Essays in Industrial Organization

A DISSERTATION
SUBMITTED TO THE FACULTY OF THE GRADUATE SCHOOL
OF THE UNIVERSITY OF MINNESOTA
BY

Keaton Stanley Miller

IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY

Amil Petrin, Adviser

June 2015

Acknowledgements

I'd like to thank my committee, Amil Petrin, Thomas Holmes, Robert Town, and Joel Waldfogel for their support and encouragement. Thanks to Kailin Clarke, Boyoung Seo, Ethan Clarke, Thomas Youle, Thomas Quan, Dominic Smith, Matt Shapiro, Bitmaro Kim and the members of the Minnesota IO Workshop for helpful comments and suggestions. Thanks to Michael Chernew, Harvard Medical School, John Keifer and Jake Barney for data. This work was supported in part by the University of Minnesota Supercomputing Institute.

Abstract

In two essays, we examine several problems in industrial organization. In the first essay, we study the effectiveness of partially-privatized Medicare by estimating the costs that private firms face when providing care equivalent to that of the public sector. In contrast to previous studies, we take a dynamic approach, driven by the idea that consumers face large switching costs. We find that private firms face higher costs than the government after adjusting for patient characteristics and generosity of benefits. The second essay focuses on the effectiveness of U.S. merger policy by studying the acquisition behaviors of cable telecommunication companies. We construct a novel dataset of acquisitions in the cable industry from 2000-2012 and find the Hart-Scott-Rodino disclosure threshold only affects firm behavior when acquiring firms with overlapping geographic coverage areas.

Contents

List of Tables

List of Figures

Chapter 1

Do private Medicare firms have lower costs?

1.1 Introduction

Medicare, the U.S. federal health insurance program for seniors, provides a standard level of coverage for hospital and medical expenses with several well-known gaps (Moon et al., 2000). Seeking to improve the welfare of seniors by expanding the available choices to fill these gaps, Congress introduced a partial-privatization program known as Medicare Advantage (MA). Under this policy, the government offers subsidies to private insurers, who then offer Medicare-replacement plans to seniors during an open enrollment period each year. These plans are required to cover the same services as Medicare covers, but may offer some additional benefits, such as coverage for dental or vision services.[1] To help pay for these benefits, firms may charge a premium on top of the subsidy they receive from the government. Today, over 30% of eligible seniors are enrolled in a MA plan.

Proponents of the program point to the prevalence of supplemental benefits as proof that the program is welfare enhancing. Additionally, private insurers may face

[1] The Medigap program offers insurance that solely supplements traditional Medicare. In contrast, MA plans replace traditional Medicare benefits and also provide supplemental benefits.

lower costs than the government for traditional Medicare benefits through superior negotiation with local providers and "managed-care" restrictions, such as referral and network requirements.[2] Detractors have voiced concerns about the high administrative costs of the program and the relative value of the supplemental benefits offered to consumers (Pear and Bogdanich, 2003).

In this work, I study the overall welfare of the MA program taking into account the subsidies paid to firms by the government, the supplemental benefits offered, firm costs, and traditional Medicare expenditures. Currently, the subsidies are set at a level higher than Medicare's average costs. At the same time, almost all plans provide additional benefits beyond those of traditional Medicare. Therefore, to compare the programs, I must put them on an equal footing and estimate both the cost to firms of providing Medicare-equivalent benefits as well as the consumer welfare generated by the supplemental benefits.

I follow the literature[3] and employ a revealed preference approach: I construct a model of supply and demand for Medicare Advantage and estimate the model's parameters using panel data on consumers and plans from 2008-2010. I use detailed information on traditional Medicare expenditures to compare the costs of MA plans to Medicare's cost. In contrast to previous work on Medicare Advantage, I allow for the presence of switching costs on the demand side to impact firm behavior in an imperfectly competitive environment, which requires a dynamic model. To achieve tractibility, I extend the Oblivious Equilibrium concept of Weintraub et al. (2008) to models with switching costs.

I find that, on average, firms spend \$5,293 per year to insure a healthy individual and \$14,609 to insure an unhealthy individual when they provide Medicare-equivalent

[2] Medicare is organized as a Fee-For-Service system, whereas most Medicare Advantage plans are offered by Preferred Provider Organziations or Health Maintenance Organizations. The Medicare Advantage system is largely based on the idea of managed competition promulgated by Enthoven (1978).

[3] See, for example, Curto et al. (2014); Lustig (2011)

coverage. In contrast, Medicare spends an average of \$4,390 on healthy individuals and \$11,453 on unhealthy individuals. While these costs are higher than Medicare's costs, they are below the average subsidy rate (\$5,826 for healthy individuals and \$16,419 for unhealthy individuals), allowing firms to capture significant profits despite the presence of many competitors. On average, firms spend \$184 on supplemental benefits, generating an average of \$328 in welfare for individuals enrolled in their plans.

My findings are driven by the interactions between several relevant market features, including large switching costs. Miller et al. (2014, hereafter MPTC) use a comprehensive panel survey of Medicare recipients along with detailed information on plan features to estimate the demand for MA plans and find seniors who wish to switch plans incur a cost of up to \$1,300, well above the mean annual premium of \$825. These switching costs are largely driven by network restrictions: switching insurance providers often involves establishing relationships with an entirely new set of care-givers. The existence of switching costs implies that a firm's current customer base is a large determinant of its future profits (Farrell and Klemperer, 2007). Firms therefore face an intertemporal tradeoff between "locking in" customers with attractive but relatively unprofitable plans and "harvesting" their market share with unattractive but more profitable offerings.[4]

In addition to switching costs, the MA market is characterized by a high degree of heterogeneity in health among seniors, which, when combined with a requirement that firms offer the same products at the same prices to all seniors, leads to adverse selection: firms have an incentive to enroll the most profitable individuals through careful management of plan prices and benefits. To address this incentive,

[4]By default, seniors are automatically re-enrolled in their current plan if they take no action during the open enrollment period. While there are restrictions on the changes firms can make to their set of offered plans each year, they essentially amount to a requirement that the rank-ordering of plans with respect to price stays constant. In other words, a firm can raise prices on its enrollees from year-to-year and those enrollees must take action in order to avoid these price increases.

the Medicare Advantage system adjusts the payments offered to firms based upon the characteristics of each of their enrollees. These adjustments, however, are based upon Medicare's expenditures on individuals with similar characteristics. If firm costs differ from Medicare's costs unevenly, firms will still have an incentive to enroll particular groups of individuals.[5]

Finally, unlike the highly regulated Medigap program, MA firms may offer a wide variety of supplemental benefits on top of the minimum "package" required by law. Firms may set a number of parameters for their plans, ranging from the number of days patients are allowed to spend in the hospital per year to the copay for primary care visits. Most firms offer more than one plan with a range of prices and characteristics to appeal to individuals with different preferences. Since these plans generally use the same network, firms must worry about potential self-cannibalization: if a lower priced plan is too generous, consumers may switch.

My model incorporates these features into a dynamic environment with imperfect competition. In the model, consumers are described by their health status and a number of demographic characteristics. They face a discrete choice between a number of different plans defined by a price and a generosity index. Choices that require consumers to change insurance providers incur a switching cost. Firms choose product characteristics for multiple plans and take into account the dynamic incentives generated by the switching costs. Firms face marginal costs and subsidies that vary according to the demographics and health of their enrollees.

I estimate the model in two steps. First, I estimate the demand-side parameters largely following the procedure outlined in MPTC. which in turn is based on the discrete choice literature (Berry, 1994; Berry et al., 1995; Goolsbee and Petrin, 2004).

[5] For example, suppose there are two health conditions in the market. Condition A costs firms \$200 to treat and condition B costs firms \$500 to treat. The risk-adjusted payments to firms for individuals with conditions A and B are \$250 and \$700, respectively. Since condition B is more profitable for firms, they have an incentive to use whatever information they have about demand to attract individuals with condition B.

In the second stage, I form moments from the model's predictions of optimal firm behavior and compare them to their sample equivalents to estimate the parameters of the firms' cost function (Newey and McFadden, 1994). I take advantage of the fact that the government pays firms different subsidies for similar individuals in different counties to assist in identifying county-level parameters.

My cost estimates are driven by the combination of factors, including the price elasticity of demand and the size of the switching costs. In some sense, the existence of switching costs creates a group of consumers over which the firm faces far less competition. In the preferred specification, these factors combine with the intertemporal dynamics to determine the markup the firm can set. The preferred estimates differ significantly from those obtained with a static version of my model. If switching costs are ignored on the demand side, the estimated costs are up to 20% lower than the preferred specification. This result is largely driven by the low enrollment rate: a majority of seniors in my sample never enroll in a Medicare Advantage plan. The absence of switching costs biases the rest of the demand system and generates a much lower mean price elasticity. This in turn implies the premiums I observe are marked up well above firm costs.

On the other hand, if switching costs are included but firms are assumed to be myopic, the estimated costs are up to 25% higher. This is driven by the connection between market share and firm behavior. When firms are myopic, they will freely charge high prices without regard to what effects those prices will have on their future market power. The inclusion of the dynamic incentive dampens that connection.

In a counterfactual simulation, I explore the impact of a 50% reduction in the switching costs incurred by consumers. This increased "fluidity" in the demand system implies that firms with a low market share don't need to offer plans that are as attractive as the baseline to win customers. In equilbrium, firms with larger shares respond to this change by lowering the attractiveness of their plans as well – as the competitive pressure they face has weakened. As the plans in the market become less attractive, consumers respond by leaving – the overall Medicare Advantage enrollment rate is almost cut in half. The exodus is biased: the consumers that remain in Medicare Advantage are less healthy than those in the baseline scenario, which leads to higher mean per-enrollee profits for firms. Given the relatively low value consumers place on the supplemental benefits offered by MA plans, the change increases overall welfare by 8%, driven by the movement of consumers from expensive MA plans to cheaper Medicare Fee-for-Service (FFS) benefits.

A second counterfactual explores the effect of a 5% reduction in the subsidies offered to firms, similar to the policy included in the Affordable Care Act (Kaiser Family Foundation, 2014). The change in subsidies implies that zero-premium plans are no longer profitable for firms. Firms stop offering these plans and reduce the attractiveness of their remaining plans to compensate for the reduced subsidies. These effects combine to produce a drastic reduction in MA enrollment, consumer welfare and total firm profits. These reductions, however, are more than offset by a large reduction in total Medicare expenditures.

I simplify firms' information and action sets to achieve computational tractibility. Instead of explicitly tracking the full states and actions of each of their competitors, firms keep track of the average 'competitive environment' they face. This approach is similar to the Oblivious Equilibrium concept of Weintraub et al. (2008); though I allow for a continuous state space and firms in the model keep track of detailed information on the effect of competitors' actions on individuals across the demographic spectrum.[6] Instead of modelling each feature of plans, I follow Lustig (2011) and use a generosity index to measure the relative desirability of different combinations of benefits.

This paper contributes to the literature on private Medicare plans and competition

[6]Qi (2013) uses an alternative modification of the Oblivious Equilibrium concept to estimate a model of dynamic oligopoly in cigarette advertising.

within health care, an area reviewed by Gaynor and Town (2011). Curto et al. (2014) also estimate costs for MA firms. They estimate a logit demand model and firm costs under the assumption that the subsidy system perfectly captures the relative heterogeneous risk of enrollees. Duggan et al. (2014) study the effect of changes in the subsidies offered to firms in particular metropolitan areas. Aizawa and Kim (2013) study the effect of advertising on demand for MA plans in a static setting, taking into account the risk-adjustment system.

The concept of switching costs I use is broadly related to the literature on consumer inertia in health. Ho et al. (2014) study consumer inertia in the Medicare Part D (prescription drug) market and calculate a static counterfactual environment in which inertia is removed. They find, as I do, the elimination of interia would lead to substantial savings for both consumers and the government. Abaluck and Gruber (2013) also study Medicare Part D and focus on decomposing observed consumer inertia into demand- and supply-side factors. They find little improvement in the ability of consumers to choose plans over time. Handel (2013) studies the interaction between inertia and selection in employer-provided insurance and finds that a reduction in inertia leads to increased selection. Cebul et al. (2011) study search frictions in the commercial insurance market and find that frictions increase premiums and insurance turnover.

My results on the relative profitability of different types of individuals are similar to those of Brown et al. (2011), who use data on Medicare expenditures to understand the changes in incentives brought about by the introduction of the risk adjustment system used by the government to compute Medicare Advantage payments. They find the risk adjustment system significantly increased the profitibility of unhealthy people and led firms to change their selection patterns. Newhouse et al. (2014) futher examines the current behavior of Medicare Advantage plans and finds evidence of selection.

Finally, I contribute to the empirical study of dynamic firm behavior in environments with endogenous product characteristics, an area reviewed by Crawford (2012). My estimation procedure allows firms to simultaneously choose characteristics of a "line-up" of plans each period. Many recent studies of firm behavior[7] use two-step estimators in which policy and transition functions are estimated semi- or non-parametrically and structural parameters are recovered separately;[8] however, given the number of covariates I include, my sample is too small to reliably estimate the transition matrix. Instead, I compute the information sets of firms directly from the data and solve a single firm problem.

The remainder of the paper is structured as follows. Section 2 provides a brief history and description of the Medicare Advantage program. Section 3 introduces my model of dynamic competition. Section 4 describes my data. Section 5 details the empirical implementation of my model including details of estimation, counterfactual, and computation. Results are described in section 6. Section 7 concludes.

