Essays in Applied Microeconomics

Essays in Applied Microeconomics

A dissertation presented

by

Tilman Cornelius Dette

to

The Department of Economics

in partial fulfillment of the requirements

for the degree of

Doctor of Philosophy

in the subject of

Economics

Harvard University

Cambridge, Massachusetts

April 2016

Dissertation Advisor:
Professor Andrei Shleifer

Author:
Tilman Cornelius Dette

Essays in Applied Microeconomics

Abstract

This thesis combines three essays in applied microeconomics. The first essay studies hospital responses to price changes and the introduction of DRG reimbursements; using a large administrative data set on all inpatient hospital admissions in Germany from 2005 to 2013, we find that hospitals respond stronger to financial incentives in areas of higher medical discretion. The second essay studies the effect of two UK compulsory schooling law changes; deriving an optimal pooled regression outcome and pooling data across 50 surveys, I show that the two reforms had no measurable impact on a large set of job market outcomes. The third essay studies the benefit of observing more customer data on optimizations of a large online retailer; predicting optimal product display ranks based on smaller data sets than the actually observed data, we estimate the effect of less data on click and order conversion rates.

Contents

Abstract iii
Acknowledgments ix

Introduction **1**

1 G-DRG Side Effects: Hospital Responses to Germany's Inpatient Reimbursement System **4**

1.1 Introduction 4
1.2 Empirical Setting and Data 9
1.3 Growth in Admissions 15
1.4 Baseline Price Elasticity Regression 20
1.5 Heterogeneous Price Elasticities 27
1.6 Growth Rate and Elasticity Correlations 34
1.7 A Model of Two Financial Incentives and Medical Discretion 37
1.8 Conclusion 44

2 No Significant Returns to Compulsory Schooling: The UK 1947 and 1972 Law Changes Revisited **46**

2.1 Introduction 46
2.2 Estimates Based on Oreopolous (2006) Data 49
2.3 An Alternative Model that Utilizes More Data 54
2.4 LFS Data 58
2.5 $\mathrm{E}\left[s_i \mid y_i\right]$ Estimation and FRD Specification 61
2.6 Results 62
2.7 Conclusion 68

3 Returns to Scale of Data: Evidence from Online Retail **69**

3.1 Introduction 69
3.2 Empirical Setting and Data 72
3.3 Empirical Model of Product Quality 75
3.4 Results 78

3.4.1 Quality Model 79
3.4.2 Effects on Precision 81
3.4.3 Effects on Clicks, Purchases, and Repeat Visits 85
3.5 Discussion 90

References **92**

Appendix A Appendix to Chapter 2 **97**
A.1 Proof of Results in Linear Model 97
A.2 Additional Figures 100

List of Tables

1.1 Summary statistics for §21 G-DRG data 14
1.2 Summary of year-over-year changes in log points 15
1.3 Overall price elasticity estimates for $3 \times 3 \times 5$ alternative specifications. 23
1.4 DRG-price elasticity regression estimates with different sets of controls 26
1.5 Year-over-year price elasticity estimates by hospital department 31
1.6 Year-over-year price elasticity estimates by CCS code 32
1.7 Correlates of elasticity estimates 36
1.8 Correlations of elasticity and annual growth estimates 37

2.1 Replication of Oreopolous (2008) results and alternative specifications 50
2.2 Standard error estimates of z-Score of placebo law changes 51
2.3 Summary statistics for UK Labour Force Surveys 60
2.4 Relative influence of education predictors 63
2.5 OLS and regression discontinuity results for pooled outcome 65

3.1 Effect of out-of-sample quality prediction on economic outcomes 89

List of Figures

1.1 Distribution of $\Delta \ln p_{drg}$ by year weighted by number of affected admissions . 12
1.2 Hospital statistics before and after the 2005 G-DRG system introduction . . . 16
1.3 Admission, surgery, and discharge numbers by hour of day, weekday, and year 18
1.4 Hospitalization rates by age and gender 19
1.5 Scatter plot of $\Delta \ln p_{drg}$ and $\Delta \ln q_{drg}$ 21
1.6 Elasticity estimates by year and by month 28
1.7 Elasticity estimates by age group and gender 30
1.8 Heterogeneous elasticity vs annual growth rate estimates 35

2.1 The impact of two UK compulsory schooling increases on school leaving age 48
2.2 Distributions of z-scores of estimated placebo laws in Oreopoulos (2008) FRD design 51
2.3 Statistics by birth year for data used by Oreopoulos (2008) 53
2.4 Graphical result for effect of compulsory schooling increases on average school leaving age: actual versus the prediction, based on adult outcomes 64
2.5 Graphical result for OLS regression, based on weighted observations included in main regression specification 67

3.1 Position effect (first page) relative to first position 80
3.2 Quality histogram for example category 81
3.3 Variability in quality estimates for example categories and different fractions of training data 82
3.4 Variability in estimates by average number of impressions of products displayed (ϕ_c) 84
3.5 Average out-of-sample quality by position for sort order based on different training data sizes 85
3.6 Gain in expected number of clicks from larger training data 86
3.7 Actual number of clicks against page quality/predicted number 87
3.8 Purchases and repeat impressions against quality 88

A.1 Graphical summary of re-weighted Oreopolous (2008) data 101

A.2 Individual LFS outcomes, averaged by birth cohort. 102
A.3 Observations and control variables for LFS data 102
A.4 Cross-validated MSE by bandwidth and specification 103
A.5 Z-score distribution of placebo test 103

Acknowledgments

I am greatly indebted to my advisors Andrei Shleifer, Ariel Pakes, and David Cutler for their invaluable support, guidance, and feedback on my research. I thank my co-authors Matthias Bäuml and Daniel Pollmann for the great collaboration and the inspiring time we spent together on essays one and three. I am grateful for the advice and support of Johann Blauth, Raj Chetty, Roland Fryer, Nathan Hendren, Greg Lewis, Robert McMillan, Karl Neumar, Jonas Schreyögg, Doug Staiger, Elias Papaioannou, Kamilya Tazhibayeva, and the participants of Harvard's Industrial Organization research seminar.

Introduction

Recent improvements in computational power have increased the opportunities for exciting empirical research. All three essays of this thesis deduce empirical results from data sets of sizes that would have been difficult to analyze ten years ago. The first essay requires the computation of counterfactual reimbursement prices for 144 million hospital admissions, based on detaild discharge records. The second essay uses boosted regression trees to optimally combine detailed survey responses to 50 UK-wide labor force survey waves with 200 outcomes each. The last essay relies on a regression that uses five billion user-product impressions. Thus, each essay leverages computational methods that allow for the estimation of highly granular results. The three papers are briefly summarized below.

G-DRG Side Effects: Hospital Responses to Germany's Inpatient Reimbursement System[1]

Germany's 2005 introduction of a universal DRG system coincided with an increase of both annual admission and cost growth rates, with rates more than doubling to 2.02 and 4.06 percent, respectively. Using 2005 to 2013 administrative data that cover all inpatient hospital admissions in Germany, we study hospital quantity responses to annual changes in financial incentives. We estimate an average price elasticity of 0.259 and elasticity differences by patient age, primary diagnosis CCS, and department specialization. The heterogeneous elasticity estimates correlate positively with our subjective assessment of medical necessity as well as the heterogenous annual growth rates in admissions. We derive a simple static

[1] Co-authored with Matthias Bäuml

model to explain the empirical correlations. In the model, both a switch from per-diem to per-admission reimbursement as well as price increases lead to hospitals admitting more patients; the magnitude of both effects increases with the share of marginal admissions, which we interpret as discretionary medical treatment.

No Significant Returns to Compulsory Schooling: The UK 1947 and 1972 Law Changes Revisited

Previous studies that instrumented years of schooling with the 1947 and 1972 UK compulsory schooling law changes likely underestimated standard errors – replicating the Oreopoulos (2006) analysis with a less biased local linear regression discontinuity design yields a much wider 95 percent confidence interval that includes both zero and the least squares estimate. To attain more statistical power, I pool an average of 194 outcomes on earnings, occupation, family, and health measures into a single score that minimizes the variance of the ratio of instrumental variable and least squares return estimate. Combining 50 waves of the UK Labor Force Survey, I estimate that the 1947 and 1972 UK compulsory schooling law changes had no statistically significant effect on the pooled outcome, with a 95 percent confidence interval of just [-7.4, 23.2] percent relative to the least squares return estimate. The narrow confidence interval is biased corrected (Calonico *et al.*, 2014) and exhibits near correct coverage in a falsification test.

Returns To Scale of Data: Evidence from Online Retail[2]

For a large online retailer, we estimate the returns to observing the search behavior of additional consumers when this data is used to optimize the ranking of products displayed to all consumers, affecting their search and demand. We estimate hypothetical firm beliefs over product quality on training data sets of different size using a revealed-preference approach, and analyze the prediction quality of these estimates on hold-out data. Given a large and heterogeneous product catalog, there is a substantial size-precision gradient, which translates

[2]Co-authored with Daniel Pollmann

into quality differences in the predicted optimal ranking that affect consumer search and purchases beyond the initial visit.

Chapter 1

G-DRG Side Effects: Hospital Responses to Germany's Inpatient Reimbursement System[1]

1.1 Introduction

Compared to cost-plus-based reimbursement, diagnosis related group (DRG) based reimbursement incentivizes hospitals to reduce per-admission costs via yard-stick competition and fewer marginal treatment incentives. However, a switch to DRG reimbursement may also influence hospitals to alter admission decisions, which in turn may offset the per-admission cost savings. Indeed, Germany's 2005 G-DRG introduction was accompanied by an increase in the annual hospital admission growth rate to 2.02 percent. Further, shifts in the composition of admission volumes to more expensive DRGs resulted in annual cost increases of 4.06 percent. To better understand hospitals' admission decision responses to G-DRG financial incentives, we also estimate price elasticities with respect to annual variation in relative DRG prices, inherent to the G-DRG system. With an average elasticity estimate of .259, we find that elasticity differences by patient age, primary diagnosis CCS, and department specialization all correlate

[1] Co-authored with Matthias Bäuml

positively with differences in the respective annual admission growth rates. E.g., the annual number of in-hospital births did not increase since 2005 nor correlate with changing DRG prices, while the number of admissions of 80-year-olds has increased at an annual rate of 5.68 percent and has a relative price elasticity of .361. Contrary to findings on the 1983 introduction of Medicare's DRG system, Germany's DRG system appears to have incentivized immediate increases in discretionary hospital treatments.

Replacing a cost-plus-based system, Germany adopted DRG reimbursements with the primary goal of "improving the cost-effectiveness of hospital care" (Braun *et al.*, 2007). In the new system, all inpatient hospital admissions are mapped to one of around 1,100 DRGs based on diagnoses, demographics and treatment type and intensity information. For each admission, hospitals receive a DRG-specific, predominantly flat rate payment. With payment rates based on average cost reports from a large sample of German hospitals, the DRG systems' inherent yardstick competition rewards hospitals for being more cost-efficient than other hospitals for the same type of admission (Shleifer, 1985). Further, the predominantly flat rate payment eliminates many marginal treatment incentives of the previous cost-plus-based system. Indeed, the average length of stay of German hospital admissions – one driver of costs – has been decreasing at an annual rate of 2.01 percent since 2005; controlling for DRG composition shifts increases the magnitude of this effect to an annual rate of 3.05 percent. Yet, to the degree that DRG-reimbursed average costs exceed the marginal costs of additional admissions, the DRG flat rate payments also incentivize positive extensive margin responses.

While the introduction of Medicare's DRG system did not yield such response (Coulam and Gaumer, 1992),[2] a number of studies identify extensive margin responses to other changing financial incentives for medical care provision. Duggan (2000) finds that both for-profit and not-for-profit hospitals aggressively cream-skim profitable Medicaid admissions away from state-run hospitals in response to a steep increase in the associated reimbursement. Clemens and Gottlieb (2014) estimate a positive price elasticity of physician services of 1.5 in the years

[2] Coulam and Gaumer (1992) provide a summary of studies analyzing the first few years of the introduction of Medicare's prospective payment system. The literature they survey finds significant drops in Medicare admissions immediately following the 1983 DRG system introduction.

following a reimbursement base rate adjustment by Medicare. Yip (1998) on the other hand estimates a backwards bending supply curve for the provision of CABG surgeries in response to Medicare fee cuts.

Acemoglu and Finkelstein (2008) study Medicare's switch from per-diem to DRG-based reimbursements. Based on a simple neoclassical model, they derive that DRG payments increase the returns to capital; consistent with this model, they find that hospitals most affected by the new payment system most strongly increased their technology investments.[3] Consistent with Acemoglu and Finkelstein (2008), we find that German hospitals increased the efficiency of their operating rooms, with the number of coded surgeries per admission increasing annually by an additional 1.96 percent, beyond the 2.02 percent growth in admissions. Further, the growth in hospital admissions is concentrated during the Monday to Friday morning hours while the resulting discharge growth occurs in the afternoon, increasing each admission's share of time spent in the hospital during regular business hours. Similarly, the growth in reported surgical procedures per admission occurs primarily during regular business hours, with annual increases of 2.57 percent Monday to Friday from 6 am to 3 pm, compared to .78 percent during off hours.

This growth in increasing surgical treatment may not be socially optimal. Deyo *et al.* (2010) show that a complex treatment of spinal stenosis increased 15-fold from 2002 to 2007 for Medicare patients, in lieu of two less expensive treatment options; concurrently, the associated rate of major complications, 30-day mortality, and resource use all increased. Jena *et al.* (2015) study hospitalizations with acute cardiovascular conditions at teaching hospitals at the time of the two largest cardiology meetings; they find that treatment intensity falls, while mortality does not change or decreases for all studied conditions. Cutler (1995) finds that payment increases due to Medicare's PPS introduction yielded decreases in in-hospital and 30-day mortality, though one-year mortality did not change. Similarly, Duggan (2000) finds no improvements in one-year mortality due to a great increase in reimbursement for Medicaid

[3] McGuire (2000) provides a broader summary of different incentives of DRG-payments, than the focussed analysis of Acemoglu and Finkelstein (2008).

patients.

The lack of outcome improvements due to additional spending may not be surprising, as hospital treatments increasingly appear to be discretionary (Chandra *et al.*, 2012). Cutler (2014) summarizes that an estimated one third of medical spending in the US may be wasteful, not yielding any medical benefit. Chandra and Skinner (2012) characterize wasteful spending on new unproven technologies as a main source of the rapid growth in health care costs; yet, citing the findings of Schreyögg *et al.* (2009), they also note that unproven technologies may turn out highly beneficial after some initial adoption period. This paper contributes to this larger literature showing that the financial incentives of G-DRG reimbursements appear to be a driver of growth in discretionary care provision.

To link growth in admissions to the strength of financial incentives, we first estimate year-over-year price elasticities at different levels of disaggregation. Estimating elasticities at the level of patient age and gender, primary diagnosis group, and hospital department specialization yields heterogenous elasticity estimates. Notably, areas associated with greater shares of elective surgery and treatment – including CCS codes for prostate cancer, cataract, and hip fracture, and the geriatrics and orthopedics departments – yield some of the highest price elasticity estimates from .375 to .591. At the other end of the spectrum, the pediatrics and pediatric surgery departments yield statistically insignificant point estimates of .017 and .036. More generally, based on our subjective assessment of the heterogenous elasticity estimates, hospitals' responses to annual price changes are stronger in areas of greater medical discretion.

The year-over-year price elasticity estimates are unlikely to only measure changes in hospital admission and treatment decisions, but may also be due to up-coding response as documented by Dafny (2005) and Silverman and Skinner (2004). Hospital doctors and coding staff may have discretion in the exact diagnoses they report, as well as the co-morbidities, and the choice of primary diagnosis for multi-morbid patients. To counteract up-coding incentives, the German regulator requires hospitals to prove medical reasons for annual increases in their case mix index – the average DRG payment rate by admission (Braun *et al.*, 2007). Further, Germany's public insurers have designated 'medical offices' that assess around 13 percent of all submitted hospital claims for correct coding.[4] Nonetheless, not all up-coding may be easy to detect. Coding changes to diagnosis and procedure codes may be especially difficult to detect in areas of medical uncertainty. Undetected up-coding may similarly be more likely in areas of medical discretion. Thus, differences in price elasticities due to up-coding may once more indicate areas of greater medical discretion.

Whether due to up-coding or real treatment changes, the heterogenous elasticity estimates also correlate positively with the different annual growth rates in admissions. Jointly with the different price elasticities, we estimate differences in the admission growth rates since 2005. For distinct regressions by age group, department, and CCS code, as well as a joint regression model that includes all three differentiators, we find that elasticity and growth rate estimates correlate positively at each level. We derive a simple model to explain this empirical result. In the model differences in medical discretion yield differences in price elasticities. Separately, a switch from per-diem to flat-rate reimbursement yields an increase in admissions that also increases in magnitude with greater medical discretion. Thus, differences in medical discretion may explain the observed correlation of the two hospital responses.

The estimated two effects are distinct from early findings on the introduction of Medicare's PPS. Medicare's introduction did not seem to yield increases in Medicare admissions (Coulam and Gaumer, 1992), nor did it yield a G-DRG level of annual price variation that can be used to estimate year-over-year price elasticities. The former effect may be explained by differences in the two payment system's coverage – unlike Medicare's PPS, the G-DRG system covers nearly all in-patient admissions nationwide. The latter result is due to the G-DRG system's price setting mechanism and the detailed data collection efforts by the German regulator. A further analysis of our findings will ideally focus on linking hospital response estimates to medical outcome measures. That is, while our findings indicate that much of the recent G-DRG admission increase may have been in areas of elective treatment, such spending may

[4] The designated 'medical offices' examined 2,533,239 hospital services in 2014 (MDK, 2015) compared to 20,026,210 admissions.

or may not be wasteful. In the conclusion of our paper we suggest one approach to link our findings to mortality outcomes at the age and regional level.

The rest of the paper is structured as follows. Section 1.2 provides more detail on the empirical setting, the sources of annual price variation, and the G-DRG data. Section 1.3 analyzes the growth in admissions and procedures since 2005. Section 1.4 introduces the primary year-over-year price elasticity regression and robustness specifications, while section 1.5 reports more granular elasticity estimates for three differentiating variables. Section 1.6 correlates the elasticity and growth rate estimates. Section 1.7 presents a model to explain the empirical results. Section 1.8 concludes.

1.2 Empirical Setting and Data

DRG payment systems have been adopted by many countries. Starting with Medicare's US-wide prospective payment system (PPS) introduction in 1983, DRG-based systems have become the primary basis of US hospital compensation, with MS-DRGs being used by Medicare and the further refined AP-DRGs and APR-DRGs being used by many private insurers. Following the US, most European countries adopted modified DRG versions to fit their country's need (Busse *et al.*, 2011). Germany developed the G-DRG version using Australia's AR-DRG system as a basis (Schreyögg *et al.*, 2006).

The G-DRG system replaced a largely per-diem-based hospital reimbursement system in 2005, following government legislation in 2000 (Schreyögg *et al.*, 2006). Since 2005, G-DRG reimbursement applies to all inpatient admissions in Germany, irrespective of insurance type.[5] 94 percent of patients in our data have statutory health insurance plans, which feature no deductible and a per-diem copay of 10 Euro that is waived for children, indigents, and long-stays; thus, these patients are not directly affected by changes in G-DRG reimbursement

[5] Only psychiatry and psychosomatic medicine admissions are excluded from the G-DRG system, with a separate DRG-based reimbursement system being developed. Further, while DRG rates also apply to private insurance plans, most of these premium plans reimburse additional preferential services with rates being set independent of the DRG schedule. For related work on Schreyögg *et al.* (2014), we had access to insurance information by admission; applying our elasticity estimation methodology only on the subset of non-statutory health insurance patients yielded near zero elasticity estimates, not statistically different from zero.

rates. Patients can be directly admitted to hospitals for emergency care, while needing a referral by a specialist otherwise; hospitals have some discretion in classifying admissions as emergencies, though. Within three days of admitting a patient, hospitals must submit an 'admission record;' within three days of discharge they must submit a 'discharge record;' subsequently hospitals file a much more detailed 'invoice record.'[6] The final record includes primary and secondary ICD-10 diagnoses codes, Germany-specific OPS procedure codes (similar to Medicare's ICD-9-PCS codes), age, gender, admission and discharge dates and reason codes by department, birth weight for births, and hours of ventilation. Insurance companies keep the final records both for documentation and to input them into a G-DRG 'grouper' - a logic tree that assigns a unique DRG to each admission.

Based on the assigned DRG insurance companies reimburse hospitals. 84.6 percent[7] of hospital remuneration is due to the DRG-specific flat-rate price component p_{drg} that gets multiplied by a hospital-specific "base case value."[8] Another 3.7 and 3.5 percent are due to per-diem deductions and surcharges to p_{drg} for short- and long-stayers, respectively. The remaining 8.2 percent of remuneration are primarily due to additional fees for new services not covered by DRG payments, and surcharges for teaching hospitals. For our analysis, we define the reimbursement price r_i for admission i at hospital h as

$$r_i \quad = \quad p_{drg(i,t)} \left(1 + \nu_{i,t}\right) bcv_{h(i),t} + \xi_{i,t}$$

where ν_i denotes the potential per-diem based adjustments and ξ_i captures the additional fees. Subscript t accounts for the year of admission to reflect that DRG price and grouper logic, base case value and additional adjustments can change every year.

To estimate price elasticities, we compute year-over-year changes in log prices. That is,

---

[6] Requirements for these records are specified by paragraph 4 of SGBV 301 (DKG, 2015).

[7] Percentage calculations are based on data pertaining to the 144,514,504 admissions used in our analysis. Further deduction and surcharge amounts are considered in absolute terms to compute percentages in terms of overall reimbursement volume.

[8] The DRG-specific flat-rate component p_{drg} is also commonly referred to as DRG cost weight or relative weight, denoting only a relative price. Since our price elasticity estimates are based on relative variation in these 'weights,' we denote them as p_{drg} for straightforward comprehension.

we approximate changes in log reimbursements with

$$\begin{aligned}\Delta \ln r_{i,t} &= \ln\left(p_{drg(i,t)}\left(1+\nu_{i,t}\right)bcv_{h(i),t}+\xi_{i,t}\right)-\ln\left(p_{drg(i,t-1)}\left(1+\nu_{i,t-1}\right)bcv_{h(i),t-1}+\xi_{i,t-1}\right)\\ &\approx \Delta\ln p_{drg(i,t)}+\Delta\ln bcv_{h(i),t}+\Delta\ln\left(1+\nu_{i,t}\right).\end{aligned}$$

Given the myriad of regulations involved in determining the $\xi_{i,t}$ error term, and the fact that these additional fees only account for 8 percent of remuneration, we choose to compute price elasticities only with respect to the DRG-based compensation.[9] From 2005 to 2009, $bcv_{h(i),t}$ was negotiated at the hospital level to facilitate an 'alignment' from hospitals' previous per-diem reimbursement levels to common DRG levels at the state level. Including year fixed effects in all regressions, we treat additional variation in $\ln bcv_{h(i),t}$ and $\ln p_{drg(i,t)}$ as orthogonal – for robustness, we include a specification with hospital-year fixed effects, which does not seem to affect our coefficient estimates meaningfully. We also treat variation in $\nu_{i,t}$ and $\ln p_{drg(i,t)}$ as orthogonal, an assumption we also verify in one regression, by explicitly computing the year-over-year changes per-diem deduction and surcharge rates. For all other regressions, we measure elasticity based on variation in $\ln p_{drg(i,t)}$ only, while assuming the other price components into the error term. Noting the dependency of the grouper classification on t, we run all regression at the level of price variation, aggregating admission counts either at the $(drg\,(i,t)\,, drg\,(i,t-1))$ level or at the $(drg\,(i,2005)\,,\ldots, drg\,(i,2013))$ level depending on whether we are exploiting the panel structure of the data.[10]

Relative DRG prices p_{drg} vary from year to year for a number of reasons. To set relative prices, the regulator collects 2-year old cost reports for admissions from around 20 percent of hospitals who participate voluntarily, averaging the reported costs by DRG.[11] Thus, changes

[9] Compared to overall reimbursement elasticities, our estimates will exhibit attenuation bias. Further, by ignoring the additional, albeit small reimbursement $\xi_{i,t}$, our approximation of $\Delta\ln p_{i,t}$ is based on changes relative to a lower average compensation, upward biasing $\Delta\ln p_{i,t}$ and thus downward biasing our elasticity estimates further.

[10] 4.7 percent of admission will be mapped to differing DRG codes in years t and $t-1$ due to grouping logic refinements.

[11] Cost reports for all admissions are collected based on the discharge year. Calculations of average costs are performed the following year and the derived DRG weights are applicable to all hospital patients being admitted yet another year later (InEK, 2007).

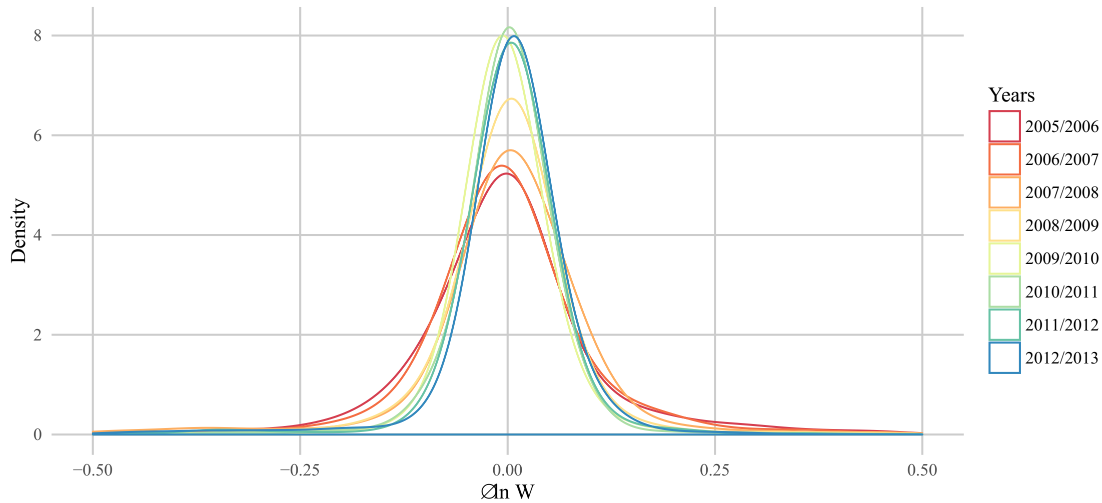

Figure 1.1: *Distribution of* $\Delta \ln p_{drg}$ *by year weighted by number of affected admissions*

in the average reported cost between years $t-3$ and $t-2$ will impact the associated DRG weights in years $t-1$ and t proportionally. Multiple sources cause changes in the average reported cost. First, the finite sample of costs reports and cost heterogeneity within each DRG leads to variance in the average cost estimates. Second, cost reporting practices and the accuracy in cost attribution changed over time (Schreyögg *et al.*, 2006). Third, the set of hospitals reporting costs changed over time. Fourth, year-to-year adjustments in the grouper logic will cause admissions to be grouped together with admissions of different costs, leading to differences in the associated average cost by DRG grouping. And fifth, the actual costs of treating patients of a given DRG may change. Figure 1.1 shows the distribution of resulting year-over-year price changes. To the degree that the direction of the change is not persistent over time, all these changes may be plausible exogenous to changes in hospital costs or patient demand two years down the road.

Some persistence in the actual changes of costs is not unlikely; e.g., doctors may slowly gain more experience with a new technology, realizing increasing productivity over time. In this case, changes in costs from year $t-3$ to $t-2$ may be predictive of changes in costs from year $t-1$ to t. However, we would expect changes in costs from year $t-2$ to $t-1$ to be similarly predictive of changes in costs from year $t-1$ to t. Thus, we can at least partially control for this type of endogeneity by adding $\Delta \ln p_{drg(i,t+1)}$ as an additional regressor.[12] Similarly, we can also add $\Delta \ln p_{drg(i,t+2)}$ as a regressor to reflect changes in average costs from year $t-1$ to t, as measured by the regulator, and we can add observed changes in cost proxies as further controls. In both instances we might be over-controlling, though. That is, reported average costs may also be a function of $\Delta \ln p_{drg(i,t)}$ and our outcome variable, as hospitals may adjust treatment and admit different marginal patients in response to price changes. Still, we find that neither control changes our estimates substantially, yielding only slightly higher point estimates.

Our estimates are based on aggregations of the complete set of G-DRG 'invoice-records' for hospital admission between 2005 and 2013.[13] These records cover all inpatient hospital stays and include diagnosis, procedure, treatment date, patient demographic, and hospital department variables. Records do not include physician or patient identifiers that would allow tracking beyond a single treatment episode. We restrict our analysis to 'primary' hospital admissions, excluding the 8.3 percent of reported 'semi-residential' and occupancy ward admissions.[14] We also exclude another 1.06 percent of admissions that are not reimbursed based on DRG flat rate payments. Table 1.1 reports basic summary statistics for the remaining 144,514,504 admissions which are linked to up to 1,672 hospitals in a given year.

We map each admission to their respective 2005 to 2014 DRG codes based on regulator certified DRG grouping software. We also use crosswalk tables from the Germany's ICD-10-GM diagnosis codes to Clinical Classifications Software (CCS) codes by the Agency for Healthcare Research and Quality, assigning each admission to one of 259 clinical indication

---

[12] This statement can be formalized, modeling changes in actual costs as an auto-regressive process. Given that adding $\Delta \ln p_{drg(i,t+1)}$ has no measurable impact on our estimates, we omit this formalization.

[13] This main data source is made available by the Research Data Centers of Germany's Federal Statistical Office. See `http://www.forschungsdatenzentren.de/bestand/drg/index.asp` for more information. The data have been checked for consistency by the regulator.

[14] While hospitals are also compensated via DRG flat rate payments for 'semi-residential' and occupancy ward admissions, compensation and incentives can differ substantially. The english translations "primary," "semi-residential," and "occupancy ward" refer to the German terms "Hauptabteilung," "teilstationär," and "Belegabteilung."

Table 1.1: *Summary statistics for §21 G-DRG data*

	Admissions		Average per admission							Distinct	
Year	Total	/1000	Age	Female	LOS	ICD	OPS	OPS5	CMI	DRGs	Hs
2005	14,703,449	178.5	53.3	0.534	7.84	4.97	2.55	0.761	1.08	844	1,672
2006	15,029,997	182.7	53.7	0.532	7.66	5.08	2.63	0.784	1.08	912	1,648
2007	15,431,815	187.9	54.0	0.532	7.51	5.20	2.73	0.810	1.09	1,035	1,631
2008	15,842,720	193.4	54.3	0.531	7.32	5.35	2.80	0.821	1.10	1,089	1,613
2009	16,133,547	197.3	54.7	0.530	7.18	5.56	2.95	0.853	1.12	1,146	1,588
2010	16,401,731	202.4	54.9	0.529	7.03	5.76	3.04	0.877	1.12	1,154	1,574
2011	16,707,679	207.7	55.2	0.527	6.87	5.92	3.12	0.889	1.12	1,149	1,550
2012	17,029,945	211.2	55.5	0.526	6.77	6.08	3.16	0.895	1.12	1,148	1,525
2013	17,233,621	212.8	55.6	0.526	6.66	6.30	3.15	0.889	1.12	1,142	1,512
All	144,514,504	197.1	54.6	0.530	7.18	5.60	2.92	0.844	1.11	9,619	†

Notes: Summary statistics are by year of admission date. The "/1000" column reports the number of admissions per 1000 residents. Abbreviated columns denote: LOS – length of stay, ICD – number of diagnosis codes, OPS – number of procedure codes, OPS5 – number of surgical procedure codes, CMI – case mix index, DRGs – DRG codes, Hs – hospitals. †The number of distinct hospitals is calculated based on discharge year, as hospital identifiers are discharge year-specific and can different between years. Thus, we do not calculate the total number of distinct hospitals across all years. Data source, description, and restrictions are in the text.

groups based on their primary diagnosis (HCUP, 2009). We use DRG assignments to compute the actual and nine counterfactual DRG reimbursement prices, while using the other mapping to estimate heterogeneous price elasticity.

