# Consumer Heterogeneity and Medical Care Price Responsiveness: Evidence and Implications for Optimal Insurance Design*

Zarek C. Brot-Goldberg
Economics Department, UC Berkeley

Amitabh Chandra
Kennedy School, Harvard University and NBER

Benjamin R. Handel
Economics Department, UC Berkeley and NBER

Jonathan T. Kolstad
Wharton School, University of Pennsylvania and NBER

PRELIMINARY AND INCOMPLETE: DO NOT DISTRIBUTE

October 15, 2014

**Abstract**

Measuring consumer responsiveness to medical care prices is a central issue in health economics and a key ingredient in the optimal design and regulation of health insurance markets. We study consumer price responsiveness to medical care prices leveraging a natural experiment that occurred at a large self-insured firm that forced all of its employees to switch from an insurance plan that provided free health care to a non-linear high deductible insurance plan with markedly higher cost sharing. We find that the switch caused a 16.45% reduction in total yearly medical expenditures at the firm, leading to approximately $123 million in savings per year. We use this plan switch to study heterogeneity in price responsiveness as a function of (i) consumer demographics (ii) consumer health status and (iii) types of medical services consumed. We use both regression and machine learning techniques to estimate heterogeneous demand elasticities for medical care, and use our results to study the welfare implications of the plan switch as well as counterfactual analyses of (i) optimal menu design and (ii) optimal value-based contract design. We find that the plan switch led to a welfare increase between $604 and $794 per individual, with the benefits from more efficient medical spending outweighing the incremental loss from reduced risk protection.

*We thank Microsoft Research for their support of this work. We also thank seminar participants at Microsoft Research. Finally, we thank Ishita Chordia for excellent research assistance. All errors are our own.

# 1 Introduction

Spending on health care services in the United States has grown rapidly over the past 50 years, increasing from 4.6% of GDP in 1960 to 16.4% in 2011 (Congressional Budget Office (2013)). As the rise in health care spending has increased, policymakers, large employers, and insurers have grappled with the problem of how to limit growth in health care spending without substantially reducing the quality of medical care consumed. To this end, decision-makers have explored both (i) supply side policies that aim to directly restrict the technologies and services consumers can access and (ii) demand side policies whereby consumers with insurance face a greater portion of the full price for health care services. Though many nations around the world have focused on supply side policies that ration care, in the United States there has been a limited public appetite for such policies, and, consequently, more of a burden on demand side policies to control health care spending.

As a result, employers offering health insurance in the United States (covering the majority of consumers nationwide) have increasingly moved towards high-deductible health insurance plans (HDHPs) that increase the percentage of medical care that consumers pay for. Such plans force consumers to pay the full cost of their care initially, prior to hitting a deductible, and offer more generous coverage as the amount spent increases beyond the deductible and consumer risk protection becomes more of a pressing concern than consumer price sensitivity. Beyond the value such plans provide in terms of health care cost savings, they were made more attractive by the 2003 Medicare Modernization Act that allows plans with sufficiently high deductibles to be bundled with health savings accounts (HSAs) that allow employees to save tax-exempt money to be used for health care spending at any point in their lives. It is clear that such plans have become increasingly popular with employers just recently: enrollment in such plans through employers increased from 15% to 33% of all relevant consumers from 2010 to 2014, while 16% of employers *only* offered a HDHP to employees in 2014, a number that is expected to rapidly increase to 30% in 2015 (Towers Watson (2014)).

Given these recent trends, it is important for insurers and policymakers to precisely understand the welfare implications of shifting consumers towards insurance with higher cost sharing. In the spirit of Zeckhauser (1970), this involves understanding (i) the potential welfare gains from more efficient health care utilization in HDHPs (relative to more comprehensive insurance options) and (ii) the potential welfare losses from additional risk exposure in HDHPs. While a range of recent studies have focused on understanding the welfare implications of different insurance allocations for risk exposure (see e.g. Cohen and Einav (2007), Handel (2013), Barseghyan et al. (2013), or Handel and Kolstad (2014)) it has been difficult to comprehensively characterize consumer price sensitivity across (i) the range of health care services typically covered by health insurance and (ii) a range of consumer preference types. This not only has made it difficult to effectively evaluate the welfare implications of the shift towards HDHPs, but has also made it more difficult to answer a range of questions related to optimal design and regulation of insurance markets.

In this paper we comprehensively quantify consumer price responsiveness for a large range of

medical services and consumer types, leveraging a proprietary dataset from a large self-insured firm covering approximately 52,000 employees and 150,000 lives. The data, spanning 2009-2013, contain detailed information on employee (and dependent) demographics, line by line health claims, and insurance plan enrollment along with several other linked benefits decisions. Prior to 2013, the firm offered a menu of insurance options, and approximately 85% of employees and covered lives were enrolled in a plan with no consumer cost sharing (i.e. the plan covered 100% of all medical spending). For 2013, the firm discontinued this popular option, forcing all of its employees enrolled in that plan into the other primary option, a high-deductible plan. Employees at the firm paid 22% of total medical expenses under this plan in 2013, a substantial increase from the 0% employees paid previously under the PPO. Moreover, the HDHP allowed access to exactly the same network of physicians and medical treatments (at essentially the same base prices) in 2013 as the PPO had in 2012, implying that the change to coverage was purely in the form of financial cost-sharing. This large forced shift provides exogenous variation in cost sharing (and the entire non-linear contract) from 2012 and before to 2013, which we use to quantify the degree of consumer price sensitivity for a range of medical services and for different consumer types. In the raw data, spending for the 85% of consumers forced to switch to the HDHP from the free PPO option decreased by 19.61% from 2012 to 2013, reducing total health care spending at the firm by approximately $150 Million overall.

While the change in spending in the raw data reflects the absolute spending change from 2012 to 2013, we implement several additional layers of analysis to obtain a causal impact of the switch to the HDHP on total medical spending at the firm. First, we adjust spending over time from 2009-2013 with a medical CPI that controls for general time trends in medical spending during this time period. Second, we implement a model adjust for aging in our sample that occurs over time. These two adjustments change the treatment effect of the forced HDHP switch to a 22.90% reduction in total medical spending.

Importantly, it is clear in the raw data that there is also anticipatory intertemporal substitution of medical spending from the beginning of 2013 (when claims fall under the HDHP) to the end of 2012 (when claims are free for consumers). This type of substitution has been documented in the domain of dental care (see Cabral (2013) but, to our knowledge, not rigorously in the domain of health insurance. We leverage our panel of claims from 2009-2013 to empirically quantify an excess mass of medical spending in the final months of 2012, and a corresponding absence of claims from the beginning months of 2013. We apply this methodology both to claims in aggregate, and to specific kinds of claims and types of consumers. We estimate an average excess mass of $166 per consumer combined over November and December 2012 (95% CI of [$122.57, $212.21]), equal to approximately 20% of average per consumer spending in those two months for each of the preceding three years. We develop a causal impact of the switch to the HDHP on total health spending by adjusting for this anticipatory spending: without incorporating this adjustment our analysis would overestimate consumer price responsiveness because 2012 spending would be high relative to a normal year in the PPO and 2013 spending would be low relative to a normal year in the HDHP. Incorporating this adjustment, our primary causal treatment effect estimate finds that

the forced HDHP switch reduced total medical spending by 16.45%, equivalent to approximately $123 million at the firm we study. We translate this estimate into a semi-arc elasticity so that it can be directly compared to prior work on consumer price responsiveness in health care. Our primary causal treatment effect corresponds to a semi-arc elasticity of -0.484, approximately half of the semi-arc elasticity from two of the primary comparisons conducted in the RAND Health Insurance Experiment (results summarized in Newhouse and the Insurance Experiment Group (1993)).[1]

One of the primary advantages of the depth of our data, combined with the exogenous variation in health plan enrollment (from the individual perspective), is the ability to comprehensively model heterogeneous responses to the change in cost sharing, both with respect to types of medical spending and types of consumers. We document heterogeneity in price responsiveness both in the raw data and with our causal treatment effect estimates incorporating anticipatory spending, the medical CPI, and population aging. For consumer types based on observable demographics, our primary causal estimate reveals fairly uniform effects as a function of consumer age. As a function of income, we find that consumers who earn between 0 and $100,000 per year are three times as price responsive as consumers earning over $200,000 per year, reflecting the fact that lower income consumers are more price sensitive. We find the both employees and their spouses are more likely to reduce their own medical spending than that of their dependents. In terms of ex ante health status measures, we find that consumers predicted to be in the top quartile of spending reduce spending by double all other employees, with similar results for employees with 3+ chronic conditions relative to those with fewer chronic conditions.

For different types of medical services, we find not surprisingly that consumers are less responsive with respect to inpatient hospital claims relative to most other claims categories. Our primary causal estimates indicate a 14% reduction in spending on inpatient hospital services, and an 11% reduction in outpatient hospital spending, which together account for approximately 40% of total spending. Interestingly, consumers reduce emergency room spending by 23% as the result of increased cost sharing, suggesting that such spending is fairly manipulable. Consumers reduce spending on branded drugs by 20% and generic drugs by 21%, though much of the reduction for generic drugs comes from consumers substituting to cheaper generics while for brands the effect comes from consumers substituting to generics. We find a 17% reduction in spending on mental health services. Interestingly, we find a 12% reduction in spending on preventive health care services, which account for 9% of total spending and are free for consumers under the HDHP. Thus, despite there being no actual price change for these services, consumers reduced spending in this domain by 12%, potentially suggesting a lack of information about these services being free or that these services are bundled with other services that are not free for a given office or hospital visit.

We present a series of robustness analyses revealing that these treatment effect results are robust to a range of alternative sample definitions and alternative methods for quantifying intertemporal

[1]The most typically cited elasticity estimates in the literature are arc elasticities, not semi-arc elasticities. We use semi-arc elasticities, because, for a change starting from (or ending in) a health plan with 0 price for consumers, an arc elasticity yields an estimate that does not reflect the magnitude of the price change. We note that we compute semi-arc elasticities using statistics in Newhouse and the Insurance Experiment Group (1993), and that those numbers may change after discussions with the authors of that work.

substitution. In addition, we show that the change in medical spending in 2013 after the introduction of the HDHP is not due to changes in provider prices, which remained essentially constant over time.[2] We also analyze whether or not there is attrition from the firm after either (i) the announcement of the forthcoming HDHP switch or (ii) the HDHP switch itself. We find that employee entry and exit patterns look similar in each year throughout our sample period and that there is limited to no attrition of spouses, dependents, or specific types of consumers.

In order to be able to conduct welfare analysis and counterfactual analysis related to different insurance plan offerings, we develop a series of models to predict individual-level price responsiveness. First, we estimate regression models that take into account consumer demographics, consumer health risk, and type of medical services consumed. These regressions predict similar overall treatment effects to those described in our primary causal analysis mentioned above, and incorporate heterogeneity in medical services and consumer types using a typical econometric approach whereby the researcher selects the form of heterogeneity to be studied. Second, we use an unsupervised learning model to search over variables that describe consumer heterogeneity in our sample, classifying consumers into types that best predict differential utilization effects. These techniques draw on insights from machine learning and allow for greater flexibility in investigating heterogeneous responses to the forced switch to the HDHP. We discuss these methods and illustrate the value they add in describing heterogeneity in the population and subsequent value added in our optimal insurance design counterfactuals.

We use these heterogeneous treatment effect estimates for consumer price responsiveness, together with estimates of consumer health spending risk and risk aversion (drawn from work with the same data in Handel and Kolstad (2014)), to study several distinct questions related to optimal insurance design.[3]. First, we investigate the welfare impact of the within-sample forced switch to the HDHP in 2013. We compute this welfare impact for a range of assumptions on the value of foregone medical care in the HDHP, and the investigate how the welfare benefits from more efficient utilization of medical care trade off with the welfare losses from reduced risk protection. Our primary results reveal that the welfare gain from more efficient medical care utilization ranges from \$678 to \$868 per person, while the welfare loss from increased risk exposure is \$74 per person on average, implying an overall mean welfare impact in the range [\$604.1, \$794.8] per individual. The lower bound occurs when the value of foregone medical care equals the average coinsurance rate for that care in the population, while the upper bound occurs when that care is purely wasteful. Depending on the assumption on the foregone value of medical spending, a % reduction in spending between 1.4% and 6.4% would make the forced switch break even from a welfare perspective.

Next, we investigate two optimal insurance design counterfactuals using our price elasticity estimates together with estimates of risk aversion and health risk from Handel and Kolstad (2014). First, we investigate a value-based insurance design counterfactual where we ask what insurance contract the firm should offer if (i) it wants to maximize social welfare and (ii) it can vary cost shar-

[2]The insurance administrator for the HDHP is the same as the insurance administrator for the PPO, which makes this ex ante likely as well.

[3]The model of health spending risk and estimation of risk aversion parameters is described in depth in Handel and Kolstad (2014)

ing by type of medical service. In this analysis, we partition the space of potential medical services into categories, and allow different cost sharing rates for each category. The welfare analysis then trades off risk protection and efficient medical care utilization in a category specific manner, with aggregation in terms of overall consumers risk exposure. We first solve for the optimal value-based contract that holds plan actuarial value constant at the same level as the HDHP in our sample. Next, we find the optimal value based contract regardless of actuarial value. We discuss the welfare improvements from our optimal value-based contract relative to the HDHP, and discuss the underpinnings for those improvements. Finally, we also investigate an additional counterfactual that ask what the optimal menu of two standard non-linear insurance contracts is, given consumer heterogeneity in both price responsiveness and risk aversion. Consumer heterogeneity is a primary motivation for menu offerings (as in the ACA) relative to just offering one contract, and we characterize optimal offerings of two plans for a population given our estimates.

There is a long history of past work studying consumer price responsiveness with respect to health care prices. The most cited study in this literature is the RAND Health Insurance Experiment (results summarized in Newhouse and the Insurance Experiment Group (1993)) which randomized consumers into different health insurance plans and tracked their subsequent utilization. While this kind of randomized evaluation provides very clean exogenous variation in consumer exposure to different medical prices, the relatively small sample size in the Experiment made it unable to explore consumer response heterogeneity in the way we are able to in our analysis here, where we also leverage detailed line by line claims data and consumer demographics. Our analysis here is also able to compare consumer responses to different arms of non-linear contracts, in the spirit of Einav et al. (2013a) and Aron-Dine et al. (2012), since we observe exogenous variation between completely free health care and a non-linear HDHP typically found in health insurance markets. Our analysis is also able to take advantage of time series variation whereby we observe the same consumers in different plans over time, which the RAND HIE was not able to. Finally, given the changing nature of health care spending over the past 35 years since the RAND HIE was conducted, the current environment may also be quite different empirically in terms of medical services and consumer preferences.

In more recent work, Finkelstein et al. (2012) study health care utilization in the context of Medicaid, with exogenous variation in Medicaid coverage due to a coverage lottery conducted in Oregon in 2008. The authors find consumer utilize more health care services when enrolled in Medicaid, and only investigate other socio-economic outcomes for these consumers. This project and the RAND HIE project are the two projects we are aware of that try to use exogenous variation in coverage to comprehensively study the range of services typically covered under health insurance. There are a range of projects that use natural experiments to study price responsiveness to specific types of medical services. For example, Chandra et al. (2010) studies the effect changes in copayments for individuals enrolled in California's insurance program for civil servants. They find price elasticities in the range of the RAND HIE, for specific services such as physician office visits and certain types of prescription drugs. Other work in this vein has specifically studied the effect of copayment changes on use of office visits (Cherkin et al. (1989)), emergency room use (Selby

et al. (1996)), and prescription drug use (Gaynor et al. (2007)), finding modest effects. Relative to RAND and the Oregon experiment, these papers focus in on specific service and smaller scale price changes on the range of $5-$20 copayment changes for specific services. Chandra et al. (2008) provides a comprehensive summary of prior studies that study smaller scale natural experiments to study price responsiveness for specific services and types of consumers as well as prior observational work that correlates coverage and utilization.

A recent series of papers uses structural techniques and more nuanced identification strategies to examine consumer price responsiveness. Einav et al. (2013a) study consumer price responsiveness in Medicare Part D prescription drug insurance using local variation in utilization around kink points in non-linear contracts and within-year variation in the time consumers arrive to calender year contracts.[4]. Kowalski (2013) finds high consumer price responsiveness for a large employer-insured sample of consumers, using identifying variation from shocks to family members' health for families with multiple members covered in one non-linear insurance contract. While these papers are able to identify price responsiveness for consumers near kink points in non-linear contracts or for larger families with certain types of contracts, they require structural assumptions to extrapolate those results to their full samples and deal with selection issues related to their identification strategies.

There are also papers that specifically focus on studying high-deductible health plans since the Medicare Modernization Act of 2003. Buntin et al. (2011) study data on health insurance choice and utilization from 53 employers, 28 of whom offered an HDHP. They focus on utilization for people in firms offering HDHPs, choosing HDHPs, as compared to similar individuals (based on propensity score matching) in firms who did not offer HDHPs. While only 5% of consumers in sample choose an HDHP, implying potentially substantial selection issues for the analysis within and across firms, this study is one of the first to carefully investigate take-up and utilization of HDHPs in employers. Sood et al. (2013), Haviland et al. (2011b), and Haviland et al. (2011a) study some specific aspects of HDHP offering and takeup on utilization. Lo Sasso et al. (2010) study the impact of different employer account contributions for utilization in HDHPs.

Overall, our study is (i) able to leverage exogenous variation in insurance offerings for a large sample of consumers to overcome selection concerns and (ii) use detailed data on consumer health and demographics together with new techniques to model heterogeneous price responses. Modeling this heterogeneity in depth is essential for questions of optimal insurance design, both in terms of the optimal design of payments within one insurance contract offered and in terms of what insurance contracts should be included in a menu offering for a given population, both questions that we investigate here.

The rest of the paper proceeds as follows. Section 2 describes our empirical setting and the data we use to conduct our analysis. Section 3 presents our primary descriptive results for consumer price responsiveness in the context of the forced switch to the high-deductible health plan from free health care. Section 4 describes our different models designed to measure heterogeneity in price responsiveness across types of medical services and types of consumers. Section 5 presents

[4]Kowalski (2012) and Marsh (2011) also study price responsiveness using local variation around kink points in non-linear contracts

our analysis of the welfare impact of the forced HDHP switch and several optimal insurance design counterfactuals. Section 6 concludes.

## 2 Data and Setting

**Administrative Data.** We analyze administrative data for a large self-insured firm with approximately 56,000 U.S. employees (as of 2013), covering approximately 160,000 lives. We observe these employees over the time period from 2009 to 2013. Our dataset includes three major components. First, we observe each individual's enrollment in a health insurance plan for each month over the course of these years, including their choice of plan and level of coverage. Second, we observe the universe of line-item health care claims incurred by all employees and their dependents, including the total payment made both by the insurer and the employee, and detailed codes indicating the diagnosis, procedure, and service location associated with the claim. Finally, we observe rich demographic data, encompassing not only standard demographics such as age and gender, but also detailed job characteristics and income, as well as the employee's participation in and contributions to health savings account (HSA), flexible spending account (FSA), and 401k savings vehicles. These data are similar in content to other detailed data sets used recently in the health insurance literature, such as those in, e.g., Einav et al. (2013b), Handel (2013), or Carlin and Town (2009). The data we use here have a particular advantage for studying moral hazard in health care utilization due to a policy change that occurred during our sample period, which we discuss in detail shortly.

The first column of Table 1 presents summary statistics for the entire sample of $52,445$ employees (and $147,388$ dependents) present at the firm in 2012.[5] The employee population at this firm is heavily male (74.6%), young (13.1% $\leq$ 29 years), and high income (53.2% $\geq$ \$125,000 per year) relative to the general population. 23.5% of employees have insurance only covering themselves, with 20.0% covering one dependent and 56.5% covering two or more. Mean total medical expenditures (including payments by the insurer and the employee) for an individual in the plan (an employee or their dependent) were \$5,120 in 2012. While the sample of employees and dependents is not externally valid for any specific policy-relevant population, it is partially representative of other large firms nationwide, many of which are in the process of transitioning their health benefits programs (see Towers Watson (2014)). Moreover, given the high income of employees at the firm, it is quite likely that our results can be interpreted as lower bounds on the utilization impact of cost sharing relative to a lower income population.

**Policy Change.** From 2009 on, employees at the firm had two major insurance options.[6] Table

[5] These numbers only count employees enrolled in the PPO or HDHP insurance plans, the primary options for employees in 2012. It does not include employees enrolled in the HMO option available to some employees, for whom we do not observe medical claims. It also does not include employees who otherwise did not have access to the same menu of plans (e.g., because they were part-time employees). The number of employees for whom this is true is roughly 3,000, and is stable over time.

