Price and Income Effects of Hospital Reimbursements*

Matthias Bäuml[a] Tilman Dette[b] Michael Pollmann[c d]

December 7, 2021

Abstract

Health insurance systems in many countries reimburse hospitals through fixed prices based on the diagnosis-related groups (DRGs) of patients. We quantify the effects of price and income changes for the full spectrum of hospital services as average and heterogeneous elasticities of quantities (number of admissions) and quality-related outcomes. For our empirical analysis, we use data on over 160 million hospital admissions, constituting the universe of hospital admissions in Germany between 2005 and 2016. Our identification strategy is based on instruments exploiting a two-year lag in regulatory price setting. The strategy lends itself to a placebo test demonstrating that our instruments do not have substantive anticipatory direct effects. We find that the compensated own-price elasticity of quantity is positive (0.2), while the income elasticity is negative (-0.15). On net, increasing *all* prices increases costs due to a behavioral response of larger quantities in addition to the mechanical increase.
Keywords: hospital; diagnosis-related groups; price elasticity; income elasticity
JEL: H21, H51, I11, I18, L51.

[a]University of Hamburg, Esplanade 36, D-20354 Hamburg, Germany. Email: matthias.baeuml@uni-hamburg.de.
[b]QuantCo, Inc. 955 Massachusetts Ave., Cambridge, MA 02139, United States. Email: dette@quantco.com.
[c]Stanford University, 579 Jane Stanford Way, Stanford, CA 94305, United States. Email: pollmann@stanford.edu.
[d]Corresponding author.
* This paper uses confidential hospital claims data. Researchers can request access to the data through the Research Data Centre of the German Federal Statistical Office, see `https://www.forschungsdatenzentrum.de/en` for details. We thank Rebecca Diamond, Mark Duggan, Liran Einav, Caroline Hoxby, Petra Persson, and participants in seminars at Stanford University, as well as two anonymous referees for helpful feedback and suggestions. We gratefully acknowledge a grant by The Europe Center at Stanford University. We have no conflicts of interest to declare.

1 Introduction

In many developed countries, hospitals are reimbursed for treating patients based on diagnosis-related group (DRG) systems. In these systems, a hospital receives a fixed price based on the diagnoses and other characteristics of the patient, irrespective of the incurred cost. Such prices per patient in a DRG are typically set by a regulatory agency rather than through market forces. The effects of price-setting, as well as improvements to how prices are set, are therefore of immediate policy relevance.

A multitude of empirical studies has documented distortive incentive effects of prices for individual DRG. Oftentimes, such effects are plausibly nominal, in that they reflect changes in reporting, but not necessarily in treatment.[1] However, comprehensive credible empirical evidence on price and income effects, covering the full spectrum of hospital services for a representative (patient) population and demonstrating real responses in who is treated, is limited.

In this paper, we estimate the price and income effects of prices for the full spectrum of hospital services in the German DRG system using quasi-random variation due to regulatory price setting. The German DRG system shares many similarities with other DRG systems such as the Medicare Prospective Payment System in the United States, such that our findings and empirical strategy may have broader applicability. Focusing on the German hospital system for our empirical analysis has two key advantages: First, we observe the universe of hospital admissions from 2005 to 2016. Observing all hospital admissions allows us to accurately measure income and to address substitution between hospitals, DRGs, and insurance types. Second, we exploit the institutional details of the German price setting mechanism for credible causal identification for all DRGs. Prices each year are set according to a two-year lag of average costs in the DRG. We argue that price and income shocks due to this two-year lag in costs are plausibly exogenous conditional on fine fixed effects. Empirically, we demonstrate that a *one*-year lag of the same variation does not have a meaningful direct effect, such that the exclusion restriction is plausibly satisfied when using the *two*-year lag as an instrument. In other words, we show that our instruments for income and prices, which are based on these historical cost estimates, do not cause substantive anticipatory effects. Our estimated effects therefore are plausibly identifying the effects of changes in income and prices, rather than direct effects of our instruments. We use a "simulated instrument" (cf. Currie and Gruber, 1996a,b) for income that measures the income a hospital department would realize under new prices if it treated patients with the same characteristics as in a previous year. The instrument isolates and uses only the exogenous variation in income due to the lag in price setting, rather than the endogenous variation due to changes in the number and composition of the patients treated in the hospital department. Since we observe yearly variation in prices for all DRGs and in income for all German hospitals in a panel of 12 years, we are able to also estimate heterogeneous elasticities for clinically defined super-groups of DRG based on the Clinical Classification System categories. We make these heterogeneous elasticity estimates available to other researchers for further use. Overall, our results show that not only changes in relative prices affect hospital behavior, but that increasing or decreasing the price level uniformly also leads to behavioral responses and affects real treatment decisions due to income effects.

We find a positive average price elasticity of quantity[2] and a negative income elasticity. Specifically, we find that hospitals respond to a compensated 1% increase in the price of a single DRG by treating approximately 0.2% more such patients, reflecting the increased financial incentive to treat patients of this particular DRG. The income elasticity, in contrast, is negative: If a hospital department's income increases by 1%, then – holding the price paid for a DRG fixed – it treats about 0.15% *fewer* patients of

[1] A prime example of nominal effects are those found on birth weights: Reported birth weights exhibit biologically implausible bunching around thresholds where reimbursements change discontinuously (e.g. Jürges and Köberlein, 2015; Reif et al., 2018). For other recent work on upcoding, see, for instance, Cook and Averett (2020).

[2] Throughout this paper, quantity refers to the number of hospital admissions. We sometimes write patients rather than admissions for more accessible language.

a given DRG. This effect is not driven by upcoding (cf. Silverman and Skinner, 2004) responses: The income elasticity remains largely unchanged if the income shock is due to clinically similar or dissimilar DRGs, which differ in their suitability for upcoding. This suggests that we measure real changes in who is treated rather than upcoding or minor changes in treatment.

We estimate a net positive effect on quantity when the price for all DRGs treated by a department increases. A 1% increase in all prices causes hospital departments to treat over 0.1% more patients. We find that this increase reflects a real increase in the number of patients. For this analysis, we aggregate patients at the department level, such that upcoding does not lead to increases in measured quantity. The effect cannot be explained by reallocation of patients between departments or hospitals: If the price level increases for hospital departments in a given hospital service area, then more patients from the corresponding zip codes are treated in hospitals. These results highlight the importance of calculating not only the mechanical effect of price changes but also analyzing behavioral responses to payment reforms.

While prices in DRG systems most directly incentivize treating more or fewer patients, we also find effects on measures of intensity and quality. Within a DRG, the average length of stay decreases as prices for other DRGs increase income. In addition to potential selection effects, this likely reflects stickiness of both the number of beds and personnel. In particular stickiness of personnel likely drives adverse effects of income shocks on quality. We find mild evidence suggesting that both the incidence of hospital-acquired conditions, such as hospital-acquired pneumonia and infections of surgical wounds, potentially leading to sepsis, and mortality rates increase as a response to changes in prices and income. When hospital departments treat more patients with the same number of staff, nurses and doctors likely have less time per patient, resulting in more preventable mistakes in care and deaths.

We contribute to the literature on hospital responses to price and income changes by estimating average and heterogeneous effects for the full spectrum of hospital services for a representative patient population. A substantial literature has documented responses to incentives for upcoding and intensifying service for specific DRGs (Gilman, 2000; Silverman and Skinner, 2004; Papanicolas and McGuire, 2015; Jürges and Köberlein, 2015; Foo et al., 2017; Di Giacomo et al., 2017; Einav et al., 2018; Geruso and Layton, 2020, among others). Dafny (2005) uses price variation due to a 1988 policy reform that affected 43% of Medicare admissions across a wider range of DRGs. However, the population of Medicare beneficiaries differs systematically from the overall population, and represents only a part of hospital income. In our paper, we are able to study the effects of hospital income and the overall price level by observing the universe of hospital admissions rather than those that are part of any particular insurance system. Observing the universe of admissions is particularly important if income gains or losses from one part of admissions can be offset by price and quantity changes for other, unobserved, admissions (Dranove et al., 2013; Feldman et al., 2015; Finkelstein et al., 2019). Duggan (2000, 2002) and Baicker et al. (2013) highlight the importance of such reallocation and spillover effects between and within hospitals. Our estimates of the effects of increasing the overall price level are in the same direction as those reported by Clemens and Gottlieb (2014) for physician services, but quantitatively substantially smaller.

The remainder of this paper is structured as follows. In Section 2, we provide a brief introduction to DRG systems and their implementation in Germany. We describe our data in Section 3. We discuss our empirical specifications that isolate the source of identifying quasi-random variation in Section 4. In Section 5, we present our empirical estimates and show robustness to alternative specifications and interpretations. In Section 6, we propose a sufficient statistics approach to the design of optimal DRG systems based on the estimated elasticities. We conclude with a brief discussion of the implications of our findings as well as their limitations in Section 7.

2 Background: Hospital Reimbursements in Germany

Since 2005, hospital reimbursements in Germany are based on a DRG system similar to those of many other developed countries. The German hospital system has two key features that are important for studying price and income effects. First, administrative data contain the universe of hospital admissions allowing us to accurately measure most hospital income, and largely eliminates substitution or spillovers between different insurers, insurance types, or hospitals. With 12 years of data (2005 – 2016) and approximately 1500 hospitals in the cross-section, the German setting also offers sufficient sample size for our purposes. Second, the price setting mechanism used by the German regulator induces quasi-random variation in prices and income. Since prices are set annually, the 12-year long panel offers substantial variation within DRGs and hospitals.

Health insurance in Germany is universal, and hospitals are reimbursed at the same DRG rates irrespective of a patient's insurer. For the inpatient hospital admissions in our data, patients face a copayment of €10 per night for up to 28 nights per calendar year to be paid to the insurer irrespective of the DRG and treatment. Variation in the real value of the copayment is mostly be absorbed by year and hospital fixed effects in our empirical specifications. Since we are interested in the effect and optimal design of DRG reimbursements, we therefore focus on the hospital side of the market, without modeling patient choice.

2.1 Diagnosis-Related Groups

In DRG systems, patients are classified into groups primarily based on diagnosis codes and demographic characteristics. The reimbursement a hospital receives for treating a patient is then fixed within group, regardless of actual expenditure. The German regulator scales reimbursements by hospital-specific (state-specific starting in 2010) factors to account for differences in cost levels across hospitals as well as inflation.

In the German DRG system, introduced in 2005,[3] the regulator designs the classification algorithm. Since reimbursements are fixed within DRG (and year), the regulator attempts to minimize within-DRG cost heterogeneity with annual updates to the algorithm. Consequently, the number of DRG has increased steadily from 878 in 2005 to 1220 in 2016. Since the definition of DRGs changes over time, two hypothetical patients with similar but not identical characteristics may be classified into the same DRG in one year and into separate DRG in the next year. Compared to the Medicare Prospective Payment System of around 500 DRG, the German system is more granular.

In practice, hospitals employ trained coding staff who enter patient characteristics after the end of a hospital stay into certified software that computes the classification based on the algorithm that was in effect at admission. Due to the large number of DRG, small changes to coding, such as the choice of "primary" and "secondary" diagnoses, can affect DRG and hence reimbursements. Some software automatically computes reimbursements for such hypothetical similar patients, potentially encouraging upcoding behavior.

2.2 The German Hospital Sector

We view the formal organization of hospitals as consisting of three layers: 1) ownership, 2) physical location, 3) hospital department. In Germany, there are a number of hospital chains with multiple hospitals owned by the same company.[4] However, our data do not allow us to distinguish between

[3] Reimbursements prior to 2005 were primarily based on *per diem* rates that create strong financial incentives for hospitals to increase the number of days per stay.

[4] The largest such chain owns 110 hospitals across most German States.

different types of ownership; our estimates should be considered as averaging over ownership-specific effects. In total, there were around 1,500 hospitals in 2014; approximately 40% of them are private not-for-profit, 30% are public, and 30% are private for-profit (InEK, 2015). Entry and exit in the hospital market are regulated by state governments to ensure at least a minimum level of coverage. During our sample period, the public discourse has mostly focused on excess capacity and privatization, with the number of hospitals slowly decreasing from year to year. In our main analysis, we abstract away from such dynamics by focusing on a balanced panel of hospitals, but show that results are similar in an unbalanced panel.

Most hospitals comprise several specialized departments, which open and close more frequently than entire hospitals, such as cardiology, pediatrics, and urology departments. Hospital departments in Germany are typically fairly autonomous business units (Fleßa and Nickel, 2010). Hospital beds, rooms, and wings are assigned to specific departments, and operating room schedules often have fixed time slots allocated to departments for planned surgeries. In the medium- and long-run, these resources can be reallocated between departments in the same hospital, allowing some forms of internal competition. In interviews we conducted, hospital department heads stated that decisions are made within the department, with budgets and income targets for individual departments set at the hospital level. We therefore regard the hospital department as the most important organizational level to study in our context. In other contexts or countries, however, different levels of organization may be more important, and the hospital department of an admission are not always available to the researcher.

2.3 Price Setting

In the German DRG system, the regulator updates prices annually. Prices are proportional to two-year old average reported costs, but scaled each year in an attempt to hold overall inflation-adjusted expenditure constant.

Specifically, prices for calendar year t are set in the fall of year $t-1$ based on reported costs for patients discharged in year $t-2$. These cost reports come from around 20% of hospitals which participate voluntarily,[5] and are checked for plausibility and correctness by the regulator. The identity of cost-reporting hospitals is not public beyond a larger list of hospitals which have signed an agreement that is required for cost reporting eligibility. However, coarse summary statistics about reporting hospitals published by the regulator suggest some year-to-year variation in which of the eligible hospitals report costs. For cost reports from year $t-2$ that pass the plausibility tests of the regulator, the regulator uses the classification algorithm for year t to group admissions according to the updated DRG classification.

The regulator takes the unweighted average of reported costs within each DRG to determine relative prices. Relative prices in year t are therefore proportional to average costs in year $t-2$. Next, prices for year t are scaled to hold aggregate real expenditure constant from year $t-1$ to year t if both years had the same admissions as year $t-2$. The regulator publishes the new prices and classification algorithm for calendar year t in the fall of year $t-1$.

We focus our analysis on the relative price $p_{g(t),t} \equiv \bar{c}_{g(t),t-2} n_t$, where $\bar{c}_{g(t),t-2}$ is the average of reported costs for admissions from year $t-2$ according to the DRG $g(t)$ of the classification algorithm for year t, and n_t is the normalization factor holding aggregate real expenditure constant if admissions were the same each year. These prices best reflect the variation in reimbursements that is exogenous to hospital choices. The actual reimbursements paid to a hospital scale $p_{g(t),t}$ by a hospital- or later state-specific adjustment factor, surcharges and discounts if length of stay differs substantially from the diagnosis mean, and minor fees paid, for instance, to teaching hospitals.[6]

[5] Starting in 2017, after our main sample, a small number of hospitals is required to report costs in order to increase representativeness. See Appendix A.2.4 for a brief analysis of the effects of this new policy on prices.

[6] We consider length of stay separately below. We have found, in separate analyses available from the authors by request,

3 Data: Universe of Hospital Admissions 2005 – 2016

Our primary data source is administrative claims data collected by the German regulator. We also use supplemental data on the cost reports published by the regulator. We merge additional data on the physical locations of hospital departments to accurately track departments over time despite mergers and internal reorganizations. Finally, our heterogeneity analysis is based on the clinical classifications software (CCS) developed by the Agency for Healthcare Research and Quality in the United States.

3.1 Diagnosis and Reimbursement Data

We use administrative claims data collected by the regulator. Insurers receive these claims to calculate the reimbursements owed to hospitals. The data include all inpatient hospital admissions in Germany between 2005 and 2016. For each hospital admission, we observe demographic information (e.g. age, gender, zip code), primary and secondary diagnosis codes, procedure codes, as well as a hospital identifier and standardized department codes. These data allow us to calculate actual reimbursements as well as counterfactual reimbursements that hospitals would have received for each admission if it occurred in any year. The regulator collects copies of these data mainly to analyze refinements to the German DRG system. The German Federal Statistical Office makes anonymized data available for research at its Research Data Centre.

To address the yearly redefinition of DRGs by the regulator, we define more granular "sub-DRG." For each admission, we apply the DRG classification algorithm of every year, irrespective of the year the admission actually occurred in. Specifically, suppose $\text{drg}_{i,t}$ is the DRG admission i would be classified as in year t. Admissions i and j are in the same sub-DRG if and only if $\text{drg}_{i,s} = \text{drg}_{j,s}$ for $s = 2005, \ldots, 2016$. A sub-DRG g is therefore defined by its sequence of DRGs $(\text{drg}_{2005}, \ldots, \text{drg}_{2016})$. We assign each admission to the sub-DRG that matches its sequence of DRGs.

Suppose, for instance, that there are two DRGs, A and B, and we have two years of data. The regulator changes the definitions of the DRG system such that some admissions that would have been classified as A in the first year, are classified as B in the second year, and vice-versa. Then we create up to four sub-DRGs. The first sub-DRG contains all admissions that would have been in A in both years. The second sub-DRG contains all admissions that would have been in A in the first year but in B in the second year. The third sub-DRG contains all admissions that would have been in B in the first year but in A in the second year. The fourth sub-DRG contains all admissions that would have been in B in all years. If B is an entirely new DRG in the second year, the third and fourth sub-DRG are empty. In general, if there are G_t DRGs in year t, there are $\prod_t G_t$ potential sub-DRGs. Most of these sub-DRGs, however, would be empty because no admission classified as suffering from heart failure in one year would be classified as suffering from an entirely different condition in another year. We therefore only need to create a sub-DRG if there is at least one actual admission in it.

This definition creates time-invariant sub-DRGs that exist through all years in our sample despite changes to the classification algorithm. Importantly, it allows us to assign a unique price $p_{g,t}$ to every sub-DRG g and year t, based on the price of the unique DRG that all admissions in sub-DRG g are classified as in year t. Similarly, we merge data about the cost reports, which the regulator publishes at the DRG-year level. Our empirical analysis then relates the number of admissions in sub-DRG g in year t in a hospital department d, $q_{d,g,t}$, to the price for such admissions in that year, $p_{g,t}$, and the hospital department income.

---

the other components of the final reimbursement received by hospitals, such as hospital- and state-specific base rates, to be orthogonal to the changes in relative prices we study in this paper and hence do not focus on them in more detail in the present paper, though they may be of interest in their own right (for instance Salm and Wübker, 2020).

3.2 Hospital Location Data

In the administrative data, an institutional code identifies the hospital for each admission. For hospitals with multiple locations, the location releasing the patient is also indicated by a numerical variable, which is consistent within but not necessarily across years. In addition, we observe which department within the hospital treats the patient.[7] However, multiple hospital locations may share the same institutional code, and codes may change when hospitals are acquired by hospital chains or when hospitals merge. We therefore use additional data on the physical location of hospital departments to accurately track departments over time and use the panel structure of our data.

Hospitals in Germany are mandated to fill out "quality report cards" on a biannual (2006 – 2012) or annual (since 2013) basis. These reports include three parts: Structure and performance data for the hospital (Part A), structure and performance data for each hospital department (Part B), and a set of quality indicators (Part C). The structural data from parts A and B include the institutional codes used in the administrative claims data, as well as addresses of the hospitals and departments. Through geocoding, these addresses allow us to determine which institutional codes and departments are part of the same hospital complex. If a hospital changes its institutional code due to a merger or acquisition, we can track it through its common physical location pre and post merger.