Previewing results of the next section, table 1.2 summarizes some of the observed year-over-year changes in log point differences. Notably, the case mix (CM) – the sum of normalized[15] DRG prices $p_{drg(i)}$ – grew at an even greater annual rate than admissions. That is, the average severity of hospital admissions as proxied for by $p_{drg(i)}$ increased over time. Concurrently, the average length of stay decreased at an annual rate of 2.02 log points. This effect is amplified if we control for the compositional shift of DRG admissions towards DRGs of greater clinical severity, by holding $(drg\,(i,t)\,, drg\,(i,t-1))$ shares to $t-1$ levels in our aggregations. We also note the 10.4 log point standard error of $\Delta p_{drg(i,t)}$; that is the G-DRG price variation is both sufficient for estimation and makes the elasticity estimates of section 1.4 economically

[15] Relative DRG prices were initially normalized so that the average hospital admission has a relative price of 1. For subsequent years, the relative prices are normalized so that the case mix of admission under the new weights equals the case mix of the previous year, on the basis of the admission composition from two years prior (InEK, 2007).

Table 1.2: *Summary of year-over-year changes in log points*

Change	Adm	CM	LOS (1)	LOS (2)	ICDs (1)	ICDs (2)	OPSs (1)	OPSs (2)	OPS5s (1)	OPS5s (2)	s.e. of $\Delta \ln p$
2005/2006	2.46	2.70	-2.27	-3.38	2.30	1.38	3.29	1.92	2.89	1.78	13.00
2006/2007	2.61	3.46	-2.07	-3.73	2.44	1.27	3.57	2.06	3.19	1.77	12.52
2007/2008	2.51	3.55	-2.50	-3.62	2.82	2.06	2.62	1.47	1.18	-0.11	12.91
2008/2009	1.78	2.73	-1.97	-3.40	3.86	2.86	4.94	3.57	3.76	2.07	9.73
2009/2010	1.59	2.02	-2.07	-3.22	3.53	2.67	3.32	2.14	2.77	1.62	8.44
2010/2011	1.84	2.16	-2.28	-2.75	2.66	2.12	2.33	1.71	1.29	1.26	7.79
2011/2012	2.01	1.70	-1.52	-1.95	2.62	2.14	1.36	1.04	0.91	1.18	8.67
2012/2013	1.23	1.67	-1.48	-2.38	3.69	2.87	-0.14	-0.57	-0.44	0.16	9.86
Average	2.00	2.50	-2.02	-3.05	2.99	2.17	2.66	1.67	1.94	1.22	10.37

Notes: Changes in Admission (Adm) and case mix (CM) are computed as $\Delta \ln \sum_{i \in S_t} 1$ and $\Delta \ln \sum_{i \in S_t} p_{drg(i)}$, where S_t denotes the set of year t admissions. For admission averages of length of stay (LOS), and number of diagnosis (ICDs), procedure (OPSs), and surgical codes (OPS5s) we compute two averages: (1) the year-by-year averages over all admissions, (2) the DRG specific year-by-year average, averaged by weighting by the admission counts of the first year. The respective columns report the differences across subsequent years in log points. The final column reports the standard deviation of DRG weight changes, weighted by the affected revenue in adjacent years.

relevant.

1.3 Growth in Admissions

The introduction of the G-DRG system changed hospitals' financial incentives. With marginal costs generally lower than average reimbursed cost and reimbursement changing from per-diem to per-admission payments, the G-DRG system rewards hospitals for admitting more patients while decreasing the length of stay for a given medical episode. Indeed, since the 2005 introduction, admission numbers have been growing, while average length of stay has been decreasing. To account for general trends preceding the G-DRG introduction, figure 1.2 plots trends in four statistics from 1996 to 2014. Contrary to findings on the introduction of Medicare's PPS, admission growth rates increased significantly post-G-DRG introduction, while length of stay already declined at similar rates prior to 2005. At the risk of over-interpreting simple difference estimates, we note that inpatient hospital costs have been growing at an annual rate of 4.06 percent since 2005 compared to prior rate of 1.88 percent. The introduction of the G-DRG system also coincided with a halt in the decline of hospital

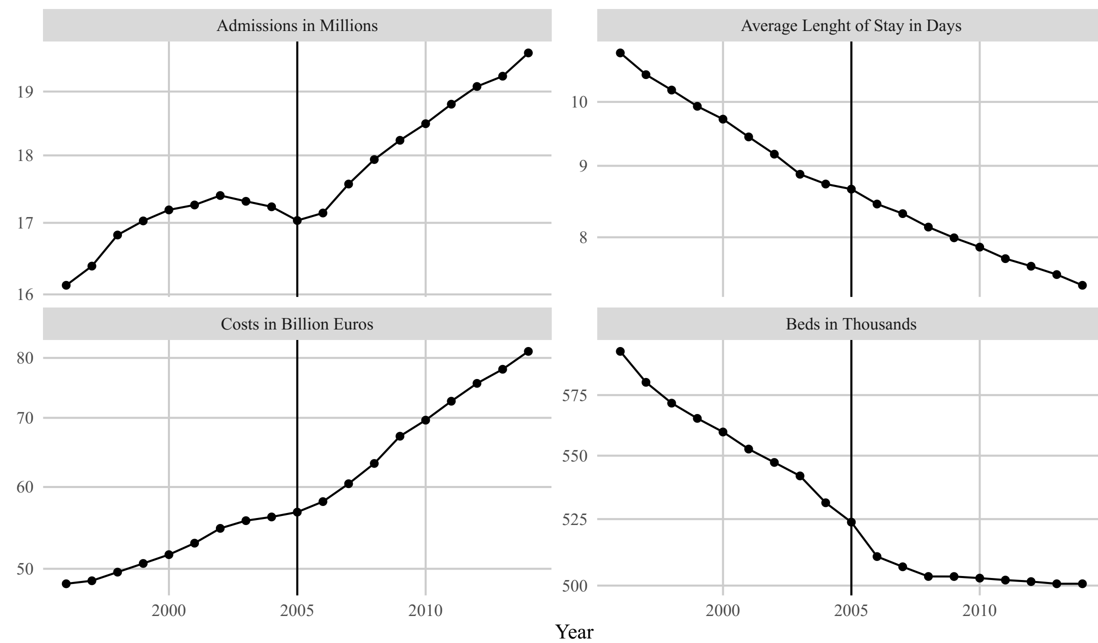

Figure 1.2: *Hospital statistics before and after the 2005 G-DRG system introduction*

Notes: Based on two hospital statistics publications by Germany's Federal Statistical Office ("Grunddaten der Krankenhäuser 2014" and "Kostennachweis der Krankenhäuser 2014").

bed capacity, accommodating the increased admission growth rates.

To better document the admission increases since 2005, figure 1.3 shows admission, surgery and discharge numbers by hour of day, weekday, and year. For each plotted year, admission numbers increase the strongest during Monday to Friday morning hours between 6 and 8 am; surgery numbers increase the most between 7 am to 3 pm; by comparison, discharge numbers increase most in the later hours of Monday to Friday, and on the weekend. Thus, over the years patients tend to spend more of their time in the hospital during regular business hours. Surgeries also increase most during regular business hours, even in relative terms, with per admission increases of 2.62 percent during regular business hours compared to a .82 percent increases during off hours. Consistent with the findings of Acemoglu and Finkelstein (2008), German hospitals seem to have increased their regular business hour efficiency since the introduction of G-DRG reimbursements.

Some of the admissions growth under G-DRG reimbursements may also be explained by Germany's aging population requiring more medical services. While Germany's population has not increased since 2005, life expectancy and average age have increased (table 1.1). Thus, a simple model of increasing medical utilization by age may also predict increases in medical care. Merging population statistics by year, age, and gender, we estimate the decomposition

$$E\left[\ln m_{a,g,t} - \ln pop_{a,g,t}\right] = \rho_{a,g} + \delta_{a,g}\left(t - \bar{t}\right) \tag{1.1}$$

via least squares, where a denotes age in years, g denotes gender, t denotes year, $m_{a,g,t}$ denotes a medical utilization metric, and $pop_{a,g,t}$ denotes the number of residents by grouping. $\rho_{a,g}$ and $\delta_{a,g}$ estimate averages for base utilization rate and annual increases for each age-gender group. Figure 1.4 shows $\rho_{a,g}$ and $\delta_{a,g}$ estimates for two utilization metrics as well as similar estimates for the unnormalized utilization metrics. Indeed, estimated annual increases $\delta_{a,g}$ are lower for per resident metrics. Compared to the raw rates of 2.53 percent and 4.78 percent, the case-mix index and the number of coded procedures increased at annal rates of 1.78 and 4.17 percent after controlling for age and gender specific resident number changes. Thus, the impact of Germany's aging population on admission number increases seems limited.

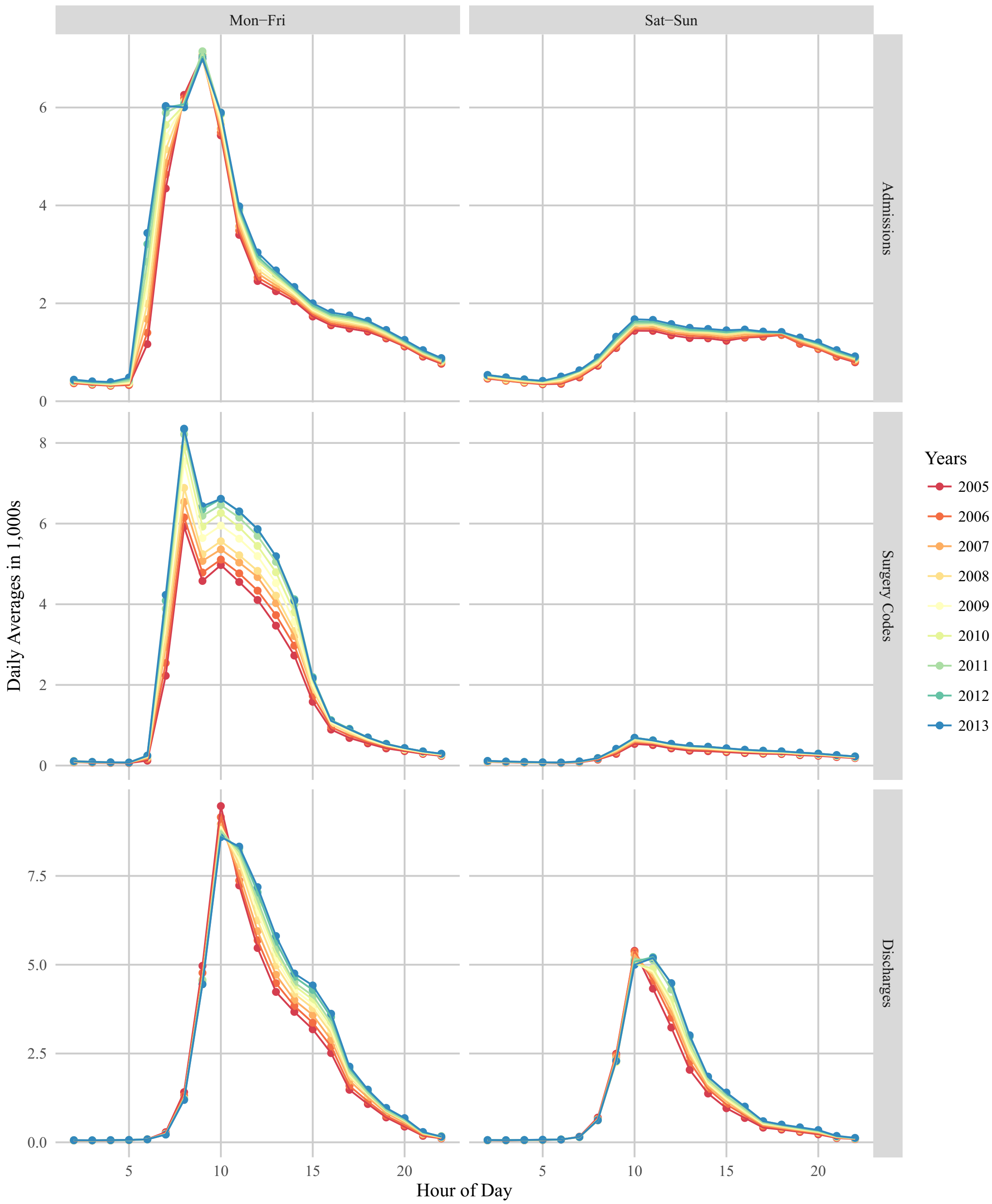

Figure 1.3: *Admission, surgery, and discharge numbers by hour of day, weekday, and year*

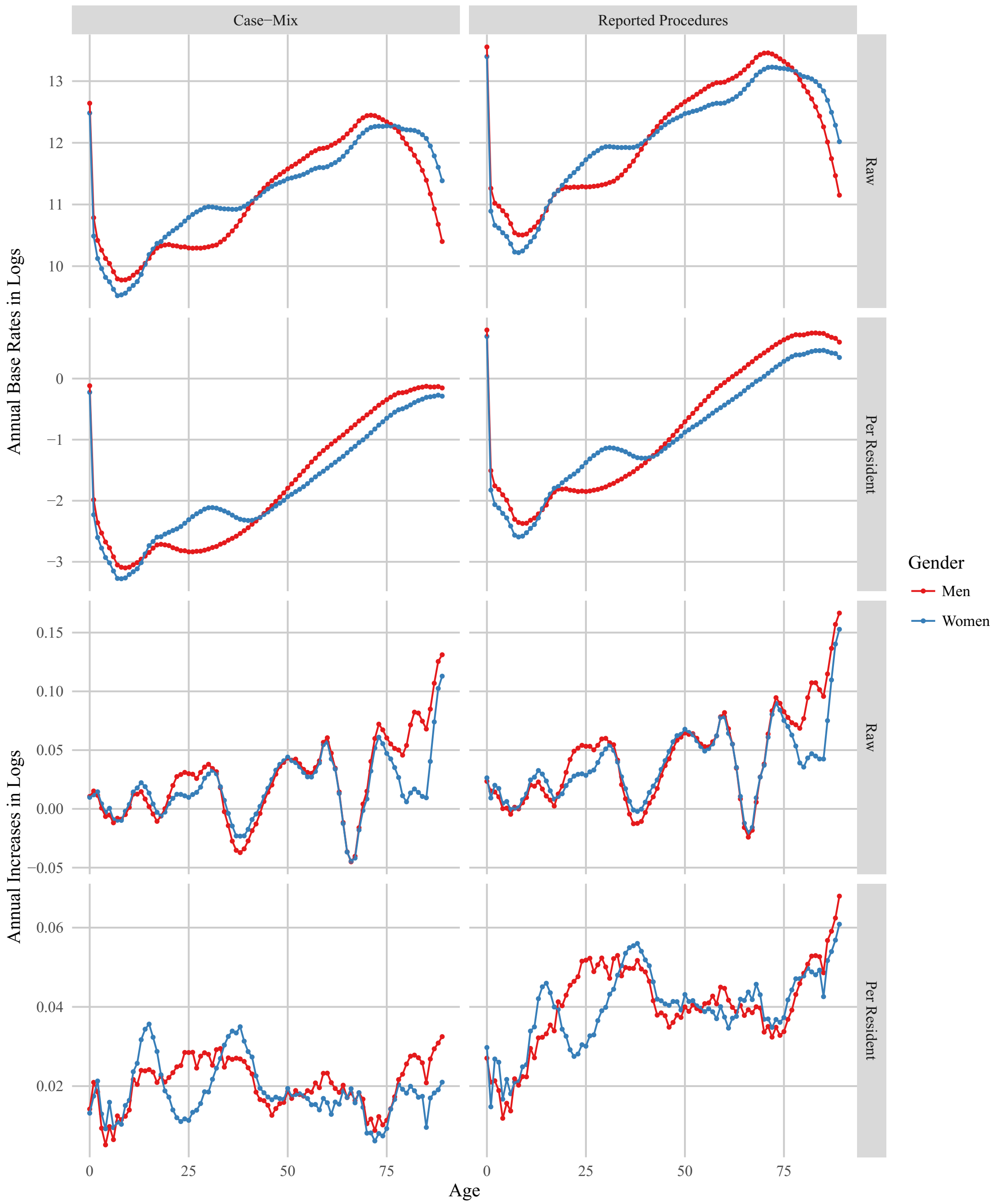

Figure 1.4: *Hospitalization rates by age and gender*

1.4 Baseline Price Elasticity Regression

We quantify hospital admission responses to financial incentives, positing a reduced form, constant elasticity model. Specifically, for aggregation level j in year t the expected number of admissions $\mathrm{E}\left[q_{j,t}\right]$ is a function of aggregation level fixed effect A_j, year fixed effect B_t, aggregation level annual growth rate δ_j and price $p_{j,t}$, with

$$\mathrm{E}\left[q_{j,t}\right] = A_j B_t e^{\delta_j t} p_{j,t}^{\varepsilon_j}, \tag{1.2}$$

where ε_j denotes the aggregation level specific constant price elasticity and B_1 is normalized to 1. For our primary regression setup, we aggregate admissions at the level of year-over-year price changes ($\Delta \ln p_{x,t} = \ln p_{drg(x,t)} - \ln p_{drg(x,t-1)}$), that is, we let j denote $(t_j, drg_{j,1}, drg_{j,0})$-combinations. We compute admissions counts as

$$\begin{aligned} q_{j,t} &= \sum_i 1\left\{t_i = t \wedge drg\left(i,t\right) = drg_{j,1} \wedge drg\left(i,t-1\right) = drg_{j,0}\right\} \\ q_{j,t-1} &= \sum_i 1\left\{t_i = t-1 \wedge drg\left(i,t\right) = drg_{j,1} \wedge drg\left(i,t-1\right) = drg_{j,0}\right\} \end{aligned}$$

for all occurring $(t_j, drg_{j,1}, drg_{j,0})$-combinations for which $q_{l,t_l} + q_{l,t_l-1} \geq 1000$.[16]

To visualize the empirical dependency of $q_{j,t}$ on $p_{j,t}$ non-parametrically, figure 1.5 plots a binned scatter-plot of changes in log quantities and prices. Specifically, we aggregate the 9,044 $(t_j, drg_{j,1}, drg_{j,0})$-combinations into 40 bins based on the corresponding $\Delta \ln p_{j,t}$ for $|\Delta \ln p_{j,t}| \leq .5$. For observations in each bin we compute the weighted mean of $\Delta \ln p_{j,t}$ and log difference in the weighted sum of $q_{j,t}$'s and $q_{j,t-1}$'s.[17] The figure visualizes two primary

[16]The minimum quantity restriction only excludes .8 percent of admissions, reduces sensitivity to outlier observations, and reduces computational requirements.

[17]For the aggregations of figure 1.5 we effectively weight each $(t_j, drg_{j,1}, drg_{j,0})$-cell by $\sqrt{q'_{j,t} + q_{j,t-1}}$: we split cells into 40 bins of equal weight along their associated value of associated $\Delta \ln p_{j,t}$. For cells j in any given bin B, we compute the weighted mean of $\Delta \ln p_{j,t}$ and define

$$\Delta \ln q_B = \ln\left(\sum_{j \in B} q_{j,t}\left(q_{j,t} + q_{j,t-1}\right)^{-\frac{1}{2}}\right) - \ln\left(\sum_{j \in B} q_{j,t-1}\left(q_{j,t} + q_{j,t-1}\right)^{-\frac{1}{2}}\right).$$

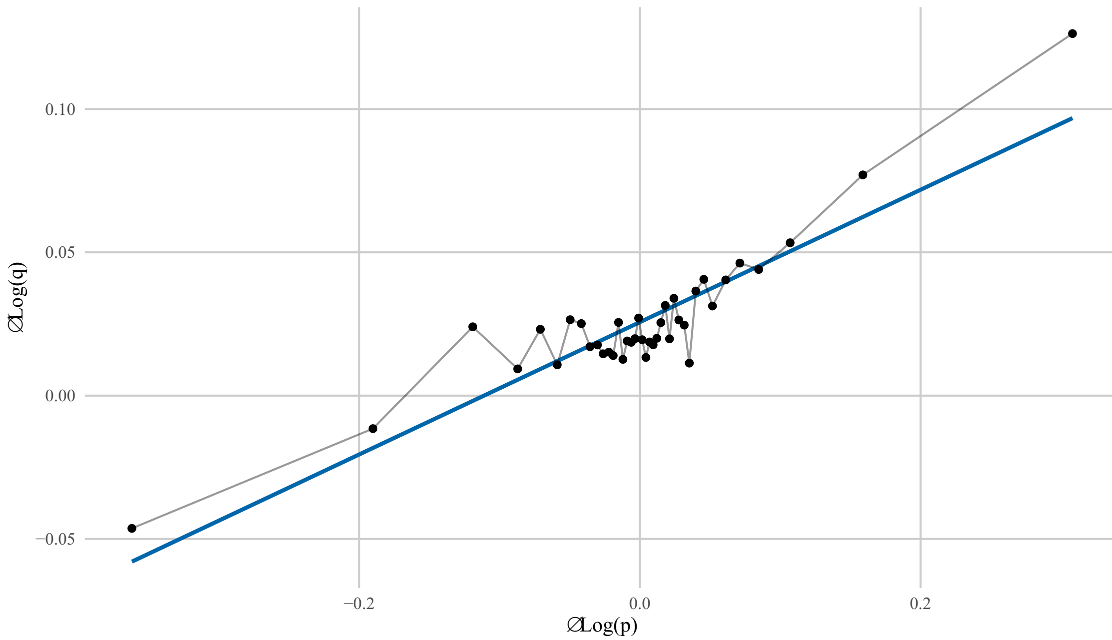

Figure 1.5: *Scatter plot of* $\Delta \ln p_{drg}$ *and* $\Delta \ln q_{drg}$

Notes: Plotted points are based on aggregations of observations from our main regression into 40 bins. Line visualizes least squares fit to plotted points.

findings of this paper. It shows a positive price elasticity – the least square fit has a slope of .23 – and annual growth in overall admissions, with a least square intercept at .026. Notably, the relationship between $\Delta \ln p_{drg}$ and $\Delta \ln q_{drg}$ does not appear fully linear; price declines of less than 10 percent seem to yield little to no negative response relative to a 0 percent price change.[18] Nonetheless, overall the binned scatter plot shows a strong positive correlation between $\Delta \ln p_{drg}$ and $\Delta \ln q_{drg}$.

To estimate the implied elasticity, we rely on a fixed effect poisson regression. The maximum likelihood estimator of the poisson regression is equivalent to the GMM estimator of equation (1.2) using all right hand side variables directly as instruments; Gourieroux *et al.* (1984) provide an alternative justification for using a poisson regression to estimate equation (1.2). We estimate

$$\mathrm{E}\left[q_{j,t}\right] = \exp\left(\varepsilon_j \ln p_{j,t} + \delta_j t + \ln A_j + \ln B_t\right)$$

by conditioning on $\sum_t \mathrm{E}\left[q_{j,t}\right]$ to partial out the $\ln A_j$ fixed effects (Hausman *et al.*, 1984).[19] For our primary regression we estimate a single constant elasticity estimate $\hat{\varepsilon}$, additionally controlling for overall year-over-year changes with the time fixed effects ($\Delta \ln B_t$). Estimates for $\hat{\varepsilon}$ are presented in table 1.3. For all of our regressions, we compute standard errors clustered at the base-DRG level.[20]

To check the robustness of our poisson regression estimates, we also estimate equation (1.2) via two different linear approximations. For the "deviation" approximation, we use a

[18] While this finding may allow for interesting behavioral interpretations, such interpretations are outside of the scope of this paper.

[19] (Hausman *et al.*, 1984) show that the resulting estimator coincides with McFadden (1973)'s multinomial logit specification. We use this equivalency to estimate the fixed effect poisson regression model using existing logit software packages. To add a second set of nuisance fixed effects with many levels while keeping estimation computationally tractable, we rely on a maximization routine by Turner and Firth (2007) and compute the clustered variance-covariance matrix using the inverse formula of block matrices.

[20] The G-DRG groupings are denoted via a 'letter–two digits–letter' code. The first three characters denote the "base-DRG," while the last letter may differentiate between severity levels where applicable. Due to year-to-year DRG grouping logic adjustments, 4.7 percent of admissions would receive a different DRG code if admitted in the previous or subsequent year. Yet, only 1.0 percent of admissions would see a change in "base-DRG" in our regression specifications, motivating our choice of clustering. Our findings are not sensitive to excluding this 1.0 percent of admissions.

Table 1.3: *Overall price elasticity estimates for $3 \times 3 \times 5$ alternative specifications.*

Model	Observation Weights				
	Case Mix	LOS	Adm	$\sqrt{\text{Adm}}$	1
FE Poisson	0.258***	0.289***	0.205***	0.228***	0.246***
	(0.028)	(0.032)	(0.023)	(0.021)	(0.023)
Deviations	0.271***	0.305***	0.213***	0.241***	0.265***
	(0.031)	(0.038)	(0.025)	(0.026)	(0.029)
Log-Linear	0.263***	0.294***	0.210***	0.235***	0.255***
	(0.029)	(0.033)	(0.023)	(0.022)	(0.024)
Price change restriction: $\lvert\Delta \ln p\rvert \leq .5$, $N = 8{,}679$					
FE Poisson	0.259***	0.283***	0.219***	0.240***	0.251***
	(0.033)	(0.034)	(0.027)	(0.021)	(0.019)
Deviations	0.266***	0.289***	0.223***	0.246***	0.259***
	(0.035)	(0.036)	(0.028)	(0.023)	(0.021)
Log-Linear	0.260***	0.285***	0.221***	0.242***	0.253***
	(0.033)	(0.034)	(0.027)	(0.021)	(0.019)
Price change restriction: $\lvert\Delta \ln p\rvert \leq .1$, $N = 6{,}695$					
FE Poisson	0.169***	0.213***	0.168***	0.176***	0.173***
	(0.042)	(0.046)	(0.037)	(0.027)	(0.028)
Deviations	0.169***	0.213***	0.169***	0.177***	0.173***
	(0.042)	(0.046)	(0.037)	(0.027)	(0.028)
Log-Linear	0.169***	0.213***	0.168***	0.176***	0.173***
	(0.042)	(0.046)	(0.037)	(0.027)	(0.028)

Notes: This table reports admission price elasticity estimates with respect to DRG price changes for three types of regression models, five alternative observation weightings, and three different data tructations. The three regression models and five observation weight definitions are described in the text. All regressions are based on aggregations of 144,514,504 hospital admission at $N = 9{,}044$ (t, drg_t, drg_{t-1})-levels. Standard errors – clustered at level of 663 'Base-DRGs' – are reported in parentheses. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

tailor approximation around $\bar{p}_t = \frac{1}{T}\sum_t p_{j,t}$, and take first differences to obtain

$$\left(E\left[q_{j,t}\right] - E\left[q_{j,t-1}\right]\right) / \left. E\left[q_j\right]\right|_{p_{j,t}=\bar{p}_j} \approx \varepsilon_j \left(p_{j,t} - p_{j,t-1}\right)/\bar{p}_j + \delta_j + \Delta \ln B_t \tag{1.3}$$

Approximating $\left. E\left[q_j\right]\right|_{p_{jt}=\bar{p}_j}$ with observed mean outcome $\bar{q}_t = \frac{1}{T}\sum_t q_{j,t}$, we can estimate equation (1.3) via least squares.[21] Alternatively, for the "log-linear" approximation, we further restrict our regression dataset to all j for which $\min_t q_{j,t} \geq 50$, excluding only .047 percent of the underlying admissions data. Substituting $q_{j,t}$ for $\mathrm{E}\left[q_{j,t}\right]$ in equation (1.2) and taking logs and first differences yields regression equation

$$\Delta \ln q_{j,t} \approx \varepsilon_j \Delta \ln p_{j,t} + \delta_j + \Delta \ln B_t + \eta_{j,t}$$

with additional error term $\eta_{j,t}$. Table 1.3 reports elasticity estimates for all three regression models limited to estimating a single $\hat{\varepsilon}$ and year fixed effect controls. Notably, across all weighting and censoring specifications, the three models yield very similar estimates with differences that are much smaller than estimated standard errors; indeed, for the specifications limited to price changes less than 10 log points coefficient estimates are near identical.[22]

We estimate each model using five different effective weightings for each (j,t)-aggregate. We define

$$w_j^{Case\text{-}Mix} = T_j^{-2}\sum_t q_{j,t}\sum_t p_{j,t}, \qquad w_j^{Los} = T_j^{-2}\sum_t q_{j,t}\sum_t \frac{los_{j,t}}{q_{j,t}},$$

$$w_j^{Adm} = T_j^{-1}\sum_t q_{j,t}, \qquad w_j^{\sqrt{Adm}} = \left(T_j^{-1}\sum_t q_{j,t}\right)^{\frac{1}{2}}, \qquad w_j^1 = 1.$$

Thus, the different w_j's weight observations by the affected Case-Mix – the number of admission times the relative reimbursement price, the affected hospital days – the number of admissions times the average length of stay (los), the number of affected admissions, the square root of

[21] The latter approximation is upward biased on average by Jensen's inequality. Similarly, the taylor approximations for the effect of changes in $p_{j,t}$ yield similar bias, both of which will be small for small changes in $p_{j,t}$.

[22] The two linear approximations offer computational advantages, when using existing software packages. Further, the deviations approximation also allows for zero $q_{j,t}$ observations, only requiring the $\Delta \ln p_{j,t}$ to be 'small' to limit bias. Thus, the deviations model may allow a computationally tractable estimation of more granular specifications than presented in this paper, without additional hardware requirements.

the number of affected admissions, and equal weight. While the first three definitions are economically motivated, the fourth definition can be the variance minimizing estimator under a certain set of assumptions.[23] The fifth definition is included for robustness purposes. The poisson regression effectively weights observations by $q_{j,t}$, while the two linear approximations do not. Thus, for the poisson regression, we divide all weights once more by $T_j^{-1} \sum_t q_{j,t}$.

Differently specified observation weights yield some differences in elasticity estimates. As the different columns of table 1.3 show, estimates for the case-mix and the hospital days weighting are slightly higher than estimates for the admissions only weights. To the degree that admissions with higher price or longer length of stay offer more absolute optimization reward per admission, a higher overall elasticity estimate is plausible. Further, the square root admission and equal weight regression also yield slightly higher elasticity estimates than the admission weighting. Both these weightings assign more weight to observations with larger price changes,[24] which in turn are associated with greater elasticities. The case-mix weighting best reflects financial relevance to hospitals. To a lesser degree it may also best reflect medical severity – to the degree that prices reflect average costs which in turn are associated with medical severity. Thus, we choose to use the case-mix weighting for all subsequent regressions.