[6] Depending on the location of the office within the U.S., a subset of employees could also choose a Health Maintenance Organization (HMO) option. Since fewer than 5% of employees in the relevant locations choose this option (remaining steady over time) we exclude those who choose the HMO from our analysis.

| **Sample Demographics** | PPO or HDHP in 2012 | PPO in 2012 | Sample in 2012 |
|---|---|---|---|
| **N** - Employees | 52,445 | 44,711 | 31,293 |
| **N** - Emp. & Dep. | 147,388 | 129,183 | 97,022 |
| Enrollment in PPO in 2012 | 85.21% | 100% | 100% |
| **Gender** - Employees (% Male) | 76.6% | 76.0% | 76.8% |
| **Gender** - Emp. & Dep. | 51.9% | 51.5% | 51.2% |
| **Age** - Employees | | | |
| 18-29 | 12.0% | 10.3% | 5.6% |
| 30-54 | 83.2% | 84.8% | 89.1% |
| $\geq 55$ | 4.8% | 4.9% | 5.3% |
| **Age** - Emp.& Dep. | | | |
| $< 18$ | 34.5% | 35.3% | 36.8% |
| 18-29 | 12.3% | 11.5% | 9.8% |
| 30-54 | 50.1% | 50.1% | 50.1% |
| $\geq 55$ | 3.1% | 3.1% | 3.3% |
| **Income** | | | |
| Tier 1 ($<$ \$75K) | 1.8% | 1.8% | 2.0% |
| Tier 2 (\$75K-\$100K) | 6.6% | 6.4% | 6.2% |
| Tier 3 (\$100K-\$125K) | 30.1% | 29.8% | 30.5% |
| Tier 4 (\$125K-\$150K) | 34.9% | 35.1% | 35.3% |
| Tier 5 (\$150K-\$175K) | 15.5% | 15.6% | 15.2% |
| Tier 6 (\$175K-\$200K) | 6.3% | 6.4% | 6.2% |
| Tier 7 (\$200K-\$225K) | 2.8% | 2.8% | 2.7% |
| Tier 8 (\$225K-\$250K) | 1.0% | 1.0% | 0.9% |
| Tier 9 ($>$ \$250K) | 1.1% | 1.1% | 1.0% |
| **Family Size** | | | |
| 1 | 23.7% | 21.4% | 17.8% |
| 2 | 19.6% | 19.1% | 18.3% |
| 3+ | 56.7% | 59.5% | 64.0% |
| **Individual Spending, 2012** | | | |
| Mean | \$5,020 | \$5,401 | \$5,811 |
| 25th Percentile | \$609 | \$687 | \$722 |
| Median | \$1,678 | \$1,869 | \$1,978 |
| 75th Percentile | \$4,601 | \$5,036 | \$5,219 |
| 95th Percentile | \$18,256 | \$19,367 | \$20,201 |
| 99th Percentile | \$49,803 | \$52,872 | \$56,624 |

Table 1: This table presents summary demographic statistics for (i) employees enrolled in the PPO or HDHP plan options at the firm in 2012; (ii) employees enrolled in the PPO plan option at the firm in 2012; and (iii) our final sample, which is restricted to employees present over the time horizon 2009-2013, and their dependents. This sample is described below.

| **Health Plan Characteristics**<br>**Family Tier** | PPO | HDHP |
|---|---|---|
| Premium | $0 | $0 |
| Health Savings Account (HSA) | No | Yes |
| HSA Subsidy | - | $3,750* |
| Max. HSA Contribution | - | $6,250** |
| Deductible | $0*** | $3,750* |
| Coinsurance (IN) | 0% | 10% |
| Coinsurance (OUT) | 20% | 30% |
| Out-of-Pocket Max. | $0*** | $6,250* |

*Values for family coverage tier (2+ dependents). Single employees (those with one dependent) have .4x (.8x) the values given in this table.
**Single employees have a legal maximum contribution of $3,100. Employees over 55 can contribute an extra $1,000 in 'catch-up' contribution. This includes the employer subsidy.
***For out-of-network spending, the PPO has a deductible of $100 per person (up to $300) and an out-of-pocket maximum of $400 per person (up to $1200).

Table 2: This table presents key characteristics of the two primary plans offered at the firm we study. The PPO option has more comprehensive risk coverage while the HDHP option gives a lump sum payment to employees up front but has a lower degree of risk protection. The numbers in the main table are presented for the family tier (the majority of employees) though we also note the levels for single employees and couples below the main table. Both plan options were present at the firm from 2009-2012, but the PPO option was removed in 2013, forcing employees to join the HDHP in that year.

2 lists features of the two plans, side by side. The first was a popular broad network PPO plan with unusually generous first-dollar coverage. This plan had no up front premium and no employee cost-sharing for in-network medical services. The second primary option was a high-deductible health plan (HDHP) with the same broad network of providers as the PPO. Enrollees in this plan face cost-sharing for medical expenditures, with a deductible, coinsurance arm, and out-of-pocket maximum typical of more generous high-deductible health plans (in 2013, this plan paid 78.1% of ex post total medical expenditures). Despite this higher cost sharing, this plan was potentially attractive relative to the PPO option because it offered a substantial subsidy to enrollees that was directly deposited into their health savings account that was directly linked to the HDHP. As shown in table 1, in 2012 85.2% of employees chose the PPO with the remainder choosing the HDHP. Consumer choice between these two plan options is discussed in depth in **?**. For this paper, it is only important to note that the large majority of employees were enrolled in the PPO prior to the forced plan switch that occurred at the firm for 2013.

| **Policy Change: Price Impact 2012 Total Spending** | | | | | |
|---|---|---|---|---|---|
| Coverage Tier | Avg. HDHP Price | % Under Deductible | % Over Ded., Under OOP Max. | % Over OOP Max. | Actuarial Value |
| 0 Dependents | 0.428 | 37.92% (< $1,500) | 49.16% ($1,500 - $11,500) | 12.92% (> $11,500) | 78.31% |
| 1 Dependent | 0.293 | 23.22% (< $3,000) | 61.08% ($3,000 - $23,000) | 15.70% (> $23,000) | 76.59% |
| 2+ Dependents | 0.201 | 13.30% (< $3,750) | 68.40% ($3,750 - $28,750) | 18.30% (> $28,750) | 78.24% |
| All Tiers | 0.249 | 18.42% | 64.46% | 17.12% | 78.05% |

Table 3: This table presents results from the simple counterfactual exercise discussed in the text. For the employees in our primary sample, we take their 2012 health care spending and calculate the amount that they would have paid out-of-pocket if they spent the same amount while enrolled in the HDHP. We present the average % of total spending paid, as well as the likelihood of reaching each arm of the non-linear HDHP contract. Below each percentage is the range of allowed expenditures required to be in that arm of the insurance plan for that tier of coverage, if the employee only received care in-network (which is typical for most employees).

In October 2010, the firm announced to its employees that it would discontinue the PPO option as of 2013. This forced those employees still enrolled in the PPO in 2012 to switch to the HDHP option in 2013. For these employees, this policy change represented a substantial and exogenous change to the marginal prices they faced for health care services. Moreover, because of the PPO plan structure, the employees that were forced to switch into the HDHP had a zero marginal price for medical care prior to the switch, implying that we observe true cost-free demand for health care services as our baseline.

Table 3 presents evidence on the change in cost-sharing faced by the $31,293$ employees in our primary sample (described below) forced to move into the HDHP in 2013. We take the spending of all employees enrolled in the PPO in 2012, and assume that they had instead been enrolled in the HDHP in that year. We then determine what arm of the plan they would have ended up in. This simple counterfactual is intended to illustrate the price change: the statistics presented here will change somewhat once consumer price elasticity of demand for health care services is accounted for.[7] Employees and dependents paid 0% of all in-network expenses under the PPO, while under the HDHP, the average consumer would have paid for 24.9% of these total expenses. Table 3 breaks down the change in consumer prices by coverage tier, and illustrates the end-of-year marginal price that they face by showing which arm of the non-linear contract they would have reached by the end of the year. 18.42% of employees would have been under the HDHP deductible in 2012, 64.46% would have passed the deductible but not reached the out-of-pocket maximum, and 17.12% would have reached the out-of-pocket maximum. Those not passing the deductible would have faced the

[7]Throughout the paper, our analysis takes into account the fact that preventive services are always free under the HDHP. Such spending accounts for 8.87% of total spending.

full marginal price of care at the end of the year, those who passed the deductible but not the out-of-pocket maximum a marginal price of 10%, and those who passed the out-of-pocket maximum a marginal price of zero. This simple evidence illustrates the substantial average and marginal price changes for employees from 2012 to 2013 due to the firm's insurance benefits redesign.[8] Finally, we note that the actuarial value of the HDHP is 78.05% in 2013, implying that the plan pays for 78.05% of total medical spending at the firm in that year, with employees paying the remainder.[9]

**Primary Sample.** For the majority of our forthcoming analysis, we use the sample of employees who (i) were present at the firm for the whole sample period, 2009-2013; and (ii) were enrolled in the PPO prior to the forced switch in 2012. We use this sample to ensure that we have a substantial time series of information on the health status of employees we analyze. We additionally use these employees, as early switchers are selected on youth, relative health, and information about health insurance (see Handel and Kolstad (2014))

Column 3 of table 1 shows the summary statistics for this primary sample, which can be compared to the full sample of employees present in 2012 presented in Column 1. There are 31,293 employees in the primary sample covering 97,022 dependents. This is approximately 55% of employees and dependents present in any given year at the firm. Relative to our larger sample (in Column 1), these employees are somewhat older (only 5.6% $\leq$ 29) and somewhat more wealthy (61.3 $\geq$ \$125,000 per year), but differ only slightly on gender and health spending. In Appendix 6 we present summary statistics and some of our core results for alternative samples, including those that restrict the sample to employees present from 2010 and 2011 onwards (instead of 2009). Our main results are essentially unchanged for these alternative samples. In addition, in that appendix we present analysis on whether or not there is substantial attrition from the firm after the announcement of the switch to the HDHP (in late 2010) or after the actual forced switch to that plan in 2013. This analysis shows that (i) employee entry and exit patterns look simple in each year throughout the sample period and (ii) there is limited to no attrition of dependents (or existing employees) from health insurance at the firm.

# 3 Evidence: Impact of Increased Cost Sharing

**Policy Impact: Change in Total Spending.** Our descriptive analysis investigates the change in total health care spending resulting for the forced change of consumers to the HDHP in 2013. We present a series of analyses, beginning with a description of the raw data and ending with a complete analysis that is intended to reflect a causal impact of the contract change.

Figure 1 plots mean monthly spending at the individual level for our primary sample over the time period 2009-2013. The vertical line in the figure represents the beginning of 2013, the time when the policy change took effect and all individuals in the primary sample switched from the free

[8] We note that, with reductions in total medical expenditures in the HDHP due to a positive price elasticity of demand, the marginal prices consumers would have faced in 2012 would be larger than the numbers given here.

[9] At the firm overall, the % of claims paid by employees is lower than the average of cost sharing across employees, because the former is weighted by total spending per individual while the latter is not.

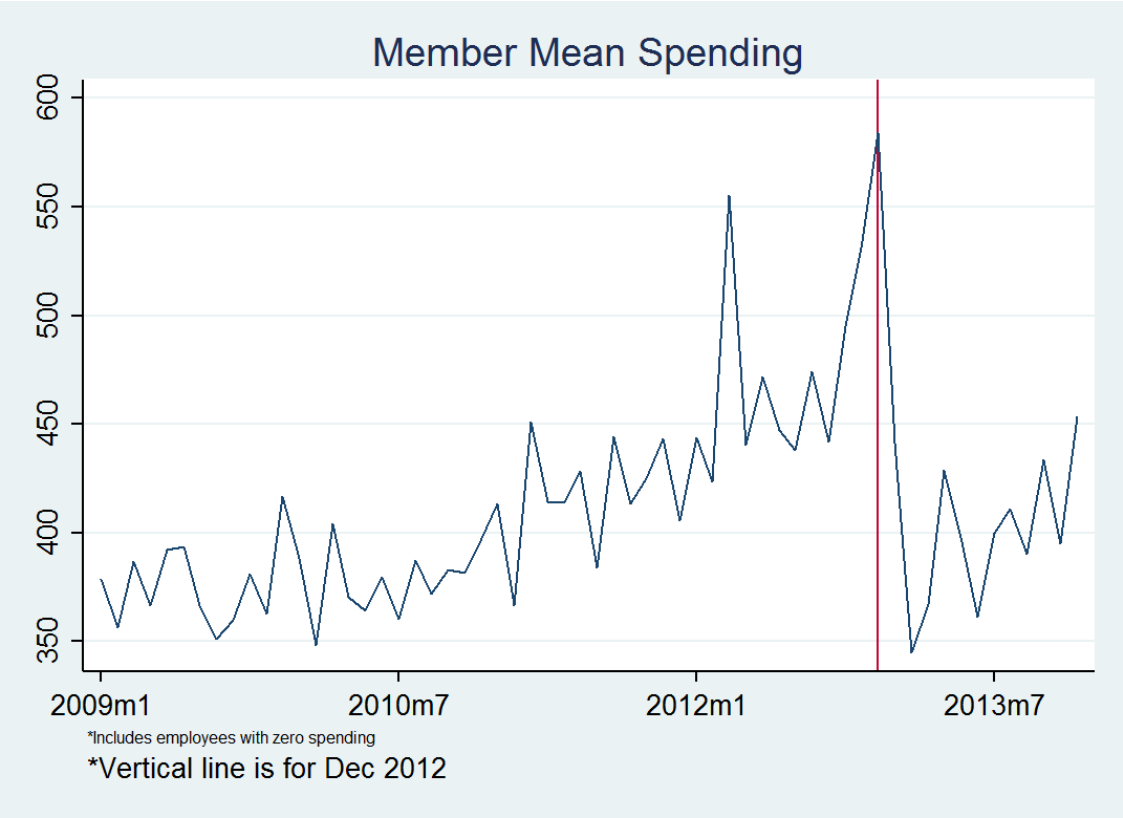



Figure 1: This figure plots mean monthly spending by individuals in our sample over the time period we observe, for our primary sample.

PPO to the HDHP. The figure clearly illustrates that spending drops after the forced switch to the HDHP: the average yearly spending for an individual dropped from $5714.63 in 2012 to $4575.17 in 2013. This constituted a year on year 19.61% drop in spending in the raw data, effectively returning nominal spending to just below 2009 spending levels for this sample. The figure also reveals that there was a substantial increase in spending at the end of 2012, indicating excess spending in that month due to the impending increase in the price of health care. We formally account for this inter-temporal substitution of expenses later in this section. Finally, the figure illustrates that the drop in spending is persistent on a monthly basis over the course of 2013. Table 4 supplements Figure 1 and presents average monthly individual spending rates for our primary sample for select months and each year in this five year timespan.

**Age and Inflation Adjustments.** The raw spending data just presented increases steadily over time. We attribute this to two factors in our environment. The first is that our sample is a fixed panel, and thus is aging across the five years. The second is that the price of care has naturally risen over time, due to inflation and other factors such as the introduction of new medical technologies. Both of these factors will inflate spending in later periods of our data relative to earlier periods.

To adjust spending for age, we take monthly individual-level spending in January 2009 and regress it on age. The resulting coefficient, which equals roughly $7.50 per year, provides an

| **Average Individual Spending**<br>**Select Months: 2009-2013** | March | June | September | December |
|---|---|---|---|---|
| 2009 | 397.88 | 400.39 | 365.17 | 419.11 |
| 2010 | 441.24 | 393.72 | 406.18 | 415.99 |
| 2011 | 475.41 | 453.67 | 423.39 | 414.02 |
| 2012 | 569.63 | 456.52 | 449.43 | 601.22 |
| 2013 | 382.60 | 369.43 | 404.84 | 445.17 |

Table 4: This table presents average individual monthly spending for select months over the time period 2009-2013, for our primary sample. It shows a clear drop in spending in 2013, the year everyone in this sample was forced to switch to the HDHP.

estimate of the increase in spending that comes about from aging one year in our sample.[10] We divide this coefficient by 12 to get the monthly effect of aging on spending, and adjust spending over time by this factor such that all spending is in January 2009 dollars. Figure A11 in Appendix 6 reproduces the plot of monthly raw spending from Figure 1, and compares it to this age-adjusted spending, which is very similar.

To adjust for inflation in medical spending we use the Consumer Price Index (CPI) for medical care for each month.[11] We divide spending for a given month by a factor derived from this CPI that reflects overall growth in medical spending since January 2009, thus converting spending in each month into January 2009 dollars. The resulting graph, combining this inflation adjustment with the prior age adjustment, is given in Figure 2. These two adjustments are intended to remove the time trend in spending from the raw spending graphs and clearly illustrate the impact of the forced HDHP switch in January 2009 age and medical spending dollars. The figure clearly illustrates the drop in average monthly individual spending following the forced HDHP switch. As presented later in table 5 once these adjustments are accounted for average individual monthly spending drops by 22.90% from 2012 to 2013 when consumers are forced to switch to the HDHP.

One natural worry in our setting is that our sample may experience different medical price inflation over time relative to the nation at large. This could be either because of differential price inflation and medical service growth by region, or because our insurer negotiated prices in a specific way over time during our sample period from 2009-2013. We perform a series of analyses on the price changes in our environment for medical services over time to ensure that our assumptions about medical price inflation are sensible. We investigate both (i) how prices for a broad basket of services change over the course of our sample and (ii) make sure that there is not substantial heterogeneity in the price changes across medical services and providers. We present these analysis

[10] One would normally expect a nonlinear relationship between age and health spending that is flatter at younger ages and steeper at older ages. The relative youthfulness of our sample is a key reason for the low estimated impact of aging here. Using a nonlinear adjustment or estimating on a different time period gives similar results.

[11] This comes from the index collected by the Bureau of Labor Statistics. A time series of this index can be found at `http://research.stlouisfed.org/fred2/series/CPIMEDNS`. A description on how this is collected can be found at `http://www.bls.gov/cpi/cpifact4.htm`. It does not directly account for technological change, but provides a time series of price changes which match that found in our data.

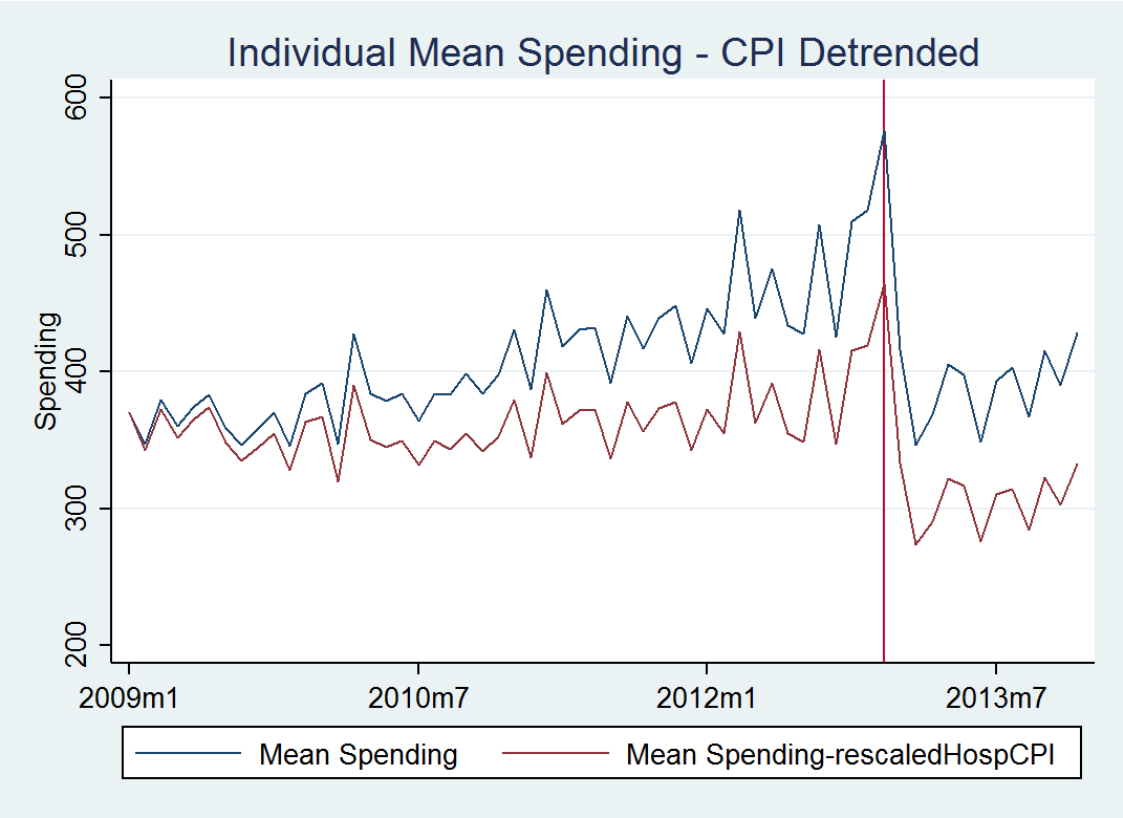



Figure 2: This figure plots spending for families in a given month, adjusted and unadjusted for age and price trends.

in full in Appendix 6. The results are reassuring: on average the price for the broad basket of medical services increased by 2% over the time period 2009-2013 in our sample (close to the CPI rate) while the heterogeneity in price changes was limited and normally distributed, indicating no concern about systematic negotiated price changes for the time period surrounding the forced switch to the HDHP in 2013.

**Intertemporal Substitution.** One apparent feature of the age and inflation adjusted spending over time, shown in Figure 2, is the larger than normal spending at the end of the year in 2012, just before employees were forced to switch from the PPO to the HDHP. This increase in spending likely reflects anticipatory spending, whereby consumers shift services whose timing is flexible from 2013, where they will be subject to cost sharing, to the end of 2012, when they are still free. Since the forced change to the HDHP was announced well in advance, it is clear that consumers would have known about the forced change in advance and could have shifted spending accordingly. This kind of intertemporal substitution of claims is documented carefully by Cabral (2013) in the context of dental insurance and dental care, though, to our knowledge, such anticipatory spending has not yet been carefully documented in the domain of health insurance.

