Population Aging and the Ballooning Cost of Nursing Homes

Francisco Cabezon*
International Monetary Fund

Mateus Dias †
Catolica-Lisbon SBE

First draft: 24th August 2023
Current draft: 9th April 2025

Abstract

How does population aging affect the prices of nursing homes? We answer this question by exploiting regional variation in the dynamics of elderly population growth rates generated by the demographic composition in the distant past combined with novel data on the prices of private and semi-private rooms in nursing homes. We estimate an elasticity of prices to the number of people aged over 75 of 0.49. We argue that this cross-sectional estimate is robust to the presence of heterogeneous treatment effects and is a lower bound for the national estimate; thus, the demand increase generated by population aging explains at least 40% of the last decade's real increase in nursing home prices, and it will further increase them between 18% and 25% during the next decade, depending on the type of room. We study the mechanism behind our results and find evidence that regulation (CON laws), not the inelastic supply of nurses, drives most of the price response to demand increase.

*fcabezon@imf.org
†matdias@ucp.pt

We thank Owen Zidar, Janet Currie, Kate Ho, Ilyana Kuziemko, and Henrik Kleven for fruitful discussions. We also thank the participants of the Princeton Brown Bag Health Seminar and the Lisbon Micro Group.

1 Introduction

The pressure to finance long-term care for the old (henceforth, LTC) in the wake of population aging is one of the main challenges economies worldwide will face. In the US, the median price of a room in a nursing home is more than double the median income of the elderly; therefore, it is not surprising that more than 60% Americans name the ability to afford LTC as one of their three main long-term concerns (OECD, 2021). Nursing homes are particularly important, representing 36% of Medicaid expenditures in 2020. This pressure will get worse because the group over 75 years of age represents 60% of the demand for nursing homes and this group will more than double in size by 2050 (United Nations, Department of Economic and Social Affairs, Population Division, 2024; Lendon et al., 2023). Thus, the effect of population aging on nursing home prices is a pressing input for public policy and household decisions.

However, estimating this effect empirically is challenging. First, it is necessary to find a compelling exogenous variation in population aging to isolate its effects on the LTC market. Second, although there is extensive administrative information on U.S. nursing homes; out-of-pocket prices charged by them are missing. We break new ground in this crucial estimation by overcoming these two issues. To deal with endogeneity, we use demographic variation from the distant past at the regional level as an instrument for variations in the current elderly population. To overcome the lack of price data, we obtained out-of-pocket price data from Genworth, an LTC insurance company, at the regional level for 2015 to 2020, which we combined with the rich available administrative data on nursing homes.

We find that the increase in demand generated by population aging has a large and significant effect on prices. From the 16% to 18% increase in the real prices of beds between 2010-2020, at least 40% of it can be attributed to the increase in demand generated by population aging. Overall, our results indicate that the response of prices to population aging is about twice as large as the demand response. Hence, the price response is responsible for about two-thirds of the increase in the expenditures on nursing homes caused by population aging. Furthermore, we find strong evidence that regulation– in the form of Certificate of Need laws–, not the inelastic supply of nurses, drives most of the price response to aging.

The data on nursing homes is obtained from three distinct sources. First, LTCFocus provides data on CMS-certified nursing homes, which includes details such as the number of beds, occupancy rates, nurse type and quantity, and patient severity, among other variables, at the facility and county levels. Second, employment and wage data pertaining to *Nursing Care Facilities* (NACIS 6231), both at county and MSA levels, comes from the County Business Patterns data. Finally, as mentioned above, Genworth's Cost of Care Survey, which surveyed over 12,000 nursing homes annually from 2015-2020, provides median room price data at the MSA level. To the best of our knowledge, these data on prices have not been used in previous work related to the nursing home and long-term care industry.

To estimate the effect of population aging on the Nursing Homes market and prices, we use the heterogeneous variation in the size elderly population in different geographic zones in the U.S. We leverage the *population accounting equation* to instrument the yearly changes in the elderly population size of each area by the change in the demographic composition of that area in the distant past. More specifically, we instrument the yearly changes in the number of people aged over 75 by the yearly changes in the number of people between 35 and 55 years old who lived in the same geographic unit (county or MSA) forty years ago. The identifying assumption is that changes in the number of people aged between 35 and 55 years old in a region does not affect the changes of the nursing home industry forty years after, other than through their effect on the change in the number of eledery people. The relevance condition for our instrument is satisfied; the instrument is strong and robust to geographic definition and controls.

We also provide evidence supporting the exclusion restriction of our instrument. Using data from censuses, births, and death certificates at the county level, we show that births, migration, and deaths 50 years ago of the relevant age group — those aged between 25 and 45 in 1970 and between 35 and 55 in 1980 — do not affect the nursing home industry other than through the current population above 75 years old. We also perform a placebo analysis and find that the response of other health industries to population aging is in line with how relevant elderly demand is for that industry. Thus, we discard that specific health industry dynamics drive our estimates. Similarly, our estimates are robust to the geographic definition, population health controls (years of potential life lost and share of local population reporting being in bad health), and socioeconomic controls at the geographic level in 2000 interacted with year dummies, namely share of male individuals, shares of the population identifying as black, white, and Hispanic, the share of people with high school, the share of people with a bachelor degree, the share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. Finally, we discard the possibility that migration dynamics are driving our results by performing an autocorrelation analysis of net migration and finding that the autocorrelation disappears after two decades.

Implementing our empirical strategy, we find that the price of beds in nursing homes is rising in response to the increase in demand generated by local increases in the elderly population. For an increase of 10% in the population aged over 75, the number of occupied nursing home beds increases by 2.4%, the number of nurse hours in the industry by 4.4%, and the prices of private rooms by 4.9%. This effect has significant economic consequences. The growth in the elderly population can explain at least 40% of the increase in nursing home prices between 2010 and 2020 (41% for semi-private rooms and 48% for private rooms), and our estimates imply that population aging alone will increase prices by an additional 18.6% and 25% for semi-private and private room, respectively, from 2020 to 2030[1]. This increase in price, which is much harder to estimate, is significantly more important than the increase in quantity demanded in this market. Based on our preferred estimates, the response of expenses to population aging without taking price responses into account would be approximately one-third of the full response. For example, Wharton's Budget Model Projection for Medicaid's Long-Term Care Expenditures does not incorporate this price increase Penn Wharton Budget Model (2022). Thus, our estimation of the price response compounding the effect of aging on the expenditures on long-term care is a key element for public health policy and policymakers.

We also study the mechanism behind this price increase and find evidence that it is not

[1]These calculations were made based on data from the World Bank DataBank Population estimates and projections. These data indicate that the population aged 75 years old or more grew about 18% between 2010 and 2020 and will grow 50%. Data from the Genworth Cost of Care survey report from 2010 and 2020 indicate the price of private and semi-private rooms increased, in real terms, by 16% and 18.6%, respectively, taking the cumulative inflation in the period as 18.7%. We then use our preferred estimates of the effect of population aging on these prices to do our back-of-the-envelope calculations.

driven by an increase in costs. We find that population aging has no effect on nurses' wages, which represent more than 80% of the total cost for the average nursing home. This aligns with previous studies showing that nursing staff employment at nursing homes and their wages are unresponsive to each other (Matsudaira, 2014; Shields, 2004). Thus, nurses, and their wages, are not the bottleneck driving the price response to the increases in demand. In contrast, we find that CON laws effectively constrain the elasticity of bed supply, driving most of the price response of nursing home rooms to population aging.

The main contribution of our paper is to connect two important branches of the literature. On the one hand, a large amount of literature estimates the demand for long-term care and its relationship with age (De Meijer et al., 2011; Balia and Brau, 2014). On the other hand, the demography literature documents that the population is aging at an increasing pace in virtually every country, a phenomenon called "the Third Demographic Transition" (Lee, 2002; Galor, 2012; World Health Organization, 2015). Both strands of the literature identify the rising long-term cost caused by population aging as a crucial topic. The gap between these branches of the literature is partially covered by studies that provide projections of long-term care expenses (Karlsson et al., 2006; de la Maisonneuve and Martins, 2015; Penn Wharton Budget Model, 2022). However, despite being very informative, the relationships between aging and costs estimated in these studies are not causal, and they omit the effect of demand on prices– one of the crucial outcome variables. Hence, the information these projections can provide for the design of policies tackling rising long-term care costs is severely limited. Our paper overcomes this limitation by providing a (and, to the best of our knowledge, the first) causal estimate of the effect of population aging on long-term care industry prices for patients.

This paper also relates to other strands of the literature. First, there is a rich literature on the macroeconomic effects of population aging. Previous studies have assessed the impact of an aging population on economic growth and productivity (Acemoglu and Restrepo, 2017; Maestas et al., 2023; Eggertsson et al., 2019; Aksoy et al., 2019), aspects related to taxation and public finance in general (Razin et al., 2002; Elmendorf and Sheiner, 2017; Ferraro and Fiori, 2020; Butler and Yi, 2022), among other macroeconomic variables or trends of interest (Lee, 2014; Cravino et al., 2022). This paper adds to this literature by providing estimates of the effects of aging on one of the industries most affected by it, the nursing home industry, which is also a relevant part of healthcare spending in many countries and, consequently, of overall fiscal spending. Regarding the empirical strategy, we are particularly close to Maestas et al. (2023). However, in our case, we use population data from further back in time, which strengthens the argument in favor of the instrument's exogeneity, and we provide strong evidence supporting the exogeneity condition in our specific analysis.