As figure 1.5 indicates, elasticity estimates may vary with the size of price changes. We estimate all our specifications for all 9,044 $(t_j, drg_{j,1}, drg_{j,0})$-combinations described above, as well as two truncations based on the magnitude of relative price change $\Delta \ln p_{j,t}$. Indeed, restricting our main regression to price changes of 10 log points or less yields a lower elasticity estate of .169. Censoring the 365 potential outlier observations with price changes greater than 50 log points does not seem to affect our estimates. To avoid sensitivity to these observations in our subsequent, more granular regressions, we continue to use this censoring. Thus, for our

[23] Without deriving this result for a specific set of assumptions, we note that the square root admission weight indeed tends to yield the lowest standard errors. Yet, as our single elasticity estimate yields an average over heterogenous elasticities, we prefer a weighting that may yield a less precisely estimated average elasticity but is economically motivated.

[24] Cells with a larger amount of admissions have price changes of smaller magnitude on average. Based on the sources of price changes described in section 1.2, this result is not surprising given that most price change sources will have have less of an impact whenever the average cost calculation is based on more admissions.

Table 1.4: *DRG-price elasticity regression estimates with different sets of controls*

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
$\Delta \ln p$	0.259***	0.251***	0.270***	0.284***	0.288***	0.261***	0.245***
	(0.033)	(0.035)	(0.038)	(0.035)	(0.039)	(0.043)	(0.037)
$\Delta_{+1} \ln p$		-0.109***	-0.091*		-0.084*	-0.097**	
		(0.030)	(0.036)		(0.035)	(0.036)	
$\Delta_{+2} \ln p$			-0.233***		-0.207***	-0.204***	
			(0.033)		(0.032)	(0.034)	
Δ ln LOS				-0.118	-0.013		
				(0.092)	(0.097)		
Δ ln PCCL				-0.868***	-1.102***		
				(0.178)	(0.256)		
Δ ln #OPS				-0.282***	-0.244**		
				(0.081)	(0.082)		
Hospital-Year FEs						X	
Full DRG price							X
Effective N	2732	2728	2372	2732	2372	1532803	5381

Notes: All regression estimate the elasticity of year-to-year admission volume at the $l = (DRG_t, DRG_{t+1})$ level, based on aggregations of 162,138,540 hospital admissions. Regressions are estimated via a two period poisson fixed effect regression, include year fixed effects and are weighted by the affected average case-mix. Regression specification (7) estimates the price elasticity with respect to log changes in the total sum of DRG-related compensation, rather than just the DRG flat-rate. Standard errors – clustered at level of 621 'Base-DRGs' – are reported in parentheses. $^*p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$. The effective N statistic reports the inverse of the sum of squared, normalized weights $\left(\sum_l \omega_l^2 / \left(\sum_l \omega_l\right)^2\right)^{-1}$.

final base line specification regression – the case-mix weighted, fixed-effect poisson regression – we estimate an average price elasticity of .259.

Starting from this base line regression, table 1.4 presents estimates after adding the different sets of controls discussed in section 1.2. Adding future price changes – to control for a persistence in the direction of price changes – does not affect our main point estimate of .259 significantly. Notably, the future price change that reflects changes in current average reported costs yields a negative coefficient of similar magnitude to the price elasticity coefficient. We might interpret this correlation as an indication that changes in average costs have a similar impact on admission numbers as changes in price. Yet, changes in reported average costs are a direct result of hospitals's changes in who they admit; thus, simultaneity bias limits any causal interpretation of the observed correlation.

Adding three cost proxies to our main regression increases our elasticity point estimate to .284. The signs of the coefficients of all three proxies are negative, as expected. Interestingly, the coefficient estimate for changes in average length of stay is not significant. However, changes in average patient clinical complexity level (PCCL)[25] have a very significant negative impact with a t-statistic of 4.7. Similarly, changes in the average number of reported procedure codes are associated with a highly significant elasticity estimate of -.282. As in the case of changes in the average reported cost, the procedure code estimate has the same magnitude as our main price elasticity estimate. Though once more, due to potential simultaneity, we do not interpret this correlation as necessarily causal. Additionally including our future price change controls further increases the price elasticity point estimate to .288. Yet, even this larger point estimate is still within one standard error of our main elasticity estimate of .259. Our subsequent robustness tests also yield point estimates within one standard error of .259, including controlling for changes in 'base-case-values' via hospital year fixed effects, for trends by DRG-path, and explicit calculations of changes in per-diem length of stay discounts and surcharges. Thus, in subsequent regressions we treat the primary price changes as plausibly exogenous. That is, we do not add cost controls when estimating more granular year-over-year elasticities.

1.5 Heterogeneous Price Elasticities

Allowing for heterogeneity in price elasticities, we interact $\ln p_{j,t}$ with several indicator variables to estimate elasticities at different levels. Figure 1.6 plots elasticity coefficient estimates by year-over-year change and by month of admission. For the former differentiation, we estimate distinct $\hat{\varepsilon}_{t,t-1}$ coefficients for year changes '05/'06 to '12/'13. As indicated by the least squares fit, the average year-over-year price elasticity appears to be falling over the years, although the decline is not statistically significant. The $\hat{\varepsilon}_{t,t-1}$ coefficients are less precisely estimated

[25] PCCL aggregates clinical complexity levels associated with a patients primary and secondary diagnoses, see appendix C of InEK (2015) for definition. The scaling of this index has no immediate interpretation. PCCL cutoff values are used by the DRG logit tree.

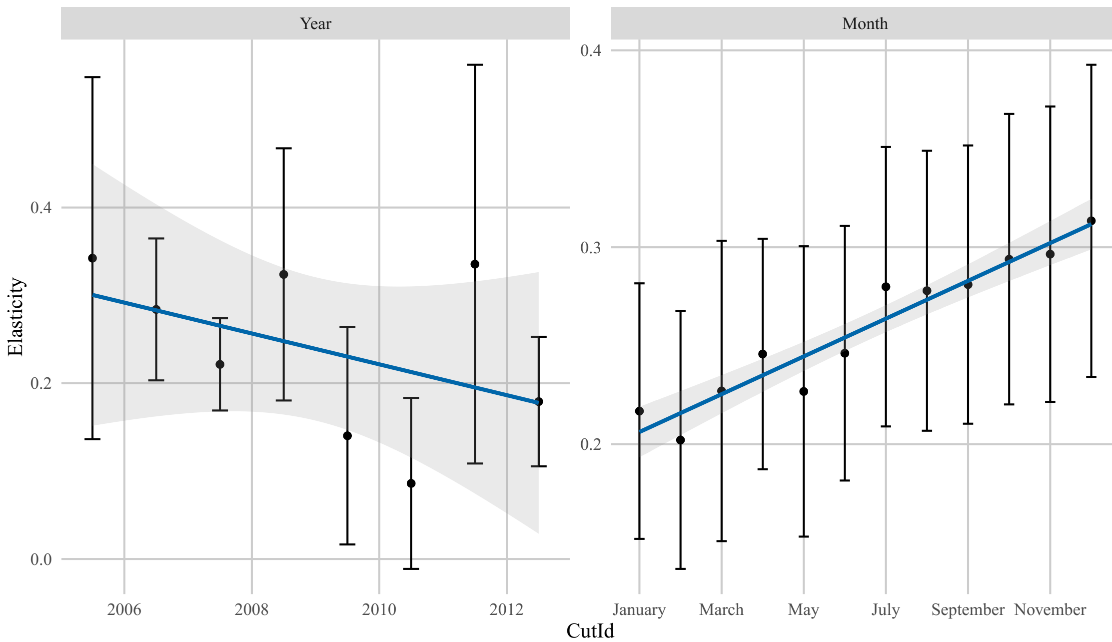

Figure 1.6: *Elasticity estimates by year and by month*

Notes: Error bars indicate 95 percent confidence intervals based on clustered standard error estimates at the base DRG level. The gray shaded areas show the 95 percent confidence interval of the least square fit lines.

than the overall estimate, as each $\hat{\varepsilon}_{t,t-1}$ is estimated off one ninth of the identifying annual price variations.

For all further heterogeneous price elasticity estimations, we disaggregate the base regression data at the levels of the corresponding indicator variables. That is, for the month regression we aggregate observations at the $(m_j, t_j, drg_{j,1}, drg_{j,0})$-level, differentiating our base aggregation by the calendar month of admission dates. Figure 1.6 shows the 12 elasticity estimates increasing from an elasticity of .21 in January to .31 in December.[26] Here, the least square fit yields a statistically significant positive slope. New DRG prices and the associated grouper logic are usually published by the regulator in September of the preceding year and effective for all admissions from January 1. Thus, we offer two interpretations for the empirical result of increasing hospital responses over time. Certain adjustments may be easy to implement right away in January including marginal treatment adjustments and ex-post up-coding, while more extensive adjustments including changes at the extensive admission margin may take more time to implement across the hospital. Alternatively, some hospitals may be faster to realize certain adjustment opportunities and only slowly share this information via informal networks. In either case, these increasing elasticity estimates highlight that our year-over-year estimates only measure a short term response. Longer term hospital responses to changes in financial incentives are likely elicit larger responses similar to findings by Clemens and Gottlieb (2014) for physicians.

Figure 1.7 shows year-over-year elasticity estimates by age group and gender. The age groups were chosen to yield 20 bins of equal admission numbers. Elasticity estimates do not differ significantly by gender. However, hospitals' price elasticity increases strongly with patient age, with point estimates of less than .15 for patients age 30 and younger and elasticities greater than .4 for the oldest age group. With the availability of different treatment options generally increasing with the age of the patient, the increasing elasticity estimates are consistent with the overall findings of this paper – hospital responses to financial incentives

[26]For all disaggregated regressions we also include group specific growth rate controls that estimate $\hat{\delta}_{g(j)}$; we drop our $\Delta \ln B_{2006}$ estimates to not introduce collinearity.

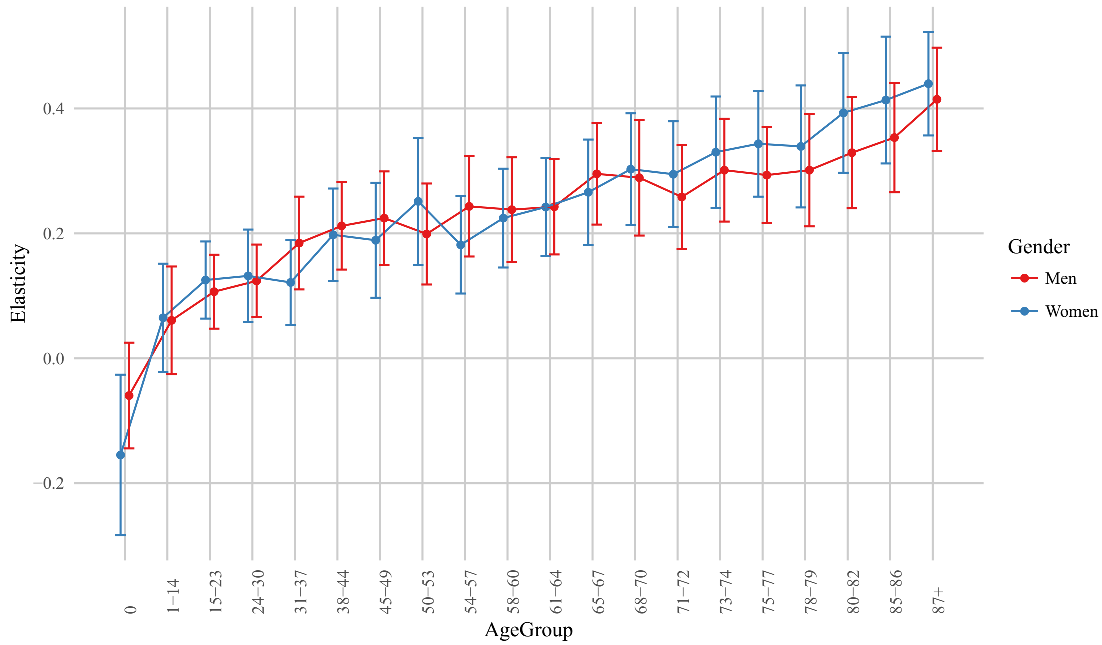

Figure 1.7: *Elasticity estimates by age group and gender*

tend to be greater in areas of greater medical discretion.

Table 1.5 reports price elasticity estimates by hospital department specialization. Specifically, we associate each admission with the hospital department the patient spends most of her or his time at, aggregating observations at the $(dept_j, t_j, drg_{j,1}, drg_{j,0})$-level. Mirroring our elasticity by age results, pediatrics yields the lowest elasticity point estimate of .017, while geriatrics is associated with the highest elasticity point estimate of .591. Of the two largest departments by admission numbers, internal medicine is more price elastic at .30 than general surgery at .19. With a price elasticity of .38, orthopedics is the second most elastic department, primarily providing elective hip and knee replacements and spinal surgery. While obstetrics is associated with the second lowest elasticity estimate of .028. From first glance, departments associated with areas of greater medical discretion are also associated with higher year-over-year price elasticity estimates.

Table 1.6 provides estimates at the more granular CCS level. CCS groups are assigned via AHRQ's grouping of patients' primary diagnosis (HCUP, 2009). We restrict reporting

Table 1.5: *Year-over-year price elasticity estimates by hospital department*

Department specialization	Coef.	S.e.	t-stat	Admissions
Pediatrics	0.017	0.041	0.409	7,539,493
Gynecology and obstetrics	0.028	0.079	0.355	15,211,740
Pediatric surgery	0.036	0.098	0.365	920,159
Ophthalmology	0.126	0.057	2.206	2,477,666
Pneumology	0.132	0.077	1.727	1,144,474
Radiotherapy	0.165	0.073	2.267	660,028
Cardiac surgery	0.169	0.098	1.719	845,785
Ear, Nose, Throat	0.176	0.039	4.458	3,807,086
Nephrology	0.180	0.060	2.985	569,027
General surgery	0.187	0.029	6.391	24,308,876
Vascular surgery	0.193	0.047	4.120	1,216,692
Plastic surgery	0.203	0.094	2.154	511,715
Intensive Care	0.203	0.061	3.328	1,146,714
Respiratory medicine	0.205	0.071	2.906	614,492
Hematology and medical oncology	0.221	0.038	5.752	1,482,952
Gastroenterology	0.227	0.029	7.967	2,111,968
Cardiology	0.249	0.046	5.406	5,768,087
Urology	0.251	0.059	4.253	5,460,158
Oral and Maxillofacial Surgery	0.256	0.070	3.639	759,680
Dermatology	0.287	0.051	5.633	1,593,279
Traumatology	0.289	0.066	4.353	5,872,460
Internal medicine	0.300	0.025	11.994	41,559,436
Obstetrics	0.362	0.140	2.590	933,358
Neurosurgery	0.368	0.093	3.961	1,596,816
Orthopedics	0.375	0.119	3.146	5,335,988
Other Department	0.476	0.091	5.211	640,594
Pseudo-Code for Readmissions	0.580	0.095	6.097	851,123
Geriatrics	0.591	0.154	3.848	1,779,707

Notes: This table reports admission price elasticity estimates with respect to DRG price changes by hospital department specialization, sorted by coefficient estimates of column 1. Column 2 reports standard errors, clustered at level of 663 'Base-DRGs.' Column 3 reports the resulting t-statistic. Column 4 reports the number of 2005-2013 admissions associated with each department specialization, underlying the regression. Reported estimates are limited to departments with at least 500,000 associated admissions and at most a standard error of 0.2.

to the 76 CCS groups with at least 500,000 associated admissions and standard errors of less than .2. Notably, elective procedure categories “cancer of prostate,” “cataract,” and “spondylosis” yield the second, fourth, and sixth most elastic point estimates. While birth related groups, “appendicitis,” and “fractures of upper limb” have some of the lowest price elasticity spots. Thus, price elasticity heterogeneity by CCS also seems to positively correlate with our subjective assessment of medical necessity.

Table 1.6: *Year-over-year price elasticity estimates by CCS code*

CCS Name	Coef.	s.e.	t-stat	Admissions
Short gestation; low birth weight; [...]	-0.237	0.138	-1.715	504,810
Anal and rectal conditions	-0.182	0.099	-1.831	685,249
Other complications of pregnancy	-0.154	0.152	-1.013	913,945
Other perinatal conditions	-0.104	0.079	-1.327	662,410
Appendicitis and other appendiceal conditions	-0.069	0.081	-0.849	1,119,936
Diverticulosis and diverticulitis	0.008	0.058	0.132	1,067,704
Intestinal infection	0.014	0.130	0.107	2,209,912
Fracture of upper limb	0.043	0.125	0.342	2,090,911
OB-related trauma to perineum and vulva	0.055	0.041	1.334	702,624
Coronary atherosclerosis and other heart disease	0.061	0.089	0.685	4,017,348
Joint disorders and dislocations; trauma-related	0.073	0.063	1.152	1,533,340
Biliary tract disease	0.077	0.039	1.948	2,337,956
Neoplasms of unspecified nature or uncertain behavior	0.078	0.068	1.148	747,094
Other eye disorders	0.087	0.061	1.441	509,817
Other and unspecified benign neoplasm	0.089	0.045	1.959	1,486,796
Non-Hodgkin‘s lymphoma	0.100	0.054	1.844	531,398
Other complications of birth; [...]	0.103	0.030	3.422	1,514,655
Superficial injury; contusion	0.123	0.063	1.932	1,177,146
Other circulatory disease	0.125	0.055	2.265	592,927
Intestinal obstruction without hernia	0.127	0.044	2.852	943,343
Polyhydramnios and other problems of amniotic cavity	0.131	0.165	0.792	544,045
Other non-epithelial cancer of skin	0.131	0.039	3.378	517,070
Cancer of head and neck	0.136	0.081	1.667	630,022
Cancer of rectum and anus	0.136	0.042	3.227	727,740
Cancer of bronchus; lung	0.151	0.046	3.266	1,658,960
Calculus of urinary tract	0.154	0.113	1.359	1,028,380
Abdominal hernia	0.158	0.083	1.894	2,184,678
Intracranial injury	0.164	0.069	2.383	2,059,181
Other gastrointestinal disorders	0.165	0.047	3.467	1,032,978
Other diseases of kidney and ureters	0.166	0.061	2.704	803,722
Varicose veins of lower extremity	0.173	0.142	1.225	757,500
Gastritis and duodenitis	0.181	0.038	4.757	980,580
Cancer of colon	0.186	0.070	2.675	809,344
Esophageal disorders	0.192	0.045	4.243	656,042

Continued on next page

Table 1.6 – continued from previous page

CCS Name	Coef.	s.e.	t-stat	Admissions
Peripheral and visceral atherosclerosis	0.196	0.063	3.109	1,664,390
Pancreatic disorders (not diabetes)	0.197	0.088	2.224	608,167
Complication of device; implant or graft	0.200	0.034	5.811	1,561,093
Alcohol-related mental disorders	0.202	0.048	4.243	1,240,720
Thyroid disorders	0.211	0.035	5.996	1,059,398
Gastrointestinal hemorrhage	0.217	0.055	3.972	819,438
Conditions associated with dizziness or vertigo	0.220	0.087	2.520	919,958
Benign neoplasm of uterus	0.226	0.095	2.364	593,968
Other nervous system disorders	0.234	0.142	1.646	2,274,335
Other fractures	0.240	0.102	2.352	1,438,232
Skin and subcutaneous tissue infections	0.242	0.055	4.412	1,538,222
Hypertension with complications [...]	0.245	0.064	3.834	537,912
Cancer of bladder	0.249	0.069	3.605	765,641
Epilepsy; convulsions	0.255	0.061	4.169	1,406,118
Urinary tract infections	0.255	0.026	9.648	1,384,172
Deficiency and other anemia	0.256	0.079	3.261	626,548
Acute and chronic tonsillitis	0.276	0.071	3.913	919,874
Fracture of lower limb	0.292	0.089	3.259	1,656,832
Syncope	0.295	0.065	4.547	1,339,057
Osteoarthritis	0.305	0.144	2.114	3,080,115
Pneumonia (not caused by [...])	0.316	0.037	8.533	2,481,927
Complications of surgical procedures or medical care	0.327	0.095	3.443	1,117,326
Congestive heart failure; nonhypertensive	0.348	0.030	11.594	3,163,431
Other connective tissue disease	0.350	0.089	3.918	1,867,563
Essential hypertension	0.352	0.058	6.051	1,746,307
Abdominal pain	0.354	0.089	3.978	1,036,161
Acute myocardial infarction	0.355	0.073	4.835	1,893,564
Septicemia (except in labor)	0.364	0.060	6.082	682,552
Secondary malignancies	0.389	0.074	5.225	978,056
Other bone disease and musculoskeletal deformities	0.396	0.197	2.013	685,586
Cardiac dysrhythmias	0.397	0.099	4.019	3,126,595
Chronic obstructive pulmonary disease [...]	0.412	0.074	5.586	1,822,075
Other upper respiratory disease	0.415	0.131	3.177	1,047,682
Crushing injury or internal injury	0.428	0.120	3.574	686,946
Fracture of neck of femur (hip)	0.442	0.071	6.245	1,235,881
Diabetes mellitus with complications	0.443	0.049	9.057	1,405,756
Spondylosis; intervertebral disc disorders; [...]	0.456	0.138	3.298	4,027,175
Allergic reactions	0.504	0.137	3.687	502,334
Cataract	0.504	0.181	2.792	665,381
Retinal detachments; defects; [...]	0.528	0.174	3.033	573,742
Cancer of prostate	0.534	0.117	4.545	653,112
Transient cerebral ischemia	0.933	0.047	19.968	930,529

Notes: This table reports admission price elasticity estimates with respect to DRG price changes by CCS code of patients' primary diagnosis, sorted by coefficient estimates of column 1. Column 2 reports standard errors, clustered at level of 663 'Base-DRGs.' Column 3 reports the resulting t-statistic. Column 4 reports the number of 2005-2013 admissions associated with each department, underlying the regression. Reported estimates are limited to departments with at least 500,000 associated admissions and at most a standard error of 0.2.

For all elasticity estimates in this section, we do not identify hospitals' different adjustment margins. Hospitals may adjust at the extensive margin by changing inpatient vs outpatient treatment for certain diagnoses. They may also change their treatment intensity, achieving higher reimbursed DRG codes via additional procedures. Alternatively, they may simply upcode patients into more profitable DRGs. Irrespective of the adjustment margin, we find that heterogenous elasticity estimates correlate positively with our subjective assessment of areas of greater numbers of elective procedures and alternative treatment options. That is, hospitals respond stronger to financial incentives in areas of greater medical discretion.

1.6 Growth Rate and Elasticity Correlations

For the heterogeneous elasticity regressions of the last section, we also estimated group specific admission growth rates $\hat{\delta}_{g(i)}$. The left column of figure 1.8 plots these growth rate estimates by age group, hospital department, and CCS code against the corresponding elasticity estimates.[27] For each of the three differentiation levels, the two sets of estimates correlate positively. Similarly, regressing the elasticity estimates on growth rate estimates yields a positive, statistically significant coefficient estimates, as reported in the first row of table 1.7. Thus, in areas where hospitals respond stronger to the financial incentive of year-over-year price variation, they also seem to be responding stronger to the general G-DRG incentive to increase admissions.

For additional robustness, we also estimate growth and elasticity coefficients for all three differentiation variables jointly. Specifically, we propose to estimate the reduced-form model

$$\begin{aligned} \mathrm{E}\left[q_{j,t}\right] &= \exp\left(\left(\varepsilon_0 + \varepsilon_{age(j)} + \varepsilon_{dept(j)} + \varepsilon_{ccs(j)}\right)\ln p_{j,t}\right) \times \\ &\quad \exp\left(\left(\delta_0 + \delta_{age(j)} + \delta_{dept(j)} + \delta_{ccs(j)}\right) t + \ln A_j + \ln B_t\right), \end{aligned}$$

[27]For figure 1.8 and subsequently reported correlation coefficients, we censor two outlier observations based on the growth and elasticity point estimates; the two observations would otherwise be the primary driver of the observed positive correlations. In particular, we drop the "Pseudo-Code for Readmissions" department and the "Transient cerebral ischemia" CCS-code with elasticity, growth rate estimates of (0.560, 0.197) and (0.933, 0.054) respectively. We also drop the gender differentiation given that elasticity does not vary significantly by that differentiation. Results are robust to omitting these modifications.

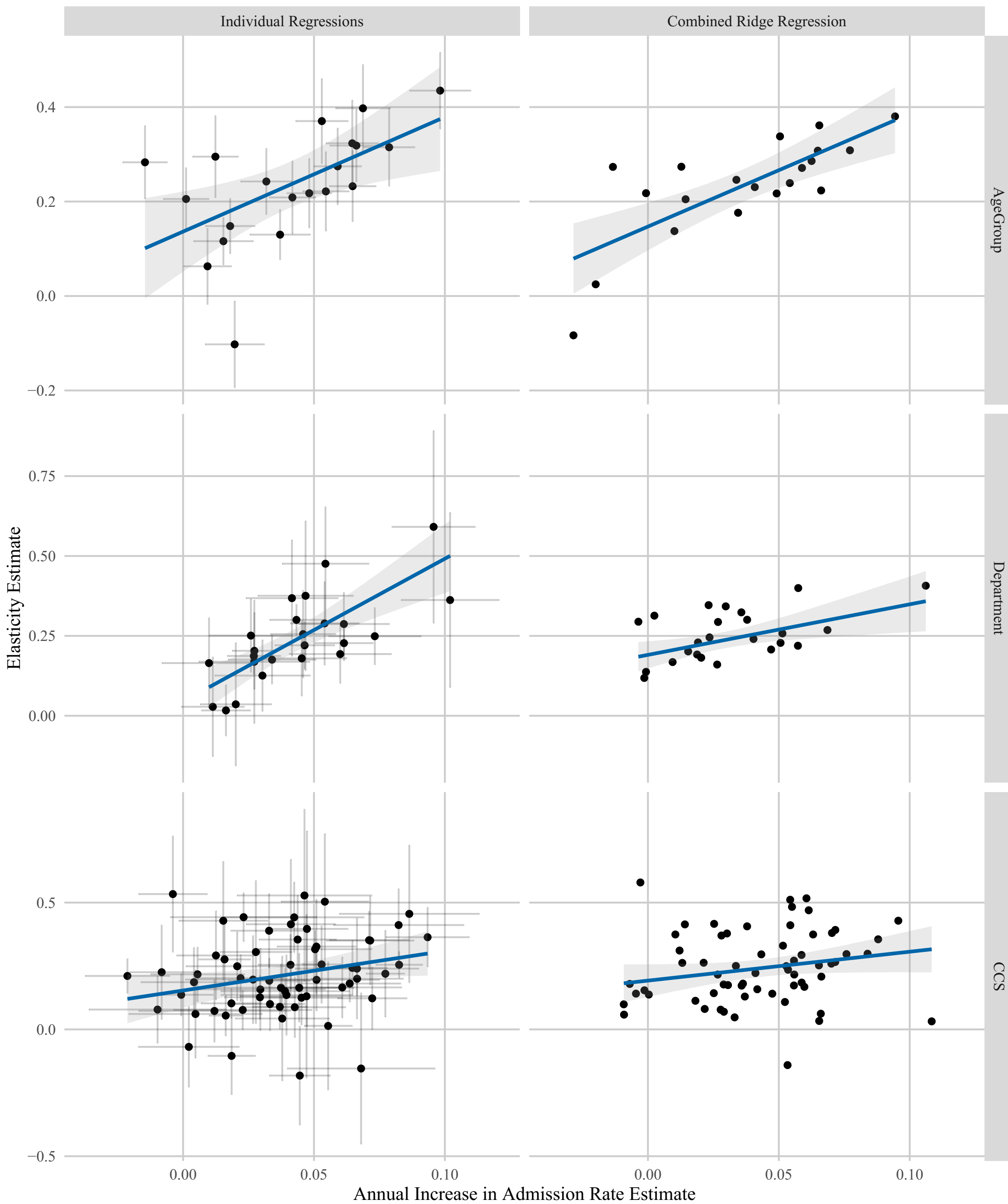

Figure 1.8: *Heterogeneous elasticity vs annual growth rate estimates*

Notes: Error bars show 95 percent confidence intervals for the coefficient estimates of the single regressions, while confidence intervals are not defined for the combined regression estimates. Each plot includes a least fit line with a 95 percent confidence interval shaded-in in grey.

Table 1.7: *Correlates of elasticity estimates*

	Separate elasticity estimate by			Joint elasticity estimate by		
	Age	Dept	CCS	Age	Dept	CCS
Growth estimate	2.426**	4.464***	1.559*	2.386***	1.582**	1.140
	(0.814)	(0.819)	(0.728)	(0.491)	(0.566)	(0.640)
Growth estimate	0.139	3.595***	1.201	1.195*	1.610**	1.356*
	(0.649)	(0.759)	(0.688)	(0.549)	(0.580)	(0.632)
Avg. length of stay	0.046***	0.015**	0.018**	0.026**	-0.002	0.011*
	(0.008)	(0.005)	(0.006)	(0.008)	(0.004)	(0.005)
Growth estimate	0.419	3.647***	1.457*	0.958*	1.921**	1.416*
	(0.541)	(0.742)	(0.696)	(0.472)	(0.598)	(0.658)
Avg. length of stay	0.038***	0.013**	0.018**	0.023**	-0.003	0.011*
	(0.007)	(0.005)	(0.006)	(0.007)	(0.004)	(0.005)
Off-hour admission share	-0.572**	-0.224	-0.207	-0.510**	-0.215	-0.045
	(0.189)	(0.158)	(0.127)	(0.182)	(0.140)	(0.125)

Notes: This table reports esimtates from regressing six sets of heterogenous elasticity estimates on three sets of variables for the 20 age groups, 27 departments, and 75 CCS groups. Standard errors are reported in parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

estimating elasticity and growth rate differences by age group, hospital department specialization, and CCS code.[28] Given the collinearity introduced when estimating elasticity and growth rate coefficients jointly for all levels, we add the ridge penalty

$$\lambda \left(\varepsilon'_{age}\varepsilon_{age} + \varepsilon'_{dept}\varepsilon_{dept} + \varepsilon'_{ccs}\varepsilon_{ccs} + 10 \left(\delta'_{age}\delta_{age} + \delta'_{dept}\delta_{dept} + \delta'_{ccs}\delta_{ccs} \right) \right)$$

to the maximum likelihood estimation. The right column of figure 1.8 shows the coefficient estimates of the joint regression for the limit as $\lambda \to 0$ with the un-penalized average effect estimates ε_0 and δ_0 added back in. For non-zero λ the ridge regression also has the Bayesian interpretation of assuming normal priors for the penalized coefficients and estimating the maximum posterior. Table 1.8 reports correlations between the estimates of the separate regressions of section (1.5) and the combined ridge regressions with $\lambda \to 0$ and deviance minimizing λ as estimated via 10-fold cross validation (Friedman *et al.*, 2010). As shown in figure 1.8 and table 1.8, the combined regressions also yield positively correlated price

[28]For this regression, admissions are aggregated at the levels of the three grouping variables, the level of DRG price variation, and by year.