3.3 Clinical Classification Software

Developed by the Agency for Healthcare Research and Quality in the United States, the Clinical Classifications Software (CCS) is a powerful tool for clustering patients into clinically meaningful categories. Our heterogeneity analysis is based on the 2016 ICD-10 version, which matches the final year of our sample. In total, there are 283 CCS groups constituting 18 top-level categories (such as "Diseases of the circulatory system") and 136 subcategories (such as "Hypertension" and "Diseases of veins and lymphatics").

We aggregate our sub-DRGs into CCS groups for heterogeneity analyses. The CCS assigns each diagnosis code to a unique CCS group. We determine the most common primary diagnosis code of each sub-DRG and assign the sub-DRG to the corresponding CCS group. Since specialized departments and institutions for psychiatric and psychosomatic care are not reimbursed through the DRG system, we may have few if any sub-DRGs for these CCS groups and categories. The estimated effects for such CCS groups and categories are therefore only valid for the average such admission in our data.

3.4 Summary Statistics

Our sample contains claims and reported costs data from the years 2005 – 2016. Overall, we use data on about 162 million inpatient hospital admissions, linked to approximately 1,500 German hospital locations. The number of hospitals is significantly lower than the 1849 institutional codes, reflecting the need for our secondary data source to combine different institutional codes that belong to the same hospital location. Our numbers of hospital admissions and hospitals generally match those reported by the regulator each year (e.g. InEK, 2015, for 2014). The number of DRG defined by the regulator grows steadily over time, from 878 when the G-DRG system was introduced in 2005, to 1,220 at the end of our sample.

For our primary analysis, we restrict ourselves to a balanced panel: We only keep regression observations of hospital departments and sub-DRGs with hospital admissions for every year in our sample. Importantly, we always calculate income (and its instrument, described in more detail in Section 4) based on *all* admissions in the hospital department year, even if the admission is in a sub-DRG that does not meet

[7] If a patient is treated by multiple departments, we assign the patient to the department by which she is treated the longest. Our results are robust to instead assigning patients to the admitting or releasing department.

Table 1: Summary statistics for the baseline sample with a balanced panel.

	0.25 quantile	median	0.75 quantile	Maximum	Std. Dev.
departments per hospital	1	3	5	20	3.13
sub-DRGs per DRG	1	1	2	11	0.96
admissions per sub-DRG-hospital-department-year	26	42	73	7394	100.36
sub-DRGs per CCS	23	38	62	347	46.82
relative price changes	-1.90	0.00	2.00		3.94
department income changes	-1.93	3.26	8.78		9.68
instrument for department income changes	-1.40	-0.49	0.47		1.60

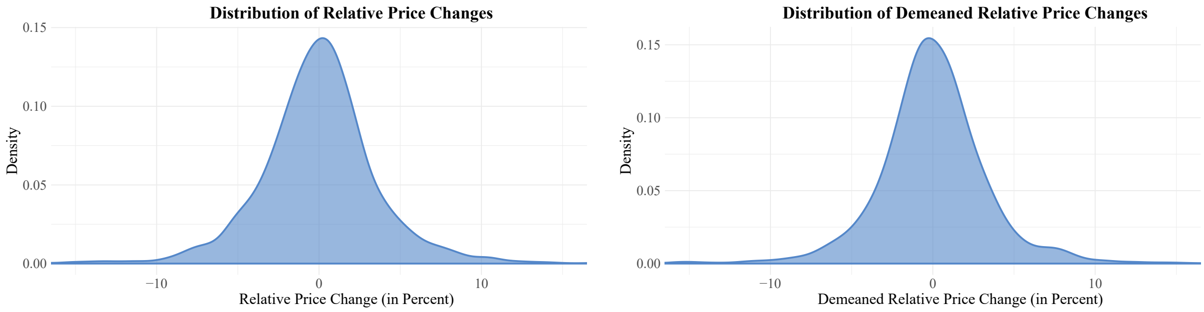
the balanced panel requirements to be included as a regression observation. Our regression specification and instruments require non-zero quantities in the previous two years. In an unbalanced panel, we would need to treat the opening and closing of departments or clinical services asymmetrical: Due to lack of pre-data, opening would be skipped, while closing would be included. We therefore use a balanced panel that treats zero-quantities early and late in the sample symmetrically. After balancing the panel, around 70% of all admissions remain; dropped admissions are in particular due to hospital (departments) that close during our sample period as well admissions in sub-DRG that are infrequently treated by the treating hospital. We do not take a stance on whether less "permanent" hospitals and more infrequent sub-DRG are more or less responsive to short-term changes in financial incentives. It is unclear whether yearly price and income changes systematically affect which departments open and close due to potentially large fixed costs and required regulatory approval. In our robustness analysis, we find that an unbalanced panel yields very similar results.

We show some summary statistics for our balanced panel in Table 1. A substantial fraction of hospitals consists of a single department (typically labeled as internal medicine) treating all patients, and most remaining hospitals are subdivided into a small number of departments. Balancing the panel drops sub-DRG-hospital-department pairs that do not have patients every year. As a result, for many DRGs only a single sub-DRG remains. However, the split into the more granular sub-DRGs remains necessary for uniquely defined price series. The interquartile range for the number of admissions in a sub-DRG in a hospital department per year ranges from 26 to 73, or between one every two weeks and three over a two week period. However, there is a substantial right tail, with a maximum of approximately 20 patients per day of a sub-DRG in a hospital department. Most CCS groups contain a few dozen sub-DRGs (conditional on containing at least one). The more sub-DRGs we have per CCS, the larger the sample size and exogenous variation in our heterogeneity analysis.

Finally, we illustrate the magnitude and variation in our primary explanatory variables. The last three rows of Table 1 show the quartiles and standard deviation of the year-on-year percentage changes in sub-DRG prices, hospital department income, and our simulated instrument for income, which we describe in detail in Section 4. Figures 1 and 2 plot the corresponding densities.

Panel a of Figure 1 shows the density of year-over-year changes in sub-DRG prices, our main source of identifying variation, weighted by the number of admissions. Price changes of more than 5% are not uncommon. As panel b of the same figure demonstrates, these changes are not dominated by sub-DRG-specific trends, which are removed by fixed effects in our empirical specifications; see Section 4 for details. Substantial variation remains around the within-sub-DRG average growth rate of prices.

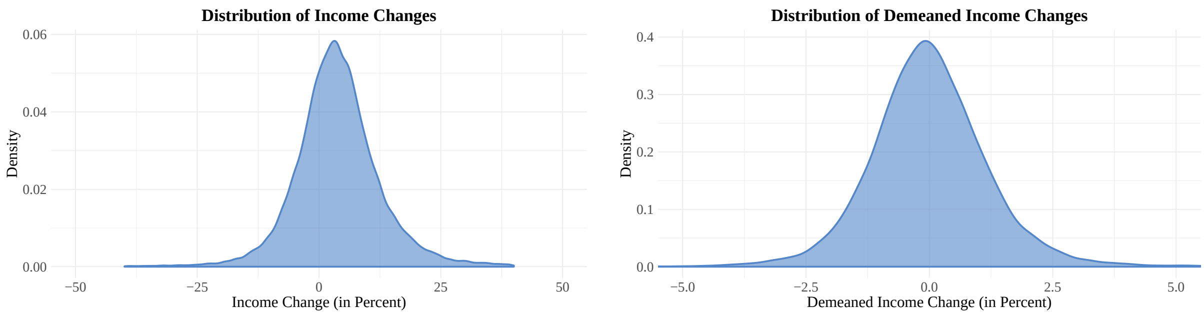
Department income, shown in panel a of Figure 2 varies substantially. This variation is primarily due to changes in the number of admissions, and is therefore likely offset by corresponding changes in costs. To identify causal effects of changes in income, we rely on a simulated instrument that isolates the exogenous variation due to changes in prices, holding the number of admissions fixed; see Section 4 for details. As panel b illustrates, the remaining variation, after also removing the fixed effects of our empirical specifications, is substantially smaller. However, these regulator-induced exogenous income changes are



(a) Raw Relative Price Changes

(b) Demeaned Relative Price Changes

Figure 1: Densities of relative price changes. Panel a shows the distribution of raw price changes, weighted by the number of admissions affected. Panel b first demeans the price changes by the fixed effects of our empirical specifications.



(a) Department Income Changes

(b) Instrument For Department Income Changes

Figure 2: Densities of department income changes. Panel a shows the distribution of raw department income changes, weighted by the number of admissions affected. Panel b instead shows the distribution of our simulated instrument for department income changes, demeaned by the fixed effects of our empirical specifications.

still economically meaningful. For comparison, KPMG (2013) estimates average profit margins of around 5%, with substantial variation across hospitals.

4 Empirical Strategy

We estimate price and income elasticities by regressing year-on-year percentage changes in outcome variables on corresponding percentage changes in price and hospital department income. Since the change in income mechanically is a function of the change in quantity, we use a simulated instrument e(cf. Currie and Gruber, 1996a,b) that fixes quantities and isolates the variation due to regulatory price setting. We assess the validity of our instrument and exogeneity of prices in general with a placebo test. We test alternative interpretations of our income elasticity estimates as substitution between departments or hospitals through different aggregations of admissions.

4.1 Regression Specifications

We regress year-on-year percentage changes in the outcome variable on corresponding percentage changes in price and income. Let g denote the sub-DRG (see Section 3 for the definition of sub-DRGs, which are time-invariant and more granular than DRGs), d the hospital department treating the patients, and t the calendar year. Our baseline regression specification is

$$q^{\%}_{d,g,t} = \beta_0 + \beta_1 p^{\%}_{g,t} + \beta_2 \text{inc}^{\%}_{d,t} + \delta_{d,g} + \delta_t + \epsilon_{d,g,t} \tag{1}$$

where the superscript % denotes year-on-year percentage changes. For instance, the percentage change in the outcome variable, quantity, is calculated as $q^{\%}_{d,g,t} \equiv (q_{d,g,t} - q_{d,g,t-1})/q_{d,g,t-1}$.[8] In all regressions, we weight observations by the number of admissions they represent such that we obtain average effects at the admission-level rather than averages dominated by the large number of sub-DRGs representing a small fraction of admissions.

Throughout, our regressions include cross-sectional fixed effects $\delta_{d,g}$ and year fixed effects δ_t. Since the outcome (and regressors) are year-on-year percentage changes, cross-sectional fixed effects control for percentage *trends* in both the outcome and in price and income shocks, allowing such trends to vary flexibly across sub-DRGs and hospital departments. Our specification relates yearly deviations from long-run trends in quantities to yearly deviations from long-run trends in prices and income. Hence, the specification does not compare departments and sub-DRG in specializations with growing income and prices to those with declining income and prices. Instead, it compares the same department and sub-DRG to itself in years where the income and price changes are relatively good or bad for the department. Below, we argue that these yearly fluctuations in prices and income around long-run trends represent quasi-random variation. Year fixed effects δ_t furthermore absorb common changes in prices and income, for instance due to inflation.

The coefficients of interest are the price elasticity β_1, that is the percentage effect of a one percent increase of price p on quantity q, and the similarly defined income elasticity β_2. There are three key channels through which hospitals generate changes in quantity in response to changes in prices and income: changes in coding, changes in treatment, and changes in who receives (any) inpatient treatment. Any of these channels are relevant for hospital reimbursement systems. While we cannot, in general across all DRG, disentangle these channels, we provide evidence from aggregate specifications, described in more detail below, suggesting substantial "real" (cf. Dafny, 2005) responses in who is treated.

We argue that (demeaned) variation in price changes is exogenous. Recall that price changes are proportional to changes in reported costs two years prior. Changes in reported costs from earlier years have little direct effect on current outcomes for four key reasons: First, our specifications in percentage changes include cross-sectional fixed effects and thereby remove long-run trends that past cost changes may otherwise predict. Second, the regulator only obtains cost reports from a sample of hospitals, resulting in relatively few reported admissions per DRG. This creates non-negligible yearly sampling variation in the average cost reports. Third, the sample of hospitals reporting costs to the regulator is self-selected and varies from year to year. If a high-cost hospital newly reports costs for its admissions, the average of reported costs (and later the price) rises more for DRGs where the high-cost hospital reports relatively more admissions. Fourth, DRG redefinitions change, for a given admission, the composition of admissions in its DRG from one year to the next. If an admission is grouped together with relatively low-cost admissions in one year and high-cost admissions in the next, its reimbursement increases without changes to costs or how such a patient is treated. We quantify these sources of random variation in prices in Appendix A.2.

Variation in income, in contrast, is *mechanically* endogenous in regressions with quantity changes as the outcome variable, irrespective of the price setting mechanism. Specifically, the issue is that the

[8] Alternative specifications using the average quantity in the denominator, differences in natural logarithms, or Poisson fixed effect models lead to quantitatively similar estimates. We prefer the specification in percentage changes because it most closely corresponds to estimating average elasticities when there is heterogeneity in effects and percentage changes are large such that differences in logs provide poor approximations.

change in income is related to both changes in prices and changes in quantities:

$$\text{inc}^{\%}_{d,t} = \frac{\sum_g p_{g,t} q_{d,g,t} - \sum_g p_{g,t-1} q_{d,g,t-1}}{\sum_g p_{g,t-1} q_{d,g,t-1}} + \epsilon_{d,t}$$

where $\epsilon_{d,t}$ includes the inflation adjustment, surcharges or discounts for extreme length of stay, and other minor fees. Note that income, as well as the instrument described below, includes income from all sub-DRGs g, whether or not they meet the balanced panel restrictions discussed in Section 3.4 to be included as observations in our primary analysis. OLS regression of $q^{\%}$ on $\text{inc}^{\%}$ yields coefficient estimates close to 1 because (uniform) quantity changes *cause* (one-to-one) income changes, creating simultaneity bias.

The income elasticity β_2, however, should capture the effect of exogenous variation in income; that is, variation in income that the regulator induces by (exogenous) changes in the total reimbursements to a hospital department. To estimate the income elasticity, we therefore use a simulated instrument. The instrument "simulates" income in a world where the endogenous variable, here quantity, is held constant, and only the quasi-random variation in prices creates changes in income. Unless specified otherwise, we construct our simulated instrument as

$$\widetilde{\text{inc}}^{\%}_{d,t} = \frac{\sum_g p_{g,t} q_{d,g,t-2} - p_{g,t-1} q_{d,g,t-2}}{\sum_g p_{g,t-1} q_{d,g,t-2}} \tag{2}$$

where we fix quantities at their level $t-2$.[9] By using the same aggregation of prices as in actual income, albeit with slightly different (quantity-) weights, we obtain an instrument that is highly correlated with the endogenous variable (strong first stage), but only uses the exogenous variation (exclusion restriction). While in principle we can use all prices as separate instruments, the simulated instrument is constructed to combine them in an efficient and parsimonious way based on our knowledge of the institutional setting.

We can alternatively interpret the simulated instrument as the *expected* income change, which hospital departments can calculate and respond to. When the prices for year t are announced in fall of year $t-1$, hospitals can forecast their income for the next year based on admissions of the most recent year with complete data available to them at the time, $t-2$. Such income forecasts are possible in typically-used software for coding patients into DRGs, and can be used to set income targets for departments. We see this as a conceptually attractive mechanism generating income effects, so we also report the reduced form effects of the instrument in addition to the IV estimates. However, we cannot distinguish between these mechanisms in the data, so we remain agnostic about the ultimate source of the estimated effects.

In addition to sub-DRG level regressions, we also estimate specifications based on aggregating admissions at the hospital department level. The aggregated regressions are of the form

$$q^{\%}_{d,t} = \beta_0 + \beta_2 \text{inc}^{\%}_{d,t} + \delta_d + \delta_t + \epsilon_{d,t} \tag{3}$$

Here, $q^{\%}_{d,t} \equiv (\sum_g q_{d,g,t} - q_{d,g,t-1})/(\sum_g q_{d,g,t-1})$ measures the percentage change in aggregate admissions in department d from year $t-1$ to year t. The coefficient β_2 captures the effect of increasing department income on the aggregate number of patients treated in the department, irrespective of the patients' sub-DRGs.

Since upcoding by definition shifts patients between DRGs but does not affect the number of patients treated in the department, we can think of β_2 as measuring real rather than nominal responses (cf. Dafny, 2005). As the regulator increases the overall price level, hospitals face a stronger financial incentive to treat otherwise marginal patients. We therefore refer to β_2 as the uncompensated average price elasticity,

[9] Alternative specifications holding q fixed at some other level, such as initial-period or average, yield similar estimates.

which measures the effect of increasing reimbursements for all DRGs uniformly. Beyond direct effects of the price of a DRG on the number of admissions of the same DRG, complementarities in treatment of different DRGs can generate real changes in the aggregate number of patients treated, and are also included in the effect size β_2. While it is neither a standard price nor income elasticity, the uncompensated average price elasticity has independent policy relevance: This elasticity captures the changes in the number of patients treated net of reporting responses, and is thus informative about the net change in admissions due to making a hospital reimbursement system uniformly more or less generous.

The standard errors reported in this paper reflect the design-based variation (cf. Abadie et al., 2020) in our estimates due to quasi-random assignment of price (and income) shocks. We cluster standard errors at the levels of treatment assignment (Abadie et al., 2017), which are the department-year and sub-DRG-year levels. This reflects the repeated sampling thought experiment where each year the price and income changes, net of long-run trends (the fixed effects in our regression specifications), are randomly assigned, holding the hospital departments and sub-DRGs in our sample fixed. Departments of the same specialization still differ substantially in the composition of admissions, such that the correlation of income shocks between departments of the same specialization is not of first-order concern empirically. Income shocks due to prices, holding quantities fixed, are also not correlated across departments within the same hospital because there is even less overlap in the DRGs treated. Correlations of price and income changes net of long-run trends also show close to no serial correlation (see Appendix A.2 for price changes), so correlation within cross-sectional unit over time is also empirically not relevant. Since we observe the universe of hospitals and admissions in Germany, these standard errors are conceptually attractive, yielding confidence intervals for the average effect in the population of German hospitals and admissions (Abadie et al., 2017). Alternative sampling-based standard errors that cluster at the hospital instead of the hospital-year level instead reflect a thought experiment where German hospitals are i.i.d. draws from a hypothetical super-population of hospitals. For our baseline regressions, such standard errors are 10%-20% larger than the design-based standard errors we report. However, since we observe the universe of hospitals and admissions, we do not find this hypothetical super-population to be of particular interest (cf. Abadie et al., 2020).

4.2 Assessing the Identifying Assumption

In the ideal experiment, the regulator randomly changes relative prices across DRGs as well as the price level across hospitals from one year to the next. In our observational data, price changes today correspond to changes in reported costs two years earlier. To identify the causal effects of prices and income, the exclusion restriction requires that these lagged cost changes only affect outcomes through prices and incomes, but not otherwise.

