of each residual  $\rho(X_{it}, \alpha)$  with  $Z_{it}^{d_{znt}}$ , a vector of transformations of its information vector  $I_{it}$ . So, we have generated  $d_{zn}$  unconditional moments from  $T - 1$  conditional moments. This, along with [Equation 12](#), imply the following increasing number of unconditional moment restrictions:

$$\mathbb{E}[\psi_i^{d_{zn}}(\alpha)] = 0, \quad d_{zn} \rightarrow \infty. \quad (14)$$



In this paper, we study asymptotics at a researcher specified number of moment conditions fixed at  $d_z$  that does not grow with sample size. So,  $d_{zn} = d_z$  for all  $n$ , and we drop this superscript from here on. While common in practice, this may result in inefficiency (as the estimator is not exploiting all the information in our model) and potential identification concerns. As we will show in the next subsection, even with a fixed number of moment conditions, our eventual GMM estimator will be non-standard because of the penalty and shape-constraints we will impose on it.

The usual GMM objective for an  $O_p(1)$  positive definite  $d_z \times d_z$  weight matrix  $W_n(\alpha)$ :

$$Q_n^0(\alpha) = \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\alpha) \right]' W_n(\alpha) \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\alpha) \right]. \quad (15)$$

### 3.2 Shape-Constrained and Penalized GMM Estimator

In [subsection 3.1](#), we proposed an econometric model for production function estimation that embeds in [Blundell and Bond \(2000\)](#) a team-based specification for the labor index. In this labor index, there are  $2^{d-1}$  allocation parameters for each worker type,  $2^{d-1}$  team-time elasticity parameters for each worker type, and  $2^d - 1$  match surplus parameters, one for each non-empty team. Other than that, we have three additional parameters not related to the labor index: the constant  $\delta_0$ , the output elasticity of capital  $\beta_k$ , and the auto-regressive coefficient for productivity  $\delta_1$ . Thus, the total number of parameters in the model is given by  $d_\alpha = d(2^{d-1}) + d(2^{d-1}) + 2^d - 1 + 3 = (d+1)2^d + 2$ . This grows exponentially with the number of worker types. We could end up in a high-dimensional regime with  $d_\alpha \gg n$ . Fortunately, the information we get from the model in [Section 2](#) will help us considerably restrict our parameter space. That is our goal in this subsection.

It makes sense for a worker to contribute some strictly positive amount to the effective units of time for them to be meaningfully considered part of a team. Thus, team-time elasticities are bounded below by a small known strictly positive number  $\underline{\gamma}$ . Further, we require that  $\gamma_{gj}$  sums up to at most unity across workers which ensures weakly diminishing returns to scale for each team, so  $\gamma_{gj} \geq \underline{\gamma}$ , and  $\sum_{j=1}^d \gamma_{gj} \leq 1$ . Since adding teams is always beneficial, as seen in equation -, we require the match surplus from each team to be bounded below by some small but strictly positive known number  $\underline{v}$ . Allocation is non-negative and respects

the time budget constraint, adding up to exactly unity for each worker type  $j$ , so  $a_{gj} \geq 0$  and  $\sum_{g=1}^G a_{gj} = 1$ . Among other things, these constraints ensures that the overall labor index is monotonic and concave in its arguments.

Also, for the productivity process to converge to a stationary distribution, we require the auto-regressive coefficient to be less than unity in absolute value. Finally, for the output elasticity of capital to be meaningful, it has to lie in the unit interval.

Note that the parameter restrictions described thus far are simple linear equality and inequality constraints. So, the shape-constrained parameter space  $\tilde{\mathcal{A}} \subset \mathcal{A}$  is convex. From a computational perspective, it is straightforward to implement this in a constrained optimization routine. This parameter space is characterized as:

$$\tilde{\mathcal{A}} = \left\{ \alpha \in \mathcal{A} \equiv \mathbb{R}^{2^d(d+1)+2} : \begin{array}{l} (1) \ a_{gj} \geq 0 \text{ for each } j, \text{ and } g \in \mathcal{G}^*, \\ (2) \ \|a_{\bullet j}\|_1 = 1 \text{ for each } j, \\ (3) \ \gamma_{gj} \geq \underline{\gamma} > 0 \text{ for each } j, \text{ and } g \in \mathcal{G}^*, \\ (4) \ \|\gamma_{g\bullet}\|_1 \leq 1 \text{ for each } g \in \mathcal{G}^*, \\ (5) \ \nu_g \geq \underline{\nu} > 0 \text{ for each } g \in \mathcal{G}^*, \\ (6) \ \beta_k \in [0, 1], \\ (7) \ \delta_1 \in [-1, 1] \end{array} \right\}. \quad (16)$$

Next, the costs associated with implementing a particular organization design as in [Equation 8](#) promote sparsity in the collection of allocation vectors  $\{a_{g\bullet}\}_{g \in \mathcal{G}^*}$  in the sense that  $\sum_{g \in \mathcal{G}^*} \infty \{a_{g\bullet} \neq \mathbf{0}\} \ll |\mathcal{G}^*|$ . Let  $p_g \equiv \dim(a_{g\bullet})$  represent the number of workers in team  $g$ , and  $\|a_{g\bullet}\|_2 = \sqrt{\sum_{j \in \mathcal{G}} a_{gj}^2}$  be the  $\ell_2$ -norm of the allocation vector. Although the exact cost structure is not known to the researcher, we know that it has two arguments—coordination and management costs—and increases monotonically in both. We propose the a penalty based on these considerations:

$$\text{Penalty} = \lambda_n \sum_{g=1}^G \sqrt{p_g} \|a_{g\bullet}\|_2. \quad (17)$$

We now examine the components of this penalty. Firstly, coordination cost implies that organizations are unlikely to maintain large teams, which we capture in our penalty via

the term  $\sqrt{p_g}$ —increasing the penalty at a diminishing rate as team size grows. Secondly, management cost suggests that the number of teams in an organization should remain limited, reflected in the term  $\lambda_n$  which we leave as a researcher-specified tuning parameter. The penalty targets the  $\ell_2$ -norm of the allocation vectors, incentivizing the estimator to eliminate entire allocation vectors for some teams. While we could have applied this selection process to the match surplus parameters  $v_g$ , doing so would risk allocating non-zero worker hours to teams with zero match surplus, potentially invalidating the time budget constraint in Equation 16. Moreover, firms’ optimizing behavior suggests that allocation in teams not part of the chosen design will be zero, whereas match surplus is a primitive we expect to remain positive, even if small, irrespective of the chosen design.

Besides aligning with the economic objective, this penalty also serves as a statistical regularization criterion that promotes model parsimony. So, we might end up eliminating a team that is part of the true organization design if the value from retaining it in terms of the GMM loss is not sufficiently high given our sample.

By integrating the shape-constraints from Equation 16 and the penalty from Equation 17 into the usual GMM objective of Equation 15, we are now ready to characterize our proposed penalized and shape-constrained GMM estimator as follows:

$$\hat{\alpha} = \arg \min_{\alpha \in \tilde{\mathcal{A}}} Q_n(\alpha) \equiv Q_n^0(\alpha) + \lambda_n \sum_{g \in \mathcal{G}^*} \sqrt{p_g} \|a_g\|_2. \quad (18)$$

Note that the penalty has similarities with the Group-LASSO penalty studied in Yuan and Lin (2006). Their focus is on a linear regression setting and does not involve shape constraints. Our implementation of the estimator takes the form of a projected gradient descent algorithm which is motivated by the following proposition, in which we formalize the conditions under which our proposed estimator promotes group-wise sparsity in the parameter vector. It is a direct consequence of the Karush–Kuhn–Tucker conditions.

**Proposition 1.** *Consider the parameter vector  $\alpha = (\alpha_1, \dots, \alpha_p) \in \mathbb{R}^p$ , partitioned into  $G$  disjoint groups  $\alpha_g (g = 1, \dots, G)$ . The penalized and constrained GMM estimator  $\hat{\alpha}$  is defined as:*

$$\hat{\alpha} = \arg \min_{\alpha} \left( Q_n^0(\alpha) + \lambda \sum_{g=1}^G \|\alpha_g\|_2 \right)$$

subject to the constraints: (1)  $\sum_{j=1}^p \alpha_j = 1$ , and (2)  $\alpha_j \geq 0$ ,  $\forall j = 1, \dots, p$ , where  $Q_n^0(\alpha) = m_{1n}(\alpha)^T W_n(\alpha) m_{1n}(\alpha)$ , and  $m_{1n}(\alpha)$  is a vector of empirical moment conditions, with  $W_n(\alpha)$  as the weighting matrix.

Then, there exists a  $\lambda^* > 0$  such that for  $\lambda \geq \lambda^*$ , the solution  $\hat{\alpha}$  exhibits group-wise sparsity, i.e., there exist groups  $g$  for which  $\hat{\alpha}_g = 0$ .

Given the empirical motivation in [Section 1](#), our first order of business is to obtain consistent estimates of the production function. That is what we formally explore in the next subsection.

### 3.3 Statistical Properties of the Proposed Estimator

Extending the arguments in [Caner \(2009\)](#) to group-LASSO, we are able to obtain consistency for our proposed estimator. We first state the assumptions standard in LASSO and GMM settings. Note that  $\psi_i(\alpha) \equiv \rho(X_i, \alpha) * Z_i$  as in [Equation 13](#).

#### Assumption 3.2.

For all  $1 \leq i \leq n, n \geq 1$ ,

- (a)  $\psi_i(\alpha)$  is identically and independently distributed;
- (b)  $|\psi_i(\theta_1) - \psi_i(\theta_2)| \leq B_i |\theta_1 - \theta_2|$   
with  $\lim_{n \rightarrow \infty} \sum_{i=1}^n \mathbb{E}[B_i^d] < \infty$ , for some  $d > 2$ ;
- (c)  $\sup_{\alpha \in \mathcal{A}} \mathbb{E}[|\psi_i(\alpha)|^d] < \infty$   
for some  $d > 2$ .

#### Assumption 3.3.

Define  $\mathbb{E}[\frac{1}{n} \sum_{i=1}^n \psi_i(\alpha)] = m_{1n}(\alpha)$ .

- (a) Assume  $m_{1n}(\alpha) \rightarrow m_1(\alpha)$  uniformly over  $\mathcal{A}$ ,  $m_{1n}(\alpha)$  is continuously differentiable in  $\alpha$ ,  $m_1(\theta_0) = 0$ , and  $m_1(\alpha) \neq 0$  for  $\alpha \neq \theta_0$ ;  $m_1(\alpha)$  is continuous in  $\alpha$ .
- (b) Define the following  $\mathcal{J} \times p$  matrix:  $R_n(\alpha) = \partial m_{1n}(\alpha) / \partial \alpha'$ . Assume that

$$R_n(\alpha) \xrightarrow{P} R(\alpha)$$

uniformly in a neighborhood  $N$  of  $\theta_0$ . Here  $R(\theta_0)$  is of full column rank, and  $R(\alpha)$  is continuous in  $\alpha$ .

**Assumption 3.4.**  $W_n(\alpha)$  is a positive definite matrix that is continuous in  $\alpha \in \mathcal{A}$ , and  $W_n(\alpha) \xrightarrow{p} W(\alpha)$  uniformly in  $\alpha$ . The matrix  $W(\alpha)$  is a symmetric nonrandom  $\mathcal{J} \times \mathcal{J}$  matrix that is continuous in  $\alpha$  and is positive definite for all  $\alpha \in \mathcal{A}$ .

We are now ready to state the consistency result for the proposed estimator.

**Theorem 1** (Consistency of the Proposed Estimator). *Under Assumptions 1-3,*

1. *If  $\lambda_n/n \rightarrow \lambda_0 \geq 0$ , then*

$$\hat{\alpha}_n \xrightarrow{p} \arg \min_{\alpha \in \Gamma} Z(\alpha),$$

$$\text{where } Z(\alpha) = m_1(\alpha)'W(\alpha)m_1(\alpha) + \lambda_0 \sum_{g=1}^G \left\| \alpha^{(g)} \right\|_2.$$

2. *If  $\lambda_n = o(n)$ ,*

$$\hat{\alpha}_n \xrightarrow{p} \alpha_0.$$

In the next section, we simulate data using real-world parameter calibrations to evaluate how our proposed estimator fares in practice and against common strategies employed in empirical research.