Table 1.8: *Correlations of elasticity and annual growth estimates*

	Correlation of						
	$(\hat{\varepsilon}, \hat{\varepsilon}_{R0})$	$(\hat{\varepsilon}, \hat{\varepsilon}_{R1})$	$(\hat{\delta}, \hat{\delta}_{R0})$	$(\hat{\delta}, \hat{\delta}_{R1})$	$(\hat{\varepsilon}, \hat{\delta})$	$(\hat{\varepsilon}_{R0}, \hat{\delta}_{R0})$	$(\hat{\varepsilon}_{R1}, \hat{\delta}_{R1})$
Age Group	0.977	0.942	0.941	0.968	0.568	0.749	0.698
Department	0.547	0.877	0.678	0.732	0.791	0.564	0.326
CCS	0.892	0.851	0.854	0.887	0.360	0.328	0.487

Notes: Correlations of coefficient estimates of price elasticit and annual growth rates for three regression models: the basic poisson fixed effect regression run separately for each differentiating variable with $\hat{\varepsilon}, \hat{\delta}$, the combined ridge regression with penalty factor $\lambda \rightarrow 0$ with $\hat{\varepsilon}_{R0}, \hat{\delta}_{R0}$ and cross-validation deviance minimizing λ with $\hat{\varepsilon}_{R1}, \hat{\delta}_{R1}$. Correlations are weighted by number of admission by differntiating variable.

elasticity and admission growth rate estimates with all correlations exceeding .3.

Exploring alternative explanations for the observed elasticity differences, table 1.7 also reports estimates of two additional regression specifications, iteratively adding the average length of stay and a medical necessity proxy. The average length of stay variable yields a statistically significant and positive coefficient in all but one case; similar to findings of section 1.4, hospitals tend to respond stronger to year-over-year price changes for admissions with higher length of stay. One explanation for this result could be that longer treatment episodes may offer hospitals more opportunities to adjust treatments to achieve a more lucrative DRG reimbursement. Further, the coefficient estimate on the share of patients admitted outside of regular business hours is negative for all three elasticity outcomes; hospitals seem to respond stronger to prices in areas with regular business hour admissions. In any case, for all regression specifications the association of elasticity estimates and growth rate estimates is positive and in most cases significantly different from zero.

1.7 A Model of Two Financial Incentives and Medical Discretion

To offer additional interpretation of the empirical result that price elasticities and growth rates correlate positively, we propose a simple model on hospitals' optimization of whom to admit and for how long. In the model, hospitals will increase admissions in response to both a reimbursement increase and a shift from per-diem to per-admission based payments. We subsequently show that the magnitude of both responses is increasing in our medical discretion parameter.

In our model, hospitals admit up to a unit mass of patients conditional on per-patient admission utility u^{hos} exceeding the normalized outside option value of zero. u^{hos} is the sum of patient utility u^{pat} and per-patient payment p minus costs c^{hos}. u^{pat} in turn is the health benefit θh minus patient costs c^{pat}. All terms are a function of hospital chosen length of stay l as follows.

$$\begin{aligned}
u^{hos}(l,\theta) &= u^{pat}(l,\theta) + p(l) - c^{hos}(l) \\
u^{pat}(l,\theta) &= \theta h(l) - c^{pat}(l) \\
\theta &\sim F, \quad F : \mathbb{R}^{+} \to [0,1] \\
h(0) &= 0, \quad h'(0) > 0 \quad h' \geq 0, \lim_{x\to\infty} h'(x) = 0, \quad h'' < 0 \\
c^{pat}(0), c^{pat\prime}, c^{pat\prime\prime} &> 0 \\
p(l) &= p_0 + p_1 l \\
c^{hos}(0), c^{hos\prime} &> 0, \quad c^{hos\prime\prime} \leq 0
\end{aligned}$$

That is, patients only differ in their potential maximum health benefit θ, with their realized health benefit increasing in l at a decreasing rate. Patients incur costs for mere admission to the hospital that increase furhter at an increasing rate with l.[29] Hospitals are reimbursed via a linear contract in l that thus includes the possibility of both G-DRG per-admission and prior per-diem payments. We model hospital costs as increasing in l at a decreasing rate, as expensive procedures and diagnostics generally occur at the start of treatment episodes. To ensure an interior solution for hospitals' admission decision, we also assume

$$c^{hos\prime\prime} \geq -c^{pat\prime\prime}, \qquad \forall l,\ p(l) < c^{pat}(l) + c^{hos}(l), \tag{1.4}$$

that is, marginal costs to the patient grow at a greater rate than decreases in marginal hospital

[29]We may motivate these assumptions as a patient's daily costs of not following their regular routine, as well as nocebo effects of admission and extensive length of stay.

costs, and admitting patients that realize no medical benefit yields negative utility for any choice of l.

For each admitted patient, hospitals choose utility maximizing length of stay l_θ, which is defined by the first order condition

$$0 \;=\; \theta h'(l_\theta) + p_1 - c^{pat\prime}(l_\theta) - c^{hos\prime}(l_\theta), \tag{1.5}$$

which has a unique solution as $\partial^2/\partial l^2 \left(u^{hos}(l,\theta)\right) = \theta h''(l) - c^{pat\prime\prime}(l) - c^{hos\prime\prime}(l)$ is negative for all l. Thus, by the implicit function theorem

$$\frac{\partial l_\theta}{\partial \theta} \;=\; \frac{h'(l_\theta)}{\left(c^{pat\prime\prime} + c^{hos\prime\prime} - \theta h''\right)(l_\theta)} \;\geq\; 0. \tag{1.6}$$

where the inequality is strict for all $\theta > \left(c^{pat\prime}(0) + c^{hos\prime}(0)\right)/h'(0)$. Further, using the envelope theorem and assumption (1.4), we can show that

$$u^{hos}(l_0, 0) < 0 \qquad \frac{\partial}{\partial\theta} u^{hos}(l_\theta,\theta) \;=\; h(l_\theta) \;>\; 0,$$

Thus, there exists a unique cutoff value $\underline{\theta}$ that makes hospitals indifferent between admitting and not admitting a patient, with

$$\underline{\theta} \;=\; h\left(l_{\underline{\theta}}\right)^{-1}\left(c^{pat} + c^{hos} - p\right)\left(l_{\underline{\theta}}\right)$$

determining the share of admitted patients $q = 1 - F(\underline{\theta})$.

To calculate hospitals' price elasticity with respect to changes in reimbursement schedule $p\left(l_{\underline{\theta}}\right)$, we rely on the implicit function theorem and the envelope theorem to get

$$\begin{aligned} \frac{\partial\underline{\theta}}{\partial p\left(l_{\underline{\theta}}\right)} \;&=\; -h\left(l_{\underline{\theta}}\right)^{-1} + \frac{\partial\underline{\theta}}{\partial l_{\underline{\theta}}}\frac{\partial l_\theta}{\partial p\left(l_{\underline{\theta}}\right)} \;=\; -h\left(l_{\underline{\theta}}\right)^{-1} \;<\; 0 \\ \therefore \qquad \frac{\partial \ln q}{\partial \ln p} \;&=\; \frac{f(\underline{\theta})}{1 - F(\underline{\theta})} \times \frac{p\left(l_{\underline{\theta}}\right)}{h\left(l_{\underline{\theta}}\right)} \;>\; 0 \end{aligned} \tag{1.7}$$

That is, hospitals' response to price changes is positive and proportional to the inverse Mills ratio at admission cutoff value $\underline{\theta}$.

To derive comparative statics for the switch from per-diem to per-admission based reimbursements, we define a price setting mechanism that approximates the "yard-stick competition" goal of the G-DRG regulator.[30] That is, based on the regulator's expectation function E^R, we define

$$\begin{aligned}
p_0 &= \alpha C \\
p_1 &= \frac{(1-\alpha)\, C}{E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]} \\
C &= \left(E^R_{\theta \geq \underline{\theta}}\left[c^{hos}\left(l_\theta\right)\right] + k\left(1 - F\left(\underline{\theta}\right)\right)^{-1}\right) \;=\; E^R_{\theta \geq \underline{\theta}}\left[p\left(l_\theta\right)\right]
\end{aligned}$$

where k denotes hospital fixed costs that are independent of admission decisions. That is, the regulator intends to reimburse average cost. Increasing from zero to one, primitive α maps out the continuous transition from per-diem to per-admission based payments. To map out the effect of the transition on admission cutoff value $\underline{\theta}$, we derive

$$\begin{aligned}
\frac{\partial \underline{\theta}}{\partial \alpha} &= \frac{\partial \underline{\theta}}{\partial p_0}\frac{\partial p_0}{\partial \alpha} + \frac{\partial \underline{\theta}}{\partial p_1}\frac{\partial p_1}{\partial \alpha} \\
&= -h\left(l_{\underline{\theta}}\right)^{-1}\left(1 - \frac{l_{\underline{\theta}}}{E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]}\right) C && (1.8) \\
&\quad + h\left(l_{\underline{\theta}}\right)^{-1}\left(\frac{(1-\alpha)\, l_{\underline{\theta}}}{E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]}\right)\left(\frac{\frac{\partial}{\partial \alpha} E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]}{E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]}\right) C && (1.9) \\
&\quad - h\left(l_{\underline{\theta}}\right)^{-1}\left(\alpha + \frac{(1-\alpha)\, l_{\underline{\theta}}}{E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]}\right)\frac{\partial C}{\partial \alpha}. && (1.10)
\end{aligned}$$

Thus, three terms determine the financial incentives for the marginal admission.

Term (1.8) denotes the difference in effective compensation for the marginal admission compared to the average admission. To derive the sign of this term, we note inequality (1.6), which is guaranteed to be strict as $\alpha \to 1$ by equation (1.5), and subsequently we have $l_{\underline{\theta}} < E_{\theta \geq \underline{\theta}}\left[l_\theta\right]$. Thus, if the regulator does not underestimate the average length of stay substantially, we have that $l_{\underline{\theta}} < E^R_{\theta \geq \underline{\theta}}\left[l_\theta\right]$, yielding a negative sign for term (1.8). That is, as we transition from per-diem to per-admission reimbursement, reimbursement for the marginal admission increases, incentivizing hospitals to admit more patients.

Terms (1.9) and (1.10) are both proportional to changes in the regulator's expectation

[30] Germany's switch from cost-plus to DRG reimbursements also introduced effective yard-stick competition. For analytic tractability, we only model the elimination of "per-diem" marginal incentives here.

with respect to α. Thus, if the regulator's expectations are independent of α – e.g., $E^R_{\theta\geq\underline{\theta}}[\cdot]$ is only based on past observations – both terms will be zero. Alternatively, the regulator may have correct expectations with $E^R_{\theta\geq\underline{\theta}}[\cdot] = \mathrm{E}_{\theta\geq\underline{\theta}}[\cdot]$.[31] As

$$\frac{\partial l_\theta}{\partial\alpha} = \frac{\partial p_1}{\partial\alpha}\left(c^{pat\prime\prime} + c^{hos\prime\prime} - \theta h''\right)(l_\theta)^{-1} < 0,$$

we then have

$$\begin{aligned}
\frac{\partial}{\partial\alpha}E^R_{\theta\geq\underline{\theta}}[l_\theta] &= \underbrace{E^R_{\theta\geq\underline{\theta}}\left[\frac{\partial l_\theta}{\partial\alpha}\right]}_{<0} + \underbrace{\frac{f(\underline{\theta})}{(1-F(\underline{\theta}))}\left(E^R_{\theta\geq\underline{\theta}}[l_\theta] - l_{\underline{\theta}}\right)}_{>0}\frac{\partial\underline{\theta}}{\partial\alpha} \\
\frac{\partial}{\partial\alpha}E^R_{\theta\geq\underline{\theta}}\left[c^{hos}(l_\theta)\right] &= \underbrace{E^R_{\theta\geq\underline{\theta}}\left[c^{hos\prime}(l_\theta)\frac{\partial l_\theta}{\partial\alpha}\right]}_{<0} + \underbrace{\frac{f(\underline{\theta})}{(1-F(\underline{\theta}))}\left(E^R_{\theta\geq\underline{\theta}}\left[c^{hos}(l_\theta)\right] - c^{hos}(l_{\underline{\theta}})\right)}_{>0}\frac{\partial\underline{\theta}}{\partial\alpha}.
\end{aligned}$$

Thus, if $\partial\underline{\theta}/\partial\alpha < 0$, term (1.9) amplifies the effect of term (1.8); that is, in anticipation of hospitals' negative length of stay response to increases in α, the regulator lowers per-diem compensation rate p_1 more gradually, which in turn increases reimbursement for the marginal admission further. Term (1.10) on the other hand, has the opposite sign; if the regulator correctly anticipates hospitals' per-admission cost savings due to reductions in l_θ, he will lower overall compensation. Subsequently, we cannot sign $\partial\underline{\theta}/\partial\alpha$ conclusively without further assumptions. We nonetheless surmise that the effect of (1.10) may be smaller than the effects of the former two terms as the expected cost savings are likely to be low in relative terms, with $c^{hos\prime}(l) < c^{hos}(l)/l$ and hospital fixed costs k increasing per-admission compensation further.

---

[31] To guarantee a unique solution to the price setting mechanism in this case, we also need that $\partial\left(C/\mathrm{E}_{\theta\geq\underline{\theta}}[l_\theta]\right)/\partial p_1 < 1$. That is, as hospitals increase l_θ in response the higher per-diem payments, average treatment costs per day grow at a slower rate. This assumption may seem intuitive, given our assumption of marginal costs being lower than average daily costs. Yet, our weak concavity assumption on c^{hos} does not seem sufficient to yield this result. However, if we assume that for all $\theta 1, \theta 2$

$$l_{\theta 2}c^{hos\prime}(l_{\theta 1})\frac{\partial l_{\theta 1}}{\partial p_1} + l_{\theta 1}c^{hos\prime}(l_{\theta 2})\frac{\partial l_{\theta 2}}{\partial p_1} < \frac{\partial l_{\theta 1}}{\partial p_1}c^{hos}(l_{\theta 2}) + \frac{\partial l_{\theta 2}}{\partial p_1}c^{hos}(l_{\theta 1})$$

then we can even show that $\partial\left(C/\mathrm{E}_{\theta\geq\underline{\theta}}[l_\theta]\right)/\partial p_1 < 0$. This extra result follows trivially if $c^{hos\prime\prime} = 0$ from our assumptions on c^{hos}, as $c^{hos\prime}(l) < c^{hos}(l)/l$.

The empirical results of section 1.3 also suggest that the effect of term (1.8) dominates the effects of terms (1.9) and (1.10). While length of stay has been decreasing in our data, the annual declines predate the introduction of the G-DRG system. The number of admissions on the other hand has been growing at a greater rate since 2005. Without increasingly decreasing length of stay, hospitals achieved the additional growth, halting the previous decline in bed capacity (figure 1.2). Linking these findings back to our model, the effects due to changes in length of stay as captured by terms (1.9) and (1.10) are likely to have been small, while the positive extensive margin response as captured by term (1.8) seems to have been much larger.

Notably, the growth in admissions response has been gradual with high admission growth rates sustained throughout the period we have data for. To link this second finding to our static model, we note that hospitals cannot increase their bed capacity without additional approval by the state.[32] Thus, one cause for the lagged response may be hospitals's short run capacity constraints as length of stay adjustments take time – two effects outside of our model.

To use our model to analyze the impact of medical discretion, we note that

$$\frac{\partial \ln q}{\partial \alpha} = \frac{f\left(\underline{\theta}\right)}{1 - F\left(\underline{\theta}\right)} \times \frac{\partial \underline{\theta}}{\partial \alpha}$$

is also proportional to the inverse Mills ratio, just as the price elasticity derived in equation (1.7). That is, hospitals respond stronger to both incentives if the number of marginal admissions $f\left(\underline{\theta}\right)$ relative to the total number of admissions $\left(1 - F\left(\underline{\theta}\right)\right)$ is greater. Intuitively, we may label an increase in this ratio as an increase in medical discretion, with more patients near the cutoff value $\underline{\theta}$ at which hospitals are indifferent between admitting and not admitting patients. We can state this notion mathematically, by parameterizing the distribution of θ. Assuming F to be the exponential distribution with $F(x) = 1 - \exp(-\lambda x)$, we interpret inverse mean and shape parameter λ as our measure of medical discretion. In this case, as λ

[32] Germany's states restrict the number of beds hospitals may have as mandated by §6 of the "hospital financing law" (`http://www.gesetze-im-internet.de/khg/index.html`).

increases, the average medical benefit to all patients falls while

$$\frac{\partial}{\partial\lambda}\left(\frac{f(\underline{\theta})}{1-F(\underline{\theta})}\right) = 1$$

and the relative number of marginal admissions increases with our medical discretion parameter.

Modeling medical discretion via changes in the distribution F also affects our derivation of $\partial\underline{\theta}/\partial\alpha$. Ceteris paribus, as the density of marginal admissions increases, $E^R_{\theta\geq\underline{\theta}}[l_\theta]$ will fall, which in turn reduces the magnitude term (1.8). Yet, with $\partial^2 l_\theta/\partial\theta^2 \leq 0$ and assuming that the effects on the regulator's expectations C are negligible,[33] we can show that

$$\begin{aligned}\frac{\partial}{\partial\lambda}\left(\ln\frac{\partial \ln q}{\partial\alpha}\right) &\approx \frac{\partial}{\partial\lambda}\ln\left(\frac{f(\underline{\theta})}{1-F(\underline{\theta})}\right)+\frac{\partial}{\partial\lambda}\ln\left(1-\frac{l_{\underline{\theta}}}{E^R_{\theta\geq\underline{\theta}}[l_\theta]}\right)\\ &= \frac{1}{\lambda}+\frac{l_{\underline{\theta}}\frac{\partial}{\partial\lambda}E^R_{\theta\geq\underline{\theta}}[l_\theta]}{E^R_{\theta\geq\underline{\theta}}[l_\theta]\left(E^R_{\theta\geq\underline{\theta}}[l_\theta]-l_{\underline{\theta}}\right)}\\ &\leq \frac{1}{\lambda}-\frac{l_{\underline{\theta}}}{\lambda E^R_{\theta\geq\underline{\theta}}[l_\theta]} < 0.\end{aligned}$$

Thus, the relative effect of increasing medical discretion on term (1.8) is smaller in magnitude than the resulting increase in the inverse Mills ratio. Subsequently, hospitals respond stronger to the admission growth incentives of the G-DRG system for higher levels of medical discretion.

This final medical discretion result relies on our exponential distribution assumption for F. For alternative distributions, the off-setting effects of decreases in term (1.8) may potentially be larger. Further, our model also does not explain the annual increases in admission prior to the G-DRG introduction. Albeit at lower average rate, prior admission growth rates may have

[33] For $\partial^2 l_\theta/\partial\theta^2$ to be less or equal to zero, we need

$$\underbrace{h''(l_\theta)}_{<0}\underbrace{\left(c^{pat\prime\prime}+c^{hos\prime\prime}-\theta h''\right)(l_\theta)}_{>0}-\underbrace{h'(l_\theta)}_{>0}\left(c^{pat\prime\prime\prime}+c^{hos\prime\prime\prime}-\theta h'''\right)(l_\theta) \leq 0,$$

thus requiring additional assumptions on the third derivative of the cost and health benefit functions. Further, the effect of λ on expected average costs may not be negligible if both fixed costs k are small and $c^{hos\prime}$ is not negligible either. In this case, a change in λ has two additional, off-setting effects: (i) an increase in medical discretion will yield lower average treatment levels for a given θ, lowering C and subsequently $p(l_\theta)$, which in turn has a decreasing effect on the magnitudes of results (1.7) and (1.8). However, decreases in $p(l_\theta)$ yield an increase in cutoff value $\underline{\theta}$ which once more increases the inverse Mills ration, thereby increasing the magnitude of the two effects.

already been growing at different rates by level of medical discretion. Nonetheless, our model illustrates a distinct incentive for hospitals to increase admission numbers under the new per-admission payment system and when per-admission payment increases, further it provides a clear link for the magnitude of these two effects via differences in medical discretion.

1.8 Conclusion

The 2005 introduction of DRG reimbursement in Germany yielded different hospital responses than Medicare's 1983 PPS introduction. Unlike Medicare's introduction, the G-DRG introduction did not seem to yield additional length of stay declines. Further, the G-DRG introduction coincided with a great increase in inpatient admission and cost growth rates. To accommodate these increases, German hospitals seem to have halted the prior trend of declines in bed capacity, with even greater admission growth increases seemingly halted by state governments' restrictions on the maximum number of hospital beds.

Concurrently, G-DRG reimbursement also resulted in year-over-year variation in relative reimbursement rates across DRGs. We use this variation to estimate an average hospital price elasticity of .259. Thus, German hospitals seem to respond to financial incentives at one-year rates similar to US physicians as measured by Clemens and Gottlieb (2014). Our price elasticity estimates differ by patient age, hospital department spcialization, and CCS code, and seem to correlate with areas of elective medical care. Further, differences in elasticity also seem to explain differences in admission growth rates across all levels of differentiation that we analyze. With the aid of a simple model, we deduce that G-DRG reimbursements seem to have increased admission growth especially in areas of great medical discretion.

The G-DRG system introduced yard-stick competition to the German hospital market, which incentivizes hospitals to lower their average costs. In part, German hospitals seem to have responded by admitting more marginal patients with marginal costs below average costs, and for whom inpatient treatment may not have been the most cost effective option. As Holmstrom and Milgrom (1991) show, a distortionary result may be unavoidable, given that medical outcomes are difficult to measure and even more problematic to contract on.

Falk and Szech (2013) provide another argument that the G-DRG's increased focus on cost efficiency may have unintended consequences.

Thus, to better evaluate the effects of G-DRG reimbursement, further research will ideally focus on linking important medical outcomes to the G-DRG data. One possible direction might be to apply our heterogeneous elasticity and growth rate estimation approach to Germany's 402 government regions; one could then correlate differences in elasticity estimates with differences and changes in mortality trends, as recorded in the official "cause of death statistics," which also differentiates mortality by region, age, and gender.[34] A different direction for further research may be to examine the market-wide forces of the G-DRG reimbursements, as the system and the available data cover all inpatient services at German hospitals.[35]

[34]This data is maintained by "Statistisches Bundesamt, Wiesbaden."

[35]See Gaynor *et al.* (2015) for a review of several important market forces in the health case sector from an industrial organization perspective.

Chapter 2

No Significant Returns to Compulsory Schooling: The UK 1947 and 1972 Law Changes Revisited

2.1 Introduction

Mincer's least square wage regression estimates (Mincer, 1958, 1974) have been followed by a large literature of instrumental variable estimates and other identification strategies, with the goal of determining how much of the least square estimate is due to the causal effect of schooling and how much of it is driven by unobservable bias. Card (1999) summarizes that the average causal marginal return does not appear to be much lower than least square estimates, while several instrumental variable estimates yield local average treatment effects exceeding least squares estimates by 20 to 40 percent.[1] Others have studied to what degree least square estimates of differences in health by years of schooling are causal (Cutler and

[1] Imbens and Rosenbaum (2005) show that weak instrumental may have yielded biased point and standard error estimates in the instrumental variable literature cited by Card (1999).

Lleras-Muney, 2010, Clark and Roayer, 2013). To the degree that the effects of education on wage and health outcomes are related and suffer from similar unobservable bias, we may be able to pool the different outcomes for a joint instrumental variable regression. In this paper, I derive an optimal pooled outcome for estimating the average causal effect relative to the least squares estimate. Pooling employment, health, and family outcomes, I apply the derived framework to estimate the effect of two changes to UK compulsory schooling law (CSL).

A number of papers have used CSL changes as an instrument for years of schooling. Acemoglu and Angrist (2001) and Oreopoulos (2006, section 3) use multiple state specific changes across the US.[2] This regression design faces three challenges: a common trend assumption across states,[3] serially correlated errors that lead to underestimated standard errors (Bertrand *et al.*, 2004), and a week first stage (Imbens and Rosenbaum, 2005). Oreopoulos (2006, section 2) and a number of other studies[4] use a fuzzy regression discontinuity (FRD) design based on two UK CSL changes that had a much stronger first stage impact; the two laws increased the school leaving age for around 40 and 25 percent of the cohorts at the cutoff, respectively (see figure 2.1).[5] Yet, as section 2.2 of this paper shows, Oreopolous' global FRD design also severely underestimates standard errors. Similarly, Calonico *et al.* (2014) show that the local linear FRD design employed by several more recent studies likely underestimates standard errors. Correcting the standard error estimates may explain the divergence of results in the literature – zero return estimates to those that exceed OLS estimates – as sampling variance and sensitivity to differing regression specifications.

---

[2]See also Oreopoulos *et al.* (2006), Oreopoulos (2007), Oreopoulos and Salvanes (2011), Lance Lochner (2004). Pischke and von Wachter (2008), Kemptner *et al.* (2011) use state specific law changes across Western Germany.

[3]The strength of the first stage derives from a problematic common time trend assumption in schooling and wage growth across states, rather than the small impacts of the law changes. Indeed, Oreopoulos (2006) includes an additional specification in column 4 of his table 4 where he allows for differing controls and time trends; in this specification his standard error estimates increase by factor 60, rendering his point estimates insignificant.

[4]See also Clark and Roayer (2013), Devereux and Hart (2010), Machin *et al.* (2011)

[5]Oreopoulos (2006) uses the 1947 CSL change in Britain and a 1957 CSL change in Northern Ireland. Limiting myself to the more detailed British data, I study the 1972 CSL change from age 15 to 16 in addition to the 1947 change.

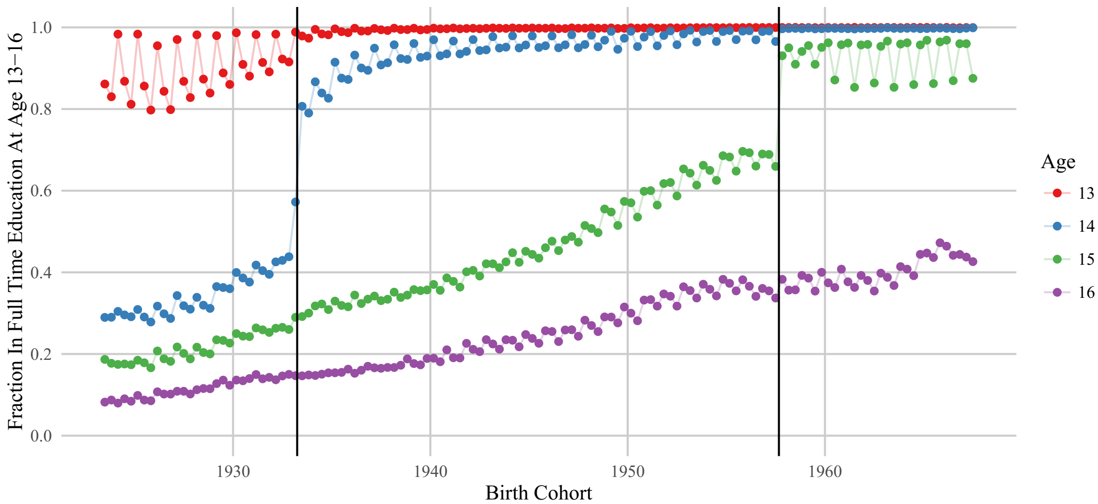

Figure 2.1: *The impact of two UK compulsory schooling increases on school leaving age*

To gain greater statistical precision, I develop an alternative returns outcome that pools a number of outcomes associated with education into a single score. That is, as a number of outcomes including employment, family planning, and health outcomes are associated with schooling attainment, a pooled outcome may be more predictive of a person's human capital than any measure individually. Section 2.3 shows that, based on a simplified linear model, the expected schooling attainment conditional on all observed adult outcomes optimally pools information; that is, this score minimizes the variance when estimating the relative size of a causal instrumental variable return estimate compared to the least squares estimate that does not control for unobserved ability. Based on the result from the simplified model, I propose a non-linear methodology to pool an average of 194 employment, family, and health outcomes for each of 50 UK labour force surveys (LFS) – described in section 2.4 – into a joint attainment score that is strongly correlated with educational attainment.

Based on this approach, section 2.5 describes the empirical implementation and the Calonico *et al.* (2014) FRD design. Section 2.6 presents returns estimates to the two UK CSL changes. While precision for any single outcome measure is insufficient to reject either estimates of zero or the OLS estimates, the 95 percent confidence interval for the pooled outcome is just $[-7.4, 23.2]$ percent of the OLS return estimate. With this confidence interval exhibiting near correct coverage in a falsification test, I conclude that the two law changes had little to no impact in aggregate on the many outcomes recorded by the UK LFS relative to what simple OLS regression would predict by not controlling for important unobservables, such as innate ability or a sustained impact of socio-economic background.

2.2 Estimates Based on Oreopolous (2006) Data

The Oreopoulos (2006) returns estimates are not robust to minor regression specification changes. After being alerted by colleagues who could not replicate his published results, Oreopolous posted a corrigendum[6] with revised estimates. The results of his preferred specification are replicated in table 2.1 row 1, predicting wage increases of 11.4 log points for one additional year of schooling, with a clustered standard error estimate of 3.4 log points, for the 1947 CSL change. Yet, running the same regression specification on the same data to predict returns to total household income instead of earnings yields a point estimate of -8.3 log points with a standard error of 5.0 log points. Similarly, replicating Oreopolous's regression design for the 1972 CSL change yields negative return estimates, and Devereux and Hart (2010) find 3 and 7 log point return, re-running Oreopolous's regression design separately for men and women. Thus, the different estimates exceed the 95 percent confidence interval Oreopoulos (2008) estimated.

To test for potential under-coverage in the confidence interval estimates, I estimate point estimates $\hat{\beta}_i$ and standard errors $\hat{\sigma}_i$ for all permutations i of a second placebo law change within the Oreopoulos (2008) regression design. Given the modeling assumption of no additional common shocks but only a smooth quartic trend across birth cohorts, the z-scores of these estimates $(\hat{\beta}_i/\hat{\sigma}_i)$ should follow a standard normal distribution. Yet, as figure 2.2 shows, the actual distribution is much wider for all of the specifications Oreopoulos (2008) reports. Table 2.2 reports estimated standard deviations of 2.3 and 2.0 for the first stage and the reduced

[6] See Oreopoulos (2008). In this paper I will refer to the corrigendum when referencing the Oreopoulos (2008) results, and the original paper when referencing the Oreopoulos (2006) regression specification.

Table 2.1: *Replication of Oreopolous (2008) results and alternative specifications*

	Log Earnings G. Britain		Log Earnings N. Ireland		Log Earnings Combined		Log Household Income	
Oreopolous (2006) Design								
1947 / 1957 law	0.114***	(0.034)	0.181**	(0.062)	0.136***	(0.027)	-0.083	(0.050)
		[73,951]		[8,954]		[82,905]		[65,646]
1972 law	-0.154	(0.157)	-0.076	(0.117)	-0.147	(0.128)	-0.084	(0.099)
		[91,662]		[8,888]		[100,550]		[49,626]
Local Linear Regression								
3 year bandwidth	-0.054	(0.099)	-0.115	(0.243)	-0.056	(0.099)	0.006	(0.052)
		[26,219]		[4,363]		[30,582]		[21,988]
5 year bandwidth	-0.055	(0.063)	0.072	(0.155)	-0.047	(0.061)	0.048	(0.036)
		[43,262]		[7,199]		[50,462]		[35,922]
10 year bandwidth	-0.020	(0.046)	-0.016	(0.127)	-0.019	(0.044)	0.023	(0.035)
		[84,911]		[13,336]		[98,247]		[66,383]

Notes: The four columns report returns to compulsory schooling estimates for: log annual earnings in Great Britain, Northern Ireland, a pooled regression of both regions, and household income in Great Britain. The first row replicates the two-stage least-squares regression of Oreopolous (2008), instrumenting "Age left full time education" with a dummy for cohorts born after 1932 and quartic controls in birth year and age. The second row replicates the Oreopolous (2008) regression for the 1972 compulsory schooling law change for birth cohorts from 1944 to 1974. Rows three to five report estimates based on a local linear regression discontinuity design using both law changes for three different bandwidths; the minimum cross-validation minimizing mean squared error bandwidth is 5 years. Sample restrictions and data described in Oreopolous (2008) and in the text. Standard errors clustered at the region-birth-cohort level in round parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. Number of observations in square parentheses.