While the exclusion restriction is untestable, we can assess its plausibility with a placebo test on a "pseudo-outcome" (Athey and Imbens, 2017). We use the one-year lag of the outcome variable as a pseudo-outcome that is known not to be affected by the two-year lag change in costs if a stronger version of the exclusion restriction holds. Intuitively, suppose the "treatment," $\bar{c}^{\%}_{g,t-2}$ and the aggregated simulated instrument, has no effect on the quantity change in year $t-1$, $q^{\%}_{d,g,t-1}$. Then it is plausible that there are no unobserved differences between those observations with large treatment and those with low treatment; that is, the treatment satisfies the exclusion restriction / exogeneity. By a similar argument, even if $\bar{c}^{\%}_{g,t-2}$ has an effect on $q^{\%}_{d,g,t-1}$, we expect its effect on $q^{\%}_{d,g,t}$ to be smaller in magnitude. The longer the time difference between changes in cost reports and changes in outcomes, the smaller any direct effect is expected to be. Any persistence in cost changes is small in magnitude, as we demonstrate in Appendix A.2.

4.3 Alternative Interpretations of Income Effects

We assess whether our estimate of the income effect is driven by substitution between DRGs, between departments, or between hospitals, rather than real changes in who is treated.

To interpret the coefficient on the simulated instrument as an income effect, we need to rule out that the underlying variation in prices instead picks up cross-price elasticities. The price of one sub-DRG can affect the outcome for another directly if there are coding responses or complementarities in treatment, such as shared resources. To reduce the influence of cross-price elasticities through the simulated instrument for income, we remove the price variation due to those sub-DRGs which are most likely to create non-zero cross-price elasticities.

Arguments for non-zero cross-price elasticities typically rely on clinical similarity of DRGs: If two DRGs are clinically similar, more complementarities in treatment may exist. Similarly, if hospitals respond to financial incentives by upcoding, any upcoding plausibly only occurs between clinically similar DRGs. Hospitals cannot easily engage in upcoding resulting in a DRG appearing clinically dissimilar from the alternative coding. As our primary strategy to address cross-price elasticities, we therefore remove price variation from the simulated instrument that is due to sub-DRG that share the same CCS group. This instrument is given by

$$\overset{\smile}{\text{inc}}^{\%}_{d,g,t} = \frac{\sum_{g':\, \text{ccs}(g') \neq \text{ccs}(g)} (p_{g',t} q_{d,g',t-2} - p_{g',t-1} q_{d,g',t-2})}{\sum_{g':\, \text{ccs}(g') \neq \text{ccs}(g)} p_{g',t-1} q_{d,g',t-2}}$$

where g is the sub-DRG of the regression observation, and $\text{ccs}(g)$ is the CCS group of g as defined in Section 3. By summing over sub-DRGs g' with CCS group different from that of g, the instrument only uses variation in prices of clinically dissimilar sub-DRGs. Suppose clinically similar sub-DRGs have non-zero compensated cross-price elasticities while compensated cross-price elasticities for dissimilar sub-DRGs are zero or at least smaller in magnitude. Then the estimator using $\overset{\smile}{\text{inc}}^{\%}_{d,g,t}$ as the instrument is closer to the true income elasticity. This specification of the instrument also alleviates concerns of correlation between income and price shocks. While the change in own-price is included in income with weight proportional to the number of admissions in the sub-DRG, it is by construction excluded from the instrument when we remove variation from sub-DRG that are in the same CCS group. We discuss alternative strategies for excluding "similar" DRG and report additional results in Appendix A.3.

We also assess whether our estimates reflect real changes in the number of patients treated or changes in where the patients are treated. First, we consider substitution between departments within the same hospital. In the data, we observe patients treated for the same sub-DRG in multiple departments of the same hospital. Hence, there may be scope for substitution between departments. Second, we consider substitution between departments of different hospitals. Such substitution may affect the quality of care, but does not have first order effects on the aggregate costs at the system level. Hence, determining whether such substitution is driving our results is of policy relevance. In our robustness analysis, we therefore assess whether substitution between departments and hospitals is the primary source of the effects we estimate.

To focus on substitution between departments within the same hospital, for each sub-DRG we aggregate all admissions irrespective of the department within the hospital. By aggregating all admissions of a sub-DRG irrespective of the treating department, we eliminate the margin of which department treats a patient of a "shared" sub-DRG. We then assign each sub-DRG to the department within the hospital that most frequently treats these patients. We estimate our regression with admissions aggregated at the department level for these "pseudo-departments." This strategy excludes any shifting of patients between departments, unless the shifting moves a patient into a different sub-DRG that is more frequently treated by a different department. This remaining between-department shift is arguably substantially smaller.

The analysis of between-hospital substitution additionally aggregates all hospitals within the same hospital service area. A hospital service area is a geographic area based on zip codes, such that within a hospital service area, all zip codes share the same hospital as the primary provider of hospital care.[10] To avoid measuring substitution between departments within a hospital service area, we aggregate all admissions of a sub-DRG irrespective of the treating hospital and department for all patients coming from zip codes within the area. We then assign each sub-DRG to the department-type within the hospital service area that most frequently treats these patients, aggregating patients across hospitals. Since coding changes and the choice of hospital department or hospital does not affect the zip code a patient lives in, effects at this level imply changes in the aggregate number of patients from the hospital service area.

5 Results

5.1 Price and Income Elasticities

We illustrate our baseline results for the causal effect of changes in prices and expected income on quantities nonparametrically in Figures 3, 4, and 5. Table 2 summarizes these reduced form relationships as elasticity estimates and adds instrumental variables estimates of income elasticities. In Table 3, we estimate price and income effects on length of stay, hospital-acquired conditions, and in-hospital mortality rates. While these outcomes are not directly incentivized in the DRG system, they are suggestive of mechanisms and important for overall welfare implications.

Figure 3 shows the relationship between compensated price changes and quantity changes in a binscatter plot with the linear regression fit superimposed. Each observation is given by the price and quantity changes of a sub-DRG in a hospital department from one year to the next. The binscatter groups observations based on their price changes into bins of 1% width, and plots the average quantity change of the sub-DRGs as a nonparametric estimate of the relationship between prices and quantities (Cattaneo et al., 2019). Consistent with our empirical strategy outlined in Section 4, we weight observations by the number of admissions they represent and residualize prices and quantities with year and hospital department sub-DRG fixed effects. To approximate the compensated price effects, we also include our simulated instrument for income linearly in the residualizing regressions. Under our identifying assumption of no direct effects of lagged cost changes on quantity changes (see Section 4 for details), we interpret the binscatter as the causal effect of price changes on quantity changes, holding income constant. Overall, the causal effect of prices on quantity is positive: When hospitals receive larger reimbursements for a given sub-DRG, they treat more patients of it.

Figure 4 similarly summarizes the (reduced form) effect of changes in the income instrument on the number of admissions in a residualized binscatter plot. Here, the residualizing regressions include the price change of the sub-DRG linearly in addition to two-way fixed effects. Controlling for price changes avoids confounding the income effect of interest with the direct effect of the own-price. Figure 4 illustrates an overall negative relationship between changes in our instrument for income and changes in quantities. As we can see in Figure 2b, most of the observations face a residualized income shock between -1% and $+1\%$, where Figure 4 shows a negative relationship between income and quantity. It remains an interesting question for future research whether large positive income shocks have an effect that deviates from the negative linear relationship. Unfortunately, our quasi-random variation in prices does not create sufficient variation in income to credibly estimate the effects of such large shocks.

While not present for profit maximizing firms, income effects are plausible under alternative models of firm behavior such as revenue maximization or income targets (cf. Evans, 1974; McGuire and Pauly, 1991;

[10] We define hospital service areas for Germany based on the algorithm in The Dartmouth Atlas of Health Care (Wennberg, 1996) for the United States.

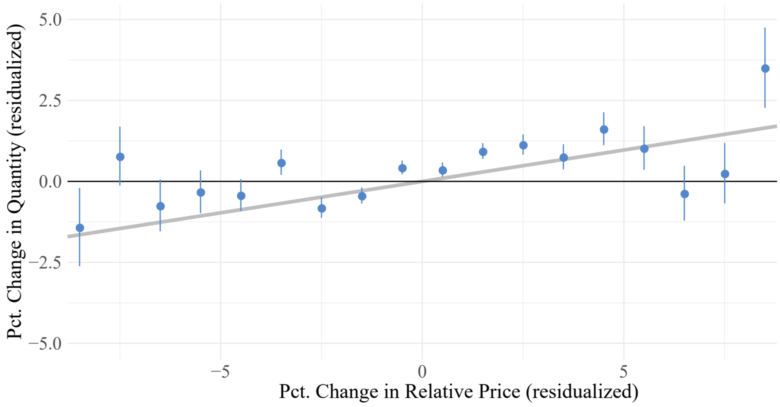


Figure 3: Residualized binscatter showing the effect of changes in relative price on changes in quantity at the department-sub-DRG-year level.

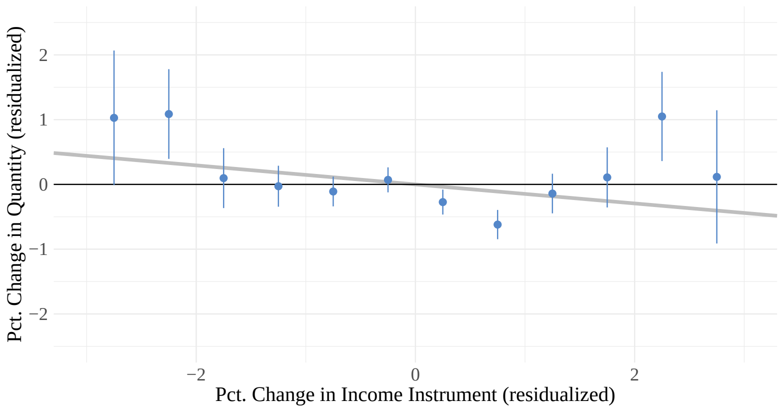


Figure 4: Residualized binscatter showing the reduced form effect of changes in simulated department income (instrument) on changes in quantity at the department-sub-DRG-year level.

McGuire, 2000, in the context of physician behavior). Since only about 30% of German hospitals are private for-profit (InEK, 2015), such alternatives are likely relevant for describing the average behavior of hospitals in our data. Income per admission is salient in DRG systems. In contrast, fixed costs are large, and variable costs are less salient and difficult to assign to particular admissions. These factors facilitate a focus on revenue, rather than profits, for hospitals. Additionally, income targets were explicitly mentioned by some department heads we interviewed, and are consistent with negative income effects: When income is exogenously *reduced*, hospital departments need to treat *more* patients to meet their income targets, holding prices fixed. Since actual income is not known to hospitals until after all admissions have been realized, we can interpret the reduced form effect of the instrument on the outcome as the effect of *expected* changes in income. Hospitals can compute similar metrics to our instrument for their internal planning after price changes are announced but before they observe future admissions.

With a positive compensated price effect and a negative income effect, as well as possible substitution and complementarities between sub-DRGs, the effect of an overall increase in the price level is a priori ambiguous. The binscatter in Figure 5 suggests a small positive effect of the average price level on aggregate admissions. To identify net aggregate effects, we aggregate admissions and prices at the department-year level. Consistent with the previous figures, we residualize the percentage change in admissions and average price using two-way (hospital department and year) fixed effects. At this level of aggregation of the outcome variable, our income instrument is best interpreted as an uncompensated change in the average price; that is, the effect of a change in all prices, inclusive of any resulting income

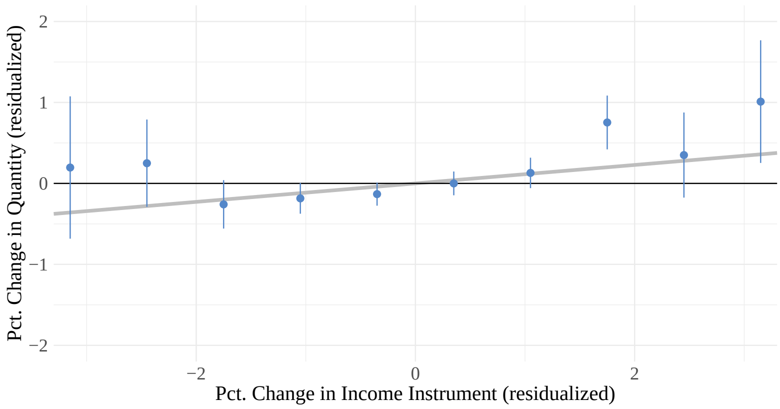


Figure 5: Residualized binscatter showing the reduced form effect of changes in simulated average price (instrument) on changes in department quantity at the department-year level.

effects. Figure 5 illustrates that, as the average price received by hospital departments increases, they treat more patients overall.

We present our baseline estimates of the price and income elasticities of quantity in Table 2. Columns (1) and (3) show the effect on the quantity of a given sub-DRG when the price for that sub-DRG or the overall income of the hospital department are changed. Column (1) shows the reduced form effect of the simulated instrument, while in column (3) we use it to instrument for actual department income. Since the first stage coefficient on the income instrument is close to 1 (column 4), the two columns are similar. The estimated price elasticity is positive and close to 0.2. If the relative price for a sub-DRG increases by 1%, hospitals treat 0.2% more patients of that sub-DRG. The income elasticity is negative and approximately -0.15. If a hospital receives a positive income shock without an increase in the reimbursement for a sub-DRG, the number of patients treated of the given sub-DRG decreases. Column (2) shows that the hospital response is strongest (largest in absolute value) for the *smallest* hospitals. The smallest hospitals tend to be highly specialized, which may leave them better positioned but also in more need to optimize their response to price and income changes, compared to large hospitals offering services across the full DRG-spectrum and thereby less affected by changes in reimbursements for individual DRG. Hospital size may also be a rough proxy for ownership type, with the smallest most specialized hospitals often being private, and the largest hospitals being (public) university hospitals in Germany. Previous work focusing on the United States (for instance, Duggan, 2000; Silverman and Skinner, 2004; Dafny, 2005) has connected for-profit hospitals to stronger responses to a variety of financial incentives.

The remaining columns of Table 2 show the effect of a uniform increase in prices on aggregate quantities. As shown in columns (5) and (6), increasing the prices for all patients treated in a hospital department by 1% causes an increase in quantity of the department of about 0.1%. The overall budget impact of a 1% increase in all prices is therefore approximately a 1.1% increase in expenditure. For these two regressions, we aggregate the data at the department level, such that the resulting effect is net of possible substitution in treatment (treating more patients of one sub-DRG but fewer of another) or nominal reporting changes. If the average price (and hence income) for the services a department provides increases, the department treats more patients. The reduced form elasticity in column (5) most closely captures the effect of increasing all reimbursements by a fixed percentage. The IV estimate in column (6) instead scales the elasticity by the implied change in actual income. Since the first stage coefficient is again close to 1 (column 7), both elasticities are similar in magnitude.

While DRG systems most directly incentivize responses in quantities, the quality of care is of substantial interest in health care settings as well. Here, we focus on three margins to measure how

Table 2: Compensated price and income elasticities at the sub-DRG level and uncompensated average price elasticity at the department level.

	All Variables in Percentage Changes						
	Diagnosis-Group Level				Department Level		
	(1) (OLS)	(2) (OLS)	(3) (IV)	(4) (First Stage)	(5) (OLS)	(6) (IV)	(7) (First Stage)
Outcome Variable:	Quantity	Quantity	Quantity	Income	Quantity	Quantity	Income
Relative Price	0.194^{***}	0.215^{***}	0.197^{***}	0.018^{***}			
	(0.016)	(0.019)	(0.017)	(0.005)			
Relative Price × Hospital Size		-0.032^{*}					
		(0.018)					
Department Income (instrument)	-0.147^{**}	-0.207^{***}		0.917^{***}	0.114^{***}		0.877^{***}
	(0.057)	(0.061)		(0.044)	(0.029)		(0.034)
Dept. Income (instr.) × Hospital Size		0.103^{**}					
		(0.049)					
Department Income (actual)			-0.160^{**}			0.130^{***}	
			(0.065)			(0.030)	
Observations	650,890	650,890	650,890	650,890	43,950	43,950	43,950

Notes: All variables are in year-on-year percentage changes as described in Section 4. In the interaction terms, denoted by "× Hospital Size," Hospital Size refers to Percent Difference between the size of the hospital and the size of the median hospital, such that this variable equals 0 for the median hospital and 1 for a hospital twice the size of the median hospital. All specifications include cross-sectional fixed effects (department-sub-DRG for columns (1) – (4), department for columns (5) – (7)) and year fixed effects. Robust standard errors, clustered at the level of treatment assignment (department-year and sub-DRG for columns (1) – (4), department-year for columns (5) – (7)), are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

price and income incentives affect how patients are treated: average length of stay (LOS), the fraction of patients with hospital-acquired conditions (HAC),[11] and in-hospital mortality rates.[12] Below, we highlight key institutional features that can explain our findings. Note, however, that the estimated effects can reflect both selection and changes in outcomes; our variation in prices and income does not allow us to disentangle these.

We show the results for these measures of intensity and quality in Table 3. Throughout, we estimate *semi-elasticities*; that is, we estimate the average level change in LOS (in days), the change in the probability of an admission developing a HAC, and the change in the mortality rate, in response to a 1% increase in price or department income. Results for elasticities (not reported), that is the percentage effects on the average LOS, fraction of admissions with HAC, and mortality rate, are qualitatively similar.

In column (1) of Table 3, we find a negative effect of department income on average length of stay. This result is consistent with hospital capacity, given by the number of hospital beds, being sticky.[13] When hospitals operate at close to full capacity, the positive effect on *overall* department quantity we find in columns (5) and (6) of Table 2 implies that each patient stays in the hospital for a shorter period of time. This allows the hospital department to treat more patients despite a fixed capacity. Even when a hospital department is not at full capacity, doctors may have a preference to retain a minimum number of open beds for emergency or other unplanned stays, similarly creating downward pressure on the average length of stay.

We estimate a positive price effect on the average length of stay. This may pick up that the surcharges and discounts for extreme length of stay are related to prices. Hence, for some admissions close to a LOS surcharge threshold, an increase in price creates a direct incentive to keep the patient in the hospital for

[11] We infer hospital-acquired conditions from diagnosis codes, see Appendix A.1 for details.

[12] In contrast to Medicare data in the United States, (30-day) readmissions are not identified in German reimbursement data.

[13] We do not directly observe the number of hospital beds in our data. However, German hospitals have limited autonomy since capacity is typically set by state governments (Geissler et al., 2011).

Table 3: Reduced form compensated price and income semi-elasticities of treatment and quality-related outcomes at the sub-DRG level.

Outcome Variable:	avg. LOS	HAC prob.	Mortality
	(1)	(2)	(3)
Relative Price	0.289***	0.014**	0.002**
	(0.012)	(0.003)	(0.001)
Department Income (instrument)	−0.236***	0.030**	−0.001
	(0.042)	(0.013)	(0.002)
Observations	650,890	650,890	650,890

Notes: The dependent variables are year-on-year changes in levels (column 1) and changes in the fraction of admissions (columns 2 and 3), while the independent variables are year-on-year percentage changes. LOS refers to "length of stay," HAC to "hospital-acquired conditions," and mortality refers to in-hospital mortality. All specifications include cross-sectional fixed effects (department-sub-DRG) and year fixed effects. Robust standard errors, clustered at the level of treatment assignment (department-year and sub-DRG), are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

additional nights to gain the surcharge for additional days. Similarly, for patients near the threshold for low LOS discounts, an increase in price creates a larger incentive to avoid the discount for short LOS. Alternatively, if doctors consider the fixed DRG-reimbursements per patient as a "budget" that can be spent on care for that patient, an increase in this budget may be spent at least in part on additional length of stay.