## 4 MONTE-CARLO SIMULATIONS

Consider the case of two worker types: clinical nursing staff and administrative staff. The number of hours employed of each in firm  $i$  at time  $t$  is given by  $L_{1it}$  and  $L_{2it}$  respectively. The fixed capital assets are denoted by  $K_{it}$ . We consider a sample of  $n = 1000$  facilities for  $T = 3$  years, yielding a balanced panel with 3,000 observations. With two worker types, we have 3 different teams under consideration: the nurse working alone, the admin working alone, and both working together. Then, the labor index  $H(L; \theta_0)$  with all three of these teams represented is given by:

$$H(L; \theta_0) = v_1(a_1 L_1)^{\gamma_1} + v_2(a_2 L_2)^{\gamma_2} + v_3(a_{31} L_1)^{\gamma_{31}} (a_{32} L_2)^{\gamma_{32}}, \quad (19)$$

where  $\theta_0 \equiv (a_1, a_2, a_{31}, a_{32}, \gamma_1, \gamma_2, \gamma_{31}, \gamma_{32})'$  is the true parameter vector that characterizes the labor index. Suppose both of the workers work alone and the team in which they work together is not deployed. So,  $a_{31} = a_{32} = 0$  in Equation 19. Since the time budget constraint has to be satisfied, this implies that  $a_1 = a_2 = 1$ .

We consider a calibrated data generating process in which we start with the empirical distribution of the 3 worker types and evolve capital based on dynamic considerations with a one period gestation lag. We allow hiring to be flexible which is based on minimizing cost subject to achieving a target level of net revenue. There is exogenous variation in the hiring of each type induced by evolution of the wage rates, with the starting rates matched to what we observe in the data compiled by the US Bureau of Labor Statistics. We assume productivity of follow an auto-regressive process. We provide step-by-step details on this data generating process in Appendix G.

We are interested in estimating not just the production function  $F$ , but certain functionals of it, such as the direct partial elasticity of substitution. For any pair of inputs  $(j_1, j_2)$ , evaluated at the input vector  $x$ , it has the following expression:

$$\sigma_{j_1 j_2}(x) = \frac{F'_1 F'_2 (x_1 F'_1 + x_2 F'_2)}{x_1 x_2 \left[ 2F'_1 F'_2 F''_{12} - (F'_2)^2 F''_{11} - (F'_1)^2 F''_{22} \right]}, \quad (20)$$

where  $F'_1$  and  $F''_1$  are the first and second order derivatives respectively with respect to  $j_1$ , and  $F''_{12}$  is the cross partial derivative with respect to  $j_1$  and  $j_2$ . It is analogous for  $F'_2$  and  $F''_2$ . All of these are evaluated at the input vector  $x$ . Thus, this formula for elasticity is measured keeping the other inputs fixed at some given level. Note that the Hicks neutrality assumption baked into Equation 39 implies that we do not need to know what the unobserved productivity is to compute Equation 20 as it would simply drop out. In multi-factor production functions, there are other ways of measuring this, such as the Allen and shadow partial elasticity of substitutions. Literature on this goes back to the 1960s with work by McFadden (1963), later consolidated in Silberberg (1978) and Sydsæter, Strøm, and Berck (2005).

We consider four models with  $S = 1000$  simulations. First, we estimate parameters using standard GMM under the true model. Second, we aggregate the two worker types into a single index  $L_{it} = L_{1it} + L_{2it}$  and apply GMM with a Cobb-Douglas specification. Third,

we use LASSO with a Bernstein polynomial specification. Lastly, we estimate with our proposed PSC-GMM (Penalized and Shape-Constrained GMM). As can be seen in Table 3, our method performs well compared to Polynomial LASSO, while the index aggregation model performs the worst<sup>3</sup>. We also check stability of selection with respect to the choice

Table 2: RMSE for the Production Function and Elasticities Across Models

Model	RMSE			
	<b>H</b>	$\sigma_{L_1L_2}$	$\sigma_{L_1K}$	$\sigma_{L_2K}$
True Specification Known	1.013	0.341	0.062	0.091
Aggregated Index + Cobb-Douglas	2.804	-	0.345	1.988
Bernstein Polynomial + LASSO	2.095	1.853	0.174	0.359
Proposed Estimator (PSC-GMM)	1.504	0.509	0.124	0.217

of the tuning parameter and the sampling distribution of the parameters in Appendix G. We considered other simulation designs as well, including when we increase the number of worker types from 2 to 3. Qualitatively, the story remains the same. Our proposed estimator performs better than an overly flexible specification with an off the shelf regularization as well as an ad-hoc grouping of nursing vs non-nursing staff. For the configuration paths, we

Table 3: RMSE for the Production Function and Elasticities Across Models

Model	RMSE						
	<b>H</b>	$\sigma_{L_1L_2}$	$\sigma_{L_1L_3}$	$\sigma_{L_2L_3}$	$\sigma_{L_1K}$	$\sigma_{L_2K}$	$\sigma_{L_3K}$
(1)	0.231	0.125	0.031	0.033	0.030	0.029	0.010
(2)	1.081	—	0.323	0.197	0.329	0.179	0.093
(3)	0.817	0.398	0.174	0.185	0.174	0.313	0.112
(4)	0.615	0.153	0.080	0.117	0.046	0.096	0.027

(1): True Model, (2): Agg Index, (3): Bernstein LASSO, (4): Prop Estimator (PSC-GMM)

plot the  $\ell_2$ -norm of the team allocation coefficients as a function of the tuning parameter.

<sup>3</sup>Note that by the very construction of the index,  $\sigma_{L_1L_2}$  will be infinity, so its RMSE is undefined.

The interaction between the time budget constraint and the penalty strength implies that when the latter is set too high, the estimator tends to select the team with both the workers together, even when this is not the correct specification. This occurs because the reduction in the GMM loss is outweighed by the increase in costs associated with selecting more teams.

## 5 EMPIRICAL APPLICATION TO THE US NURSING HOME INDUSTRY

### *5.1 Institutional Details and Background*

Nursing home expenditures totaled \$210 billion in 2023, about 5% of total health care spending, up from 3% in 1965. Over the next decade, the U.S. population, like that of many advanced economies, will age rapidly. A report by the Population Reference Bureau projects a 75% increase in the number of Americans aged 65 and older needing nursing home care, equating to about 2.3 million residents in 2030. Nursing home expenditures are thus expected to grow at an annual rate of 5.3%. Government payers account for 75% of their revenue— 60% is from Medicaid, 15% from Medicare, with out-of-pocket expense and private insurance covering the rest. Furthermore, about 70% of nursing homes are for-profit<sup>4</sup> and a similar number are in urban neighborhoods. Compared to hospitals, the role of government payers and share of for-profit owners is much larger in this industry. Only 4% of nursing homes are hospital-based and 16% have any special care unit. Policymakers have long been concerned about the low quality of care at nursing homes, often attributing this issue to for-profit ownership (see [Grabowski et al. \(2013\)](#)).

As of June 2022, 20% of Covid-19 related deaths in the US occurred in long-term care facilities<sup>5</sup>. Reports that followed identified understaffing as a key factor in the inadequate care provided. In response, on April 18, 2023, the Biden Administration issued an [executive order](#) for the development of minimum staffing standards. On April 22, 2024, the Centers for Medicare & Medicaid Services (CMS) issued these standards targeting only the nursing staff. This is an aggressive measure as it is a binding constraint for 80% of facilities currently. It will be gradually rolled out nationwide over the next five years and many of

---

<sup>4</sup>The remaining 30% are owned by non-profit entities (churches, charities, government).

<sup>5</sup>see [Chidambaram and Burns \(2022\)](#).



its aspects are subject to revision. Evaluating such a policy thus necessitates the need for a structural model.

We obtain facility-level annual data between 2017 and 2019 from publicly available CMS sources<sup>6</sup>. In each year we observe about 15,000 unique skilled nursing homes, for a total of approximately 45,000 observations. For production function estimation, our outcome of interest is net revenue and the inputs are the number of hours worked by the different worker types as well as capital assets of the firm. The *Payroll Based Journal (PBJ) Nurse Staffing and Non-Nurse Staffing* datasets provide information on the hours that different worker types are employed for in each facility for each date. It was recorded only from 2017 onwards, which is why our analysis begins from that year. We end in 2019, before the Covid-19 pandemic fundamentally altered the industry. While it records 32 different occupations, many of these distinctions stem from accounting practices rather than genuine differences in their functional roles within the production process. We consolidate them into five broad production relevant worker types: administrative staff, clinical working staff, specialists, therapists, and wellness staff. This also makes the problem computationally feasible, since the number of teams (which, as we will show later are the building blocks of our production model) grow exponentially with the number of worker types. Other ways of constructing the worker types is discussed in [Appendix D](#). Data on the financial performance of these facilities comes from the *Skilled Nursing Facility Cost Report*. Since non-profit facilities may have concerns besides profit, we focus only on only for-profit facilities for the counterfactual. After data cleanup, we are left with 8,144 nursing homes for each of the three years (2017 – 2019), yielding a balanced panel with 24,432 observations.

To quantify care quality, we estimate a model of health outcomes of patients. Nursing homes typically provide two kinds of care: short and long term. We argue that short-term care quality can be reasonably measured using a one-quarter ahead rate of successful discharges, since short term residents typically stay for less than a quarter and a non-successful discharge implies either re-hospitalization or the patient passing away which are both unambiguously bad outcomes that we wish to prevent. On the other hand, long term care can be reasonable measured by a one year ahead percentage of patients with ADL decline. These outcomes depends on the staffing mix of the facilities, which we have already talked about. But, they also depends on various facility level attributes, such as its location, size, ownership

---

<sup>6</sup>Many of the datasets we will mention have been organized and made available for research by the Long Term Care Focus research center at Brown University. See [www.ltcfocus.org](http://www.ltcfocus.org) for details.

structure, and various aspects of resident acuity. We obtain the quality measures that form the basis for the outcomes and resident acuity from the *Minimum Dataset (MDS)* and the Data from *Medicare Claims*. They are based on the average level of a nursing home's performance in certain areas of care for all the residents in that facility. We obtain facility attributes from the *Provider Information* dataset. Summary statistics of some of the variables can be found in [Table 4](#) below.

We now summarize key behaviors of nursing facilities documented in the literature that we aim to capture in modeling their production function and counterfactual responses. On October 5, 2020, Massachusetts increased the minimum nurse staffing from 0 to 3.58 total nurse staff HPRD. A penalty of 2% reduction in quarterly Medicaid payments was introduced from the following year for non-compliance. The treatment group was thus the subset of nursing homes with strongest incentives to respond: those with high Medicaid resident shares ( $\geq$  75th percentile) and initial staffing below the policy threshold ( $\text{HPRD} \leq 3.58$ ). Meeting these criteria are 1,617 out of 15,333 nursing homes nationally and 40 out of 373 nursing homes in Massachusetts. Following [Abadie \(2021\)](#), a CMS study uses synthetic control to construct a suitable control group for Massachusetts (optimally weighting donor states to match pre-treatment trends from 2015Q1 - 2020Q3). They found a statistically significant effect of this policy on the treatment group, and disproportionate hiring of less expensive nurse types to meet requirement. Two key takeaways emerge. First, firms act in line with cost minimizing behavior and second, substitution effects between the different worker types can cause a change in the staffing of non-targeted workers as well.

During our period of study, the reimbursement system<sup>7</sup> in place was RUG-IV (resource utilization group, version IV) which allowed significant flexibility to facilities in the quantity of different services they provide and get reimbursed for. This was changed later to the PDPM (patient driven payment model) system which restricts what firms can get reimbursed for based on patient characteristics. Note that adjustments to patient load occur with a lag due to limited bed capacity, flow of referrals from hospitals, and patient recovery timelines. We will use this institutional detail in our counterfactual to argue for not worrying about patient load choice when analyzing short-run response of nursing homes staffing regulations.

Next, in a time-use survey conducted by CMS in 2023, a group of clinicians shadowed

---

<sup>7</sup>Medicaid/Medicare reimbursement involves a complex process that takes into account various services provided to patients, which we are abstracting away from. It is a combination of direct billing for specific services and bundled payments (daily rates) that cover a range of care services.