Finally, our paper relates to both the economic and public health branches of the literature that analyze the effects of changes in the supply of long-term care on different outcomes. In both cases, the studies disproportionately focus on the effect that different policy instruments and certain characteristics of nursing homes have on the quality of the service provided at such facilities, both in isolation from one another or considering their interactions. Examples include the effect of changes in Medicaid reimbursement rates (Hackmann, 2019; Grabowski, 2001), Certificate of Need and pass-through laws (Ching et al., 2015; Foster and Lee, 2015), nurse staffing (Lin, 2015; Matsudaira, 2014; Foster and Lee, 2015), and ownership status (Chou, 2002; Braun et al., 2020, 2021; Comondore et al., 2009; Grabowski and Hirth, 2003; Grabowski et al., 2013). Even though the quality of care is a fundamental aspect of the service provided by nursing homes, another crucial variable are prices. Our paper adds to this literature by studying how prices and the supply of beds are affected in equilibrium by changes in demand driven by aging and how supply-side variables and policies can mediate this effect.

The rest of the paper is structured as follows. Section 2 describes the data. Section 3 shows the empirical strategy and introduces the population accounting equation. Section 4 presents the main results and discusses their implications for the national increase in nursing home prices. Section 5 studies the mechanism behind the increase in prices. Section 6 concludes.

2 Data

We combine data from several sources. First, we use data from the Census Bureau's Population Estimates Program (PEP) as estimates of the population for five-year age groups at the county and metropolitan levels. This data is available through the Surveillance, Epidemiology, and End Results Program (SEER) for 1969-2020 (National Cancer Institute, DCCPS, Surveillance Research Program, 2022). This is the data that we use to build our independent variable of interest, the number of elderly people, and instrument, demographic composition 40 years in the past. Related to this data, we use data on migration, deaths, and births at the county level from the Applied Population Laboratory of the University of Wisconsin-Madison for the period 1950-2020.

Second, we use data on nursing home facilities from Certification And Survey Provider Enhanced Reporting (CASPER), Minimum Data Set (MDS), and Residential History File (RHF) on the number of beds and occupancy rate at the facility level. We also use data on the average total nurse hours per facility. These data are provided by LTCfocus for the years between 2000 and 2020, with some variation depending on the variable considered (LTCFocus, 2023).

Third, we use data from the Occupational Employment and Wage Statistics by the Bureau of Labor Statistics (Bureau of Labor Statistics, U.S. Department of Labor, 2023). Aside from information on employment, these data provide statistics of the wage distribution (e.g., median and average) for different occupation categories, defined by the Office of Management and Budget's 2018 Standard Occupational Classification (SOC) system, at the Metropolitan Statistical Areas level. We use data from 1996 to 2020.

Fourth, we use data on long-term care prices from Genworth's Cost of Care Survey, obtained through web scraping. This survey is conducted annually by reaching out to LTC providers and asking for their prices. For example, the 2021 survey contacted 67,742 providers by phone to complete 14,698 surveys of nursing homes, assisted living facilities, adult day health facilities, and home care providers, asking them about their prices for six services: private room, semi-private room, assisted living facility, adult day health care, home health aide, and homemaker services. The responses are aggregated at the regional level, with regions defined largely based on MSAs.[2] For simplicity, we will refer to these regions as MSAs. We have access to these data for the period between 2015 and 2021[3] and

[2]The main difference is that, for each state, some MSAs are aggregated in a single unit called "Rest of State," which we do not use in our analysis.

[3]For our back-of-the-envelope calculations, we could access the 2010 report and use aggregated data.

focus on the prices of private rooms and semi-private rooms, which are more directly related to nursing homes. To match the rest of our data, we use data from 2015 to 2020.

Fifth, to build socioeconomic controls at each geographic level, we use data from the Applied Population Laboratory of the University of Wisconsin from the project called *Net Migration Patterns for US Counties* (Egan-Robertson et al., 2023). This database provides not only population data but also socioeconomic variables at the county and MSA levels. We select our controls using data from 2000 to be sure that these variables are not affected by the variation in population aging that we are using in our estimations.

Finally, we complement our controls with local health data from the University of Wisconsin Population Health Institute's County Health Rankings & Roadmaps (CHR&R) from 2010 to 2020 (University of Wisconsin Population Health Institute, 2024). Specifically, we use this dataset information on the share of each county surveyed population reporting poor health and the years of potential life lost (YPLL). For our MSA-level results, we aggregate the share of the local population reporting poor health using the SEER population data as weights[4].

Regarding sample selection, we discard the five states that currently receive the largest number of elderly people: Florida, Arizona, Texas, North Carolina, and South Carolina. In these cases, there is a substantial variation in the elderly population that should not be correlated with our instrument but may generate variation in outcomes in which we are interested. Therefore, excluding these states should, in principle, give us more statistical power[5].

Unfortunately, we do not have access to the data disaggregated by region for years before 2015.

[4]When using the county-level data, we need to combine some counties to match the SEER data. However, we do not have local population data for the 2010 to 2020 period matching the county division the CHR&R data uses. We then aggregate the share of the population reporting poor health in these combined counties using the data we have for 2000. Although this might generate some errors, these should be minimal considering the number of counties for which this is relevant is low. Notice also that this is not an issue for YPLL, since this measure is a sum, not an average.

[5]In practice, all our main results remain if we do not exclude these states.

3 Empirical Strategy

The goal is to understand how outcomes of the nursing home industry are affected when the local elderly population grows — in other words, how the nursing home industry responds to an increase in demand. Hence, our main independent variable is the number of people 75 years old or above each year at the relevant geographic unit, county, or MSA. Naturally, variations in the elderly population in a given area tend to be correlated with a variety of economic forces that also affect the nursing home industry. Therefore, we rely on an instrumental variable strategy to estimate the causal effect of increases in the elderly population over nursing home industry outcomes. Namely, our instrument is constructed based on the demographic composition of that geographic unit in the distant past, 40 years prior. We compute the number of people between 35 and 54 years old who lived in that geographic unit 40 years ago. We argue that this instrument variable is relevant and exogenous to our dependent variables.[6]

The relevance of our instrument comes from the population accounting identity. This identity states that the number of elderly people who live today in a geographic unit is given by the number of people in the relevant age group who lived in the same area in the past minus the net out-migration and minus the number of deaths over time of that age group. In other words, and to make the point clearer with an example, the identity states that the number of people who are at least 75 years old in a given location is equal to the number of people who were living in the same area 40 years prior and were at least 35 years old, minus the net migration out of the region over the last 40 years of people who were at least 35 years old at that point in time, minus the deaths of residents in the same area and in the same age group that occurred during the prior 40 years. We find that our first stage is strong and robust to different specifications.

We also need to assume that our instrument is exogenous — i.e., it should not affect nursing home industry outcomes in the present through channels other than the current elderly population. We argue this is true because of a combination of two reasons. First, the number of young people who lived in a given area in the past is affected by the location's

[6] We use an age upper bound for the instrument to guarantee that this variation is not being driven by older people who are retiring and could have considered using long-term care.

socioeconomic characteristics. However, to affect the current state of the nursing home industry, these effects would have to be strongly persistent over time to affect outcomes some 40 years later. The second point is that if any of these characteristics are persistent enough to survive such a long time, they are likely akin to fixed characteristics and, hence, absorbed by fixed effects. In any case, we later explain our instrument in more detail and provide some suggestive evidence that the exogeneity assumption is valid.

We then use two-stages least squares combined with fixed effects to estimate the effects of an increase in the elderly population over local nursing home industry outcomes. Since we are using the demographic composition 40 years prior as an instrument, the first stage is:

$$ln(75^{+}_{gt}) = \tilde{\alpha}_g + \tilde{\alpha}_t + \tilde{\alpha}_1 ln([35\_55]_{g(t-40)}) + \epsilon_{gt} \tag{1}$$

Here, 75^{+}_{gt} is the number of people 75 years old or more in geography unit g at year t; $[35\_54]_{g(t-40)}$ is the number of people between 35 and 54 years old in geographic unit g 40 years before year t; $\tilde{\alpha_g}$ and $\tilde{\alpha_t}$ are geographic unit and year fixed effects, respectively.

The second stage is:

$$ln(y_{gt}) = \alpha_t + \alpha_g + \gamma \widehat{ln(75^{+}_{gt})} + \varepsilon_{gt} \tag{2}$$

In the equation above, y_{gt} is the dependent variable (beds, occupancy, prices, nurses' wages, and employment), and $\widehat{ln(75^{+}_{gt})}$ is the log number of elderly people in the region g at year t predicted by the first stage. We also control for geographic unit and year fixed effects, respectively α_g and α_t, and $epsilon_{gt}$ and $\varepsilon_{g,t}$ are the errors.

We also consider long-difference specifications using the first and last years available, in which we can include our socioeconomic controls at the local level for 2000. In this case, the first stage is

$$\Delta ln(75^{+}_{g}) = \tilde{\eta} \Delta ln([35\_55]_g) + \tilde{\beta} X_g + \tilde{\zeta}_g \tag{3}$$

Here, X_{gt} are socioeconomic controls from 2000, and $\Delta ln([35\_55]_g)$ is the difference of the logarithm of the local population in the relevant age range 40 years ago. We also include as controls two health measures, measured contemporaneously: years of potential life lost (YPLL) and the share of the population reporting poor health. We include these two measures to account for the possibility that our instrument is correlated with groups of local elderly populations who are healthier or sicker than the average[7].

The second stage is given by

$$\Delta ln(y_g) = \eta \Delta \widehat{ln(75_g^+)} + \beta X_g + \zeta_g \tag{4}$$

3.1 Population Dynamics

At the core of our identification is the evolution of the age distribution in a geographic zone. In this section, we formalize the age composition evolution of the population in a given area and show how we can provide suggestive evidence that the exogeneity assumption is satisfied.