We show the estimated effects of price and income changes on the fraction of patients with hospital-acquired conditions in column (2) of Table 3. Increases in price and income lead to a larger fraction of patients experiencing HACs. Note that our measure of hospital-acquired conditions includes in particular nursing sensitive outcomes; that is, conditions that can oftentimes be prevented with sufficient nursing care. The results are therefore consistent with not just hospital beds, but also the number of hospital staff, including nurses, being difficult to adjust in the short-run. When a regulator-induced one-year increase in the average price leads to an increase in admissions (columns (5) and (6) of Table 2), departments may be unable or unwilling to respond by hiring additional staff. Germany has experienced staff shortages both for clinicians as well as nursing staff for many years. The patient-to-nurse ratio, for instance, equals 13 to 1 in Germany, compared to 5.3 to 1 in the United States (Simon and Mehmecke, 2017). Furthermore, German labor laws may make it unattractive for hospitals to hire staff in response to short-term (yearly) changes in prices and income.

Our results on in-hospital mortality rates, shown in column (3) of Table 3, are more mixed. Only the price effect is statistically significant: We estimate that a 1% increase in the price for a sub-DRG causes an increase in mortality rate within the sub-DRG of 0.2 percentage points. While the effect is precisely estimated, it is substantively close to 0, and may reflect differences in clinical condition between marginal patients whose sub-DRG may be affected by prices, and average patients (selection effects). The estimated income effect is slightly negative and not statistically significant. Taken at face-value, an income effect of −0.001 on the mortality rate would imply a cost of around €35,000 per life saved.[14] Overall, the results on mortality suggest relatively small and ambiguous effects. While the overall increase in work load due to income increases appears to cause a substantial increase in hospital-acquired conditions, potentially due to limited staffing, quality management appears to be sufficient to prevent increases in mortality of similar magnitude.

[14]Based on an average price of €3,500, a 1% increase in income through prices is approximately €35 per patient. If it reduces the probability of death by 0.1 percentage points, such that the cost per life saved is $\frac{€35}{0.001 \text{ lives}} = €35{,}000$ per life.

5.2 Heterogeneity

Many hospitals provide a diverse set of medical services, including emergency care and elective procedures. While the preceding analysis pools all DRGs that are reimbursed through the German DRG system to provide precisely estimated average effect sizes, researchers and policy-makers also have substantive interest in the heterogeneity in the estimated effects. We therefore provide elasticity estimates for clinically defined subgroups. In Section 6, we show that these elasticities are relevant in the context of optimal hospital reimbursement. Since these heterogeneous elasticities may also be of interest and use to researchers for other analyses, we provide a complete listing of the estimates in an Online Appendix.

To estimate heterogeneous effects, we run separate regressions on subsamples defined by the CCS groups. Specifically, each subsample corresponds to one CCS group and consists only of observations belonging to sub-DRGs assigned to that CCS group as described in Section 3.3. The regression specifications are unchanged from the previous section. The full results are shown in the Online Appendix and also available to download as spreadsheets.

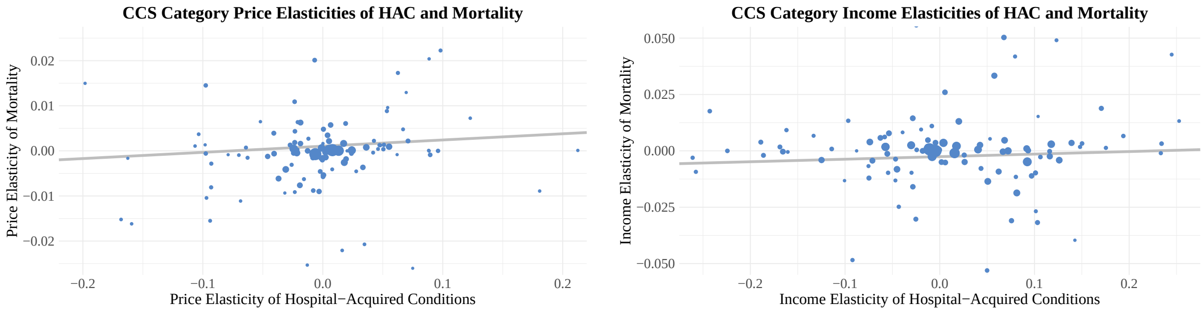
We can validate the estimated patterns of heterogeneity by comparison to the existing literature. For instance, Birkmeyer et al. (2013) identify surgical removal of early-stage prostate cancer as a preference-based treatment with large geographic variation in treatment rates. Indeed, we find a particularly large positive price elasticity for prostate cancer treatment (CCS category 29) of 0.48 (s.e. 0.17) and large negative income elasticity of -2.3 (s.e. 0.8), in line with the suggested electiveness of some of the associated treatment options. In contrast, treatment for acute myocardial infarction (CCS category 100) predominantly requires acute care (cf. Jena et al., 2015). Neither the estimated price elasticity (-0.15, s.e. 0.11) nor the estimated income elasticity (-0.29, s.e. 0.43) is statistically significantly different from zero.

The heterogeneous elasticities allow us to address responses in the quality of care in more detail. Specifically, for many DRGs mortality is an extreme and unlikely outcome irrespective of most physician behavior. The near-zero and statistically insignificant estimated average effect may therefore mask meaningful mortality effects in some subgroups. Similarly, while we attempt to measure diverse hospital-acquired conditions, the rates at which these conditions occur vary substantially across sub-DRG. Hence, there may be a sub-DRG for which these conditions are less relevant, such that any effects for them are likely to be zero, and which moves the *average* effect towards zero.

Figure 6 shows that CCS groups where we estimate larger price and income semi-elasticities of hospital-acquired conditions tend to also be characterized by larger effects on the mortality rates. In panel (a), we show a scatterplot of the estimated price semi-elasticities of hospital-acquired conditions and mortality rates, with each point corresponding to one CCS group. Panel (b) repeats the same figure but with income elasticities rather than price elasticities. The size of each point is inversely proportional to the product of standard errors of the two estimated elasticities, such that CCS groups where we estimate the elasticities most precisely are more prominent visually. While the estimated mortality effects are all close to zero (with the exception of some imprecise estimates), mortality effects are larger for those CCS groups where we also estimate a larger effect on hospital-acquired conditions. There is no direct or mechanical relationship between our definitions of hospital-acquired conditions and mortality, such that the relationship is unlikely to purely be driven by the covariance of estimation errors. Independent sampling variation in the estimated coefficients instead likely attenuates the relationships due to measurement error.

In Section 6, we show how our estimated elasticities may enter into a new model of optimal reimbursements in DRG systems.[15] Our model features a social planner who sets prices to maximize social welfare,

[15] We note the close conceptual and analytical connection between optimal DRG prices and optimal linear commodity taxation (Ramsey, 1927; Diamond, 1975; Mirrlees, 1976).



(a) Price Elasticities

(b) Income Elasticities

Figure 6: Scatter plot of estimated CCS group level elasticities of hospital-acquired conditions and mortality. Each point corresponds to one CCS group, with larger points corresponding to smaller standard errors of the estimates. For both the price elasticities (top figure) and income elasticities (bottom figure), CCS groups where we estimate a larger response of hospital-acquired conditions also show a larger response of the mortality rate.

taking into account that the representative hospital responds to price and income changes. The model improves upon the model of "yardstick competition" (Shleifer, 1985), which is often cited as motivation for setting prices equal to average costs in DRG systems, by flexibly allowing for (compensated) own-, cross-price, and income elasticities of quantities and qualities. The estimated elasticities are part of the reduced form sufficient statistics (cf. Chetty, 2009) of this model.

In our model, optimal prices are such that an "index of encouragement" (cf. Mirrlees, 1976) is constant across DRGs: The ratio between the approximate net welfare benefit of setting the optimal price for a particular DRG compared to a 0 price, and the budget share of the same DRG, is constant. This also implies that the marginal social welfare of treating patients is not equalized across DRGs in general. For instance, if the price of a particular DRG leads to socially undesirable responses for other DRGs (such as upcoding, which is financially costly to the social planner and potentially without benefits to the patient), its price may remain relatively low compared to the marginal benefit of treating patients in the DRG. Such across-DRG considerations are currently absent from price setting, at least in the German DRG system.

5.3 Assessing the Identifying Assumption

The exclusion restriction states that two-year lagged cost changes only affect the outcome through prices and income. In this section, we assess the plausibility of this assumption. In summary, since one-year lagged cost changes (the lead of price and income changes) do not have a large direct effect on the outcome, two-year lagged cost changes also plausibly do not have a large direct effect (other than through prices and income) on the outcome. Lagged cost changes, after the demeaning of our regression specifications, do not have direct effects on the outcome because they are not predictive of present cost changes, which do have an effect on the outcome.

The regression specifications in Table 4 parallel the specification of our baseline regression in column (1) of Table 2. The costs used in the regressions in Table 4 are average reported costs that are collected and published ex-post by the regulator for use in price setting two years later. Similar to our instrument for income (Department Income$_t$ (instrument)), the Aggregate Cost$_t$ (instrument) variable also aggregates sub-DRG-specific cost changes using $t-2$ admissions.

In the first column of Table 4, we control for current changes in costs and aggregate costs. This addresses concerns that price, the two-year lag of cost changes, is correlated with current cost changes, which are likely to have a direct effect on current changes in the outcome variable. As expected, the coefficient on current cost changes is negative, -0.047, such that relatively higher treatment costs are associated with fewer treated patients. The estimated compensated price elasticity, 0.187, and income elasticity, -0.147, are largely unchanged from their values in Table 2 because the correlation between two-year lagged percentage changes and current percentage changes is negligible in our fixed effect specifications. See Appendix A.2 for details on the relationship between current and lagged cost changes. Our estimates of price and income elasticities therefore appear not to be biased by correlation with current changes in costs.

The magnitude of the price elasticity is more than three times the coefficient on the percentage change in present costs, Cost_t, in column (1) of Table 4. While we do not have an instrument to identify the causal effect of cost changes, we find two explanations particularly credible: First, regulatory price changes affect all hospitals directly, while changes in average reported costs for a DRG may only weakly correlate with changes in actual costs of individual hospitals. We expect this to hold in particular for hospitals which did not report costs themselves. The coefficient on Cost_t may then be attenuated due to this "measurement error," relative to the elasticity of admissions with respect to actual cost of treatment in the particular hospital department. Second, changes in prices are substantially more salient than changes in costs. While prices are published ex-ante by the regulator, the ex-post accounting for cost per admission at the DRG level is based on a variety of indirect measures such as staffing costs and nurse-to-patient ratios. Hence, even if we had a precise measure of (and instrument for) actual cost, hospitals may still respond more to changes in prices than to changes in costs due to salience at the time of decision-making. When *aggregating* the sub-DRG changes in prices and costs, however, hospitals appear to respond similarly to changes in aggregate income, -0.147, and changes in aggregate costs, 0.170. The averaging across sub-DRGs likely averages out some part of the measurement error, and changes in aggregate costs are likely to be more salient. Hence, the responses to changes in prices and costs may in principle be of the same magnitude.

We can use the estimates in column (2) to assess a plausible upper bound on the biases due to direct effects of the two-year lag of changes in costs. For this analysis, we use the one-year lag of costs (the lead of prices) as a "placebo treatment" holding actual prices and income fixed. Controlling for changes in prices and income as well as current changes in costs and aggregate costs, the direct effects of the one-year lag of changes in costs, 0.057, and aggregate costs, -0.037, are relatively small in magnitude. Compared to the cost changes of year $t-1$, cost changes of year $t-2$ are arguably likely to have substantially smaller direct effects on outcome changes in year t. Hence, we expect the direct effect (not through prices) of cost changes of year $t-2$ to be substantially smaller (in absolute value) than 0.057, for a price elasticity between 0.15 and 0.25. Similarly, we expect the direct effect (not through income) of changes in aggregate costs of year $t-2$ to be substantially smaller (in absolute value) than -0.037, suggesting a bias-adjusted income elasticity estimate would be between -0.1 and -0.2.

In columns (3) and (4), we check whether our "treatments," prices and income, have direct effects on a "pseudo-outcome" (Athey and Imbens, 2017), which they are known not to affect. In our context, we test whether there are anticipatory effects. Since price (and expected income) changes for year t are unknown until the fall of year $t-1$ and do not come into effect until January of year t, effects on the outcome in year $t-1$ would be suggestive of violations of the identifying assumption that there are no effects of lagged cost changes through other channels. We therefore check whether there are anticipatory effects of the lead of price and income changes. Column (3) controls for current cost changes as in column (1) of Table 4, while column (4) corresponds to our baseline specification (column (1) of Table 2) without this control. In both regressions, the coefficients on "Relative Price_{t+1}" and "Department Income_{t+1}" are small in magnitude and not statistically significant at the 5% level, suggesting no anticipatory effects of the treatment on the outcome. To summarize, because the price changes in year $t+1$ (cost changes in $t-1$) do not affect changes in the outcome variable in year t, but price changes in year t (cost changes in

Table 4: Estimating the coefficients on sub-DRG price and income changes at different times.

	All Variables in Percentage Changes Outcome Variable: Quantity$_t$			
	(1)	(2)	(3)	(4)
Relative Price$_t$	0.187***	0.196***		
	(0.016)	(0.017)		
Department Income$_t$ (instrument)	$-$0.147**	$-$0.150**		
	(0.057)	(0.059)		
Cost$_t$	$-$0.047**	$-$0.037**	$-$0.053***	
	(0.019)	(0.019)	(0.019)	
Aggregate Cost$_t$ (instrument)	0.170***	0.166***	0.165***	
	(0.059)	(0.060)	(0.060)	
Relative Price$_{t+1}$		0.057***	0.021	0.027*
		(0.016)	(0.015)	(0.015)
Department Income$_{t+1}$ (instrument)		$-$0.037	$-$0.025	$-$0.051
		(0.062)	(0.059)	(0.058)
Observations	650,890	650,890	650,890	650,890

Notes: Year t price and income changes reflect cost changes in year $t-2$, and year $t+1$ price and income changes reflect cost changes in year $t-1$. Coefficients for the price and income changes in year t are largely unchanged from Table 2 column (1) when controlling for current cost changes, and the one-year lead of prices and income has little direct effect on the outcome. All Department Income (instrument) and Aggregate Costs (instrument) variables, irrespective of leads, aggregate sub-DRGs based on admission counts in year $t-2$. All variables are in year-on-year percentage changes. All specifications include cross-sectional fixed effects (department-sub-DRG) and year fixed effects. Robust standard errors, clustered at the level of treatment assignment (department-year and sub-DRG), are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

$t-2$) do have an effect, this latter effect is likely only through prices and income.

5.4 Robustness to Alternative Specifications and Interpretations

We address the robustness of our estimated price and income elasticities to alternative specifications and interpretations in Table 5.

Since hospitals self-select into reporting costs on a yearly basis, some hospitals may strategically choose to submit cost reports to influence prices. In supplementary analyses (see Appendix A.2.4), we show that adding even a few hospitals to the set of cost-reporting hospitals does affect prices. To rule out that our estimates are dominated by strategic behavior of cost-reporting hospitals, column (1) of Table 5 drops all hospitals from the sample that ever signed an agreement to become eligible for cost reporting. For the sample of non-reporting hospitals, the price elasticity remains mostly unchanged, while the income elasticity becomes slightly more negative, but within one standard error of the original estimate in Table 2. Since the cost-reporting hospitals are not fully representative of the population of German hospitals (InEK, 2015), small differences in coefficients may also be due to compositional differences in the samples.

In column (2) of Table 5, we present results using an unbalanced panel. Most of the additional regression observations in the unbalanced panel represent two or fewer patients in a sub-DRG-department-year cell. For these cross-sectional units, there are some years without any such patient leading to their exclusion from the balanced panel. Since we weight observations by the number of admissions they represent, they receive little weight even in the unbalanced panel. Entry and exit of entire hospital departments, however, may be correlated with price and income shocks, and estimates from the unbalanced panel and balanced panel may both be biased (Olley and Pakes, 1996). Intuitively, the departments (and sub-DRGs) that are most responsive to changes in prices and income may also be more likely to open or close, such that they are disproportionately dropped from the balanced panel. In practice, since the hospital market in Germany is heavily regulated by state governments (Geissler et al., 2011), the resulting biases are likely minor. We restrict the unbalanced panel to treat entry and exit of hospital departments and services symmetrically. Since the income instrument requires a two-year lag of quantities, hospital departments can only appear in the regression two years after first opening. We therefore automatically exclude rapid growth from the first to second year of operation if the department was only open for part of its first year. To symmetrically avoid biases due to mid-year closure of departments, we require that departments remain open for another year. Since we cannot determine with certainty whether departments remained open past the final year of our sample, 2016, we also drop that year from the regression for all hospitals. The estimated price and income elasticities for the unbalanced panel in column (2) are within one standard error of our baseline estimates, but slightly larger in magnitude.

We check whether the estimated income elasticity is driven by direct effects of prices for other sub-DRGs, such as upcoding, in column (3) of Table 5. When hospitals engage in upcoding behavior, an increase in the price for other sub-DRGs can lead to a decrease in admissions of a given sub-DRG, as relatively more patients are coded into the sub-DRGs with price increases. Since our instrument for department income aggregates the price changes of all sub-DRGs, such effects may bias the baseline income elasticity estimate. However, upcoding is unlikely to affect the more general medical condition or disease classification of the patient. In column (3), we therefore create an alternative instrument for income that for each sub-DRG only aggregates the price shocks of clinically *dis*similar sub-DRG. Specifically, for an observation of sub-DRG g, we only aggregate sub-DRGs g' with g' assigned to a different CCS group than g. This eliminates any bias due to cross-price effects, such as upcoding, of sub-DRGs within the same CCS group. The point estimate for the income elasticity, -0.141, is close to the baseline income elasticity of -0.160 in column (3) of Table 2 in the analogous instrumental variables regression specification. In Appendix A.3, we show that excluding price changes of sub-DRGs within more aggregate CCS categories or based on various measures of regulatory similarity yields similar estimates.

The interpretation of changes in the number of patients within a hospital department in response to price and income shocks fundamentally depends on whether these patients are alternatively treated in different hospitals or departments, or are patients that otherwise do not receive inpatient care. In many hospitals, patients of a particular sub-DRG can be treated in different departments. The positive average price elasticity at the department level in column (5) of Table 2 may therefore reflect that departments with a positive price shock tend to treat a larger share of patients in these "shared" sub-DRGs.

For column (5) of Table 5, we therefore assign within each hospital all patients of a sub-DRG to the hospital department that most frequently treats such patients. As a result, when a positive average price shock causes an increase in admissions for such an artificial hospital department, there cannot be any offsetting reduction in admissions of the same sub-DRG in other departments of the same hospital. The point estimate estimate, 0.197, suggests that the increase in admissions we document in response to increases in average prices is not offset by corresponding reductions in admissions of the same sub-DRGs in other departments of same hospital.