Table 4: Summary Statistics of Variables Used in Production and Health Models

Variable	Count	Mean	Std. Dev.	25%	50%	75%
<b>Facility Level Resident Attributes</b>						
Number of Beds	45 040	106.9	60.7	65.0	100.0	128.0
Number of Residents	44 682	79.3	49.0	47.0	71.0	98.0
Average Age (years)	44 682	79.0	7.3	75.4	80.0	83.8
% with Low CFS	39 389	38.3	12.7	30.0	37.2	45.5
% with High CFS	16 478	15.1	13.2	8.2	14.1	20.3
% Bedfast	21 914	20.9	13.9	13.5	19.8	27.1
% Female	43 830	65.5	12.1	58.7	67.0	73.8
% White	43 552	79.6	22.0	68.5	88.0	96.8
% Under 65	21 950	22.9	17.9	12.1	20.6	31.3
% Bladder Incontinent	44 140	79.1	13.2	71.4	81.0	88.9
% Bowel Incontinent	43 130	63.6	15.4	53.1	63.9	74.4
% with Dementia	13 989	50.3	15.0	40.5	50.5	60.3
% with Hypertension	44 091	76.4	10.8	71.0	78.0	83.8
% Schizophrenic/Bipolar	18 037	18.9	18.3	7.3	15.5	25.0
% Reporting Daily Pain	13 155	7.3	5.0	3.8	6.7	10.0
% Obese	37 100	29.4	8.3	23.8	28.8	34.3
% with ADL Decline	13 785	14.2	8.0	8.6	13.2	18.8
% Rehospitalized	44 850	16.9	7.1	12.6	16.5	20.9
% Successfully Discharged	44 450	55.0	18.9	45.0	58.4	68.6
<b>Employment by Worker Type (Thousands of Hours)</b>						
Administrative Staff	46 181	11.0	8.4	6.0	9.1	13.5
Nursing Staff	46 181	101.7	68.8	56.3	88.2	129.6
Specialist	46 181	6.3	9.6	0.8	2.5	8.2
Wellness Staff	46 181	7.8	6.4	3.9	6.3	9.9
Therapist	46 181	12.6	11.7	4.8	9.7	16.8
<b>Financial Metrics (Millions of USD)</b>						
Gross Revenue	44 896	12.1	13.2	5.7	9.5	15.2
Operating Expense	44 896	2.0	8.4	0.2	1.2	3.1
Capital Assets	40 214	6.8	10.3	1.4	2.8	7.1

<sup>a</sup> A unit of observation in this sample is a (facility, year) pair.

<sup>b</sup> % ADL decline: long-stay residents with increased need for help with activities of daily living.

<sup>c</sup> CFS is Cognitive Function Scale score. Low is 1 and high is  $\geq 4$  (severe cognitive impairment).

the nursing staff for their 8 – 12 hours shifts and recorded the frequency and duration of different tasks performed. The O\*NET dataset, managed by the US Bureau of Labor

Statistics, also floats a questionnaire to workers in which they rate their skill levels and their ability to perform certain tasks. Both of these surveys point us towards a common qualitative view that providing health services and treatments to patients requires workers to collaborate with one another. However, they do so selectively—failures in coordination and task prioritization, as is evident from delayed and omitted care, point to barriers in team formation.

## 5.2 Estimation of Value Added Production Function

First, we consider the case in which we do not treat for any endogeneity and simply run an OLS regression while retaining the proposed specification, shape constraints and penalty. We first determine the value of the tuning parameter  $\lambda_{GMM}$  using K-fold cross validation for the GMM loss. Then, we find a corresponding value of the tuning parameter in the OLS setting which selects the same number of teams as we do under GMM so that the results are comparable. This turns out to be  $\lambda_{OLS} = 0.1$ . Out of the 31 teams, only 8 were selected, with our regularization term shrinking the allocation vectors corresponding to the remaining 23 teams to zero. The parameter estimates for the surviving teams are given in [Table 5](#) below:

Team	Allocation (a)	Elasticity $\gamma$	Match Surplus (v)
admin, therapist	0.55, 0.51	0.51, 0.49	4.15
specialist, therapist	0.64, 0.49	0.50, 0.50	2.75
admin	0.44	1.00	2.26
nurse	0.51	1.00	2.05
specialist	0.36	1.00	2.04
nurse, wellness	0.49, 0.79	0.54, 0.46	1.95
wellness	0.20	1.00	1.65
admin, wellness	0.01, 0.01	0.58, 0.42	1.03

Table 5: Parameter Estimates of the Surviving Teams Under OLS

In the table, the selected teams are ranked according to the match surplus they are associated with. The results from the OLS method are quite puzzling. Most of the teams are very small. The nurse, in particular, only forms a team with the wellness staff. The time-team elasticities are always hitting unity, suggesting that this constraint binds for all teams as all of them are exhibiting constant returns to scale. It is also surprising for the nurse and

specialists to not be in as many high value teams. All of this suggests that the problem of endogeneity is likely quite severe. The estimates of the auxiliary parameters: the constant and the output elasticity of capital, can be found in [Table 7](#) below. The output elasticity of capital is abnormally low. This is in line with other studies that suggest that in the presence of endogeneity, we often end up underestimating this parameter when using OLS.

Next we run the proposed penalized and shape constrained GMM procedure as laid out in [Section 3](#). The estimates we obtain for the surviving teams are in [Table 6](#) and those for the auxiliary parameters (output elasticity of capital, constant, and the auto-regressive coefficient) are recorded in [Table 7](#). Some more details on both the OLS and GMM results can be found in [Appendix F](#).

Table 6: Parameter Estimates of the Surviving Teams Under GMM

Team	Allocation (a)	Elasticity $\gamma$	Match Surplus (v)
nurse, wellness, specialist	0.15, 0.09, 0.76	0.34, 0.27, 0.10	1.87
wellness	0.81	0.99	1.76
admin, therapist	0.07, 0.22	0.39, 0.45	1.72
admin, nurse, wellness	0.20, 0.06, 0.10	0.59, 0.10, 0.20	1.60
admin	0.57	1.00	1.33
admin, specialist, therapist	0.06, 0.24, 0.16	0.38, 0.16, 0.21	1.29
admin, nurse	0.10, 0.79	0.10, 0.80	0.85
therapist	0.62	0.89	0.25

Table 7: Estimates of Auxiliary Parameters Under OLS and GMM

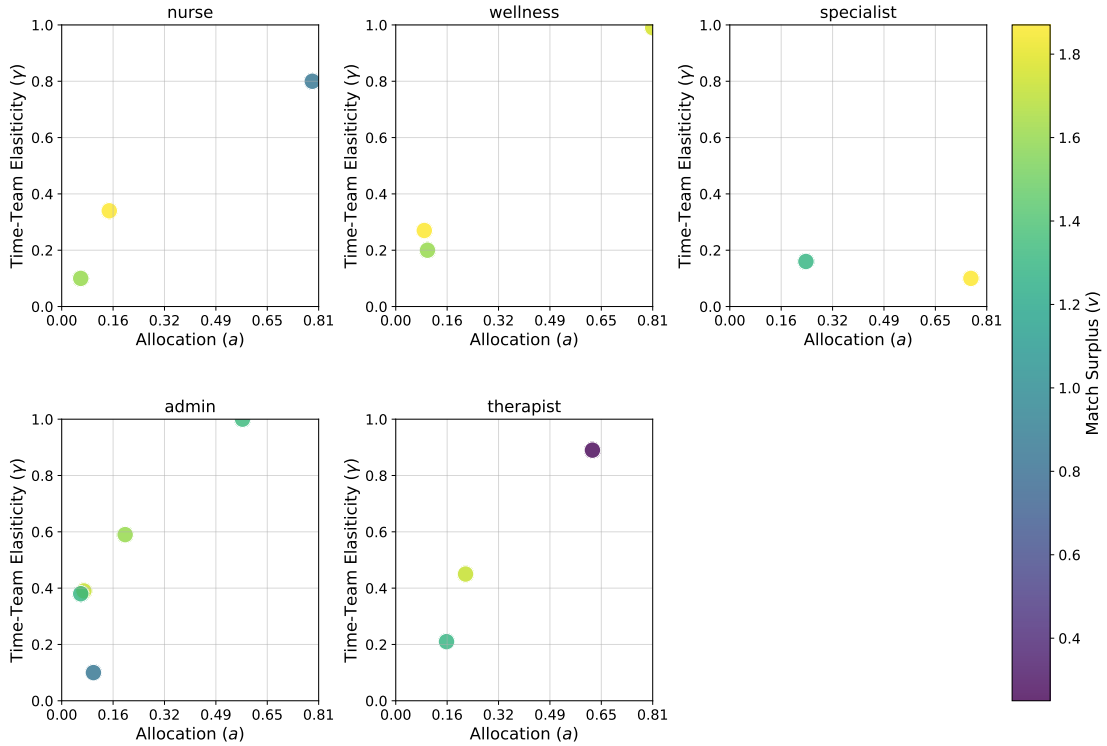
Parameter	Estimated Value (OLS)	Estimated Value (GMM)
$\beta_k$	0.14	0.35
$\delta_0$	2.71	1.26
$\delta_1$	-	0.67

The results are much more sensible now. The total number of selected teams is 8, which is one-quarter of the set of all possible team configurations. The average team size is 2, with a standard deviation of 0.87, indicating a moderate variability in the distribution of team sizes. The administrative staff is the most dispersed, participating in 5 teams. On the other hand, the specialist is the most focused, being in just 2 teams. However, those are both

large teams and around 80% of the specialist's hours are devoted to the team with the high-match surplus. This can be driven by (1) the team having a high skill level and performing complex tasks, or (2) the density of tasks allocated to this team is very high i.e. they are performing many tasks. There is an average overlap of 0.71 worker types per pair of teams. Teams having common members implies a certain level of redundancy or shared responsibilities, which could be a strategy for ensuring coverage and flexibility within the organization. Although the goal of this paper is not inference on the individual parameters or how they might identify certain primitives of the economic model, it is encouraging that the estimates are economically interpretable and align with our qualitative understanding of the organization design.

Figure 2 visualizes the estimated parameters for each worker type. Each point represents

Figure 2: Parameter Plots by Worker Types



a team, with the color indicating the team's match surplus, and the position reflecting the worker's estimated allocation and team-time elasticity parameters. The fact that the points are quite dispersed and not clustered suggests that the role and impact of each worker type

are highly team-dependent. The monotonic relationship between allocation and team-time elasticity suggests that workers who allocate more hours to a team also contribute more to its overall effective time. There is no clear link between allocation and surplus—nurses, for example, may allocate significant time to teams with lower surplus, as their tasks, such as patient monitoring, do not always drive revenue. Specialists, however, are more often involved in higher-surplus teams due to the targeted nature of their contributions.

### 5.3 *Counterfactual Labor Demand and Health Outcomes Under Minimum Staffing*

On April 22, 2024, the Centers for Medicare & Medicaid Services (CMS) issued new minimum staffing standards, which will have a staggered rollout over the next five years all across the United States. Such policies typically have three components: (1) target base of worker types, which can be specific (e.g., registered nurses) or broad (e.g., clinical nursing staff); (2) minimum threshold, expressed in hours per resident day (HPRD); and (3) consequences for non-compliance, such as a reduction in Medicaid/Medicare reimbursements. The standards stipulate a total nurse staffing requirement of 3.48 hours per resident day (HPRD), a threshold that poses a binding constraint for 80% of nursing facilities nationwide. It further specifies that out of the total 3.48 HPRD, 0.55 HPRD must be provided by registered nurses (RNs) and 2.45 HPRD by nurse aides (NAs). The remaining 0.48 HPRD can be fulfilled through any combination of RNs, licensed practical nurses (LPNs), and NAs. The American Health Care Association (AHCA) estimates<sup>8</sup> that the proposed mandate would require nursing homes to hire more than 100,000 additional nursing at an annual cost of 6.8 billion. Since we lump all nurses into the clinical nursing worker type for our empirical exercise, we will primarily be imposing the overall constraint of 3.4 HPRD on that group. In our data, we see that 60% of nursing hours come from NAs, while RNs and LPNs contribute 20% each. Therefore, when we implement the minimum staffing threshold, we implicitly assume that the hours for each worker type in the targeted group increase proportionally, maintaining the existing staff composition. In other words, we implicitly require that NAs meet 2.1 HPRD, while RNs and LPNs each meet 0.65 HPRD. This is not too far off from the actual policy requirements.

We abstract away from issues of the firm dynamically making decisions regarding patient load and mix between long and short stay patients. We take that as being decided with a

---

<sup>8</sup>details on <https://ahcancal.org/Data-and-Research/Pages/Staffing-Mandate-Analysis.aspx>.

one period lag, similar to the capital assets. Therefore, the only thing the firm is changing is the assortment of services it provides to the existing set of patients which affects its revenue, which, as described earlier, is reasonable in the RUG-IV payment scheme that was in place during this time period. We also assume that all facilities fully comply, since CMS hasn't yet formally laid out the repercussions if a facility fails to meet these standards in a stipulated time-frame. If the failure stems from exogenous conditions like a tight local labor market, then the firms are granted temporary exceptions. There are also many infractions that are dealt with on a case-by-case basis. To isolate the effect of staffing mix on welfare, we assume a partial equilibrium framework with firms facing a fixed price for their services as well as fixed wages and rents in their cost schedule. Implicitly, we assume that the demand for services in the output market and the supply of labor and capital in the input market are flat and perfectly elastic at those fixed prices. A more comprehensive welfare analysis is beyond the scope of this paper but represents an important direction for future research. Details of the counterfactual can be found in [Appendix H](#) but we provide a high level overview of the steps here.