Despite the fact that the plan change was announced well in advance, spending increases over time at what looks like a regular rate, until November and December 2012, when it is much higher

than expected. This spike is particularly large relative to baseline predicted spending, taking into account the fact that November and December are historically lower utilization months at the firm (likely due to winter holidays). We calculate the extent of this spike by comparing actual spending in those months against trend. To this end, we estimate the following regression:

$$\bar{y_t} = \beta t + \lambda_t + \bar{\epsilon}_t$$

where $\bar{y}$ is the average individual-level spending at the firm in a given month, $\beta$ is a linear time trend, $\lambda$ represents monthly fixed effects, and $\bar{\epsilon}$ is the average of idiosyncratic individual health shocks at the firm for the given month.[12] We estimate this regression on the time period from January 2009 to December 2011, to ensure we do not include observations where substitution from 2013 is likely.

We calculate the 'excess mass' induced by intertemporal substitution by calculating the difference between these two lines for the months November and December 2012. This mass is calculated at \$61.96 for November 2012 and \$105.42 for December 2012. This measure of the mass is a noisy one, because it incorporates both the effect of intertemporal substitution and idiosyncratic aggregate shocks to health spending. The latter can be quite large in magnitude, as indicated by the visible month-to-month variation. The mean squared error of the regression model is a consistent estimator for the variance of these shocks. Using this, and assuming no autocorrelation between shocks, we can construct a 95% confidence interval on excess mass for this two month period at the end of 2012, equal to $\pm 2\sqrt{2}\times$ the standard deviation of $\bar{\epsilon}$. The standard error term is 44.82, giving a 95% confidence interval for the two month excess mass of [\$122.57,\$212.21] per individual.

While understanding intertemporal substitution in spending is interesting in and of itself (and we investigate how different claims are impacted by this later) we are primarily interested in quantifying this so that we can quantify the causal impact of the forced HDHP switch on total spending. If we don't formally account for this intertemporal substitution, our estimates of the HDHP impact will overstate the impact of increased cost sharing on total spending, because 2012 spending will seem higher (and 2013 spending lower) due to this substitution. To this end, we must maintain several assumptions. First, we must maintain assumptions about the months that substitution occurs to in 2012, and away from in 2013. Second, we must maintain an assumption on whether the excess mass entirely represents claims that would have occurred in 2013, or whether some of that mass represents claims that would not have occurred under the new, higher prices.

For the time frame of substitution, the excess mass calculations by month in 2012 (in the context of their standard errors) strongly suggests that November and December 2012 are the months when consumers substitute claims to (or increase claims for). For 2013, it is less clear what months claims are substituted away from. This is in part because we only have a single year of data after the change, in comparison to the four years of data we have before the change. However, an examination of the type of claims consumer substitute to 2012 suggests that such substitution occurs relative

[12] We use average spending at the firm because central limit theorem-like properties should guarantee that $\bar{\epsilon}$ is roughly normal, whereas $\epsilon$ alone would not be, as individual health spending is strongly right-skewed.

to a shorter time horizon. As shown later in this section in our heterogeneity analysis, when we recreate our excess mass analysis for different service types (such as prescription drugs, outpatient procedures, etc.), it suggests that there is a far stronger effect of intertemporal substitution on prescription drugs and office visits than other types of services. Since these are more likely to be services consumed over a shorter time frame (rather than expensive one-off elective services) we focus our analysis of substitution on the first few months after the policy change (we present additional robustness checks in 6).

For the proportion of excess mass that would have been consumed, we investigate several approaches. For our primary approach, we maintain the assumption that all excess mass in 2012 would have been consumed in 2013. This approach places a lower bound on the absolute impact of the policy change on spending. We maintain the assumption that this excess mass was substituted away from January through June 2013 only, and assign it evenly across these six months. In Appendix 6, we investigate other assumptions, such as spreading the excess mass over three months, and spreading that excess mass in a declining manner from January 2013 on. For our overall analysis of the impact of the policy change on spending, this assumption about the months being substituted from do not affect our conclusions over the assumption that all substitution comes from 2013. In Appendix 6 we investigate two additional methods for assigning excess mass to 2013. The first is the same as our primary method, but assumes only 80% of excess mass would have actually been spent in 2012 (roughly reflecting our treatment effect estimates of the HDHP in the raw data). The second estimates a trend in the post-policy period in 2013 from March through October. This allows us to predict January and February 2013 spending based on that trend, and compute the 'missing mass' in January and February relative to those predictions. These approaches yield slightly different results by design, but have a minor impact on our main results / conclusions.

Figure 3 graphs average monthly spending over time given the above corrections for intertemporal substitution. It removes predicted excess mass from November and December 2012, and assigns it evenly to the first six months of 2013.. The figure presents the 95% confidence interval for spending in November and December 2012, since our excess mass estimates occur relative to mean predicted spending in those months, as well as for the first six months of 2013, which reflects our uncertainty about the missing spending in 2013 that comes from our uncertainty about the 'excess mass' in 2012. Figure 4 compares monthly average individual spending over time with and without adjustments for intertemporal substitution to illustrate the impact of those adjustments.

As presented in Table 5, once these intertemporal substitution adjustments are accounted for, the impact of the forced HDHP switch on spending is -16.45%, implying this adjustment reduces the treatment effect estimate by approximately 6 percentage points (approximately 26% of total effect). Relative to the analysis that just adjusts for aging and medical cost inflation, the treatment effect accounting for intertemporal substitution is a better reflection of the causal impact of the forced HDHP switch on total spending.

**Treatment Effect Estimates.** Table 5 gives an overview of our main reduced form results for the impact of the forced switch of consumers to the HDHP on total medical spending. Each

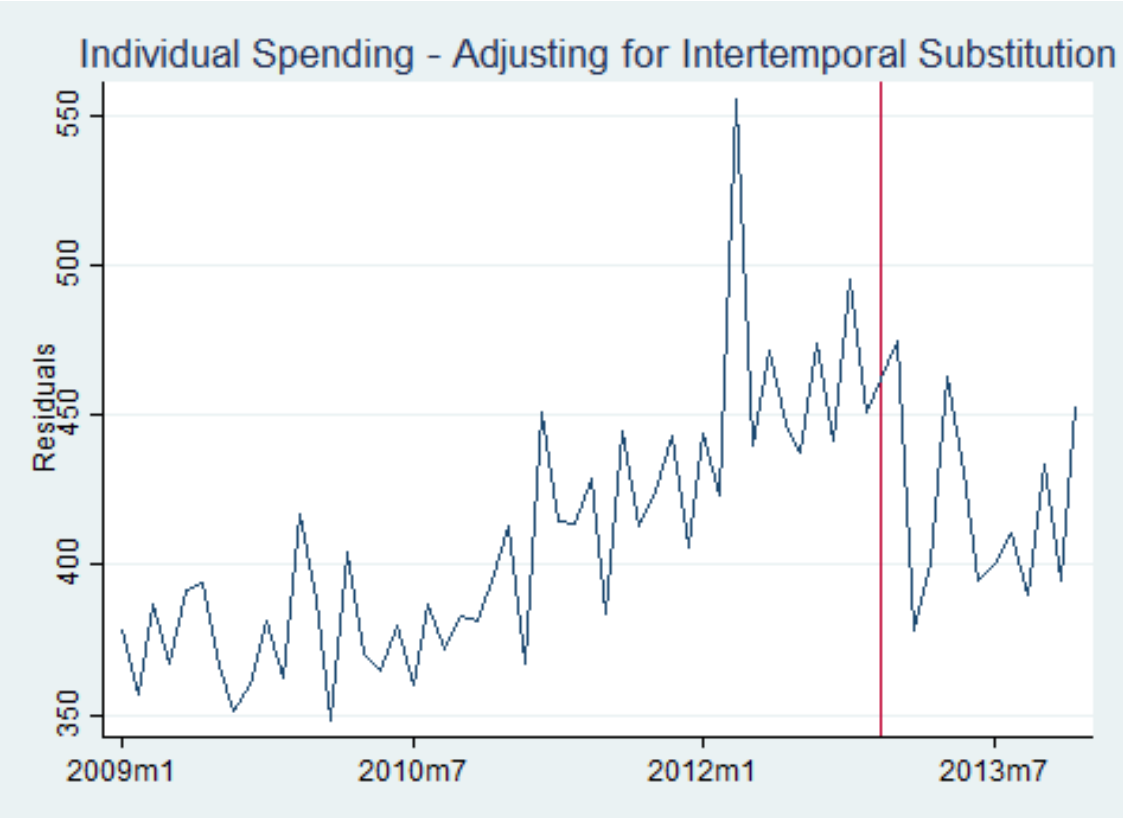



Figure 3: This figure plots average individual spending for across our sample period, with relevant months adjusted for the presence of intertemporal substitution. The figure presents the 95% CI for spending in November and December 2012, since our excess mass estimates occur relative to mean predicted spending in those months.

column in the table represents a different approach to estimating this treatment effect, corresponding to different sets of adjustments described above, as well as to some regression models described momentarily. Column (1) presents the actual nominal yearly spending in our sample, column (2) presents the same, with the aging and CPI adjustments, and column (3) presents the same model as column (2) plus the intertemporal substitution adjustments described above. The rows of the table describe predicted average yearly spending per individual under each framework, as well as the associated percentage change in spending from 2012 to 2013, which are our main treatment effect estimates of the HDHP switch on spending.

The results from these first three columns, which summarize the approaches described thus far, indicate a real and substantial drop in medical spending as a result of the firm forcing all employees to join the HDHP in 2013. The nominal spending data indicates a 19.61% reduction in yearly total medical spending, while this number is 22.90% once we adjust for aging and medical inflation. Finally, once we model and quantify intertemporal substitution, the effect becomes 16.45%, which should be interpreted as our primary treatment effect estimate for analyses relying primarily on the data alone. As shown in Table 5, a back of the envelope calculation that multiplies total spending at the firm (roughly $750 million) by this treatment effect indicates that the firm reduced total

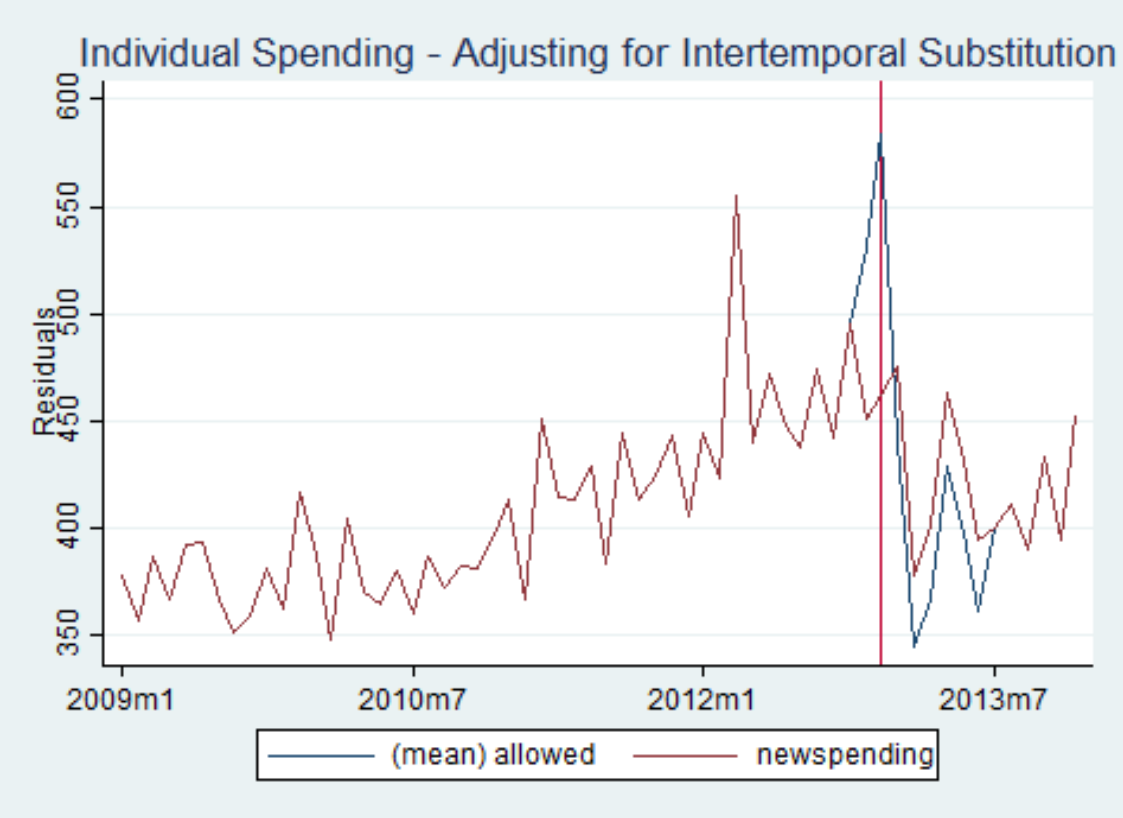



Figure 4: This figure compares monthly average individual spending over time with and without adjustments for intertemporal substitution.

| **HDHP Switch Spending Impact** | | | | Model | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4a) | (4b) | (4c) |
| | – | CPI | Intertemp. | Regression | Regression | Regression |
| Year | | | Substitution | (1 stage,logs) | (2 stage, logs) | (2 stage, levels) |
| 2009 | 4372.54 | 4244.68 | 4244.68 | 2011.67 | 2234.22 | 4464.82 |
| 2010 | 4709.95 | 4273.05 | 4273.05 | 1982.54 | 2208.80 | 4428.72 |
| 2011 | 5159.41 | 4434.72 | 4434.72 | 1953.37 | 2182.02 | 4391.76 |
| 2012 | 5811.48 | 4764.97 | 4597.58 | 1970.46 | 2185.85 | 4402.56 |
| 2013 | 4671.73 | 3673.75 | 3841.14 | 1242.87 | 1500.18 | 3580.19 |
| % Decrease, 2012-2013 | -19.61% | -22.90% | -16.45% | -36.92% | -31.37% | -18.68% |
| $ Impact (million) | -$147.09 | -$171.76 | -$123.40 | -$276.94 | -$235.27 | -$140.09 |

Table 5: This table details the treatment effect of the forced HDHP switch under different frameworks (the raw data is presented in column (1)). Under each framework we display the predicted values for mean yearly individual spending, for each year as well as the predicted % change in this spending as a result of the forced HDHP switch over the period 2012-2013.

medical spending by $123.39 million dollars per year by forcing this switch to the HDHP.

A typical metric used to compare price sensitivity estimates in medical spending is the arc elasticity, which is given by:

$$\frac{(q_2 - q_1)/(q_2 + q_1)}{(p_2 - p_1)/(p_2 + p_1)}$$

Here, $q_2$ can be thought of as the quantity of medical spending in plan 2, given price level $p_2$, while $q_1$ and $p_1$ are the analogous quantities for a second health plan, plan 1. One attractive feature of arc elasticity as a measure of price sensitivity is that it only depends on these prices and quantities, and not the direction of which plan is being switched to empirically.[13] While we can compute an arc elasticity in our setting, the computation is not meaningful because the price of the PPO is 0, implying that the arc elasticity calculation in our setting would be $\frac{q_2-q_1}{q_1+q_2}$. Thus, given a price of 0 for one plan, the arc elasticity just tells us the change in quantity for any size price change, which is not a constructive measure of price sensitivity in any sense.

As a result, we measure price sensitivity and compare our results to the literature using a *semi-arc elasticity* metric, defined as:

$$\frac{(q_2 - q_1)/(q_2 + q_1)}{(p_2 - p_1)}$$

This measure divides the change in quantity, normalized by average baseline quantity, divided by the actual change in price. In general, as with the arc-elasticity measure, one might want to normalize the price change as well to reflect differences in scale (e.g. comparing changes of $5 to $10 versus $5000 to $10000). In our setting, this is not an issue because we define price as the share of firm-wide costs that fall on the employee, following past work on moral hazard (see e.g. Manning et al. (1987)). Since this percentage is a relative measure already, this scaling issue does not arise when using the semi-arc elasticity measure. We follow convention in the literature and define the 'price' for each plan as their respective actuarial values subtracted from 1, equal to the percentage of total spending that consumers pay out-of-pocket, if all consumers were enrolled in one plan. We calculate this actuarial value independently of ex-post moral hazard, so we calculate this value under 2012 spending levels, while all consumers were enrolled in the PPO. For this semi-arc elasticity calculation, the PPO price is $P_{PPO} = 0$ while the HDHP price is $P_{HDHP} = .219$.[14] We use these prices, together with the quantity changes described in table 5 to compute the semi arc-elasticity for the forced HDHP switch, which we report in table 6. For comparison purposes

[13] The definition for the standard notion of point price elasticity is $\frac{\Delta Q/q_1}{\Delta p/p_1}$. This depends on which plan is the initial plan, and so is not as attractive as the arc-elasticity measure described in the text. In addition, when the price of the PPO is 0, this regular price elasticity cannot be calculated. Moreover, it induces two different possible measured elasticities from the same treatment effect.

[14] As noted in Einav et al. (2013a), the notion that a health plan can be described by one price for a population of heterogeneous consumers represents a substantial simplification of the prices consumers actually face / perceive they face. First, the true rational price faced by consumers is their expected end-of-year marginal price (which actuarial value does not represent in any specific way). Second consumers may emphasize their spot price in decision-making instead of this expected end-of-year marginal price. As a result, these elasticity estimates are presented for comparison purposes, and not to represent underlying price sensitivity in any structural manner.

| **HDHP Switch** **RAND Comparison** | (1) – | (2) CPI | (3) Intertemp. Substitution | (4b) Regression (2 stage, levels) | (R1) RAND 0 - 25 | (R2) RAND 25 - 95 |
|---|---|---|---|---|---|---|
| % Change Spending | -19.61% | -22.90% | -16.45% | -18.68% | -29% | -29% |
| Price Change | 0 to .219 | 0 to .219 | 0 to .219 | 0 to .219 | 0 to .16 | .16 to .31 |
| Semi-Elasticity | -0.407 | -0.469 | -0.347 | -0.387 | -1.059 | -1.130 |

Table 6: This table details the semi-arc-elasticity estimates from our treatment effect estimates, as well as similar results from the RAND experiment, using percent changes in expenditures given by Keeler and Rolph (1988).

we also present the semi-arc elasticities for some of the most cited treatment effects in the RAND Health Insurance Experiment.

As Table 6 reveals, the semi-arc elasticity for the three primarily descriptive treatment effects in table 5 lie in the range -0.35 to -0.47. These are lower than those we compute based on two distinct RAND treatment effects. The first RAND comparison studies the difference in quantity consumed for a 0 cost-sharing plan (similar to our PPO option) and a plan with 25% coinsurance up until an out-of-pocket maximum of $1000 (over $6000 in 2014 dollars) which had an actuarial value of .16. The second comparison studies movement from this same 25% coinsurance plan to a plan with 95% coinsurance (almost no coverage) and the same out-of-pocket maximum (actuarial value of .31). For these RAND treatment effects, our computed semi-arc elasticities are -1.06 and -1.13 respectively, indicating that consumers are less price sensitive in our setting relative to the RAND estimates. It is important to point out that there are many potentially important differences between these two settings, and that these semi-arc elasticities are meant to be illustrative and help provide some context for our results.

**Heterogeneous Treatment Effects.** The rich claims data we observe, together with our large sample size, allow us to investigate the heterogeneous impact of the forced HDHP switch in substantial detail. We investigate these heterogeneous treatment effects across multiple dimensions of heterogeneity including (i) demographics (ii) health status and (iii) types of medical services. In this section, we replicate the treatment effect analysis from the previous sections and present these heterogeneous treatment effects using (i) the data alone (ii) age and CPI adjusted spending data and (iii) age, CPI, and intertemporal substitution adjusted spending data. For simplicity, in the figures we will present the age and CPI adjusted results, while in table 7 we present the results from all three of these approaches.

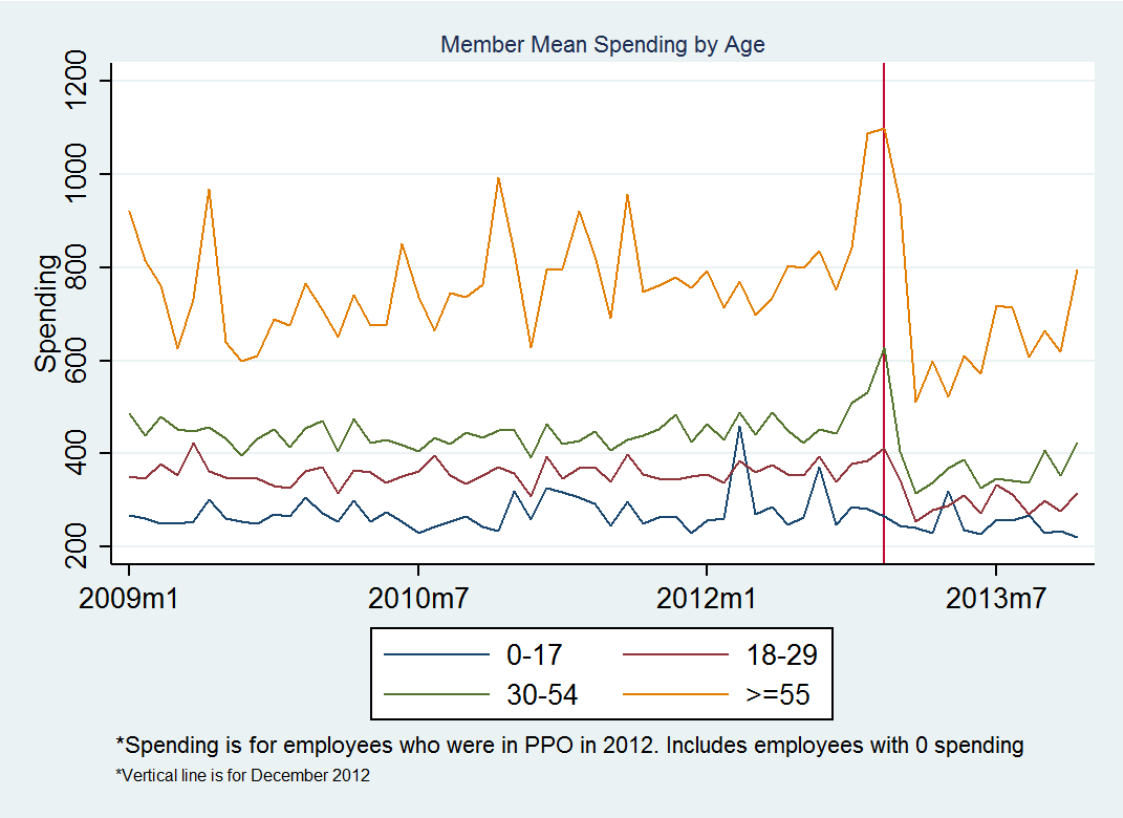



Figure 5: This figure plots adjusted spending for individuals in a given month, by age bracket.