Following the population accounting equation, the number $N_{A,t}$ of people of age A who live in a certain area[8] in year t is defined by:

$$N_{A,t} = b_{t-A} + \sum_{\tau=t-A}^{t} m_{\tau-(t-A),\tau} + \sum_{\tau=t-A}^{t} d_{\tau-(t-A),\tau} \tag{5}$$

where b_s is the number of births in year s, $m_{a,s}$ is net out-migration of people aged a at year s, and $d_{a,s}$ is the number of deaths at year s of people of age a. We can rewrite the right-hand side of the equation (5) as follows, for any k:

$$N_{A,t} = \underbrace{b_{t-A} + \sum_{\tau=t-A}^{t-k} m_{\tau-(t-A),\tau} + \sum_{\tau=t-A}^{t-k} d_{\tau-(t-A),\tau}}_{=N_{A-k,t-k}} + \sum_{\tau=t-k+1}^{t} m_{\tau-(t-A),\tau} + \sum_{\tau=t-k+1}^{t} d_{\tau-(t-A),\tau}$$

$$N_{A,t} = N_{A-k,t-k} + \sum_{\tau=t-k+1}^{t} m_{\tau-(t-A),\tau} + \sum_{\tau=t-k+1}^{t} d_{\tau-(t-A),\tau} \tag{6}$$

In other words, we rewrote $N_{A,t}$ as a function of the number of people who, k years in the past, were aged $A-k$, plus the out-migration and death dynamics since $t-k$. In the context

[7] We experimented with a set of controls excluding these health measures in previous versions of the paper and the point estimates remain remarkably similar.

[8] We suppress the geographic unit index for ease of exposition.

of our empirical strategy, $N_{A-k,t-k}$ is the instrument we use. Our preferred choice of k is $k = 40$, which is as far back as we can go with the SEER population data. With equation (6), we can better understand our instrument. The relevance condition holds if the migration and deaths of the relevant age groups in the last 40 years are not large enough relative to the number of people of that age group 40 years ago. Empirically, the relevance condition holds and our instrument is strong. Figure shows, at the county and MSA level, the relationship between the percent change in the population 35-55 years old between 1960-1970 and that of the population 75-95 years old between 2010-2020. The relationship is strong, as indicated by the Kleibergen-Paap rk Wald F statistic.

Equation (6) also sheds light on the plausibility of the exogeneity condition. First, it is important to notice that our instrument is exogenous to the last 40 years of migration and deaths at the local level. Second, our instrument is defined by the number of births between 75 and 95 years ago plus the accumulated migration and deaths of the relevant age groups between 95 and 40 years ago. Therefore, our identifying assumption is that the dynamics of births, migration, and death in the distant past do not affect the dynamics of the nursing home industry today through channels other than the elderly population today. We can test this assumption.

We have limited data on the births, migration, and deaths at the local level in the distant past. However, in the data from the Applied Population Laboratory of the University of Wisconsin (Egan-Robertson et al., 2023), we observe these variables 50 years ago at the county level. Since these migration rates are calculated for each decade (and are not annual), we focus on long differences[9]. We use these data to test that our identifying assumption holds for what happened 50 years ago. Then, under the assumption that the effect of migration, deaths, and births on variables today gets weaker as we move further back — e.g. the dynamics of migration 60 years ago are less relevant for today than those of 50 years ago —, the exogeneity assumption holds for our instrument defined using $k = 40$ years ago.

In Tables A.3, A.1, and A.2, we show the effect that the dynamics of migration, deaths, and births 50 years prior have on the current dynamics of the number of beds, occupancy

[9] For migration, we assign the net migration in the 2000-2010 period to 2010 and the net migration in 2010-2020 to 2020.

rate, and nursing hours in the nursing home industry. We can see that the dynamics of migration, deaths, and births 50 years ago barely affect the dynamics of key nursing home industry outcomes, and in the rare cases where they do, it happens through the current elderly population. This rules out that these variables may affect the nursing home industry through channels independent of the current elderly population, lending more credibility to the exogeneity assumption.

To strengthen our point, we also explore in more detail the dynamics of migration, which, among the determinants of population dynamics, is the one that is more responsive to socioeconomic conditions. Specifically, we use past migration data to see the correlation between present and past migration. Figure 2 shows the serial correlation of net migration at the county level for the 30-60 and 65-75 age groups. We can see that migration is serially correlated, but this correlation vanishes after two decades. This pattern is the same across age groups — 30-60 and 65-75 years old — and time — the 2010s, 2000s, and 1990s decades. Thus, the dynamics of migration are not driving our results.

With these pieces of evidence supporting our instrument, we estimate the effect of population aging on the nursing home industry in the next section.

4 Results

4.1 Main Results

In this section, we show the paper's main results, namely, the effect of population aging on the demand for nursing homes and its consequences on nursing home bed prices. We rely on our IV approach as specified in equations (3) and (4), estimating them using two-stage least squares. The first stages are reported in Table 1[10].

In Tables 2 and 3, we show the estimated effects of population aging on the number of beds and occupancy rates of nursing facilities, respectively. We present the second stage results using counties and MSAs for the whole period of analysis, 2010 to 2020, both with

[10]Due to some missing data, the number of observations for different outcomes and specifications may vary conditional on the geographical unit used. Since none of these impact in the first stage in any meaningful way, we report only two specifications for each geographical unit, with and without controls. We report the Kleinbergen-Paap F statistic relative to each result in the tables with the main results.

and without controls. All our F statistics indicate that our instrument is strong. Also, our results are robust regarding the geographic definition. In our preferred specification, controlling for several socioeconomic characteristics at the MSA level, we get an elasticity of the number of beds to the elderly population of 0.240. We do not observe any meaningful effect on occupancy rates, which indicates that the management of excess capacity is not affected significantly.

As the demand for beds increases with the growth of the elderly population, it is reasonable to expect that facilities also hire more nurses in response to the same growth. Table 4 shows the estimated effect of population aging on the average total number of nurse hours per facility, where nurse hours include registered nurses and licensed and practical vocational nurses. We find that the number of nurse hours increases when the elderly population grows. In our preferred specification (column 4), the elasticity of the number of nurses to the number of elderly people at the MSA level equals 0.442. Our results so far are expected. Increases in the elderly population cause increases in the demand for long-term care. Hence, facilities increase their number of beds (and occupied beds) and hire more nurses to be able to provide services to more people.

Having established that the quantity of nursing home care increases when the elderly population increases, the next step of our analysis is to assess what happens with prices. Table 5 shows the effect that increases in the elderly population have on the prices of private and semi-private rooms in nursing home facilities. We find positive and significant elasticities for both types of rooms in all of our specifications, though our results are somewhat noisier since we have less data for prices. However, given the standard errors, the point estimates are reasonably stable across specifications. Our preferred specifications (columns 2 and 4) indicate that the elasticities of private and semi-private rooms to the size of the elderly population are, respectively, 0.493 and 0.367.

To provide further evidence that our instrument is capturing the effect of population aging on the nursing home industry, we estimate the impact that variations in the elderly population have on employment in other health industries. Figure 3 shows the point estimates for employment in nine health industries defined at the four-digit NAICS code. As expected, the estimates are larger for industries where the elderly demand is more relevant as a share of the whole demand. For example, we find that the elderly population does not affect the child-care industry, has a small effect on hospitals, and has a large effect on nursing homes and social services for the elderly.

4.2 The Missing Intercept: Cross-sectional and National Estimates

The estimation of the effect that population aging has on the price comes from cross-sectional variation. Thus, its interpretation of the consequences of the population aging at the national level is affected by *the missing intercept problem*. We argue that our findings are a lower bound for the national effects of population aging, and this lower bound has significant economic consequences.

To understand why the micro elasticity (cross-sectional) is a lower bound for the macro elasticity (national), we must consider factor mobility. Let us suppose that there are two identical regions, i and j, that the elderly population increased a lot in the region i while it stayed constant in the region j, and that the number of nurses is the only input factor and its national supply is fixed. There are two polar cases: (i) nurses can move between regions with no cost (perfect mobility) and (ii) nurses cannot move (immobility).

In the case of perfect mobility, the increase in the demand for nursing homes in region i will prompt nurses to move from region j to i. Since there are no switching costs, this migration of nurses will keep the wages and nursing home prices the same in both regions, although the overall level of both wages and prices is now higher. Therefore, the increase in the elderly population increases prices overall, but an estimation that relies on a cross-sectional comparison between these two regions would give us zero price elasticity.

In the case of perfect immobility, the increase in the demand for nursing homes in the region i will have no effect on the price of nursing homes of the region j and will be completely absorbed by the price of the region i. Thus, a cross-sectional estimation would find a positive price elasticity.

The real world is likely in between these two polar cases. There is some mobility of production factors (nurses and capital), but probably not perfect. A similar analysis can be done with the mobility across regions of elderly people. If the price of nursing homes goes up in region i, some elderlies will move to region j thus increasing the demand and price of nursing homes in that region. Both, the demand and supply spillovers, go in the same direction of decreasing the effect that local age variation has on prices. Thus, the missing intercept is positive and the cross-sectional estimate is smaller than the national estimate, providing a lower bound for the effect that population aging has on the price of long-term care. This is, because the Stable Unit of Treatment Assumption (SUTVA) provides a lower bound for the effect that population aging has on prices.

Note that the same analysis can not be done for the number of nurses and beds, where the demand and supply spillovers go in opposite directions, and the missing intercept sign is ambiguous. Hence, we cannot claim that our cross-sectional estimates are lower or upper bounds for the national estimates.

4.3 Two-Way Fixed Effects and Difference-in-Differences with Heterogeneous Treatment Effects

Our estimator can be interpreted as a difference-in-differences estimator with continuous treatment values, where the treatment is the (exogenous part of the) variation of the elderly population. As pointed out by De Chaisemartin and D'Haultfœuille (2020), when we use a two-way fixed effects estimator in this setting as we do, heterogeneous treatment effects may threaten the identification of our coefficients of interest if we are unwilling to make some strong assumptions about their homogeneity and stability.