Let  $M_{it}$  be the health outcome of interest, which is successful discharge rate for short term care quality or rate of ADL decline for long term care quality. We assume that  $M_{it}$  depends on the staffing vector  $L_{it}$ , capital assets  $K_{it}$ , firm attributes  $X_{it}$  through a parametric conditional distribution specification which we estimate using maximum likelihood method, yielding  $\hat{G}_{M|L,K,X}$ . Next, we obtain the post policy staffing mix of each firm. We suppress the  $t$  subscript in the rest of this subsection with the understanding that all variables are taken to be at their 2019 levels. We obtain the before-tax wage schedule  $w_i$  faced by the nursing home  $i$  from the US-Bureau of Labor Statistics. It varies only at the state level but does show considerable variation across states due to sticky local labor market conditions. Firm  $i$  solves the cost minimization problem which simplifies to the following:

$$L_i^* = \arg \min_{L \in \mathbb{R}_+^d} w'_s L \quad \text{subject to } L \geq \bar{L}, \text{ and } h(L; \theta) = h(L_i; \theta) \quad (21)$$

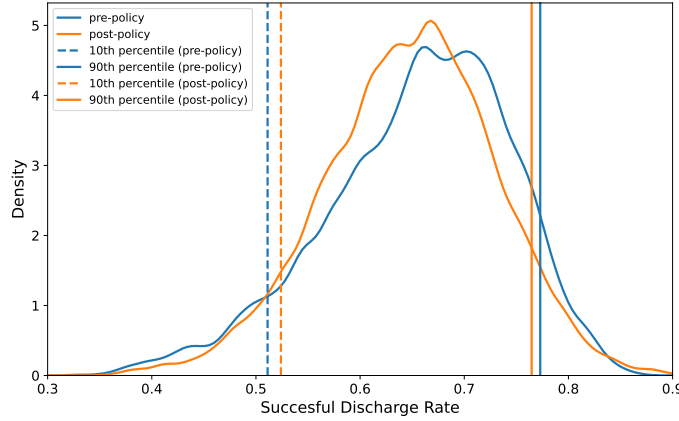
where  $\bar{L}_i$  captures the minimum staffing requirements for the labor vector. We compute the above at  $\hat{\theta}$  from our proposed estimator. The counterfactual staffing and within-facility percent change in employment of different worker types is given in . Let  $\{(L_i^*, X_i)\}_{i=1}^n$  represent the post-policy staffing vector  $L_i^* \in \mathbb{R}_+^d$  and firm attributes  $X_i \in \mathbb{R}^p$ . The empirical CDF estimator for the joint distribution of  $L_i^*, X_i$ . The estimator for the post-policy marginal



distribution of  $M_{it}$  is:

$$\hat{F}_{M^*}(m) = \int \hat{\mathbb{E}} [\mathbb{1}(M_i \leq m) \mid L_i^*, X_i] \hat{F}_{L^*, X}(L^*, X) = \frac{1}{n} \sum_{i=1}^n \hat{F}_{M \mid L_i^*, X_i}(m), \quad (22)$$

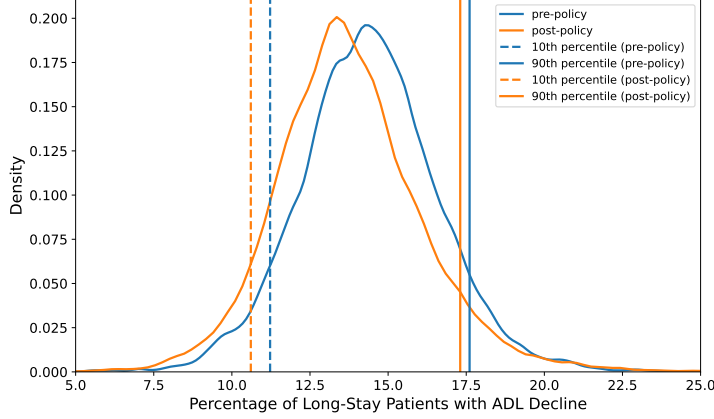
Figure 3: Pre and Post Policy Density of Successful Discharge Rate



Following the steps leading upto [Equation 52](#), we plot the pre and post policy marginal density of the successful discharge rate in the population of nursing homes in [Figure 3](#). As discussed before, it is appropriate to measure short-term quality of care.

We see an improvement in this rate by about 2 percentage points among the lower decile firms as well as drop in the rate among the upper decile by about 1 percentage point. These are economically significant effects since in absolute numbers they suggest difference in discharges of thousands of patients per year. The mean rate fell while the disparity between the facilities also reduced. These trade-offs render common decision criteria— such as first and second order stochastic dominance, inadequate to be able to conclude whether the policy is welfare enhancing. For the impact on care quality of long-stay patients, we use the proportion of patients who report a decline in their activities of daily living (ADL) as the health outcome of interest. ADL captures essential life tasks like eating, bathing, and mobility, which are critical to the well-being and quality of life of long-stay patients. This is shown in [Figure 4](#). We see an unambiguous leftward shift of the distribution, suggesting that the policy improves health outcomes for long stay patients despite reducing in hiring of non-targeted worker types. This makes intuitive sense since most of the short stay patients have a specific treatment goal which often involves non-nursing staff. On the other hand, a large proportion of long stay patients overwhelming require services that are exclusively

Figure 4: Pre and Post Policy Density of Percentage of Patients with ADL decline



provided by the nursing staff.

## 6 CONCLUSION AND NEXT STEPS

In this paper, we develop a framework for robust estimation of production functions that incorporates a disaggregated vector of endogenously chosen labor inputs. We construct a latent model of production characterized by the division of labor, where tasks are allocated to interdisciplinary teams of worker types operating under time constraints. Even without directly observing tasks and team structures, our model yields a production function specification that captures skill complementarities among team members, adheres to economically meaningful constraints (non-negativity, monotonicity, concavity, and supermodularity), and includes a novel "team sparsity" penalty to reflect firms' optimal team choices as they balance the benefits from employing teams with the costs of having too many or overly large teams. We propose a shape-constrained and penalized GMM estimator and establish its consistency. Through this specific setting, we demonstrate how domain-specific institutional knowledge can be combined with machine learning for robust structural estimation.

We apply our methodology to estimate the generation of net revenue in the US nursing home industry, considering five different worker types: administrative staff, nursing staff, specialists, wellness staff, and therapists. Our estimator selects 8 out of the 31 possible team configurations, suggesting barriers to team formation. Specialists are the most concentrated—working in only a few but high value large teams. On the other hand, the administrative staff are the most diffuse and agnostic to team size.

We then examine the effects of a counterfactual minimum staffing mandate targeting nursing staff. Firms respond by increasing nurse hiring but reducing other staff types, particularly specialists. For long-stay patients, care outcomes improve overall, while for short-stay patients, outcomes are mixed—rendering first and second order stochastic dominance criteria inadequate for comparing between the pre and post policy distributions. These findings underscore the need for the policymaker to take into account the potential for unintended consequences in care quality of the minimum staffing interventions based on how the firms re-calibrate their staffing mix. We also briefly explore how our methodology can be applied to other structural contexts where the theory informs us of latent group structures, such as for robust estimation of skill formation functions in labor economics.

Some next steps involve embedding our framework in a semi-nonparametric setting, so as to estimate auxiliary functions non-parametrically, improve efficiency by leveraging all the information in the conditional moment restrictions, and accommodate a high-dimensional regime with finer input categorizations. Additionally, we are interested in doing inference on various functionals of interest, obtaining selection consistency, and optimally choosing the tuning parameter. Also of interest is extending our methodology to do robust estimation in multi-output settings of the the production possibility frontier or the transformation function since, given limited productive capacity, firms would face a trade-off between quantity (patient load served) and quality of short and long term care provided. This can lead us towards asking what an optimal policy might look like, from among a viable set of reimbursement rates and staffing ratios.

## REFERENCES

- Abadie, Alberto. 2021. “Using Synthetic Controls: Feasibility, Data Requirements, and Methodological Aspects.” *Journal of Economic Literature* 59 (2): 391–425.
- Acemoglu, Daron, and David Autor. 2011. “Skills, Tasks and Technologies: Implications for Employment and Earnings.” Working Paper 16082, National Bureau of Economic Research.
- Ackerberg, Daniel A. 2023. “Timing Assumptions and Efficiency: Empirical Evidence in a Production Function Context.” *The Journal of Industrial Economics* 71 (3): 644–674.
- Ackerberg, Daniel A., Kevin Caves, and Garth Frazer. 2015. “Identification Properties of Recent Production Function Estimators.” *Econometrica* 83 (6): 2411–2451.
- Ackerberg, Daniel, Jinyong Hahn, and Qingsong Pan. 2022. “Nonparametric Identification Using Timing and Information Set Assumptions with an Application to Non-Hicks Neutral Productivity

- Shocks.” *Working Paper*.
- Almagro, Milena, and Elena Manresa. 2021. “Data-Driven Nests in Discrete Choice Models.”
- Andrews, Donald W.K. 1994. “Chapter 37 Empirical process methods in econometrics.” vol. 4 of *Handbook of Econometrics*, 2247–2294: Elsevier.
- Becker, Gary S., and Kevin M. Murphy. 1992. “The Division of Labor, Coordination Costs, and Knowledge.” *The Quarterly Journal of Economics* 107 (4): 1137–1160.
- Blundell, Richard, and Stephen Bond. 2000. “GMM Estimation with persistent panel data: an application to production functions.” *Econometric Reviews* 19 (3): 321–340.
- Blundell, Richard, Xiaohong Chen, and Dennis Kristensen. 2007. “Semi-Nonparametric IV Estimation of Shape-Invariant Engel Curves.” *Econometrica* 75 (6): 1613–1669.
- Bonhomme, Stephane. 2021. “Teams: Heterogeneity, Sorting, and Complementarity.”
- Bonhomme, Stéphane, and Elena Manresa. 2015. “Grouped Patterns of Heterogeneity in Panel Data.” *Econometrica* 83 (3): 1147–1184.
- Burton, Richard M., Børge Obel, and Dorthe Døjbak Håkonsson. 2021. *Organizational Design: A Step-by-Step Approach*. 4th ed.: Cambridge University Press.
- Caner, Mehmet. 2009. “Lasso-Type GMM Estimator.” *Econometric Theory* 25 (1): 270–290.
- Caner, Mehmet, Xu Han, and Yoonseok Lee. 2018. “Adaptive Elastic Net GMM Estimation With Many Invalid Moment Conditions: Simultaneous Model and Moment Selection.” *Journal of Business & Economic Statistics* 36 (1): 24–46.
- Caplin, Andrew, David J Deming, Søren Leth-Petersen, and Ben Weidmann. 2023. “Allocative Skill.” Working Paper 31674, National Bureau of Economic Research.
- Chernozhukov, Victor, Whitney K. Newey, and Andres Santos. 2023. “Constrained Conditional Moment Restriction Models.” *Econometrica* 91 (2): 709–736.
- Chetverikov, Denis, and Daniel Wilhelm. 2017. “Nonparametric Instrumental Variable Estimation Under Monotonicity.” *Econometrica* 85 (4): 1303–1320.
- Chidambaram, Priya, and Alice Burns. 2022. “10 Things About Long-Term Services and Supports (LTSS).” Accessed: 2024-06-08.
- Ching, Andrew T., Fumiko Hayashi, and Hui Wang. 2015. “Quantifying the Impact of Limited Supply: The Case of Nursing Homes.” *International Economic Review* 56 (4): 1291–1322.
- Demirer, Mert. 2020. “Essays on Productivity, Misallocation, and Firm Dynamics.” Job Market Paper, University of Pennsylvania, Department of Economics.
- Dessein, Wouter, Desmond (Ho-Fu) Lo, and Chieko Minami. 2022. “Coordination and Organization Design: Theory and Micro-Evidence.” *American Economic Journal: Microeconomics* 14 (4): 804–43.
- Einav, Liran, Amy Finkelstein, and Neale Mahoney. 2022. “Producing Health: Measuring Value Added of Nursing Homes.” Working Paper 30228, National Bureau of Economic Research.
- Gandhi, Ashvin. 2023. “Picking Your Patients: Selective Admissions in the Nursing Home Industry.” (3613950).
- Gillen, Benjamin J, Sergio Montero, Hyungsik Roger Moon, and Matthew Shum. 2019. “BLP-2LASSO for aggregate discrete choice models with rich covariates.” *The Econometrics Journal* 22 (3): 262–281.