We first assess heterogeneity as a function of consumer demographics. Figure 5 plots mean individual monthly spending for different age buckets over time, for our primary sample. For this analysis we group individuals into four age buckets (corresponding to their 2012 ages): 0-17, 18-29, 30-54, and 55+. These brackets, respectively, make up 36.84, 9.83, 50.05, and 3.28% of the sample. The figure reveals that the oldest age bracket has both the highest mean spending and the largest drop in medical spending as a result of the forced switch to the HDHP, before all adjustments are taken into account. As shown in table 7, the percent drop in spending for the oldest age bucket when adjusting for age, inflation, and intertemporal substitution is 15%. For the 0-17, 18-29, and 30-54 age buckets these drops in spending are 16%, 21%, and 18% respectively. These patterns are also reflected in the heterogeneous impact by type of enrollee (employee, spouse, or dependent) as shown in Figure A14 in appendix 6. Dependents, who are primarily in the 0-17 age range, see a relatively lower change in spending as a result of the forced HDHP switch.

We investigate the utilization impact as a function of income, and, as shown in table 7, find that price responsiveness in decreasing in income. Spending decreases by 23% (adjusting for intertemporal substitution) for employees earning between \$0-\$100,000, 15% and 18% for those earning between\$100,000-\$150,000 respectively, and only 8% for those earning over \$200,000. Figure A15 in Appendix 6 plots our event study analysis as a function of income. This result reflects the intuition that those with higher income may be less price sensitive because they have similar values

| **Heterogeneous HDHP Spending Impact** | | | | Treatment Effect | | |
|---|---|---|---|---|---|---|
| | Group % | Spending % | 2012 Mean Spending | (1) Nominal Spending | (2) CPI | (3) Intertemp. Substitution |
| **Age 0-17** | 36.84 | 27.16 | 4285.09 | -0.12 | -0.16 | -0.16 |
| **Age 18-29** | 9.83 | 7.84 | 4638.26 | -0.19 | -0.22 | -0.21 |
| **Age 30-54** | 50.05 | 57.84 | 6716.19 | -0.23 | -0.26 | -0.18 |
| **Age 55+** | 3.28 | 7.15 | 12658.71 | -0.29 | -0.32 | -0.15 |
| **Income $0-100K** | 6.92 | 7.37 | 6383.32 | -0.26 | -0.29 | -0.23 |
| **Income $100-150K** | 63.69 | 61.75 | 5815.48 | -0.19 | -0.22 | -0.15 |
| **Income $150-200K** | 23.95 | 22.09 | 5532.83 | -0.16 | -0.20 | -0.18 |
| **Income $200K+** | 5.44 | 4.99 | 5505.47 | -0.07 | -0.11 | -0.08 |
| **Employee** | 32.25 | 34.34 | 6188.31 | -0.22 | -0.25 | -0.15 |
| **Spouse** | 24.45 | 32.60 | 7750.05 | -0.23 | -0.26 | -0.21 |
| **Dependent** | 43.30 | 33.05 | 4436.21 | -0.13 | -0.17 | -0.12 |
| **2011 ACG Quartile 1** | 24.41 | 7.36 | 1862.08 | -0.07 | -0.11 | -0.13 |
| **2011 ACG Quartile 2** | 25.51 | 11.57 | 2800.47 | -0.11 | -0.14 | -0.14 |
| **2011 ACG Quartile 3** | 25.02 | 19.25 | 4752.00 | -0.23 | -0.27 | -0.15 |
| **2011 ACG Quartile 4** | 25.06 | 53.73 | 13241.58 | -0.23 | -0.26 | -0.29 |
| **2011 ACG Top 1%** | 0.94 | 10.12 | 62403.85 | -0.21 | -0.24 | -0.38 |
| **0 Chronic Cond.** | 61.74 | 35.02 | 3296.16 | -0.15 | -0.19 | -0.12 |
| **1-2 Chronic Cond.** | 33.66 | 46.60 | 8046.15 | -0.24 | -0.27 | -0.21 |
| **3+ Chronic Cond.** | 4.60 | 18.38 | 23206.57 | -0.46 | -0.48 | -0.40 |
| **Inpatient** | | 19.20 | 1115.71 | -0.14 | -0.17 | -0.09 |
| **Outpatient Hosp.** | | 17.67 | 1026.84 | -0.18 | -0.22 | -0.12 |
| **ER** | | 2.92 | 169.41 | -0.27 | -0.30 | -0.29 |
| **Office Visit** | | 7.02 | 407.99 | -0.19 | -0.22 | -0.17 |
| **RX** | | 12.25 | 712.14 | -0.22 | -0.25 | -0.22 |
| **RX-Brand** | | 11.70 | 679.94 | -0.20 | -0.23 | -0.20 |
| **RX-Generic** | | 3.87 | 224.88 | -0.19 | -0.23 | -0.21 |
| **Mental Health** | | 9.02 | 524.21 | -0.09 | -0.13 | -0.16 |
| **Preventive** | | 8.87 | 515.32 | -0.11 | -0.15 | -0.11 |
| **Other** | | 23.06 | 1339.86 | -0.31 | -0.33 | -0.24 |

Table 7: This table summarizes our descriptive evidence for the heterogeneous treatment effects of the forced HDHP switch for 2013. The table presents the results for different (i) demographics (ii) health status measures and (iii) types of health services. The first column reports the % of people within a given demographic group or health status group for categories (i) and (ii), and the % of total spending a given service spending is for category (iii). The second column reports average mean individual yearly spending for categories (i) and (ii), and average mean individual spending for each type of service for category (iii). The second through fourth columns present, for each respective framework, the % change in spending (for each demographic group, or type of service) as a result of the forced HDHP switch over the period 2012-2013.

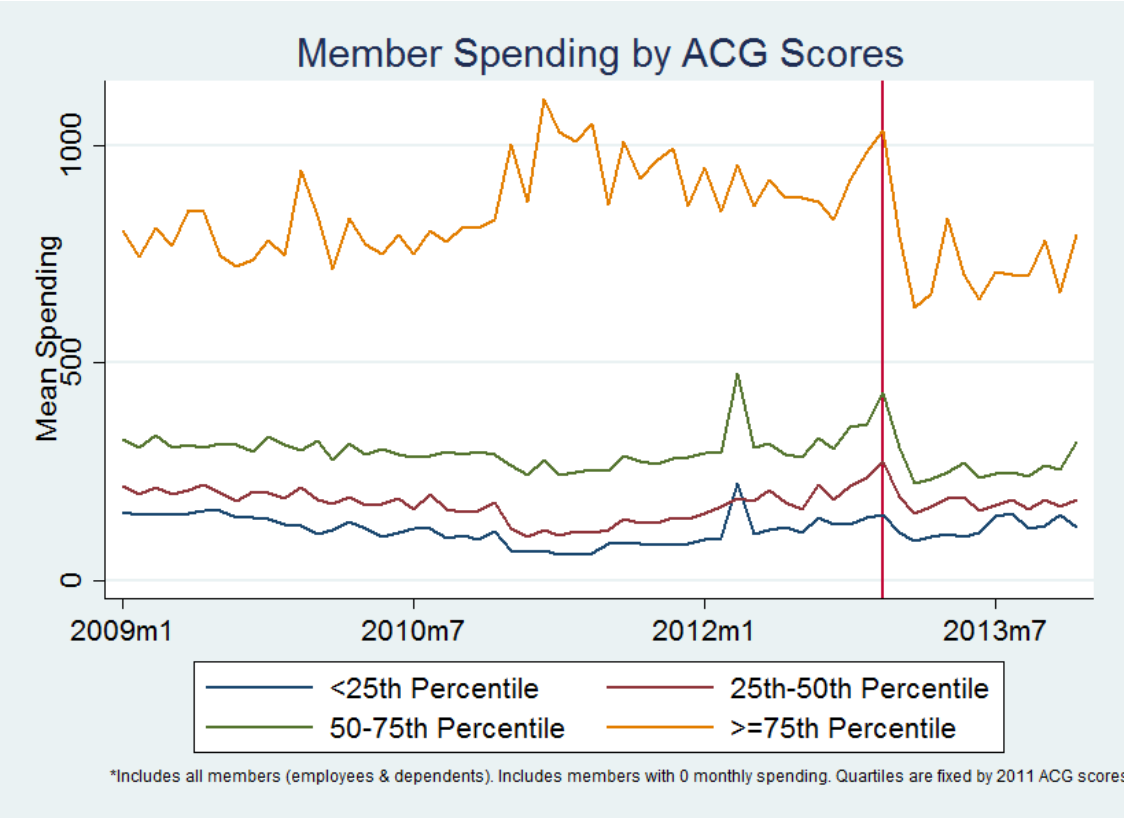



Figure 6: This figure plots adjusted spending for individuals in a given month, by 2011 ACG score quartile.

of health spending but lower marginal utilities from a dollar spent on other products or services.

We next investigate treatment effect heterogeneity as a function of individual ex ante health status measures. For these purposes, we leverage the Johns Hopkins ACG software, which assigns each individual a predictive score, based on their past year of detailed claims data, for the upcoming health year. This score reflects the type of diagnoses that an individual had in the past year, along with their age and gender, rather than relying on past expenditures alone. In Figure 6 we plot age and CPI adjusted spending over time for individuals by the quartile of their ACG score in 2011 (plotting by 2012 ACG score yields similar results).

The impact of the HDHP switch on spending is clearly largest in both % terms and in absolute value for the sickest quartile of consumers based on this ex ante risk score. We decompose the impact within this quartile further in Figure 7 since we want to know whether the treatment effect is largest for the very sickest consumers in the population. This is interesting from a behavioral response perspective, since the very sickest consumers should know in advance that they will pass the HDHP out-of-pocket maximum, and thus not be concerned about incremental total spending. The figure illustrates a noisy time series for the top 1% of sickest consumers, not surprisingly, but identifies that the main drop in medical spending comes from those between the 90th and 99th percentiles in ACG, and not the sickest 1%. As shown in Table 7 the treatment effects for the quartiles 1 (healthiest), 2, 3 and 4 are 13%, 14%, 15%, and 29% reductions ins spending respectively,

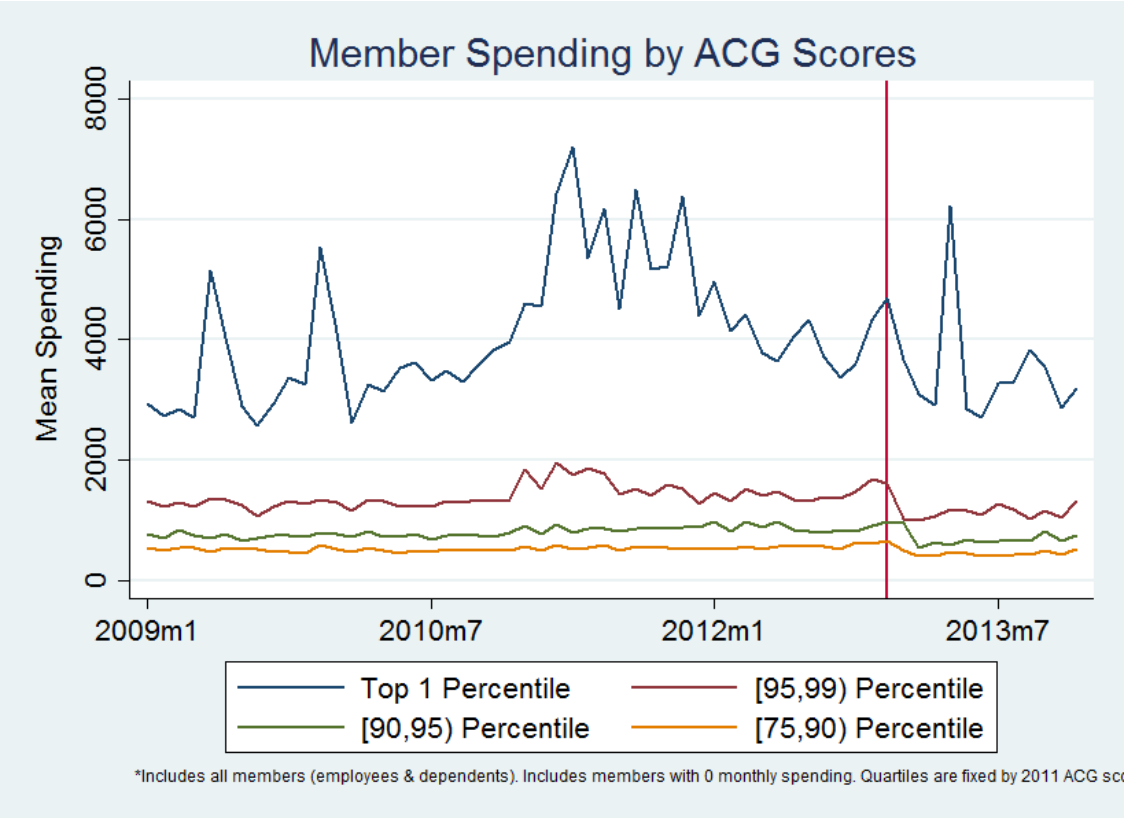



Figure 7: This figure plots adjusted spending for individuals in the top 2011 ACG score quartile in a given month, by quantile.

with all adjustments taken into account. For the top percentile this effect is 38%. Table 7 also reports results for individuals categorized by number of chronic conditions present in 2012.

Next, we observe the specific types of services consumed. Our first categorization groups services into five mutually exclusive categories roughly corresponding to point of service: (i) inpatient hospital (20.6% of 2012 spending) (ii) outpatient hospital (18.75%), emergency room (6.73%), physician office (31.17%), and prescription drug (16.16%).[15] Table 7 presents the age, CPI, and intertemporal substitution adjusted impacts of the forced HDHP switch on medical spending for each of these five categories. Reassuringly, there is a smaller impact (14%) for inpatient hospital spending, which makes sense since these are likely necessary or emergency procedures for patients likely to exceed the out-of-pocket maximum. There are more marked impacts for most of the other four categories: outpatient hospital spending drops by 11%, emergency room spending drops by 23%, office visit spending drops by 20%, and drug spending drops by 19%. The event series figures for age and CPI adjusted spending are included in Appendix 6 in Figures A16 through A20. We present the figure for emergency room use over time here in the main text to illustrate these service specific charts.

The emergency room results are quite interesting, and likely reflect the fact that many people

[15]These categories are almost exhaustive, as fewer than 1% of claims by spending remain once claims are categorized into these five categories.

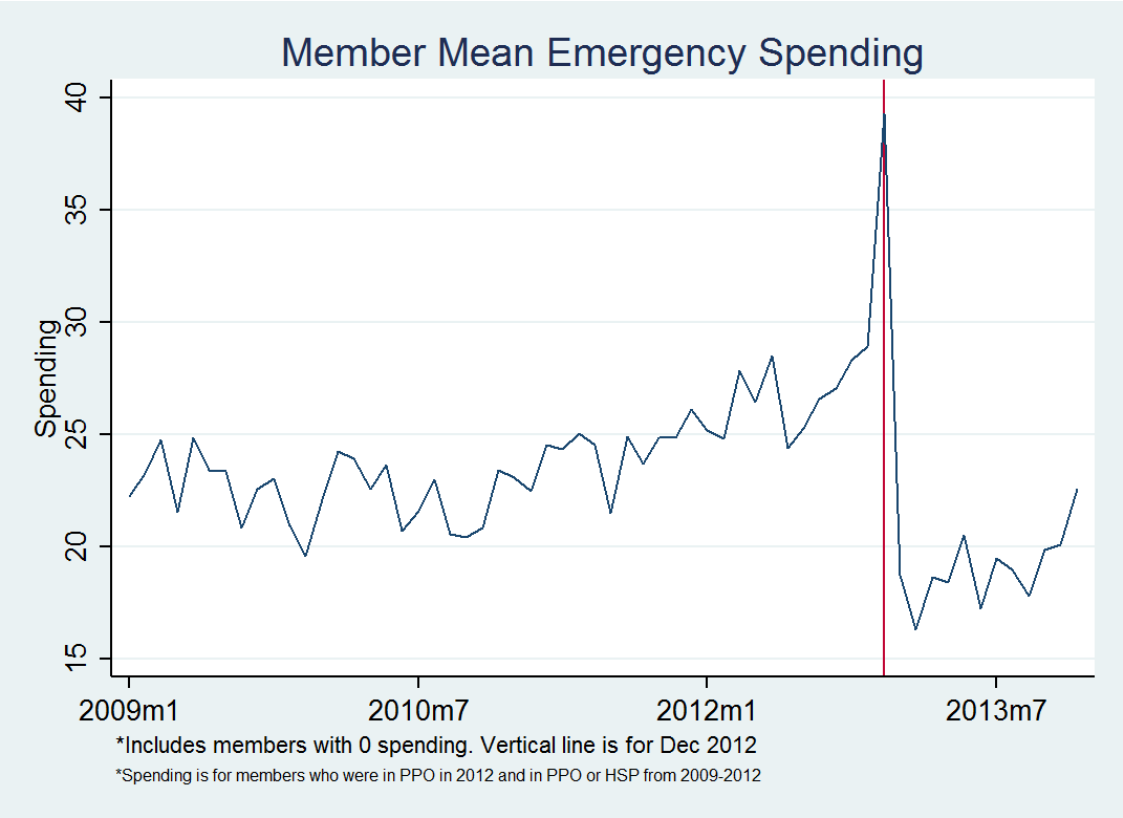



Figure 8: This figure plots age and CPI adjusted spending for individuals on emergency room services over time.

use the emergency room for routine care (in Gindi et al. (2012), around 50% of privately insured individuals who went to the emergency room went because their doctor's office was not open). The 23% drop as a result of the HDHP switch indicates the potential for cost-sharing to reduce this kind of utilization, even among high income populations.

It is also interesting to decompose the impact of RX spending into spending on branded drugs vs. spending on generic drugs. Encouraging consumers to substitute from branded drugs to equivalent generic drugs is a clear and oft-discussed way for firms and policymakers to reduce medical spending without any incremental and concurrent negative health impacts (see e.g. Bronnenberg et al. (2014)). We are able to categorize approximately 85% of our prescription drug data (weighted by spending) into one of these two categories. Table 7 reveals that the forced HDHP switch impact on branded drug spending (with all adjustments) is 20 % (off a base of $ 680 per individual) while that for generic drugs is 21 % (off a base of $224 per individual). Figure 9 illustrates the impact for branded drugs, while the analogous figure for generic drugs is contained in Appendix 6 (Figure A21). While the % spending drop in branded and generic drugs is similar, Figure A22 in Appendix 6 illustrates that there is a drop in number of branded prescriptions filled due to the HDHP switch, but not a drop in the number of generic prescriptions filled. This indicates that the drop in branded drug spending is primarily on the extensive margin, while that in generic spending is primarily on the intensive margin.