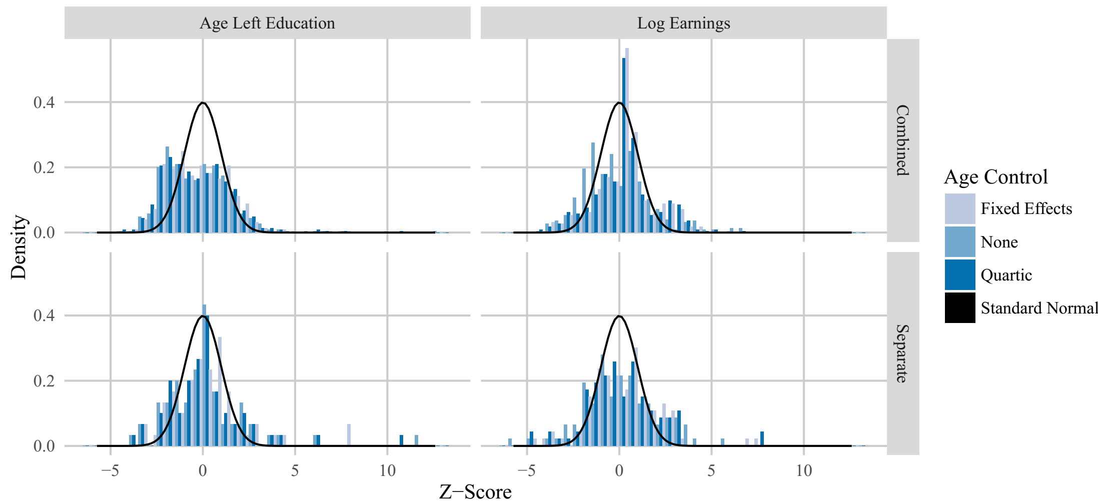

Figure 2.2: *Distributions of z-scores of estimated placebo laws in Oreopoulos (2008) FRD design*

Table 2.2: *Standard error estimates of z-Score of placebo law changes*

Outcome	Age Left Education		Log Earnings	
Region	Combined	Separate	Combined	Separate
No age controls	1.856	2.374	1.867	2.011
Quartic age control	1.845	2.281	1.683	2.064
Age fixed effects	1.819	2.101	1.593	1.916
Average	1.840	2.252	1.714	1.997

Notes: This table reports $\sqrt{\frac{1}{N}\sum\left(\hat{\beta}_i/\hat{\sigma}_i\right)^2}$ for the estimates of 449 and 60 placebo law changes added to the Oreopolous (2008) regression specifications for the combined and separate first stage and the reduced form regressions. All placebo laws are distinct from the two actual law changes and occur at least 9 years away from the data ends.

form for the region-specific specifications. By this measure, standard error estimates seem to be downward biased by around 50 percent.[7]

A number of factors may bias the Devereux and Hart (2010), Oreopoulos (2006) regression specification. Imbens and Lemieux (2008) advise against the global control regression discontinuity design, as the design imposes unnecessarily strict identification assumptions. (i) It requires the law treatment to have an additive effect to any pre-existing trend.[8] (ii) Global methods are sensitive to a correctly specified trend assumption in the forcing variable far away from the discontinuity. Similarly, the quartic control in the birth year variable function may not control flexibly enough for endogenous trends, creating serially correlated errors. Finally, the Oreopoulos (2006) specification estimates standard errors off of only 31 clusters without a needed standard error bias adjustment (Cameron *et al.*, 2008).

An imbalance in controls may introduce yet another source of bias in the Oreopoulos (2006) regression design. Figure 2.3 shows the number of observations by birth cohort that are included in the Oreopoulos (2008) Great Britain regression. Notably, the number of observations drops below the pre-existing trend right after the 1947 CSL change that affected the 1934 and later born cohorts. Similarly, the average survey year of observation changes non-smoothly by birth cohort. This lack of smoothness is largely due to Oreopolous pooling data across UK general household surveys from different years and an age based truncation. That is, Oreopolous restricts the GHS data to respondents aged 32 to 65 at the time of the survey. Thus, data from surveys of a given year will discontinuously start and end as a function of the year-of-birth forcing variable. And to the degree that responses or questions differed significantly across different surveys, these discontinuities will violate the necessary smoothness assumption of regression discontinuity designs.

The majority of the above identified biases can be easily addressed. (i) Rather than

[7] A complete estimate of standard error bias for the instrumental variable estimates also needs to account for bias in the covariance estimate of the first stage and reduced form, as well as the delta method that yields the final estimate. Yet, for the illustrative purposes of this section the more straight forward bias estimates for first stage and reduced form estimates may suffice.

[8] As the first column of appendix figure A.1 shows, the slope average "age left education" by birth-cohort changes distinctly after the second CSL change.

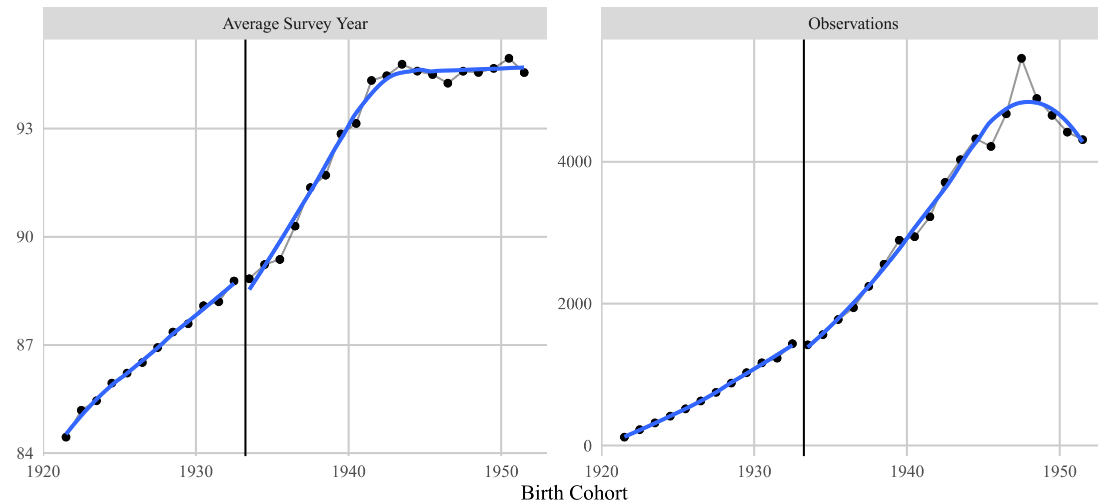

Figure 2.3: *Statistics by birth year for data used by Oreopoulos (2008)*

discontinuously truncating each survey's data by age, I decrease the importance weighting of 'boundary' observations smoothly to zero.[9] Appendix figure A.1 shows that average survey year and number of observations vary more smoothly by birth cohort for the re-weighted Oreopoulos (2008) dataset. (ii) On the re-weighted dataset, I follow Imbens and Lemieux (2008) to estimate a local linear FRD design that does not rely on the additive effect assumption, nor information from observations far away from the discontinuity. I estimate the 2SLS returns jointly for both law changes with a CV-MSE bandwidth of 5 years;[10] I also restrict the dataset to the affected population – respondents who left full time education before age 19.[11] The

[9] When pooling data across surveys, I smoothly decrease the importance weight for boundary observations as a function of birth cohort via a twice-differentiable function f over a seven-year bandwidth. That is for x = |(dob - boundary)/7|,

$$w(x) = \begin{cases} 2x^2 & \text{for } x < .5 \\ 1 - 2(x-1)^2 & \text{for } .5 \leq x < 1 \\ 1 & \text{for } 1 \leq x \end{cases}$$

[10] To estimate the optimal bandwidth for the local linear regression, I estimate the 'out-of-sample' mean squared error (MSE) of local linear regressions that predicts average school leaving age and log earnings for observations near the CSL cut-offs. That is, for each bandwidth b and each birth year y within 4 years to the left and right of the cut offs, I predict average school leaving age and log earnings based on a prediction model that was estimated on birth-cohorts $[y-b, y)$ and $(y, y+b]$, respectively. Following Imbens and Lemieux (2008), I select the smaller bandwidth of the two MSE minimizing bandwidths.

[11] I also lowered the treatment indicator for the birth-cohort year that was only partially affected to the

second part of table 2.1 reports the resulting estimates. Notably, estimated standard errors for the 5-year bandwidth design are already more than twice as large as the Oreopoulos (2008) estimates at around 6 log points. These standard error estimates are still likely to be downward biased, due to a low number of clusters and a too wide bandwidth (Calonico *et al.*, 2014). Thus, this corrected design does not yield sufficiently precise estimates to distinguish between the differing return estimates of Devereux and Hart (2010) and Oreopoulos (2006).

2.3 An Alternative Model that Utilizes More Data

To estimate returns to CSL increases, we can look at a larger number of adult outcomes that are positively correlated with educational attainment. Different studies have tried to estimate the effect of education on health outcomes, family planning, and crime using compulsory schooling age changes as an instrument.[12] However, as none of these outcomes is much more correlated with educational attainment than log wages, a robust FRD design on a dataset similar to that of Oreopoulos (2008) is unlikely to yield smaller standard error estimates relative to the magnitude of the OLS return. Even for the larger dataset constructed for this paper, robust FRD estimates to all tested individual outcomes yield 95 confidence intervals that include both zero and the OLS estimate (see table 2.5). Yet, testing for a pooled effect can yield more statistical power. This section derives a pooled outcome that maximizes the statistical power of the IV estimate relative to OLS estimate for a simple linear model. Further, the optimal pooled outcome can be generalized to utilize observations from different datasets with differing sets of adult outcomes. Based on this result, I suggest a more general approach that pools all relevant observed outcomes via a k-fold gradient boosting mechanism across multiple surveys.

In the simple linear model, individual i has human capital $\tilde{y}_i$, which increases with her innate ability a_i and her schooling attainment s_i, Her schooling in turn is also function of a_i

share of the cohort that was affected.

[12] Health: Albouy and Lequien (2009), Clark and Roayer (2013), Kemptner *et al.* (2011), Teenage childbearing: Black *et al.* (2008), Crime: Machin *et al.* (2011)

and CSL z_i. While both $\tilde{y}_i$ and a_i are latent, the econometrician observes M labor market outcomes y_{mi} that are a function of $\tilde{y}_i$. The following set of three equations describes the linear dependencies,

$$\begin{aligned}
s_i &= \alpha_0 + \alpha_1 a_i + \alpha_2 z_i + \varepsilon_{1i} \\
\tilde{y}_i &= \beta_0 + \beta_1 a_i + \beta_2 s_i + \varepsilon_{2i} \\
\begin{bmatrix} y_{1i} \\ \vdots \\ y_{Mi} \end{bmatrix} &= \begin{bmatrix} \gamma_{01} \\ \vdots \\ \gamma_{0M} \end{bmatrix} + \begin{bmatrix} \gamma_{11} \\ \vdots \\ \gamma_{1M} \end{bmatrix} \tilde{y}_i + \begin{bmatrix} \xi_{1i} \\ \vdots \\ \xi_{Mi} \end{bmatrix}
\end{aligned}$$

where all error terms are iid and independent with respect to the respective RHS variables, and $a_i \perp z_i$. We also write the last equation in vector form as $y_i = \gamma_0 + \gamma_1 \tilde{y}_i + \xi_i$, and denote $\mathrm{Var}(\xi_i) = \Sigma$. Finally, δ denotes the coefficients of the short regression

$$\begin{aligned}
\mathrm{E}[\tilde{y}_i \mid s_i] &= \delta_0 + \delta_2 s_i \\
\therefore \quad \delta_2 &= \frac{\mathrm{Cov}(\tilde{y}_i, s_i)}{\mathrm{Var}(s_i)}
\end{aligned}$$

The goal of the econometrician is to estimate the causal effect of schooling on human capital relative to the short regression estimate (β_2/δ_2).

If $\tilde{y}_i$ was observed, one could estimate (β_2/δ_2) via 2SLS and OLS, to get the sample equivalents of

$$\frac{\mathrm{Cov}(\tilde{y}_i, z_i) / \mathrm{Cov}(s_i, z_i)}{\mathrm{Cov}(\tilde{y}_i, s_i) / \mathrm{Var}(s_i)} = \frac{\beta_2}{\delta_2}.$$

Alternatively, we can construct the pooled observed outcomes $y_i^* = c_0 + \theta' y_i$, for some constant c_0 and vector θ. Then, as long as $\theta'\gamma_1 \neq 0$,

$$\frac{\mathrm{Cov}(y_i^*, z_i) / \mathrm{Cov}(s_i, z_i)}{\mathrm{Cov}(y_i^*, s_i) / \mathrm{Var}(s_i)} = \frac{\mathrm{Cov}((\theta'\gamma_1)\tilde{y}_i, z_i) / \mathrm{Cov}(s_i, z_i)}{\mathrm{Cov}((\theta'\gamma_1)\tilde{y}_i, s_i) / \mathrm{Var}(s_i)} = \frac{\mathrm{Cov}(\tilde{y}_i, z_i) / \mathrm{Cov}(s_i, z_i)}{\mathrm{Cov}(\tilde{y}_i, s_i) / \mathrm{Var}(s_i)},$$

yielding an alternative consistent estimator. For sample estimators, let

$$
\begin{aligned}
\hat{\beta} &= \left(Z'S\right)^{-1} Z'Y \\
\hat{\delta} &= \left(S'S\right)^{-1} S'Y
\end{aligned}
$$

with

$$
Z = \begin{pmatrix} 1 & z_1 \\ \vdots & \vdots \\ 1 & z_N \end{pmatrix} \quad S = \begin{pmatrix} 1 & s_1 \\ \vdots & \vdots \\ 1 & s_N \end{pmatrix} \quad Y = c_0 + \begin{pmatrix} {y_1}' \\ \vdots \\ {y_N}' \end{pmatrix} \theta.
$$

Appendix A.1 shows the derivations for the covariance matrix of $\begin{pmatrix} \hat{\beta}' & \hat{\delta}' \end{pmatrix}'$ and based on the delta method that

$$
\mathrm{Var}\left(\frac{\hat{\beta}_2}{\hat{\delta}_2}\right) = c_1 + c_2 \frac{\theta' \sum \theta}{\left(\theta' \gamma_1\right)^2}
$$

where constants c_1, c_2 are functions of the model primitives, with $c_2 > 0$. Setting the FOC to zero for θ, yields

$$
\theta = \Sigma^{-1} \gamma_1 c_3
$$

for any $c_3 \neq 0$. That is, any y_i^* satisfying the above equation minimizes the variance of the desired estimate.

Appendix A.1 further shows that for some constants c_4, c_5

$$
\mathrm{E}\left[s_i \mid y_i\right] = c_4 + c_5 \Sigma^{-1} \gamma_1
$$

Thus, any affine transformation of reverse regression function $\mathrm{E}\left[s_i \mid y_i\right]$ minimizes the variance of $\left(\hat{\beta}_2 / \hat{\delta}_2\right)$, yielding the optimally pooled outcome for the linear model.

We can extend our model to also allow for optimal pooling of data across multiple surveys with different outcomes. Suppose that we observe several independent samples drawn from the same population. However, for each sample l we might observe a different vector of adult outcomes $y_{l,i}$, for which create the pooled outcome

$$\begin{aligned}
y^*_{l,i} &= \frac{\mathrm{Var}\left(s_i\right)}{\mathrm{Var}\left(\mathrm{E}\left[s_i \mid y_{l,i}\right]\right)}\mathrm{E}\left[s_i \mid y_{l,i}\right] + \left(1 - \frac{\mathrm{Var}\left(s_i\right)}{\mathrm{Var}\left(\mathrm{E}\left[s_i \mid y_{l,i}\right]\right)}\right)\mathrm{E}\left[s_i\right] \\
&\therefore \\
\mathrm{E}\left[y^*_{l,i}\right] &= \mathrm{E}\left[s_i\right] \\
\mathrm{Cov}\left(y^*_{l,i}, s_i\right) &= \mathrm{Var}\left(s_i\right)
\end{aligned}$$

Thus, for any two subsets k and l and their respective coefficients vectors θ_k and θ_l

$$\begin{aligned}
\mathrm{Cov}\left(y^*_{i,k}, s_i\right) &= \mathrm{Cov}\left(y^*_{i,l}, s_i\right) \\
\therefore \quad \mathrm{Cov}\left({\theta_k}' y_{k,i}, s_i\right) &= \mathrm{Cov}\left({\theta_l}' y_{l,i}, s_i\right) \\
\therefore \quad {\theta_k}'\gamma_{1k} &= {\theta_l}'\gamma_{1l}
\end{aligned}$$

And for a combined regression on $y^*_{i,l}$, in the limit as the number of observations $\sum_l N_l \to \infty$

$$\frac{\sum_l N_l \mathrm{Cov}\left(y^*_{l,i}, z_i\right) / \mathrm{Cov}\left(s_i, z_i\right)}{\sum_l N_l \mathrm{Cov}\left(y^*_{l,i}, s_i\right) / \mathrm{Var}\left(s_i\right)} = \frac{\sum_l \left({\theta_l}'\gamma_{1l}\right) \mathrm{Cov}\left(\tilde{y}_i, z_i\right) / \mathrm{Cov}\left(s_i, z_i\right)}{\sum_l \left({\theta_l}'\gamma_{1l}\right) \mathrm{Cov}\left(\tilde{y}_i, s_i\right) / \mathrm{Var}\left(s_i\right)} = \frac{\beta_2}{\delta_2}$$

and similarly to before, for some constants $c_6, c_{7l} > 0$

$$Var\left(\frac{\hat{\beta}_2}{\hat{\delta}_2}\right) = c_6 + \frac{\sum_l N_l c_{7l} {\theta_l}' \Sigma \theta_l}{\sum_l N_l \left({\theta_l}'\gamma_{1l}\right)^2}$$

which once again is minimized at the proposed pooled outcome for the linear model.

The actual data generating process may not be linear, nor reducible to a single human capital scalar. Nonetheless, the results from the linear model motivate a similar approach for the more complicated case. That is, if additional compulsory schooling significantly improves an individual's human capital, which in turn has a positive impact on some job market outcomes, these outcomes should be predictive of changes in compulsory schooling requirements. To test this prediction, I estimate $g\left(y_i\right) = \mathrm{E}\left[s_i \mid y_i\right]$ flexibly to also pick up on non-linear links of compulsory schooling changes and changes in job market outcomes. Just as in the simple linear model, I subsequently test whether the CSL changes had any significant impact on the estimation of $\hat{g}\left(y_i\right)$, comparing IV and OLS estimates.

In this more general approach, the interpretation of the IV estimate divided by the OLS estimates may be ill-defined. In particular, if y_i predicts s_i non-linearly, $\hat{g}(y_i)$ is not guaranteed to be a linear function of s_i, either. The least squares regression will still estimate an average linear association of $\hat{g}(y_i)$ and s_i over the whole distribution of s_i. Yet, the IV regression may only estimate a local average treatment effect, that may be specific to a narrower range of s_i. To the degree that the OLS association of $\hat{g}(y_i)$ and s_i is differs over this range, the $\hat{\beta}_2/\hat{\delta}_2$ estimator suggested above will no longer estimate the relative size of the causal CSL effect relative to the relevant OLS prediction. This caveat may nonetheless be overcome, by re-weighting the observations of the OLS regression to estimate the effect for the same distribution of s_i as affected by the IV regression.[13] Subsequently, we can once more compare how much of the OLS estimate on the pooled outcome is causal. Alternatively, if we can graphically verify that $\hat{g}(y_i)$ approximately increases linearly with respect to s_i, this problem may not be of first order concern.

2.4 LFS Data

The 1975-2001 UK LFS data fit into the framework derived in the previous section (UK Data Service, 2004). The LFS was initially conducted biennially, starting in 1975; from 1981 to 1991 it was conducted annually; since, it has been conducted quarterly. The survey asks a great number of questions about respondents' employment, occupation, training, family, and housing situation. Importantly, the survey also asks each respondent of working age at what age they left full-time education, as well as their birth year and birth month.[14] That is, the LFS provides sufficiently granular date of birth information to be used as the forcing

[13] In the case of the two UK CSL changes studied here, the re-weighting amounts to a restriction of the data to individuals that left full time education aged 14 and 15 for the '47 CSL change, and 15 and 16 for the '72 CSL change.

[14] Starting with the 2001 surveys, the birth-month information has been removed from the standard End User License. An on-site special access license can be requested to gain access to this information for later surveys as well. Further, most of the GHS data used in Oreopolous also has secure access information that includes granular date of information, unlike in the public use data files that Oreopoulos (2008) uses. Gaining access to these additional data sources is beyond the scope of this paper.

variable in the FRD regression discontinuity design, and information on the endogenous regressor, consistently coded across surveys. While earnings information is only recorded for a small random subset of respondents, all surveys record a great number of other outcomes that are strongly associated with education, including occupation codes and socioeconomic classifications.

To employ the above regression framework, I restrict the LFS data in two ways. First, I only include responses from women and men, aged 25 to 59/64,[15] that were born in the UK, currently live in Great Britain, and are British nationals – that is, I restrict to the majority of individuals of working age that are likely to have been affected by the law. I also top code the "age left full time education" variable at 22 and drop observation with inconsistent age and date of birth information, bringing the final number of responses to 3,293,434.[16] Second, I include all employment, household/family, and health outcome variables that do not include direct information on the respondents' age, nor their training, educational attainment, or qualifications. I further exclude any variable that varies for less than 2.5 percent of observations. Beyond that, I do not alter the coding of any outcome variables, as the R gbm package implementation of gradient boosted trees allows for the inclusion of categorical variables and variables with missing data (Ridgeway, 2007). The final dataset is summarized by survey date in table 2.3. Respondent characteristics do not vary much across surveys.[17] However, the number of potentially relevant outcome variables does vary from year to year. As the above derived methodology does allow for differing outcome variables by survey, this is not of concern.

[15]More than 99 percent of respondents left full-time education before age 25, and until recently, 60 and 65 were the UK retirement age for women and men.

[16]The quarterly LFS tries to include individuals for five consecutive surveys. Clustered standard errors by birth cohort account for potentially resulting correlated errors.

[17]Survey characteristics only vary meaningfully for the 1983 survey. In this year the "age left education" was coded slightly differently and only asked of respondents aged 50 and younger.

Table 2.3: *Summary statistics for UK Labour Force Surveys*

		Average				Predictors				Correlation	
Survey	Responses	Age	Male	Edu	Work	(1)	(2)	(3)	(4)	(1)	(2)
1975	85597	39.423	0.506	15.492	0.721	44	7	11	0	0.655	0.647
1977	90543	40.192	0.507	15.315	0.754	24	19	5	0	0.610	0.571
1979	84118	40.994	0.508	15.279	0.756	36	23	13	0	0.609	0.586
1981	87019	41.207	0.506	15.307	0.735	58	28	72	0	0.645	0.607
1983	58244	36.425	0.503	16.294	0.726	63	41	36	0	0.698	0.688
1984	61586	42.008	0.520	15.910	0.701	96	59	90	4	0.696	0.659
1985	63604	42.085	0.525	15.993	0.703	105	65	56	4	0.698	0.658
1986	63596	42.027	0.522	16.049	0.720	117	65	62	5	0.703	0.666
1987	62829	42.080	0.518	16.108	0.713	118	66	64	5	0.687	0.649
1988	63779	42.106	0.520	16.146	0.734	113	65	70	5	0.691	0.655
1989	64601	41.972	0.520	16.228	0.752	116	67	68	4	0.690	0.654
1990	63902	41.969	0.518	16.329	0.760	117	67	66	3	0.688	0.657
1991	63312	42.019	0.517	16.390	0.752	127	69	60	3	0.676	0.642
1992 Q2	65607	41.944	0.515	16.466	0.737	73	50	19	6	0.670	0.647
1992 Q3	65950	41.857	0.515	16.486	0.736	72	50	19	6	0.675	0.653
1992 Q4	66890	41.861	0.516	16.504	0.734	72	50	20	6	0.681	0.657
1993 Q1	67195	41.816	0.514	16.522	0.730	87	54	20	0	0.673	0.649
1993 Q2	67138	41.844	0.514	16.522	0.735	87	53	26	0	0.676	0.652
1993 Q3	66569	41.868	0.514	16.519	0.737	89	55	19	0	0.677	0.651
1993 Q4	66340	41.866	0.515	16.536	0.738	87	51	20	0	0.676	0.651
1994 Q1	65755	41.814	0.513	16.552	0.737	91	53	20	5	0.674	0.651
1994 Q2	65154	41.891	0.513	16.564	0.740	98	106	27	2	0.675	0.651
1994 Q3	65117	41.884	0.512	16.573	0.742	98	56	21	0	0.675	0.652
1994 Q4	65034	41.953	0.511	16.590	0.743	100	58	19	0	0.673	0.650
1995 Q1	65364	41.999	0.511	16.598	0.741	97	57	22	0	0.673	0.650
1995 Q2	65398	42.093	0.512	16.596	0.745	99	56	29	0	0.675	0.651
1995 Q3	65319	42.090	0.511	16.612	0.745	99	58	29	1	0.673	0.650
1995 Q4	65748	42.151	0.511	16.615	0.749	100	58	28	0	0.676	0.654
1996 Q1	67535	42.137	0.511	16.629	0.745	115	59	34	18	0.679	0.656
1996 Q2	67204	42.165	0.510	16.645	0.747	99	60	33	9	0.681	0.658
1996 Q3	66518	42.212	0.509	16.651	0.748	98	59	34	9	0.677	0.654
1996 Q4	65977	42.194	0.509	16.657	0.754	109	61	33	9	0.673	0.651
1997 Q1	65686	42.187	0.508	16.686	0.754	122	73	37	20	0.671	0.649
1997 Q2	65026	42.201	0.508	16.716	0.757	102	68	33	11	0.676	0.654
1997 Q3	64310	42.292	0.508	16.725	0.758	102	68	32	11	0.675	0.653
1997 Q4	64019	42.375	0.510	16.736	0.760	97	51	32	13	0.674	0.653
1998 Q1	64209	42.443	0.510	16.755	0.760	99	52	32	13	0.674	0.653
1998 Q2	63893	42.449	0.510	16.777	0.763	98	64	32	15	0.671	0.650
1998 Q3	63574	42.429	0.510	16.777	0.764	97	65	32	15	0.671	0.651
1998 Q4	63523	42.479	0.510	16.797	0.769	102	62	32	14	0.671	0.650
1999 Q1	63492	42.542	0.509	16.812	0.768	105	65	32	14	0.675	0.654
1999 Q2	62870	42.577	0.509	16.823	0.771	106	65	32	14	0.672	0.652
1999 Q3	62025	42.630	0.508	16.831	0.770	105	65	32	14	0.671	0.649
1999 Q4	61688	42.688	0.509	16.845	0.774	106	64	32	13	0.673	0.652

Continued on next page

Table 2.3 – continued from previous page

Survey	Responses	Average				Predictors				Correlation	
		Age	Male	Edu	Work	(1)	(2)	(3)	(4)	(1)	(2)
2000 Q1	61385	42.671	0.509	16.861	0.772	114	59	32	13	0.667	0.647
2000 Q2	60667	42.826	0.510	16.869	0.776	105	60	32	14	0.664	0.644
2000 Q3	59776	42.925	0.510	16.888	0.776	104	59	32	14	0.661	0.639
2000 Q4	58983	43.077	0.508	16.894	0.777	106	61	32	14	0.662	0.642
2001 Q1	59875	42.981	0.506	16.938	0.775	84	49	32	14	0.596	0.568
2001 Q2	59891	42.962	0.506	16.958	0.778	140	64	35	14	0.657	0.634
Average	65868	41.978	0.512	16.487	0.749	96.0	57.2	34.6	7.1	0.671	0.646

Notes: Summary statistics for the UK Labour Force Survey data, restricted to all British nationals, born in the UK after 1920 and aged 25 to 60 and 65 men and women with data on 'age left full time education' (Edu). Number of predictor variables are by section: (1) employment, (2) unemployment, (3) household and family characteristics, (4) health. Predictors are restricted to variables not pertaining directly to age or qualifications earned, and that have more than 2.5 percent distinct values. The final two columns report the correlation of the education variable and the out of sample GBM prediction (1) unconditionally, and (2) conditional on gender and a quartic polynomial in age.

2.5 $\mathrm{E}\left[s_i \mid y_i\right]$ Estimation and FRD Specification

To approximate the optimal pooled outcome of section 2.3, I estimate $g\left(y_i\right) = \mathrm{E}\left[s_i \mid y_i\right]$ via a k-fold gradient boosted regression tree design. Boosted regression trees have great predictive properties, leveraging prediction trees' ability to detect non-linearities without overfitting as quickly (Friedman, 2001). To further avoid overfitting to individual observations, I split my final data randomly into K folds; for all observations i in fold k I estimate $\hat{g}_{-k}$ only based on observations in the other $K-1$ folds, to compute 'out-of-sample' predictions $\hat{g}_{-k}\left(y_i\right)$. Finally, since $\mathrm{Var}\left(\mathrm{E}\left[s_i \mid y_i\right]\right)$ and $\mathrm{Var}\left(s_i\right)$ do not differ much across the different LFS's, I omit estimating the final adjustment in favor of avoiding another source of in-sample error.[18]

For the final FRD regression, I use a robust local polynomial design following Calonico *et al.* (2014). That is, I estimate

$$y_i^* = \beta_2 s_i + \sum_{l=1}^{2} 1\left\{law_i = l\right\} \left(\sum_{p=1}^{P} \left(\alpha_{1lp}\left(dl_l - dob_i\right)^p 1\left\{dob_i < dl_l\right\} + \alpha_{2lp}\left(dob_i - dl_l\right)^p 1\left\{dl_l < dob_i\right\}\right) + Z_i\gamma_l \right) + \varepsilon_i$$

[18] At this point, I have not had the time to verify that the k-fold prediction approach is sufficient to not need any additional error adjustments beyond the Calonico *et al.* (2014) correction; although, the falsification test does not indicate any under-coverage problems. Re-running the final robust FRD design on the adjusted score only produces marginally lower standard errors, and near identical point estimates.

near the two policy change relevant cutoffs $dl_1 = 1933\frac{1}{4}$ and $dl_2 = 1957\frac{2}{3}$. Here y_i^* is the pooled job market outcome score, law_i indicates which policy cutoff is closer, dob_i is the birth year plus the birth month minus one half divided by 12, Z_i holds additional controls, and s_i is the age individual i left full-time education and will be instrumented with the two indicators $1\{dob_i > dl_1\}\, 1\{law_i = 1\}$ and $1\{dob_i > dl_2\}\, 1\{law_i = 2\}$. I calculate law-specific mean squared error minimizing bandwidths using the cross-validation criterion of Imbens and Lemieux (2008) for a linear specification ($P = 1$) and a triangular kernel.[19] Yet, following remark 7 of Calonico *et al.* (2014), I use a quadratic specification ($P = 2$) for the actual regression, to avoid bias due to having selected a too wide bandwidth. Following Imbens and Lemieux (2008), I cluster standard error estimates at birth cohort level. Results for the two-stage least squares, and least squares estimates for different specifications are presented in the next section.