- 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 (1): 12–21. Epub 2012 Sep 14.
- Grieco, Paul L. E., and Ryan C. McDevitt. 2016. "Productivity and Quality in Health Care: Evidence from the Dialysis Industry." *The Review of Economic Studies* 84 (3): 1071–1105.
- Gupta, Atul, Sabrina T Howell, Constantine Yannelis, and Abhinav Gupta. 2023. "Owner Incentives and Performance in Healthcare: Private Equity Investment in Nursing Homes." *The Review of Financial Studies* 37 (4): 1029–1077.
- Haag, Berthold R., Stefan Hoderlein, and Krishna Pendakur. 2009. "Testing and imposing Slutsky symmetry in nonparametric demand systems." *Journal of Econometrics* 153 (1): 33–50.
- Hackmann, Martin B. 2019. "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry." *American Economic Review* 109 (5): 1684–1716.
- Institute of Medicine. 2001. *Crossing the Quality Chasm: A New Health System for the 21st Century*. Washington, DC: National Academies Press (US).
- Kasahara, Hiroyuki, Paul Schrimpf, and Michio Suzuki. 2023. "Identification and Estimation of Production Function with Unobserved Heterogeneity."
- Lazear, Edward P., and Michael Gibbs. 2014.
- Levinsohn, James, and Amil Petrin. 2003. "Estimating Production Functions Using Inputs to Control for Unobservables." *The Review of Economic Studies* 70 (2): 317–341.
- Liao, Zhipeng. 2013. "Adaptive Gmm Shrinkage Estimation with Consistent Moment Selection." *Econometric Theory* 29 (5): 857–904.
- Lin, Haizhen. 2015. "Quality Choice and Market Structure: A dynamic analysis of nursing home oligopolies." *International Economic Review* 56 (4): 1261–1290.
- Lindenlaub, Ilse. 2017. "Sorting Multidimensional Types: Theory and Application." *The Review of Economic Studies* 84 (2): 718–789.
- Matzkin, Rosa L. 2013. "Nonparametric Identification in Structural Economic Models." *Annual Review of Economics* 5 (1): 457–486.
- McFadden, Daniel. 1963. "Constant Elasticity of Substitution Production Functions." *The Review of Economic Studies* 30 (2): 73–83.
- Menzel, Konrad. 2022. "Structural Sieves."
- Ocampo, Sergio. 2022. "A Task Based Theory of Occupations."
- Olenski, Andrew. 2023. "Reallocation and the (In)efficiency of Exit in the U.S. Nursing Home Industry."
- Olley, G Steven, and Ariel Pakes. 1996. "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica* 64 (6): 1263–97.
- Silberberg, E. 1978. *The Structure of Economics: A Mathematical Analysis.*: McGraw-Hill.
- Su, Liangjun, Zhentao Shi, and Peter C. B. Phillips. 2016. "Identifying Latent Structures in Panel Data." *Econometrica* 84 (6): 2215–2264.
- Sydsæter, Knut, Arne Strøm, and Peter Berck. 2005. *Economists' Mathematical Manual*. 4th ed. Berlin, Heidelberg: Springer.
- Vaart, A. W. van der. 1998. *Asymptotic Statistics*. Cambridge Series in Statistical and Probabilistic

- Mathematics: Cambridge University Press.
- Valmari, Nelli. 2023. "Estimating Production Functions of Multiproduct Firms." *The Review of Economic Studies* 90 (6): 3315–3342.
- Wang, Ao. 2023. "Sieve BLP: A semi-nonparametric model of demand for differentiated products." *Journal of Econometrics* 235 (2): 325–351.
- Yuan, Ming, and Yi Lin. 2006. "Model selection and estimation in regression with grouped variables." *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* 68 (1): 49–67.

## A ALTERNATIVE IDENTIFICATION STRATEGIES

We introduce a set of inputs called materials which enter the gross production function. Consider a version of the value-added production (in logs) in [Equation 9](#):

$$y_{it} = h_0(L_{it}, K_{it}) + \omega_{it} + \varepsilon_{it} \quad (23)$$

**Assumption A.1** (Timing of Input Selection).

1. Materials  $M_{it}$  are perfectly flexible inputs and chosen at time  $t$ .
2. Capital  $K_{it}$  is a quasi-fixed input and chosen at time  $t - 1$ .
3. Partially flexible labor  $L_{it}$  chosen between capital & materials at  $t - b$ ,  $b \in (0, 1)$ .

Materials maximize flow profits/minimize short run production costs. Capital with dynamic implications (LOM:  $K_{it} = \delta K_{it-1} + I_{it-1}$ ). Timing assumptions require capital and labor to be chosen before current period productivity is drawn but allows their choice to have dynamic considerations (like capital following a law of motion of the form  $K_{it} = \gamma K_{it-1} + I_{it-1}$ ) while abstracting away from the need to model it.

**Assumption A.2** (Properties of the Unobserved Productivity).

1. Productivity  $\omega_{it}$  is Markovian, so  $\omega_{it} = h_1(\omega_{it-1}) + \xi_{it}$ , where  $\mathbb{E}[\xi_{it} | I_{it-1}] = 0$ .
2. Intermediate input demand  $m_{it} \equiv \psi_t(\ell_{it}, k_{it}, \cdot)$  is strictly monotonic in  $\omega_{it}$ .

Let  $V_t \equiv (X_t, Z_t)$  be the vector of all random variables, with  $X_t \equiv (W_t, Z_t^*)$ . Here  $W_t = (y_{t+1}, \ell_{t+1})$  is endogenous while  $Z_t^* = (y_t, \ell_t, k_t, k_{t+1})$  is exogenous and is a subset of all the instruments  $Z_t \equiv I_t$ . We get these  $2T - 1$  conditional moment restrictions:

$$\mathbb{E}[\rho(X, \alpha_0) | Z] \equiv \{\mathbb{E}[\rho_t(X_t, \alpha_0) | I_t]\}_{t=1}^T = 0 \quad (24)$$

where  $\rho_t(Z, \alpha) = (\varepsilon_t, \xi_{t+1})'$  so  $d_{\rho_t} = 2$  for  $t = 1, \dots, T - 1$  and  $\rho_T(Z, \alpha) = \varepsilon_t$  so  $d_{\rho_T} = 1$ . Here,  $\alpha \equiv (h_0, h_1, h_2) \in \prod_{i=1}^3 \mathcal{H}_i \equiv \mathcal{A}$ , where  $h_0$  is the value added production function,  $h_1$

is the markovian function describing the law of motion of productivity, and  $h_2$  is a nuisance function. The residual functions  $\varepsilon_t$  and  $\xi_{t+1}$  are:

$$\begin{aligned}\varepsilon_t &= y_t - h_2(\ell_t, k_t, m_t) \\ \xi_{t+1} &= h_2(\ell_{t+1}, k_{t+1}, m_{t+1}) - h_0(\ell_{t+1}, k_{t+1}, \omega_{t+1}) - h_1(h_2(\ell_t, k_t, m_t) - h_0(\ell_t, k_t, \omega_t))\end{aligned}$$

## B SIEVE INTERPRETATION

Suppose the labor index  $H$  emerges from the performance of  $M$  teams. We also assume that the output of team  $m$  is characterized by the CES technology  $\phi_m$  as:

$$\phi_m(L; \theta_m^{(1)}) = \left[ \sum_{j=1}^d \gamma_{mj} \alpha_{mj}^{\rho_m} L_j^{\rho_m} \right]^{\tau_m / \rho_m}, \text{ for } m = 1, \dots, M,$$

where  $\theta_m^{(1)} \equiv (\gamma_{m\bullet}, \alpha_{m\bullet}, \rho_m, \tau_m)$ . Here,  $\sigma_m = 1/(1 - \rho_m)$  captures the substitutability between inputs in team  $m$ ,  $\gamma_{mj}$  and  $\alpha_{mj}$  are the relative productivity and the fraction of hours of input  $j$  in team  $m$  respectively.  $\tau_m$  is the returns to scale in team  $m$ . Let  $\phi(L; \theta^{(1)}) \equiv (\phi_1(L; \theta_1^{(1)}), \dots, \phi_M(L; \theta_M^{(1)}))$ . These team outputs are combined to yield the labor index  $H$  using the CES technology  $\psi$  as follows:

$$H(X; \theta) \equiv \psi(\phi(L; \theta^{(1)}); \theta^{(2)}) = \left[ \sum_{m=1}^M \gamma_m (\phi_m(L; \theta_m^{(1)}))^{\rho} \right]^{\tau / \rho},$$

where  $\theta^{(2)} \equiv (\gamma_{\bullet}, \rho, \tau)$  and  $\theta \equiv (\theta^{(1)}, \theta^{(2)})$ . As before,  $\sigma = 1/(1 - \rho)$  captures the substitutability between teams,  $\gamma_m$  is the relatively productivity of team  $m$ , and  $\tau$  is the returns to scale over the teams. By allowing  $M$  to be a tuning parameter that increases with the sample size, we get a nested sequence of parametric models,  $\mathcal{H}_1^d \subset \mathcal{H}_2^d \subset \dots$  which forms our sieve. Without restrictions, this sieve is very flexible and dense in the space of all continuous functions, as formalized in [Theorem 2](#):

**Theorem 2** (Universal Approximation of NG-CES DNN). *Consider the sieve*



$$\mathcal{H}_M^d = \left\{ H : [0, 1]^d \rightarrow \mathbb{R} \left| H(L; \theta) = \left( \sum_{m=1}^M \gamma_m \left[ \left( \sum_{j=1}^d \gamma_{mj} \alpha_{mj}^{\rho_m} L_j^{\rho_m} \right)^{\tau_m / \rho_m} \right]^\rho \right)^{\tau / \rho} \right. \right. \\ \left. \left. \text{with } \theta \in \Theta_M^d \equiv \left\{ \theta \in \mathbb{R}^{M(2d+3)+2} \left| \gamma_{mj}, \alpha_{mj}, \tau_m \geq 0, \sum_{m=1}^M \alpha_{mj} = 1, \rho_m \leq 1 \right. \right\} \right. \right\}$$

Here,  $\Theta_M^d$  removes redundancies and implements normalizations. Then,  $\mathcal{H}^d = \bigcup_{M=1}^{\infty} \mathcal{H}_M^d$  is dense in the space of all continuous functions  $C[0, 1]^d$  under the sup-norm metric.

The intuition is that a polynomial of any order can be generated through a specific combination of sieve coefficients for a finite  $M$ . The result then follows from the polynomials being dense in the space of continuous functions as given by the Stone Weierstrass theorem.

## C PROOFS

*Proof of Theorem 1.*

- (i) Note that  $\mathcal{G}^* = \mathcal{P}(\{1, \dots, d\}) / \emptyset$ . Let the cardinality of this set be given by  $G$ . Next, we index this collection of sets by subscript  $g = 1, \dots, G$ . This will help simplify our notation. Now, let us start by decomposing the sample average of the random vector  $\psi_i(\alpha)$  into two terms as follows:

$$\frac{1}{n} \sum_{i=1}^n \psi_i(\alpha) = \underbrace{\frac{1}{n} \sum_{i=1}^n (\psi_i(\alpha) - \mathbb{E}[\psi_i(\alpha)])}_{\text{Term 1}} + \underbrace{\frac{1}{n} \sum_{i=1}^n \mathbb{E}[\psi_i(\alpha)]}_{\text{Term 2}}.$$

We will analyze these two terms one by one. Term 2 is what we defined as  $m_{1n}(\alpha)$  in [subsection 3.3](#). Using part (a) of [Assumption 3.3](#), we know that:

$$\mathbb{E} \left[ \frac{1}{n} \sum_{i=1}^n \psi_i(\alpha) \right] \equiv m_{1n}(\alpha) \xrightarrow{p} m_1(\alpha). \quad (25)$$

Using [Assumption 3.2](#) and empirical process results from [Andrews \(1994\)](#),

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\psi_i(\alpha) - \mathbb{E}[\psi_i(\alpha)]) = O_p(1) \quad (26)$$

Then by [Assumption 25](#) and [Assumption 3.4](#), and using  $\lambda_n/n \rightarrow \lambda_0 \geq 0$ ,

$$\begin{aligned} Z_n(\alpha) &= \frac{1}{n} U_n(\alpha) = \left[ \frac{1}{n} \sum_{i=1}^n \psi_i(\alpha) \right]' W_n(\alpha) \left[ \frac{1}{n} \sum_{i=1}^n \psi_i(\alpha) \right] + \frac{\lambda_n}{n} \sum_{g=1}^G \|\alpha^{(g)}\|_2 \\ &\xrightarrow{p} m_1(\alpha)' W(\alpha) m_1(\alpha) + \lambda_0 \sum_{g=1}^G \|\alpha^{(g)}\|_2 \\ &= Z(\alpha). \end{aligned} \quad (27)$$

So using Equation (27) and Assumptions [3.2](#), [3.3](#), and [3.4](#), we have

$$\arg \min_{\alpha \in \mathcal{A}} Z_n(\alpha) \xrightarrow{p} \arg \min_{\alpha} Z(\alpha).$$

(ii) When  $\lambda_n = o(n)$ , equation (32) is modified as follows:

$$Z_n(\alpha) = \frac{1}{n} U_n(\alpha) \xrightarrow{p} m_1(\alpha)' W(\alpha) m_1(\alpha). \quad (28)$$

Since  $\alpha_0$  is identified by [Assumption 3.3](#) and [Assumption 3.4](#) and it is the unique minimum of the limit in [Assumption 28](#), using Theorem 5.7 of [Vaart \(1998\)](#) we have the consistency result:

$$\hat{\alpha}_n = \arg \min_{\alpha \in \mathcal{A}} Z_n(\alpha) \xrightarrow{p} \arg \min_{\alpha \in \Theta} [m_1(\alpha)' W(\alpha) m_1(\alpha)] = \alpha_0.$$

■

*Proof of [Proposition 1](#).*

The proof is structured as follows: (1) Define the Lagrangian with both penalty and constraints. (2) derive the Karsh-Kuhn-Tucker (KKT) conditions for optimality, and (3) establish conditions for group-wise sparsity.