We conduct tests with a slightly modified version of our instrument to assess how heterogeneous effects might affect our estimates, focusing on the reduced form. For the results on the number of beds and occupancy rates, we focus on the specification at the county level and estimate the model using our instrument as a continuous treatment. For the results at the MSA and regional level, we do not have enough observations to estimate the specification using a continuous treatment. Therefore, we discretize our instrument to generate a treatment as follows. We first calculate, for every unit and year, the difference between the current value of the instrument (the logarithm of the population in the relevant age range 40 years prior) and the value at the baseline year, 2010 (except for prices, since these data start in 2015). Let σ denote the overall standard deviation of this calculated difference. Then, we define our discretized instrument in the following way: (i) equal to 0 if $|instr_{it} - instr_{i,2010}| \leq 0.1\sigma$; (ii) equal to 1 if $instr_{it} - instr_{i,2010} > 0.1\sigma$; (iii) and -1 if $instr_{it} - instr_{i,2010} < -0.1\sigma$. We report results using our specification (TWFE) and the heterogeneous effects estimator proposed by De Chaisemartin and D'Haultfœuille (2020). For the heterogeneous difference-in-differences estimator, we show the average cumulative effect per unit, estimating the effects until ten years after treatment, also reporting the average number of periods for which effects are accumulated.

Results are shown in Tables 6 and 7. Overall, the estimates using the two-way fixed effects and the heterogeneous difference-in-differences estimators are very similar. This indicates that our main estimates are not significantly affected by the so-called "forbidden comparisons." It is also important to notice that, although the results using the discretized instrument lack significance, we do not put much weight on this; the discretization procedure naturally reduces the statistical power, and hence we focus on the comparison between the point estimates.

4.4 Economic Interpretation of the Results

We now turn to the economic interpretation of our results. First, the effect of population aging on nursing home bed prices is economically relevant. Moreover, our results show that the supply of beds is far from being perfectly elastic, leaning towards being inelastic depending on the specification considered. [11]

As discussed above, our estimate for the elasticity of prices to the population aged 75 years and more is a lower bound. Thus, using our preferred coefficient of 0.493 in a back-of-the-envelope calculation, we get that, depending on which type of room we consider, 41% or 48% of the increase in nursing home prices of the last decade can be explained by the change

[11] A back-of-the-envelope calculation of the price elasticity of supply would be as follows. Let Q be the number of occupied beds, p the price of a bed, and N the elderly population. Then, $\varepsilon_p = \frac{d\log p}{d\log Q} = \frac{d\log p}{d\log N}\frac{d\log N}{d\log Q}$. Using our estimates for number of beds at the MSA level, for the price of private rooms, and that $\frac{d\log N}{d\log Q} = \left(\frac{d\log Q}{d\log N}\right)^{-1}$, $\varepsilon_p = 0.493/0.240 = 2.054$.

in the population's age distribution. Extrapolating for the next decade, population aging alone can further increase real prices of beds by 25% (private) and 18.6% (semi-private).

A last interesting point to notice regards the difference between short- versus long-term responses. In principle, our results look at the contemporaneous effects of increases in the elderly population on our outcomes of interest. Therefore, one can be tempted to interpret our estimates as pure short-term responses. This would be relevant if we consider that it should be easier to change prices than increase capacity and, thus, the response of prices should be lower in the medium- and long-run. However, the variation of the elderly population we use to estimate the effects is based on past population changes, which should be easily predictable by nursing homes. Therefore, inasmuch as managers are sophisticated enough to use this past demographic information in their decisions, our estimates should reflect more long-run effects than short-term, transitory effects.

5 Mechanisms

In this section, we explore the mechanism behind the price increase driven by increases in the elderly population. We explore how costs and regulation affect the response of outcomes to population aging. Our findings support that regulation, and not increases in costs, is the main driver of price increases in the nursing home industry in response to the increases in demand.

5.1 Costs: Nurses Wages

To assess the role of costs in explaining price increases, we focus on wages paid by the nursing home industry. The rationale for this is that payroll is the most relevant cost for nursing homes, representing more than 85% of the average facility's total cost (Dalal, 2023). We study the effect that the increase in demand driven by population aging had on the wages at nursing home facilities, using data from the US Census Bureau's County Business Patterns. It is important to notice that nurses' wages represent more than 90% of nursing homes' total payroll cost. We use the same two-stage estimation strategy defined in section 3 to estimate the effect of population aging on the wages of nursing home facilities and nurses.

Table 8 shows the results at the industry and occupation level. We find that neither the average wage in nursing facilities nor nurses' wages increase when the local elderly population grows. One potential interpretation of this result is that nurses are mobile; therefore, the cross-section estimation would be zero even with the national estimation being positive. However, this interpretation is not consistent with the data since the national average wage of the nursing home industry has decreased in real terms during the last decade, while the number of people aged 75 and more increased by 15%.

Another explanation for the unresponsiveness of wages to the increase in demand is that this industry is regulated through Medicaid's reimbursement conditions. We test this by checking the effects' heterogeneity based on pass-through laws' existence. These laws are adopted at the state level, allowing nursing facilities to reimburse nurse wage increases. Figure 4 shows which states and when they adopted this law. To incorporate this in the analysis, we run a modified version of equations (3) and (4), adding pass-through and interaction-independent variables that change over time and across states. The results number of beds, prices of private rooms, and wages and employment at the industry level are in Table 9[12]. Based on our results, the presence of pass-through laws did not make average wages more responsive to the increase in the elderly population, nor did it affect the effect of population aging on employment or prices. The laws do appear to make the number of beds more responsive to population aging, but this effect is not quantitatively meaningful. These results suggest that nursing shortages due to low wages in the sector cannot explain the increases in prices in response to population aging.

5.2 Regulation: Certificate of Need Laws

Several states have Certificate of Need (CON) laws, which are designed to ensure a minimum level of quality of care but, in turn, restrict the entry and expansion of nursing homes. These laws were implemented between 1972 and 1990 to limit increases in demand for long-term care that could be deemed unnecessary and thus curb the spending on Medicaid reimbursements that would be considered wasteful (National Conference of State Legislatures

[12] We also present a version of the regression with pass-through laws and the other mechanism we explore at the same time. None of the results change in these specifications.

(2023))[13]. Figure 5 shows a map identifying states with and without CON laws. We leverage the fact that these laws were implemented decades before the time frame we study and use this variation on the existence of CON laws at the state level to assess heterogeneity in the effects we find based on the existence or absence of CON laws. The rationale is that CON laws should restrict supply, making it more inelastic. Therefore, we should expect to see the number of beds (quantities) being less responsive to variations in the local elderly population in states with CON laws. In contrast, prices should be more responsive in these states.

In Table 9, we also report the effects of CON laws on the supply of beds and prices. We can see that the bed supply is significantly less responsive in states with CON laws. At the same time, prices are significantly more responsive to variations in the elderly population in states with CON laws – in fact, our estimates suggest that states with these laws drive our results on prices. Interestingly, CON laws do not seem to affect employment or average wages in the nursing home industry. In sum, our results suggest that restrictions on supply expansion are a key source of the large elasticity of nursing home prices to an increase in the demand driven by population aging.

6 Conclusion

Population aging is a widespread phenomenon affecting almost every country in the world. Despite the many advances we have seen in medicine, it is still true that the demand for healthcare increases with age. Therefore, the increase in the share of the elderly population entails a significant increase in the demand for healthcare. Since the healthcare needs of the elderly population are not homogenous across different types of care, some areas tend to be more affected than others. One of the most important and affected of these areas is long-term care and nursing homes in particular. Thus, a natural question is to understand how this industry is affected by the increase in the elderly population.

[13] *"A primary objective of state CON laws is to control health care costs by avoiding unnecessary expansion or duplicative services within an area. Proponents of CON laws contend that if excess healthcare service capacity exists, price inflation may occur to compensate for new, underused healthcare services or empty beds. Some states also require CON approval for health facility mergers, acquisitions or changes in ownership, which many researchers say contributes to higher health care costs and prices." (National Conference of State Legislatures (2023), p. 1)*

In this paper, we investigate the effects of population aging on the US nursing home industry. Our results show that increases in the elderly population significantly affect the long-term care industry. Although this increase in demand does induce facilities to expand their capacity, we also find a significant increase in prices for private and semi-private beds in long-term care facilities, which indicates a supply far from perfect elasticity. We also study what can explain this effect on prices. Contrary to some arguments made in the public debate, we do not find cost pressure coming from nurse salary increases, which puts into question the effectiveness of pass-through laws to some extent. Instead, we find that a lack of competition and consequent concentration of market power on incumbents is a key driver of prices, particularly in states with CON laws.

Our results show the importance of understanding the market structure and the role of market power in understanding the recent trends in this market and how it is likely to evolve in the near future. If policymakers want to control the increases in prices and expenses that are likely to occur with the increase in the elderly population share, it is paramount to study market design and policies that can curb market power in the nursing home industry. This is a challenging but promising avenue for future research since there are other aspects that we do not study in this paper but are of substantial importance to the market, like quality of care and its interaction with market power.

References

Acemoglu, Daron and Pascual Restrepo, "Secular Stagnation? The Effect of Aging on Economic Growth in the Age of Automation," *American Economic Review: Papers Proceedings*, 2017, *107* (5), 174–179.

Aksoy, Yunus, Henrique S. Basso, Ron P. Smith, and Tobias Grasl, "Demographic Structure and Macroeconomic Trends," *American Economic Journal: Macroeconomics*, 2019, *11* (1), 193–222.

Autor, David H and David Dorn, "The Growth of Low-Skill Service Jobs and the Polarization of the US Labor Market," *American Economic Review*, 2013, *103* (5), 1553–

1597.

Balia, Silvia and Rinaldo Brau, "A country for old men? Long-term home care utilization in Europe," *Health Economics*, 2014, *23* (10), 1185–1212.