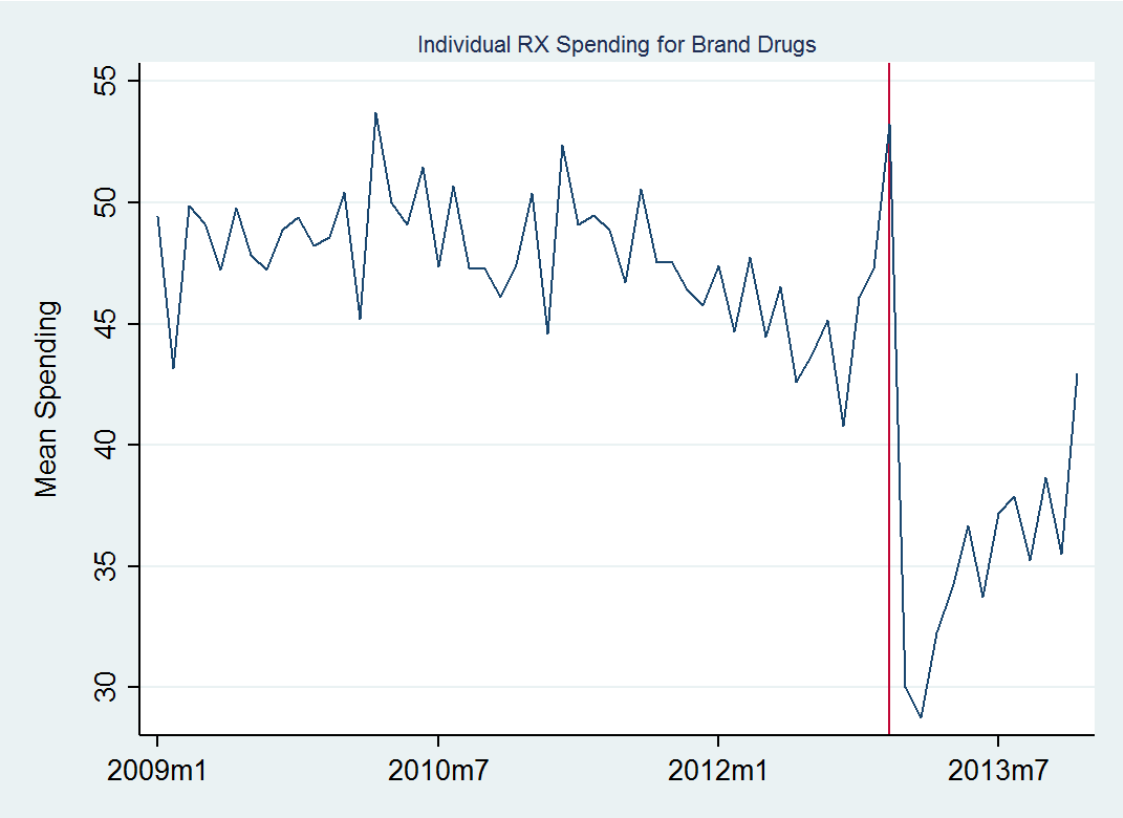



Figure 9: This figure plots age and inflation adjusted spending for individuals on branded prescription drugs.

One additional spending category of substantial interest is preventive care, which is care that the firm / insurer wants consumers to use in order to prevent the onset of diseases and medical spending that will occur down the road. Also, in the HDHP (as in most high-deductible plans) preventive care is always free for consumers, so there is actually no change in cost sharing for preventive care when switching from the PPO to the HDHP (though consumers may or may not accurately perceive this when making care decisions). Using a detailed mapping from our insurer to identify preventive services, we are able to identify such services and examine the impact of the HDHP switch on preventive care spending. Such care makes up 8.87% of 2012 spending. Figure 10 plots spending on preventive care over time. As Table 7 reveals, once adjusted for age, inflation, and intertemporal substitution, spending on preventive care decreases by 12 %, the lower treatment effect we find. This effect is consistent with the results in Baicker et al. (2013), who find that individuals do economize on preventive care in the context of cost-sharing, even though those decisions are likely bad from a social perspective.

Table 7 also includes treatment effect estimates for mental health services, where total spending is reduced by 17 %. Reassuringly, there is no impact on the proportion of individuals with extreme health events over time (Figure A24 in Appendix 6 plots the time series of this proportion).

Finally, we also investigate whether the overall drop in spending occurs primarily on the extensive margin (the decision whether or not to get care at all) or on the intensive margin (the intensity

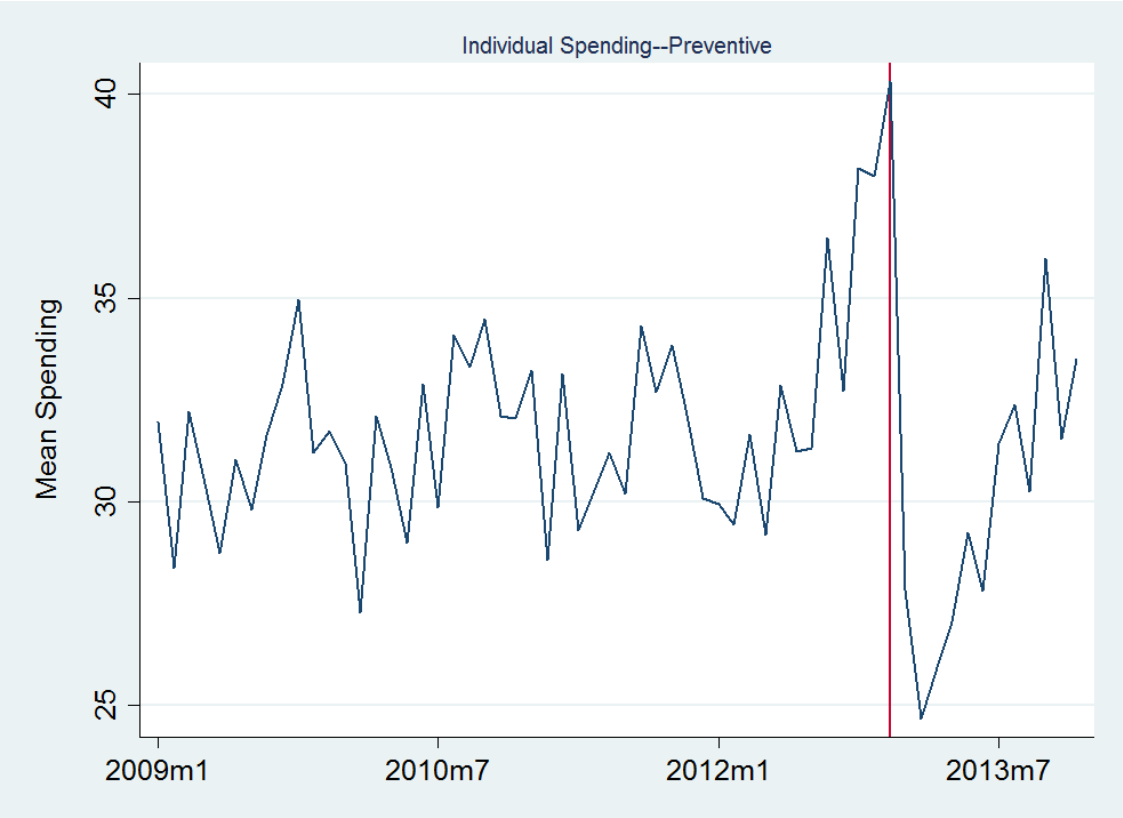



Figure 10: This figure plots age and inflation adjusted spending for individuals on preventive care.

of care to get; this could potentially also include changes in price-shopping behavior). Figure A26 in Appendix 6 plots the percentage of individuals who spend any amount in the given month while Figure A27 plots the mean spending for individuals over time, excluding those who did not spend anything in a given month. In both graphs we see similar % decreases, implying that reductions in spending are occurring on both margins.

# 4 Modeling Treatment Heterogeneity

In Section 3, we provided much descriptive evidence showing that employee response to the policy change varied with both consumer observable demographics and with the type of medical services consumed. In this section we explore two methods for develop individual treatment effect estimates. The first involves estimating regressions that predict individual level spending changes as a function of demographics, medical service types, and health status. The second uses techniques from machine learning to model heterogeneity in the most flexible way possible. Our first approach to flexibly modeling the impact of the HDHP is to estimate a set of tree models (see e.g. Hastie et al. (2013), Varian (2014)). We discuss estimation in more detail below but these models flexibly account for the interactions of variables in order to maximize predictive power, measured by out-of-sample R-squared.

### 4.1 Regression Analysis

Before turning to statistical/machine learning methods that are less familiar to most economists, we begin by analyzing treatment heterogeneity using linear models of the impact of the introduction of the HDHP. The goal of all of these analyses is to capture the impact of the HDHP across the entire population in as flexible a way possible. One of the unique features of our study setting is the detailed data not only on health but on jobs and demographics. Using these data we estimate a linear model of the differential change in spending in 2012. The model is as follows

$$(y_{i,2013} - y_{i,2012}) = \alpha + \beta X_i^{health} + \gamma X_i^{demo} + \psi X_i^{job} + \sum_{t=2010}^{2012} \tau_t (y_{i,t} - y_{i,t-1}) + \epsilon_i \quad (1)$$

where the dependent variable is the differential change in spending between 2013 — when the HDHP was implemented — and the prior year. We incorporate observable characteristics of the population as a vector $X_i^g$ where $g \in \{demographics, health, job\}$ captures the different types of characteristics that might drive response to the HDHP. We allow for an auto-regressive structure for spending changes year over year by flexibly including all prior year changes in annual spending for individual $i$.

We are interested in the results of this model in two capacities. First, we want to understand how much of the variation in the response to the HDHP can be explained by observable characteristics of the population. Second, we are interested in understanding heterogeneity and modeling it in a flexible manner that allows us to generate counterfactual scenarios that capture the potential impact of policies where we are not only interested in the mean impact but in the distributional impact in the population (e.g. menu design). To address our first goal we begin by estimating a series of versions of equation 1.

We present the results of those models in table 4.5 below. For each model we present the in- and out-of-sample (OOS) $R^2$ averaged across all 5 folds of the cross validation. That is, we estimate the model 5 times on a training sample of 80% of the observations. For each model, we then predict spending changes in the left out group between 2012 and 2013 for each individual and compute the share of actual observed variation in change in spending explained by the model.

### 4.2 Tree Models and Extensions

We next turn to estimating a variety of models based on machine learning approaches. These approaches have not been widely used in economics but are gaining traction given their tractability and ability to capture high-dimension heterogeneity as well as to rely on statistical variables selection approaches rather than intuition (see e.g. Belonie et al. (2014), Varian (2014)). The setting we study is ripe for the application of this type of approach in that we i) have very rich data on consumers that likely interacts in complex ways (e.g. responses may depend on working near a person of similar health and age specifically not just the age, health or job location by themselves) and ii) detailed data on the exact type of care consumed (not just total spending). Furthermore, as we saw in section 4.1 allowing for heterogeneity on observables in the linear model does not

capture much of the observed variation. While this could be due to unobservable heterogeneity, before making the kinds of parametric assumptions required to model such heterogeneity we seek to determine the degree to which high dimension interactions of observables may actually capture more of the heterogeneity in response than a standard econometric model.

We estimate a number of different models, all of which have at their core a tree model. Therefore, we begin by introducing the basic estimation approach and then turn to refinements that enhance model performance. We define an observation as an individual $i$ characterized by $(x_i, y_i)$ where $x_i = (x_{i,1}, x_{i,2}...x_{i,p})$ for p inputs included in the vector of observables as discussed above and $y_i = y_{i,2013} - y_{i,2012}$ for parsimony. The goal is to select values of particular inputs $p$ at which to partition the sample into $M$ regions $R_1, R_2...R_M$ in order to best match actual spending changes. For each region $M$ we take the mean spending change as the prediction of $y$; define that as $c_m$. Therefore, our model is:

$$f(x) = \sum_{m=1}^{M} c_m I(x \in R_m)$$

in order to find the optimal partition consider a splitting variable $j$ and a split point $s$. Our goal is to find a split point solving the equation:

$$\min_{j,s}[\min_{c_1} \sum_{x_i \in R_1} (y_i - c_1)^2 + \min_{c_2} \sum_{x_i \in R_2} (y_i - c_2)^2]$$

That is, we wish to select the variables and partitions of those variables that minimize the sum of squared residuals within each partition. We repeat the process within each group, selecting yet further branches to the tree. Repeating the process allows us to "grow" a tree to a very deep length. We then "prune" the tree back in order to reduce undesirable properties such as over-fitting. Using this approach we can allow for an arbitrarily complex relationship between the observable characteristics of the population and the resulting response to the HDHP.

The process of selecting a model and tuning requires obtaining the best fit for the model given actual data. The primary focus is on balancing bias and variance. In our specific setting, this amounts to trying to model the heterogeneity in the response to the HDHP that most effectively captures the true heterogeneity that can be modeled based on observables without selecting a model that generates predictions that are specific to the particular idiosyncrasies of the population but are not general: over-fitting. To see this, we can decompose the expected MSE of any estimator as (see e.g. Hastie et al. (2013) for the derivation):

$$E(y_0 - \hat{f}(x_0))^2 = Var(\hat{f}(x_0)) + [Bias(\hat{f}(x_0))]^2 + Var(\epsilon)$$

where the variance in an estimator ($Var(\hat{f}(x_0))$) captures the potential for a model to give different predictions on different data samples. This can be a particular problem in models with more flexibility, precisely what we apply in our case. The bias term captures the error introduce by mis-specifying the model. In our setting, bias could include the error introduced by applying a linear model to an inherently more complex relationship between the observables and the response

to the introduction of the HDHP. Finally, we have variation that is inherently random and cannot be modeled. In our setting we expec this to be relatively large; health spending is inherently highly variable conditional on observables (CITE).

Given these varied goals, and the tradeoff between variance and bias in an estimator, cross validation allow us to test the different models. Estimating models repeatedly across a number of sample allows us to capture variance in the estimator. Using OOS fit as a metric to asses models we can capture both bias and variance. In addition to a simple tree model, in order to improve performance we estimate a variety of different models that are known to reduce overfitting that is often the results of a single complex model. The basic notion of these approaches is to use a combination of bootstrap sampling with estimating many trees to reduce both variance and bias (Hastie et al. (2013)).

### 4.3 Unsupervised Learning Approaches

The models we develop in section 4.2 focus primarily on selecting combinations of observables that best predict the outcome of interest - spending in a particular year and, therefore, the change following the treatment. An alternate approach we propose is to select a representation of heterogeneity in high dimensions that is not specifically focused on fitting the data. Rather it is the most parsimonious representation of the covariate structure. These approaches are generally referred to as unsupervised learning - there is no outcome to target the model in selecting a partition of covariates. In our case we specifically estimate a clustering model that selects the partition of the data in $p$ dimensional space that minimizes the squared Euclidean distance amongst the points in a cluster.

This approach is a powerful approach in that it allows us to traceably represent the population in a very high dimensional space. We then estimate a cluster by cluster treatment effect yielding a non-parametric distribution of the treatment effects. To be clear, we follow our model above and represent the observable characteristics of individual $i$ as $x_i = (x_{i,1}, x_{i,2}...x_{i,p})$ for p inputs. The goal is to select a set of partitions in to k clusters, denoted by $C_k$, the solve the following:

$$\sum_{k=1}^{K} \frac{1}{|C_k|} \sum_{i,i' \in C_k} \sum_{j=1}^{p} (x_{i,j} - x_{i',j})^2$$

where we define $|C_k|$ as the number of observations in cluster $k$. That is, for a given number of clusters ($K$) we select a categorization that groups the most similar points, across the vector of observables; the minimize the sum of pairwise Euclidean distances within each cluster.

Having selected a set of clusters each individual $i$ in year $t$, $y_i, t$, is assigned a predicted spending value:

$$\hat{y}_{i,t} = \frac{1}{n} \sum_{x_i \in C_k} y_{i,t}$$

That is, we assign the mean outcome in cluster $k$ to that group. We repeat the process for a

variety of numbers of clusters and allowing for alternate dimensions of heterogeneity in $g$.

An important thing to note in this approach is that the decision on the number of clusters is a modeling decision made by the econometrician. To optimize the selected value for every set of attributes ($X^g$) we estimate the model at a variety of number of clusters. For our primary models we select the number of clusters where the change in MSE from adding clusters drops. That is, we seek to find the point at which splitting any particular cluster does little reduce within cluster variation in pairwise distance within the cluster. Put differently, splitting a cluster produces two subsequent clusters with similar within cluster heterogeneity.

### 4.4 Estimating Elasticities with Heterogeneous Response

The sections above focus primarily on modeling the heterogeneous response to the introduction of the HDHP. While this is the main object of interest, we also estimate elasticities that can vary in the population as a function of observables We do so as it allows us to provide unit free measures of the response that can be compared to the large existing literature (e.g. the RAND study). It also allows us to model welfare under different insurance contract designs and offerings in our counterfactual analysis.

To estimate the elasticity of response to the HDHP we need the change in health care spending as well as the change in underlying exposure to prices. For our average treatment effect estimation we use the average change in the actuarial value of the benefit between the PPO and HDHP plans (100% to 78%). We use a similar approach allowing for heterogeneity. Specifically, we estimate utilization in 2011 for each individual in the population as a function of observables. Based on these predicted values we then compute the change in actuarial value for that individual. Using these estimates we are able to generate an individual specific change in actuarial value as a function of observables. We define the individual specific change in price/actuarial value as:

$$\hat{p}_{2,i} - \hat{p}_{1,i} = (1 - HDHP(E(c_i, 2011|X_{i,t})) - (1 - PPO(E(c_i, 2011|X_{i,t}))$$

where $HDHP()$ and $PPO()$ are functions generating out-of-pocket spending based on a level of cost $c_i$ for each of the respective plans. Combing estimates for prices changes with individual specific changes in utilization gives us a individual specific semi-arc elasticity of:

$$\frac{(\hat{q}_{2,i} - \hat{q}_{1,i})/(\hat{q}_{2,i} + \hat{q}_{1,i})}{\hat{p}_{2,i} - \hat{p}_{1,i}}$$

where $\hat{q}$ and $\hat{p}$ reflect the fitted values from the particular model used (e.g. OLS, tree, random forest).

### 4.5 Model Performance Results

Here we present results assessing the performance of the different approaches used to estimate the heterogeneous impact of the introduction of the HDHP. We begin by estimating three different versions of our model: an OLS model, a tree model and a random forest model. Each model was

estimated sequentially allowing for different elements to enter. We being with a simple model of the change in spending between pre- and post-HDHP (2012 to 2013) as a function of year-over-year changes in spending in prior years. In subsequent models we allow for different observables by groups $g \in \{demographics, health, job\}$. Depending on the model, these additional elements can interact in different ways. In the simple OLS they merely control for the average effect of having a particular covariate but in tree models each observable can interact. Thus, for example, in a random forest model adding in demographic and health observables allows for heterogeneous impacts that can vary by demographic by disease states.

Table 4.5 presents the out-of-sample performance for three different modeling approaches. The first column presents the results for the OLS model. We see relatively high $R^2$ in the baseline model of .277. That is, previous spending changes can explain roughly 27% of the variation in spending changes observed between 2012 and 2013. Put differently, if we assume that the linear model is an appropriate approach, much of the heterogeneity in the impact of the HDHP is, in fact, merely underlying variation in year-over-year spending that was observed in the PPO plan. Turning to the second row of column 1 suggests incorporating demographic does nothing to improve the model. Including ACG scores actually lowers the $R^2$ slightly to .266 and subsequently incorporating job characteristics leaves that unchanged. Taken together, these results suggest that the OLS model accounts for the underlying year-over-year heterogeneity in utilization but that incorporating the characteristics of individuals does not capture any variation in the impact of the HDHP.

The second column present results from a simple tree model. As discussed above, the tree model allows us to more flexibly model high dimension impacts of the HDHP based on observables but also has the potential to over-fit the data. The results suggest that this is, in fact, the case. All of the out-of-sample $R^2$ value are negative. The negative value captures that fact that using the model fit on different subsets of the data to predict in the remaining population is worse than no model at all (a straight line in a regression sense)! While the results vary somewhat by specification we find that overfitting is an important problem in the tree model we estimate.

Finally, column 3 presents for a random forest estimating using 50 trees [NOTE: CHANGE THIS AS WE ESTIMATE THE FULL MODEL WITH 500 TREES+]. A few things are immediately clear. First, we see that we are getting the improved performance in terms of reducing model variance that a random forest provides. Recall that a random forest is a set of tree models that are estimated on bootstrapped samples. Further, each subsequent model oversampled poorly fitted data in order to reduce bias. Finally, the results are based on a 'voting' across models, thereby reducing variance in the estimator. This is clearly the case as the OOS $R^2$ from the random forest is close to the OLS model and a marked change from the simple tree approach. The $R^2$ values across models demonstrate some of the advantages of this approach to modeling heterogeneous treatment effects. Row 1 show that much of the variation in the estimated impact of the HDHP can be explained by flexibly modeling an individual's prior year spending changes across all of the previous years (2009-2012). The non-linear model explains slightly less of the variation in the 2012-2013 change this way that the OLS model (.186 vs. .277 for OOS $R^2$). Adding in additional demographics to the random forest model, however, changes the fit of the model in a material way.

| **Out-of-sample $R^2$ by Modeling Approach** | OLS | Tree | Forest |
|---|---|---|---|
| Prior Spending | .277 | -.277 | .186 |
| $+X^{demog}$ | .277 | -.3636 | .221 |
| $+X^{health}$ | .266 | -.298 | .188 |
| $+X^{job}$ | .266 | -.230 | .218 |

This is in juxtaposition to the OLS model. We see allowing for variation in the impact of the HDHP based on both demographics and prior year spending changes (row 2 of column 3 of table 4.5). Fit improve substantially after including demographics. Subsequently including health measures (ACG scores) in row 3 explains less of the variation in impact. Finally, including all types of observables generates an OOS $R^2$ of .218. While the OOS $R^2$ in the random forest model is lower than in a simple OLS we also note that there seems to be an impact of the higher dimension interaction of observables on HDHP impact. Since this is not permitted in the OLS specification we are interested in some of the heterogeneity that seems to be captured in the random forest model but not in the OLS. That is, the OLS model in the first version attributes much of the effect to individual level prior year spending changes. This may obscure some of the underlying heterogeneity that actually exists since in the OLS setup observables can only enter linearly, conditional on the baseline model. This finding is consistent the results in Varian (2014) in that he find simple fit between logit model and a tree model in predicting mortgage lending decisions. However, he finds that the observable race —the main focus of the original paper and his subsequent assessment — differs dramatically between the approaches, underscoring the value of machine learning approaches relative to the standard econometric models in capturing treatment effect heterogeneity.