### Step 1: Define the Lagrangian

The Lagrangian that incorporates the Group LASSO penalty and the constraints:

$$\mathcal{L}(\alpha, \nu, \mu) = Q_n(\alpha) + \lambda \sum_{g=1}^G \|\alpha_g\|_2 + \nu \left( \sum_{j=1}^p \alpha_j - 1 \right) - \sum_{j=1}^p \mu_j \alpha_j, \quad (29)$$

where  $\nu \in \mathbb{R}$  is the Lagrange multiplier associated with the equality constraint  $\sum_{j=1}^p \alpha_j = 1$  and  $\mu_j \geq 0$  ( $j = 1, \dots, p$ ) are the Lagrange multipliers for the non-negativity constraints  $\alpha_j \geq 0$ .

## Step 2: Derive the KKT Conditions

The Karush-Kuhn-Tucker (KKT) conditions consist of:

1. Stationarity Condition: The gradient of the Lagrangian with respect to  $\alpha_g$  must be zero:

$$\nabla_{\alpha_g} \mathcal{L} = \nabla_{\alpha_g} Q_n(\alpha) + s_g + \nu \mathbf{1}_{p_g} - \mu_g = 0, \quad \forall g = 1, \dots, G, \quad (30)$$

where  $s_g$  represents the subgradient of the Group LASSO penalty  $\lambda \|\alpha_g\|_2$ , which is defined as:

$$s_g = \begin{cases} \lambda \frac{\alpha_g}{\|\alpha_g\|_2}, & \text{if } \alpha_g \neq 0, \\ u_g, & \text{if } \alpha_g = 0 \text{ and } \|u_g\|_2 \leq \lambda. \end{cases} \quad (31)$$

Here,  $\nabla_{\alpha_g} Q_n(\alpha)$  represents the partial derivative of the GMM objective with respect to the group vector  $\alpha_g$ , and  $\mathbf{1}_{p_g}$  is a vector of ones of length  $p_g$ .

2. Primal Feasibility:

$$\sum_{j=1}^p \alpha_j = 1, \quad \alpha_j \geq 0, \quad \forall j = 1, \dots, p. \quad (32)$$

3. Dual Feasibility:

$$\mu_j \geq 0, \quad \forall j = 1, \dots, p. \quad (33)$$

4. Complementary Slackness:

$$\mu_j \alpha_j = 0, \quad \forall j = 1, \dots, p. \quad (34)$$

### Step 3: Establish Conditions for Sparsity

To understand under what conditions the solution exhibits group-wise sparsity, let us analyze the stationarity condition for each group  $\alpha_g$ .

- Case 1:  $\alpha_g \neq 0$

When  $\alpha_g \neq 0$ , the subgradient  $s_g = \lambda \frac{\alpha_g}{\|\alpha_g\|_2}$ , and the stationarity condition for group  $g$  becomes:

$$\nabla_{\alpha_g} Q_n(\alpha) + \lambda \frac{\alpha_g}{\|\alpha_g\|_2} + \nu \mathbf{1}_{p_g} = \mu_g. \quad (35)$$

Since  $\alpha_g \neq 0$ , by complementary slackness,  $\mu_j = 0$  for all  $j \in g$ . Thus, we have:

$$\nabla_{\alpha_g} Q_n(\alpha) + \lambda \frac{\alpha_g}{\|\alpha_g\|_2} + \nu \mathbf{1}_{p_g} = 0. \quad (36)$$

This equation implies that the gradient of the objective function must balance with the regularization term and the effect of the equality constraint.

- Case 2:  $\alpha_g = 0$

When  $\alpha_g = 0$ , the subgradient  $s_g = u_g$  satisfies  $\|u_g\|_2 \leq \lambda$ . The stationarity condition becomes:

$$\nabla_{\alpha_g} Q_n(\alpha) + u_g + \nu \mathbf{1}_{p_g} = \mu_g, \quad (37)$$

with  $\|u_g\|_2 \leq \lambda$ .

since  $\alpha_g = \mathbf{0}$ , by complementary slackness,  $\mu_j > 0$  for all  $j \in g$ . 3. Expressing  $u_g$  by

rearranging the stationarity condition:

$$u_g = \mu_g - \nabla_{\alpha_g} Q_n(\alpha) - \nu \mathbf{1}_{p_g}$$

The subgradient norm condition implies:

$$\|u_g\|_2 = \|\mu_g - \nabla_{\alpha_g} Q_n(\alpha) - \nu \mathbf{1}_{p_g}\|_2 \leq \lambda$$

Since  $\mu_g$  is non-negative, it follows that a sufficient condition for  $\alpha_g = \mathbf{0}$  to be optimal is:

$$\|\nabla_{\alpha_g} Q_n(\alpha) + \nu \mathbf{1}_{p_g}\|_2 \leq \lambda$$

This condition states that if the gradient of the objective function for group  $g$ , combined with the influence of the Lagrange multiplier  $\nu$ , is sufficiently small (i.e., less than or equal to the penalty parameter  $\lambda$ ), then the entire group  $\alpha_g$  will be set to zero, resulting in group-wise sparsity. ■

*Proof of [Theorem 2](#).*

Consider the case of a scalar input  $\ell \in [0, 1]$ . Results for vector-valued input follow. We know that the coefficients are constrained to be in  $\Lambda_d^{(K_1)}$  but let us restrict them further in the following way: For each node  $j \in \{1, \dots, K_1\}$  in the latent layer, let  $A_j^{(1)} = 0$ ,  $r_j^{(1)} = 1$  and  $\tau_j^{(1)} = j$ . Then, the output of each latent node  $j$  will be  $\ell^j$ . Now, constrain the output layer to just be an affine transformation, so we have  $r^{(2)} = 1$  and  $\tau^{(2)} = 1$ . The output of such a neural network with  $K_1$  nodes in the hidden layer is of the form  $h(\ell) = A^{(2)} + \gamma_1^{(2)}\ell + \dots + \gamma_{K_1}^{(2)}\ell^{K_1} \equiv \sum_{j=0}^{K_1} \gamma_j^{(2)}\ell^j$  where  $\gamma_0^{(2)} \equiv A^{(2)}$  and  $\gamma_j^{(2)} \in \mathbb{R}$  for all  $j$ .

Now, consider any continuous function over the unit interval,  $f \in C([0, 1])$  and choose any target precision  $\varepsilon > 0$ . By the Stone–Weierstrass theorem, we know that there exist  $K_1 < \infty$  and a vector of coefficients  $\lambda \equiv \gamma^{(2)} \in \mathbb{R}^{K_1+1}$  (and a corresponding function  $h(\ell, \lambda) \in \mathcal{H}_1$ ) such that the following holds:

$$\rho(f, h) \equiv \sup_{\ell \in [0, 1]} \left| f(\ell) - h(\ell, \lambda) \right| < \varepsilon.$$

So,  $\mathcal{H}_1$  is dense in the space of all continuous functions over  $[0, 1]$  when coefficients are

unconstrained (or constrained to lie in  $\tilde{\Lambda}$ ). We can extend this to vector valued  $\ell$  using the binomial expansion theorem. ■

## D DETAILS OF THE DATASET

Table 8: Summary Statistics of Employment Hours of All Worker Types

	mean	std	min	25%	50%	75%	max
beds	99.87	48.46	1.00	66.00	96.00	120.00	1361.00
bed_days	35.19	17.67	0.05	22.26	33.58	43.80	496.76
assets	9.11	25.45	-20.66	1.11	2.25	5.43	242.86
grossrev	9.37	5.65	0.13	5.26	8.13	12.04	42.73
hrs_rn_donadmin	1.75	0.52	0.00	1.61	1.86	2.00	3.92
hrs_rnadmin	3.16	2.86	0.00	1.04	2.41	4.50	17.94
hrs_rn	9.37	6.80	0.00	4.52	7.77	12.49	47.05
hrs_lpn_admin	1.74	2.21	0.00	0.00	0.99	2.64	12.41
hrs_lpn	19.89	11.78	0.00	10.84	17.96	27.11	76.53
hrs_cna	52.54	26.62	0.00	32.61	47.94	67.47	189.84
hrs_na_trn	0.70	1.50	0.00	0.00	0.00	0.59	9.24
hrs_medaide	1.82	3.37	0.00	0.00	0.00	2.14	16.66
hrs_admin	2.46	2.37	0.00	1.67	1.93	2.10	17.29
hrs_meddir	0.09	0.12	0.00	0.00	0.05	0.13	0.87
hrs_othmd	0.02	0.08	0.00	0.00	0.00	0.00	1.43
hrs_pa	0.00	0.02	0.00	0.00	0.00	0.00	0.50
hrs_np	0.03	0.12	0.00	0.00	0.00	0.00	1.31
hrs_clinnrsspec	0.00	0.03	0.00	0.00	0.00	0.00	0.67
hrs_pharmacist	0.08	0.08	0.00	0.01	0.07	0.12	1.04
hrs_dietician	0.60	0.75	0.00	0.10	0.30	0.83	6.40
hrs_feedasst	1.23	3.77	0.00	0.00	0.00	0.00	22.56
hrs_ot	1.70	1.51	0.00	0.49	1.42	2.27	9.73
hrs_otasst	2.16	1.76	0.00	0.85	1.82	3.12	10.77
hrs_otaide	0.02	0.17	0.00	0.00	0.00	0.00	2.28

	mean	std	min	25%	50%	75%	max
hrs_pt	1.71	1.55	0.00	0.50	1.40	2.29	10.39
hrs_ptasst	2.54	1.96	0.00	1.14	2.08	3.62	12.38
hrs_ptaide	0.25	0.58	0.00	0.00	0.00	0.02	3.86
hrs_respther	0.10	0.64	0.00	0.00	0.00	0.00	10.28
hrs_resptech	0.00	0.05	0.00	0.00	0.00	0.00	2.17
hrs_spcangpath	1.21	0.98	0.00	0.39	1.04	1.77	5.42
hrs_therrecspec	0.10	0.41	0.00	0.00	0.00	0.00	3.58
hrs_qualactvprof	1.37	0.99	0.00	0.36	1.68	1.93	6.65
hrs_othactv	2.65	2.70	0.00	0.42	1.96	3.89	16.34
hrs_qualsocwrk	1.37	1.20	0.00	0.02	1.57	1.95	7.26
hrs_othsocwrk	0.74	1.08	0.00	0.00	0.00	1.54	5.73
hrs_mhsvc	0.01	0.09	0.00	0.00	0.00	0.00	2.36

Table 9: Correlation Matrix of Various Variables

	<b>adm</b>	<b>nurse</b>	<b>spcl</b>	<b>well</b>	<b>ther</b>	<b>assets</b>	<b>grev</b>	<b>beds</b>	<b>bed days</b>
<b>adm</b>	1.00	0.51	0.15	0.38	0.47	0.04	0.44	0.12	0.37
<b>nurse</b>	0.51	1.00	0.21	0.50	0.64	0.09	0.76	0.22	0.64
<b>spcl</b>	0.15	0.21	1.00	0.22	0.23	0.03	0.23	0.04	0.13
<b>well</b>	0.38	0.50	0.22	1.00	0.31	0.02	0.42	0.12	0.35
<b>ther</b>	0.47	0.64	0.23	0.31	1.00	0.11	0.68	0.13	0.40
<b>assets</b>	0.04	0.09	0.03	0.02	0.11	1.00	0.21	0.04	0.10
<b>grev</b>	0.44	0.76	0.23	0.42	0.68	0.21	1.00	0.20	0.60
<b>beds</b>	0.12	0.22	0.04	0.12	0.13	0.04	0.20	1.00	0.32
<b>bed days</b>	0.37	0.64	0.13	0.35	0.40	0.10	0.60	0.32	1.00

After data cleanup (e.g., removal of outliers, missing values), we have data for 8,144 nursing homes spanning three years (2017-2019), yielding a balanced panel with 24,432 observations.