1.2 Medicare Advantage

Medicare was created in 1965 to address the lack of health insurance among senior citizens.[9] While the original law provided basic hospital (Part B) and medical (Part B) insurance to seniors (age 65 or older), reforms have slowly expanded Medicare's role within the U.S. health care system. Eligibility was extended to individuals under 65 with certain disabilities and illnesses. The range of services covered by the program increased. These expansions had a serious impact on the cost of the program. In 1970, Medicare composed about 0.5% of GDP. By 1980, Medicare had more than doubled

[7]For examples, see Ryan (2012); Collard-Wexler (2013); Youle (2014)

[8]For examples, see Hotz and Miller (1993); Bajari et al. (2007); Aguirregabiria and Mira (2007); Pakes et al. (2007)

[9]Much of the historical information in this section is compiled from `http://www.cms.gov/`

in size to 1.1% of GDP.[10]

This growth led policy-makers to begin experimenting with different cost-containment and care-delivery strategies in the 1980s. While many efforts focused on broad reforms, such as changing the way Medicare reimbursed care-givers, the Centers for Medicare and Medicaid Services (CMS, then known as the Health Care Financing Administration) began a series of limited trial programs based in part on the ideas of Enthoven (1978) in which the government contracted with Health Maintenance Organizations (HMOs) to manage the care of select groups of enrollees.

HMOs, which had become popular after the passage of the Health Maintenance Organization Act of 1973, provided health care to their customers under a fundamentally different model. Previously, most health insurance was operated under a Fee-for-Service (FFS) model, in which doctors charged patients and insurers for each individual service performed. Given the level of asymmetric information present in the doctor-patient relationship, many feared the system made it too easy for doctors to perform unnecessary procedures (Arrow, 1963; Chernew, 2003). HMOs changed that by signing pre-paid contracts with physicians and hospitals (Markovich, 2003).

Today, HMOs take advantage of a number of other components of the so-called managed care model (Glied, 2000). Patients are generally required to see a primary care physician for a referral before they can visit a more expensive specialist. Preventative care is usually provided at little-to-no charge to enrollees with the idea that regular checkups can detect illnesses before expensive procedures are necessary. HMOs regularly review the performance of their doctors to ensure the number of services they each perform are in line with expectations. Preferred Provider Organizations (PPOs) have arisen as a slightly more flexible but more expensive alternative (Gabel et al., 1988).

[10] See `http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/NationalHealthExpendData/NationalHealthAccountsHistorical.html` for information on health care spending in the U.S.

The Balanced Budget Act of 1997 expanded and formalized the Medicare managed care program into Part C. The new program closely followed common models of employer-provided health insurance and had several key components. Each spring, firms submitted county-by-county plan proposals to CMS. CMS verified that the plans met the minimum requirements and covered the same conditions as Medicare, though many plans chose to offer additional benefits. In the fall, CMS operated an open-enrollment period, during which Medicare recipients could freely choose between original Medicare or any of the plans available in their county of residence. Coverage began on January 1st, and firms received a flat subsidy from CMS, known as a capitation payment, each month (CMS, 2014).

After several enrollment periods passed, policy-makers grew concerned about the direction of the program (Pear and Bogdanich, 2003). Entry by firms was largely limited to suburban areas and many rural and inner-city residents did not have access to the new plans. Additionally, since the capitation payment was the same for all enrollees, firms had an incentive to tailor their plans to appeal to only the healthiest (and therefore most profitable) consumers. Enrollment hovered around 15% of eligible seniors, or six million people (Kaiser Family Foundation, 2014).

The Medicare Prescription Drug, Improvement, and Modernization Act of 2003 sought to address these concerns by reforming Part C. Plan providers were given new flexibility to manage the care of their enrollees – particularly with respect to the provision of non-emergency care. The subsidies offered to plans were significantly increased to encourage entry in more geographic areas. Finally, the reimbursement system became risk-adjusted. Under the new system, firms submit demographic and diagnostic information about their enrollees each month. CMS "scores" each enrollee's risk according to the cost of similar individuals enrolled in the traditional Medicare system – the average senior has a risk score of 1.0. CMS sets a benchmark rate for each county and multiplies this rate by the individual risk score for each enrollee to determine the subsidies paid to firms (CMS, 2014).

The changes have had the desired effects: now almost all seniors have the option of at least one Medicare Advantage plan and most can choose between two or more (in addition to traditional Medicare). Studies of the risk adjustment system have concluded that it effectively reduced the tendency of firms to prefer healthy enrollees (Brown et al., 2011). Enrollment surged and today over 30% of seniors, almost 16 million people, are enrolled in a Medicare Advantage plan (Kaiser Family Foundation, 2014).

1.3 Model

Taking these institutional details into account, I build a model of supply and demand for Medicare Advantage. On the demand side, heterogeneous consumers face a discrete choice of plans described by a price and a generosity index. They incur a cost if they switch insurers. Following Handel (2013), I assume consumers are myopic and do not take into account future switching costs.[11] On the supply side, symmetric firms simultaneously choose the price and generosity of multiple plans. They solve a recursive value function that takes into account the dynamic tradeoff generated by the switching costs.

I simplify the problem to ensure that my estimation exercise is tractible while maintaining a degree of flexibility in my specifications of utility and firm costs. In particular, I allow firms to keep track of their market shares of healthy and unhealthy individuals (as opposed to allowing firms to keep track of the full, joint distribution of enrollee characteristics). Firms calculate their per-enrollee profits from a distribution over consumer characteristics conditioned on their health status. This allows firms

[11] In essence, allowing this would require consumers to forecast future prices – meaning consumers would have to keep track of who will purchase which plans – and their own future health statuses over a potentially lengthy time-frame Handel (2013).

to keep track of the dimension most correlated with cost while keeping a relatively low-dimension state space. while firms face a recursive problem – which allows firms to alter their markups based upon the switching costs faced by consumers – entry and exit is assumed to be exogenous.[12] While firms store summarized market information – which allows firm behavior across markets to differ based on the degree of competition – they do not keep track of the details of each of their competitors.

1.3.1 Environment

Time is discrete and denoted by t; each period represents a year. The world is divided into a number of discrete markets (representing individual U.S. counties) denoted by m. Each market contains N_m consumers, with individual consumers denoted by i. Each market also has a vector of observables Y_m, including the benchmark subsidy rate B_m.

In each period, for every market, there is an independent set of firms F_m^t, determined exogenously. Individual firms are denoted by f, and these firms offer plans in a set J_f denoted by j, where the size of the set is determined exogenously. Each plan j consists of a premium p_j and a generosity index g_j. Firms enter each period with a market share of healthy people s_{fh} and a market share of unhealthy people s_{fu}. In all markets, the outside good, good 0, is Medicare, and the set of all plans within a market is denoted J_m.[13]

[12]In my modification of the Oblivious Equilibrium concept (explained in detail below), firms keep track of the competition they face in terms of the utility of competing plans. In some sense, firms don't care who those products are offered by (or how many competitors there are) – just the level of competition in the market.

[13]There are two other substantial components of the post-65 insurnace system: Private plans provided by employers as part of a retirement package or pension system, and Medigap, which provides supplementary insurnace on top of traditional Medicare benefits. As individuals with employer-provided plans have very low Medicare Advantage enrollment rates, I treat the employer-based system as a separate entity and remove those individuals from the market. Since Medigap plans are highly regulated and consistent across geographies, I abstract from the variance in those plans and consider them part of the outside good.

1.3.2 Timing

In each period t, actions are taken as follows:

1. Firms observe their information set I_f^t, described in detail below.

2. Firms simultaneously choose actions $\sigma_f = \{p_j, g_j\}_j$.

3. Consumers choose plans from J_m.

4. Profits and the next state of the market are realized.

1.3.3 Consumers

Individual i has a vector of demographic characteristics Z_i which includes age, gender, health status, education, race, and current plan enrollment. Z_i has distribution C_m. Additionally, each individual is assigned to one of several income groups, represented by the dummy variables d_{wi}.

Consumer i considering plan j faces the following choice-specific utility function:

$$\begin{aligned} U_{ij} = & \alpha_0 p_j + \sum_w \alpha_w p_j d_{wi} + SW_k * 1\{\text{switch}_{ij}\} + \beta_z Z_i + \beta_g g_j \\ & + \beta_{zg} Z_i g_j + \xi_j + \epsilon_{ij} \end{aligned} \tag{1.1}$$

In this equation, $\alpha = \{\alpha, \alpha_w\}$ represents income-specific price sensitivity. β_z captures tastes for the inside good that vary by demographic characteristics. β_g captures the mean taste for generosity and β_{zg} captures demographic specific tastes for generosity. $SW_k * 1\{\text{switch}_{ij}\}$ represents the cost SW_r that consumer i must pay if j is offered by a different firm than its current plan. SW_k is allowed to vary across switch types: switching between traditional Medicare and Medicare Advantage may incur a different cost than switching between different Medicare Advantage providers. Finally, ξ_j

represents the component of plan quality that is unobserved to the econometrician, and ϵ_{ij} represents the individual choice-specific unobservable, which is assumed to be independently drawn according to a Type-I extreme value distribution.

Following Berry et al. (1995), I can decompose the utility obtained from good j into a mean:

$$\delta_j = \alpha_0 p_j + \beta_g g_j + \xi_j$$

and an individual specific deviation:

$$\mu_{ij} = \sum_w \alpha_w p_j d_{wi} + SW * 1\{\text{switch}_{ij}\} + \beta_{zg} Z_i g_j + \epsilon_{ij}$$

Given a set of plans J_f, the probability that consumer i chooses plan j (often known as the share function) can now be written as:

$$Pr(i \text{ chooses } j) = s_{ij} = \frac{\exp(\delta_j + \mu_{ij})}{1 + \sum_{j' \in J_m} \exp(\delta_{j'} + \mu_{ij'})} \tag{1.2}$$

1.3.4 Firms

As mentioned previously, firms choose prices p_j and generosity g_j for some number of plans J. The firm's problem requires the firm to evaluate the expected profits of different combinations of prices and generosities as a function of the information they have about the market. I develop this problem in stages, starting with the firm's per-enrollee cost function. Computational restrictions prevent me from considering a full-information model. I therefore adapt the Oblivious Equilibrium concept of Weintraub et al. (2008, 2010) to this environment and define the information set of firms based on what they need to calculate demand and cost and, thus, profits. The translation is made more complicated by the inclusion of switching costs on the demand side and therefore the need for firms to track their market shares.

With the information set established, I describe the market timing and write down the firm's problem as a recursive value function.

Cost function

In order to evaluate the outcomes of various actions, the firm must first evaluate the cost of providing insurance. The per-capita expected cost of providing plan j to individual i is a function of the generosity index g_j and is conditioned on the demographic characteristics Z_i and market characteristics Y_m, such as the population density, the doctor/population ratio, and the average per-capita income:

$$c_{ijm}(g_j|Z_i, Y_m) = (\gamma + \gamma_m Y_m) + \gamma_z Z_i + (\gamma_g + \gamma_{gm} Y_m) g_j + \gamma_{gz} Z_i g_j + \gamma_{g2} g^2 + \zeta_f \quad (1.3)$$

In this equation, γ represents the average marginal cost of providing Medicare-level benefits and γ_m represents deviations due to market-specific factors. γ_z represents the deviations in marginal cost due to consumer demographics (e.g. individuals in poor health cost more to insure). Similarly, γ_g is the mean marginal cost of generosity (with γ_{gm} capturing market-specific factors) and γ_{gz} captures the deviations from that mean due to consumer demographics. I allow for the possibility of a quadratic component to the cost of generosity.[14] Finally, firms receive an i.i.d. cost shock ζ_f each period.

Information set

The contents of the information set of firms is driven by the need (of the firms) to calculate the expected profits for a given action and therefore the need to calculate demand and costs. The demand for a particular plan j can be calculated by integrating

[14] Intuitively, if a more generous plan makes individual claims more attractive to the consumer, the consumer will make additional claims, each of which may cost the firm more money. This quadratic component provides additional curvature in the firm's value function which is particularly helpful during the estimation procedure.

out the share function s_{ij} over consumers:

$$s_j(p_j, g_j) = \int_i s_{ij}(p_j, g_j, Z_i)dC(Z_i)di$$

Since costs may be particularly correlated with health, firms can individually calculate the fraction of healthy and unhealthy people who enroll in the plan:

$$s_{jh} = \int_i s_{ij}(p_j, g_j, Z_i)dC(Z_i|h)di$$
$$s_{ju} = \int_i s_{ij}(p_j, g_j, Z_i)dC(Z_i|u)di$$

These plan-level shares can be aggregated into firm-level shares: $s_f = \sum_j s_j; s_{fh} = \sum_j s_{jh}; s_{fu} = \sum_j s_{ju}$. For simplicity, I drop s_{fu} and consider only s_{fh} – the same calculations are made for s_{fu}.

To compute the numerator of equation (1.2), the firm must know how its products map into the utility obtained by consumers $\delta_j + \mu_{ij}$. Since consumers who are currently enrolled in other MA firms or original Medicare face switching costs when considering the firm's products, the firm must know its own shares s_{fh} and how many consumers are enrolled in traditional Medicare versus other MA plans. To calculate the denominator of equation (1.2), the firm must know something about the other products available in the market.

The classic dynamic oliogoply approach would be to allow firms to observe the full state space and action set of their competitors in the manner of Ericson and Pakes (1995). Equilibrium would be defined as a set of policy functions that obtained the supremum of the recursive value function. As Medicare Advantage markets often have more than 10 incumbents, computational limitations prevent me from calculating optimal firm actions as a function of all of their competitors' states and actions.[15]

[15] Even with recent improvements in solution techniques for Markov-Perfect Nash Equilibria, such as the stochastic algorithm introduced by Pakes and McGuire (2001), the number of unique states

To reduce the computational burden, I adopt the Oblivious Equilibrium solution concept developed by Weintraub et al. (2008, 2010). However, there are key differences between their model and mine that I must address.