2.6 Results

Table 2.4 summarizes the relative influence of different types of outcomes in predicting the age at which respondents left full time education via gradient boosted trees. Relative influence is computed as the in-sample variance reduction due to tree splits of a given variable (Friedman, 2001). Notably, over 50 percent of the prediction seems to be driven by an average of 4.76 variables encoding the respondent's occupation for each survey. The fact that income variables yield much less 'influence' on the boosted tree prediction is due to income information only being recorded for around 10 percent of observations. Beyond employment related variables, household structure (e.g., Marital status, and number of children) has an influence score exceeding 10 percent. Table 2.3 also reports the correlation of the survey specific k-fold out-of-sample predictions with the original schooling variable. Despite relying on differing

[19] This is equivalent to running the linear regression specification above, but re-weighting observations by

$$w_i = \max\left(0, 1 - \left|dl_{law_i} - dob_i\right| / bw\right)$$

for bandwidth bw. For the actual specification, I also multiply these weights by the age specific observation re-weighting, described above.

Table 2.4: *Relative influence of education predictors*

Variable Section/Subsection	Predictors	Influence
Employment		
Occupation	4.76	0.514
Industry	4.20	0.127
Employment status (last year)	2.62	0.039
Employment status	6.32	0.034
Hours	9.68	0.023
Socio-economic group	1.28	0.018
Current employment	2.76	0.013
Location	0.98	0.007
Income	3.66	0.005
Household		
Household structure	14.04	0.104
Housing tenure	4.70	0.030
Household details	1.14	0.001
Unemployment		
Occupation in last job	2.18	0.040
Employment status in last job	2.70	0.014
Health		
Health problems	2.48	0.006

Notes: For all GBM predictions of 'age left education', this table reports the average number of selected predictor (by variable subsection and section) and the relative influence on the overall prediction (measured as the relative reduction of insample squared error due to tree splits of the respective variables Friedman (2001)). Subsections with an influence smaller than .5 percent are ommited outcome variables, for all surveys this correlation is between 0.596 and 0.703. Finally, this correlation also remains strong after controlling for gender and a quartic trend in age. As expected, the pooled outcome – expected educational attainment given adult outcomes – is much more correlated with schooling than any individual recorded outcome.

Figure 2.4 presents the main graphical result, plotting both actual and predicted education

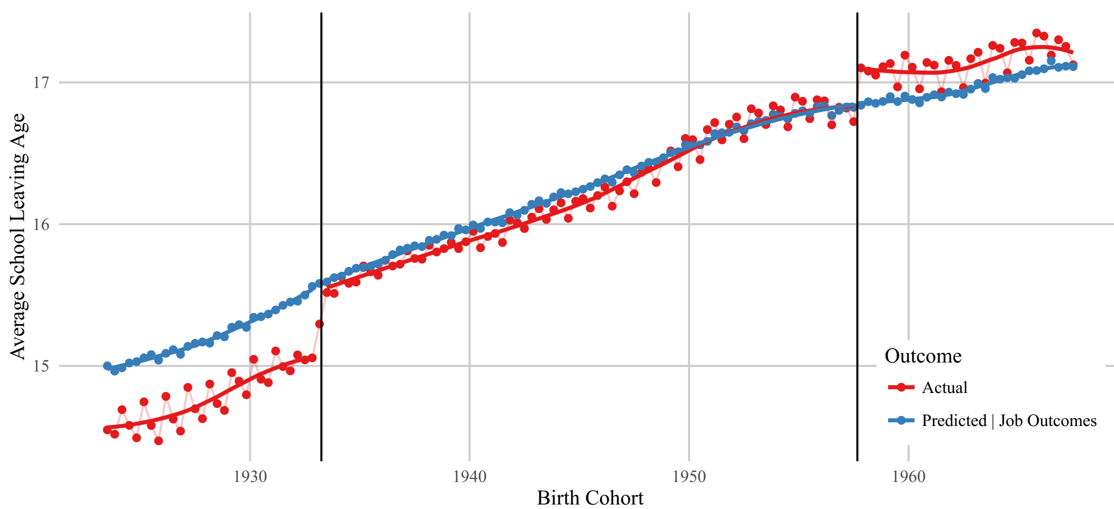

Figure 2.4: *Graphical result for effect of compulsory schooling increases on average school leaving age: actual versus the prediction, based on adult outcomes*

by birth cohort. The graph also includes fitted local polynomials for both series that are discontinued at the law change relevant date of birth cutoffs. The direct effect of the two law changes on attained education is easily discernible, leading to two strong discontinuous increase in the average observed education. Yet, no discontinuity is apparent in the expected educational attainment given adult outcomes. Appendix figure A.2 shows the average attainment by birth cohort for four alternative outcomes, here too, no clear discontinuity is apparent at the law change cutoffs. Finally, appendix figure A.3 shows that the number of observation by birth cohort and outcome, as well as the averages for three control variables, all seem to be balanced at the two cutoffs. Thus, the FRD design fulfills the requirements of a strong first stage and balance in controls; yet, the law change does not seem to have a discernible effect.

Table 2.5 presents OLS and IV schooling return estimates for a number of specifications.

Table 2.5: *OLS and regression discontinuity results for pooled outcome*

	OLS	2SLS	FS 1	FS 2	RF 1	RF 2	N
Baseline Regression							
	0.431***	0.034	0.425***	0.262***	0.022	-0.001	835491
	(0.003)	(0.034)	(0.025)	(0.025)	(0.018)	(0.015)	
Alternative Controls							
None	0.436***	0.058	0.414***	0.374***	0.032	0.019	831895
	(0.002)	(0.036)	(0.062)	(0.062)	(0.024)	(0.019)	
Birth Month, Gender	0.430***	0.035	0.426***	0.262***	0.023	-0.002	835491
	(0.003)	(0.034)	(0.025)	(0.025)	(0.018)	(0.015)	
Birth Month, Gender, Age	0.418***	0.029	0.415***	0.256***	0.015	-0.006	1083062
	(0.002)	(0.032)	(0.021)	(0.025)	(0.015)	(0.015)	
Alternative Bandwidths							
1/2 MSE bandwidth	0.429***	0.012	0.436***	0.286***	0.019	-0.023	406253
	(0.004)	(0.046)	(0.036)	(0.039)	(0.023)	(0.023)	
2x MSE bandwidth	0.432***	0.052	0.415***	0.294***	0.011	0.026	1764909
	(0.002)	(0.027)	(0.020)	(0.025)	(0.014)	(0.015)	
Left Education Age Rstriction							
By Age 18	0.410***	0.041	0.439***	0.361***	0.012	0.022*	663096
	(0.005)	(0.021)	(0.011)	(0.009)	(0.013)	(0.011)	
Aged 14/15 and 15/16	0.316***	-0.001	0.452***	0.352***	-0.024	0.027*	533454
	(0.010)	(0.024)	(0.013)	(0.016)	(0.013)	(0.013)	
Single Law Change							
‘47 law change	0.405***	0.052	0.425***		0.022		300117
	(0.004)	(0.040)	(0.025)		(0.018)		
‘72 law change	0.446***	-0.005		0.262***		-0.001	535374
	(0.002)	(0.059)		(0.025)		(0.015)	
Alternative Outcomes							
Log(Earnings)	0.120***	0.063		0.211***		0.013	70346
	(0.001)	(0.042)		(0.055)		(0.011)	
Working	0.024***	0.007	0.425***	0.262***	0.004	0.001	835491
	(0.001)	(0.011)	(0.025)	(0.025)	(0.006)	(0.005)	
Healthy	0.017***	-0.009	0.452***	0.240***	0.006	-0.013**	552997
	(0.001)	(0.016)	(0.031)	(0.031)	(0.009)	(0.004)	
Single	0.007***	-0.006	0.425***	0.262***	-0.003	-0.001	835491
	(0.001)	(0.007)	(0.025)	(0.025)	(0.003)	(0.004)	

Notes: Each row reports the estimates of a ‘returns to one additional year of schooling’ regression dicontinuity design. The baseline design estimates the impact on a pooled job market outcome score for all 1975-2001 UK Labour Force Survey respondants with ages between 25 and 64, who are white, british nationals, and born and living in Great Britain. Estimation is based on a local linear regression design around two CSL relevant brith-cohort cut-offs with a triangular kernel and 4-year bandwidth (CV-MSE minimizing for linear controls). Each regression includes both linear and quadratic controls in the forcing variable left and right of each cutoff to correct for finite sample bias (Calonico et al 2014) and controls for the birth month unless otherwise specified. The columns report the least square, IV, two first stage, and two second stage estimates with standard errors clustered at the birth-cohort level in round parentheses. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$. The final column reports the number of responses included in the regression design.

The main specification estimates that one additional year of education is associated with a pooled outcome score .43. However, the FRD only estimates a causal return of .034. To get the 95 percent confidence interval for the ratio, we can use the delta method and approximate

$$Var\left(\frac{\hat{\beta}_2}{\hat{\delta}_2}\right) \approx \frac{1}{\hat{\delta}_2^2}Var\left(\hat{\beta}_2\right) - 2\frac{\hat{\beta}_2}{\hat{\delta}_2^3}Cov\left(\hat{\beta}_2, \hat{\delta}_2\right) + \frac{\hat{\beta}_2^2}{\hat{\delta}_2^4}Var\left(\hat{\delta}_2\right) \approx \frac{1}{\hat{\delta}_2^2}Var\left(\hat{\beta}_2\right)$$

as both $Var\left(\hat{\delta}_2\right), Cov\left(\hat{\beta}_2, \hat{\delta}_2\right) \ll Var\left(\hat{\beta}_2\right)$ and $\hat{\beta}_2/\hat{\delta}_2^2 \approx .18$ is close to zero. That is, the FRD estimates a relative return of 7.9 percent with a 95 percent confidence interval of $[-7.4, 23.2]$ percent relative to the OLS estimate.

The estimated confidence intervals are bias corrected following Calonico *et al.* (2014). That is, I estimate CV-MSE minimizing bandwidths of 5.25 years and 3.5 years for the local linear design, while estimating the final design with quartic controls.[20] Further, I estimate point estimates and standard errors for the first stage and reduced form for the set of all possible placebo laws, using the average bandwidth of 4.35 years and only data that lies between the two policy changes. Given that the null hypothesis for these placebo laws should be true, the estimate z-values should be distributed normally. Appendix figure A.5 plots the actual distributions against a standard normal distribution, which seems to fit well. Further the two distributions have root mean squared z-value estimates very close to one (.94 and 1.12). Thus, unlike the regression design in section 2.2, the estimated standard errors do not seem substantially downward biased here.

To verify that the main result is not driven by a non-linear relationship of s_i and pooled outcome $\hat{g}_{-k}(y_i)$, figure 2.5 plots OLS regression for observations used in final regression. Encouragingly, the increase in the average pooled outcome from school leaving age 14 to 15 for the '47 CSL and 15 to 16 for the '72 CSL seems to follow the OLS fit over the entire range, approximately. As a further robustness check, I re-estimated the main OLS and FRD designs on two restricted subsets: individuals leaving school before the age 18, and individuals leaving school aged 14 to 15 and 15 to 16 for the two CSL changes. While the OLS point

[20]Appendix figure A.5 shows the CV criterion for each law, the first stage, the reduced form, and each of the different control specifications.

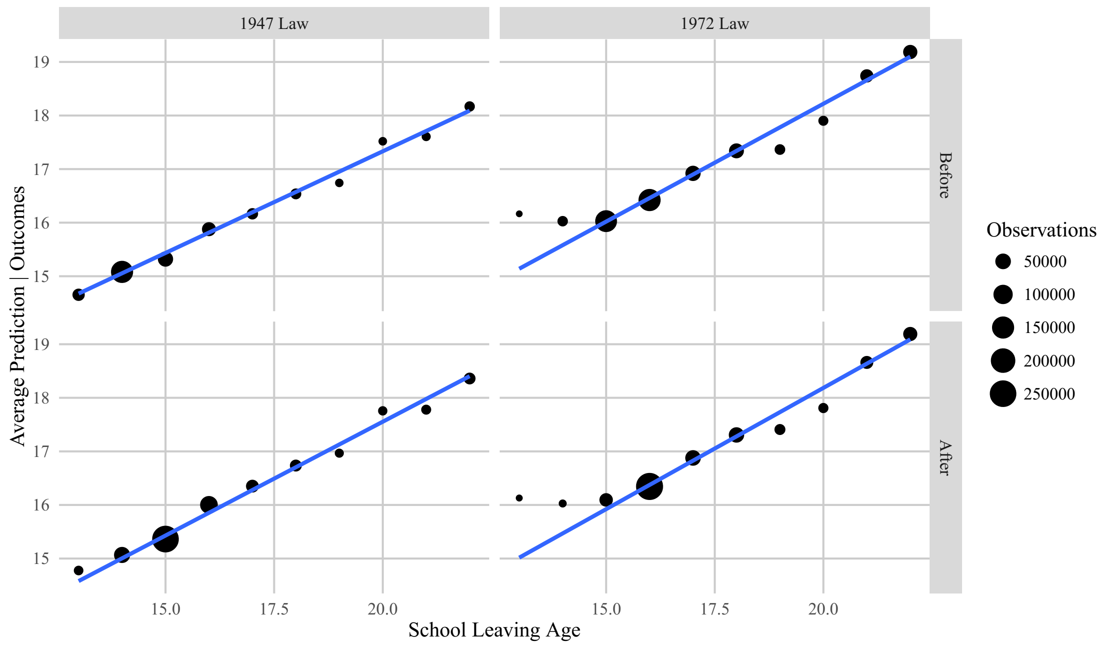

Figure 2.5: *Graphical result for OLS regression, based on weighted observations included in main regression specification*

estimate decreases slightly, the overall result does not change, with 95 percent confidence interval estimates of $[-0.1, 20.2]$ and $[-14.9, 14.5]$ percent of the OLS return, respectively.

Table 2.5 also presents four additional sets of alternative specifications. Changing the set of controls, adding a female indicator and a quartic control in age, or removing the birth month control,[21] does change the result qualitatively. Similarly, halving or doubling the bandwidth does not yield significant FRD estimates, nor does estimating returns to a single law change. I also estimate returns to four individual outcome variables for all surveys that included them. None of these estimates is significantly different from zero either, though the estimated standard errors are much larger yielding 95 percent confidence intervals that also include the OLS return. Yet, for all pooled outcome specifications none of the 95 percent confidence intervals even include half the OLS estimate.

[21]As birth month influences the reporting of the age left education variable, I include it in all other specifications to reduce variance and improve the fit of the first stage.

2.7 Conclusion

Estimating a zero return to CSL changes is no longer novel. Pischke and von Wachter (2008) estimated zero earnings returns to CSL changes in Western Germany. Clark and Roayer (2013) estimate no significant health returns to the UK CSL changes. And, Devereux and Hart (2010) found near zero returns to two UK CSL changes, while closely following the regression design of Oreopoulos (2008). Reconciling these conflicting results with the previous literature that found substantial positive estimates may seem challenging. Yet, as highlighted in section 2.2, these discrepancies may simply be due to too rigid identification assumptions and severely downward biased error estimates. To gain more statistical power, I derive a pooled outcome that aggregates a great number of adult outcomes that are associated with educational attainment, and allows for pooling data from different surveys. Constructing this pooled outcome on data from 50 LFS allows me to compute a narrow 95 percent confidence, bounding the causal returns to the two UK CSL increases to only $[-7.6\%, 23.4\%]$ of the OLS return estimate.

While the LFS data does not allow for a precise return estimate on earnings alone, the low returns estimate to the pooled labor outcome makes a distinct, more substantial effect on earnings unlikely. I can offer three defenses for nonetheless positive schooling returns. (i) The regression analysis only identifies a LATE for one additional year of schooling for individuals for whom the CSL increase is binding. (ii) The returns were attenuated due to an inability to adequately accommodate the substantial number of students staying an extra year in the short run. (iii) While the additional schooling had no measurable effect on a pooled measure of labor market outcomes it improved non-labor market outcomes, not measured by the LFS. On the other hand, we might also argue for the unlikeliness of an entire third of a British cohort myopically under-investing in education. In this light, the recent focus in the literature on quality rather than mere quantity of schooling appears to be more than appropriate.

Chapter 3

Returns to Scale of Data: Evidence from Online Retail[1]

3.1 Introduction

As firms collect growing amounts of data about the markets in which they operate, the competitiveness of any one firm can be affected by exactly what and how much data it can draw on for its decision-making. In consumer markets, for instance, a firm with a larger existing base of consumers will, all else equal, likely have more precise estimates of its decision-relevant parameters, such as those describing demand. Therefore, due to endogenous information sets, relative firm competitiveness can be a function of market structure, particularly in settings with large degrees of private information. Furthermore, when future information sets depend on current competitiveness – such as when a firm that offers higher quality at present will attract more consumers in the future – the resulting feedback loop can dynamically reinforce asymmetries between market participants.

We empirically document the magnitude of these effects – how great are precision gains from additional data, how do they translate into static competitiveness, and what is medium-run adoption – in the setting of a large online retailer. This firm sells a large number of

[1] Co-authored with Daniel Pollmann

products and needs to decide how to rank these when displaying its product catalog to consumers arriving at its website. The attractiveness of products is a set of parameters the retailer estimates based on observed consumer behavior, including how many views and orders a given product received. We have access to the entire browsing, price, and order data of the retailer for a long period of time, meaning that we see exactly that part of the firm's information set which is derived from privately observable consumer behavior. We know that in practice, the firm's decision-making relies heavily on this data, making this an ideal setting in which to study how a firm's ability to display and sell attractive products as well as attract repeat consumer interest depends on the inference on product quality it can draw from its existing customer data.

To isolate the return on additional consumer data, we ask what estimates of product quality hypothetical firms of different sizes would have arrived at, how that would have translated into different decision vectors, and what the resulting changes in static and dynamic outcomes would have been. We construct these counterfactuals by splitting the available data into training datasets of different sizes and analyzing outcomes on hold-out data when solving the firm optimization problem using estimates for the former. More specifically, we run regressions for whether consumers choose to click on and view products from a menu displayed to them, where the parameters of interest are product-level fixed effects. This yields estimates of product quality which approximate those used in practice by the retailer's algorithms. Importantly, we interpret the training data quality estimates not as the true parameter values, but rather the counterfactual beliefs of firms of different sizes, which we use to have each of them reoptimize. We then present several statistics of the economic value associated with the resulting decision vectors calculated on our hold-out dataset. Specifically, we use this data to reestimate the product click regression and obtain statistically independent estimates for the number of product views implied by the model for each of the hypothetical firms. We find that the hold-out estimates imply substantial returns to using larger fractions of the data in identifying the top products, with significantly larger average quality at the top of the first catalog page as well as for average quality of the entire page. We then present estimates showing that consumers who were displayed a higher quality set of products on their first visit – controlling for granular time trends and other factors – recorded significantly more clicks and purchases on the same day as well as greater engagement in the period after.

We take our reduced-form counterfactual calculations as evidence that in a realistic setting, there are large returns to observing additional consumer data. The main limitation of our approach is that the product rankings present in the data we use to form counterfactual beliefs are chosen by the firm and affect the rate of learning about the set of quality parameters. More specifically, the firm faces an optimization problem precisely because many consumers will only consider the subset of products that is most prominently displayed at the top of the first page of results. By implication, a firm will learn substantially faster about the quality of these products. Firms of different hypothetical sizes could optimize their ranking in light of expected sample sizes; small firms, for instance, may choose to experiment differently or specialize to reduce the number of relevant parameters. In addition, since the actual ranking of products in the observed data is based on the retailer's entire data, it is optimized to efficiently learn about the most relevant elements of the quality vector potentially more so than a smaller firm would know to. One avenue of further research is a more structural analysis of this effect of active learning.

We think that the general mechanism we consider extends beyond the technology and retail sectors to other parts of the economy, in particular consumer-facing firms. In credit and insurance markets, firms typically use predictive models for the risk of potential borrowers or insurees (Bundorf *et al.*, 2012, Einav *et al.*, 2013, 2012). Since rich models will require large sample sizes for training and validation, larger firms can make acceptance and pricing decisions that more accurately reflect underlying risk. Examples of other settings in which firms collect large amounts of consumer data include electronic heath records and the utilities and communication sectors.

Even more broadly, this paper connects to a growing literature in economics that highlights new opportunities for research using "big data" and develops appropriate statistical methods for doing so. Here, we look directly at firms as econometricians and analyze what they can learn from growing amounts of data in high-dimensional settings. We believe that the endogeneity of how much data a given firm observes creates interesting and potentially very important interactions with market structure.

Section 3.2 describes the empirical setting and available data in more detail. Section 3.3 sets up a stylized model of firm optimization and presents our corresponding statistical model which we use for both our own estimates and to model firm beliefs. We then present estimates from this model and the resulting effects on firm optimization in section 3.4 before concluding with a brief discussion in the final section.

3.2 Empirical Setting and Data

The data were obtained from an online retailer selling consumer durable products across a variety of categories. The majority of these categories, taken together, constitute one of the merchandise lines used by the U.S. Census Bureau to subcategorize "Electronic Shopping and Mail-Order Houses" (Bureau, 2015), which includes e-commerce. The retailer sources its products from manufacturers or their distributors and sells them directly to consumers, primarily through one main website, on which it is the only seller. It is among the largest online retailers in its product segment. The product mix varies by category from well-known brands to differentiated niche products, and its main competitors are online and offline retailers selling identical or substitute products.

Within a typical product category, we think of products as substitutes, with each potential customer having unit demand. However, consumers may purchase from several categories. Even before revenue weighting, the median product ordered in our sample exceeds the minimum amount beyond which shipping is free. The percentage of products which is returned by customers is in the single digits, so that we can generally assume that customers order only products they intend to purchase. The products are differentiated vertically, ranging from entry level to more upmarket, as well as horizontally, with taste heterogeneity playing a large role.

The retailer uses direct marketing, primarily via email to registered customers, but also advertises online on search engines or other websites as well as offline, for example through TV spots. Consumers arrive to the website either directly by entering its URL in their browser, from a search engine, or by clicking on a link or banner in an email or on another website.

We focus on consumers who arrive to the catalog page for a specific product category, where a category is sufficiently narrowly defined that all of its products are at least minimally substitutable. These category pages are the largest channel for orders, and they are the primary means by which consumers can search the differentiated product assortment. Consumers can arrive at these pages either through an external search engine, by navigating through from the homepage, or by entering a search term on the website, which, unless very specific, will refer them to one of the catalog pages, possibly with filters corresponding to their search query applied. The set of products displayed on these pages displayed on these pages, in general, does not condition on any consumer-specific information, so the firm problem studied here is that of selecting a default list of products intended to cater to the overall population of consumers.

In the data we use, each of these catalog pages shows a grid of 48 products, and consumers can click through to see additional pages with the same number of products from this category. Alternatively, they can filter the set of products displayed by attributes such as price as well as product characteristics which will vary across categories. For each product, consumers see a photo, the product and brand name, its price, and average reviews. After clicking on a product, a new page loads up that provides additional photos and information for the specific product as well as the option to add this product to the shopping cart and check out to purchase the product.

In our dataset, we observe all of the navigational choices made by the consumer, that is, all category and product pages viewed in their exact order, typically referred to as clickstream. Consumer visits can (sometimes imperfectly) be matched over time to a panel using identifiers based on browser cookies, IP addresses, or log-ins. In addition to product clicks, we also observe whether any of the products were ordered. Crucially, we also observe all products with their position on the catalog page, irrespective of whether they were clicked. Furthermore, we know the price of each product at each point in time.

The retailer sets its own prices, subject to minimum advertised prices set by manufacturers for some of the products. Price variation arises from cost shocks (either to wholesale costs or estimated total costs), explicit experimentation, business logic, and managerial decisions affecting the overall price level, which may translate differentially to the product level. In what follows, we will treat the resulting within-product price variation as conditionally exogenous after controlling for the product as well as granular time trends. This stands in contrast to relying on (less common) explicit sales periods; in fact, we discard observations for which the difference between current and median log product price is larger than .25.

In addition, the retailer controls the order in which products appear on the catalog pages. Variation in the sort order arises from product stock-outs, explicit experimentation (so-called A/B tests comparing discrete variations of sort orders as well as giving "exposure" to new or otherwise promising products), and an optimization algorithm that produces a fair amount of variation, which, if not explicitly stochastic, is nonetheless useful (and used in practice) to learn the relative attractiveness of products. Both stock-outs and the optimization algorithm, which is run at regular time intervals to compute a new sort order, generate variation over time, while experiments may lead to variation at any one point in time.

In addition to the above criteria, we restrict the sample to consumers browsing on computers rather than mobile devices, which is true for the majority of visits and an even larger fraction of orders. We also exclude sort orders that were explicitly personalized for a subset of repeat visitors based on which products they had clicked on in the past. This represents a small fraction of traffic, does not apply to new visitors, and is generally less useful for the purposes of our analysis (and this traffic is hence also excluded by the internal optimization algorithm responsible for the sort order displayed to the vast majority of consumers). We additionally exclude visitors flagged as bots by the retailer's internal logic. The final sample includes tens of millions of visits in under two years. In our analysis, we collect the product categories in our sample further into product groups based on an internal taxonomy and estimate the model separately for these groups.

For our analysis, we do not rely exclusively on experimental variation. In this paper, we study what an actual firm learns about consumer demand, and in practice firms often need to rely on observational data when their optimization and estimation problem is very high-dimensional, such as sorting a very large number of products by attractiveness or pricing using product-specific elasticities with a long tail of products. Explicit experimentation is more commonly used to estimate the treatment effect of discrete variations or different optimization algorithms, each of which produces one set of prices, sort orders, etc., in line with the purpose of the experiment (see Dinerstein *et al.* 2014 for an example). In fact, we specifically aim to show that even for realistic sizes of observational data seen by a large firm, the estimation and economic optimization problem can be sufficiently challenging such that there can be large returns to additional scale, and there simply is not enough traffic on which to run experiments for these to serve as the basis of estimation alone.

With this dataset in hand, we observe exactly what the firm observes about revealed consumer preferences, which is crucial given our empirical interest. It is based on the same underlying data the retailer uses to optimize its sort order and other aspects of its website and business.

3.3 Empirical Model of Product Quality

For each category, the firm needs to rank its products $j = 1, \ldots, J$ into sort positions $r = 1, \ldots, R$ for any consumer i that visits the page. We assume that the firm uses a simple separable model to approximate its static optimization problem over the expected profits from any such ranking:

$$\max_{\sigma} \mathbb{E}\left[\sum_{r=1}^{R} Y_{i,j(r;\sigma)} b_j\right] \quad = \quad \max_{\sigma} \sum_{r=1}^{R} \gamma_r \mu_{j(r;\sigma)} b_j, \tag{3.1}$$

where $Y_{i,j(r;\sigma)} \in \{0,1\}$ is the binary outcome of interest of product j listed in position r given a ranking (or permutation) σ for consumer i, and $b_j \in \mathbb{R}$ is the associated benefit. The expected value is assumed to be separable in two dimensions: additively across products, and multiplicatively between products and positions, which enter with parameters γ_r and $\mu_{j(r;\sigma)}$, respectively.[2]

When maximizing the expected number of clicks, $Y_{i,j(r;\sigma)}$ is an indicator for whether the product in question is clicked by consumer i, and $b_j = 1$ for $j = 1, \ldots, J$. Under the assumption that positions effects are decreasing,[3] $\{\gamma_r\}_{r=1}^{R}$ forms a decreasing sequence, and it is optimal for the firm to sort products in descending order of product effects $\{\mu_j\}_{j=1}^{J}$.

The central object of interest to the firm is thus the vector of product effects μ, which translates the question of this paper – how do firms of different size differ in their competitiveness due to the amount of data they observe? – into i) how well firms of different size can estimate and learn the vector μ, and ii) how this impacts their profits. We operationalize this task, derived from the stylized yet empirically relevant model above, by estimating product quality as a fixed effect in a logit regression that also includes a position as well as price effect alongside a rich set of controls:

$$Y_{i,j} \quad = \quad \mathbf{1}\left\{\delta_j - \alpha p_{i,j} + x'_{i,j}\beta + \epsilon_{i,j} \geq 0\right\}, \tag{3.2}$$

where $Y_{i,j}$ is the outcome of product j for consumer i, $p_{i,j}$ is the log price, $x_{i,j}$ is a vector including controls such as for position and time, δ_j is the fixed effect for product j, and $\epsilon_{i,j}$ is assumed to follow an iid $EV(1)$ distribution conditional on all the right-hand side variables.

Under this assumption, the estimated coefficients have a causal interpretation as certain average elasticities with respect to price and position. However, a truly structural model of the data-generating process should also account for dependence between products in both search and purchase decisions, which is ignored here. We nonetheless find the estimates on price and position (presented in section 3.4.1) useful both on their own and as a guide for the magnitude of effects that should be captured in a structural model.

The main purpose of this regression, however, is to deliver estimates of product quality

[2]Lahaie and McAfee (2011) use the same model to argue that in constructing an efficient ranking, some degree of shrinkage should optimally be applied to estimates of advertiser effects when these are uncertain. See Jeziorski and Segal (2015) and Jeziorski and Moorthy (2015) for empirical models weakening these separability assumptions in the sponsored-search context.

[3]See Ursu (2016) for evidence of large effects in a field experiment run by a travel intermediary.

that are qualitatively similar to those a firm would have arrived at. A Bayesian firm would use the data to update its prior, leading to posterior beliefs on product qualities, while a frequentist firm would also likely shrink its estimates if it cares about precision rather than unbiasedness. Put more practically, for a large number of products, a firm seems unlikely to favor a product that was successful in its only observation over one that consistently outperformed for many observations, at least in the static problem analyzed here. For this reason, we impose a ridge L2 penalty on the fixed-effect coefficients, leading to a penalized likelihood function which adds to the unpenalized likelihood based on model (3.2) the sum of squared fixed-effect coefficients weighted by a constant λ. The maximand of this expression is a maximum a posteriori estimate and equals the mode of the posterior distribution from using a Normal prior $\delta \sim \mathcal{N}(0, 2\lambda)$.[4]

In the logit model, for small probabilities,

$$\begin{aligned}
\Pr\left(Y_{i,j} = 1 \mid p_{i,j}, x_{i,j}\right) &= \frac{\exp\left(\delta_j - \alpha p_{i,j} + x'_{i,j}\beta\right)}{1 + \exp\left(\delta_j - \alpha p_{i,j} + x'_{i,j}\beta\right)} \\
&\approx \exp\left(\delta_j - \alpha p_{i,j} + x'_{i,j}\beta\right),
\end{aligned}$$

which factors into a product effect and a position effect, yielding an approximate correspondence to the relevant parameters of the stylized model (3.2) above, for which the optimal policy is simply assortative. Taking the fixed-effect vector δ to be our empirical analogue to μ, we thus assume that the firm sorts its products according to its vector of estimates $\hat{\delta}$, which, as we discuss above, need not be the maximum likelihood estimate. In addition to economic interpretability, the decision for the logit model is motivated by computational considerations, which loom large for a dataset of the size considered and will play an even larger role for any firm that needs to regularly update these estimates.

Finally, we note that for the parameters of the logit model (3.2) to be consistently estimated, we require the number observations per product to grow large to avoid incidental parameter

[4]We therefore choose λ based on the variance of fixed-effect estimates in a typical product category.

bias.[5] Here, we are specifically interested in an empirical setting in which fixed effects cannot all be estimated with arbitrary precision. We find it plausible, however, that the bias in the estimates of the "common" parameters, chiefly price and position, can vanish sufficiently fast to be of second order to our empirical question, while at the same time, the gain in precision in the fixed effect estimates that would result from sampling additional observations of the respective units would still be economically meaningful. We are unaware of theoretical or simulation results on incidental parameter bias in a setting such as the present in which different units of the panel are potentially sampled at very different rates. Bias considerations may furthermore be of less than usual importance since the estimation problem at hand is closer to a prediction problem, specifically for which products have the highest quality as perceived by consumers.