Braun, Robert Tyler, Hye-Young Jung, Lawrence P. Casalino, Zachary Myslinski, and Mark Aaron Unruh, "Association of Private Equity Investment in US Nursing Homes With the Quality and Cost of Care for Long-Stay Residents," *JAMA Health Forum*, 11 2021, *2* (11), e213817–e213817.

_ , Hyunkyung Yun, Lawrence P. Casalino, Zachary Myslinski, Farai M. Kuwonza, Hye-Young Jung, and Mark Aaron Unruh, "Comparative Performance of Private Equity–Owned US Nursing Homes During the COVID-19 Pandemic," *JAMA Network Open*, 10 2020, *3* (10), e2026702–e2026702.

Bureau of Labor Statistics, U.S. Department of Labor, "Occupational Employment and Wage Statistics," 2023.

Butler, Alexander W and Hanyi Yi, "Aging and public financing costs: Evidence from US municipal bond markets," *Journal of Public Economics*, 2022, *211*, 104665.

Census Bureau, "County Business Patterns," 2023. Available at `https://www.census.gov/programs-surveys/cbp.html`. Accessed: 2023-04-04.

Chaisemartin, Clément De and Xavier D'Haultfœuille, "Two-way fixed effects estimators with heterogeneous treatment effects," *American Economic Review*, 2020, *110* (9), 2964–96.

Ching, Andrew T, Fumiko Hayashi, and Hui Wang, "Quantifying the impacts of limited supply: The case of nursing homes," *International Economic Review*, 2015, *56* (4), 1291–1322.

Chou, Shin-Yi, "Asymmetric information, ownership and quality of care: an empirical analysis of nursing homes," *Journal of Health Economics*, 2002, *21* (2), 293–311.

Collelo, Kirsten J. and Isobel Sorenson, "Who Pays for Long-Term Services and Support," Technical Report, Congressional Research Service 2023.

Comondore, Vikram R, P J Devereaux, Qi Zhou, Samuel B Stone, Jason W Busse, Nikila C Ravindran, Karen E Burns, Ted Haines, Bernadette Stringer, Deborah J Cook, Stephen D Walter, Terrence Sullivan, Otavio Berwanger, Mohit Bhandari, Sarfaraz Banglawala, John N Lavis, Brad Petrisor, Holger Schünemann, Katie Walsh, Neera Bhatnagar, and Gordon H Guyatt, "Quality of care in for-profit and not-for-profit nursing homes: systematic review and meta-analysis," *BMJ*, 2009, *339.*

Cravino, Javier, Andrei Levchenko, , and Marco Rojas, "Population Aging and Structural Transformation," *American Economic Journal: Macroeconomics*, 2022, *14* (4), 479–498.

Dalal, Michal, "US Industry (NAICS) Report 62331 / Healthcare and Social Assistance: Retirement Communities in the US," 2023.

de la Maisonneuve, Christine and Joaquim Oliveira Martins, "The future of health and long-term care spending," *OECD Journal: Economic Studies*, 2015, *2014* (1), 61–96.

Eckert, Fabian, Teresa C. Fort, Peter K. Schott, and Natalie J. Yang, "Imputing Missing Values in the US Census Bureau's County Business Patterns," Technical Report, National Bureau of Economic Research 2021.

Egan-Robertson, David, Katherine J. Curtis, Richelle L. Winkler, Kenneth M. Johnson, and Caitlin Bourbeau, "Age-Specific Net Migration Estimates for US Counties, 1950-2020," *Applied Population Laboratory, University of Wisconsin - Madison*, 2023. Beta release.

Eggertsson, Gauti B., Manuel Lancastre, and Lawrence H. Summers, "Aging, Output Per Capita, and Secular Stagnation," *American Economic Review: Insights*, 2019, *1* (3), 325–342.

Elmendorf, Douglas W. and Louise M. Sheiner, "Federal Budget Policy with an Aging Population and Persistently Low Interest Rates," *Journal of Economic Perspectives*, 2017, *31* (3), 174–194.

Ferraro, Domenico and Giuseppe Fiori, "The Aging of the Baby Boomers: Demographics and Propagation of Tax Shocks," *American Economic Journal: Macroeconomics*, 2020, *12* (2), 167–193.

Foster, Andrew D. and Yong Suk Lee, "Staffing subsidies and the quality of care in nursing homes," *Journal of Health Economics*, 2015, *41*, 133–147.

Galor, Oded, "The Demographic Transition: Causes and Consequences," *Cliometrica*, 2012, *6* (1), 1–28.

Grabowski, David C., "Medicaid reimbursement and the quality of nursing home care," *Journal of Health Economics*, 2001, *20*, 549–569.

_ **and Richard A. Hirth**, "Competitive spillovers across non-profit and for-profit nursing homes," *Journal of Health Economics*, 2003, *22*, 1–22.

_ **, Zhanlian Feng, Richard A. Hirth, Momotazur Rahman, and Vincent Mor**, "Effect of nursing home ownership on the quality of post-acute care: An instrumental variables approach," *Journal of Health Economics*, 2013, *32*, 12–21.

Hackmann, Martin B, "Incentivizing Better Quality of Care: The Role of Medicaid and Competition in the Nursing Home Industry," *American Economic Review*, 2019, *109* (5), 1684–1716.

Karlsson, Martin, Les Mayhew, Robert Plumb, and Ben Rickayzen, "Future costs for long-term care: Cost projections for long-term care for older people in the United Kingdom," *Health Policy*, 2006, *75* (2), 187–213.

Lee, Ronald, "The Demographic Transition: Three Centuries of Fundamental Change," *Journal of Economic Perspectives*, 2002, *17* (4), 167–190.

Lee, Ronald D., "Macroeconomic Consequences of Population Aging in the United States: Overview of a National Academy Report," *American Economic Review: Papers Proceedings*, 2014, *104* (5), 234–239.

Lendon, Jessica P., Christine Caffrey, Amanuel Melekin, Priyanka Singh, Zhaohui Lu, and Manisha Sengupta, "Overview of Post-acute and Long-term Care Providers and Services Users in the United States, 2020," Technical Report, National Center for Health Statistics – CDC 2023.

Lin, Haizhen, "Revisiting the relationship between nurse staffing and quality of care in nursing homes: An instrumental variables approach," *Journal of Health Economics*, 2015, *37*, 13–24.

LTCFocus, "LTCFocus Public Use Data sponsored by the National Institute on Aging (P01 AG027296) through a cooperative agreement with the Brown University School of Public Health.," 2023.

Maestas, Nicole, Kathleen J. Mullen, and David Powell, "The Effect of Population Aging on Economic Growth, the Labor Force, and Productivity," *American Economic Journal: Macroeconomics*, 2023, *15* (2), 306–332.

Martin, Anne B., Micah Hartman, Joseph Benson, Aaron Catlin, and The National Health Expenditure Accounts Team, "National Health Care Spending In 2021: Decline In Federal Spending Outweighs Greater Use Of Health Care," *Health Affairs*, 1 2023, *42* (1), 6–17.

Matsudaira, Jordan D., "Monopsony in the Low-Wage Labor Market? Evidence from Minimum Nurse Staffing Regulations," *Review of Economics and Statistics*, 2014, *96* (1), 92–102.

Meijer, Claudine De, Marc Koopmanschap, Teresa Bago d'Uva, and Eddy Van Doorslaer, "Determinants of long-term care spending: age, time to death or disability?," *Journal of Health Economics*, 2011, *30* (2), 425–438.

National Cancer Institute, DCCPS, Surveillance Research Program, "U.S. Population Data 1969-2020," 2022.

National Conference of State Legislatures, "Certificate of Need State Laws," 2023.

OECD, "Main Findings from the 2020 Risks that Matter Survey," 2021.

Penn Wharton Budget Model, "Projecting Medicaid's Long Term Care Expenditures," https://budgetmodel.wharton.upenn.edu/issues/2022/3/9/projecting-medicaids-long-term-care-expenditures 2022. Accessed: 2023-08-07.

Razin, Assaf, Efraim Sadka, and Phillip Swagel, "The Aging Population and the Size of the Welfare State," *Journal of Political Economy*, 2002, *110* (4), 900–918.

Semega, Jessica and Melissa Kollar, "Current Population Reports, P60-276, Income in the United States: 2021, U.S. Government Publishing Office," *Washington, DC*, 2022.

Shields, Michael A., "Addressing Nurse Shortages: What Can Policy Makers Learn from the Econometric Evidence on Nurse Labour Supply?," *Econometric Journal*, 2004, *114* (449), 464–498.

United Nations, Department of Economic and Social Affairs, Population Division, "World Population Ageing 2020 Highlights: Living arrangements of older persons," 2020.

_ , "World Population Prospects 2024, Online Edition (Data)," 2024.

_ , "World Population Prospects 2024: Ten Key Messages," 2024b.

University of Virginia Weldon Cooper Center, Demographics Research Group, "National Population Projections (2018)."

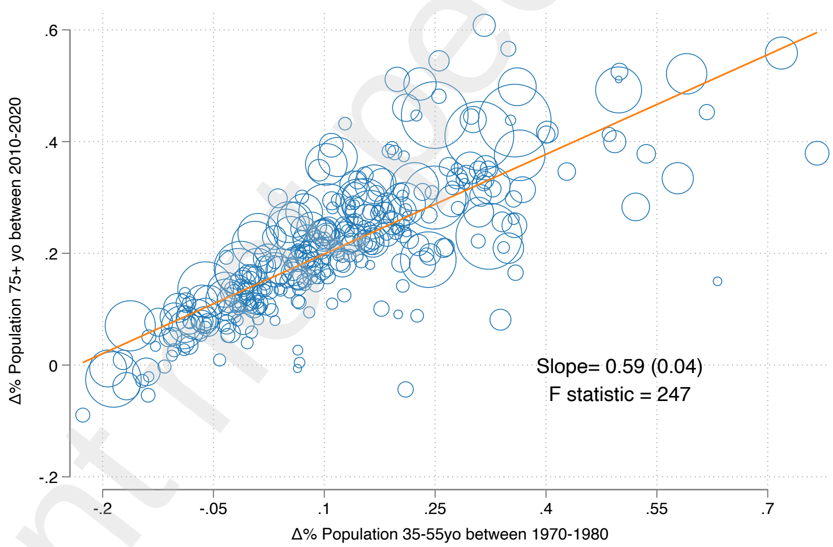
University of Wisconsin Population Health Institute, "County Health Rankings Roadmaps," 2024. Available at www.countyhealthrankings.org. Accessed: 2024-10-30.