In the Weintraub et al. (2010) version of the Ericson and Pakes (1995) dynamic oligopoly model, consumers are identical and the state space of firms is discrete, representing (depending on interpretation) their level of capital or efficiency. Weintraub et al. simplify the information set of their firms by defining a "long-run average state" vector $\bar{s}$, which is the expected number of firms (which may be fractional) with each state n at any given period in equilibrium. The components of $\bar{s}$ can be calculated as:

$$\bar{s}_n = E_t\left[\sum_f 1\{f \text{ in state } n\}\right]$$

Weintraub et al. write down their firm's problem as a function of the firm's own state, and condition on this long-run average state. In their logit application, they use this state to calculate the share s_j the firm receives as a function of the price the firm charges for its good. Recall that, for the generic logit model:

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_{j'} \exp(\delta_{j'})}$$

Since firms in their model are symmetric, which implies that each firm with the same state n will offer a good with the same $\delta_j = \delta(n)$, they can rewrite their share function using the components $\bar{s}_n$ of $\bar{s}$:[16]

$$s_j = \frac{\exp(\delta_j)}{1 + \sum_n \bar{s}_n \exp(\delta(n))} \tag{1.4}$$

visited by the market is simply too large.

[16]For ease of exposition, in the following discussion I abstract away from the effect that firm j has on the denominator of the share function. My implementation of the information set described in this section involves a small modification to $\bar{q}_i$ to remove the effect of one firm and an opposite adjustment to s_{ij} to explicitly include the effect of product j.

Unfortunately, consumers in my model are heterogenous, and have a utility composed of $\delta_j + \mu_{ij}$. However, a modification addresses the problem. Instead of taking the expected state of firms, I take the expected impact (where the expectation is over time) of the products they offer on the share equation for i:

$$\bar{q}_i = E_t\left[\sum_j \exp(\delta_j + \mu_{ij})\right] \tag{1.5}$$

In other words, given some belief about the future actions of competitors, the firm constructs $\bar{q}_i$ for each individual by summing their exponentiated utility terms for each product in every period in the future and then taking the average over all periods.[17] In the traditional OE setup, these beliefs come from the optimal strategies of firms in different states. In my estimation exercise, I form this expectation by taking the observed actions of firms in my sample period. In my counterfactual, I follow Weintraub et al. (2010) and solve for a self-consistent set of firm strategies and beliefs.

As this expression is essentially the denominator of the traditional logit share function, the share function used by the firm becomes:

$$s_{ij} = \frac{\exp(\delta_j + \mu_{ij})}{1 + \bar{q}_i} \tag{1.6}$$

In practice, I calculate integrals over the distribution of individuals by taking a number of discrete draws from the conditional dC distributions. I form a vector $\bar{Q}$

[17] This expression plays the same role in my model as equation (4.1) of (Weintraub et al., 2008, p. 1386) does in the traditional OE setup.

by calculating q_i for each draw:

$$\bar{Q} = \{\bar{q}_1, \bar{q}_2, \cdots \bar{q}_n\} \tag{1.7}$$

The interpretation of $\bar{q}_i$ is slightly different from $\bar{s}$. $\bar{q}_i$ is essentially the denominator of the share function and in some sense measures how attractive of a product firm j must offer (through δ_j) in order to achieve market share. As $\bar{q}_i$ increases, the firm must somehow increase δ_j in order to achieve the same share. Therefore, I call $\bar{Q}$ the *expected competitive pressure* of the market.[18]

Finally, as mentioned previously, the response of consumers to a product described by p_j and g_j depends upon their current enrollment since they may be subject to a switching cost if they choose plan j. The firm does not need to know where each consumer is – merely whether or not they are enrolled in original Medicare, a competitor's MA plan, or in one of the firm's own plans.[19] The firm keeps track of it's own shares by health status s_{fh}. Additionally, the firm knows the average share of individuals enrolled in any MA plan $\bar{s}_h = E_t[\sum_f s_{fh}]$. With these numbers, the firm can calculate the number of individuals enrolled in a different MA plan and the number of individuals enrolled in original Medicare. If $e \in \{0, 1, 2\}$ represents enrollment in original Medicare, a competitor's plan, and one of firm's own plans respectively,

[18]Calculating $\bar{Q}$ in this way also addresses the fact that the state space in my model is continuous.

[19]Since there is no switching cost between plans within a firm, the demand will be the same across individuals enrolled in all of a single firms' plans and therefore firms only need to keep track of shares at the firm level.

the firm can calculate $S_{e,h}$ as:[20]

$$S_{2,h} = s_{fh}; S_{1,h} = \bar{s}_h - s_{fh}; S_{0,h} = 1 - S_{1,h} - S_{2,h}$$

The firm can therefore calculate it's share function with the modified s_{ij} of equation (1.6) as

$$s_j = \sum_{e,h} S_{e,h} \int_i s_{ij} dC(Z_i|e,h) di \tag{1.8}$$

The firm knows $\bar{Q}$, the conditional distributions of consumers $dC(Z_i|e,h)$, which is invariant over time, and the average share of individuals enrolled in MA plans $\bar{s}_h$. Each period, the firm observes an information set that includes:

$$I_f = \left\{ s_{fh}^t, \zeta_f^t \right\} \tag{1.9}$$

Profit function

The firm forms expected plan-level profits by integrating over the conditional distributions of consumers. Plan-level profits are a function of the plan characteristics under consideration by the firm, and are conditioned on the information set of the firm. Firms receive a risk-adjusted subsidy for each consumer based upon the market-level benchmark and the consumer's individual risk score r_i, where $r_i = f(Z_i)$ is a

[20] It is possible that $s_{fh} > \bar{s}_h$, in which case $S_{1,h}$ will be negative. While this situation does not occur in the data, I must solve a value function across the entire state space, and the continuation values obtained for states in which this occurs will affect the value function in all other states. In the results presented below, I bound $S_{1,h}$ from below by 0 and adjust $S_{0,h}$ accordingly. Other methods of handling this situation, such as allowing negatives, do not substantially change the results.

function of the consumer's demographic characteristics:

$$\pi_j(p_j, g_j|I_f) = N_m \sum_{e,h} S_{e,h} \int_i (p_j + Br_i - c_{ij})\, s_{ij}\, dC(Z_i|e,h)di \tag{1.10}$$

In other words, the firm calculates plan-level profits by considering six different groups of individuals: healthy and unhealthy people who are enrolled in traditional Medicare, a competitor's plan, or one of the firm's own plans. For each group of people, the firm knows the distribution of their demographics, conditional on belonging to that particular group, as well as the size of that group in the market. The firm uses the $\bar{Q}$ vector to calculate s_{ij} for each individual across the conditional distribution. The firm also uses the the conditional distribution to calculate the cost incurred and subsidy obtained by insuring a particular individual.

Firm's problem

With these ingredients in hand, I can formulate a recursive value function for firms. Their value is a function of their current share of consumers across health statuses and is conditioned on their information set:

$$V(I_{fm}) = \max_{\sigma_f} \sum_j \pi_j(p_j, g_j|I_{fm}) + \beta E_{I'} \left[V \left(I'_{fm}(\sigma_f) \right) \right] \tag{1.11}$$

In this recursive problem, β is the discount factor, which is constant across firms. Dynamics are embedded in the evolution of I_{fm}. In particular, I_{fm} contains the firm's market shares and cost shock as well as the proportion of individuals enrolled in the Medicare Advantage system.

1.3.5 Equilibrium

I adopt the Oblivious Equilibrium notion of Weintraub et al. (2008) to this case. An equilibrium in my model is an integral number of firms F, a firm strategy σ and competitive pressure vector $\bar{Q}$ such that:

1. Given $\bar{Q}$, σ is the solution to equation (1.11).

2. When F firms play according to σ, $\bar{Q}$ satisfies equation (1.7).

1.4 Data

MPTC construct a comprehensive dataset of the MA market to estimate a detailed demand system. I employ this dataset with a few modifications to align with my model of firm behavior. Broadly speaking, my data falls into three categories: plans, consumers, and geographies. In this section, I briefly describe each of these categories with a focus on the differences between the data used in this paper and the data used in MPTC.

1.4.1 Sample selection

I restrict the temporal and geographic spread of my sample due to the needs of my equilibrium notion and estimation procedure. I employ a variation of the Oblivious Equilibrium concept of Weintraub et al. (2008) which imposes a notion of stationarity.[21] Additionally, my estimation procedure requires demographic-specific measures of market share. To satisfy these requirements, I select 39 markets throughout the United States where I observe a reasonable sample of individuals and do not observe significant changes in the total number of firms and plans present in the period I

[21] In particular, I must assume that the data I observe is drawn from the ergodic set of market states generated by the OE strategies.

study. Additionally, I restrict my attention to plans offered starting January 1, 2008, two years after the implementation of the current Medicare Advantage system.

1.4.2 Consumers

My data on individual consumers comes from the Medicare Current Beneficiary Survey (MCBS), an overlapping-panel survey of a nationally representative sample of Medicare recipients sponsored by the Centers for Medicare and Medicaid Services (CMS) and produced by Westat. I use the Cost and Use files from the 2007-2010 data releases to obtain data on individual plan choices[22] and demographic characteristics including age, race, education, and income. I use a self-reported health status variable to separate individuals into healthy and unhealthy categories.[23]

Summary statistics on the individuals used to produce my estimates are in Table 1.1. Despite selecting on the 39 counties with the highest sample available, individuals in my subset are similar to the fuller sample of MPTC, if very slightly richer (the average income in the full sample is $43,378 as opposed to $46,198 in my sample). The second column reports the standard deviation of the yearly means for each of the demographic characteristics I include. In particular, the demographic distribution does not shift much between periods. Though the nationwide Medicare Advantage enrollment rates increased significantly throughout the period, enrollment rates are relatively stable in the counties I consider.

The Cost and Use files also contain information on Medicare payments to service

[22] The MCBS does not track the specific plan number chosen by the individual. Instead, it reports the contract number and a number of variables (constructed from survey responses and administrative records) related to the benefits offered by the plan. Following MPTC, I rank all of the plans offered under the recorded contract by their closeness to the plan described by the survey participant and assume the true choice is the closest match.

[23] The question asks responders to rate their own health as "excellent," "very good," "good", "fair," or "poor." I group the first three responses into "healthy" and the last two (as well as any non-responses or refusals) into "unhealthy." Nyman et al. (2007) use a similar question in the Medical Expenditure Panel Survey to develop quality-of-life measures across the US population for cost-utility analyses.

Table 1.1: Summary statistics: individuals

Variable	**Mean**	**Std. dev. across years**
Income	$46,198	3391
Age	75.2	.171
Pct. healthy	56%	1.9
Pct. female	54.5%	0.96
Pct. black	6.1%	0.43
Pct. hispanic	1.5%	0.11
Pct. w/ bachelor's degree	25.5%	0.93
Pct. enrolled in MA	26.0%	0.62
Obs.	9,346	

Note: Observations are at the year-individual level. All calculations use MCBS sample weights.

providers for individuals enrolled in traditional Medicare. This data is constructed from a combination of patient interviews and administrative records.[24] Table 1.2 summarizes Medicare expenses for different groups of individuals. I use this information to construct the risk-adjusted subsidies CMS pays to firms for various individuals across the demographic distribution.[25]

1.4.3 Plan data

The Centers for Medicare & Medicaid Services maintains a public database of the characteristics of all MA plans offered each year. This database includes detailed information on plan costs, benefits, and options, at the contract-plan-segment level. For each plan, I extract the price, coverage area, and a number of plan characteristics,

[24] For more on the construction of this file, see Eppig and Chulis (1997).

[25] The true risk adjustment formula uses ICD-9 diagnostic codes (in addition to other demographic variables) to determine payments (CMS, 2014). Though the MCBS asks a number of questions about diagnoses and illnesses, it does not contain ICD-9 codes and I therefore use self-reported health status as a proxy.

Table 1.2: Average Medicare payments across groups of individuals

Category	**Mean**	**Std. dev.**
All	$6,743	16,177
Health status		
Healthy	4,390	10,761
Unhealthy	11,452	21,816
Gender		
Male	8,199	18,877
Female	7,117	15,406
Age		
65-69	4,366	14,008
70-74	6,395	18,742
75-79	8,031	15,132
80-84	8,976	15,561
85+	10,731	20,431
Obs.	6,792	

Note: Observations are at the year-individual level. All calculations use MCBS sample weights.

including copays for doctor and hospital visits, as well as flags for drug coverage, dental coverage, and vision coverage.