The estimate in column (6), which eliminates substitution between hospitals, is similar to previous estimates, suggesting that the effect is also not offset by responses in other hospitals. For this analysis, we aggregate patients by the hospital service area they live in and the type of specialized department their sub-DRG is most frequently treated by, rather than by the actual hospital and department of the patient. We find that when the average price faced by the departments in this hospital service area increases, the aggregate number of patients from the underlying zip codes increases. To summarize, increases in

Table 5: Robustness of price and income elasticities to alternative specifications.

	All Variables in Percentage Changes					
	Diagnosis-Group Level				Department Level	
	(1) (OLS)	(2) (OLS)	(3) (IV)	(4) (First Stage)	(5) (OLS)	(6) (OLS)
Outcome Variable:	Quantity	Quantity	Quantity	Income	Quantity	Quantity
Relative Price	0.187***	0.204***	0.195***	0.062***		
	(0.023)	(0.013)	(0.017)	(0.005)		
Department Income (instrument)	−0.208***	−0.193***		0.815***	0.197***	0.245***
	(0.077)	(0.044)		(0.043)	(0.028)	(0.024)
Department Income (actual)			−0.141***			
			(0.064)			
Observations	329,370	1,548,471	650,890	650,890	47,310	73,150

Note: Column (1) restricts the sample to hospitals that never signed an agreement to report costs. Column (2) is based on an unbalanced rather than a balanced sample. For column (3), the income instrument excludes price variation due to sub-DRGs of the same CCS group to limit bias due to direct price effects, such as upcoding. Since excluding variation from the instrument weakens the first stage and thereby attenuates reduced form estimates even in the absence of endogeneity, we present instrumental variables estimates that effectively scale estimates equally (to the effect of actual income) irrespective of instrument strength. Column (4) shows the first stage corresponding to column (3). For column (5), we fix the assignment of sub-DRGs to hospital departments to estimate elasticities net of offsetting responses in other departments in the same hospital. For column (6), we aggregate admissions at the hospital service area and department-type level to estimate elasticities net of offsetting responses in other hospitals. All variables are in year-on-year percentage changes. All specifications include cross-sectional fixed effects (department-sub-DRG for columns (1) – (4), department for columns (5) – (6)) and year fixed effects. Robust standard errors, clustered at the level of treatment assignment (department-year and sub-DRG for columns (1) – (4), department-year for columns (5) – (6)), are shown in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

admissions of a hospital department in response to increases in average prices and income appear not to be offset by decreases in admissions for the same sub-DRGs in other departments or hospitals.

6 A Model of Optimal Reimbursements

Price and income elasticities are relevant in the context of optimal reimbursements. Here, we sketch a simple deterministic model that draws from the rich literature on optimal taxation and sufficient statistics. Notably, this model optimizes reimbursements within a fixed DRG system. That is, a constant price is paid for each patient treated of a given DRG. Analytically, the problem of setting optimal reimbursements under the constraints of a DRG system resembles the optimal linear commodity taxation of Ramsey (1927).

A representative hospital maximizes some, potentially unknown, utility function by treating $q \in \mathbb{R}^G$ patients in G DRGs with effort $e \in \mathbb{R}^K$, measured for instance as mortality rate per DRG (such that $K = G$), and generating income Z. In a DRG system with prices $p \in \mathbb{R}^G_+$ and lump sum subsidy $L \in \mathbb{R}$, the hospital chooses quantity and effort, (q, e), as a function of price and subsidy (p, L)

$$\begin{aligned} (q^*(p, L), e^*(p, L)) \equiv \arg\max_{q,e} u(q, e, Z) \\ \text{s.t. } p \cdot q + L = Z \end{aligned} \tag{4}$$

where we abstract away from integer constraints on quantities q.

In our sufficient statistics approach, we are agnostic about hospital behavior. As a special case, the representative hospital may be profit maximizing, that is $u(q, e, Z) = c(q) - Z$, where $c(q)$ is the cost of treating q patients. Furthermore, we take a reduced form approach to the decisions that lead to the treatment of q_g^* patients of DRG g. The optimal choice of treated patients by the hospital, q^*, is the vector of reported admissions including any upcoding behavior by hospitals.

Within the constraints of a DRG system, the social planner sets prices p optimally given a budget constraint B, but cannot make lump sum subsidies:

$$\begin{aligned} p^* \equiv & \arg\max_p S(q^*(p,0), e^*(p,0)) \\ & \text{s.t. } B = p \cdot q^*(p,0) \end{aligned} \tag{5}$$

where S is a social welfare function. The budget constraint may be given by the income of the health insurance fund. In our setting, the German regulator each year sets prices such that aggregate expenditure adjusted for inflation is constant from one year to the next in the absence of strategic hospital behavior and stochastic shocks. In the deterministic model above, the social planner instead takes into account the effects that prices have on hospital behavior to meet the budget constraint exactly.

Our model is based on the Ramsey (1927) model of optimal linear commodity taxation, but adds effort e and a distinction between social welfare S and the indirect utility of the hospital. In a DRG system, the reimbursement paid to the hospital per patient is constant within DRG, similar to the linear price plus tax in the Ramsey (1927) model of commodity taxation. In our model, the commodities are DRGs, the representative individual is the hospital, and the linear taxes we optimize over are fixed reimbursements per patient in a DRG, with the revenue raising constraint replaced by a budget constraint. To allow for, for instance, investments in quality or cost reduction (cf. Shleifer, 1985), we add "untaxed" effort e to this model. In our setting, the assumption that a social planner chooses reimbursements to maximize the indirect utility function of the representative hospital is less attractive than in an optimal taxation setting, however. For instance, hospitals may perform treatments that are profitable due to large reimbursements but too costly to improve social welfare with a finite budget if there are alternative treatments that are less profitable but also less costly.

The solution to the optimization problem 5 characterizes optimal prices. The optimal prices p and Lagrange multiplier λ are the solutions to the G equations

$$\sum_{g'=1}^{G} \Big(\frac{\partial S}{\partial q_{g'}} \frac{1}{p_{g'}} - \lambda \Big) \tilde{\epsilon}^q_{g,g'} + \sum_{k=1}^{K} \frac{\partial S}{\partial e_k} \tilde{\epsilon}^e_{k,g} \frac{1}{p_g q_g^*} = -\theta \tag{6}$$

and the budget constraint in 5; see Appendix B for a detailed derivation. For ease of presentation, we suppress functional dependence of all objects on p and $L = 0$. Here, $\tilde{\epsilon}^q_{g,g'}$ is the *compensated* elasticity of the number of patients in DRG g, q_g, with respect to the price for DRG g', $p_{g'}$. For effort e_k in dimension k, $\tilde{\epsilon}^e_{k,g}$ is the compensated *semi*-elasticity; that is, the change in effort e_k in response to a (compensated) 1 percentage point change in the price of DRG g. The Lagrange multiplier λ is the marginal social benefit of increasing the budget for reimbursements by 1 unit. Its sign is not restricted: If marginal spending is socially wasteful at given prices p, then $\lambda < 0$ for these prices. The number of patients treated by the hospital is q^*, chosen optimally in response to prices p and lump sum transfer L in accordance with the hospital optimization problem in 4. Hence, $p_g q_g^*$ is the aggregate spending on DRG g.

We can interpret the parameter θ, defined as

$$\theta \equiv \sum_{g=1}^{G} \frac{\partial S}{\partial q_g} \frac{\partial q_g^*}{\partial L} + \sum_{k=1}^{K} \frac{\partial S}{\partial e_k} \frac{\partial e_k^*}{\partial L} - \lambda \Big(1 + \frac{\partial \big(\sum_{g=1}^{G} p_g q_g^* \big)}{\partial L} \Big),$$

as the marginal value of introducing a 1 unit lump sum transfer to the hospital. The marginal benefit due to inducing socially desirable responses in quantities and effort are measured by the first two terms, respectively. The final term measures the change in welfare due to changes in the budget, valued at the marginal value of additional budget, λ. The introduction of a 1 unit lump sum transfer creates a mechanical 1 unit change in the budget, which can be offset or exacerbated by behavioral responses in aggregate spending on reimbursements, $\sum_{g=1}^{G} p_g q_g^*$, to a change in the lump sum transfer, L.

By making simplifying assumptions to gain further intuition, we see that the marginal increase in social welfare of treating an additional patient is not constant across DRGs. Specifically, assume that there are no effects of prices on effort, and that compensated cross-price elasticities of quantities are zero. These assumptions set $\tilde{\epsilon}^e_{k,g} = 0$ for each effort metric k and the price of each DRG g, and $\tilde{\epsilon}^q_{g,g'} = 0$ for all DRGs g and g' with $g \neq g'$. Then optimal prices in equation 6 satisfy

$$p_g = \frac{\frac{\partial S}{\partial q_g}\tilde{\epsilon}^q_{g,g}}{\lambda\tilde{\epsilon}^q_{g,g} - \theta}.$$

Suppose, additionally, that price elasticities are constant, such that $\tilde{\epsilon}^q_{g,g} = \tilde{\epsilon}^q$ for each DRG g. Then the ratio of price, p_g, and marginal social welfare of treating an additional patient, $\frac{\partial S}{\partial q_g}$, is constant across DRGs. Hence, DRGs with a higher optimal price have larger marginal social welfare of treating additional patients. This result arises because generating the additional patient to realize the marginal welfare benefit requires a larger increase in aggregate spending if the initial price is high, when elasticities are constant in prices.[16] Empirically, the linear relationship between price changes and quantity changes in Figure 3 suggests that (income-compensated) price elasticities are indeed approximately constant around the prices set by the German regulator.

Instead, the first order condition (equation 6) implies that an "index of encouragement" is constant across DRGs.[17] This index captures the (approximate) net welfare benefits of using the price p_g of DRG g to encourage quantities $q_{g'}$ of each DRG g' and effort e_k for $k = 1, \ldots, K$, relative to the budget share of DRG g. The derivation and formal definition of the index of encouragement based on the left-hand-side of equation 6 is given in Appendix B.3. The net welfare effect considers the total effect of the optimal price p_g on welfare S, compared to a price of $p_g = 0$, holding prices p_{-g} fixed and using linear approximations to extrapolate from the optimum. It is net of the total budgetary cost of p_g through its encouragement of quantities of each DRG g', valued at the marginal welfare value of the budget constraint, λ. Since the right-hand-side in equation 6, $-\theta$, is constant across DRGs g, it follows that the index of encouragement is constant.

The reduced form income-compensated price and income elasticities are part of the sufficient statistics (Chetty, 2009) of our model of optimal reimbursements. However, in addition to own-price and income elasticities, the optimal prices in equation 6 also depend on cross-price elasticities and the effects of quantities and effort on social welfare. We can estimate cross-price elasticities with the same exogenous variation in prices used in the present paper. The primary challenge to estimation is the large number of such elasticities, requiring a principled approach to statistical regularization in estimation. In principle, estimating these elasticities resembles demand estimation with many goods, a familiar problem in industrial organization and for retailers such as Amazon experimenting with and setting prices to increase profits.

The effects of quantities and effort on social welfare fundamentally depend on how society trades off the benefits and costs of treatment. In the analytically similar setting of optimal taxation (e.g. Ramsey, 1927), such issues are partially resolved by setting social welfare equal to the utility of the representative household. A similar approach is analytically tractable but conceptually less attractive in the hospital setting. Instead, we may posit a social welfare function that takes the difference between the monetary

[16] In a model with elasticities constant in prices, the increase in aggregate spending needed to cause an infinitesimal unit increase in quantity is p/ϵ units, where ϵ is the price elasticity of quantity.

[17] We name this term in analogy to the "index of discouragement" (cf. Mirrlees, 1976) that is derived similarly in the Ramsey model of optimal taxation.

value of health benefits or improvements in quality adjusted live years and the cost of treatment for marginal patients who are otherwise untreated. A recent literature studies the relative efficiency of different treatments for the same patient and resulting inefficiencies across hospitals (cf. Chandra and Staiger, 2007, 2020; Doyle et al., 2015). Their methods and frameworks may generally be informative for valuing changes in DRG quantities resulting from changes in treatment for existing patients. For some DRGs, we may assume that marginal changes in the number of patients only reflect nominal reporting changes. If changes in quantities do not have real effects on patient outcomes, then the marginal welfare with respect to quantities is equal between DRGs, such that upcoding affects welfare only through its budgetary cost.

While practical implementation of optimal reimbursements is challenging, the model highlight trade-offs that are absent from some alternative frameworks. In contrast to the model of yardstick competition (Shleifer, 1985) that is often used to motivate setting prices equal to average costs in a DRG system, our model explicitly accounts for the effects of a price on other DRGs' quantities and quality outcomes. Such cross-price effects, for instance due to upcoding behavior, have been documented for many DRGs by an extensive empirical literature. In our model, optimal prices are such that the ratio of approximate net social welfare effect of the price of a DRG to the budget share of the DRG is constant across DRG. Acknowledging these ideas when adjusting reimbursements in existing DRG systems may reduce distortive effects of the financial incentives.

7 Conclusion

In this paper, we estimate price and income elasticities of hospital reimbursements. We find that the price elasticity of quantity, the number of patients treated, is positive, approximately 0.2. The income elasticity, in contrast, is negative, approximately -0.15. Increasing all prices by 1% has a net positive effect on quantity, leading to an increase in the aggregate number of patients of more than 0.1% (up to 0.25% in alternative specifications). This increase in patients is real; it is not driven by upcoding, reallocation of patients within the same hospital, or reallocation of patients across hospitals. Effects of prices and income on quality-related measures are more nuanced. We also provide heterogeneous estimates by Clinical Classification Software category that are available for use in other studies.

To the best of our knowledge, our estimates are the first credibly identified price and income elasticities of hospital reimbursements for a representative set of patients, DRGs, and hospitals. We use data on the universe of hospital admissions in Germany between 2005 and 2016, which are reimbursed through a DRG system. Causal identification is primarily based on a two-year lag in the regulatory price setting. We demonstrate that our prices and income instruments do not have anticipatory effects on the outcome of interest, suggesting that they indeed only affect outcomes through prices and income.

Overall, we find that how much is paid for hospital services matters. Increasing the prices of all DRGs leads not only to a mechanical increase in aggregate costs, but also to a behavioral response that further increases the number of patients and hence budgetary costs. Changes in relative prices and the relative incomes of different specialized departments affect who is treated and how they are treated. Especially when price or income shocks are temporary, hospitals may not be able to adjust nursing staff or other fixed inputs appropriately for changes in the number of patients, leading to some detrimental effects on the quality of care.

There are some aspects of the effects of prices and income that we have not addressed in this paper. Of particular interest are effects on the structure of the hospital market: Do price and income shocks cause hospitals to open or close specialized departments or even the entire hospital? If specialized departments can provide higher quality of care, such effects are of importance to patient welfare, even in the absence of effects on the aggregate number of admissions. However, our data are not well suited for studying such effects because the exogenous price and income shocks we observe are only temporary.

References

Abadie, A., S. Athey, G. W. Imbens, and J. Wooldridge (2017). When should you adjust standard errors for clustering? *NBER Working Paper Series* (24003).

Abadie, A., S. Athey, G. W. Imbens, and J. M. Wooldridge (2020). Sampling-based vs. design-based uncertainty in regression analysis. *Econometrica 88*(1), 265–296.

Athey, S. and G. W. Imbens (2017, May). The state of applied econometrics: Causality and policy evaluation. *Journal of Economic Perspectives 31*(2), 3–32.

Baicker, K., M. E. Chernew, and J. A. Robbins (2013). The spillover effects of medicare managed care: Medicare advantage and hospital utilization. *Journal of Health Economics 32*(6), 1289–1300.

Birkmeyer, J. D., B. N. Reames, P. McCulloch, A. J. Carr, W. B. Campbell, and J. E. Wennberg (2013). Understanding of regional variation in the use of surgery. *The Lancet 382*(9898), 1121–1129.

Carryer, J. B., D. Diers, B. McCloskey, and D. Wilson (2010). Effects of health policy reforms on nursing resources and patient outcomes in new zealand. *Policy, Politics, & Nursing Practice 11*(4), 275–285.

Cattaneo, M. D., R. K. Crump, M. H. Farrell, and Y. Feng (2019). On binscatter. *arXiv preprint* (1902.09608).

Chandra, A. and D. O. Staiger (2007). Productivity spillovers in health care: evidence from the treatment of heart attacks. *Journal of Political Economy 115*(1), 103–140.

Chandra, A. and D. O. Staiger (2020). Identifying sources of inefficiency in healthcare. *The Quarterly Journal of Economics*. qjz040.

Chetty, R. (2009). Sufficient statistics for welfare analysis: A bridge between structural and reduced-form methods. *Annual Review of Economics 1*(1), 451–488.

Cho, S.-H., S. Ketefian, V. H. Barkauskas, and D. G. Smith (2003). The effects of nurse staffing on adverse events, morbidity, mortality, and medical costs. *Nursing Research 52*(2), 71–79.

Clemens, J. and J. D. Gottlieb (2014). Do physicians' financial incentives affect medical treatment and patient health? *The American Economic Review 104*(4), 1320–1349.

Cook, A. and S. Averett (2020). Do hospitals respond to changing incentive structures? evidence from medicare's 2007 drg restructuring. *Journal of Health Economics 73*, 102319.

Currie, J. and J. Gruber (1996a). Health insurance eligibility, utilization of medical care, and child health. *The Quarterly Journal of Economics 111*(2), 431–466.

Currie, J. and J. Gruber (1996b). Saving babies: The efficacy and cost of recent changes in the medicaid eligibility of pregnant women. *Journal of Political Economy 104*(6), 1263–1296.

Dafny, L. S. (2005). How do hospitals respond to price changes? *The American Economic Review 95*(5), 1525–1547.

Di Giacomo, M., M. Piacenza, L. Siciliani, and G. Turati (2017). Do public hospitals respond to changes in drg price regulation? the case of birth deliveries in the italian nhs. *Health Economics 26*, 23–37.

Diamond, P. A. (1975). A many-person ramsey tax rule. *Journal of Public Economics 4*(4), 335–342.

Doyle, J. J., J. A. Graves, J. Gruber, and S. A. Kleiner (2015). Measuring returns to hospital care: Evidence from ambulance referral patterns. *Journal of Political Economy 123*(1), 170–214.

Dranove, D., C. Garthwaite, and C. Ody (2013). How do hospitals respond to negative financial shocks? the impact of the 2008 stock market crash. *NBER Working Paper Series* (18853).

Duggan, M. (2002). Hospital market structure and the behavior of not-for-profit hospitals. *RAND Journal of Economics 33*(3), 433–446.

Duggan, M. G. (2000). Hospital ownership and public medical spending. *The Quarterly Journal of Economics 115*(4), 1343–1373.

Einav, L., A. Finkelstein, and N. Mahoney (2018). Provider incentives and healthcare costs: Evidence from long-term care hospitals. *Econometrica 86*(6), 2161–2219.

Evans, R. G. (1974). Supplier-induced demand: some empirical evidence and implications. In *The Economics of Health and Medical Care*, Chapter 10, pp. 162–173. Springer.

Feldman, R., R. Coulam, and B. E. Dowd (2015). Medicare's role in determining prices throughout the health care system. *Mercatus Working Paper*.

Finkelstein, A., N. Hendren, and E. F. P. Luttmer (2019). The value of medicaid: Interpreting results from the oregon health insurance experiment. *Journal of Political Economy 127*(6), 2836–2874.