World Health Organization, *World report on ageing and health*, World Health Organization, 2015.

Yearby, Ruqaiijah, Spencer Gardner Nathan adn Davis, Kate Tomey, Anat Goss, Maya Lawson, Josie Finch, Katelyn Garland, Laura Doud, Crystal Lewis, Fawn Pettet, Chereka Klutz, Hannah Clark, and Sara Reinoso, "Direct Care Worker Wage Pass-Through Law Final Report," *The Institute for Healing Justice and Equity*, 2020.

Figures

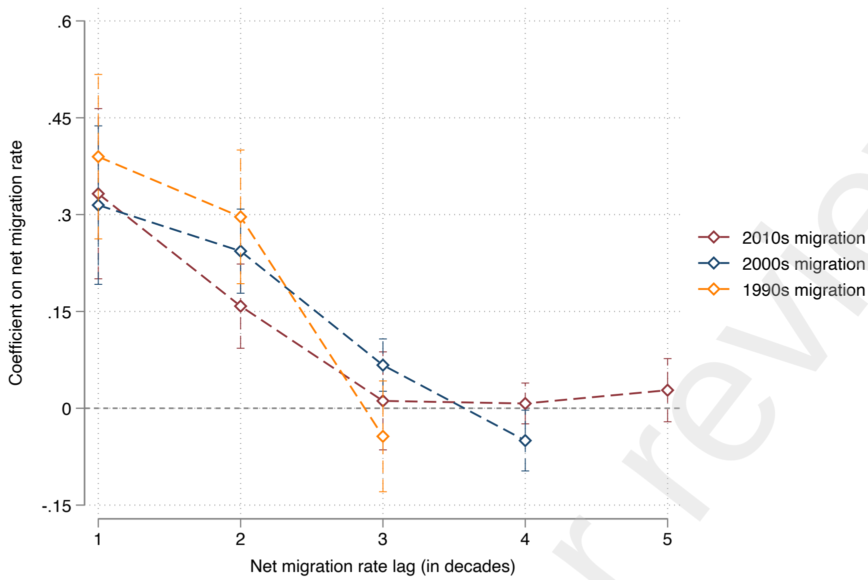
(a) Counties



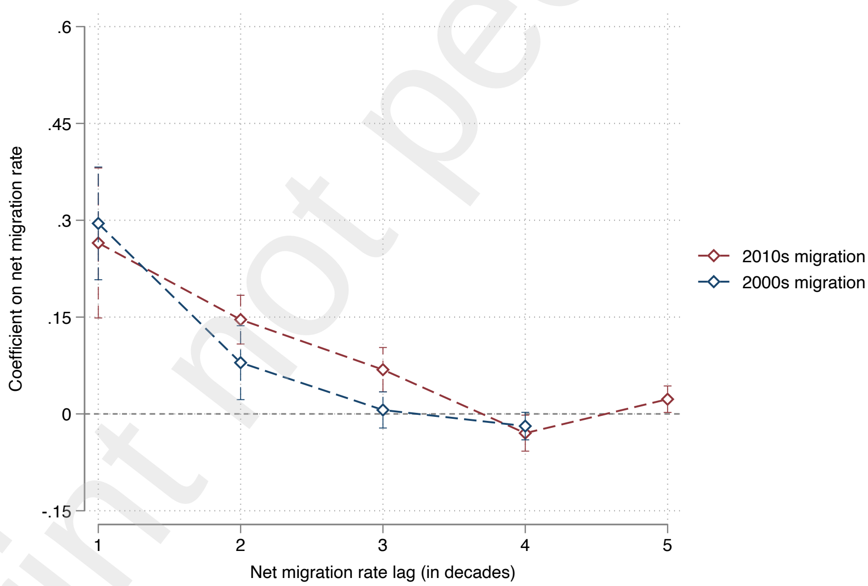
(b) MSAs

Figure 1: Relationship between population 40 years ago and population today

Note: This Figure shows the scatter and linear fit between the percentual change in the 35-55 year-old age group during the 1970s and the percentual change in the 75 and older age group size during the 2010s. The scatter size is defined by the elderly population size.



(a) 30-60 years old



(b) 65-75 years old

Figure 2: Autocorrelation of the net migration rate

Note: This figure shows the coefficient of regressing the 30-60-year-old net migration rate at the county level with its lags. The available data goes back to the migration during the 1960s decade for the 30-60-year-old age group and until the 1970s for the 65-75-year-old age group. Confidence intervals at 95% confidence with standard errors clustered at the state level.

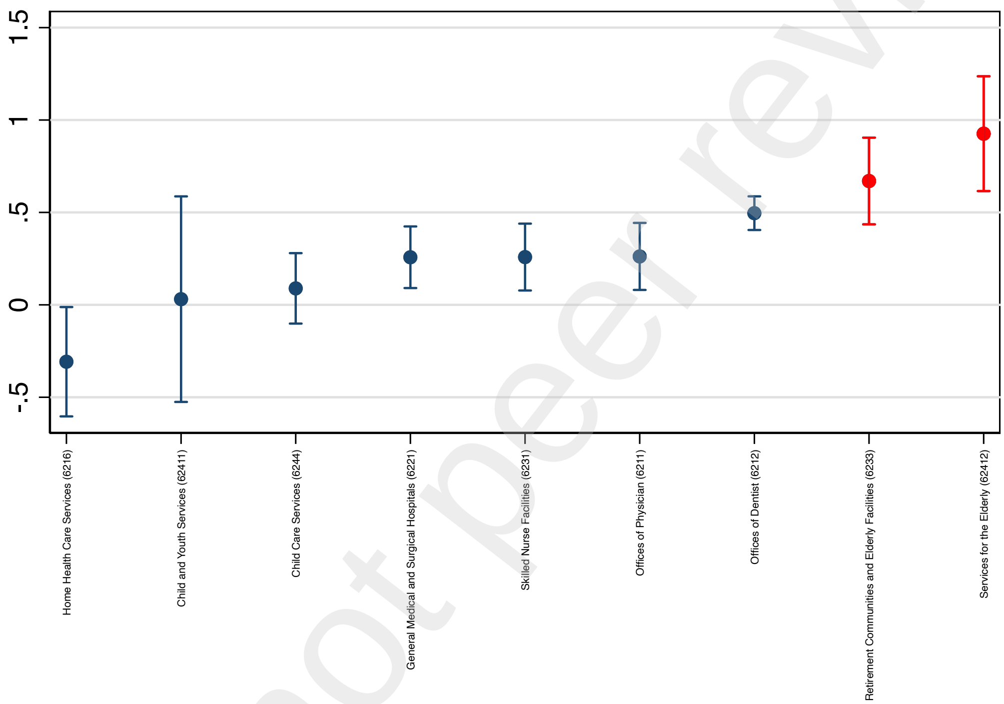


Figure 3: Population Aging and Employment in the Health Industry

Note: This figure shows the point estimate of γ of equation (4) using the long difference estimation instead of the fixed effects. The dependent variable is the log of employment for each industry in the health sector.