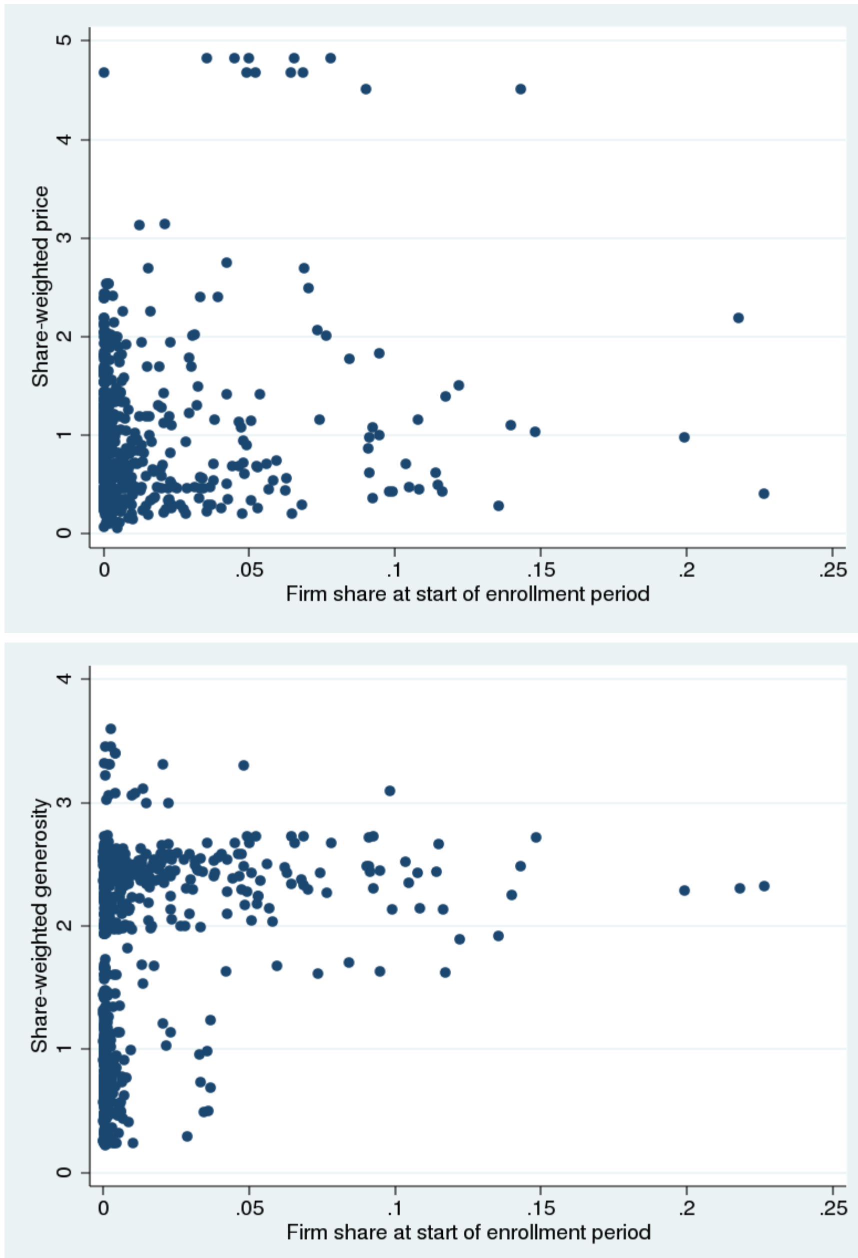
MPTC estimate the relative preferences of individuals for these different plan features. I use their estimates to transform the multi-dimension plan characteristics into a single generosity index g_j, for each plan j.[26] Since MA plan providers are required to offer coverage for each service that Medicare covers, I define $g_j = 0$ to be Medicare-level coverage and linearly transform the generosity index to ensure non-negativity. Examples of plans with different generosities can be found in Table 1.4.

Table 1.3 contains summary information on plans. Notably, 81% of firms offer more than 1 plan, and 73% of those have at least one plan with zero annual premium. Indeed, 43% of all plans offered across markets do not charge a premium. I use these facts to simplify the firm's problem in my estimation routine: I fix the number of plans per firm at 2 and restrict the first plan's premium to zero. For firms with more than two plans, I weight the attributes of the plan by the plan's individual share. Figure Figure 1.1 shows the results of this weighted average for plans with positive prices.

Table 1.5 shows the average number of firms and plans for each of the years I consider. While there is some variance in firms from year to year, the yearly means do not differ substantially from the overall mean of about 15 firms. There is a drop in plans between 2009 (21.8 plans on average) and 2010 (16.8 plans). The biggest single change was in Franklin County, Ohio, (in which Columbus, Ohio is located) which went from 28 plans in 2009 to 17 plans in 2010.

[26] This essentially assumes that, near the selections offered, consumers and firms view different plan attributes as perfect substitutes.

Figure 1.1: Firm shares and actions



Note: Price measured in thousands of dollars per year. Each dot represents one firm-year. Price and generosity weighted by previous period plan-level shares. Zero price plans excluded.

Table 1.3: Summary statistics: Plans

Variable	**Mean**	**Std. dev.**
Pct. w/ 0 premium	43%	
Annual Premium (if > 0)	\$825	629
Generosity	1.736	.809
Med. plans per firm	2	
Avg. plans per firm	2.85	1.4
Pct. firms with > 1 plan	81%	
Pct. of those with 0 prem plan	73%	
Obs.	2306	

Table 1.4: Example plan generosities

Variable	**Generosity** ≈ 1	**Generosity** ≈ 2
Drug coverage	Yes	Yes
Vision coverage	No	Yes
Dental coverage	No	No
Primary care copay	\$15	\$10
Specialist copay	\$30	\$25
Out-of-pocket limit	\$4,000	\$2,500
Generosity index	.998	2.03

Table 1.5: Competition across years

Year	**Firms**	**Plans**
2008	15.6 (4.69)	20.3 (6.06)
2009	17.1 (5.23)	21.8 (6.14)
2010	13.4 (4.85)	16.8 (6.00)

Note: Figures averaged across 39 counties. Standard deviations in parentheses.

1.4.4 Geographies

I obtain data on individual markets (which are defined as counties) from the Area Health Resources Files maintained by the Health Resources and Services Administration within the US Department of Health and Human Services. The files combine and summarize data from multiple sources, including the Census Bureau and other Health and Human Services sources, into a single county-level dataset.

For each county, I extract the population, the median income, the number of practicing medical doctors, the number of hospitals, and the number of nursing homes to use as cost covariates. I additionally extract the "contiguous county" file which allows me to identify neighboring counties for construction of instruments. Finally, I extract the "benchmark" per-capita subisdy rate for MA plans. The benchmark rate is the subsidy paid to a firm for a person of average risk and varies by market according to Medicare's average costs. I use this market-specific benchmark rate along with the the average Medicare expenditures for individuals in different groups to construct individual-specific subsidy rates.

Table 1.6 summarizes my geographic data. Since I form market shares from the MCBS data, I restrict my analysis to those markets with the greatest MCBS sample size. This in turn means the county markets I consider are significantly larger than the average across the United States.

Table 1.6: Summary statistics: Markets

Variable	**Mean**	**Std. dev.**
Num. MA firms	15.21	3.633
Total population	1,435	1,835
Medicare population	180.3	207.9
Per-capita income	$40,825	8,330
Num. doctors	3,782	4,831
Num. hospitals	19.1	22.6
Num. nursing facilities	47.9	63.4
Benchmark rate	$10,267	865
Obs.	39	

Note: Markets are defined as counties. Population measured in thousands of people. Per-capita income and benchmark rate are annual figures.

1.5 Empirical implementation

I use a multistep approach to estimate the model parameters for the demand and supply sides $\theta = \{\theta^D, \theta^S\}$, and calculate counterfactuals using a modification of the Weintraub et al. (2010) algorithm. For computational simplicity, I limit each firm to offering two plans. Since 40% of plans observed in the market have no premium, I restrict one of the firm's plans to be a zero premium plan. The firm's action therefore consists of three components: $\sigma_f = \{g_0, p_1, g_1\}$.

1.5.1 Estimation

Preliminaries

Since the policy function generated by the model is dependent on the starting shares of the firm, I must construct these shares for each firm-year observation. I use the MCBS observations to form these shares for 39 markets.[27] For each firm present in

[27]The MCBS has relatively low sample size in most of the counties it observes. Gandhi et al. (2013) show that errors in market shares caused by small sample size can bias demand estimates.

the data, I form share-weighted averages of price and generosity to match the three components of the firm's action in my model and adjust the decision of consumers accordingly.

Demand

I estimate the demand model following the two-stage approach of MPTC, which builds on the discrete choice estimation approach of Berry (1994) and Berry et al. (1995) by adding panel data on consumers' choices. I start by re-writing the utility model as a combination of individual specific terms and product fixed effects δ_j:

$$U_{ij} = \sum_r \alpha_r P_j d_{ri} + F * 1\{\text{switch}_{ij}\} + \beta_z Z_i + \beta_{zg} Z_i g_j + \delta_j + \epsilon_{ij}$$

For a given guess of individual specific demand parameters $\theta_Z^D = \{\alpha_r, F, \beta_z, \beta_{zg}\}$, I use the Berry (1994) contraction to find the unique set of $\delta_j(\theta_Z^D)$ that match predicted shares to observed market shares.[28] Using the individual choice data from the MCBS, and its panel structure to calculate switches, I construct the likelihood function for an individual as follows, where C_i represents the choice of individual i:

$$l_i(j; \theta_Z^D) = s_{ij}^{C_i = j}$$

In the first step of the demand estimation, I maximize the likelihood function over the space of θ_Z^D.[29] At the point estimate, $\hat{\theta}_Z^D$, I store the unique $\hat{\delta}_j$. In the second step of the demand estimation, I regress these $\hat{\delta}_j$ on on observable product characteristic

MPTC avoid this issue by using monthly enrollment data from CMS. However, this data is only available at the aggregate level. As I must separate out shares by health status, I must read shares from the MCBS directly and therefore must limit myself to markets in which there is sufficient sample.

[28] As the choice set is the same for all individuals in the market, I need only construct market-level shares, as opposed to demographic-specific market shares. I therefore use CMS Enrollment files, which cover the entire population of a county, to construct these shares, thus avoiding measurement error problems within the Berry (1994) contraction.

[29] This step utilitizes the Broyden-Fletcher-Goldfarb-Shanno (BFGS) maximization routine.

according to the terms in the original demand equation, where ξ_j represents product-specific unobservables:

$$\delta_j = \alpha_0 P_j + \beta + \beta_g g_j + \xi_j$$

I instrument for price using the average benchmark in the surrounding counties as well as the average generosities of competitors in the same market per Berry et al. (1995).

Supply

With demand estimates in hand, I construct the information set I_f for each firm. This consists of calculating the average competitive pressure observed in each market for each draw from that market's demographic distribution. As I take the number of firms in each period as exogenous, I create distinct measures of competitive pressure for each firm-year observed in the data and then average across firm-year observations.[30]

Conditional moment restrictions

Using notation from Pakes et al. (2006), my model creates an approximation $R(\cdot)$ to the true profit function of firms $\pi(\cdot)$ in the following sense:

$$\frac{\partial \pi_{fj}}{\partial X_{fj}} = \frac{\partial R_{fj}}{\partial X_{fj}} + \nu_{1fj}$$

In this expression, X_{fj} represents all of the data I observe about firm f and plan j, including market-level characteristics and the demographic distributions. ν_{1fj}

[30] Instead of the approach described in this paragraph and the following algorithm, I could follow the approach of Weintraub et al. (2010) and solve for a self-consistent policy function and expected competitive environment, as I do for my counterfactual simulations. Unforunately, their algorithm is not guaranteed to converge and requires significant additional computational complexity. in practice, the application of their approach increases the computational effort by an order of magnitude. The difference in policy functions computed by the two approaches at the point estimates described in the subsequent section is approximately 10%.

represents both expectational and measurement errors at the firm-plan level. Expectational errors can come from a number of sources, including incomplete information on the environment or asymmetric information on the states of other firms. As $\pi\cdot$ is the result of maximizing the profit function, any error in the profit function will contribute to the error in the policy function as well.

I impose a conditional moment restriction:

$$E[\nu_{1f}(\theta_0)|I_f] = 0\,\forall f$$

This restriction states that, conditional on their information sets, firms choose their actions optimally on average – this is equivalent to a restriction on firm behavior: $E[\sigma^{DATA} - \sigma^{MODEL}|I_f] = 0$

Since I am imposing a conditional moment restriction, there are an infinite number of possible instruments – indeed any function of any component of the information set I_f is a valid instrument. Chamberlain (1987) shows the efficient set of instruments are formed by the derivative of the moment restriction with resepect to each parameter when the derivative is evaluated at θ_0:

$$\boldsymbol{H}_f = E\left[\left.\frac{\partial \boldsymbol{\nu}_{1f}(\theta_0)}{\partial \boldsymbol{\theta}}\right| I_f\right]$$

Following Berry et al. (1999), I approximate the optimal instruments by calculating these derivatives at an initial guess of the parameters and recalculating them when I update the weight matrix during the GMM procedure.

Two-step GMM

I use a a two-step GMM approach. I start by setting my weight matrix to the identity $W_1 = I$ and calculate an approximation to the optimal instruments based upon an initial guess. I minimize the GMM objective function $f(\theta^S) = g'W_1g$ to obtain an initial estimate of the parameters $\hat{\theta}_1^S$.[31] I update the weight matrix according to the sample variance-covariance matrix $W_2^{-1} = \hat{S}(\hat{\theta}_1^S) = \frac{1}{n}\sum_n gg'$. I also update the approximation to the optimal instruments.

I then minimize the modified GMM objective function $f(\theta^S) = g'W_2g$ to obtain my final parameter estimates $\hat{\theta}$.

Identifcation

I assume that, in expectation, firms choose actions optimally given their information on the markets and the incentives they face from the terms of their value function. Parameters are therefore identified through their impact on the value function of firms and its derivative. The vector of first order conditions for optimality of the firm's value function can be written as follows:

$$\mathbf{0} = \sum_j \frac{\partial \pi_j}{\partial \boldsymbol{\sigma}_f} + \beta E\left[\frac{\partial V}{\partial \boldsymbol{\sigma}_f}\right]$$

This set of first order conditions illustrates the primary incentives faced by firms. Note that σ_f is a vector consisting of the features of all of f's plans. An expansion of $\frac{\partial \pi_j}{\partial \boldsymbol{\sigma}_f}$ would explicitly reveal a cross-product cannibalization issue for firms through the mechanism of s_{ij}: the number of people you attract to an individual plan is a function of the features of all the plans you offer. Since firms must offer plans at the same price and features to every consumer in the market, these inter-plan derivatives are crucial in determining the degree to which firms can price discriminate between consumers with different preferences.