Fleßa, S. and S. Nickel (2010). *Grundzüge der Krankenhaussteuerung*. Oldenbourg.

Foo, P. K., R. S. Lee, and K. Fong (2017). Physician prices, hospital prices, and treatment choice in labor and delivery. *American Journal of Health Economics 3*(3), 422–453.

Geissler, A., D. Scheller-Kreinsen, W. Quentin, and R. Busse (2011). Germany: Understanding g-drgs. In *Diagnosis-Related Groups in Europe*, Chapter 14, pp. 243–271. Open University Press.

Geruso, M. and T. Layton (2020). Upcoding: Evidence from medicare on squishy risk adjustment. *Journal of Political Economy 128*(3), 984–1026.

Gilman, B. H. (2000). Hospital response to drg refinements: the impact of multiple reimbursement incentives on inpatient length of stay. *Health Economics 9*(4), 277–294.

Heimig, F., C. Jacobs, M. Rabenschlag, and M. Rusert (2017). Extremkostenbericht gem. § 17b Abs. 10 KHG für 2017. Technical report, Institut für das Entgeltsystem im Krankenhaus GmbH.

InEK (2015). Abschlussbericht – Weiterentwicklung des G-DRG-Systems für das Jahr 2016. *Institut für das Entgeltsystem im Krankenhaus GmbH*.

Jena, A. B., V. Prasad, D. P. Goldman, and J. Romley (2015). Mortality and treatment patterns among patients hospitalized with acute cardiovascular conditions during dates of national cardiology meetings. *JAMA Internal Medicine 175*(2), 237–244.

Jürges, H. and J. Köberlein (2015). What explains drg upcoding in neonatology? the roles of financial incentives and infant health. *Journal of Health Economics 43*, 13–26.

Kane, R. L., T. A. Shamliyan, C. Mueller, S. Duval, and T. J. Wilt (2007). The association of registered nurse staffing levels and patient outcomes: systematic review and meta-analysis. *Medical Care*, 1195–1204.

Kovner, C., C. Jones, C. Zhan, P. J. Gergen, and J. Basu (2002). Nurse staffing and postsurgical adverse events: an analysis of administrative data from a sample of us hospitals, 1990–1996. *Health services research 37*(3), 611–629.

KPMG (2013). Durchschnittliche Umsatzrentabilität deutscher Krankenhäuser nach Trägern. *KPMG Datenbank Krankenhaus 300*.

Mattke, S., J. Needleman, P. Buerhaus, M. Stewart, and K. Zelevinsky (2004). Evaluating the role of patient sample definitions for quality indicators sensitive to nurse staffing patterns. *Medical Care*, II21–II33.

McGuire, T. G. (2000). Physician agency. In A. J. Culyer and J. P. Newhouse (Eds.), *Handbook of Health Economics*, Volume 1A of *Handbooks in Economics*, Chapter 9, pp. 461–536. Amsterdam: Elsevier.

McGuire, T. G. and M. V. Pauly (1991). Physician response to fee changes with multiple payers. *Journal of Health Economics 10*(4), 385–410.

Mirrlees, J. A. (1976). Optimal tax theory: A synthesis. *Journal of Public Economics 6*(4), 327–358.

Needleman, J., P. Buerhaus, S. Mattke, M. Stewart, and K. Zelevinsky (2002). Nurse-staffing levels and the quality of care in hospitals. *New England Journal of Medicine 346*(22), 1715–1722.

Olley, G. S. and A. Pakes (1996). The dynamics of productivity in the telecommunications equipment industry. *Econometrica 64*(6), 1263–1297.

Papanicolas, I. and A. McGuire (2015). Do financial incentives trump clinical guidance? hip replacement in england and scotland. *Journal of Health Economics 44*, 25–36.

Ramsey, F. P. (1927). A contribution to the theory of taxation. *The Economic Journal 37*(145), 47–61.

Reif, S., S. Wichert, and A. Wuppermann (2018). Is it good to be too light? birth weight thresholds in hospital reimbursement systems. *Journal of Health Economics 59*, 1–25.

Salm, M. and A. Wübker (2020). Do hospitals respond to decreasing prices by supplying more services? *Health Economics 29*(2), 209–222.

Shleifer, A. (1985). A theory of yardstick competition. *The RAND Journal of Economics 16*(3), 319–327.

Silverman, E. and J. Skinner (2004). Medicare upcoding and hospital ownership. *Journal of Health Economics 23*(2), 369–389.

Simon, M. and S. Mehmecke (2017). Nurse-to-patient ratios: Ein internationaler Überblick über staatliche vorgaben zu einer mindestbesetzung im pflegedienst der krankenhäuser. *Hans Böckler Stiftung Working Paper* (27).

Twigg, D. E., C. Duffield, A. Bremner, P. Rapley, and J. Finn (2011). The impact of the nursing hours per patient day (nhppd) staffing method on patient outcomes: a retrospective analysis of patient and staffing data. *International Journal of Nursing Studies 48*(5), 540–548.

Twigg, D. E., L. Gelder, and H. Myers (2015). The impact of understaffed shifts on nurse-sensitive outcomes. *Journal of Advanced Nursing 71*(7), 1564–1572.

Weissman, J. S., J. M. Rothschild, E. Bendavid, P. Sprivulis, E. F. Cook, R. S. Evans, Y. Kaganova, M. Bender, J. David-Kasdan, P. Haug, et al. (2007). Hospital workload and adverse events. *Medical Care 45*(5), 448–455.

Wennberg, J. E. (1996). The geography of health care in the united states. In J. E. Wennberg and M. McAndrew Cooper (Eds.), *The Dartmouth Atlas of Health Care*, Chapter 1, pp. 11–36. Chicago: American Hospital Association.

A Data and Empirics

A.1 Definition of Hospital-Acquired Conditions

In the main part of our analysis, we measure quality of care by (the absence of) hospital-acquired conditions. We identify patients suffering from hospital-acquired conditions based on the diagnosis codes suggested by Kovner et al. (2002), Needleman et al. (2002), Cho et al. (2003), Mattke et al. (2004), Kane et al. (2007), Weissman et al. (2007), Carryer et al. (2010), Twigg et al. (2011, 2015). Table A.1 provides a summary of the diagnosis codes we use to identify hospital-acquired conditions. For instance, the codes in Chapter A capture incidence of sepsis, which commonly occurs as a complication of invasive procedures. Pressure ulcers, or bedsores, captured by diagnosis codes in Chapter L are indicative of insufficient nursing care that can lead to preventable hospital-acquired conditions. Some complications of surgical and medical care are directly expressed through codes in Chapter T, such as "T81.5 – Complications of foreign body accidentally left in body following procedure." By relying on diverse diagnosis codes proposed in the literature, we measure hospital-acquired conditions across a broad spectrum of DRGs, irrespective of the intensity or invasiveness of the typical course of treatment for patients in the DRG.

Our specifications, based on *within*-sub-DRG comparisons, are reasonably robust even if some of the diagnosis codes we use are present at admission, which is not specified in the data. Some of the diagnosis codes listed in Table A.1 may be present at admission, such that not all identified patients are suffering from *hospital-acquired* conditions. Counting these as hospital-acquired conditions results in measurement error in the outcome variable. However, our specifications are based on first differences in the fraction of patients with at least one of these diagnoses *within* sub-DRG and hospital department. The specifications are therefore robust to differential initial fractions of patients with hospital-acquired condition diagnoses present at admission across sub-DRGs and hospital departments. Furthermore, the inclusion of cross-sectional fixed effects allows for differential linear trends in that fraction. Similarly, year fixed effects capture non-linear trends that are common across sub-DRGs and hospitals. Bias in our estimates due to conditions present at admission therefore only arises if the deviations from sub-DRG-hospital-department trends are correlated with price and income shocks.

A.2 Sources of Price Variation

We see the strategy outlined in Section 4.2 with results in Section 5.3 as the primary evidence in favor of the identifying assumption that in our specifications there is no direct effect of lagged cost changes on current changes in outcomes. In this appendix, we discuss and illustrate four key institutional details that explain this exogeneity of lagged cost changes. First, the regulator sets prices in year t based on average costs of patients treated in year $t-2$, and lagged cost *changes* are not predictive of current *changes* in costs. Second, costs are only reported for a sample of admissions and the sample mean as an estimator of average cost is noisy due to sampling variation in this setting. Third, redefinitions of DRG by the regulator to reduce cost heterogeneity introduce jumps in prices when DRGs are split, without associated changes in costs. Fourth, the sample of hospitals reporting costs is self-selected and not necessarily representative. Hence, changes in the set of cost-reporting hospitals can affect average reported costs without reflecting actual changes in average costs.

Table A.1: List of diagnosis codes used to identify hospital-acquired conditions.

Chapter 1	Chapter 4	Chapter 5	Chapter 9	Chapter 10
A400-A403	E1001	F059	I260	J182
A408-A409	E1011	F432	I269	J8001-J8003
A410-A414	E1101	F439	I460-I461	J8009
A4151-A4152	E1111	F4488	I499	J951-J952
A4158	E86		I801	J9600-J9601
A418-A419	E870-E878		I8020	J9609
A490-A493			I8028	
A498-A499			I8280	
			I8288	
Chapter 11	**Chapter 12**	**Chapter 14**	**Chapter 18**	**Chapter 19**
K250-K253	L8900-L8909	N390	R092	T793
K259	L8910-L8919		R34	T800
K260-K263	L8920-L8929		R401-R402	T802-T803
K269	L8930-L8939		R570-R572	T811
K270-K273	L8940-L8949		R579	T813-T816
K279	L8990-L8999		R650-R653	T835
K280-K283			R659	
K289				
K290-K291				
K296				

Notes: These diagnosis codes are proposed by an extensive literature to identify hospital-acquired conditions and nursing-sensitive outcomes. Importantly, our specifications correlate *changes* in the frequency of these conditions to changes in prices and incomes, with cross-sectional (and year) fixed effects. They are therefore robust to a fraction of patients having these diagnoses present at admission, even if that fraction varies by sub-DRG and exhibits sub-DRG-specific linear trends.

A.2.1 Limited Predictive Power of Lagged Cost Changes

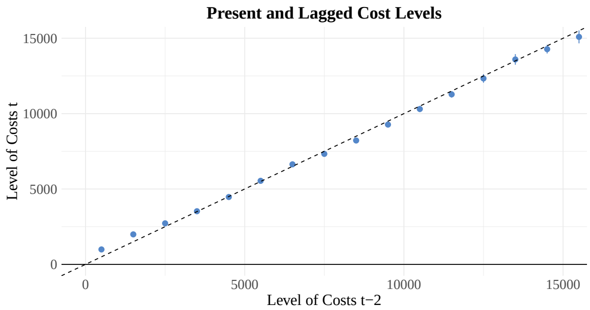
Price changes from year $t-1$ to t are based on cost changes from year $t-3$ to $t-2$. While the level of costs is mostly stable over time (see Figure A.1a), the cost changes from year $t-3$ to $t-2$ are not very predictive of cost changes from $t-1$ and t. Figure A.1b shows a binscatter of demeaned changes in reported costs, two-year lag against current, with the fixed effects included in all of our specifications projected out. The relationship between cost changes from $t-3$ to $t-2$ (price changes from $t-1$ and t) and cost changes from $t-1$ to t is extremely limited in magnitude, with even unusually large lagged cost changes on average only predicting very small current changes in costs. In particular, the linear regression corresponding to Figure A.1b has slope -0.04. Since current cost changes also only have a small effect on current outcomes (on the order of 0.05 in Table 4), excluding current costs from regressions only leads to negligible omitted variable bias.

A.2.2 Sampling Variation in Regulatory Estimates of Average Costs

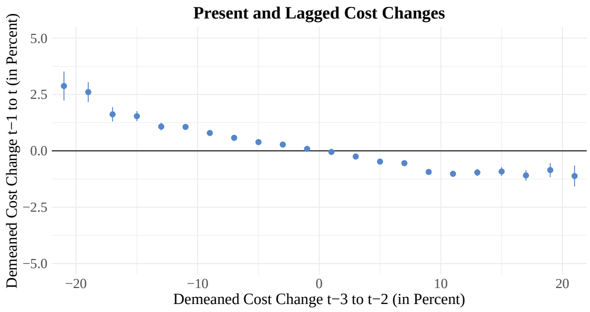
The regulator attempts to set the price of each DRG proportional to the average cost of treating a patient within the DRG. However, true average costs are not observed by the regulator; instead, prices are based on average costs in a sample of approximately 20% of admissions. This leads to noisy estimates of average costs, even under the most favorable assumption of *reported* admissions being independent draws from the distribution of costs.[18]

We illustrate this sampling variation in prices in Figure A.2. The figure shows the distribution of

[18] In practice, the setting is closer to clustered sampling: some hospitals report costs for all their admissions, while the remaining hospitals do not report any costs. Since neither the identity of reporting hospitals nor the number of reported admissions by hospital are published, we cannot properly account for clustered sampling. We instead illustrate the sampling variation assuming random sampling, which likely understates the true amount of sampling uncertainty.



(a) Current and lagged levels of costs. The binscatter shows that cost levels in year $t-2$ are highly predictive of cost levels in year t for the same sub-DRG. The 45° line is superimposed, and we zoom in on costs less than €15,000 for ease of presentation.



(b) Current and lagged demeaned cost changes. The binscatter shows the average current demeaned cost changes for admissions in sub-DRGs in 1% bins of two-year-lagged demeaned cost changes, and 95% confidence intervals around the mean. We project out the fixed effects used in all regressions to create the demeaned variables.

Figure A.1: Panel a shows that the two-year-lagged level of costs is highly predictive of current costs. In Panel b, we shift the focus from levels to changes and project out the fixed effects used in all of our regressions. Comparing the scales of the vertical and horizontal axes, past cost changes are not very predictive of current changes in costs. This suggests that the exclusion restriction holds, approximately: Lagged cost changes are used to set current prices, but have little effect on the most closely related current variable, cost changes.

price changes that could be observed if there were no changes in costs from one year to the next, but the regulator changes prices in response to receiving a new random sample of cost data for each DRG. To construct this figure, we use the average cost, standard deviation of costs, and sample size (number of cost reports) of each DRG in each year. For each DRG and year, we take two draws from a normal distribution with mean equal to the average reported costs for the DRG, and standard deviation equal to the standard deviation of reported costs divided by the square root of the number of admissions the average is based on. This normal distribution is an approximation to the sampling distribution of the average cost reports based on the central limit theorem, and would be the exact sampling distribution if costs per admission were normally distributed and cost reports constituted a random sample. To the extent that actual costs per admission are distributed with fatter tails and skewness compared to the normal distribution (cf. Heimig et al., 2017), and hospitals differ in their levels of costs, this asymptotic approximation is likely understating the true variation of sample means for DRG with a small number of cost reports and small number of reporting hospitals. We then calculate the percentage change from the first draw to the second draw. We plot the distribution of these changes, weighted by the number of actual observations (not the number of cost reports) they represent, analogous to Figure 1.

Figure A.2 reflects the distribution of price changes we would expect to see even if there were no changes in average costs and the regulator simply observed a new independent sample of cost reports to calculate prices every year. The standard deviation of the simulated distribution of price changes due only to sampling variation in cost reports is equal to approximately a third of the standard deviation of the distribution of price changes. We take this as evidence that a substantial amount of the variation in actual prices may be due to sampling variation rather than substantive changes that can have lasting direct effects on treatment decisions in hospitals.

A.2.3 Redefinition of DRG

Every year, the regulator tweaks the definitions of DRGs and creates additional DRGs with the goal of reducing within-DRG cost heterogeneity. When the definition of a DRG is changed such that, for

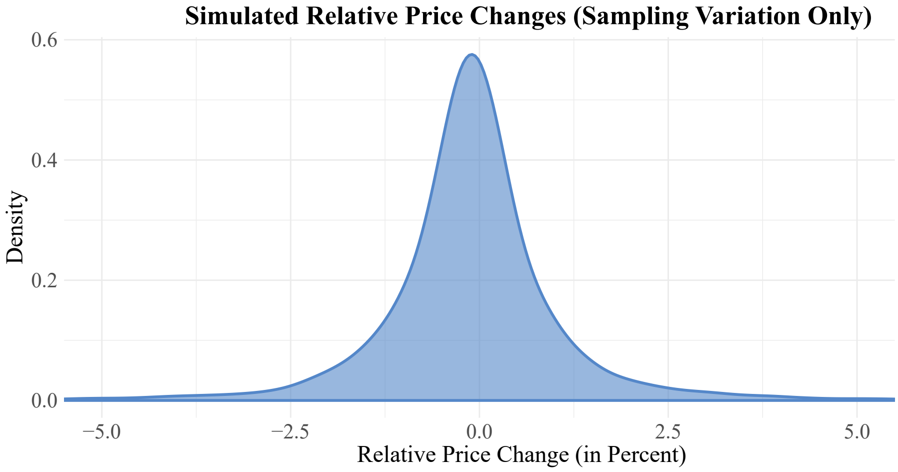


Figure A.2: Sampling variation in relative price changes. We simulate, for each DRG-year separately, the sampling variation of the price change based on the mean, standard deviation, and number of cost reports.

instance, admissions with a diagnosis that is associated with higher costs become part of a new DRG, the average cost of treating patients in the old DRG falls. Since prices are equal to (lagged) average costs, this redefinition causes a decrease in the price for admissions remaining in the old DRG, while the price increases for admissions in the new DRG. Hence, tweaks to the DRG system can create changes in prices without any changes in costs or unobservables. Instead, most redefinitions address cost heterogeneity that has existed in previous years, or are unintended consequences of changes in other parts of the classification algorithm due to its complexity. Even when the regulator changes definitions in direct response to changes in costs, the redefinitions are likely reflective of past changes in costs, but mostly uncorrelated with current changes, since the regulator only observes costs with a two-year lag. Most variation in prices created by changes in DRG definitions may therefore be plausibly exogenous in our specifications.

Figure A.3 illustrates the approximate variation caused by redefinitions of the DRG system only. For each actual DRG in a given year, we approximate the cost of the DRG in the following year in the absence of changes to the definitions: We take the average actual costs in the following year of all admissions of the DRG in the initial year. If, for instance, in the following year 80% of the DRG remain in the same DRG and 20% are classified into a second DRG, then the imputed costs without redefinitions of this hypothetical unchanged DRG is 0.8 times the average cost of the first DRG in the following year plus 0.2 times the average cost of the second DRG in the following year. This is only an approximation to the true average cost because the costs of the second DRG in the following year is based on the average cost of all admissions classified into it in that year, rather than the average of the 20% that in the initial year were classified into the first DRG. Similarly, the initial DRG in the following year may contain admissions with different average costs that initially were classified as a third DRG. However, since the regulator specifically attempts to homogenize costs within DRG, the average costs of each admission's new DRG likely reflects the costs of that admission relatively well. We then calculate the percentage difference between actual costs and hypothetical costs for each admission. By using the hypothetical cost rather than the initial cost, we remove most variation in costs that is due to actual year-to-year changes in costs.