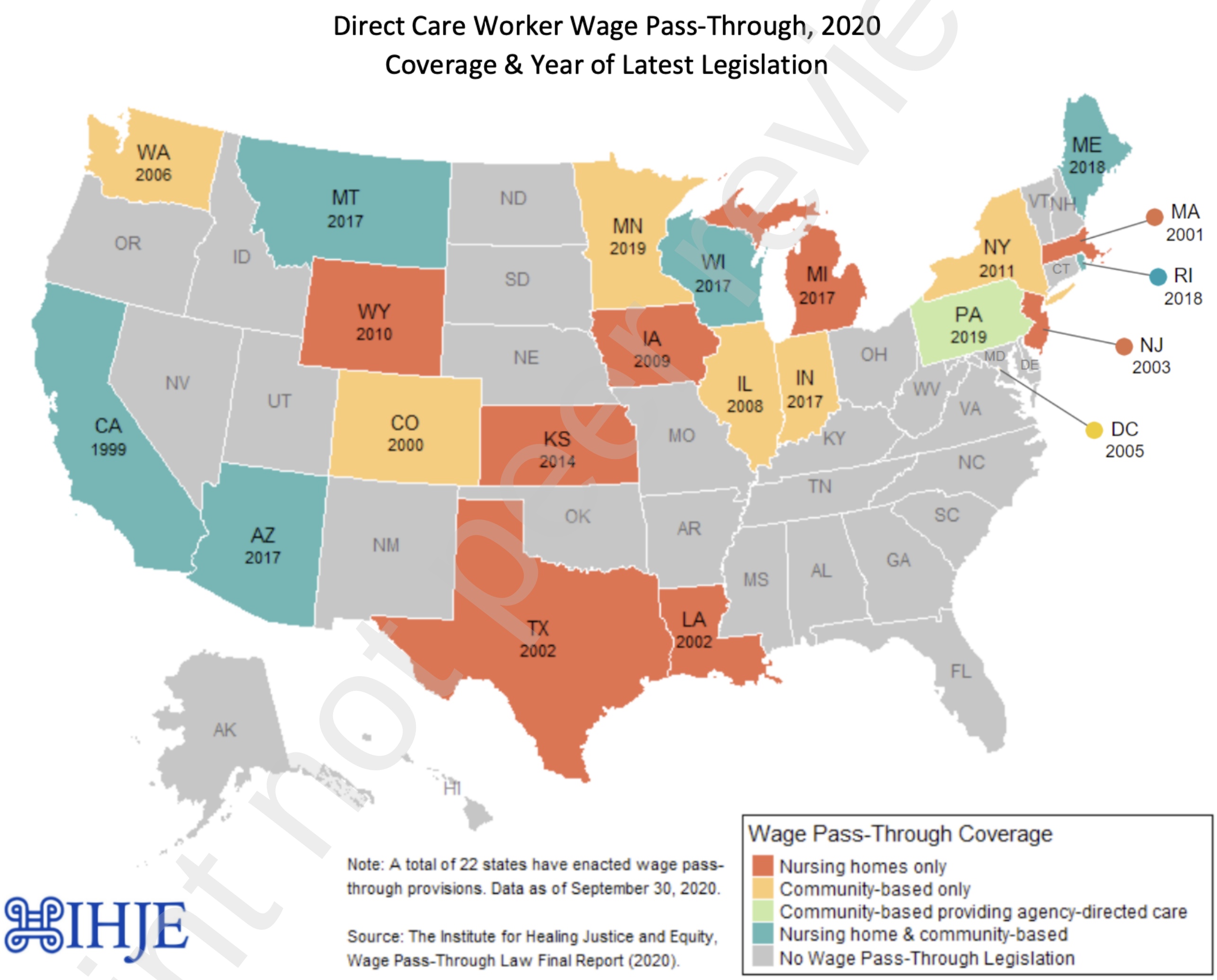


Figure 4: Pass-through regulations

Note: This figure comes from Yearby et al. (2020).

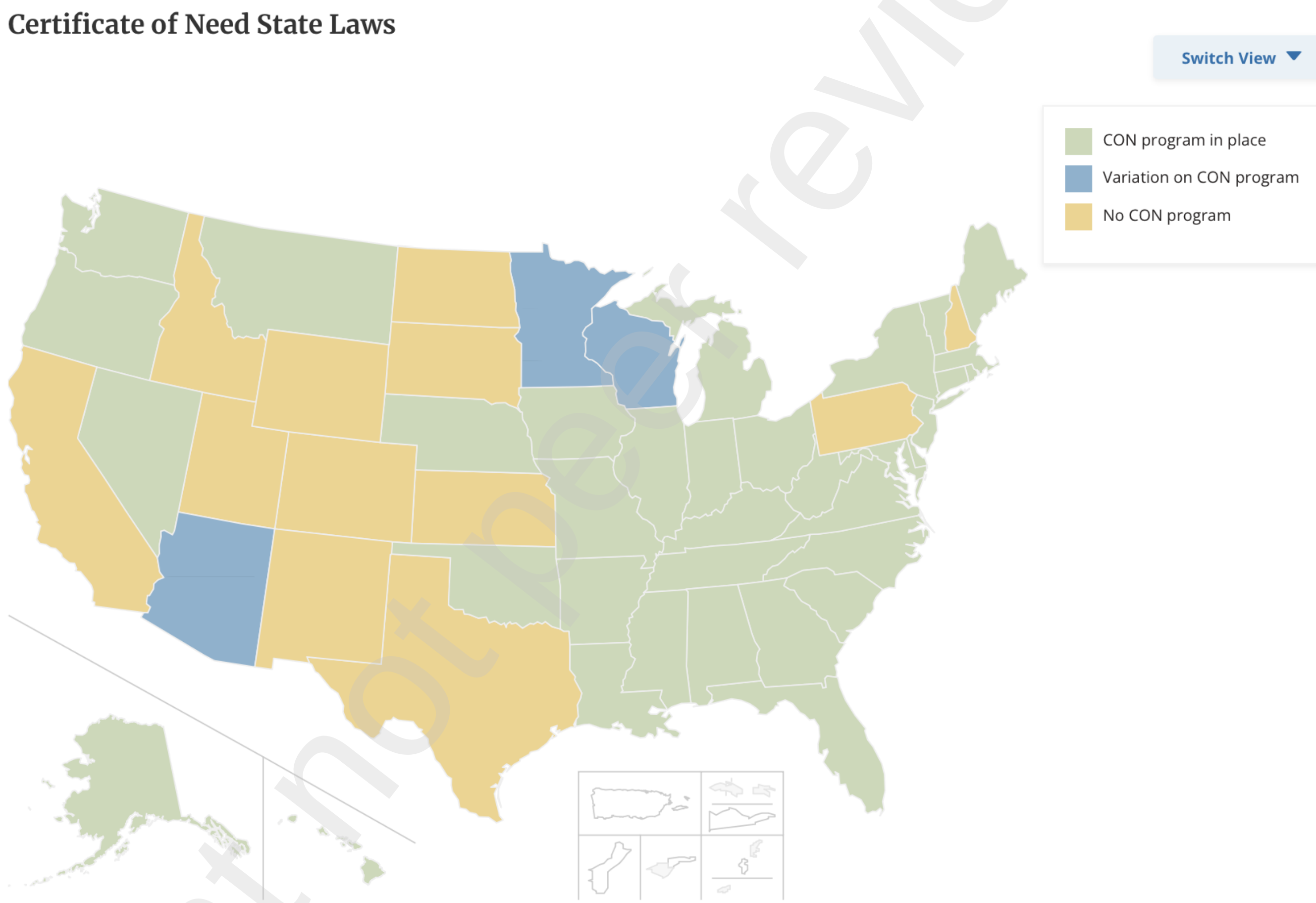


Figure 5: States with and without CON laws

Note: This figure comes from National Conference of State Legislatures (2023).

Tables

Table 1: IV, 1^{st} Stage

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(pop35-55_{t-40})$	0.601***	0.492***	0.525***	0.486***	0.491***	0.476***
	(0.016)	(0.015)	(0.036)	(0.035)	(0.032)	(0.031)
N	28,666	27,747	3,398	3,398	1,722	1,716
Geographic Unit	County	County	MSA	MSA	Region	Region
Years	2010-2020	2010-2020	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes	-	-
Year FE	Yes	Yes	Yes	Yes	-	-
Controls	No	Yes	No	Yes	No	Yes

Notes: Standard errors clustered at the unit level in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. The following socioeconomic characteristics in 2000 are interacted with year dummies and used as controls: share of male individuals, shares of population identifying as black, white, and hispanic, share of people with high school, share of people with bachelor degree, share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. We also control for years of potential life list (YPLL) and the share of people who report having poor health.

Table 2: Results on Log Number of Beds - IV, 2nd Stage

	(1)	(2)	(3)	(4)
log(*pop*75+)	0.377***	0.410***	0.251***	0.240**
	(0.043)	(0.055)	(0.093)	(0.103)
N	26,691	26,234	3,398	3,398
Number of Units	2,450	2,450	313	313
Geographic Unit	County	County	MSA	MSA
Years	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
F Statistic	1729.40	1144.44	213.72	194.04

Notes: Standard errors clustered at the unit level in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. The F statistic reported is the Kleibergen-Paap rk Wald F statistic. The following socioeconomic characteristics in 2000 are interacted with year dummies and used as controls: share of male individuals, shares of population identifying as black, white, and hispanic, share of people with high school, share of people with bachelor degree, share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. We also control for years of potential life list (YPLL) and the share of people who report having poor health.

Table 3: Results on Occupancy Rate - IV, 2nd Stage

	(1)	(2)	(3)	(4)
$\log(pop75+)$	0.006	-0.067*	0.053	-0.099
	(0.030)	(0.040)	(0.229)	(0.227)
N	26,691	26,234	3,398	3,398
Number of Units	2,450	2,450	313	313
Geographic Unit	County	County	MSA	MSA
Years	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
F Statistic	1729.40	1144.44	213.72	194.04

Notes: Standard errors clustered at the unit level in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. The F statistic reported is the Kleibergen-Paap rk Wald F statistic. The following socioeconomic characteristics in 2000 are interacted with year dummies and used as controls: share of male individuals, shares of population identifying as black, white, and hispanic, share of people with high school, share of people with bachelor degree, share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. We also control for years of potential life list (YPLL) and the share of people who report having poor health.

Table 4: Results on Log Number of Total Nurse Hours - IV, 2nd Stage

	(1)	(2)	(3)	(4)
$\log(pop75+)$	0.465***	0.606***	0.489***	0.442**
	(0.060)	(0.080)	(0.153)	(0.181)
N	25,716	25,716	3,389	3,389
Geographic Unit	County	County	MSA	MSA
Years	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
F Statistic	1719.38	1151.76	207.93	200.00

Notes: Standard errors clustered at the unit level in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. The F statistic reported is the Kleibergen-Paap rk Wald F statistic. The following socioeconomic characteristics in 2000 are interacted with year dummies and used as controls: share of male individuals, shares of population identifying as black, white, and hispanic, share of people with high school, share of people with bachelor degree, share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. We also control for years of potential life list (YPLL) and the share of people who report having poor health.

Table 5: Results on Log Prices - IV, 2nd Stage

	Private Room		Semi-private Room	
	(1)	(2)	(3)	(4)
$\log(pop75+)$	0.636***	0.493**	0.519***	0.367**
	(0.177)	(0.201)	(0.168)	(0.170)
N	1,512	1,506	1,512	1,506
Number of Units	331	331	331	331
Geographic Unit	Region	Region	Region	Region
Years	2015-2020	2015-2020	2015-2020	2015-2020
Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes
F Statistic	188.71	161.34	188.71	161.34

Notes: Standard errors clustered at the unit level in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. The F statistic reported is the Kleibergen-Paap rk Wald F statistic. The following socioeconomic characteristics in 2000 are interacted with year dummies and used as controls: share of male individuals, shares of population identifying as black, white, and hispanic, share of people with high school, share of people with bachelor degree, share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. We also control for years of potential life list (YPLL) and the share of people who report having poor health. In this table, regions are the regions reported by Genworth's Cost of Care Survey. These regions are defined largely based on MSAs (and thus we always refer to them as MSAs), but a number of them for each state are aggregated and reported as "Rest of State." We do not use these aggregations in our estimates.