The second term of these first order conditions captures the intertemporal tradeoff

[31] As analytic derivatives are unavailable and individual function evaluations are extremely expensive (see the computation subsection for more), I use the Nelder-Mead simplex search method.

between today's profits and tomorrow's share – and the effect of tomorrow's share on tomorrow's profits. This term is the key difference between this model and previous efforts to model the Medicare Advantage market. Its value at various points in strategy and state space determines whether firms are engaging in "collecting" market share or "harvesting" that share (Farrell and Klemperer, 2007, Section 2.6.2). Its scale relative to the first term determines the degree to which that behavior dominates the selection and product cannibalization concerns of the firm.

My estimator identifies the different components of the cost function through intra- and inter-market variance along with the demand distribution. The model generates a policy function for all points in the firm's state space. Differences in observed behavior of firms at different points in that space within the same market identify the constant and health-specific terms in the cost function. Differences in observed behavior of firms at similar points in markets with different attributes identify market-level cost parameters. The other demographic-specific terms in the cost function are identified by cross-market variation in the distribution of demographic characteristics.

1.5.2 Counterfactual

In order to evaluate the effects of alternative policies, I must solve for the equilibrium strategies of firms. My estimation approach uses the data to calculate the expected competitive environment faced by firms in each market. In a counterfactual scenario, however, the expected competitive environment is likely to be different from what I observe in the data. For any given strategy σ, I can calculate the expected information set $C(\sigma) = E[I_f|\sigma]$, and for any given information set C, I can calculate the optimal strategy $\sigma(C)$. The challenge is to find a fixed point of these simultaneous equations.[32]

[32] In a full-information version of the model, multiple equilibria are possible. Given the restrictions I have placed on firms' information sets, it seems likely (though not inevitable) that the number of

I proceed by fixing the number of firms to the average number of firms observed in the market[33] and adapting the algorithm of Weintraub et al. (2010). The algorithm may be outlined as follows:[34]

1. Start by initializing $\sigma_0 = \mathbf{0} \forall s_{h,f}, \zeta_f$, $t = 0$, and $x = 1$.
2. Loop until $x < tol$:
   (a) Given σ_t, calculate the expected information set C_t by forward simulation.
   (b) Given C_t, solve the single firm problem $\max_{\sigma_f} \sum_j \pi_j(\cdot) + \beta V_t(s'(\sigma_f))$ to obtain σ_{t+1}
   (c) Set $x = ||\sigma_t - \sigma_{t+1}||$ and $t = t + 1$
   (d) Set $\sigma_{t+1} = \lambda \sigma_{t+1} + (1 - \lambda)\sigma_t$

1.5.3 Computation

The primary computational challenge is calculating $\sigma^{MODEL}(s_{h,f}|\theta^S, I_{fm})$ for a given set of supply-side parameters. The primary advantage of the Oblivious Equilibrium approach is the result that I must merely solve a single firm problem for each market. To do this, I discretize the $s_{h,f}, \zeta_f$ information set space and use a value function iteration algorithm:

1. Start by initializing $V_0 = 0 \forall s_{h,f}, \zeta_f$, $t = 0$, and $x = 1$.

possible equilibrium are fewer than in a more complete information environment. In practice, I use a variety of different starting points for each counterfactual and find only a single equilibrium.

[33] Weintraub et al. (2010) include entry and exit, whereas I treat it as exogenous. While my algorithm is easily modified to include entry and exit, doing so would require assumptions about entry- and exit-relevant parameters. Therefore, for consistency with my estimation exercise, I abstract from entry and exit in the results I present below.

[34] As in Weintraub et al.'s case, this algorithm is not guaranteed to converge. Indeed, the algorithm often oscillates around the fixed point without progressing toward it. My implementation detects this condition and employs a local restart strategy with a reduction in the update parameter λ to encourage convergence. In practice, this has worked well.

2. Loop until $x < tol$:

   (a) Loop over each state $s_{h,f}, \zeta_f$:

      i. Solve the single firm problem $\max_{\sigma_f} \sum_j \pi_j(\cdot) + \beta V_t(s'(\sigma_f))$ to obtain $V_{t+1}(s_{h,f}, \zeta_f)$.

   (b) Set $x = ||V_t - V_{t+1}||$ and $t = t + 1$

To calculate the profit of various actions π_j, I numerically integrate out over the MCBS sample using their sample weights (forming $dC(Z_i)$).[35] To ensure the single firm problem is continuously differentiable, I use bicubic spline interpolation over the value function grid V_t to estimate the firm's continuation value for any future state $s'_{h,f}$. Once the value function converges, I construct sample moments from observed firm behavior by optimizing the strategy at each state observed in the data using the interpolated continuation values.

The estimation procedure spends most of its time in the profit function and its derivative calculating numeric integrals over demographic distributions. The Broyden-Fletcher-Goldfarb-Shanno extrema-finding algorithm used in step 2(a)i can require over 1000 executions of the profit function and 500 executions of the derivative to solve the single firm problem from an arbitrary starting point to the required precision. A complete execution of the value function iteration to the required precision over 1,200 grid points requires roughly 1,000,000 total executions of the profit and derivative functions. These calculations must be repeated for each market, for each guess of the cost parameters. To make matters worse, these functions must be calculated to full 64-bit floating point accuracy to ensure numerical stability of the outer-most GMM minimization routine. To enhance precision, I perform all numerical integration and moment calculations using the summation algorithm provided

[35] It is possible to construct a distribution of preferences such that the firm's problem in step 2(a)i admits multiple solutions. I perform a number of checks to ensure that the distribution created by the MCBS draws and my demand estimates leads to unique solutions of the firm's problem.

by Kahan (1965). To achieve sufficient speed, the model is implemented in C++, uses OpenMP technology to parallelize the optimization of individual states within the value function grid and uses Intel MPI technology to coordinate the simultaneous solution of the value functions for different markets across multiple nodes in a high-performance computing cluster.[36]

1.6 Results

Tables Table 1.7 and Table 1.8 contain parameter estimates for the demand side. In general, these estimates are in line with other studies of the demand for Medicare Advantage – particularly those from MPTC. The estimates imply that seniors face a cost of roughly \$1,300 when switching from Medicare to Medicare Advantage and \$1,040 when switching between firms within the Medicare Advantage system. When compared to the average annual premium of \$825, the switching costs have a large impact on the behavior of consumers. However, even when switching costs are incorporated, Medicare Advantage plans are relatively undesirably: the constant disutility of Medicare Advantage is equivalent to \$1,262. On average, a unit of generosity is worth \$284 to consumers.

Table 1.9 summarizes the taste distribution implied by these estimates. On average, unhealthy people have a lower preference for MA plans – -3.755 for healthy people versus -3.858 for unhealthy people – but have a greater taste for generosity – .745 for healthy people versus .863 for unhealthy people. This result is in line with common models of heterogenous risk and adverse selection: the greater your risk, the greater your demand for insurance against that risk.[37]

[36] My code is available upon request.

[37] Lustig (2011) estimates the demand and supply sides simultaneously. This allows for increased flexibility on the demand side with respect to individual product features while still using a generosity index on the supply side – in essence forcing the average taste for generosity to be equal to 1. Data use restrictions prevented the implementation of this approach.

Table 1.7: Estimates: First stage demand parameters

Variable	**Coefficient**	**Std. Err.**
Income-level price effects (per $1000)		
Price for medium income group	0.0069	0.072
Price for high income group	0.048	0.070
Switching costs		
Medicare-to-Medicare Advantage	-3.971	.056
Inter-contract Medicare Advantage	-3.18	.098
Demographic level effects		
Age	.0049	.012
Female	.0035	.166
Black	-.107	.4289
Hispanic	-1.857	2.255
Graduated high school	-.373	.269
Some college	-.554	.289
Bachelor's degree	-.372	.298
Healthy	.132	.187
Demographic-generosity interactions		
Age	-.001	.005
Female	-.012	.073
Black	.063	.182
Hispanic	.718	.915
Graduated high school	.118	.118
Some college	.129	.126
Bachelor's degree	-.122	.131
Healthy	-.010	.081
Weighted log likelihood	-11,241	
Sample size	12,806	

Note: All dollar amounts are in thousands per year.

Table 1.8: Estimates: Second stage demand parameters

Variable	**IV**
Premium	-3.05 (.184)
Generosity	0.865 (.105)
Constant	-3.85 (1.03)
Sample size	22,717

Note: Standard errors are in parentheses. All dollar amounts are in thousands per year.

Table 1.9: Implied taste distribution by health status

Variable	**Mean utils**	**Std. dev.**
Healthy		
α	-3.04	.020
Taste for MA plans (β_i)	-3.755	.249
Taste for generosity (β_{ig})	.745	.130
Unhealthy		
α	-3.03	.021
Taste for MA plans (β_i)	-3.858	.309
Taste for generosity (β_{ig})	.863	.138
Obs.	4122	

Note: Observations are at the year-individual level. All calculations use MCBS sample weights. α is per $1,000.

Table 1.10: Estimates: Cost parameters

Variable	Mean	Std. err.
Base cost		
Constant	4.416	0.095
Market population	0.0589	0.0238
Unhealthy	9.142	0.116
Age	0.0850	0.0100
Female	0.0272	0.0591
Generosity cost		
Constant	.0494	.0093
Market population	0.003	0.032
Unhealthy	.0309	.0269
Age	.0095	.0065
Female	-0.0118	.0245
Generosity quadratic cost		
Constant	0.0450	.0240
Unhealthy	0.0004	0.0392
Obs.	1,779	

Note: Observations are at the year-firm level. Costs are measured in thousands of dollars per enrollee per year.

Table 1.10 contain parameter estimates for the firm's cost function. These estimates imply the average base cost of insuring a healthy person is \$5,293; compared to an average subsidy of \$5,783, firms earn a 9.3% margin. The base cost of insuring an unhealthy person is \$14,609; compared to a benchmark of \$16,298, firms earn a slightly higher margin of 11.6%. The average Medicare payments for healthy and unhealthy people are \$4,390 and \$11,452, respectively. These estimates imply that firms spend an average of \$184 on benefits beyond those of Medicare as measured by the generosity index.

In Table 1.11, I compare these results to those obtained using a static model. Ignoring switching costs completely results in much lower cost estimates, shown in

Table 1.11: Estimation summary

Specification	**(I)**	**(II)**	**(III)**
Switching costs	No	Yes	Yes
Supply model	Static	Static	Dynamic
Demand			
Medicare-to-MA cost	N/A	1,302	1,302
Mean elasticity	.131	1.56	1.56
Mean inclusive value	$223	252	252
Supply			
Base cost	$3,450	5,509	4,416
Additional cost of unhealthy enrollees	$8,808	9,071	9,142
Cost per year of enrollee's age above 65	-$80	-23	85
Mean generosity expenditure	$432	213	184

Note: All dollar amounts annualized. Base cost is for a 65-year-old male in good health enrolled in a plan with 0 generosity. Inferred values (elasticity, inclusive value, generosity expenditures) calculated using observed plan characteristics and MCBS sample weights.

column I. This change is driven by the low overall enrollment rate: roughly two-thirds of seniors in my sample never enroll in a Medicare Advantage plan. The demand estimator must rationalize this behavior and does so in part by significantly lowering the price elasticity.[38] This in turn implies that firm margins are considerably higher and therefore costs are much lower. Though consumers are much less price sensitive, the estimated value of Medicare advantage is roughly the same as the specification with switching costs. Column II of Table 1.11 illustrates the results of incorporating switching costs on the consumer side but assuming firms are myopic on the firm side. This specification results in significantly higher costs for some enrollees. Both of these alternative specifications infer a negative marginal cost of aging.

The results of these alternative specifications are driven by the differences in the

[38] In some sense, the inclusion of the panel data allows the estimator to push some of the observed distaste for MA plans into a switching cost. The switching cost itself is then identified through the implied utility that would have been obtained if the consumer had switched.