## E COMPUTATION

To estimate our model of team-based production, we encode this as a custom neural network architecture that leverages momentum-based projected stochastic gradient descent. Neural networks are chosen for their computational efficiency and ability to handle complex, high-

dimensional data. This allows us to flexibly model intricate substitution patterns among different worker types while maintaining computational feasibility.

We begin by initializing the neural network parameters. The network architecture consists of two layers with a CES (constant elasticity of substitution) activation function. The initial parameters for each node, such as  $\gamma$ ,  $\tau$ ,  $r$ , and  $A$ , are set according to predefined initial values. We utilize the Adam optimizer for efficient training, initializing moving averages for gradients ( $m_0 = 0$ ) and squared gradients ( $v_0 = 0$ ), and setting the timestep to  $k = 0$ .

For each simulation, we generate a dataset  $\mathcal{D}$  using a data-generating process (DGP) characterized by parameters  $n$ ,  $\beta$ , and  $\sigma_\epsilon$ . The dataset includes observations  $\{\ell_i, y_i\}_{i=1}^n$ . During each epoch of training, we sample a mini-batch  $\mathcal{B}$  from the dataset, introducing stochasticity, which helps the model generalize better by training on random subsets of data.

In the forward pass, the neural network computes predicted outputs  $y_{pred}$  for all sampled observations by processing the input data through its layers. This involves applying the CES activation function and combining the input features according to the network's current parameters. The predicted outputs are then used to compute the penalized and shape-constrained sample Sieve-GMM objective function:

$$\mathcal{L} \equiv \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\alpha) \right]' W_n(\alpha) \left[ \frac{1}{\sqrt{n}} \sum_{i=1}^n \psi_i(\alpha) \right] + \lambda_n \sum_{g \in \mathcal{G}^*} \sqrt{p_g} \|a_g \cdot\|_2.$$

In the backward pass, we calculate the subgradients of the objective function  $\mathcal{L}$  with respect to the model parameters  $\Theta$ . These subgradients indicate how much each parameter needs to be adjusted to minimize the objective function. The Adam optimizer updates the parameters by first adjusting the moving averages of the gradients and squared gradients using the decay rates  $\beta_1$  and  $\beta_2$ :

$$m_k \leftarrow \beta_1 m_{k-1} + (1 - \beta_1) \partial_\Theta \mathcal{L}, \quad v_k \leftarrow \beta_2 v_{k-1} + (1 - \beta_2) (\partial_\Theta \mathcal{L})^2$$

These moving averages are then corrected for bias:

$$\hat{m}_k \leftarrow \frac{m_k}{1 - \beta_1^k}, \quad \hat{v}_k \leftarrow \frac{v_k}{1 - \beta_2^k}$$

The parameters are updated using the corrected moving averages, scaled by the learning



rate  $\alpha$ :

$$\Theta \leftarrow \Theta - \alpha \frac{\hat{m}_k}{\sqrt{\hat{v}_k} + \epsilon}$$

After each update, the parameters are projected onto a feasible set  $\mathcal{C}$  to enforce the parameter constraints and normalizations discussed before:

$$\Theta \leftarrow \text{Proj}_{\mathcal{C}}(\Theta)$$

After training epochs are completed, we draw a new sample  $\mathcal{D}'$  from the DGP for evaluation purposes. The model's performance is assessed by computing the root mean square error (RMSE) on this new sample:

$$RMSE_h = \sqrt{\frac{1}{B} \sum_{i=1}^B (y'_i - h(\ell'_i; \Theta))^2}$$

We repeat this process across multiple simulations, storing the RMSE and coefficients for each run. Finally, we compute the average RMSE and average coefficients across all simulations:

$$RMSE_{\Theta_j} = \sqrt{\frac{1}{S} \sum_{sim=1}^S (\Theta_{j,sim} - \bar{\Theta}_j)^2}$$



Figure 5: Linear Cost



Figure 6: Lasso-type cost

Figure 7: Firm's Optimal Organisational Structure under Different Costs

## F SUPPLEMENTAL TABLES AND FIGURES



Figure 8: Skill Formation Model Grouping Structure

Figure 9: Plots of the Data Generating Process in the Monte Carlo Simulations

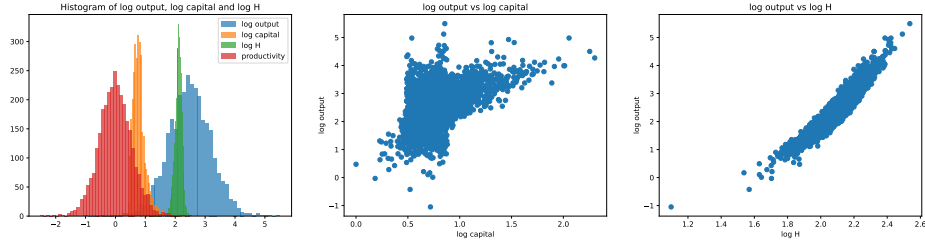


Table 10: Point Estimates and RMSE of Parameters in Nested CES Monte Carlo

Type	Coefficient	True Value	Average Estimate	RMSE
Team 1	$\gamma_{11}^{(1)}$	1	0.99	$3 \times 10^{-5}$
	$\gamma_{12}^{(1)}$	1	0.99	$6 \times 10^{-5}$
	$\gamma_{13}^{(1)}$	0	$3 \times 10^{-6}$	$3 \times 10^{-5}$
	$\gamma_{14}^{(1)}$	0	$2 \times 10^{-5}$	$2 \times 10^{-4}$
	$r_1^{(1)}$	1	0.98	0.03
	$\tau_1^{(1)}$	1	0.99	0.01
Team 2	$\gamma_{11}^{(2)}$	0	$3 \times 10^{-6}$	$4 \times 10^{-5}$
	$\gamma_{12}^{(2)}$	0	$8 \times 10^{-6}$	$6 \times 10^{-5}$
	$\gamma_{13}^{(2)}$	1	0.99	$3 \times 10^{-5}$
	$\gamma_{14}^{(2)}$	1	0.99	$2 \times 10^{-4}$
	$r_1^{(2)}$	1	0.99	0.01
	$\tau_1^{(2)}$	1	0.99	0.001
Team 3	$A^{(2)}$	1	0.99	0.008
	$\gamma_1^{(1)}$	0.3	0.30	$4 \times 10^{-3}$
	$\gamma_2^{(1)}$	0.7	0.69	$4 \times 10^{-3}$
	$r^{(2)}$	0	-0.0	$4 \times 10^{-10}$
	$\tau^{(2)}$	1	0.99	0.001

## G ADDITIONAL MONTE-CARLO RESULTS

To understand the data generating process, let us start with the first time period  $t = 1$ . For this year, for each of the variables  $L_1, L_2$  and  $K$ , we take out  $n$  i.i.d. draws from their respective empirical distributions for the year 2019. For each nursing home, we draw the

Figure 10: Non-Standard Sampling Distribution with Binding Shape Constraints



Figure 11: Configuration Paths Under Soft Time Budget (Inequality) Constraint

idiosyncratic shock  $\varepsilon_{it}$  from a mean zero normal distribution with variance  $\sigma_\varepsilon$  and assume

Figure 12: Worker-Specific Parameter Plots (OLS)



that productivity follows an AR(1) process given by:

$$\omega_{it+1} = \delta_1 \omega_{it} + \xi_{it+1}, \quad (38)$$

where  $\xi_{it}$  is an innovation shock which we assume has a mean zero normal distribution with variance  $\sigma_\xi$ . Then, for this  $t = 1$ , we initialize productivity of each nursing home to be drawn i.i.d. from its stationery distribution, which, from the normality of  $\xi$  and from [Equation 38](#), is also a mean zero normal distribution with variance  $\sigma_\xi / (1 - \delta_1)$ . Then, the output (in logs) of nursing home  $i$  at time  $t = 1$  is:

$$y_{it} \equiv \log F(X_{it}; \theta_0, \beta_0) + \varepsilon_{it} = \delta_0 + \log H(L_{1it}, L_{2it}; \theta_0) + \beta_k k_{it} + \omega_{it} + \varepsilon_{it}, \quad (39)$$

Figure 13: Team Hypergraph (GMM)

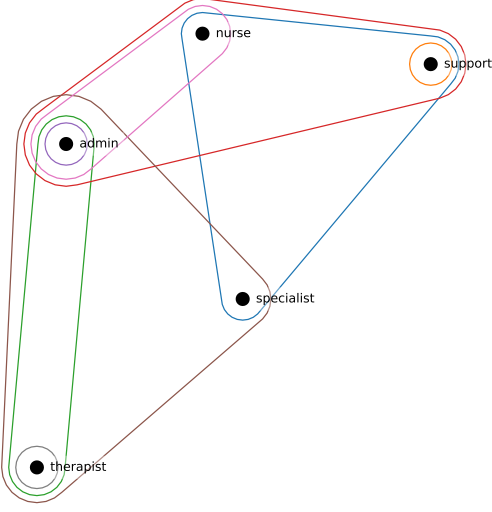
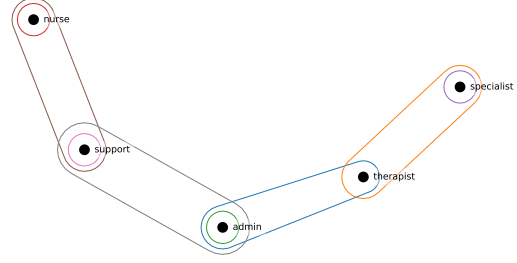


Figure 14: Team Hypergraph (OLS)



where  $X_{it} \equiv (L'_{it}, K_{it}, \omega_{it})'$ , and  $\beta_0 = (\delta_0, \beta_k)'$ . For the next time period  $t = 2$ , we assume that capital evolves according to the following law of motion:

$$K_{it+1} = (1 - \nu)K_{it} + I_{it}, \quad (40)$$

where  $\nu$  is the depreciation rate of the capital stock and  $I_{it}$  is the fresh investment made by the firm in time  $t$  which is realized after a gestation lag of 1 time period. We assume that this investment has the following functional form representation:

$$I_{it} = (\nu + \kappa \cdot \omega_{it})K_{it}. \quad (41)$$

Next, for each firm  $i$ , we draw their innovation shock from a mean zero normal distribution with variance  $\sigma_\xi^2$ . Recall that we assumed productivity  $\omega_{it+1}$  to evolve according to an AR(1) process of the form given in Equation 38. Labor is flexible and chosen to minimize short run costs subject to staying on the same isoquant:

$$\begin{aligned} L_{it+1} &= \arg \min_{L \in \mathbb{R}_+^2} w'_{it+1} L \quad \text{subject to} \\ F((L', K_{it+1}, \omega_{it+1})'; \theta_0, \beta_0) &= F((L'_{it}, K_{it}, \omega_{it})'; \theta_0, \beta_0), \end{aligned} \quad (42)$$

where the wage schedule evolves with lognormal correlated shocks as follows:

$$\log w_{it+1} = \alpha + \rho \log w_{it} + \Sigma^{1/2} \epsilon_{it+1},$$

Table 11: List of Teams Eliminated Under OLS and GMM

S. No.	Eliminated Under	Team
1	GMM	Nurse
2	GMM	Specialist
3	OLS	Therapist
4	OLS	Admin, Nurse
5	Both	Admin, Specialist
6	GMM	Admin, Wellness
7	GMM	Nurse, Wellness
8	Both	Nurse, Specialist
9	Both	Nurse, Therapist
10	Both	Wellness, Specialist
11	Both	Wellness, Therapist
12	GMM	Specialist, Therapist
13	OLS	Admin, Nurse, Wellness
14	Both	Admin, Nurse, Specialist
15	Both	Admin, Nurse, Therapist
16	Both	Admin, Wellness, Specialist
17	Both	Admin, Wellness, Therapist
18	OLS	Admin, Specialist, Therapist
19	OLS	Nurse, Wellness, Specialist
20	Both	Nurse, Wellness, Therapist
21	Both	Nurse, Specialist, Therapist
22	Both	Wellness, Specialist, Therapist
23	Both	Admin, Nurse, Wellness, Specialist
24	Both	Admin, Nurse, Wellness, Therapist
25	Both	Admin, Nurse, Specialist, Therapist
26	Both	Admin, Wellness, Specialist, Therapist
27	Both	Nurse, Wellness, Specialist, Therapist
28	Both	Admin, Nurse, Wellness, Specialist, Therapist

where  $\alpha$  is a vector of intercepts representing the overall wage growth,  $\rho$  is a diagonal matrix representing the persistence of wages for each worker type,  $\Sigma^{1/2}$  is the Cholesky decomposition of the covariance matrix  $\Sigma$ , capturing the correlations between the wages, and  $\epsilon_{it+1} \sim \mathcal{N}(0, I)$  is a vector of independent standard normal shocks. Note that the labor

choice depends on the current productivity  $\omega_{it+1}$  which means it will be endogenous. Also, we need some exogenous variation in labor for it to be identified. That is happening through the shifting of the wage schedule. The beginning wage for  $t = 1$  we use to kick-start the evolution is the wage rate prevailing in 2019 obtained from the US Bureau of Labor Statistics.