### 4.6 Estimated Treatment Effect Heterogeneity

TBD

## 5 Welfare Effects of Plan and Menu Design

#### Cost Model

TBD

#### Welfare Impact of HDHP Switch

We estimate the welfare impact of the HDHP switch using the estimates on consumer price responsiveness from this paper together with estimates of risk preferences from Handel and Kolstad (2014) which studies data from the same firm during the same time period. Average CARA risk preferences in the full model estimated in Handel and Kolstad (2014) have mean value of $9.35 * 10^{-5}$ in our

primary sample for this paper. We also use detailed estimates of consumer total health expenditure risk as an ingredient into this simulation (see Handel and Kolstad (2014) for more details.

We simulate the impact of the forced switch for all employees at the firm, extending the estimates in this paper to the entire sample of employees at the firm. The total mean predicted cost savings per individual are $868.83. This represents the upper bound on potential savings, since consumers may value foregone medical spending at something less than its new cost under the HDHP. If consumers value foregone medical spending exactly equal to the average HDHP cost sharing rate of 22%, then the value of foregone spending would be $191.14. Under the assumption that the value of foregone medical spending lies somewhere in between 0 and the average cost sharing rate for that spending, the welfare gains from more efficient medical spending as a result of the switch to the HDHP are between $677.69 and $868.83.

The predicted within sample loss from reduced risk protection is slightly higher than $74 per individual, taking into account population risk preferences, yearly ex ante health risk distributions, and the HDHP contract design. As a result, the overall welfare impact of the forced HDHP switch lies in the range [$604,$794]. Mean consumer risk aversion would have to shift up to XXXX for the HDHP switch to break even from a welfare perspective Alternatively, consumer price responsiveness would have to decrease to approximately 1.4% for (price elasticity of .06), given current estimated risk preferences, for the switch to break even.[16]

**Value-Based Insurance Design**

TBD

**Multi-Plan Menu Design**

TBD

# 6 Conclusion

TBD

[16]This analysis assumes that the value from risk protection is based on the distribution of ex ante health risk prior to accounting for the impact of moral hazard on that risk distribution.

## References


**Aron-Dine, Aviva, Liran Einav, Amy Finkelstein, and Mark Cullen**, "Moral Hazard in Health Insurance: How Important is Forward Looking Behavior?," November 2012. NBER Working Paper No. 17802.

**Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein**, "Behavioral Hazard in Health Insurance," September 2013. Harvard Working Paper.

**Barseghyan, Levon, Francisca Molinari, Ted O'Donoghue, and Joshua Teitelbaum**, "The Nature of Risk Preferences: Evidence from Insurance Choices," *American Economic Review*, 2013, *103* (6), 2499–2529.

**Belonie, Alexandre, Victor Chernozukov, and Christian Hansen**, "High-Dimensional Methods and Inference on Structural and Treatment Effects," *Journal of Economic Perspectives*, 2014, *28* (2), 29–50.

**Bronnenberg, Bart J., Jean-Pierre Dube, Matthew Gentzkow, and Jesse M. Shapiro**, "Do Pharmacists Buy Bayer? Informed Shoppers and the Brand Premium," August 2014. University of Chicago Working Paper.

**Buntin, Melinda B., Amelia M. Haviland, Roland McDevitt, and Neeraj Sood**, "Healthcare Spending and Preventive Care in High-Deductible and Consumer-Directed Health Plans," *American Journal of Managed Care*, 2011, *17* (3), 222–230.

**Cabral, Marika**, "Claim Timing and Ex Post Adverse Selection," 2013. University of Texas Working Paper.

**Carlin, Caroline and Robert Town**, "Adverse Selection, Welfare, and Optimal Pricing of Employer Sponsored Health Plans," 2009. University of Minnesota Working Paper.

**Chandra, Amitabh, Jonathan Gruber, and Robin McKnight**, "Patient Cost-Sharing, Hospitalization Offsets, and the Design of Optimal Health Insurance for the Elderly," 2008. NBER Working Paper No. 12972.

**_ , _ , and _** , "Patient Cost-Sharing and Hospitalization Offsets in the Elderly," *American Economic Review*, 2010, *100* (1), 193–213.

**Cherkin, Daniel C., Louis Grothaus, and Edward H. Wagner**, "The Effect of Office Visit Copayments on Utilization in a Health Maintenance Organization," *Medical Care*, 1989, *27* (11), 1036–1045.

**Cohen, Alma and Liran Einav**, "Estimating Risk Preferences from Deductible Choice," *The American Economic Review*, 2007, *97* (3), 745–788.

**Congressional Budget Office**, "The 2013 Long-Term Budget Outlook," September 2013.

**Duan, Naihua, Willard G. Manning, Carl N. Morris, and Joseph P. Newhouse**, "A Comparison of Alternative Models for the Demand for Medical Care," *Journal of Business and Economic Statistics*, 1983, *1* (2), 115–126.

**_ , _ , _ , and _** , "Choosing Between the Sample-Selection Model and the Multi-Part Model," *Journal of Business and Economic Statistics*, 1984, *2* (3), 283–289.

**Einav, Liran, Amy Finkelstein, and Paul Schrimpf**, "The Response of Drug Expenditure to Non-Linear Contract Design: Evidence from Medicare Part D," August 2013. NBER Working Paper No. 19393.

_ , _ , **Stephen Ryan, Paul Schrimpf, and Mark Cullen**, "Selection on Moral Hazard in Health Insurance," *American Economic Review*, 2013, *103* (1), 178–219.

**Finkelstein, Amy, Sarah Taubman, Bill Wright, Mira Bernstein, Jonathan Gruber, Newhouse Joseph P., Heidi Allen, Katherine Baicker, and The Oregon Health Study Group**, "The Oregon Health Insurance Experiment: Evidence From the First Year," *Quarterly Journal of Economics*, 2012, *127* (3), 1057–1106.

**Gaynor, Martin, Jian Li, and William B. Vogt**, "Substitution, Spending Offsets, and Prescription Drug Benefit Design," *Forum for Health Economics & Policy*, 2007, *10* (2).

**Gindi, Renee M., Robin A. Cohen, and Whitney K. Kirzinger**, "Emergency Room Use Among Adults Aged 1864: Early Release of Estimates From the National Health Interview Survey, January-June 2011," 2012. Division of Health Interview Statistics, National Center for Health Statistics. Available at `http://www.cdc.gov/nchs/data/nhis/earlyrelease/emergency_room_use_january-june_2011.pdf`.

**Handel, Benjamin R.**, "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," *American Economic Review*, 2013, *103* (7), 2643–2682.

_ **and Jonathan T. Kolstad**, "Health Insurance for Humans: Information Frictions, Plan Choice, and Consumer Welfare," August 2014. UC Berkeley Working Paper.

**Hastie, Trevor, Robert Tibshirani, and Jerome Friedman**, *The Elements of Statistical Learning*, New York, NY: Springer, 2013.

**Haviland, Amelia M., Neeraj Sood, Roland D. McDevitt, and M. Susan Marquis**, "The Effects of Consumer-Directed Health Plans on Episodes of Health Care," *Forum for Health Economics & Policy*, 2011, *14* (2).

_ , _ , _ , **and** _ , "How Do Consumer-Directed Health Plans Affect Vulnerable Populations?," *Forum for Health Economics & Policy*, 2011, *14* (2).

**Keeler, Emmett B. and John E. Rolph**, "The Demand For Episodes of Treatment in the Health Insurance Experiment," *Journal of Health Economics*, 1988, *7* (4), 337–367.

**Kowalski, Amanda**, "Estimating the Tradeoff Between Risk Protection and Moral Hazard with a Nonlinear Budget Set Model of Health Insurance," May 2012. NBER Working Paper no. 18108.

_ , "Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care," November 2013. Yale Working Paper.

**Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz**, "Health Insurance and the Demand for Medical Care: Evidence From a Randomized Experiment," *American Economic Review*, 1987, *77* (3), 251–277.

**Marsh, Christina**, "Estimating Health Expenditure Elasticities Using Nonlinear Reimbursement," February 2011. University of Georgia Working Paper.

**Newhouse, Joseph P. and the Insurance Experiment Group**, *Free For All?: Lessons from the RAND Health Insurance Experiment*, Harvard Univ. Press, 1993.

**Sasso, Anthony T. Lo, Lorens A. Helmchen, and Robert Kaestner**, "The Effects of Consumer-Directed Health Plans on Health Care Spending," *The Journal of Risk and Insurance*, 2010, *77* (1).

**Selby, Joe V., Bruce H. Fireman, and Bix E. Swain**, "Effect of a Co-payment on Use of the Emergency Department in a Health Maintenance Organization," *New England Journal of Medicine*, 1996, *334* (10), 635–641.

**Sood, Neeraj, Zachary Wagner, Peter J. Huckfeldt, and Amelia M. Haviland**, "Price Shopping in Consumer Directed Health Plans," *Forum for Health Economics & Policy*, 2013, *16* (1), 35–53.

**Towers Watson**, "Driving Performance, Connecting to Value," *19th Annual Towers Watson / National Business Group on Health Employer Survey on Purchasing Value in Health Care*, 2014.

**Varian, Hal**, "Big Data: New Tricks for Econometrics," *Journal of Economic Perspectives*, 2014, *2* (28), 3–28.

**Zeckhauser, Richard**, "Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives," *Journal of Economic Theory*, 1970, *2* (1), 10–26.

# Appendix: Additional Analysis

## 6.1 Treatment Effect Regression Analysis.

Since our sample contains the same individuals over the study time period (2009-2013), our descriptive analysis of the 2012-2013 treatment effect isn't impacted by shifting population characteristics that need to be controlled for (e.g. in a regression framework) over time. Moreover, due to our large sample of over 80,000 individuals, we are not particularly worried about idiosyncratic utilization differences: the large sample size should ensure (i) that large unusual idiosyncratic shocks at the individual-level do not meaningfully impact our treatment effects and (ii) that the underlying population distribution of health shocks is similar on a year to year basis due to the law of large numbers.[17]

Nevertheless, for comparison purposes to prior work it may be interesting to estimate a basic treatment effect with regression models.

Our first (simplest) specification is a one-stage model with the following form:

$$\log(Y_{it}+1) = \alpha + \beta Plan_{it} + \delta X_{it} + \lambda_t + \epsilon_{it}$$

Here, where $Y$ is CPI-adjusted spending by individual $i$ in month $t$ (with 1 added to allow us to provide a value for months where individuals spend zero). $Plan$ is a dummy for being enrolled in the HDHP, $X$ are demographic characteristics (such as age, gender, and income bracket) and $\lambda_t$ are month fixed effects. Since our prior analysis adjusts for intertemporal substitution at a firm-level and it is difficult to "assign" individual-level intertemporal substitution adjustments in these regression specifications we include specific fixed effects for November 2012 through February 2013 to control for excess or reduced spending in those months relative to baseline. Our main coefficient of interest in the model is $\beta$, which measures the log ratio of spending by employees in the HDHP in 2013 relative to their spending in the PPO in earlier years.

The additional two regressions estimated are two-stage models where, in the first stage, we explicitly account for the fact that there is a mass of spending at zero by estimating a probit model of the probability of spending any positive amount at all. In the second stage, we then estimate a similar regression as above, but only for those individuals who spent a positive amount. This model thus takes the form of the system of two equations:

$$\begin{aligned} P(Y_{it}>0) &= \Phi(\alpha_1 + \beta_1 Plan_{it} + \delta_1 X_{it} + \lambda_{1,t} + \nu_{it}) \\ \log(Y_{it})|Y_{it}>0 &= \alpha_2 + \beta_2 Plan_{it} + \delta_2 X_{it} + \lambda_{2,t} + \epsilon_{it} \end{aligned}$$

We estimate the two-stage model in both logs and levels. This model both accounts for the nontrivial mass of people who spend nothing on health care that tends to occur in individual-level health expenditure data, while also accounting for the skewness in the distribution of expenses for those who do spend some positive amount. Our approach follows the spirit of the analysis performed on the RAND Experiment in Manning et al. (1987), and its properties are explored in Duan et al. (1983) and Duan et al. (1984).

Columns (4a) through (4c) of Table 5 present the predicted mean individual spending by year in each of these three regression models. For the two-stage models, predicted spending is calculated by multiplying the probability of spending greater than zero by the predicted spending conditional on spending greater than zero. Modeling spending in logs weakens the impact of extreme high-

[17] This assumes no substantial population-level health shocks, which (i) would be hard to separately identify and (ii) there is no evidence of during our sample period.

| **Alternate Sample Demographics** | | |
|---|---|---|
| | Present 2012-2013 | Present 2011-2013 |
| **N** - Employees | 42,674 | 38,264 |
| **N** - Emp.&Dep. | 123,751 | 108,770 |
| | | |
| **Enrollment in PPO in 2012** | 100% | 100% |
| | | |
| **Gender - Employees(%Male)** | 76.38% | 76.66% |
| **Gender - Emp. & Dep.** | 51.55% | 51.70% |
| | | |
| **Age - Employees** | | |
| 18-29 | 11.41% | 9.09% |
| 30-54 | 83.95% | 86.02% |
| ¿=55 | 4.65% | 4.89% |
| | | |
| **Age - Emp. & Dep.** | | |
| ¡18 | 35.99% | 35.38% |
| 18-29 | 11.61% | 10.59% |
| 30-54 | 49.39% | 50.95% |
| ¿=55 | 3.01% | 3.09% |
| | | |
| **Income** | | |
| Tier 1 (¡$75K) | 2.33% | 2.07% |
| Tier 2 ($75K-$100K) | 8.82% | 7.46% |
| Tier 3 ($100K-$125K) | 33.62% | 33.29% |
| Tier 4 ($125-$150) | 32.55% | 33.65% |
| Tier 5 (150K-175K) | 13.39% | 13.92% |
| Tier 6 (175K-200K) | 5.39% | 5.57% |
| Tier 7 (200K-225K) | 2.27% | 2.37% |
| Tier 8 (225K-250K) | 0.76% | 0.78% |
| Tier 9 (¿250K) | 0.88% | 0.88% |
| | | |
| **Family Size** | | |
| 1 | 22.03% | 20.15% |
| 2 | 18.78% | 18.61% |
| 3+ | 59.19% | 61.24% |
| | | |
| **Individual Spending, 2012** | | |
| Mean | $5,596 | $5,624 |
| 25th Percentile | $663 | $717 |
| Median | $1,853 | $1,922 |
| 75th Percentile | $4,978 | $5,175 |
| 95th Percentile | $19,613 | $20,193 |
| 99th Percentile | $54,814 | $54,903 |

Table A1: This table presents summary demographic statistics for (i) employees enrolled in the PPO plan option at the firm in 2012 who were present over the time horizon 2012-2013; and (ii) the same sample, restricted to employee who were also present in 2011.

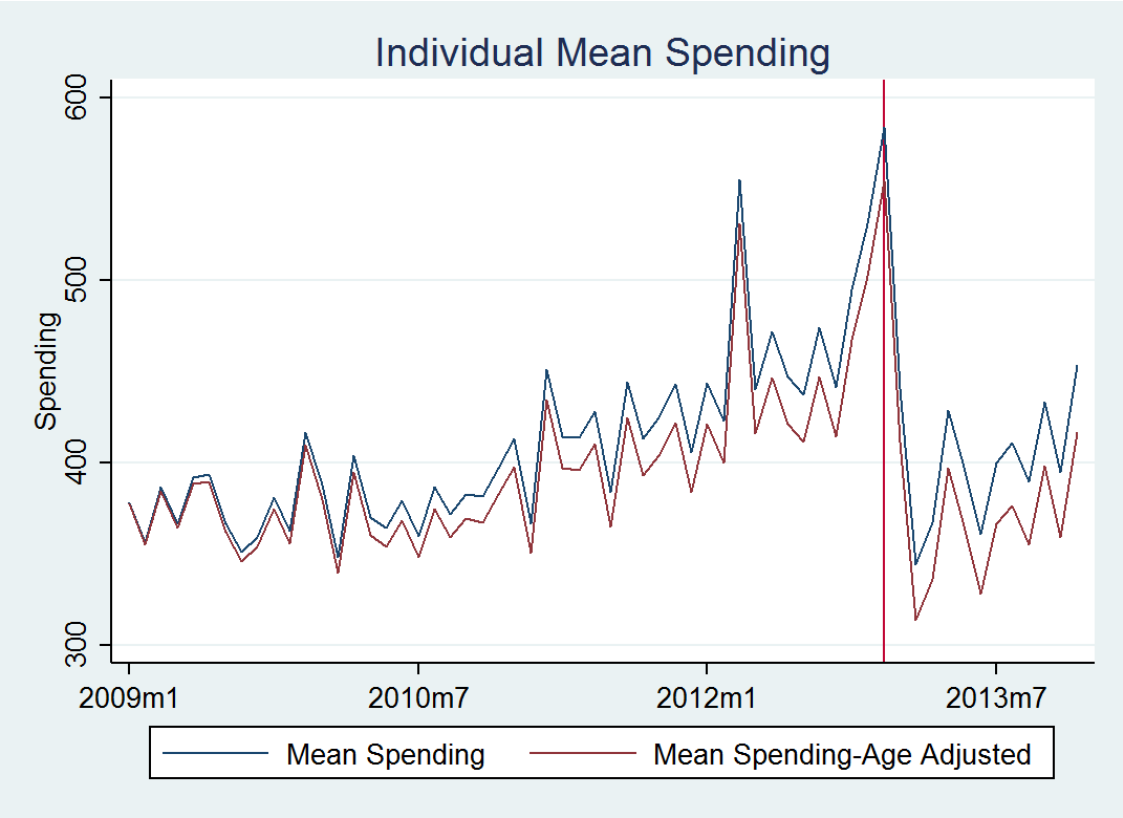



Figure A11: This figure plots spending for families in a given month, adjusted and unadjusted for age trends.

sp ending individuals on mean spending (which may be productive in reducing the impact of idiosyncratic variation at the top end of spending, where all individuals are already past the out-of-pocket maximum). Relative to the two-stage model, the one-stage model allows the mass of zero-spending individuals to have a much stronger effect than when the possibility of spending zero is modeled explicitly. The two-stage model with levels (logs) predicts a 20.08% (29.10%) reduction in spending at the individual-level due to the forced HDHP switch, while the one stage model predicts a 34.99% drop in spending.

## Figures

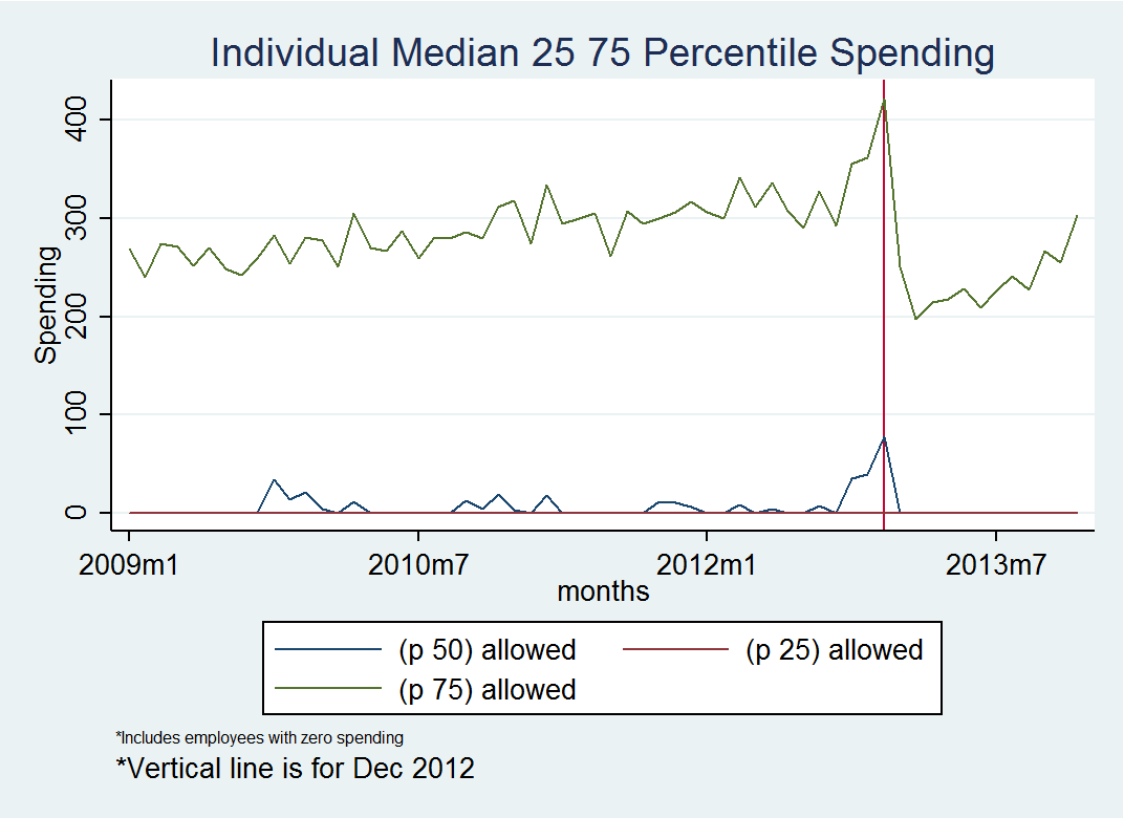



Figure A12: This figure plots the 25th percentile, median, and 75th percentile of spending on health at the individual level in a given month.