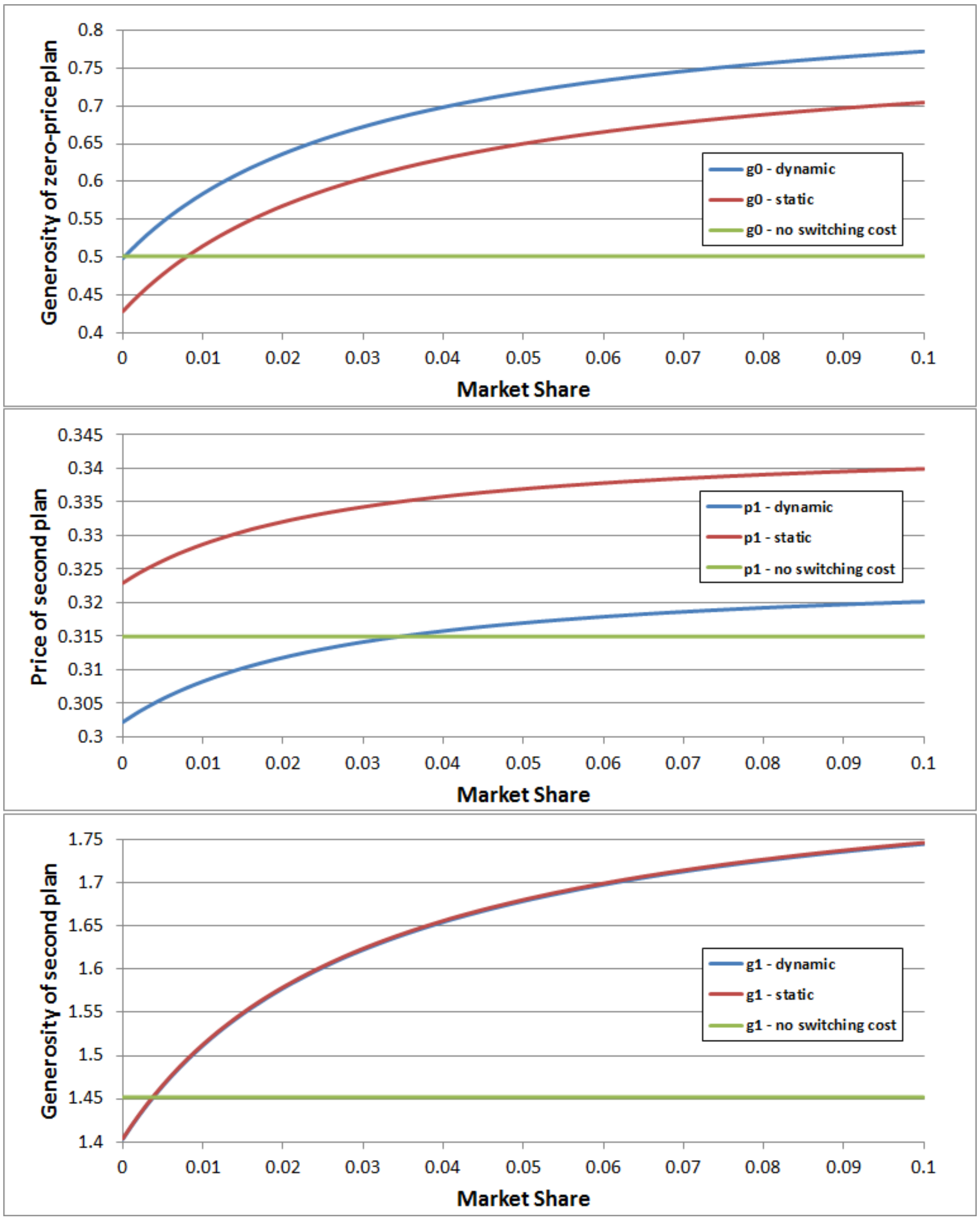
policy functions. Figure Figure 1.2 illustrates a “slice” of the three components of the policy function for a firm in Brown County, Wisconsin, using the preferred estimates of demand and cost under both models of firm behavior. When switching costs are ignored, firms behave uniformly across their possible market shares. Per Farrell and Klemperer (2007), when switching costs are included, myopic firms set higher prices and lower generosities than firms who take into account the future value of market share.

Taken together, these results reveal that MA firms do not face lower costs than the federal government. These increased costs may reflect increased administrative costs relative to Medicare, or difficulties negotiating favorable contracts with service providers.

The results of a counterfactual simulation for Brown County, Wisconsin in which the switching costs are lowered by 50% are in column 2 of Table 1.12. The reduced switching cost results in fewer consumers enrolled in Medicare Advantage overall – 11.6% in the counterfactual compared to 23.7% in the baseline scenario. The increase in consumer “liquidity” causes firms with low shares to reduce their generosities – they don’t have to offer as generous of a plan to obtain the same share. In equilibrium, firm with larger shares respond by lowering their generosities as well, as they don’t have to compete with as generous plans. This results in an overall reduction of the quality of plans offered in the market, causing many consumers to leave. Those that remain are slightly less healthy than enrollees in the baseline scenario. The difference in margins combined with lowered generosities leads to roughly a 10% increase in per-enrollee firm profits – though the total profits are much smaller, as fewer individuals are participating in Medicare Advantage.

Total welfare, including consumer surplus from the MA program, firm profits, and government spending on traditional Medicare benefits and MA subsidies, increases from negative \$224.64 million to negative \$205.82 million. This is largely driven by

Figure 1.2: Policy function comparison



Note: Policy function shown for a firm in Brown County, Wisconsin. Policy functions calculated using preferred estimates for demand and cost. Market share is of both healthy and unhealthy individuals.

changes in government spending: in the baseline scenario, the government spends \$79 million on MA subisides and \$158 million on FFS benefits.[39] With the reduction in switching costs, and accompanying reduction in MA enrollment, the government spends \$40 million on subsidies and \$176 on FFS benefits, a total savings of \$21 million.

Column 3 of Table 1.12 reports the results of a second counterfactual scenario in which subsidies are cut by 5%. This change leads to firms no longer offering the zero-premium plan and significantly increasing the premium for their second plan: the positive premium plan in the baseline scenario is offered for \$302 and the counterfactual plan is offered for \$772. Generosity is also reduced – the welfare from generosity for the second plan in the baseline is \$397 and drops to \$311 in the counterfactual plan. These effects combine to allow firms to maintain a degree of profitibility on a per-enrollee basis, but they have a drastic impact on enrollment: only 5.8% of consumers stay in Medicare Advantage.

Aggregating up to the market level, the change results in an improvement in total welfare of \$23.06 million. As in the lower switching cost scenario, the change is driven by government spending and is partly offset by a reduction in total firm profits and consumer welfare.

[39] These baseline numbers are computed using model-predicted outcomes. For comparison, CMS reported \$163 million in Medicare FFS spending for Brown County in 2009, the middle of my sample period. (Source: `http://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/Dashboard/Geo-Var-County/GeoVar_County.html`)

Table 1.12: Counterfactual: 50% decrease in switching costs

	Baseline	**Lower switching costs**	**Lower subsidies**
MA enrollment			
Healthy	24.3%	11.6%	5.7%
Unhealthy	23.2%	11.7%	5.9%
Per-enrollee:			
Welfare from generosity	$328	278	311
Average subsidy	$10,162	10,568	9,728
Mean firm costs	$9,951	10,249	10,005
Mean firm profits	$542	606	495
Total: (in millions)			
Consumer welfare	$8.30	7.97	3.49
Firm profits	$4.23	2.31	0.95
Subsidies to MA firms	$79.29	40.36	18.65
Medicare FFS spending	$157.87	175.73	187.37
Total welfare (millions)	$(224.64)	(205.82)	(201.58)

Note: Figures for Brown County, Wisconsin. Welfare, subsidies, and profits calculated using model-predicted enrollment and MCBS sample weights. Total consumer welfare is calculated using the inclusive value metric across all Medicare recipients and does not include base welfare generated by Medicare FFS benefits. "Lower subsidies" counterfactual calculated with firms only offering a single plan each with unrestricted premium.

For comparison, CMS reported 2010 Medicare FFS spending of $163 million for the county.

1.7 Conclusion

The cost of Medicare, both as a percentage of the federal budget and as a percentage of GDP, has risen steadily since its introduction despite the presence of many ‘gaps’ in coverage relative to the insurance plans sponsored by employers. Seeking in part to eliminate these gaps cost-effectively, policy makers have implemented the Medicare Advantage system, which offers subsidies to private firms who offer plans to consumers. For historical reasons, the subsidies are currently set well above Medicare’s cost – in other words a senior who switches from Medicare to Medicare Advantage increases their burden on taxpayers. However, almost all firms provide supplemental benefits on top of the mandated set of Medicare-equivalent services, meaning that a direct comparison is difficult.

To understand the welfare impact of the Medicare Advantage program, I estimate the cost structure of insurers using a revealed-preference approach. In contrast to previous work on the subject, I introduce a dynamic model of the Medicare Advantage market. The dynamic incentives in my model are driven by the existence of switching costs on the consumer side. Firms know that if they lower their prices today, they can attract a greater share to harvest tomorrow.

I find that the costs of private firms are higher than Medicare’s by a significant margin. My estimates also suggest that the risk adjustment formula used by Medicare overcompensates firms for unhealthy enrollees, relative to healthy enrollees. My results highlight the importance of switching costs in this environment – alternative estimation approaches that ignore these costs produce significantly different results.

In a counterfactual simulation, I find a reduction in the switching cost leads to a reduction in the number of Medicare Advantage enrollees. Those that remain are slightly sicker than in the baseline scenario, and firms make greater per-enrollee profits. These changes are driven by the increased “fluidity” of demand: low-share firms don't need to offer particularly generous plans in order to attract share, and high-share firms reduce their generosities as well to save costs. In equilibrium, the average level of generosity offered in the market decreases which pushes many consumers out of the system. Since the subsidies are higher than Medicare's costs, the change leads to an increase in total surplus of $18.9 million per year for a mid-sized county.

My second counterfactual simulation shows that reducing the subsidies offered to firms would also have a substantial positive effect on total welfare. However, this net effect is driven by changes in government spending and comes with a significant loss in welfare for individuals.

There are a number of avenues for future work. While the median number of plans offered by each firm is 2, many firms offer additional plans. Endogenizing the precise number of plans could lead to insights about the administrative costs of marginal plans. Endogenizing entry and exit could provide information about the overall fixed costs of the program.

These results support a view of Medicare as a relatively tax-efficient way to provide medical services to seniors in the United States. While private firms may offer attractive supplemental benefit packages, these currently come at a high cost to taxpayers.

Chapter 2

Does Premerger Notification Matter? Evidence from Cable Television

2.1 Introduction

In 1976 Congress passed the Hart-Scott-Rodino (HSR) Act as a response to several criticisms of anti-trust policy. While the Clayton Antitrust Act of 1914 had given broad powers to the Federal Trade Commission (FTC) and the Department of Justice (DOJ), these powers were largely reactive.[1] Enforcement agencies had difficulties challenging anticompetitive actions after they had occured and often found restoring a market to competitive status a costly endeavor. The HSR Act sought to address these concerns by forcing individuals and firms to report certain asset transfers or purchases and obtain pre-clearance before completing the transaction.ftc (2009)

Though the FTC and DOJ have reported the number of disclosures they've received on an annual basis (though individual filings are generally private) and used the powers granted under the HSR Act to challenge several large proposed mergers, it has been difficult to cleanly test the effectiveness of the policy. In many markets where large mergers are observed, it is difficult to obtain data on the holdings of the

[1] This chapter includes a substantial contribution from Kailin Clarke.

parties. It is also difficult to estimate the consumer level impacts of cross-industry mergers, and thus to get a sense of the degree of regulatory scrutiny transactions face. Additionally, in order to test the effectiveness of the policy we need to have a sense of the mergers that did not happen – not just those that did.

The cable television industry provides a solution to both those concerns. Firms providing cable television service must register their 'cable communities' with the Federal Communications Commission (FCC), which makes these registrations, as well as ownership changes, public. We combine this data with data from the Census Bureau to obtain a complete picture of the cable industry in the United States from 2000 to 2012. Not only does this data allow us to identify actual acquisitions, we can also construct the universe of potential acquisitions to identify model paramenters. Additionally, this data allows us to understand the degree of horizontal competition - referred to in the industry as 'overbuild' - these firms face in their individual communities. We assume any merger which includes overbuilt communities - in other words, any merger that results in a local market shift from duopoly to monopoly - would face increased scrutiny from regulators.

To test the effect of HSR, we develop a model of firm valuation which includes terms representing the cost of regulator scrutiny, particularly for acquisitions which contain a horizontal component. We then take this model to the actual and potential acquisition sets for the top four firms in the market: Comcast, Time Warner, Charter, and Cox. We concentrate on the top firms largely to reduce the assumptions we must make about their choice sets. First, given their size and the (general) strength of their balance sheets, it is reasonable to believe their marginal decision to acquire a particular small regional competitor is not based on financial constraints. This allows us to consider each decision independently, instead of considering a bundle of multiple acquisitions. Second, given their geographic spread, it is reasonable to believe they consider acquisitions across the entire extent of the United States. This allows us to remain agnostic about which small firms entire the decision set of the large firms.

Our results indicate the HSR filing requirement has the desired effect: firms pursue fewer mergers involving a horizontal component than the other characteristics of such mergers would indicate. Our results are robust to several variations of our empirical specification.

The remainder of this paper proceeds as follows: In Section 2 we provide brief background on both the cable industry and merger policy. Section 3 introduces a valuation model for potential acquisitions which leads to our simple test. In Section 4 we describe our novel dataset constructed from FCC, Census, and FTC data, highlighting the difficulty posed by limited information on former cable providers. Details on our data cleaning methods are left for an appendix. In Section 5 we detail our empirical strategy and provide the results of our test. Section 6 concludes.

2.2 Background

2.2.1 Hart-Scott-Rodino Anti-Trust Improvements Act

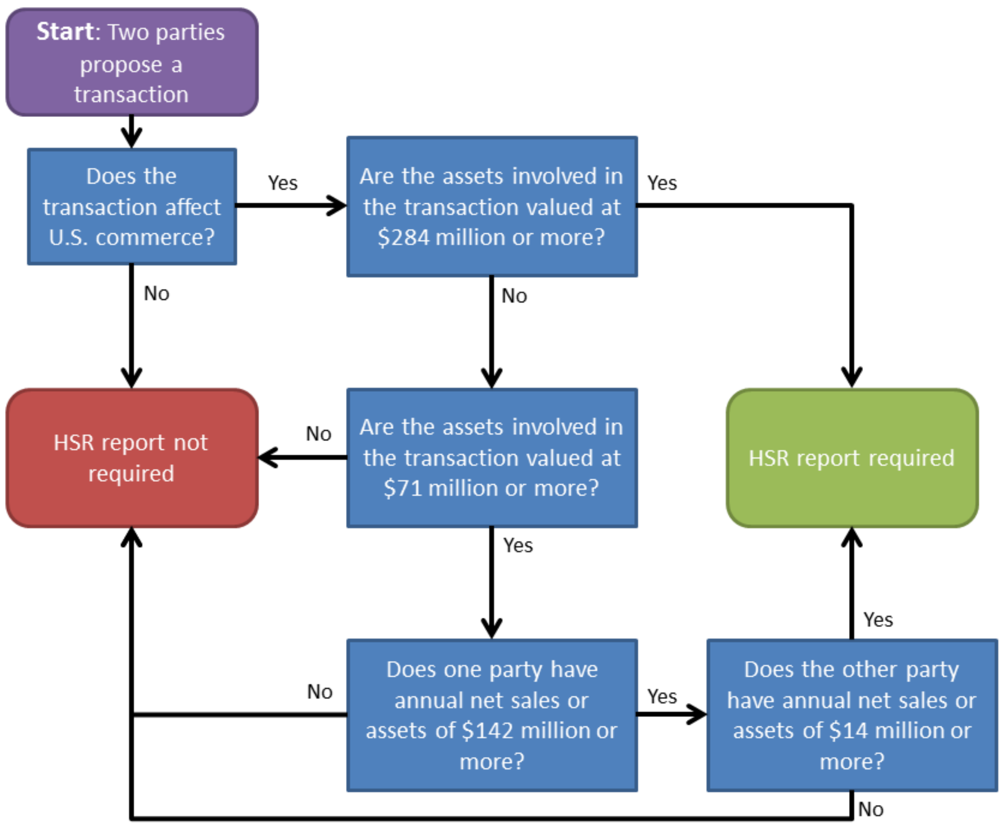
The primary effect of HSR was the creation of the FTC and DOJ's Merger Prenotification Program. Under the program, parties considering a sizeable transaction must file a "Notification and Report Form" and pay a substantial fee based upon the size of the proposed transaction. The parties must wait 30 days during which regulatory agencies may request additional information or time to review the transaction. If the reviewing agencies believe a proposed transaction violates antitrust laws, they may attempt to prohibit completion of the transaction by filing for an injunction in federal district court. Information provided to regulators during this process, including the original filing, is not subject to public disclosure, though court filings are generally available.