The distribution of price changes resulting from changes in DRG definitions is shown in Figure A.3, with each sub-DRG weighted according to the number of admissions in it, analogous to Figure 1. Most of the mass of the simulated distribution of price changes is between -1 and 1, reflecting that redefinitions of the DRG system mostly split existing DRGs that are already relatively homogeneous, and therefore do not create particularly large changes in averages costs and hence prices. However, the distribution has fat

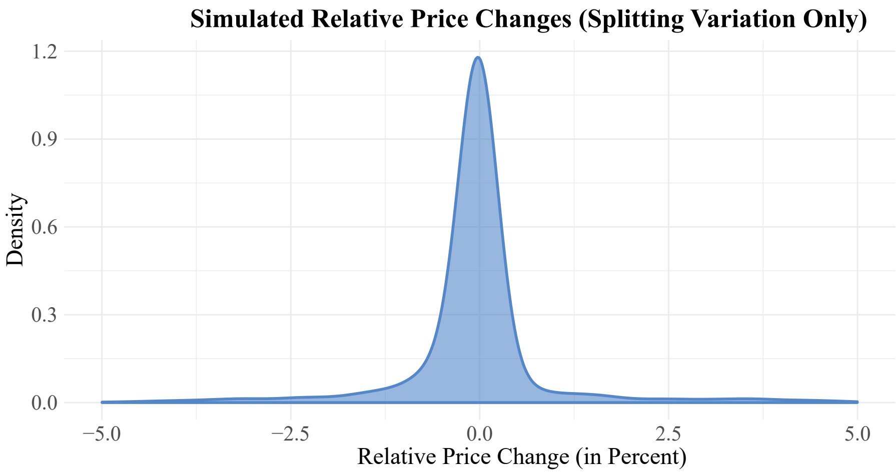


Figure A.3: Price changes due to changes in DRG definitions. The density plot shows, for each sub-DRG-year, the relative price change from the approximate price if no changes had been made to DRG definitions.

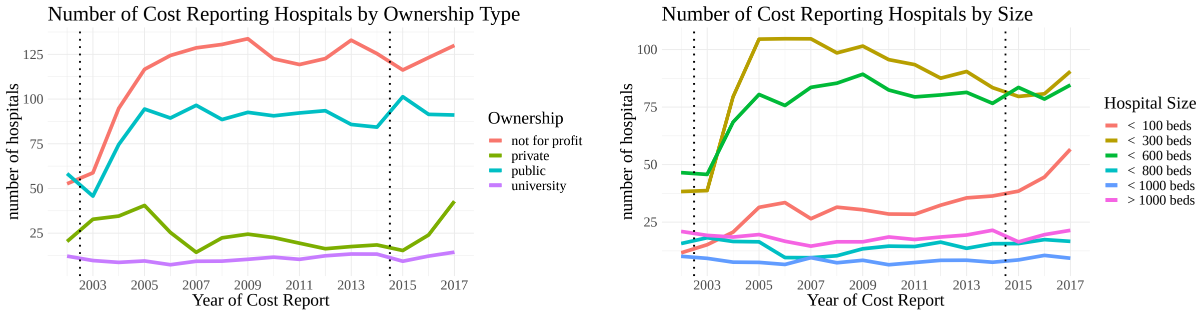
tails: Sometimes redefinitions move admissions to DRGs with prices quite different from their original DRGs. As a result, the variance of simulated price changes is approximately equal to half the variance of actual price changes.

A.2.4 Changing Composition of Cost-Reporting Hospitals

If costs differ between hospitals, changes in which hospitals report costs can affect average reported costs and thereby prices, without any direct dependence on changes in current unobservables. Specifically, if a low-cost hospital starts to report costs, we expect the average costs (and hence prices after a two-year lag) to decrease in particular for those DRGs that the hospital treats the most, relative to DRGs not treated by the low-cost hospital.

While the regulator does not publish the identity of individual cost-reporting hospitals, Figure A.4 illustrates compositional shifts in the set of hospitals reporting costs for a given year, based on official reports showing these aggregate numbers (e.g. InEK, 2015). In panel a, we see that in the early years the number of not-for-profit and public hospitals reporting costs increased drastically, while the number of private for-profit hospitals reporting costs declined starting in 2005. Panel b shows that the number of very small hospitals (less than 100 beds) reporting costs grew steadily, while the number of slightly larger hospitals (100 to 300 beds) declined systematically after a rapid increase in the first couple of years. Even in the categories of large hospitals with more than 600 beds, which have a larger influence on average costs due to the number of admissions they report, there are yearly changes in the number of such hospitals, in addition to potential within-category changes of which hospitals report costs.

Since the regulator only publishes a list of hospitals that are eligible to report costs, but not a list of hospitals actually reporting costs, it is challenging to estimate the effect particular hospitals may have on average costs. Here, we partly overcome this challenge by focusing on a regulatory change that revealed the identity of some reporting hospitals in 2016. Until 2016, participation in the cost reports was entirely voluntary, with a small financial incentive to cover administrative costs. To increase the representativeness of cost reports, the German regulator in 2016 defined a list of hospitals that were not currently reporting costs but matched certain criteria that were considered important for representativeness. From this list, 40 hospitals were chosen at random and required to submit cost reports. Of these 40 hospitals, 8 hospitals submitted complete cost reports for admissions in 2016 and were included in the sample to calculate average costs. The remaining 32 hospitals either submitted partial information that was not used directly



(a) Cost reporting hospitals by ownership type. The lines show the number of hospitals of a given ownership type reporting cost in a given year.

(b) Cost reporting hospitals by hospital size. The lines show the number of hospitals of a given size reporting cost in a given year.

Figure A.4: Panels a and b show that the number and type of hospitals reporting costs in a given year varies greatly. If hospitals of different ownership type or size face different costs, then changes in the set of hospitals reporting costs can lead to changes in prices. The dashed lines mark the start and end of cost reports that appear as prices in our sample after a two-year lag.

to calculate average costs, or were fined for failure to comply. Importantly, the identity of chosen hospitals and the status of their actual participation are both public.

We can therefore focus on 8 hospitals that are known to have participated in the cost reporting for the first time in 2016. As average costs become prices after a two-year lag, we focus on price changes from 2017, the last year when prices were not partially based on cost reports from these hospitals, to 2018, the first year when prices are based on their costs. These hospitals plausibly have different average costs from the hospitals that are voluntarily reporting costs. By including the cost reports of these 8 hospitals, average costs per DRG then change. In total, 248 hospitals reported cost data for 2016, such that the 8 newly cost-reporting hospitals constitute slightly more than 3 percent of cost-reporting hospitals. However, the newly cost-reporting hospitals tend to be larger, such that a larger percentage of cost reports comes from them. Specifically, average costs are likely to change more for those DRG with a larger fraction of cost reports from the newly cost-reporting hospitals. For DRG where the 8 newly cost-reporting hospitals only submit a negligible fraction of the total number of cost reports, no meaningful change in average costs is expected. Such an analysis relies on the (untestable) assumption that the newly cost-reporting hospitals do not systematically treat relatively more patients of DRG that are experiencing larger year-to-year changes in average costs from 2015 to 2016.

Figure A.5 shows the distribution of the share of cost reports per DRG coming from the newly cost-reporting hospitals, weighted by the aggregate number of admissions in the DRG. We see a reasonable spread with different DRGs having anywhere between 0 and 5 percent of cost reports coming from newly cost-reporting hospitals, and some DRGs having even 5 to 15 percent of their cost reports coming from the 8 new hospitals.

We estimate the effect of having a larger fraction of admissions coming from the newly cost-reporting hospitals on how much the price of a sub-DRG changed from 2017 to 2018, when the newly cost-reporting hospitals were first included in the calculation of prices. Column (1) of Table A.2 shows the coefficient estimates from a regression of the price change of sub-DRGs on the fraction of cost reports coming from the 8 newly cost-reporting hospitals. The estimated coefficient is -0.117 and statistically significant. Compared to DRGs where the newly cost-reporting hospitals did not report any costs, the DRGs where they contributed 10% of cost reports on average experienced a 1.17 percent decrease in price. Extrapolating the effect to the extreme, this suggests that DRGs where 100% of cost reports come from the 8 newly cost-reporting hospitals may have their prices decrease by close to 12%. This suggests that average costs are about 12% higher in hospitals voluntarily reporting costs than in the 8 newly cost-reporting hospitals,

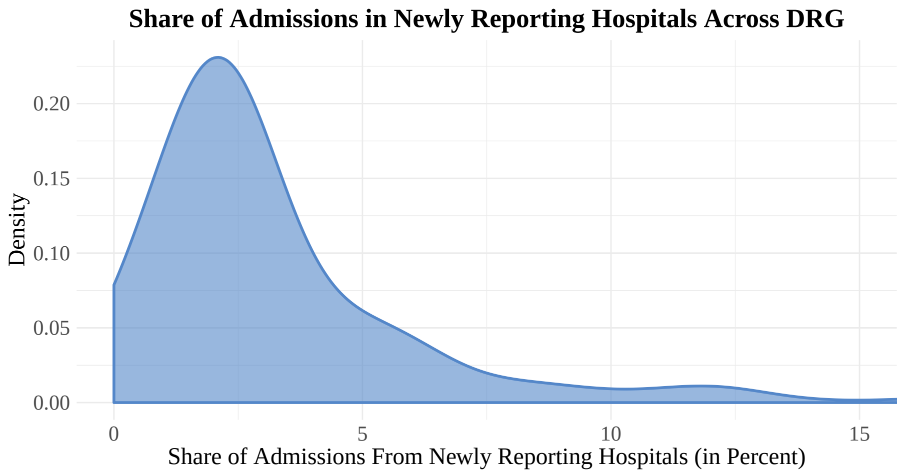


Figure A.5: Distribution of the percentage of admissions coming from newly cost-reporting hospitals across DRGs. For most DRGs, between 0 and 10 percent of admissions come from newly-reporting hospitals.

which tend to be larger. To interpret this as the causal effect of the 8 newly cost-reporting hospitals, we need to assume that the fraction of admissions of a DRG coming from newly cost-reporting hospitals is not systematically related to other reasons the price might change from one year to the next.

We assess whether the DRGs where a larger fraction of cost reports comes from the 8 newly cost-reporting hospitals are systematically different in columns (2) – (4) of Table A.2. In column (2), we add the price change from 2016 to 2017 as a regressor. The coefficient is negative, suggesting reversion to the mean: DRGs that experienced a price increase from 2016 to 2017 then were more likely to experience a (on average much smaller) price decrease from 2017 to 2018. However, the effect of the newly cost-reporting hospitals is unaffected by this conditioning; it does not meaningfully pick up mean reversion. For column (3), we instead condition on the average yearly price change of the sub-DRG between 2005 and 2017, to control for constant percentage trends in the price of sub-DRGs. Again, the coefficient on the fraction of cost reports coming from newly cost-reporting hospitals is unchanged, suggesting that it is not confounded by differences in price trends. We supplement the analysis with the regression of a pseudo-outcome (Athey and Imbens, 2017) on the treatment in column (4). The outcome variable is the price change from 2016 to 2017, before any of the 8 hospitals contributed cost reports. The estimated effect of the 8 hospitals *before* they started to actually affect prices is much smaller in magnitude and statistically insignificant, with a point estimate of about 0.02 (s.e. 0.02) compared to the estimate of about -0.12 (s.e. 0.03) in column (1). The results in columns (2) and (3) suggest that the differential price change for sub-DRGs more affected by the 8 newly cost-reporting hospitals cannot be explained by what happened to prices in the past. The placebo-test in column (4) suggests that these sub-DRGs were not different from less affected sub-DRGs right until the 8 hospitals report costs.

A.3 Excluding Similar DRGs From Income Instrument

In this appendix, we show results from alternative specifications removing variation from the instrument for income changes that is due to those price changes most likely to have compensated cross-price (direct) effects. If the price of sub-DRG g' has a direct effect on the number of patients treated in sub-DRG g (or other outcome for sub-DRG g), then in a regression of the outcome for g on the price of g and our instrument for department income, the coefficient on department income will be biased towards the direct effect of g' due to the correlation of the department income instrument and the price of g', the omitted variable. Estimation of cross-price elasticities through inclusion of prices of all other sub-DRGs

Table A.2: Effect of the fraction of cost reports from newly cost-reporting hospitals on prices.

	All Variables in Percentage Changes Outcome Variable: Price Change			
	2017/18			2016/17
	(1)	(2)	(3)	(4)
Percentage New	-0.117^{***}	-0.116^{***}	-0.119^{***}	0.019
	(0.028)	(0.028)	(0.027)	(0.019)
Relative Price Change 2016/17		-0.038^{**}		
		(0.015)		
Avg. Previous Relative Price Change			-0.062^{**}	
			(0.030)	
Constant	0.003	0.003	0.004	0.004^{*}
	(0.003)	(0.003)	(0.002)	(0.002)
Observations	1,325,634	1,325,288	1,255,194	1,325,288

Notes: Sub-DRGs with a larger fraction of admissions coming from one of the 8 newly cost-reporting hospitals experience a larger decline in price from 2017 to 2018, when the new cost reports first come into effect as prices. Observations are time-constant sub-DRGs, and standard errors are clustered at the level (drg_{2017},drg_{2018}) for columns (1) – (3) and (drg_{2016},drg_{2017}) for column (4). Robust standard errors clustered at the DRG in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

g' is infeasible in practice due to the large number of potential omitted sub-DRGs g' and their generally heterogeneous effects (the effect of the price of g' on g can differ greatly from the effect of g' on a third sub-DRG g''). Instead, we focus on removing the price variation of g' from the income instrument for observations of sub-DRG g. The department income instrument then varies also within department because we remove price variation due to different g' for observations of different g. This strategy removes the correlation between the income instrument and omitted variables, and thereby reduces bias, while remaining computationally feasible as an instrumental variables regression of outcomes on own-price and department income, using the sub-DRG-department-year-specific augmented simulated instrument as an instrument for department income.

The primary challenge for our approach is to identify which sub-DRGs g' are likely to have a direct effect on outcomes of a given sub-DRG g. We propose two strategies that are completely based on external information and avoid using outcome data to inform construction of the instrument. Specifically, our strategies are based on either clinical or regulatory similarity of sub-DRGs.

For our first strategy exploiting clinical similarity of sub-DRGs, we use the hierarchical structure of Clinical Classification Software (CCS) groups: For sub-DRG g, we remove price variation from the income instrument due to sub-DRGs g' which are part of the same CCS group (out of 283 CCS groups), part of the same more aggregate CCS subcategory (out of 136 subcategories), or part of the same top-level category (out of 18 top-level categories). Within a CCS group, patients are relatively homogeneous in clinical diagnoses, so excluding price variation due to sub-DRGs within the same CCS group acknowledges that there may be upcoding opportunities for some patients. This is effective if upcoding affects the DRG and hence reimbursement, but does not substantively alter the overall disease picture of the patients, such that the patient remains in the same CCS group. However, we may be concerned that some upcoding behavior even changes the CCS group to a related CCS group, or that the treatments for some sub-DRGs compete for or are complements for the same hospital resources. Direct effects of such behavior may better be captured within the more aggregate CCS subcategories. Using the 18 top-level CCS categories is the most aggressive approach to excluding price variation of clinically similar sub-DRGs.

Our second strategy relies on similarity of DRGs as implicitly assessed by the regulator. The principle idea is that the regulator, through redefinitions of the DRG system, implicitly defines some DRG that are similar: If two patients had the same DRG in one year but two distinct DRGs in the next year, then the patients are in principle different (otherwise they would share the same DRG in the second year), but must overall be comparable (otherwise the regulator would not have grouped them together in the first year). Furthermore, some of the changes introduced by the regulator may reflect potential choices in upcoding behavior. For instance, if in the second year a DRG is split based on the presence of a diagnosis code that was not used in the first year, then in some cases hospitals may have a choice of (not) reporting the diagnosis code in the second year to increase reimbursements. Hence, some redefinitions by the regulator can be indicative of margins for upcoding.

It is easiest to describe our definition of regulatory similarity based on viewing sub-DRGs as a network, with each node in the network representing a single sub-DRG. Each sub-DRG maps to exactly one DRG for each year. Two sub-DRGs are connected / their nodes share an edge, if for at least one year they map to the same DRG. We define the regulatory degree of separation (d.o.s.) of two sub-DRGs as the length of the shortest path between their nodes. Hence, two sub-DRGs that map to the same DRG for at least one year have d.o.s. 1. Two sub-DRGs have d.o.s. 2 if there exists a third sub-DRG connected to both sub-DRG. Two sub-DRGs have d.o.s. 3 if there exists a third sub-DRG with d.o.s. 1 from one of the sub-DRG and d.o.s. 2 from the other sub-DRG.

Results for both strategies are shown in Table A.3, with first stage regression results shown in Table A.4. Columns (1), (2), and (3) show instrumental variables estimates of the price and income elasticity using the simulated instrument excluding all price variation from sub-DRGs that are within the same CCS group, CCS subcategory, and CCS top-level category, respectively. The point estimates of -0.141, -0.219, and -0.219 are statistically indistinguishable from the comparable income elasticity estimate in column (3) of Table 2, -0.160. The pattern is similar when excluding price variation based on regulatory similarity rather than clinical similarity as shown in columns (4), (5), and (6). Only when excluding all sub-DRGs with a d.o.s. of 3 or less, in column (6), the coefficient changes substantively to -0.518. Note, however, that excluding all sub-DRGs with a d.o.s. of 3 or less comes at a substantial cost: On average, the simulated instrument for the percentage change in department income is based on the price changes of less than 30% of the admissions in that department, compared to 90% and 74% for columns (4) and (5), respectively. Excluding most of the variation comes at the cost of a substantially worse first stage fit, with a regression coefficient of only 0.27 (Table A.4 column (6)), compared to 0.85 and 0.70 (Table A.4 columns (4) and (5)). Throughout, the estimates of the price elasticity remain largely unchanged.

We briefly suggest a potential explanation for the change in the income elasticity when a substantial amount of price variation is excluded from the simulated instrument in column (6) of Tables A.3 and A.4. After excluding all sub-DRGs with a regulatory d.o.s. of up to 3, less than a third of observations of a department are used, on average, to compute the simulated instrument. However, there is substantial variation in how much variation is excluded across observations, as displayed in Figure A.6. The different panels each show the density of the fraction of admissions within the same department that are used to compute the simulated instrument for that observation. For the strategies in columns (1), (2), (4), and (5) of Tables A.3 and A.4, the simulated instrument is still based on more than 75% of admissions of the department for the majority of observations. Even when we exclude all sub-DRGs from the CCS top-level category, most observations retain price variation in their income instrument from more than 50% of their sub-DRGs. In contrast, when excluding all sub-DRGs with d.o.s. of 3 or less, the instrument is based on less than 25% of admissions for the majority of observations. However, for some observations a larger fraction of admissions are used. Since we aggregate admissions at the sub-DRG-department-year-level, observations where we only exclude a small fraction of the variation from the income instrument must be

Table A.3: Estimates of income elasticities excluding some variation from the simulated instrument for income changes.