Table 6: Results on Log Number of Beds and Occupancy Rate - Reduced Form

	Log(beds)		Occupancy Rate	
	TWFE	Het. DiD	TWFE	Het. DiD
Avg Effect	0.232***	0.248*	0.004	0.018
	(0.027)	(0.094)	(0.019)	(0.115)
N	26,691	2,440	26,691	2,440
Geographic Unit	County	County	County	County
Years	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes

Notes: Standard errors in parenthesis clustered at the unit level: *** p<0.01, ** p<0.05, * p<0.1. The table reports reduced form results for the TWFE estimator and heterogeneous difference-in-differences estimator from De Chaisemartin and D'Haultfœuille (2020). For the heterogeneous difference-in-differences estimator, we report the normalized effect to make the comparison with the TWFE estimator easier. The reported specifications do not include controls aside from the fixed effects.

Table 7: Heterogeneous DiD - Reduced Form Results - Discretized Instrument

	Log(Nurse Hrs)		Log(Avg Wages)		Log(Prov Price)		Log(Semi-priv Price)	
	TWFE	Het. DiD	TWFE	Het. DiD	TWFE	Het. DiD	TWFE	Het. DiD
Avg Effect	0.175	0.220	0.085	0.025	0.142**	0.169	0.106*	0.133
	(0.119)	(0.165)	(0.064)	(0.073)	(0.066)	(0.110)	(0.061)	(0.095)
N	3,389	2,929	2,829	2,204	1,512	1,245	1,512	1,245
Geographic Unit	MSA	MSA	MSA	MSA	Region	Region	Region	Region
Years	2010-2020	2010-2020	2010-2020	2010-2020	2015-2020	2015-2020	2015-2020	2015-2020
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Avg Num Accum Periods	-	2.67	-	2.63	-	2.13	-	2.13

Notes: Standard errors in parenthesis clustered at the unit level: *** p<0.01, ** p<0.05, * p<0.1. The table reports reduced form results for the TWFE estimator and heterogeneous difference-in-differences estimator from De Chaisemartin and D'Haultfœuille (2020) using a discretized version of our instrument. To construct this discretized version, we calculate the changes in our instrument (log number of people in the relevant age range 40 years prior) with respect to the first period and we normalize it by the overall standard deviation in the period considered, σ. Then, we assume the treatment can take one of three values: (i) 0 if $|instr_{it} - instr_{i,2010}| \leq 0.1\sigma$; (ii) 1 if $instr_{it} - instr_{i,2010} > 0.1\sigma$; (iii) -1 if $instr_{it} - instr_{i,2010} < -0.1\sigma$. For the heterogeneous difference-in-differences estimator, we consider up to 5 periods for the effects and report the average cumulative effect per unit. We also report the average number of periods over which the effects are accumulated.

Table 8: Results on Log Average Wages - IV, 2nd Stage

	Nursing Home Industry		Registered Nurses		Licensed Nurses	
	(1)	(2)	(3)	(4)	(5)	(6)
log($pop75+$)	0.097	0.099	0.001	-0.140	0.063	0.066
	(0.074)	(0.091)	(0.065)	(0.089)	(0.060)	(0.081)
N	2,829	2,829	2,797	2,797	2,966	2,966
Geographic Unit	MSA	MSA	MSA	MSA	MSA	MSA
Years	2010-2020	2010-2020	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	No	Yes	No	Yes	No	Yes
F Statistic	231.07	152.60	174.33	119.92	179.84	130.1

Notes: Standard errors clustered at the unit level in parenthesis: *** $p<0.01$, ** $p<0.05$, * $p<0.1$. The F statistic reported is the Kleibergen-Paap rk Wald F statistic. The following socioeconomic characteristics in 2000 are interacted with year dummies and used as controls: share of male individuals, shares of population identifying as black, white, and hispanic, share of people with high school, share of people with bachelor degree, share of elderly in poverty, share of population in poverty, unemployment rate, median household income, and medium value of rent. We also control for years of potential life list (YPLL) and the share of people who report having poor health.

Table 9: Mechanisms - IV, 2nd Stage

	Panel A					
	Log(Beds)			Log(Price Private Room)		
	(1)	(2)	(3)	(4)	(5)	(6)
log(*pop*75+)	0.599***	0.417***	0.597***	-0.012	0.431**	-0.117
	(0.082)	(0.057)	(0.083)	(0.287)	(0.215)	(0.290)
log(*pop*75+) × CON	-0.343***		-0.340***	0.692**		0.749**
	(0.093)		(0.097)	(0.341)		(0.346)
log(*pop*75+)× Pass-through		0.015***	0.011**		0.012	0.012
		(0.004)	(0.005)		(0.011)	(0.013)
N	26,234	26,234	26,234	1,506	1,506	1,506
Geographic Unit	County	County	County	Region	Region	Region
Number of Units	2,446	2,446	2,446	259	259	259
Years	2010-2020	2010-2020	2010-2020	2015-2020	2015-2020	2015-2020
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F Statistic	434.82	539.39	250.14	29.36	76.40	21.91

	Panel B					
	Log(Nurse Hours)			Log(Average Wages)		
	(1)	(2)	(3)	(4)	(5)	(6)
log(*pop*75+)	0.328	0.463**	0.312	0.116	0.087	0.098
	(0.220)	(0.190)	(0.220)	(0.134)	(0.093)	(0.137)
log(*pop*75+) × CON	0.204		0.241	-0.020		-0.007
	(0.277)		(0.280)	(0.148)		(0.148)
log(*pop*75+)× Pass-through		0.052***	0.052***		0.016*	0.017**
		(0.014)	(0.014)		(0.008)	(0.008)
N	3,389	3,389	3,389	2,829	2,829	2,829
Geographic Unit	MSA	MSA	MSA	MSA	MSA	MSA
Number of Units	313	313	313	268	268	268
Years	2010-2020	2010-2020	2010-2020	2010-2020	2010-2020	2010-2020
Unit FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Controls	Yes	Yes	Yes	Yes	Yes	Yes
F Statistic	105.60	95.76	68.13	74.39	78.26	52.79

Notes: Robust standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. F statistic reported is the Kleibergen-Paap rk Wald F statistic. CON is an indicator of whether the unit is in an state with a CON law, which is constant over time. Lagged for-profit share is the share of beds in for-profit facilities one year prior. Since the facility data is not reported for some small facilities, we lose some data in some of these regressions. Regressions using CON laws control for differential time trends in states with and without these laws. Pass-through regressions control for differential time trends depending on the existence or not of pass-through laws at the beginning of the sample. Regressions with lagged for-profit share allows the initial (2009) share to have different linear effects depending on the year (i.e., we interact the initial share with year dummies). None of these controls affect our estimates in our main specifications.

Appendices

A Supporting Evidence for Instrument

Table A.1: Log Number of Beds

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\text{deaths}_{1960-1970})$	-0.009	-0.012				
	(0.020)	(0.018)				
$\log(\text{migration}_{1960s-1970s})$			0.001	0.000		
			(0.001)	(0.001)		
$\log(\text{births}_{1960-1970})$					0.109***	0.064
					(0.034)	(0.039)
$\log(population75+)$		0.322***		0.297***		
		(0.070)		(0.072)		
$\log(population50-59)$						0.232***
						(0.077)
Observations	672	672	668	668	672	672
Years	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020
Unit	MSA	MSA	MSA	MSA	MSA	MSA

Notes: The dependent variable is the log of beds. Robust standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. Here, we test how the outcome is affected by past population variables and whether deaths, migration, or births are significantly correlated with the outcome when we control for the relevant age range of the current population.

Table A.2: Occupancy Rate

	(1)	(2)	(3)	(4)	(5)	(6)
$log(\text{deaths}_{1960-1970})$	-0.006	-0.007				
	(0.004)	(0.004)				
$log(\text{migration}_{1960s-1970s})$			-0.020	-0.015		
			(0.017)	(0.017)		
$log(\text{births}_{1960-1970})$					-0.034	0.047
					(0.367)	(0.417)
$\log(population75+)$		-0.664*		-0.625*		
		(0.354)		(0.335)		
$\log(population50-60)$						-0.414
						(0.470)
Observations	672	672	668	668	672	672
Years	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020
Unit	MSA	MSA	MSA	MSA	MSA	MSA

Notes: The dependent variable is the average occupancy rate of facilities in each area. Robust standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. Here, we test how our outcome is affected by past population variables and whether deaths, migration, or births are significantly correlated with the outcome when we control for the relevant age range of the current population.

Table A.3: Number of Log Average Nursing Hours

	(1)	(2)	(3)	(4)	(5)	(6)
$\log(\text{deaths}_{1960-1970})$	-0.011	-0.011				
	(0.034)	(0.034)				
$\log(\text{migration}_{1960s-1970s})$			0.000	0.000		
			(0.002)	(0.002)		
$\log(\text{births}_{1960-1970})$					-0.116	-0.105
					(0.102)	(0.116)
$\log(population75+)$		-0.067		-0.142		
		(0.160)		(0.132)		
$\log(population50-60)$						-0.055
						(0.204)
Observations	660	660	656	656	660	660
Years	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020	2010 & 2020
Unit	MSA	MSA	MSA	MSA	MSA	MSA

Notes: The dependent variable is the log of average total nurse hours per facility. Robust standard errors in parenthesis: *** p<0.01, ** p<0.05, * p<0.1. We test how our outcome is affected by past population variables and whether deaths, migration, or births are significantly correlated with the outcome when we control for the relevant age range of the current population.