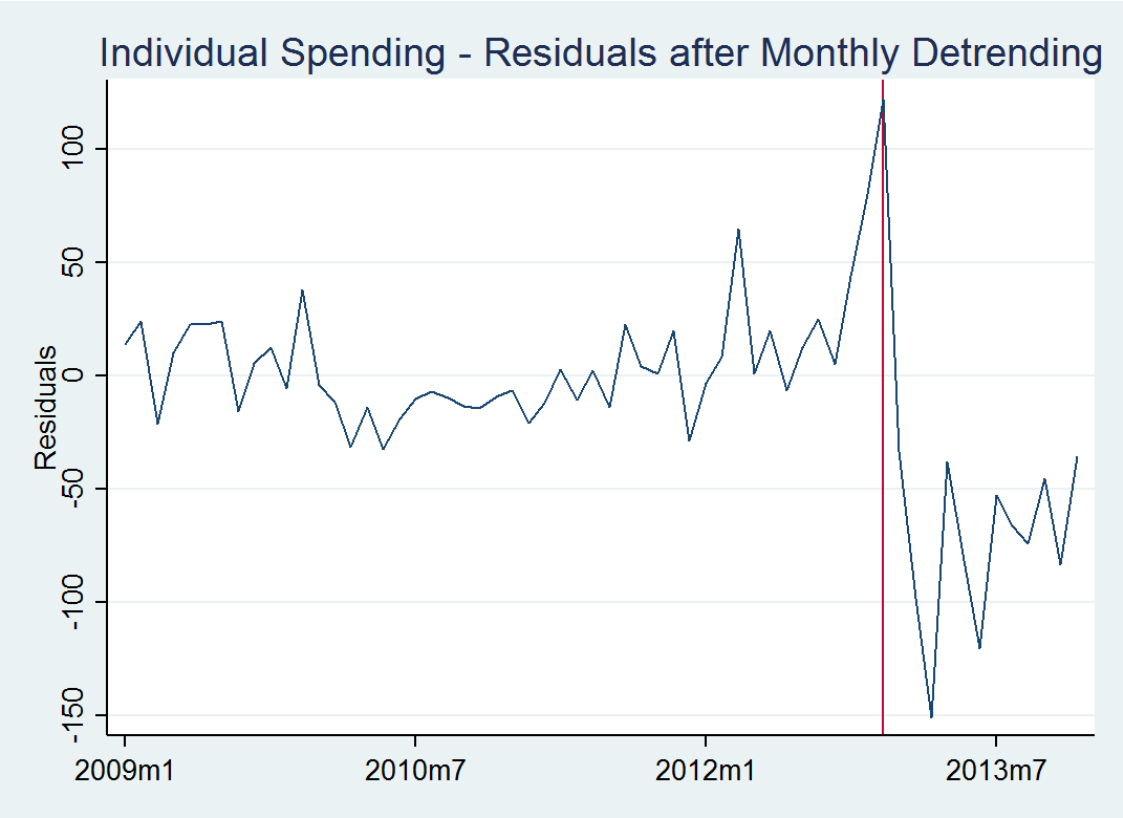



Figure A13: This figure plots the residuals from the procedure outlined above.

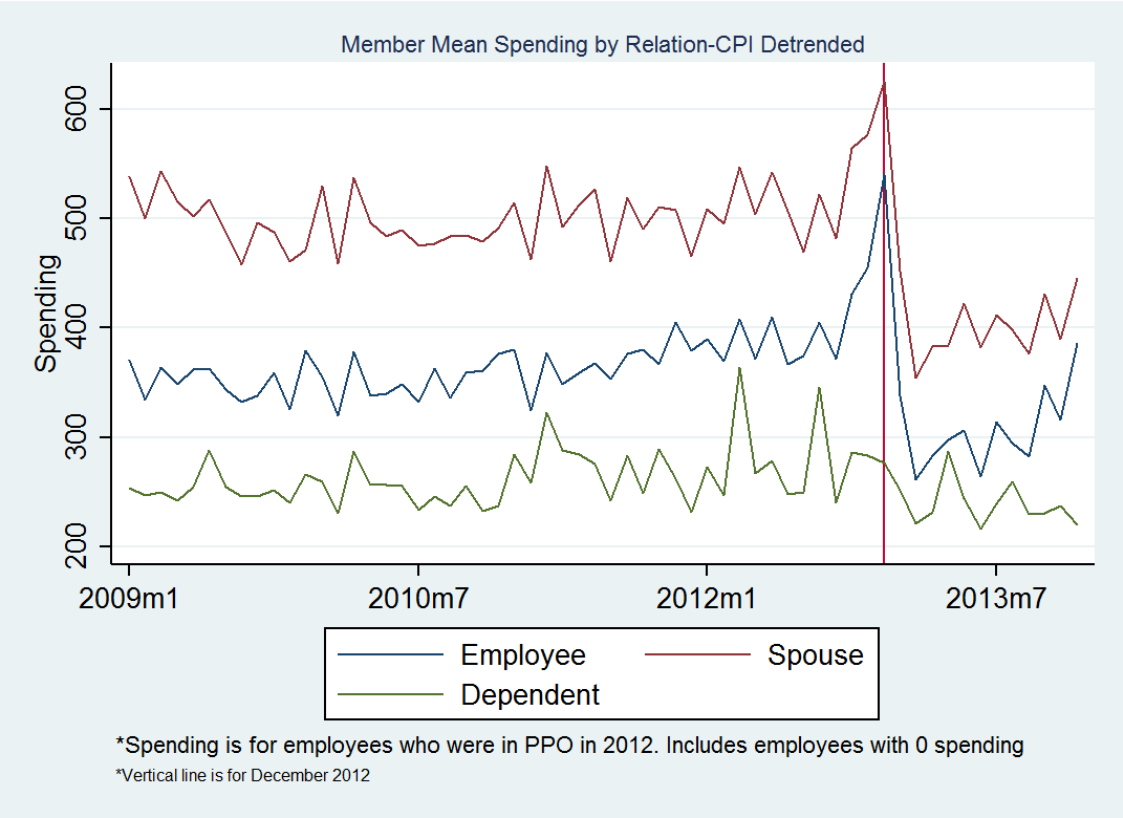



Figure A14: This figure plots adjusted spending for individuals in a given month, by relation to the employee.

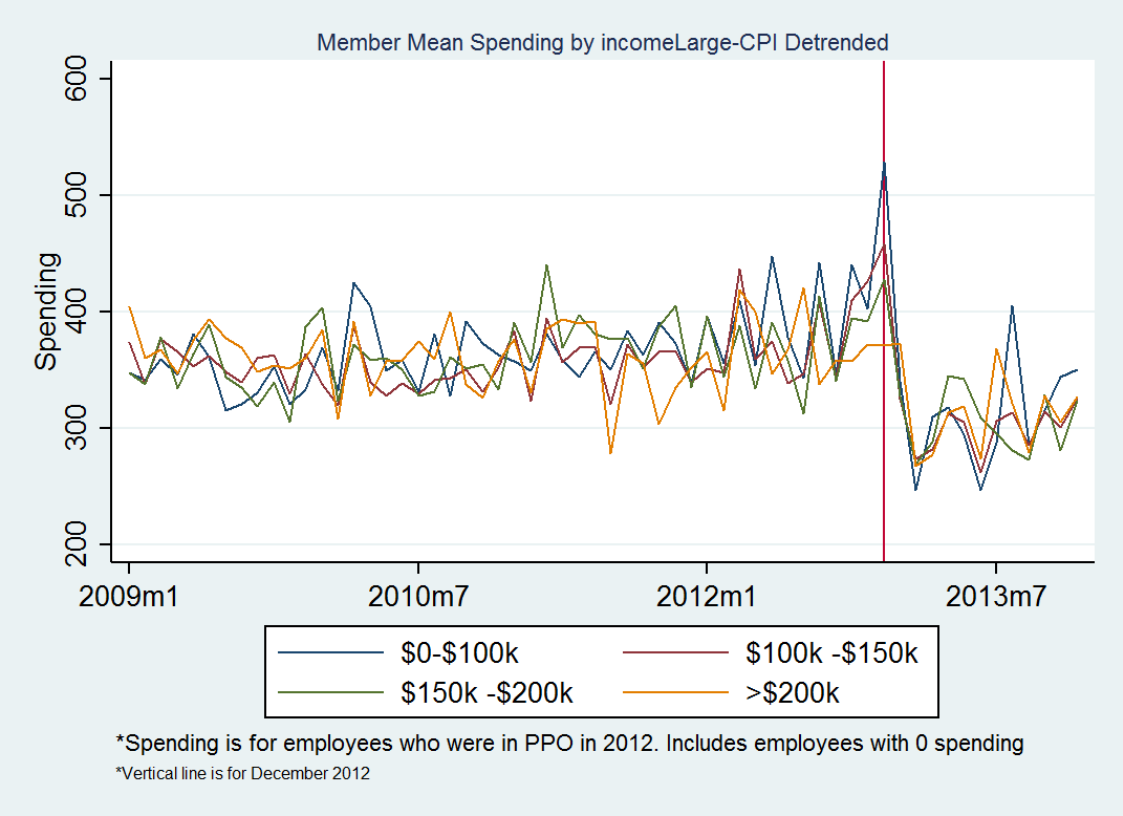



Figure A15: This figure plots adjusted spending for individuals in a given month, by income bracket.

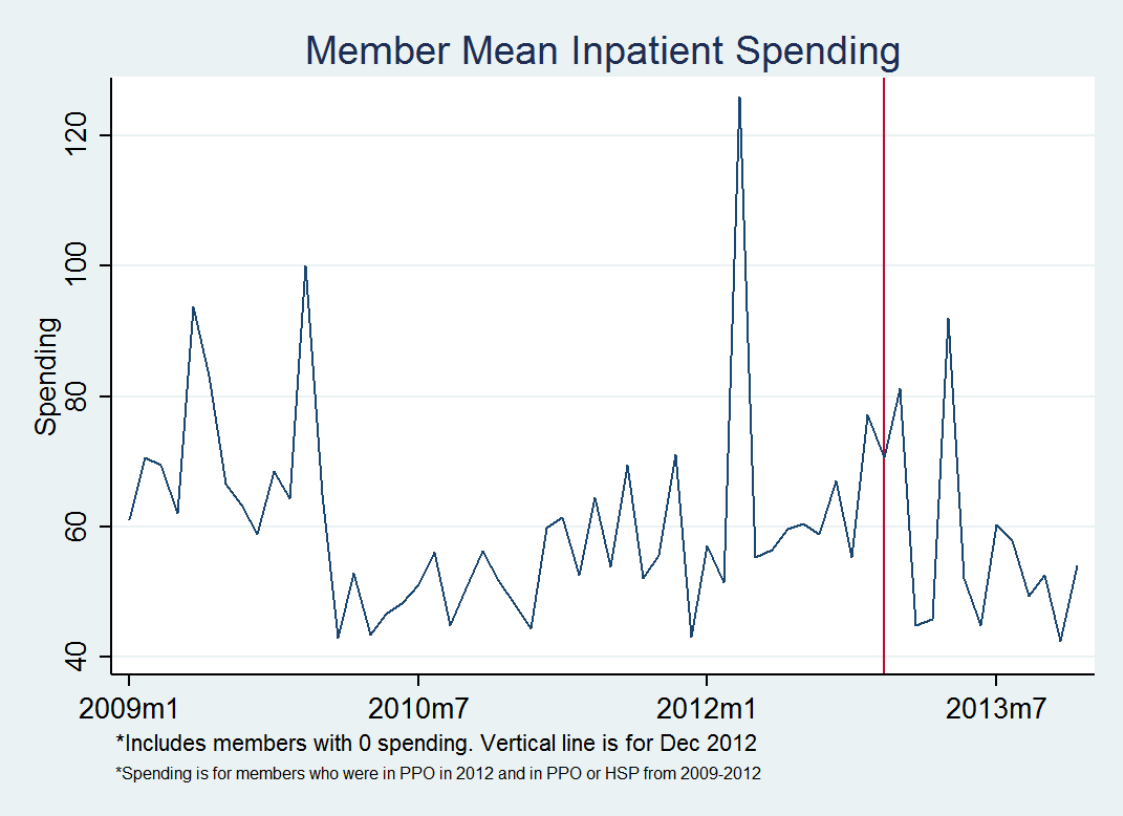



Figure A16: This figure plots adjusted spending for individuals on inpatient hospital services.

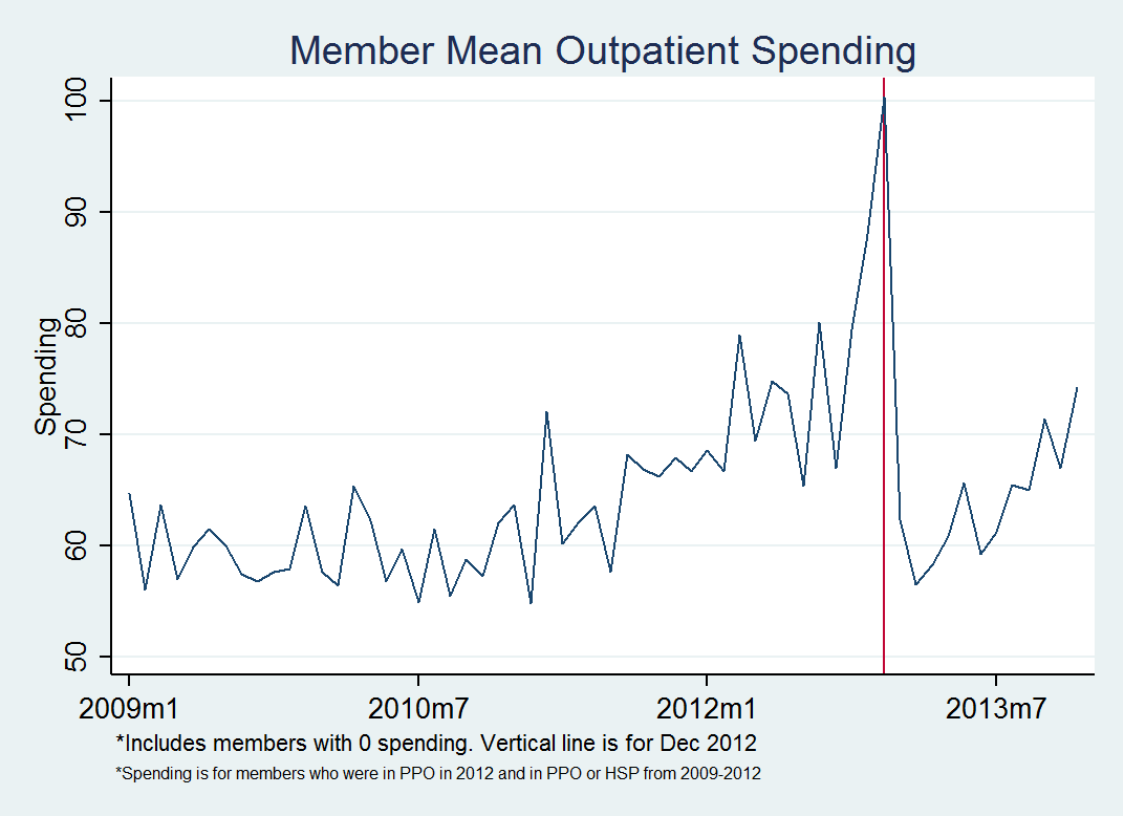



Figure A17: This figure plots adjusted spending for individuals on outpatient hospital services.

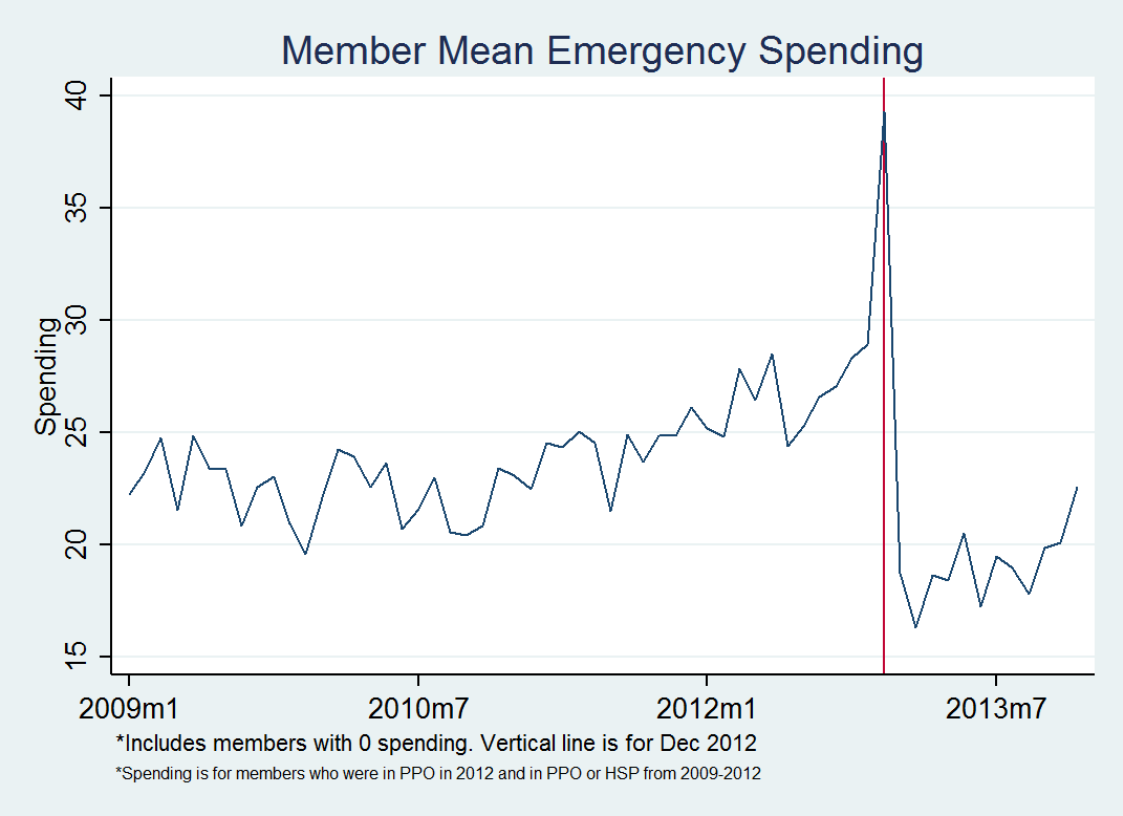



Figure A18: This figure plots adjusted spending for individuals on emergency room services.

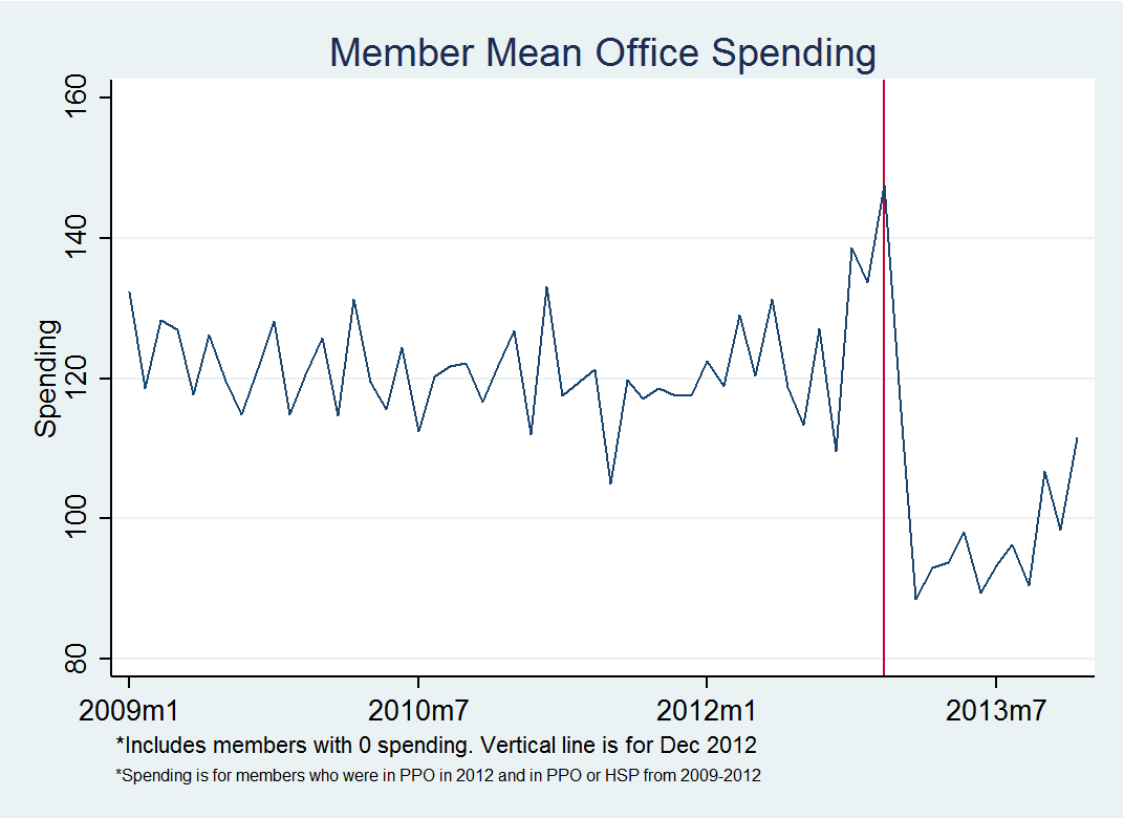



Figure A19: This figure plots adjusted spending for individuals on office services.