If the parties are conducting routine transactions or have experience with the system, they may file a request for Early Termination of the waiting period. If the Early Termination is approved, the transaction is made public as part of the Federal Register. While this data can be used to give a flavor of the types of transactions generally seen by regulators as unlikely to have anticompetitive effects, it cannot be used to identify the entire universe of attempted or successful purchases, since not all transacting parties request Early Terminations and not all Early Terminations requests are approved.

Transaction reports are necessary when either the value of the assets or the size of the parties reaches certain thresholds. These rules are designed to take effect cumulatively, so a firm which slowly acquires the assets of a competitor through multiple transactions will be forced to report even if each individual transaction is small. Thresholds are adjusted periodically by the FTC and DOJ to reflect inflation. Figure Figure 2.1 illustrates the various reporting thresholds based on the size of the parties and transaction denominated in dollars. As of 2013, reporting is required if the acquiring party will hold assets of \$281 million or more, or if one party is worth at least \$14.2 million, the other is worth at least \$142 million, and the assets transferred are worth at least \$71 million.fed (2013) An additional set of reporting requirements exist based on the percentage of assets transferred: filing is required if the transaction involves \$71 million in assets consisting of at least 50% of a company.

2.2.2 History of cable

Cable television began in the early 1950s as a way to improve the reception of over-the-air broadcast channels in remote communities. High demand for broadcast television coupled with the Federal Communications Commission's 1948 "freeze" on licenses to construct new stations led to the creation of Community Antenna Television (CATV) systems.of Broadcast Communications (2013) Instead of a separate antenna required



[p]

Figure 2.1: Flowchart of the 2013 Hart-Scott-Rodino reporting thresholds

for each household who wanted to receive broadcasts, a single, more sensitive antenna could be placed in a centralized location and connected to households through wiring.

Demand for cable systems spread rapidly, and by the 1970s even large metropolitan areas were wired for cable. Local governments executed ad-hoc franchise agreements with cable operators; in exchange for the (sometimes exclusive) right to provide cable services to the area, cable operators would guarantee certain benefits such as educational and governmental channels or special rates for particular segments of the population.Commission (2012)

Exclusive channels began appearing on cable systems starting with Home Box Office in 1972 and quickly became a large draw for subscribers. With the increased bandwidth available through wired technology, cable operators were able to offer a much wider variety to consumers than the broadcast alternative.Association (2013b); Commission (2012); Eisenmann (2000)

Today, over 90% of households have access to cable television and over 60% of households are active subscribers.Association (2013a); Nielsen (2011) Cable operators, empowered by the Telecommunications Act of 1996 discussed below, have also used the two-way properties of the communication technology to offer internet and phone services.

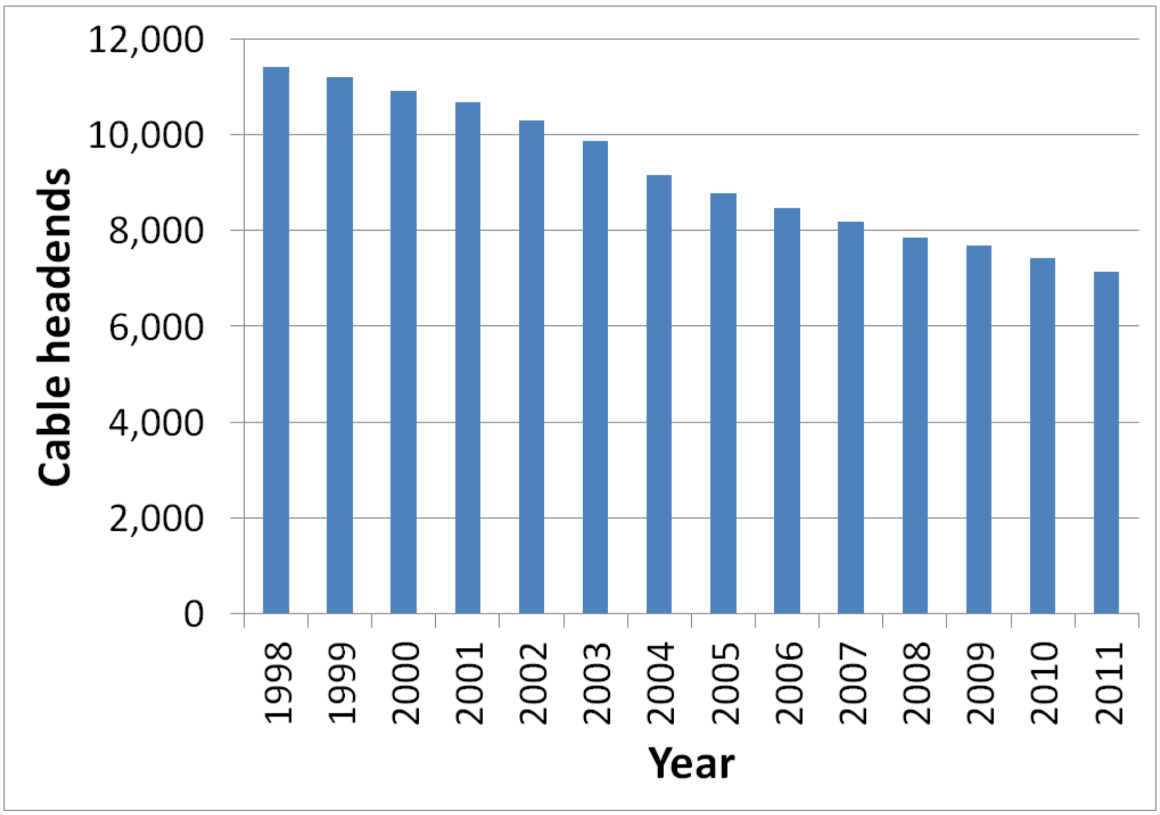
Competition in the video space comes mainly from Direct Broadcast Satellite technology, a subject previously studied in detail by Goolsbee and Petrin (2004).Goolsbee and Petrin (2004) Competition in the market for data provision comes from Digital Subscriber Line and fiber-to-the-home technologies.

2.2.3 Telecommunications Act of 1996

The Telecommunications Act of 1996, which amended the Communications Act of 1934, is the primary law regulating cable operators (as well as the rest of the telecommunications industry) today. The law's main goal was to promote competition by removing entry restrictions in telecommunications markets. In essence, the law was designed "to let any communications business compete in any market against any other."Commission (2011) Additionally, the law sought to update the FCC's regulatory authority and framework to encompass the Internet.

The 1996 Act removed most price controls from the market and encouraged local franchise authorities to allow additional firms to construct physical capital and enter local service markets. It was believed these so-called "overbuilders," along with entry from telephone service providers, would provide effective competition in major markets.Padilla (2001) These overbuilders are the source of the true horizontal purchase opportunities available to cable incumbents such as Comcast. Emmons and Prager (1997)Emmons and Prager (1997) finds empirical evidence that this change in market structure created increased incentives for monopoly power in the cable industry while Kelly and Ying (2003) examined the feasibility of overbuild and concluded profitable opportunities were rare.

Consolidation among cable providers and improvements in technology have led to a marked decrease in the number of distribution facilities required by the industry. Known as headends, these often unstaffed facilities receive channels through satellite or wired networks and re-broadcast them to the local cable network. Figure Figure 2.2 shows the number of these headends has decreased by almost 40% from 1998 to 2011.



[p]

Figure 2.2: The number of cable headends (physical locations used to recieve and distribute programming) has decreased every year since 1998. Source: Association (2013a)

2.3 Model

This section introduces a flexible model of acquisitions, focusing on the valuation the large firm makes. In our model, a large firm values individual communities individually and aggregate these valuations to value the entire set of communities served by a small firm. An acquisition is more likely to take place if the acquirer values the communities in the acquisition more than the company being acquired does, since there exists a merger price (in between the two valuations) such that both firms benefit. With this model in hand, we describe a simple test of the effectiveness of the HSR threshold which we can take to our data.

2.3.1 Environment

At the beginning of each discrete time period t there is a set J_t of large firms and a set K_t of small firms. Each firm i (where $i \in J_t \cup K_t$) operates systems which serve a set of CUIDs L_{it}. There is a set of communities M and a function $m_t(\cdot)$ which maps CUIDs to communities for a given period t. A community $c \in M$ has overbuild if two firms each have at least one CUID in the community; that is, there exist firms i, $i' \in J_t \cup K_t$, $i \neq i'$, and CUIDs $l \in L_i$, $l' \in L_{i'}$, such that $m_t(l) = m_t(l') = c$. Let $O_t(c)$ indicate whether community c is part of a community with overbuild at time t.

For a given community $c \in M$, let $R_{mt}(c)$ be the revenue that a monopolist can make in period t. Let N_c be the number of households who choose to subscribe in times of monopoly. We do not make any assumptions on how $R_{mt}(c)$ relates to N_c. However, we assume the decrease in revenue from a shift to duopoly is proportional to N_c; that is revenue decreases by $\theta_o N_c$ if the market goes from monopoly to duopoly.

Total revenue from any CUID l at time t is then[2]

$$R_t(l) = R_{mt}(m_t(l)) - O_t(m_t(l))\theta_o N_c$$

Operational costs are modeled as a per-subscriber marginal cost m_i and a fixed per-period cost $f_i(d)$ that varies by firm and depends on some measure of distance d from the CUID to the nearest other CUID that the firm owns.

In period t large firms observe which firms have operations in which communities, and information about the small firms that is unobservable to the econometrician. This information comes in two varieties: community-level information on the net per-subscriber savings in marginal costs ϵ_{jlt} and small-firm information on the fixed costs of acquisition ν_{jkt}. ϵ_{jlt} is drawn from a normal distribution with variance $\sigma_j^2, j \in J_t$ to reflect the intuition that some large firms (e.g. Comcast) produce higher gains in efficiency.

2.3.2 Valuation

Suppose a large firm j and a small firm k meet to determine whether an acquisition of k by j would be mutually beneficial. This amounts to determining whether j values k's operating systems more than k does. The difference between j's present value of k's systems and k's present value of k's systems, which we will call $DV(j,k)$, is partially made up of the sum of the differences between present values $v_j(l)$ and $v_k(l)$ of each CUID $l \in L_k$. It is also made up of the differences in j's valuations of each of its existing CUIDs $r \in L_j$, written as the difference between its pre-merger valuation

[2] For expositional and notational simplicity, we write this as if there is no difference in per-subscriber revenue between small firms and large firms. In practice, large firms may be able to obtain more revenue than small firms either through expansions of service or greater negotiating power with advertisers. In our estimation, we cannot separately identify subscriber-level changes on the cost side and the revenue side.

$v_j(r)$ and its post-merger valuation $v'_j(r)$:

$$DV(j,k) = \left(\sum_{l \in L_k} (v_j(l) - v_k(l)) \right) + \left(\sum_{r \in L_j} (v'_j(r) - v_j(r)) \right)$$

It can be shown, using both the formulation for revenue and costs above as well as the restrictions imposed by our data (detailed below), that the change in value simplifies to

$$DV(j,k) = \beta_0 s_k + \beta_1 \{\text{horiz}\} s_k + F(d) + \{\text{HSR}\}(\beta_2 + \beta_3 \{\text{horiz}\}) + \epsilon_{jlt} s_k + \nu_{jkt}$$

where

$$s_k = \sum_{l \in L_k} N_{m_t(l)}$$

is the number of subscribers acquired, $\{\text{horiz}\}$ is a flag for a horizontal or duopoloy-to-monopoly transition, $F(d)$ is the fixed cost of merging as a function of the distance d between j and k[3] and $\{\text{HSR}\}$ is a flag for the Hart-Scott-Rodino threshold. In practice, we estimate $F(d)$ according to

$$F(d) = \alpha_0 + \alpha_1 d + \alpha_2 d^2 + \xi$$

A large firm j will execute an acquisition if $DV(j,k) > 0$. This implies the probability of observing an acquisition follows a known distribution.