Second Stage IV	All Variables in Percentage Changes Outcome Variable: Quantity					
	Clinical Similarity			Regulatory Similarity		
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Price	0.195***	0.202***	0.202***	0.194***	0.199***	0.229***
	(0.017)	(0.017)	(0.017)	(0.017)	(0.018)	(0.022)
Department Income	−0.141**	−0.219***	−0.219***	−0.132**	−0.188***	−0.518***
(actual)	(0.064)	(0.072)	(0.079)	(0.062)	(0.070)	(0.143)
share included	0.88	0.79	0.63	0.90	0.74	0.29
Observations	650,890	650,890	650,890	650,890	650,890	650,890

Notes: For columns (1), (2), and (3), we exclude variation due to sub-DRGs sharing the CCS group, CCS subcategory, and CCS top-level category, respectively, for each observation in the regression. For columns (4), (5), and (6), we exclude variation due to sub-DRGs with regulatory degrees of separation of 1, 2, and 3, respectively, as defined in the text. The instrumental variables regression specifications are otherwise identical to column (3) of Table 2. The "share included" is the share of admissions in year $t-2$ whose price changes are included in the income instrument. The results alleviate concerns that our income elasticity is primarily driven by negative cross-price elasticities, for instance due to upcoding. All specifications include two-way fixed effects. Robust standard errors clustered at the level of treatment assignment in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

Table A.4: First stage regressions corresponding to the instrumental variables regressions in Table A.3. As we exclude additional variation from the income instrument, the first stage coefficient becomes smaller.

First Stage IV	All Variables in Percentage Changes Outcome Variable: Department Income (actual)					
	Clinical Similarity			Regulatory Similarity		
	(1)	(2)	(3)	(4)	(5)	(6)
Relative Price	0.062***	0.067***	0.075***	0.056***	0.073***	0.087***
	(0.005)	(0.005)	(0.006)	(0.005)	(0.006)	(0.006)
Department Income	0.815***					
(exclude CCS group)	(0.043)					
Department Income		0.706***				
(exclude subcategory)		(0.041)				
Department Income			0.563***			
(exclude category)			(0.033)			
Department Income				0.854***		
(exclude d.o.s. 1)				(0.044)		
Department Income					0.697***	
(exclude d.o.s. 2)					(0.044)	
Department Income						0.267***
(exclude d.o.s. 3)						(0.024)
share included	0.88	0.79	0.63	0.90	0.74	0.29
Observations	650,890	650,890	650,890	650,890	650,890	650,890

Notes: All specifications include two-way fixed effects. Robust standard errors clustered at the level of treatment assignment in parentheses. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.1$

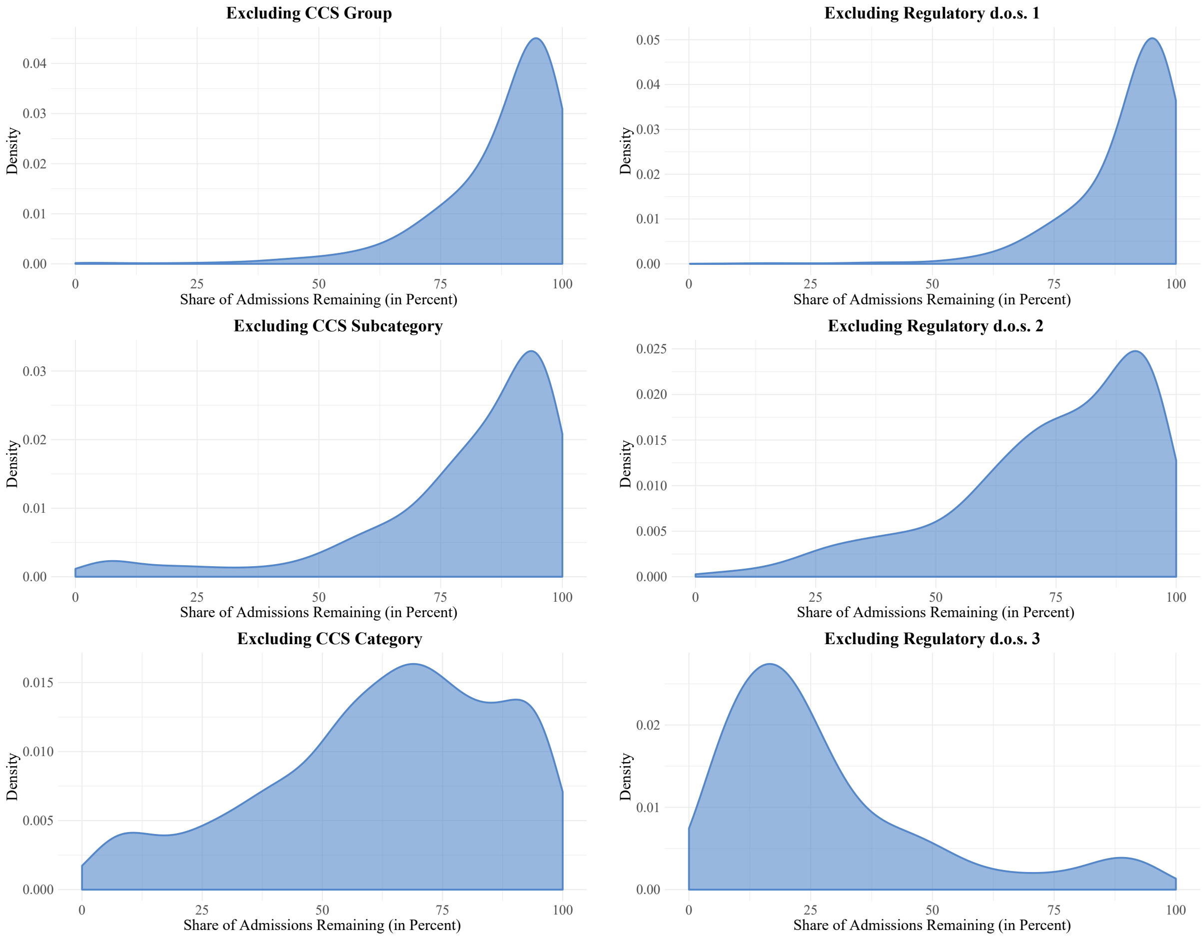


Figure A.6: Distribution of share of admissions remaining in income instrument across observations when excluding sub-DRGs that are similar either clinically (first column) or in the DRG system defined by the regulator (second column). When only a small share of admissions remains in the instrument, its correlation with actual income is reduced. Sub-DRGs with many admissions remaining are the "compliers" – for these sub-DRGs, the instrument strongly affects the endogenous variable.

sub-DRG that are atypical for the particular department.

In the context of instrumental variables regression, the instrument is most closely correlated with actual income for the observations that have only a small fraction of admissions excluded from the simulated instrument. Consequently, in these regressions the atypical sub-DRGs and diverse departments constitute "compliers;" that is, their realized treatment (actual income) most closely tracks assigned treatment (income instrument). In a setting with heterogeneous coefficients, the instrumental variables estimates are then weighted towards these compliers. The difference in estimated coefficients may therefore be explained by heterogeneity of effects, in particular with respect to sub-DRGs that are atypical for a particular department. Specifically, when a department's income increases due to increases in the prices for the majority of the services it provides, capacity constraints may particularly result in reductions for atypical sub-DRGs, which may sometimes alternatively be treated by other departments in the same hospital.

B Proofs

B.1 Solution to the Hospital Problem

The Lagrangian corresponding to the optimization of the hospital problem 4 is

$$\mathcal{L}_{\text{hospital}} = u(q, e, Z) + \alpha(p \cdot q + L - Z)$$

such that the first order condition with respect to q_g, for $g = 1, \ldots, G$, is

$$\frac{\partial u}{\partial q_g} + \alpha q_g = 0$$

while the first order condition with respect to e_k, for $k = 1, \ldots, K$, is

$$\frac{\partial u}{\partial e_k} = 0$$

and with respect to income Z

$$\frac{\partial u}{\partial Z} - \alpha = 0$$

The complementary slackness conditions are

$$\alpha(p \cdot q + L - Z) = 0$$
$$\alpha \geq 0$$

Denote the solution to this system of equations by $q^*(p, L)$, $e^*(p, L)$, $Z^*(p, L)$, and $\alpha^*(p, L)$. The indirect utility function of the hospital is defined as the maximized value,

$$V(p, L) = u(q^*(p, L), e^*(p, L), Z^*(p, L)).$$

Define the expenditure (cost) function c as

$$c(p, \bar{u}) \equiv \min_{q,e,Z} \quad Z - p \cdot q$$
$$\text{s.t. } u(q, e, Z) \geq \bar{u}$$

Let $\tilde{q}(p, \bar{u})$ and $\tilde{e}(p, \bar{u})$ be the solution to this optimization problem, Hicksian demand in consumer theory. By definition, the optimal choices satisfy $\tilde{q}(p, \bar{u}) = q^*(p, c(p, \bar{u}))$ and $\tilde{e}(p, \bar{u}) = e^*(p, c(p, \bar{u}))$.

We use the envelope theorem on the minimization problem defining c to derive a version of Shephard's Lemma:

$$\frac{\partial c}{\partial p_g}(p, \bar{u}) = -\tilde{q}_g(p, \bar{u})$$

By totally differentiating the identity $\tilde{q}_g(p, \bar{u}) = q_g^*(p, c(p, \bar{u}))$, we obtain the Slutsky equation:

$$\frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, \bar{u}) = \frac{\partial q_{g'}^*}{\partial p_g}(p, c(p, \bar{u})) + \frac{\partial q_{g'}^*}{\partial L}(p, c(p, \bar{u}))\frac{\partial c}{\partial p_g}(p, \bar{u}).$$

and similarly for effort

$$\frac{\partial \tilde{e}_k}{\partial p_g}(p, \bar{u}) = \frac{\partial e_k^*}{\partial p_g}(p, c(p, \bar{u})) + \frac{\partial e_k^*}{\partial L}(p, c(p, \bar{u}))\frac{\partial c}{\partial p_g}(p, \bar{u}).$$

Substituting Shephard's Lemma in the Slutsky equation and rearranging, we get

$$\frac{\partial q^*_{g'}}{\partial p_g}(p, c(p, \bar{u})) = \frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, \bar{u}) + \frac{\partial q^*_{g'}}{\partial L}(p, c(p, \bar{u}))\tilde{q}_g(p, \bar{u})$$

and

$$\frac{\partial e^*_k}{\partial p_g}(p, c(p, \bar{u})) = \frac{\partial \tilde{e}_k}{\partial p_g}(p, \bar{u}) + \frac{\partial e^*_k}{\partial L}(p, c(p, \bar{u}))\tilde{q}_g(p, \bar{u}),$$

where $\tilde{q}_g$ appears in both equations by Shephards Lemma.

As in consumer theory, the Slutsky matrix for quantity q is symmetric. To see this, simply differentiate Shephard's Lemma for $p_{g'}$ with respect to p_g to obtain

$$\frac{\partial^2 c}{\partial p_{g'}\,\partial p_g}(p, \bar{u}) = -\frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, \bar{u}).$$

and similarly differentiate Shephard's Lemma for p_g with respect to $p_{g'}$. Since the left hand sides are equal, we find

$$\frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, \bar{u}) = \frac{\partial \tilde{q}_g}{\partial p_{g'}}(p, \bar{u}).$$

B.2 Solution to the Social Planner Problem

The Lagrangian corresponding to the optimization of the social planner's problem in 5 is

$$\mathcal{L}_{\text{social planner}} = S(q^*(p, 0), e^*(p, 0)) + \lambda(B - p \cdot q^*(p, 0))$$

The first order condition with respect to p_g, for $g = 1, \ldots, G$, is

$$\begin{aligned} &\sum_{g'=1}^{G} \frac{\partial S}{\partial q_{g'}}(q^*(p, 0), e^*(p, 0))\frac{\partial q^*_{g'}}{\partial p_g}(p, 0) \\ &+ \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(p, 0), e^*(p, 0))\frac{\partial e^*_k}{\partial p_g}(p, 0) \\ = \lambda\Big(&q^*_g(p, 0) + \sum_{g'=1}^{G} p_{g'}\frac{\partial q^*_{g'}}{\partial p_g}(p, 0)\Big) \end{aligned}$$

Next, substitute the Slutsky equation to replace uncompensated price effects by compensated price and income effects, and group terms involving price effects on the left hand side

$$
\begin{aligned}
&\sum_{g'=1}^{G} \frac{\partial S}{\partial q_{g'}}(q^*(p,0), e^*(p,0)) \frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, V(p,0)) \\
&- \lambda \sum_{g'=1}^{G} p_{g'} \frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, V(p,0)) \\
&+ \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(p,0), e^*(p,0)) \frac{\partial \tilde{e}_k}{\partial p_g}(p, V(p,0)) \\
= q_g^*(p,0) \Big(&\lambda + \lambda \sum_{g'=1}^{G} p_{g'} \frac{\partial q_{g'}^*}{\partial L}(p,0) \\
&- \sum_{g'=1}^{G} \frac{\partial S}{\partial q_{g'}}(q^*(p,0), e^*(p,0)) \frac{\partial q_{g'}^*}{\partial L}(p,0) \\
&- \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(p,0), e^*(p,0)) \frac{\partial e_k^*}{\partial L}(p,0) \Big)
\end{aligned}
$$

where we use that Marshallian demand is equal to Hicksian demand, $q_g^*(p,0) = \tilde{q}_g^*(p, V(p,0))$. Define

$$
\begin{aligned}
\theta \equiv &\sum_{g=1}^{G} \frac{\partial S}{\partial q_g}(q^*(p,0), e^*(p,0)) \frac{\partial q_g^*}{\partial L}(p,0) \\
&+ \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(p,0), e^*(p,0)) \frac{\partial e_k^*}{\partial L}(p,0) \\
&- \lambda \Big(1 + \sum_{g=1}^{G} p_g \frac{\partial q_g^*}{\partial L}(p,0) \Big)
\end{aligned}
$$

as well as the compensated price elasticity of quantity

$$
\tilde{\epsilon}_{g,g'}^q(p,L) \equiv \frac{p_{g'}}{q_g^*(p,L)} \frac{\partial \tilde{q}_g}{\partial p_{g'}}(p, V(p,L))
$$

and the compensated semi-elasticity of effort

$$
\tilde{\epsilon}_{k,g}^e(p,L) \equiv p_g \frac{\partial \tilde{e}_k}{\partial p_g}(p, V(p,L)).
$$

Next, by symmetry of the Slutsky matrix

$$
\frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, V(p,0)) = \frac{\partial \tilde{q}_g}{\partial p_{g'}}(p, V(p,0)).
$$

Using this identity and the definitions above, we obtain after dividing by $q_g^*(p,0)$

$$\begin{aligned}
&\sum_{g'=1}^{G} \frac{\partial S}{\partial q_{g'}}(q^*(p,0), e^*(p,0)) \frac{1}{p_{g'}} \tilde{\epsilon}^q_{g,g'}(p,0) \\
&- \lambda \sum_{g'=1}^{G} \tilde{\epsilon}^q_{g,g'}(p,0) \\
&+ \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(p,0), e^*(p,0)) \tilde{\epsilon}^e_{k,g}(p,0) \frac{1}{p_g q_g^*(p,0)} \\
= &- \theta.
\end{aligned}$$

We can similarly rewrite θ by defining the income elasticity of quantity

$$\eta^q_g(p,L) \equiv \frac{B}{q_g^*(p,L)} \frac{\partial q_g^*}{\partial L}(p,L)$$

and the semi-elasticity of effort

$$\eta^e_k(p,L) \equiv B \frac{\partial e_k^*}{\partial L}(p,L).$$

Substituting these definitions for their expressions in θ, we find

$$\begin{aligned}
\theta = &\sum_{g=1}^{G} \frac{\partial S}{\partial q_g}(q^*(p,0), e^*(p,0)) \frac{q_g^*(p,0)}{B} \eta^q_g(p,0) \\
&+ \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(g,0), e^*(p,0)) \frac{1}{B} \eta^e_k(p,0) \\
&- \lambda \Big(1 + \sum_{g=1}^{G} \frac{p_g q_g^*(p,0)}{B} \eta^q_g(p,0)\Big).
\end{aligned}$$

B.3 The Index of Encouragement

We show that the social planner's first order condition permits an interpretation as a constant approximate "index of encouragement."

We start with the first order condition, rewritten to use compensated demand and the parameter θ:

$$\begin{aligned}
&\sum_{g'=1}^{G} \frac{\partial S}{\partial q_g}(q^*(p,0), e^*(p,0)) \frac{\partial \tilde{q}_g}{\partial p_{g'}}(p, V(p,0)) \\
&- \lambda \sum_{g'=1}^{G} p_{g'} \frac{\partial \tilde{q}_{g'}}{\partial p_g}(p, V(p,0)) \\
&+ \sum_{k=1}^{K} \frac{\partial S}{\partial e_k}(q^*(p,0), e^*(p,0)) \frac{\partial \tilde{e}_k}{\partial p_g}(p, V(p,0)) \\
= &- q_g^*(p,0)\theta
\end{aligned}$$

Suppressing arguments for ease of presentation, multiplying by p_g and dividing by the aggregate spending on DRG g, $p_g q_g^*$, we get

$$\frac{\sum_{g'=1}^{G}\Big(\frac{\partial S}{\partial q_{g'}} - \lambda p_{g'}\Big)\Big(\frac{\partial \tilde{q}_{g'}}{\partial p_g} p_g\Big) + \sum_{k=1}^{K} \frac{\partial S}{\partial e_k} \frac{\partial \tilde{e}_k}{p_g} p_g}{p_g q_g^*} = -\theta.$$

We can interpret $\frac{\partial \tilde{q}_{g'}}{\partial p_g} p_g$ as the approximate aggregate amount of quantity $q_{g'}$ induced by the price p_g. The approximation is exact if the derivative is constant, such that

$$\begin{aligned}\frac{\partial \tilde{q}_{g'}}{\partial p_g} p_g &= \int_0^{p_g} \frac{\partial \tilde{q}_{g'}}{\partial p_g} (\bar{p}_g, p_{-g})\, d\bar{p}_g \\ &= \tilde{q}_{g'}(p_g, p_{-g}) - \tilde{q}_{g'}(0, p_{-g})\end{aligned}$$

The product $\frac{\partial S}{\partial q_{g'}} \frac{\partial \tilde{q}_{g'}}{\partial p_g} p_g$ is the total welfare effect that the price p_g has through its effect on quantity $q_{g'}$. The interpretation as a total effect is again based on a linearization. Here, the relevant linearization is of $S(q_{g'}, q_{-g'}, e)$ around $q_{g'}$. Aggregating effects across DRGs g' gives the approximate total welfare effect of setting the price $p_{g'}$ through the quantity of patients induced by this price.

However, inducing patients through prices comes at a budgetary cost. If price p_g induces (approximately) $\frac{\partial \tilde{q}_{g'}}{\partial p_g} p_g$ patients of DRG g', the cost of this induced demand is $p_{g'} \frac{\partial \tilde{q}_{g'}}{\partial p_g} p_g$. Scaling by the Lagrange multiplier λ, which is the shadow cost of the budget constraint, translates the budgetary cost into welfare units.

The final term in the denominator is the approximate total welfare effect of price p_g through effort. This term is also based on a linearization to obtain the approximate total effect of the price p_g on each effort e_k, $\frac{\partial \tilde{e}_k}{p_g} p_g$. The approximate total effect of p_g on effort e_k is then multiplied by its marginal welfare effect, $\frac{\partial S}{\partial e_k}$ to obtain the approximate welfare effect that the price p_g has through effort e_k. Summing over efforts e_k, $k = 1, \ldots, K$, yields the approximate total effect of the price p_g on welfare through effort.