As before, we draw idiosyncratic shocks  $\varepsilon_{it} \sim \mathcal{N}(0, \sigma_\varepsilon^2)$  and obtain the output as in [Equation 39](#). We can get the data for all the later years recursively from  $t = 2$ . The list of all the parameters with their true values is given in [Table 12](#) below.

Table 12: Chosen Parameter Values for the Monte-Carlo Experiment

Parameter	True Value	Description
$a_{11}$	1.0	allocation of nurse in team (nurse)
$\gamma_{11}$	0.5	elasticity of nurse in team (nurse)
$v_1$	2.0	productivity of team (nurse)
$a_{22}$	1.0	allocation of admin in team (admin)
$\gamma_{22}$	0.8	elasticity of admin in team (admin)
$v_2$	1.0	productivity of team (admin)
$a_{31}$	0.0	allocation of nurse in team (nurse,admin)
$a_{32}$	0.0	allocation of admin in team (nurse,admin)
$\gamma_{31}$	0.4	elasticity of nurse in team (nurse,admin)
$\gamma_{32}$	0.6	elasticity of admin in team (nurse,admin)
$v_3$	0.5	productivity of team (nurse,admin)
$\beta_k$	0.3	output elasticity of capital
$\delta_0$	0.0	constant (firm and time invariant)
$\delta_1$	0.7	autoregressive coefficient
$\sigma_\varepsilon$	0.01	variance of idiosyncratic shock
$\sigma_\xi$	0.1	variance of innovation shock
$\nu$	0.1	depreciation rate of capital stock
$\kappa$	0.3	investment coefficient
$\alpha_1$	0.02	long run growth rate of nurse wages
$\alpha_2$	0.015	long run growth rate of admin wages
$\rho_1$	0.9	persistence of nurse wages
$\rho_2$	0.7	persistence of admin wages
$\Sigma_{11}$	0.01	volatility in nurse wages
$\Sigma_{22}$	0.005	volatility in admin wages
$\Sigma_{12}$	0.003	strength of correlation in wage shocks



Next, we plot the sampling distribution of the estimated parameters in Figure 15. Note that the shape constraints are binding since the true values are on the boundary of the feasible set. This leads to a non-standard sampling distribution.

Figure 15: Non-Standard Sampling Distribution with Binding Shape Constraints



## H COUNTERFACTUAL EXTENSIONS

Our counterfactual of interest is: What would the distribution of the quality of care across the universe of nursing homes in the U.S. have looked like in the year 2019 (the last year of our panel data) if minimum staffing requirements had been enforced then? Let  $M_{it}$  be the health outcome of interest, which as we argued before, is successful discharge rate for short term care quality or rate of ADL decline for long term care quality. We assume that  $M_{it}$  depends on the staffing vector  $L_{it}$  and firm attributes  $X_{it}$  through the following conditional distribution specification:

$$M_{it} \mid L_{it}, X_{it} \sim \text{Beta}(\alpha_{it}, \beta_{it}), \quad (43)$$

where  $\alpha_{it}$  and  $\beta_{it}$  are the shape parameters of the Beta distribution, which are functions of the staffing vector  $L_{it}$  and firm-level attributes  $X_{it}$ . Specifically, we assume a log-linear

specification for these parameters:

$$\alpha_{it} = \exp(Z'_{it}\pi_1), \quad \beta_{it} = \exp(Z'_{it}\pi_2), \quad (44)$$

where  $Z_{it} = (L_{it}, X_{it})$  is the combined vector of covariates, and  $\pi_1$  and  $\pi_2$  are parameter vectors to be estimated. This non-linear link function ensures that the shape parameters  $\alpha_{it}$  and  $\beta_{it}$  are strictly positive, which is required for the Beta distribution. This distribution is suitable for modeling rates, as it is defined on the unit interval and can capture different shapes of the underlying distribution. The probability density function of the Beta distribution for a given observation  $M_{it}$  is:

$$f(M_{it} | \alpha_{it}, \beta_{it}) = \frac{\Gamma(\alpha_{it} + \beta_{it})}{\Gamma(\alpha_{it})\Gamma(\beta_{it})} M_{it}^{\alpha_{it}-1} (1 - M_{it})^{\beta_{it}-1}, \quad (45)$$

where  $\Gamma(\cdot)$  denotes the Gamma function. Given a sample of  $n$  firms observed over  $T$  time periods, the log-likelihood function is written as:

$$\begin{aligned} \ell(\pi_1, \pi_2) = \sum_{i=1}^n \sum_{t=1}^T & \left[ \log \Gamma(\alpha_{it} + \beta_{it}) - \log \Gamma(\alpha_{it}) - \log \Gamma(\beta_{it}) \right. \\ & \left. + (\alpha_{it} - 1) \log M_{it} + (\beta_{it} - 1) \log(1 - M_{it}) \right], \end{aligned} \quad (46)$$

where  $\alpha_{it} = \exp(Z'_{it}\pi_1)$  and  $\beta_{it} = \exp(Z'_{it}\pi_2)$ . We can use MLE to estimate the parameter vectors  $\pi_1$  and  $\pi_2$  by maximizing the log-likelihood function as follows:

$$(\hat{\pi}_1, \hat{\pi}_2) = \arg \max_{\pi_1, \pi_2} \ell(\pi_1, \pi_2). \quad (47)$$

We have thus been able to consistently estimate the distribution of successful discharge rates  $M$  conditional on the staffing vector  $L$  and firm-time attributes  $X$  i.e.  $\hat{G}(M_{it} | L_{it}, X_{it}) \xrightarrow{P} G_0(M_{it} | L_{it}, X_{it})$  as  $\hat{\pi} \xrightarrow{P} \pi_0$ , where  $G$  is the Beta distribution, and  $\pi \equiv (\pi'_1, \pi'_2)'$ . The estimated shape parameters of the Beta Distribution are:

$$\hat{\alpha}_{it} = \exp(Z'_{it}\hat{\pi}_1), \quad \hat{\beta}_{it} = \exp(Z'_{it}\hat{\pi}_2). \quad (48)$$

Next, we obtain the post policy staffing mix of each firm. We suppress the  $t$  subscript in the rest of this subsection with the understanding that all variables are taken to be at their 2019 levels. We obtain the before-tax wage schedule  $w_i$  faced by the nursing home  $i$  from the US-Bureau of Labor Statistics. It varies only at the state level but does show considerable variation across states due to sticky local labor market conditions. Firm  $i$

solves the following cost minimization problem:

$$L_i^* = \arg \min_{L \in \mathbb{R}_+^d} w_i' L \quad \text{subject to } L \geq \bar{L}_i, \text{ and } h(L; \theta) + \beta_k k_i + \omega_i = y_i - \varepsilon_i$$

where  $\bar{L}_i$  captures the minimum staffing requirements for the labor vector. Note that  $\omega_i + \varepsilon_i$  is simply equal to  $y_i - h(L_i; \theta) - \beta_k k_i$ . Substituting that, we get:

$$L_i^* = \arg \min_{L \in \mathbb{R}_+^d} w_i' L \quad \text{subject to } L \geq \bar{L}, \text{ and } h(L; \theta) = h(L_i; \theta) \quad (49)$$

We compute the above at  $\hat{\theta}$  obtained from our proposed estimator. Let  $\{(L_i^*, X_i)\}_{i=1}^n$  represent the post-policy staffing vector  $L_i^* \in \mathbb{R}_+^d$  and firm attributes  $X_i \in \mathbb{R}^p$ . The empirical CDF estimator for the joint distribution of  $L_i^*, X_i$  can then be written as:

$$\hat{F}_{L^*, X}(L, X) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}(L_i^* \leq L, X_i \leq X), \quad (50)$$

where  $\mathbb{1}(\cdot)$  is the indicator function, which equals 1 if  $L_i^* \leq L$  and  $X_i \leq X$  hold component-wise, and 0 otherwise. It places equal probability  $\frac{1}{n}$  on each observed pair  $(L_i^*, X_i)$ . The estimator for the post-policy marginal distribution of  $M_{it}$  is:

$$\hat{F}_{M^*}(m) = \int \hat{\mathbb{E}}[\mathbb{1}(M_i \leq m) \mid L_i^*, X_i] \hat{F}_{L^*, X}(L^*, X) = \frac{1}{n} \sum_{i=1}^n \hat{F}_{M \mid L_i^*, X_i}(m), \quad (51)$$

where  $\hat{F}_{M \mid L_i^*, X_i}(m) = \text{BetaCDF}(m; \hat{\alpha}(L_i^*, X_i), \hat{\beta}(L_i^*, X_i))$ . If, instead, we were interested in estimating the post-policy marginal density, then the estimator is:

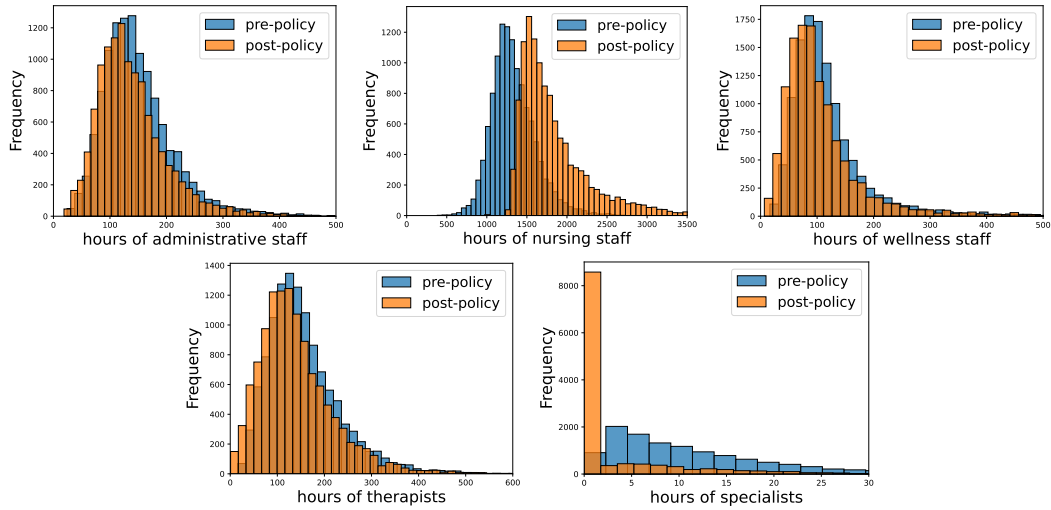
$$\hat{f}_{M^*}(m) = \int \hat{f}_{M \mid L, X}(m \mid L^*, X) d\hat{F}_{L^*, X}(L^*, X) = \frac{1}{n} \sum_{i=1}^n \hat{f}_{M \mid L_i^*, X_i}(m), \quad (52)$$

where  $\hat{f}_{M \mid L_i^*, X_i}(m) = \frac{\Gamma(\hat{\alpha}(L_i^*, X_i) + \hat{\beta}(L_i^*, X_i))}{\Gamma(\hat{\alpha}(L_i^*, X_i))\Gamma(\hat{\beta}(L_i^*, X_i))} m^{\hat{\alpha}(L_i^*, X_i)-1} (1-m)^{\hat{\beta}(L_i^*, X_i)-1}$ , and  $\hat{\alpha}(\cdot)$  and  $\hat{\beta}(\cdot)$  are as in [Equation 48](#). This is what we present in the empirical section.

## Counterfactual Staffing Mix

### Within-Firm Percentage Change in Employment of Worker Types

The intuition as to why the policy improves health outcomes for some facilities while



reducing for the others comes from the first principles of producer theory as illustrated in [Figure 16](#) above. RI is the revenue isoquant and DI is the successful discharge rate isoquant. On the axes we have 2 worker types: nurse and wellness staff. If the health isoquant is flatter than the revenue isoquant, then movement along the right on RI will cause us to be on a lower DI', as seen in the left panel. On the other hand, if DI is steeper than RI in the adjustment region, then a movement along with right on RI will put us on a higher DI', as seen in the right panel. In our counterfactual, we have 5 dimensional isoquants which can have quite complicated curvatures in different regions. So it is easy to see why some nursing homes might exhibit movement as in the left panel and others as in the right.



Figure 16: Welfare Depends on the Curvature of Production and Health Isoquants