3.4 Results

We now present estimates for how firm competitiveness varies with firm size as a result of how much data a firm has available to estimate product quality. We consider competitiveness to be the true quality the firm is able to serve consumers after solving optimization problem (3.1) when using its own (noisy) estimates. We look at both short-run click outcomes, which correspond exactly to this measure of competitiveness, and long-run (orders and future visits) outcomes, which may be directly or indirectly affected as well.

Throughout this section, we will work with different non-overlapping subsamples of the original dataset, using a hold-out sample to evaluate the prediction quality a firm would have achieved from a given training sample when running the logit regression (3.2).[6] This

[5] An alternative approach (Chamberlain, 1980), based on a conditional likelihood, does not suffer from this type of bias but has at least two other deficiencies in our context: i) it does not yield straightforward estimates of the fixed-effect parameters (which are considered nuisance parameters in many panel data models), ii) the computation of the conditional likelihood becomes extremely computationally burdensome and numerically unstable (underflow problems) once the panel dimension is moderate rather than small.

[6] This relies on numbering browser cookies based on order of arrival and assigning them to groups based on this number. For instance, to get two groups of the same size, we would divide the sample into one subsample with all even arrival numbers and another with all odd. Due to the large number of new cookies on any given day, this yields approximately random, temporally stratified assignments.

ensures that the quality estimates we assume the firm to use are statistically independent of our realized competitiveness measure, which is important since true competitiveness (3.1) is unobserved. Instead, we rely on an empirical analogues obtained by re-running regression (3.2) on the hold-out data.

By working with subsamples, we are naturally limited to counterfactual experiments in which the firm is shrunk relative to its actual size. We think that these are nonetheless interesting since the firm in question is relatively large, and because the effects in question can be expected to change continuously as a function of firm size, allowing reasonable extrapolation beyond the actual firm size. A more principled though also more model-dependent alternative approach would rely on a structural model, which could also allow for active learning through experimentation on the firm side.

We start this section by providing evidence that in our setting, changes in the sample size translate into noticeable changes in the precision of a firm's quality estimates. Then, we present the core results of this paper: a firm's quality level can be substantially affected by the amount of available data, and higher quality translates into more clicks and purchases the day of, and though measured with less precision, likely in subsequent days and weeks.

3.4.1 Quality Model

The main parameters of interest in our empirical model (3.2) are the product qualities measured by fixed effects, position effects, and click price elasticities. Across product groups, the estimated click price elasticity varies from -.68 to -2.29 with a median of -1.56. While purchase elasticities tend to be a multiple of these numbers, we nonetheless take them as evidence that consumer search on the catalog pages we consider is directed and meaningfully reflects underlying preferences. Figure 3.1 shows the estimated position effects on the first catalog page as an average over all categories, with the shaded intervals around the mean representing one standard deviation of the effect over product categories in each direction. The drop from the first position to the bottom of the page is substantial, reducing click interest by around 70%.

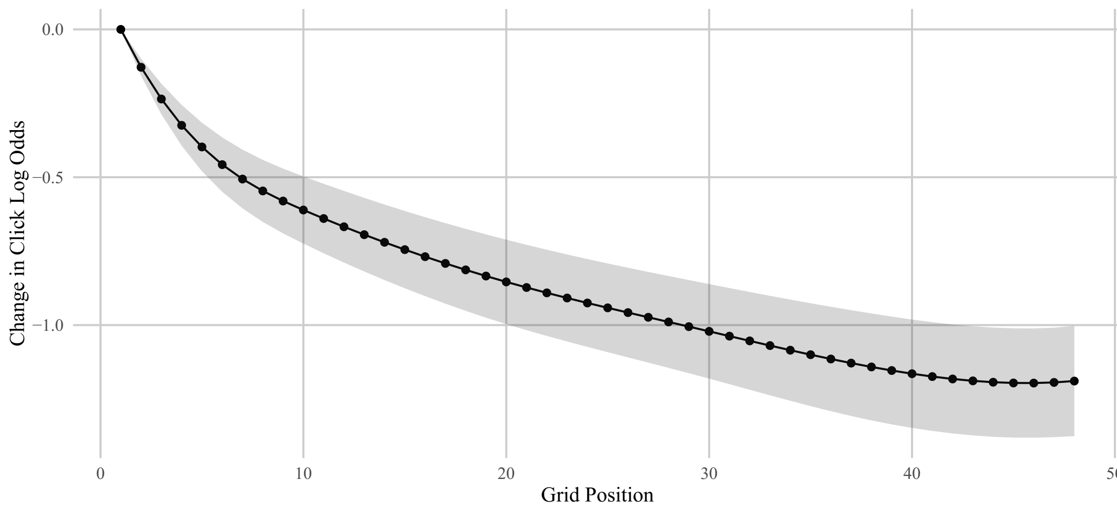

Figure 3.1: *Position effect (first page) relative to first position*

Our fixed-effect estimates suggest substantial heterogeneity in product quality, as illustrated by a histogram of the estimated product fixed effects for an example category in figure 3.2. It is important to remember, however, that these do not necessarily give the true distribution of quality parameters. The distribution of the maximum likelihood estimates would be substantially more dispersed than the distribution of the underlying parameters due to estimation error. We shrink these estimates to reflect the quality inferences a firm would have drawn, and the resulting estimates form a distribution that can have greater or lower variance than the true quality distribution. For interpretation, products to the left of the distribution are not necessarily of low quality to all consumers; the estimates are particular average levels, and with heterogeneous preferences, products may appeal to consumer types of different mass. Also, it is perhaps interesting that estimates in the right tail are shrunk less than those in the left tail because they are estimated on a greater number of impressions due to the retailer's own optimization. This is reflected in the shift of the quality distribution once reweighted by impressions.

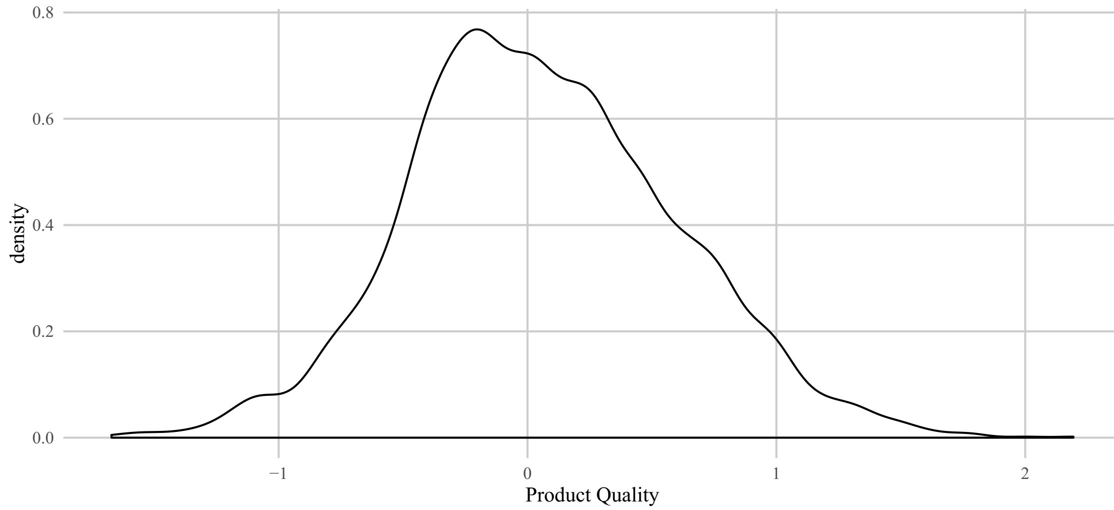

Figure 3.2: *Quality histogram for example category*

3.4.2 Effects on Precision

We first present evidence of sample variability as a function of sample size by plotting the fixed-effect estimates from regression (3.2) run on two non-overlapping subsamples, where we vary the size of each of the subsamples as a fraction of the original full dataset. This illustrates the extent to which two identical firms, represented by the two subsamples, would have agreed in their quality assessment of different products.

We compare sample variability along two dimensions. First, as stated, we vary the fraction of the original data used. Second, we perform a comparison across product categories, which vary by how many consumers visited the corresponding catalog pages in the original data as well as by how product impressions break down over products; an impression is recorded for a product whenever it appears in a menu on a page visited by a consumer. To this end, we construct a measure of the extent to which a particular product category is dominated by its head or tail products as well as of the traffic a particular category saw. Let N_c be the total number of impressions for a particular product category, and let s_j be the fraction of product impressions that were received by product j. Then, $\phi_c = N_c \cdot \sum_j s_j^2$, the Herfindahl index of product impressions in category c multiplied by the total number of product impressions

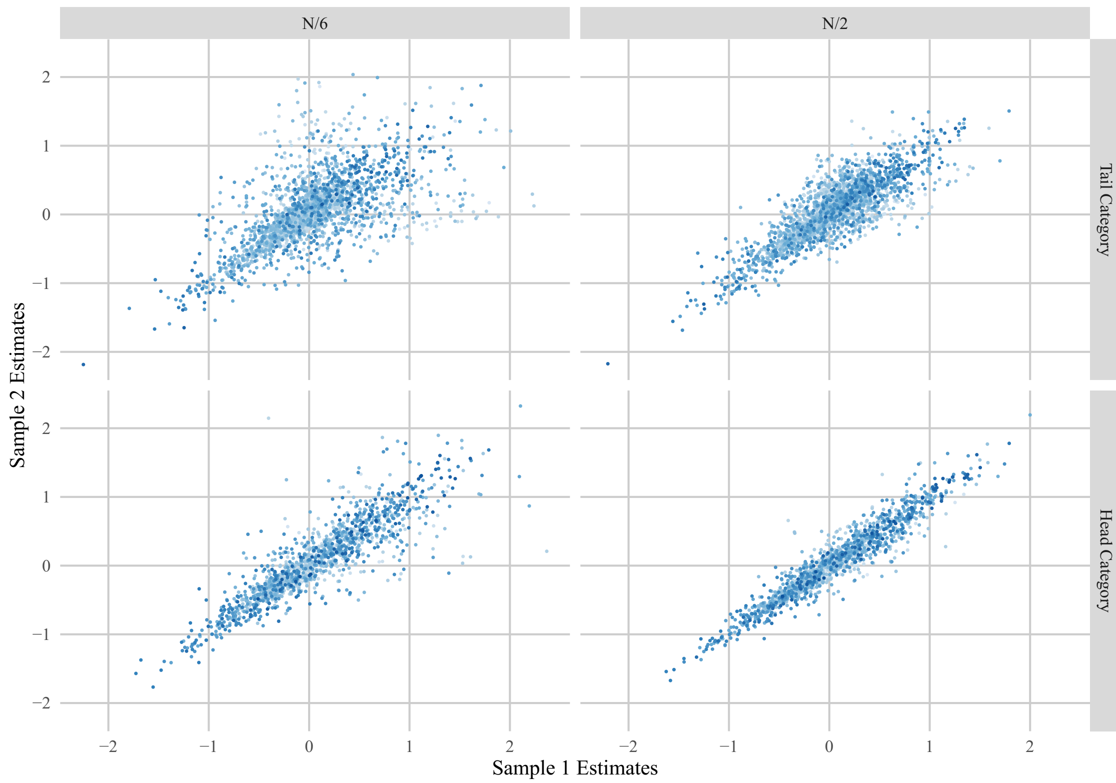

Figure 3.3: *Variability in quality estimates for example categories and different fractions of training data*

in that category, is exactly the expected number of impressions we would get if we drew a product impression at random from its empirical distribution over products and looked at the total number of impressions the associated product received in our sample. This measure is proportional to the number of consumers in a particular product category and the concentration statistic $\sum_j s_j^2$, meaning that it will be higher for product categories which are etiher dominated by its head products or see more traffic. We think of the measure as useful because it indicates how many observations the parameter estimate of the average product being displayed is based on.

Figure 3.3 provides such an assessment of sample variability in the form of a 2-by-2 plot with a less concentrated (tail) category on the top contrasted with a highly concentrated (head) category on the bottom (for roughly equal N), using one sixth of the original data on the left and one half on the right. In each of the four displays, a dot represents a products, and its first and second coordinate are equal to the estimates from two different samples. We observe in the top left display that for a tail-dominated category, the dispersion in quality estimates is substantial when using only one sixth of the data, with many of the observations far off the 45 degree line along which there would be perfect agreement between the two samples. For the optimization problem (3.1), it is of greatest relevance to identify and properly order the highest-quality products. While the estimates overall roughly align along the 45 degree line, the signal-to-noise ratio substantially deteriorates when one zooms in on the right tail of each of the four displays, where the highest-quality products are located. As we triple the sample size for the tail category to one half of the original data and consider the top right display, we see a substantial improvement in precision, as evidenced by much lower dispersion from the 45 degree line. We see a similar improvement as we move from tail to head category for each of the two data fractions in the bottom half of the plot, though some dispersion still remains.

Next, we plot the view-weighted correlation between samples as well as average standard error in product quality estimates for each category against our statistic ϕ_c (figure 3.4), again for the two data sizes of one sixth and one half of the original data. In addition, we show how the standard error in the estimates compares to the standard deviation in the quality estimates at the class level. The latter variation pools true underlying product heterogeneity and estimation error, although the estimates have been shrunk in order to reduce the effect of the latter. While its variance is increasing in ϕ_c, the interpretation is thus not entirely straightforward. We note, however, that the categories with the highest precision are actually the ones with the largest variance in the distribution of estimates, which implies that differences in the latter cannot be driven by estimation error. Rather, it may reflect either differences in true underlying heterogeneity or the fact that estimates based on more observations are shrunk less.

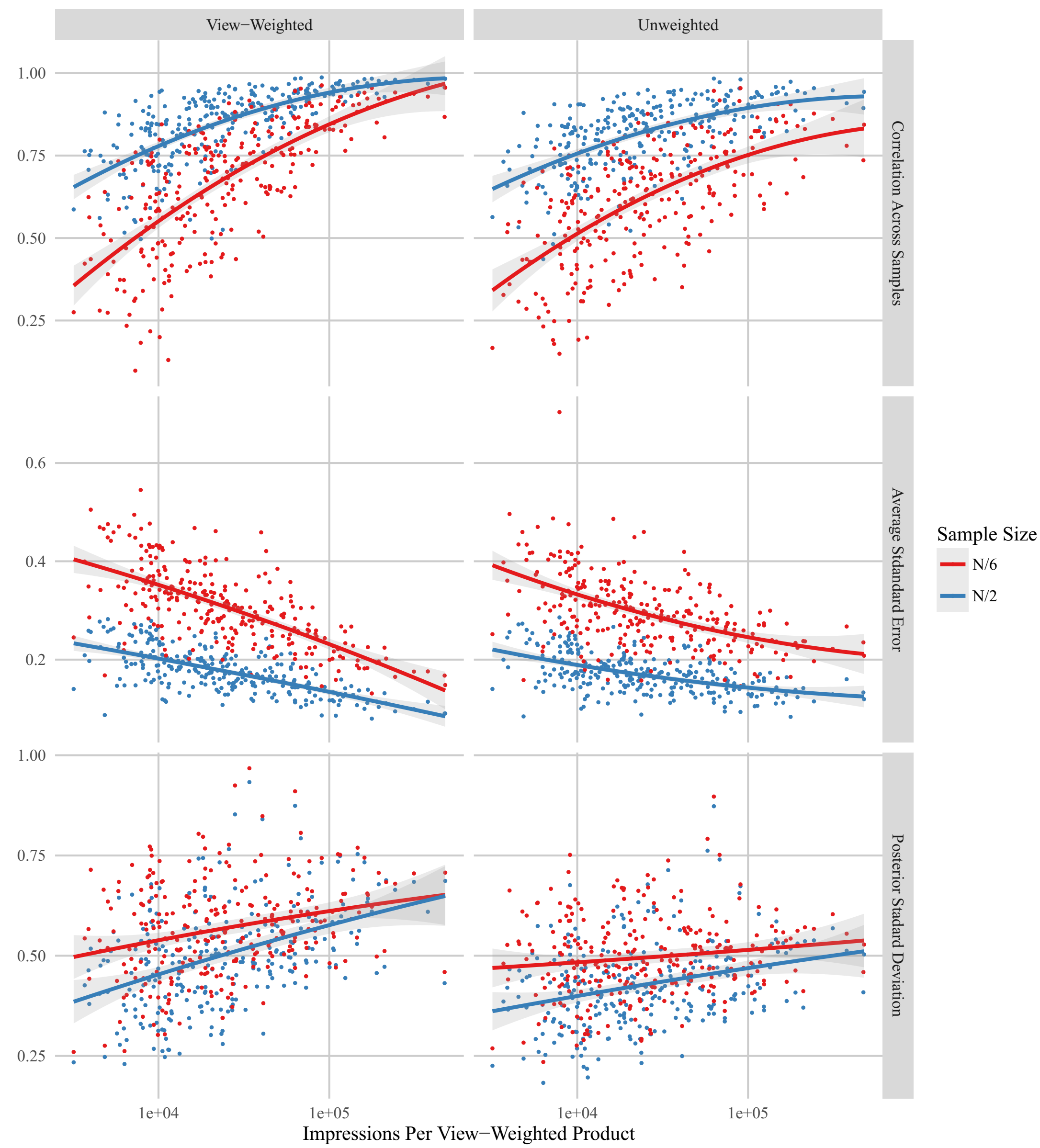

Figure 3.4: *Variability in estimates by average number of impressions of products displayed (ϕ_c)*

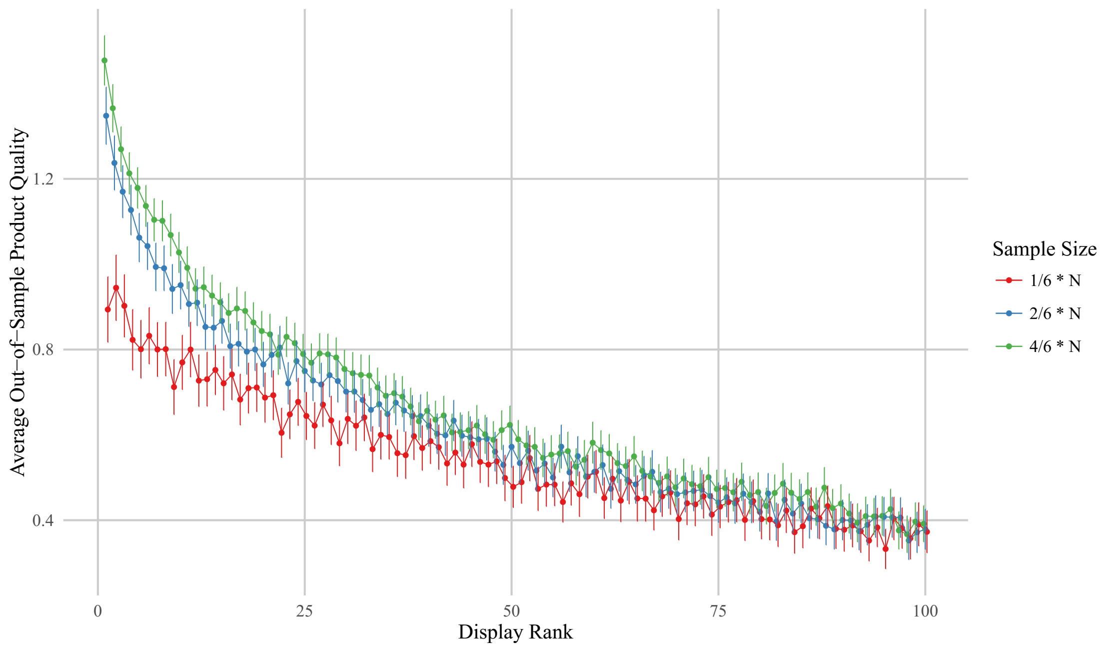

Figure 3.5: *Average out-of-sample quality by position for sort order based on different training data sizes*

3.4.3 Effects on Clicks, Purchases, and Repeat Visits

We now analyze how firm competitiveness varies as a function of the available data. To do so, we compare outcomes in hold-out data for firm decisions taken based on estimates from training datasets of different size.

We begin by calculating the quality of a counterfactual ranking chosen based on estimates from samples of different size. To this end, we estimate logit regression (3.2) on training data consisting of one sixth, one third, and two thirds of the original data as well as on hold-out data containing one sixth of the original observations. We then construct a ranking of the top 100 products according to the different training data estimates to model the solution hypothetical firms of different sizes would have used for optimization problem (3.1); this corresponds to the first two pages of products for a given category. For each product chosen, there is now a hold-out quality estimate which we can use to analyze the quality of the ranking. In figure 3.5, we plot an average across product categories of said quality, weighted by the

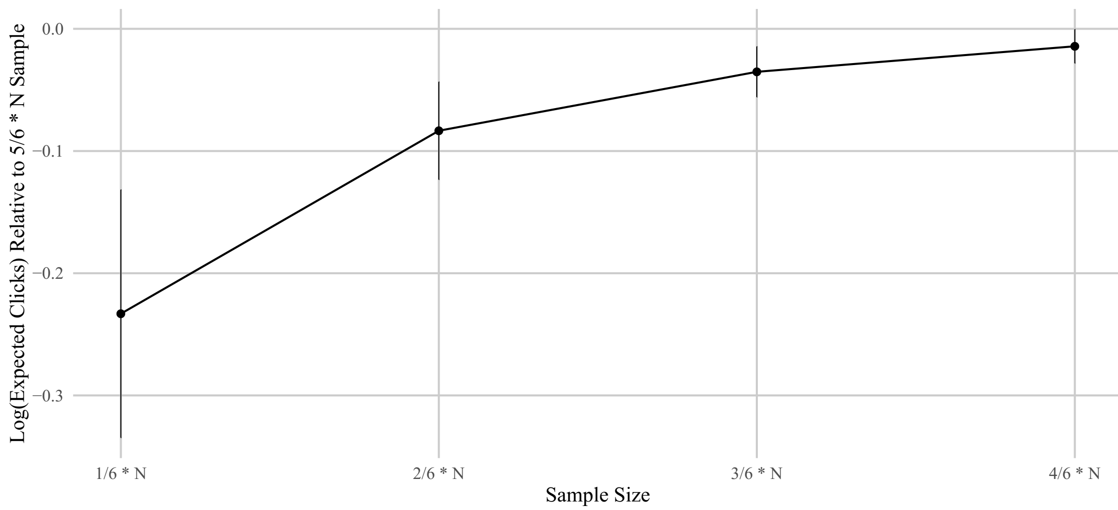

Figure 3.6: *Gain in expected number of clicks from larger training data*

respective number of consumers, for each of the 100 positions. The values on the y-axis are our estimates for the fixed effects that enter the logit link function in (3.2) relative to a mean of zero. It is evident that larger samples are particularly powerful in identifying the top products. We see that the effect decreases by position, and that the gap between one sixth and one third of the sample is much larger than that between one third and one sixth, though it still remains economically significant.

Figure 3.6 shows the average implied loss in the number of expected clicks relative to using five sixths of the data. Going from just one sixth of the data to one third brings an improvement of roughly 14 log points, while doubling the data size again to two thirds yields an additional gain of 8 log points.

Having documented a relationship between the size of training data used and quality supplied, we now turn to the effect of quality on different outcomes. Figure 3.7 plots changes in the realized number of clicks in hold-out data against changes in the expected number of clicks according to model estimates while controlling for granular time and other effects. The slope is large for both new and existing customers and very close to linear as seen by how well the averages of 20 equal-sized bins align. At roughly .4, though, it is quite different

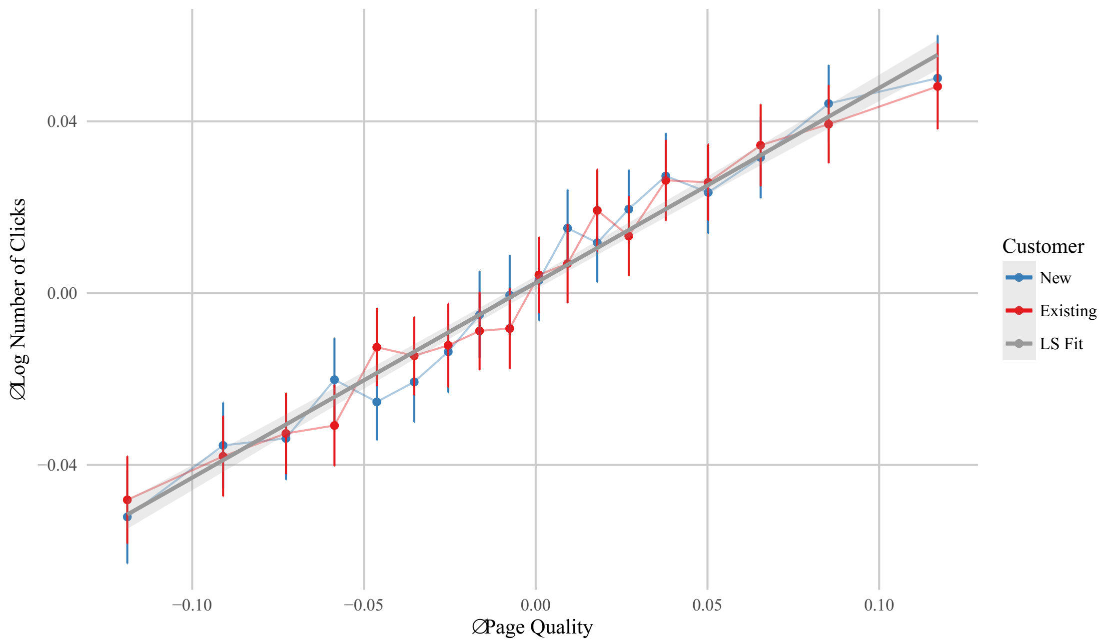

Figure 3.7: *Actual number of clicks against page quality/predicted number*

from the perfect prediction scenario of a unit slope. Possible reasons include misspecification, estimation error, and consumers finding other ways to view products. Overall, however, changes in the predicted quality strongly correlate with the actual number of clicks received.

Finally, we consider the effect of changes in the quality of a catalog page on purchases and repeat impressions. For 20 equal-sized bins, figure 3.8 shows the effect of this quality for new and existing customers on the probability of placing an order (top) or looking at an additional page (both in any product category). We see mostly increasing relationships, with larger slopes for new customers for whom we only consider the very first catalog page they saw in our data, which suggests that consumer beliefs are most sensitive to initial experience, and that continuing customers are either generally more loyal or have already accumulated a stock of positive experiences, making them less sensitive. The effect appears to persist over time, even beyond one week, though at that point, admittedly, the corresponding regression estimates, presented in table 3.1, become relatively noisy.

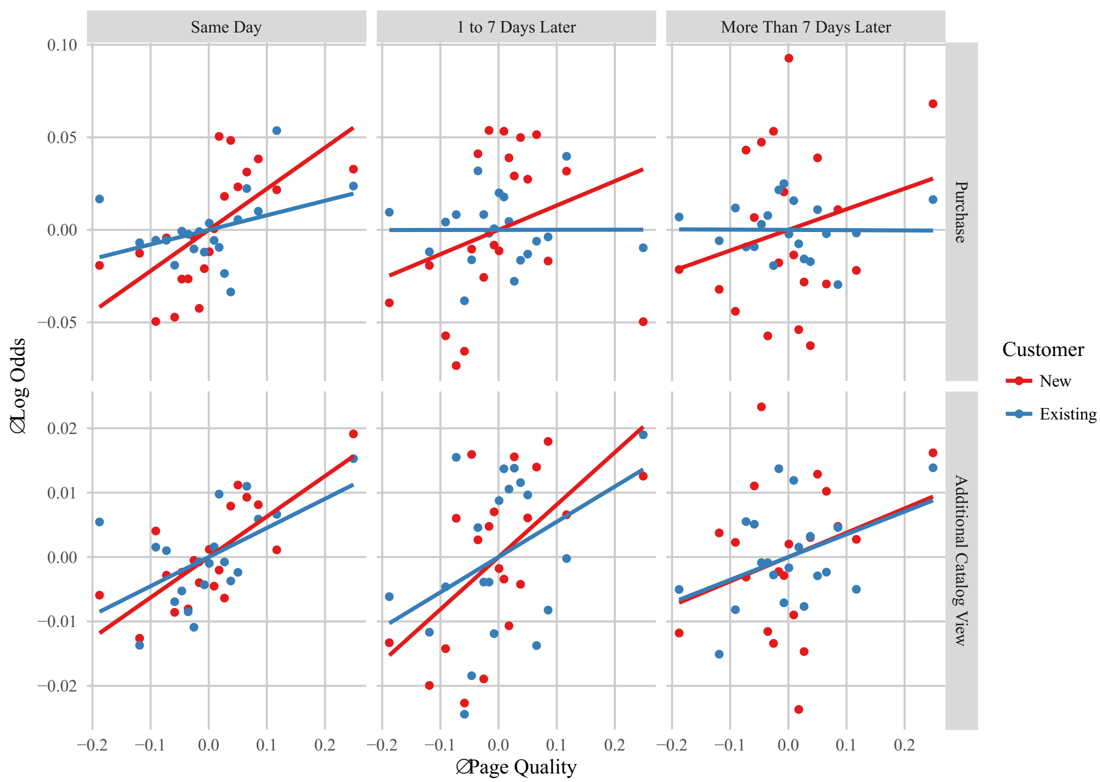

Figure 3.8: *Purchases and repeat impressions against quality*

Table 3.1: *Effect of out-of-sample quality prediction on economic outcomes*

Customer Type	Same day	1-7 days after	>7 days after
	Purchase		
New	0.210***	0.079	0.090
	(0.050)	(0.078)	(0.084)
Existing	0.045	0.012	-0.009
	(0.042)	(0.047)	(0.044)
	Additional search		
New	0.064***	0.054	0.022
	(0.014)	(0.031)	(0.029)
Existing	0.043**	0.057**	0.042**
	(0.014)	(0.018)	(0.015)
	Number of clicked products		
New	0.393***		
	(0.012)		
Existing	0.411***		
	(0.011)		

Notes: This table reports coefficient estimates on the effect of an out-of-sample estimate of the quality/expected number of clicks of a catalog page by customer type. All regressions control for category and customer type specific time trends via splines with 5 and 10 degrees of freedom, respectively. Standard errors – clustered at category-day level – are reported in parentheses. $*p < 0.05$, $**p < 0.01$, $***p < 0.001$.

3.5 Discussion

Our approach to modelling product quality is more closely related to empirical models describing differentiated product demand in product space (e.g., Hausman, 1994) rather than characteristic space (e.g., Berry, 1994, Berry *et al.*, 1995), which is perhaps the more common choice in industrial organization and related fields today. In our setting, however, the primary objects of interest are measures of product quality, which motivates estimating these explicitly. In addition, the sort order algorithm used by the firm in practice also treats product quality levels as individual parameters rather than combinations of preferences over characteristics. Part of the problem is likely the difficulty of modelling quality as a function of covariates for millions of products across a wide range of different product categories, but more importantly, many of the products are very strongly differentiated based on tastes, and these are difficult to capture on the basis of covariates alone. It would nonetheless be interesting to consider an approach that shrinks product quality levels towards a function of covariates; while this should improve precision for any size of dataset, it is not entirely clear whether it would be more beneficial for firms with more or less data. Furthermore, a covariate-based approach would run the risk of reducing the variety of top-ranked products in terms of characteristics.