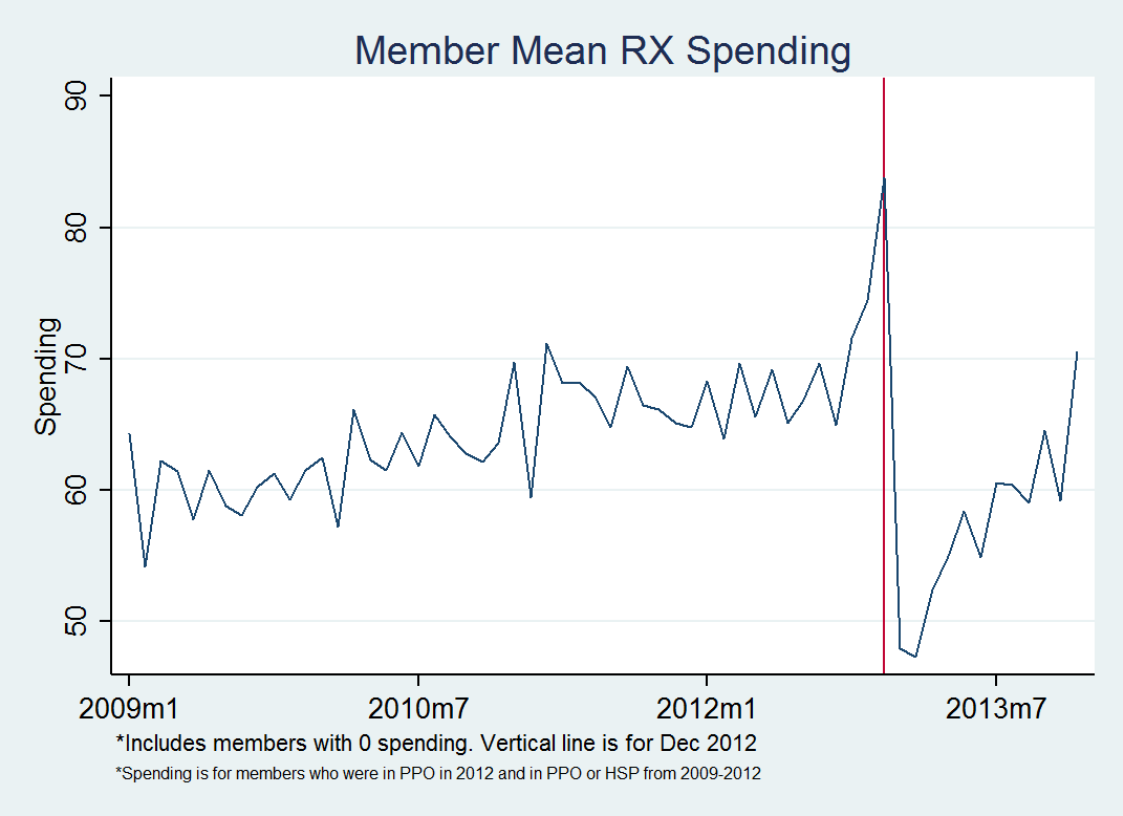



Figure A20: This figure plots adjusted spending for individuals on prescription drugs.

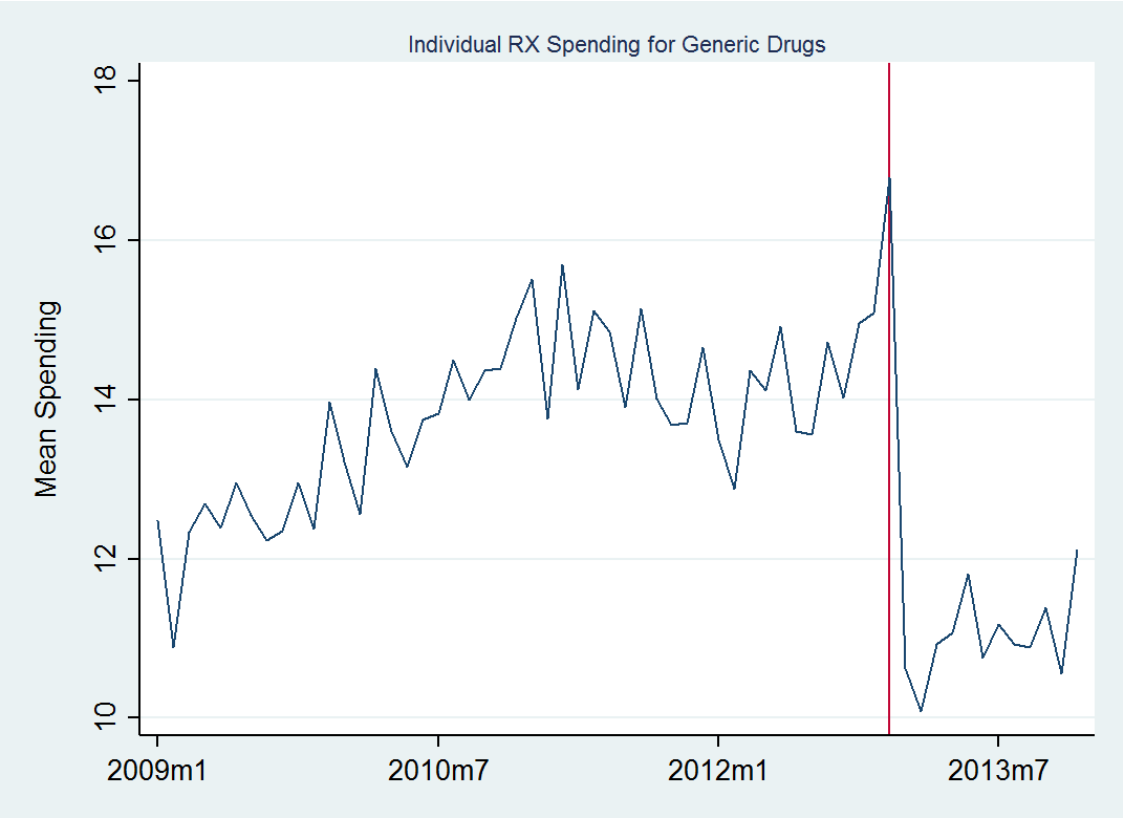



Figure A21: This figure plots adjusted spending for individuals on generic prescription drugs.

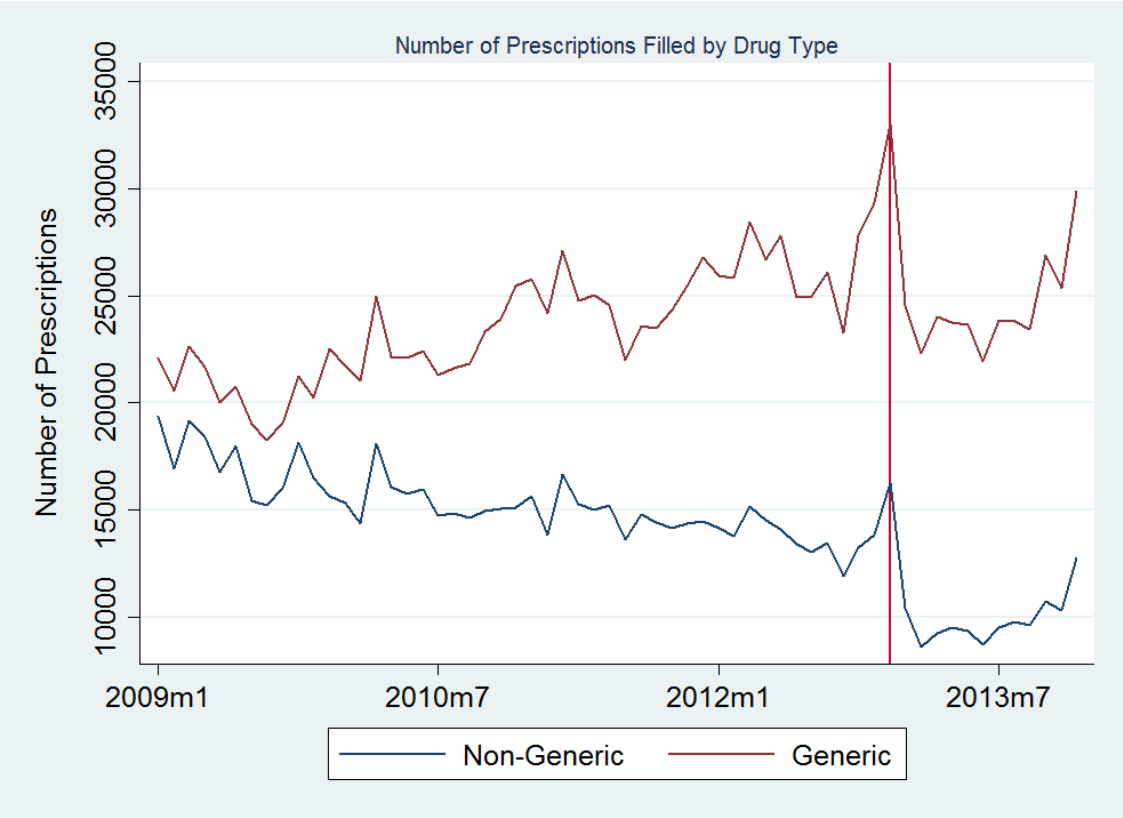



Figure A22: This figure plots the number of prescriptions filled on branded and generic prescription drugs.




Figure A23: This figure plots the proportion of individuals who spend over $10,000 (adjusted) in a single month.

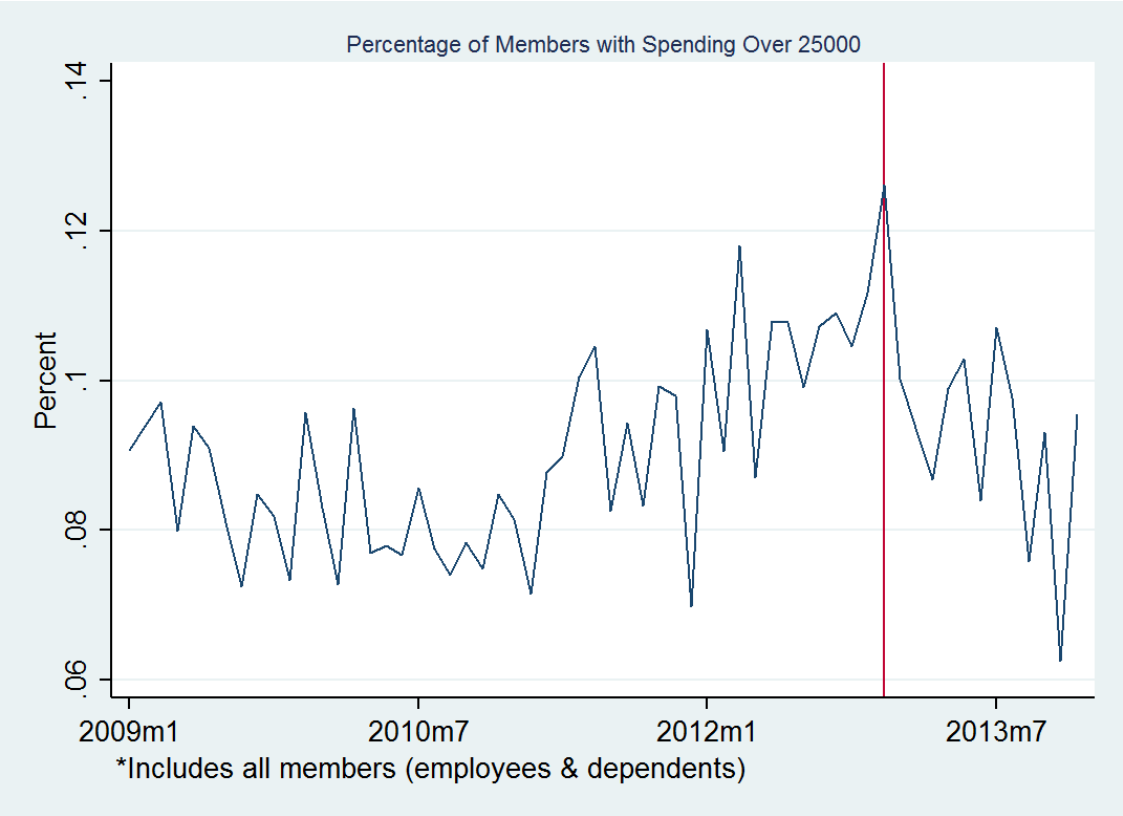



Figure A24: This figure plots the proportion of individuals who spend over $25,000 (adjusted) in a single month.

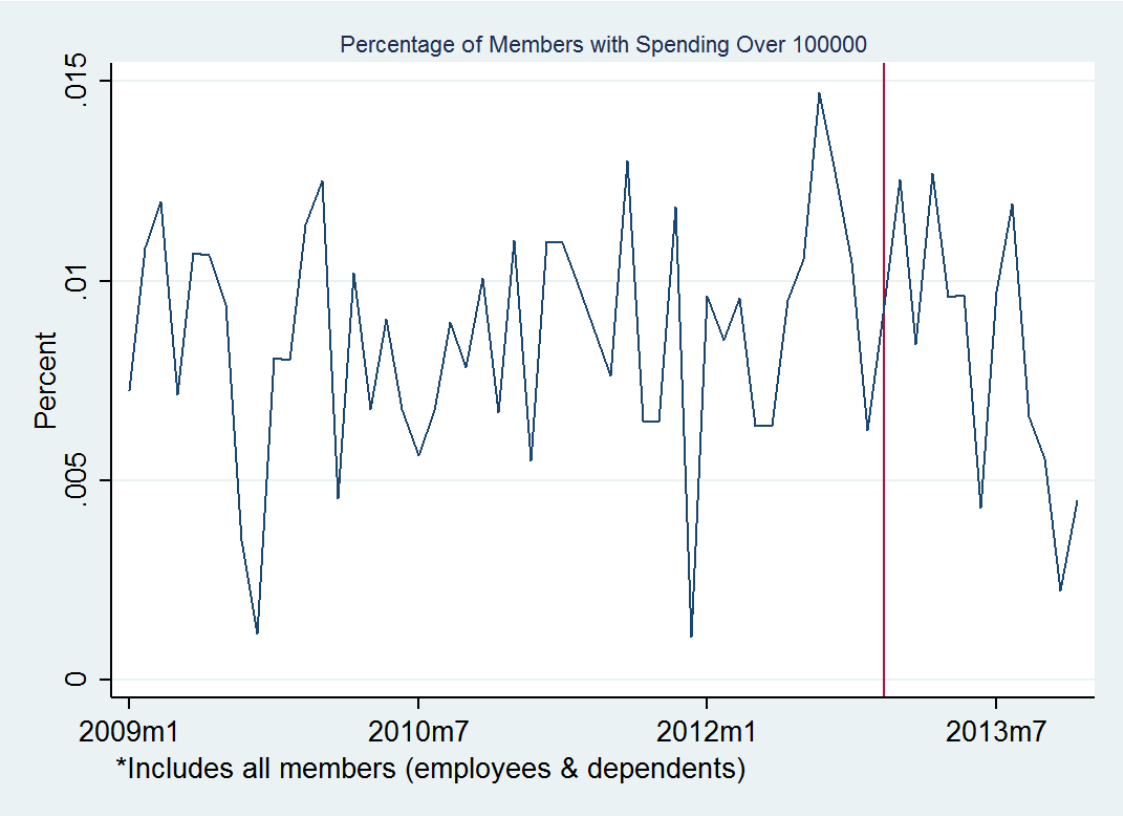



Figure A25: This figure plots the proportion of individuals who spend over $100,000 (adjusted) in a single month.

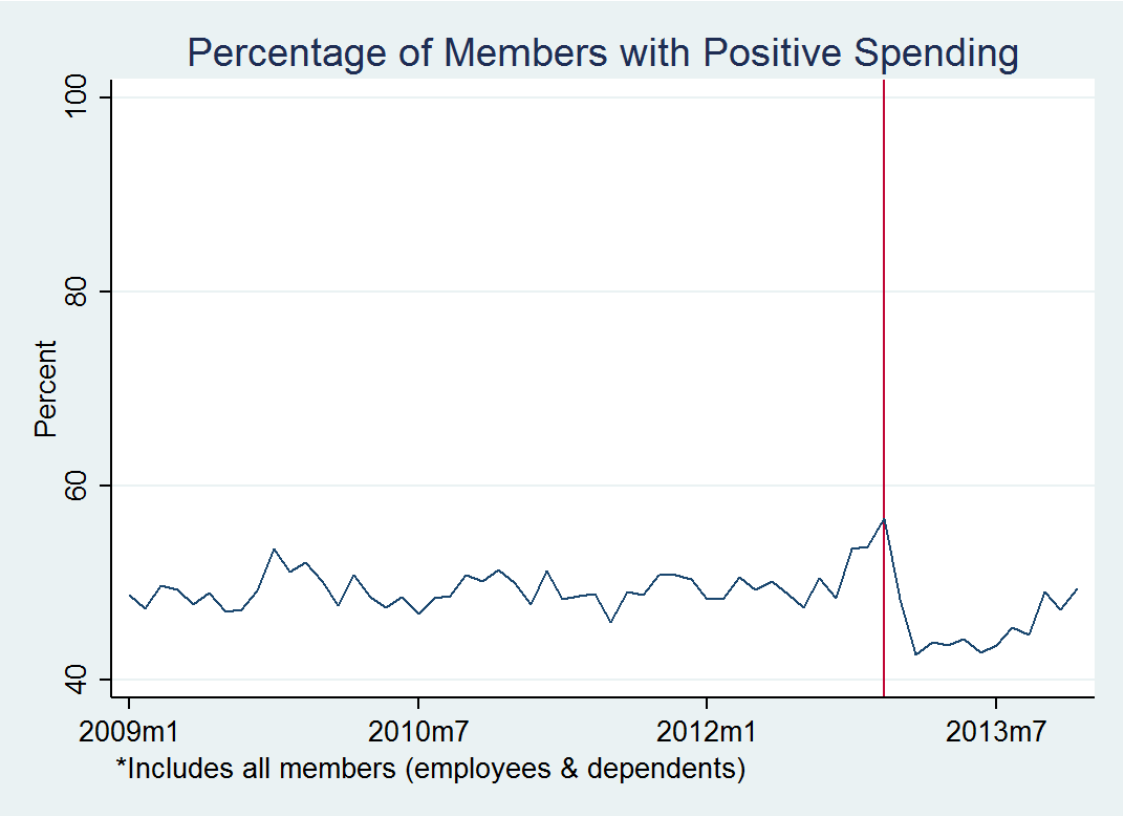



Figure A26: This figure plots the percent of individuals in a given month who spend a nonzero amount in that month.

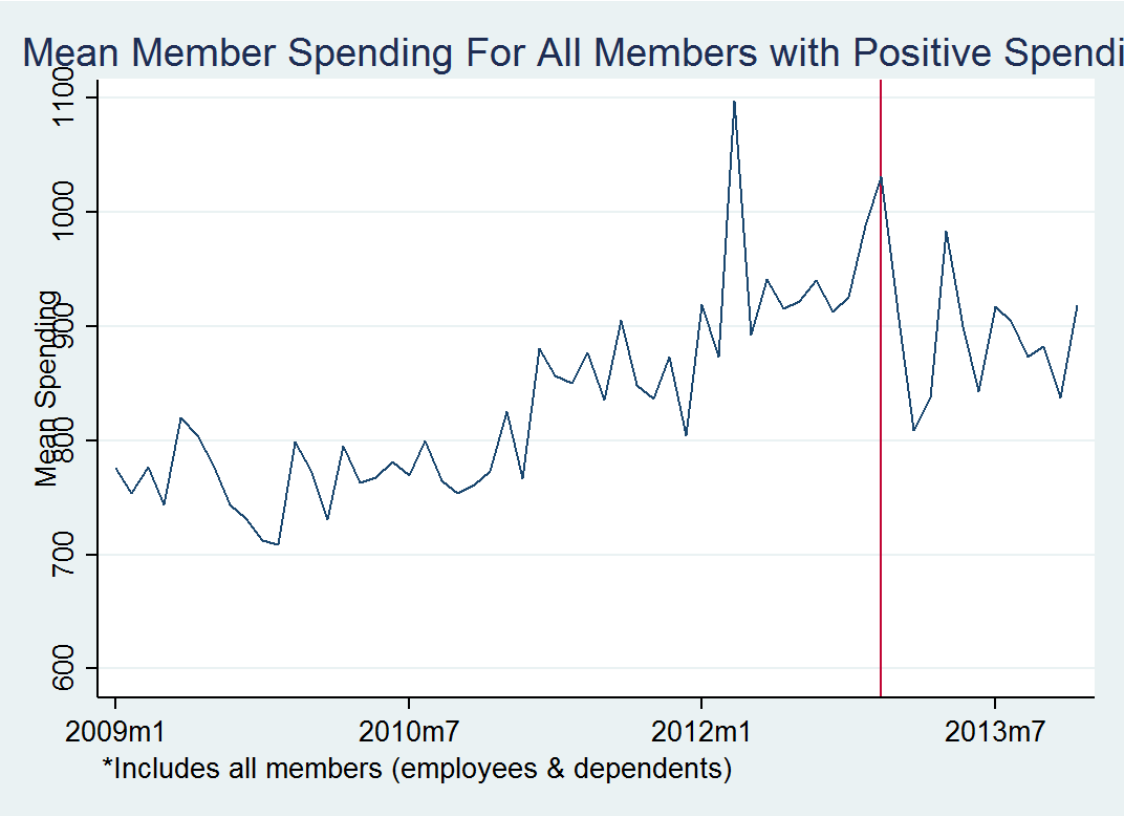



Figure A27: This figure plots spending for individuals in a given month who spent a nonzero amount during that month.