In competitive settings, the importance of data may also depend on whether these data and the resulting optimization decisions are privately or publicly observable. For instance, in the setting at hand, firms are able to observe the sort orders chosen by competitors, which allows for social learning and may reduce informational asymmetries between firms. The usefulness of this information will then depend on the overlap in product catalogs between retailers as well as on the extent to which they specialize and cater to different consumer preferences. Firms may also experiment to learn qualities resulting in statically suboptimal actions, which complicates social learning for their competitors in this setting because outcomes are privately observed and inference can only be drawn from actions played. Observability will also be affected when firms tailor or personalize their sort order by conditioning on covariates or past behavior (e.g., Fradkin, 2015). Depending on the structure of demand, this will further increase the importance of having large amounts of data, since quality needs to now be estimated conditionally, and individual cells of data can be quite small.

Finally, we note that we have only considered returns to data while keeping the available technology fixed. Many technology firms in particular have hired economists, computer scientists, and data scientists to solve complex optimization and estimation problems. In many cases, these solutions will scale relatively well, so that the associated fixed costs will ammortize significantly faster for larger firms. As a result, these firms may command better infrastructure and human capital and therefore be able to use their data resources more efficiently.

References

ACEMOGLU, D. and ANGRIST, J. (2001). *How Large are Human-Capital Externalities? Evidence from Compulsory-Schooling Laws*, MIT PRess, pp. 9–74.

— and FINKELSTEIN, A. (2008). Input and technology choices in regulated industries: evidence from the health care sector. *Journal of Political Economy*, **116** (5), 837–880.

ALBOUY, V. and LEQUIEN, L. (2009). Does compulsory education lower mortality? *Journal of health economics*, **28** (1), 155–168.

BERRY, S., LEVINSOHN, J. and PAKES, A. (1995). Automobile Prices in Market Equilibrium. *Econometrica*, **63** (4), 841–90.

BERRY, S. T. (1994). Estimating Discrete-Choice Models of Product Differentiation. *RAND Journal of Economics*, **25** (2), 242–262.

BERTRAND, M., DUFLO, E. and MULLAINATHAN, S. (2004). How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, **119** (1), pp. 249–275.

BLACK, S. E., DEVEREUX, P. J. and SALVANES, K. G. (2008). Staying in the classroom and out of the maternity ward? the effect of compulsory schooling laws on teenage births*. *The Economic Journal*, **118** (530), 1025–1054.

BRAUN, T., RAU, F. and TUSCHEN, K. H. (2007). Die drg-einführung aus gesundheitspolitischer sicht. eine zwischenbilanz. *Krankenhaus-Report*, **7**, 3–22.

BUNDORF, M. K., LEVIN, J. and MAHONEY, N. (2012). Pricing and welfare in health plan choice. *American Economic Review*, **102** (7), 3214–48.

BUREAU, U. C. (2015). U.s. electronic shopping and mail-order houses (naics 4541) - total and e-commerce sales by merchandise line (1999-2014).

BUSSE, R., GEISSLER, A. and QUENTIN, W. (2011). *Diagnosis-Related Groups in Europe: Moving towards transparency, efficiency and quality in hospitals*. McGraw-Hill Education (UK).

CALONICO, S., CATTANEO, M. D. and TITIUNIK, R. (2014). Robust nonparametric confidence intervals for regression-discontinuity designs. *Econometrica*, **82** (6), 2295–2326.

CAMERON, A. C., GELBACH, J. B. and MILLER, D. L. (2008). Bootstrap-based improvements for inference with clustered errors. *The Review of Economics and Statistics*, **90** (3), 414–427.

Card, D. (1999). The causal effect of education on earnings. *Handbook of labor economics*, **3**, 1801–1863.

Chamberlain, G. (1980). Analysis of covariance with qualitative data. *The Review of Economic Studies*, **47** (1), 225–238.

Chandra, A., Cutler, D., Song, Z. *et al.* (2012). Who ordered that? the economics of treatment choices in medical care. *Handbook of Health Economics*, **2**, 397–432.

— and Skinner, J. (2012). Technology growth and expenditure growth in health care. *Journal of Economic Literature*, **50** (3), 645–680.

Clark, D. and Roayer, H. (2013). The effect of education on adult mortality and health: Evidence from britain. *The American Economic Review*, **103** (6), 2087–2120.

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

Coulam, R. F. and Gaumer, G. L. (1992). Medicare's prospective payment system: a critical appraisal. *Health Care Financing Review*, **1991** (Suppl), 45–77.

Cutler, D. (1995). The incidence of adverse medical outcomes under prospective payment. *Econometrica*, **63** (1), 29–50.

— (2014). *The quality cure: How focusing on health care quality can save your life and lower spending too*, vol. 9. Univ of California Press.

Cutler, D. M. and Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. *Journal of health economics*, **29** (1), 1–28.

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

Deutsche Krankenhausgesellschaft (DKG) (2015). Datenübermittlung nach § 301 abs. 3 sgb v. `http://www.dkgev.de/media/file/22588.Gesamtdokumentation_SGBV_301_2015-12-22.pdf`, accessed: 2015-02-29.

Devereux, P. J. and Hart, R. A. (2010). Forced to be rich? returns to compulsory schooling in britain*. *The Economic Journal*, **120** (549), 1345–1364.

Deyo, R. A., Mirza, S. K., Martin, B. I., Kreuter, W., Goodman, D. C. and Jarvik, J. G. (2010). Trends, major medical complications, and charges associated with surgery for lumbar spinal stenosis in older adults. *JAMA*, **303** (13), 1259–1265.

Dinerstein, M., Einav, L., Levin, J. and Sundaresan, N. (2014). *Consumer Price Search and Platform Design in Internet Commerce*. Tech. rep.

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

Einav, L., Finkelstein, A., Ryan, S. P., Schrimpf, P. and Cullen, M. R. (2013). Selection on moral hazard in health insurance. *American Economic Review*, **103** (1), 178–219.

—, Jenkins, M. and Levin, J. (2012). Contract pricing in consumer credit markets. *Econometrica*, **80** (4), 1387–1432.

Falk, A. and Szech, N. (2013). Morals and markets. *science*, **340** (6133), 707–711.

Fradkin, A. (2015). *Search Frictions and the Design of Online Marketplaces*. Tech. rep.

Friedman, J., Hastie, T. and Tibshirani, R. (2010). Regularization paths for generalized linear models via coordinate descent. *Journal of Statistical Software*, **33** (1), 1–22.

Friedman, J. H. (2001). Greedy function approximation: a gradient boosting machine. *Annals of statistics*, pp. 1189–1232.

Gaynor, M., Ho, K., Town, R. J. *et al.* (2015). The industrial organization of health-care markets. *Journal of Economic Literature*, **53** (2), 235–84.

Gourieroux, C., Monfort, A. and Trognon, A. (1984). Pseudo maximum likelihood methods: Theory. *Econometrica: Journal of the Econometric Society*, pp. 681–700.

Hausman, J., Hall, B. H., Griliches, Z. *et al.* (1984). Econometric models for count data with an application to the patents-r&d relationship. *Econometrica*, **52** (4), 909–38.

Hausman, J. A. (1994). *Valuation of New Goods under Perfect and Imperfect Competition*. NBER Working Papers 4970, National Bureau of Economic Research, Inc.

Healthcare Cost and Utilization Project (HCUP) (2009). Hcup clinical classifications software (ccs) for icd-10. Agency for Healthcare Research and Quality, Rockville, MD. `http://www.hcup-us.ahrq.gov/toolssoftware/icd_10/ccs_icd_10.jsp`, accessed: 2014-06-20.

Holmstrom, B. and Milgrom, P. (1991). Multitask principal-agent analyses: Incentive contracts, asset ownership, and job design. *Journal of Law, Economics, & Organization*, **7**, 24–52.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142** (2), 615 – 635, <ce:title>The regression discontinuity design: Theory and applications</ce:title>.

— and Rosenbaum, P. R. (2005). Robust, accurate confidence intervals with a weak instrument: quarter of birth and education. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, **168** (1), 109–126.

Institut für das Entgeltsystem im Krankenhaus (InEK) (2007). Kalkulation von fallkosten. `http://www.g-drg.de/cms/Kalkulation2/DRG-Fallpauschalen_17b_KHG/Kalkulationshandbuch`, accessed: 2015-02-29.

Institut für das Entgeltsystem im Krankenhaus (InEK) (2015). German diagnosis related groups, version 2016 – definitionshandbuch. http://www.g-drg.de/cms/G-DRG-System_2016/Definitionshandbuch/Definitionshandbuch_2016, accessed: 2015-02-29.

Jena, A. B., Prasad, V., Goldman, D. P. and Romley, J. (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.

Jeziorski, P. and Moorthy, S. (2015). *Advertiser prominence effects in search advertising*. Tech. rep.

— and Segal, I. (2015). What makes them click: Empirical analysis of consumer demand for search advertising. *American Economic Journal: Microeconomics*, **7** (3), 24–53.

Kemptner, D., Jürges, H. and Reinhold, S. (2011). Changes in compulsory schooling and the causal effect of education on health: Evidence from germany. *Journal of Health Economics*, **30** (2), 340–354.

Lahaie, S. and McAfee, R. P. (2011). Efficient ranking in sponsored search. In *Proceedings of the Seventh Workshop on Ad Auctions*.

Lance Lochner, E. M. (2004). The effect of education on crime: Evidence from prison inmates, arrests, and self-reports. *The American Economic Review*, **94** (1), 155–189.

Machin, S., Marie, O. and Vujić, S. (2011). The crime reducing effect of education*. *The Economic Journal*, **121** (552), 463–484.

McFadden, D. (1973). Conditional logit analysis of qualitative choice behavior. *Frontiers in Econometrics*, pp. 105–142.

McGuire, T. G. (2000). Physician agency. *Handbook of Health Economics*, **1**, 461–536.

Medizinischer Dienst der Krankenversicherung (MDK) (2015). Zahlen – mdk – medizinischer dienst der krankenversicherung. http://www.mdk.de/314.htm, accessed: 2015-02-29.

Mincer, J. (1958). Investment in human capital and personal income distribution. *The journal of political economy*, pp. 281–302.

Mincer, J. A. (1974). Schooling and earnings. In *Schooling, experience, and earnings*, NBER, pp. 41–63.

Oreopoulos, P. (2006). Estimating average and local average treatment effects of education when compulsory schooling laws really matter. *The American Economic Review*, **96** (1), pp. 152–175.

— (2007). Do dropouts drop out too soon? wealth, health and happiness from compulsory schooling. *Journal of Public Economics*, **91** (11–12), 2213–2229.

— (2008). Estimating average and local average treatment effects of education when compulsory schooling laws really matter: Corrigendum. *The American Economic Review*.

—, Page, M. E. and Stevens, A. H. (2006). The intergenerational effects of compulsory schooling. *Journal of Labor Economics*, **24** (4), pp. 729–760.

— and Salvanes, K. G. (2011). Priceless: The nonpecuniary benefits of schooling. *The Journal of Economic Perspectives*, **25** (1), 159–184.

Pischke, J.-S. and von Wachter, T. (2008). Zero returns to compulsory schooling in germany: Evidence and interpretation. *Review of Economics and Statistics*, **90** (3), 592 – 598.

Ridgeway, G. (2007). Generalized boosted models: A guide to the gbm package. *Update*, **1** (1), 2007.

Schreyögg, J., Bäuml, M., Krämer, J., Dette, T., Busse, R. and Geissler, A. (2014). *Forschungsauftrag zur Mengenentwicklung nach § 17b Abs. 9 KHG*. Tech. rep., Hamburg Center for Health Economics/Universität Hamburg, Hamburg.

—, Bäumler, M. and Busse, R. (2009). Balancing adoption and affordability of medical devices in europe. *Health Policy*, **92** (2), 218–224.

—, Tiemann, O. and Busse, R. (2006). Cost accounting to determine prices: how well do prices reflect costs in the german drg-system? *Health Care Management Science*, **9** (3), 269–279.

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

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

Turner, H. and Firth, D. (2007). Generalized nonlinear models in r: An overview of the gnm package.

UK Data Service (2004). *Labour Force Survey, 1975 – 2001*. Tech. rep.

Ursu, R. (2016). *The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions*. Tech. rep.

Yip, W. C. (1998). Physician response to medicare fee reductions: changes in the volume of coronary artery bypass graft (cabg) surgeries in the medicare and private sectors. *Journal of Health Economics*, **17** (6), 675–699.

Appendix A

Appendix to Chapter 2

A.1 Proof of Results in Linear Model

Following the setup of section 2.3, we can write down the covariance matrix for the joint estimator as

$$Var\begin{pmatrix} \hat{\beta} \\ \hat{\delta} \end{pmatrix} = \begin{pmatrix} Var\left(r_{\beta i}\right)\left(Z'S\right)^{-1}Z'Z\left(S'Z\right)^{-1} & Cov\left(r_{\beta i}, r_{\delta i}\right)\left(S'S\right)^{-1} \\ & Var\left(r_{\delta i}\right)\left(S'S\right)^{-1} \end{pmatrix}$$

where

$$\begin{aligned}
r_{\beta i} &= \left(\theta'\gamma_1\right)\left(\beta_1 a_i + \varepsilon_{2i}\right) + \theta'\xi_i \\
r_{\delta i} &= \left(\theta'\gamma_1\right)\left(\beta_1\left(a_i\frac{\sigma_s^2 - \alpha_1^2\sigma_a^2}{\sigma_s^2} - \frac{\alpha_1\sigma_a^2}{\sigma_s^2}\left(\alpha_2 z_i + \varepsilon_{1i}\right)\right) + \varepsilon_{2i}\right) + \theta'\xi_i \\
\sigma_x^2 &= \mathrm{Var}\left(x\right) \\
&\therefore \\
Var\left(r_{\beta i}\right) &= \left(\theta'\gamma_1\right)^2\left(\beta_1^2\sigma_a^2 + \sigma_{\varepsilon_2}^2\right) + \theta'\Sigma\theta \\
Cov\left(r_{\beta i}, r_{\delta i}\right) &= \left(\theta'\gamma_1\right)^2\left(\beta_1^2\sigma_a^2\left(\frac{\sigma_s^2 - \alpha_1^2\sigma_a^2}{\sigma_s^2}\right) + \sigma_{\varepsilon_2}^2\right) + +\theta'\Sigma\theta \\
Var\left(r_{\delta i}\right) &= \left(\theta'\gamma_1\right)^2\left(\beta_1^2\left(\sigma_a^2\left(\frac{\sigma_s^2 - \alpha_1^2\sigma_a^2}{\sigma_s^2}\right)^2 + \left(\frac{\alpha_1\sigma_a^2}{\sigma_s^2}\right)^2\left(\alpha_2^2\sigma_z^2 + \sigma_{\varepsilon_1}^2\right)\right) + \sigma_{\varepsilon_2}^2\right) + +\theta'\Sigma\theta
\end{aligned}$$

and thus using the delta method, we have

$$
\begin{aligned}
Var\left(\frac{\hat{\beta}_2}{\hat{\delta}_2}\right) &= \left(\begin{array}{c} \frac{1}{\delta} \\ \frac{\beta}{\delta^2} \end{array}\right)' Var\left(\begin{array}{c} \hat{\beta}_2 \\ \hat{\delta}_2 \end{array}\right)\left(\begin{array}{c} \frac{1}{\delta} \\ \frac{\beta}{\delta^2} \end{array}\right)(\theta'\gamma_1)^{-2} \\
&= (\theta'\gamma_1)^{-2}\left(\frac{1}{\delta^2}k_1 Var\left(r_{\beta i}\right) + 2\frac{\beta}{\delta^3}k_2 Cov\left(r_{\beta i}, r_{\delta i}\right) + \frac{\beta^2}{\delta^4}k_2 Var\left(r_{\delta i}\right)\right) \\
&= c_0 + c_1(\theta'\gamma_1)^{-2}\left(\theta'\Sigma\theta\right)
\end{aligned}
$$

where c_0, c_1 are constants based only on the model primitives and $c_1 > 0$. Thus, solving this for a zero FOC wrt θ, and exploiting the symmetry of Σ, we have

$$
\begin{aligned}
\frac{\partial}{\partial\theta} Var\left(\frac{\hat{\beta}_2}{\hat{\delta}_2}\right) &= c_1\left(\frac{2\theta'\Sigma}{(\theta'\gamma_1)^2} - 2\gamma_1\frac{\theta'\Sigma\theta}{(\theta'\gamma_1)^3}\right) \\
&\therefore \\
\Sigma\theta &= \gamma_1\left(\frac{\theta'\Sigma\theta}{\theta'\gamma_1}\right) \\
&\therefore \\
\theta &= a\Sigma^{-1}\gamma_1
\end{aligned}
$$

for any $a \neq 0$.

Finally, we can solve for the reverse regression $\mathrm{E}\left[s_i \mid \{y_{mi}\}_m\right]$, by noting

$$
\begin{aligned}
\mathrm{E}\left[s_i \mid \tilde{y}_i\right] &= c_{10} + c_{11}\tilde{y}_i \\
\mathrm{E}\left[\tilde{y}_i \mid y_i\right] &= c_{20} + c_{21}\left(\Sigma^{-1}\gamma_1\right)' y_i \\
\mathrm{E}\left[s_i \mid y_i\right] &= \mathrm{E}\left[\mathrm{E}\left[s_i \mid \tilde{y}_i\right] \mid y_i\right] \\
&= c_{30} + c_{11}c_{21}\left(\Sigma^{-1}\gamma_1\right)' y_i
\end{aligned}
$$

The result for $\mathrm{E}\left[\tilde{y}_i \mid y_i\right]$ is due to a simple application of the “Sherman–Morrison formula” to the solve

$$
\begin{aligned}
\operatorname{Var}(y_i)^{-1}\operatorname{Cov}(y_i,\tilde{y}_i) &= \left(\Sigma+\gamma_1\gamma_1{}'\operatorname{Var}(\tilde{y}_i)\right)^{-1}\left(\gamma_1\operatorname{Var}(\tilde{y}_i)\right) \\
&= \left(\Sigma^{-1}+\frac{\Sigma^{-1}\gamma_1\gamma_1{}'\Sigma^{-1}}{\operatorname{Var}(\tilde{y}_i)^{-1}+\gamma_1{}'\Sigma^{-1}\gamma_1}\right)\gamma_1\operatorname{Var}(\tilde{y}_i) \\
&\quad \left(\frac{\operatorname{Var}(\tilde{y}_i)\,\gamma_1{}'\Sigma^{-1}\gamma_1}{\operatorname{Var}(\tilde{y}_i)^{-1}+\gamma_1{}'\Sigma^{-1}\gamma_1}\right)\Sigma^{-1}\gamma_1
\end{aligned}
$$

A.2 Additional Figures

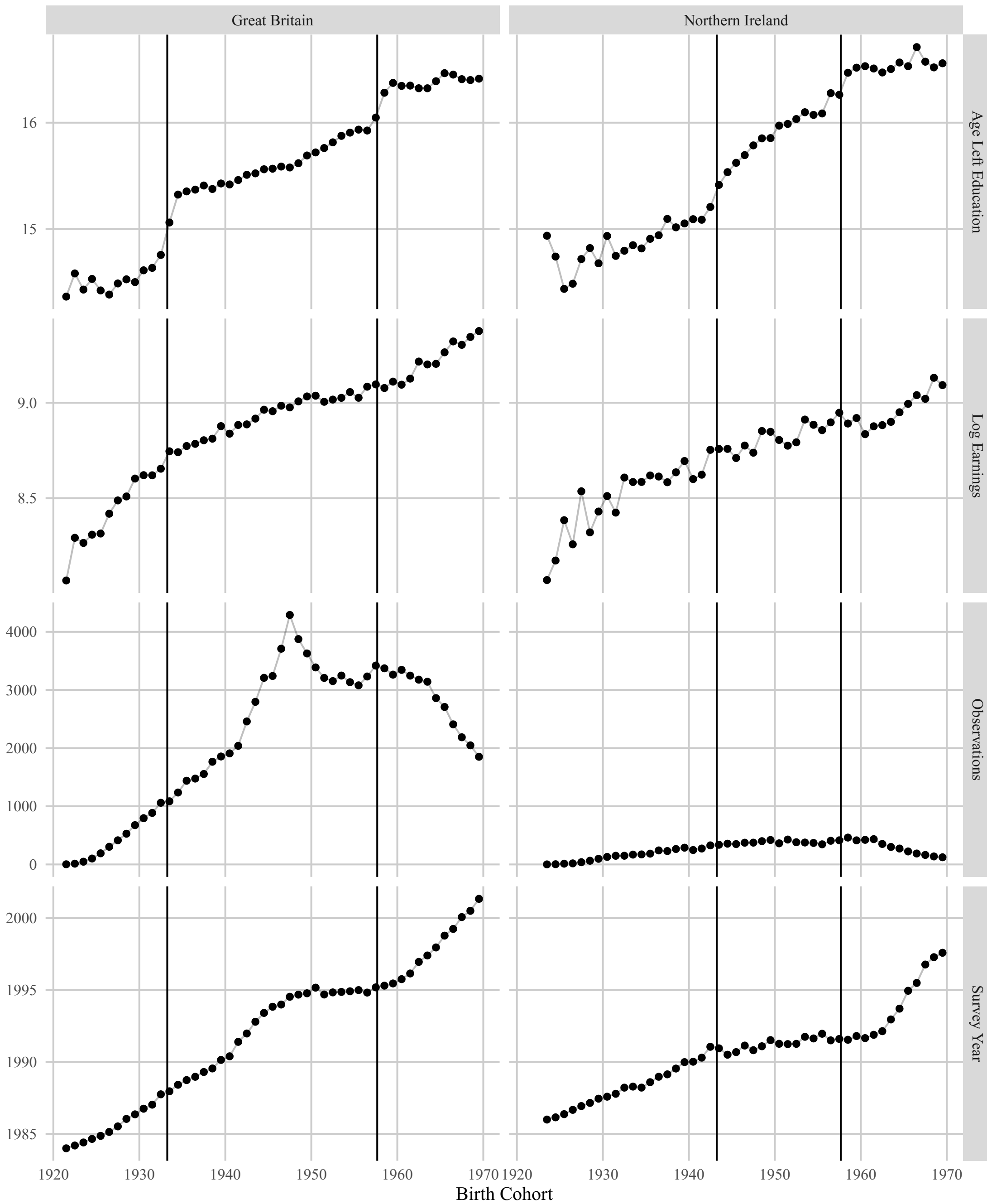

Figure A.1: *Graphical summary of re-weighted Oreopolous (2008) data*

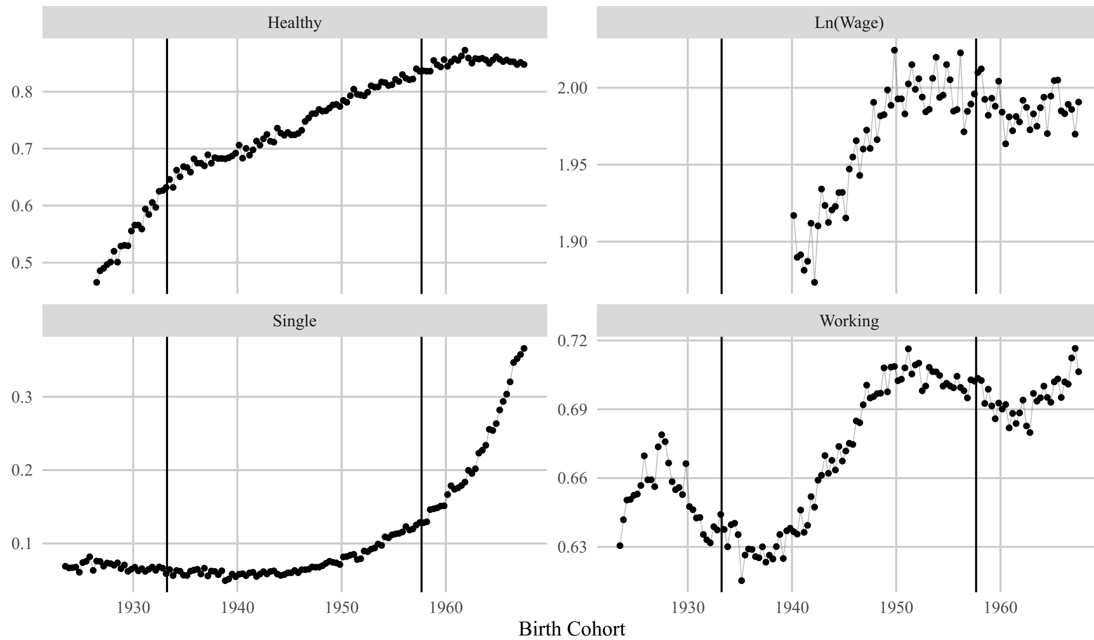

Figure A.2: *Individual LFS outcomes, averaged by birth cohort.*

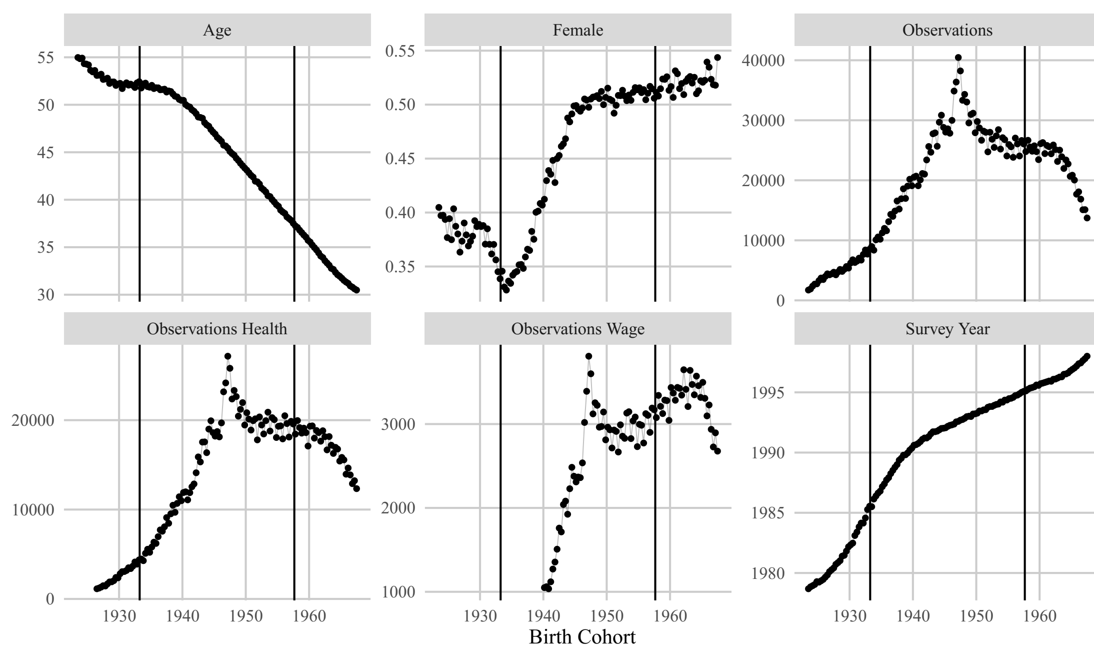

Figure A.3: *Observations and control variables for LFS data*

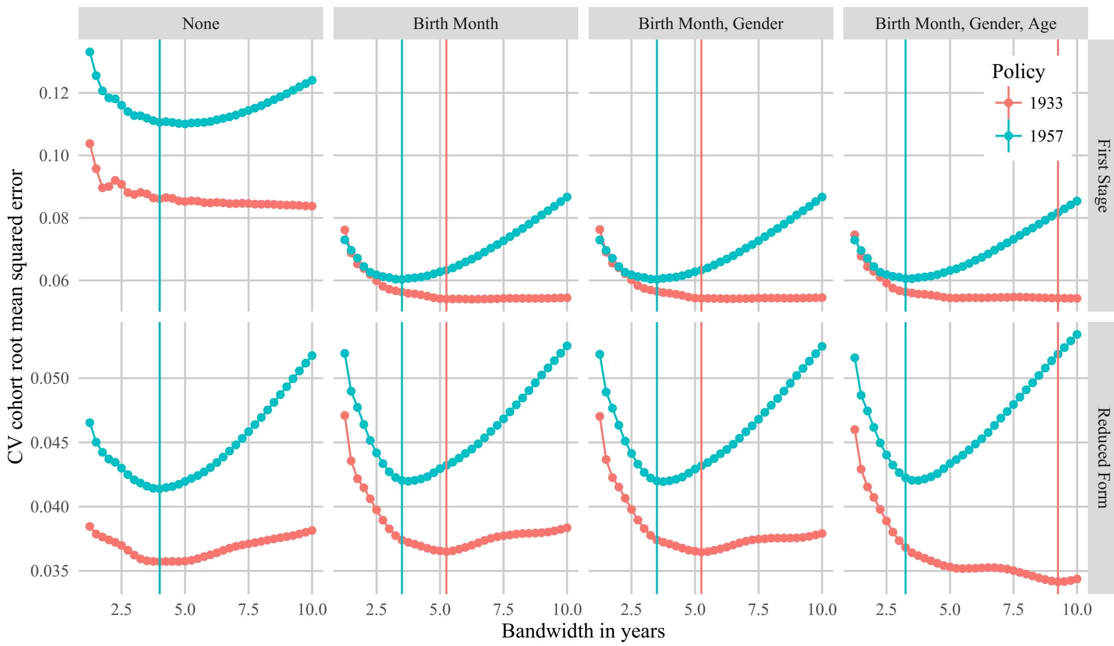

Figure A.4: *Cross-validated MSE by bandwidth and specification*

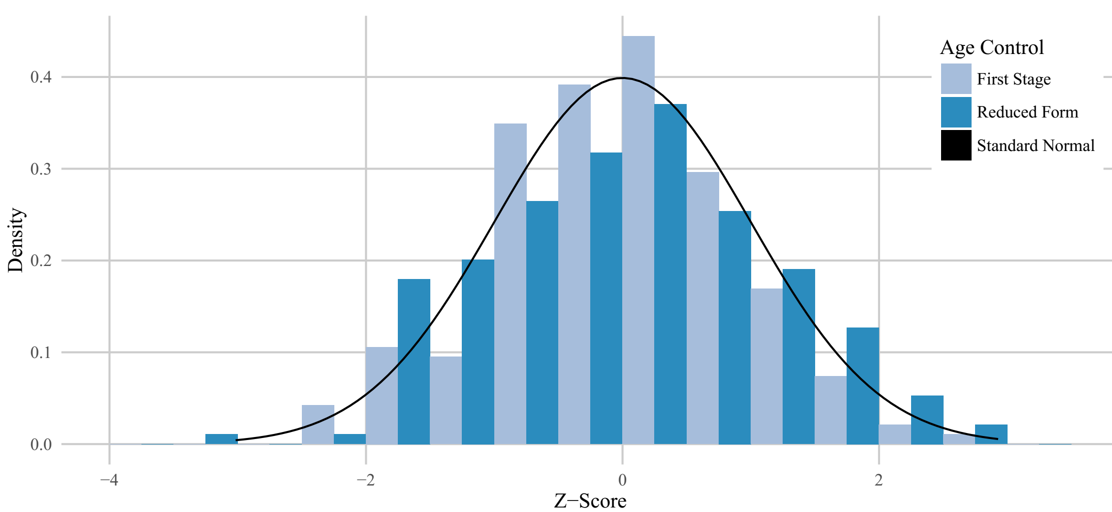

Figure A.5: *Z-score distribution of placebo test*