2.4 Data

In order to capture an accurate picture of the cable industry through time and understand the effect of merger policy on consolidation in the cable industry, we combine

[3]We set d to the minimum distance between the CUIDs operated by j and the CUIDs operated by k.

data on cable television systems from the FCC with market-level data on household counts from the Census and geographic location data from the United States Board on Geographic Names to create a novel dataset. We supplement this data with Annual Reports submitted to the FCC by cable providers, Early Termination Notices from the FTC, a series of letters Comcast wrote to the FCC informing the Commission of completed acquisitions, and a small number of public transaction size disclosures.

Our data on cable television systems was collected from FCC's internet-based Cable Operations and Licensing System (COALS) using an automated process. For a given Community Unit (known as a CUID in FCC parlance), COALS lists the current and previous service providers. COALS also provides access to administrative or regulatory filings made by the system operator that relate to the cable system, including ownership change forms and annual reports.

Table 2.1 presents a summary of the CUID ownership file. Just under half of CUIDs undergo legal-entity changes at some point throughout the study period, and the average number of unique parent companies responsible for a CUID was 1.85.

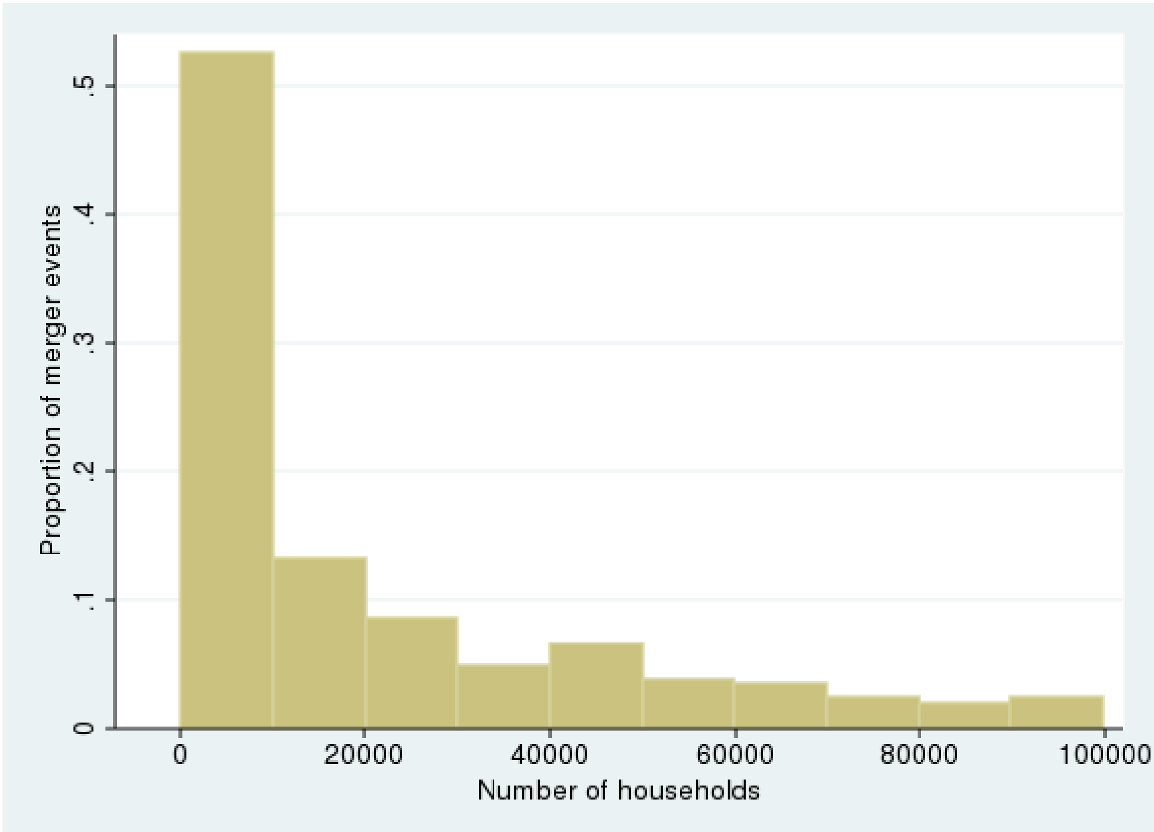
We identified individual acquisition events by looking at groups of CUIDs that switched from (say) Owner A to Owner B within a short time period. We verified our purchase identification process using data collected from a series of public disclosures Comcast made to the FCC about its acquisitions from 2003 to 2008. We distinguish between horizontal and conglomerate purchases with a simple process: For each CUID involved in the event, we examined the list of the acquiring company's existing properties at the time of the event for an exact community name match. If a match is found, the CUID is flagged as a horizontal acquisition. The remaining purchases are considered conglomerate.

To understand the value of controlling any particular cable system, we obtained population and household count data from the U.S. Census at the Census Place level. To understand the value of geographic clustering, we collected data on the location of the various systems (i.e. latitude and longitude) from the Gazetteer created by the Board on Geographic Names. We matched these data to our FCC community information by community name and county.

Finally, to understand the effect of the HSR disclosure requirement, we needed to map the financial value requirement to the context of our community-level data. We used limited public disclosures on acquisition prices to estimate a value of $4000 per subscriber and use annual report and industry data to estimate subscription rates across years. On average, the estimated acquisition value per household was $2600. We then applied the monetary threshold of $71 million to arrive at a threshold value of 27,000 households. While the monetary thresholds change throughout the study period, they are tied to the rate of inflation, which should roughly track the rate of growth in the value of a single subscriber.

Total number of CUIDs	45,146
Average number of providers per CUID	1.85
Std. dev.	1.14
CUIDs with single provider	22,986
CUIDs with more than 5 providers	690

Table 2.1: Summary of cleaned provider data.



[p]

Figure 2.3: Histogram of the size of the 712 acquisitions we study. This chart removes a small number of extremely large transactions for clarity.

2.5 Acquisitions

Table 2.2 provides summary statistics for our final set of 712 purchases made by top firms, covering 15,357 communities (or CUIDs, in FCC parlance) during our study period. Most purchases covered a relatively small area; the median number of communities involved in a single transaction was 3 and the median population affected was 31,123.[4] Figure Figure 2.3 shows the distribution of merger size as measured in households.

The existence of clustered purchases is immediately apparent: the average mean

[4]Compare to the median population of all cities and towns of the U.S. of 41,994.

distance between CUIDs involved in an acquisition and the set of CUIDs already owned by the acquiring firm was 4.7 miles. Since distances are calculated using centroids, this suggests many purchases involved systems essentially adjacent to the acquiring firm's pre-existing properties.

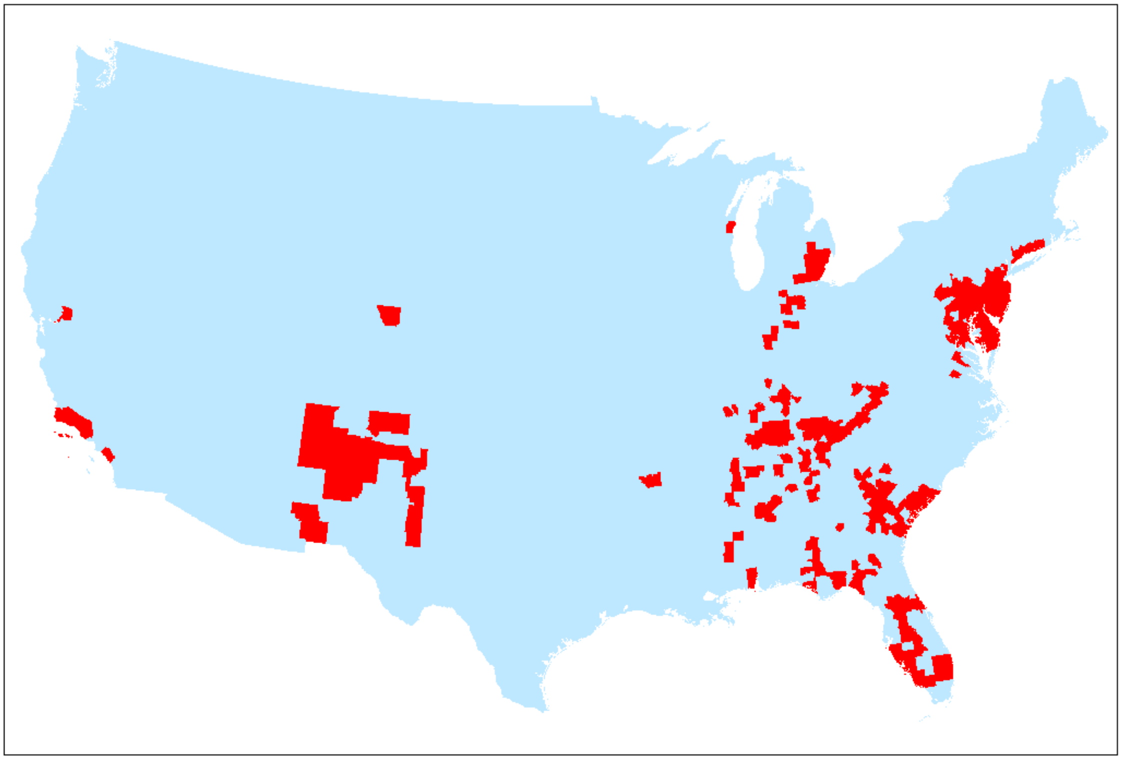
Table 2.3 provides the same summary statistics for each large firm we study. Comcast had 43% of the acquisitions covering 45% of the total acquired CUIDs and 44% of the population transferred during the period. As such, the summary statistics for Comcast largely drive the overall numbers reported in Table 2.2. That being said, the acquisition strategies for the other firms implied by the summary statistics are remarkably similar. The average number of CUIDs involved in a single event are almost identical, except for Adelphia which was impacted by its bankruptcy during the period.

The median number of households involved in purchases was below the threshold value of 27,000 for all firms except AT&T, suggesting a large amount of the consolidation in this industry was done without regulator scrutiny. Time Warner's significantly larger average purchase size was driven mostly by a few very large purchases in the New York and New England region.

Additionally, the average minimum distance between the acquired CUIDs and the firm's pre-existing CUIDs was also similar for all companies besides AT&T. Even AT&T's relatively large distance, 42.7 miles, equates to most acquisitions taking place within a space similar in size to the average US county.[5]

This clustering is apparent visually. Figure Figure 2.4 shows Comcast's holdings by county in 2001. By 2003, shown in Figure Figure 2.5, Comcast had not just consolidated its holdings in places such as Florida, it had also bought clustered operations in the Mountain West. Finally, by 2010 (Figure Figure 2.6) Comcast had expanded to

[5] In fact, this large distance is largely driven by a single acquisition 560 miles from the nearest AT&T-owned CUID.



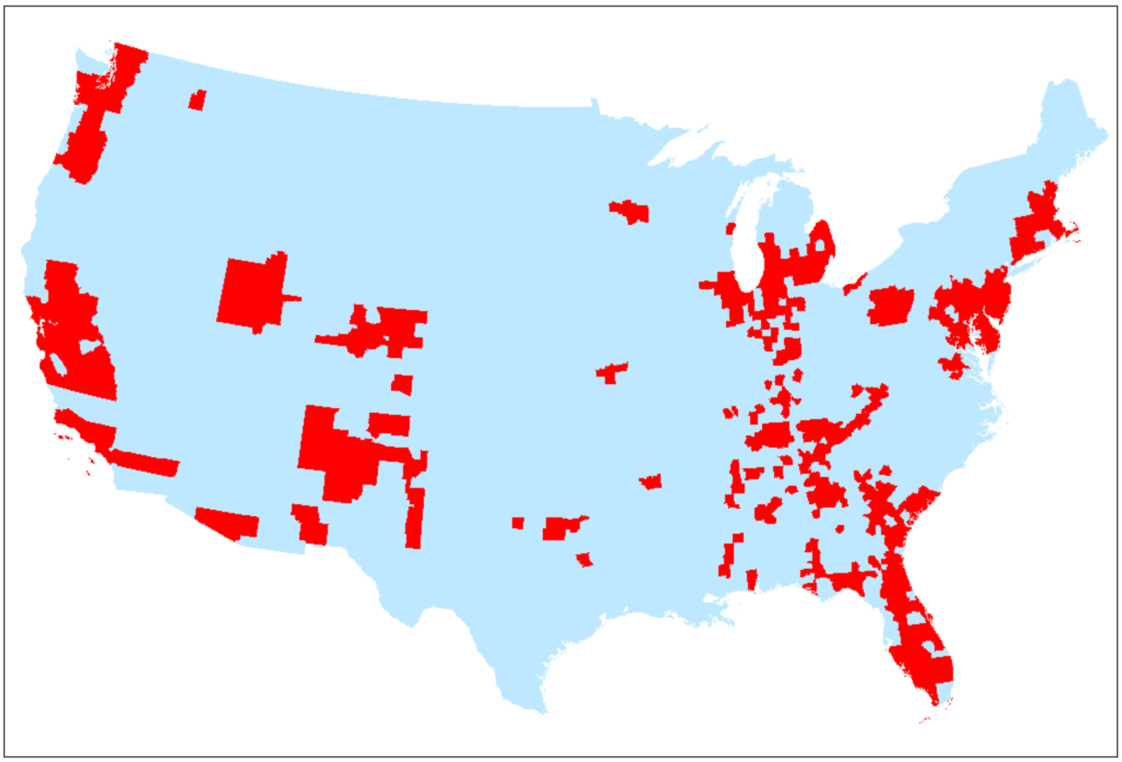
[p]

Figure 2.4: Map of Comcast's holdings by county in 2001. Counties are red if Comcast serves at least one community in the county.

the market leadership position largely through additional regional purchases. In this way, as shown in Figure Figure 2.7, Comcast has expanded its reach from roughly 10 million households to over 60 million by 2013. This implies that today, over 50% of households are in Comcast's territory (Figure Figure 2.8).

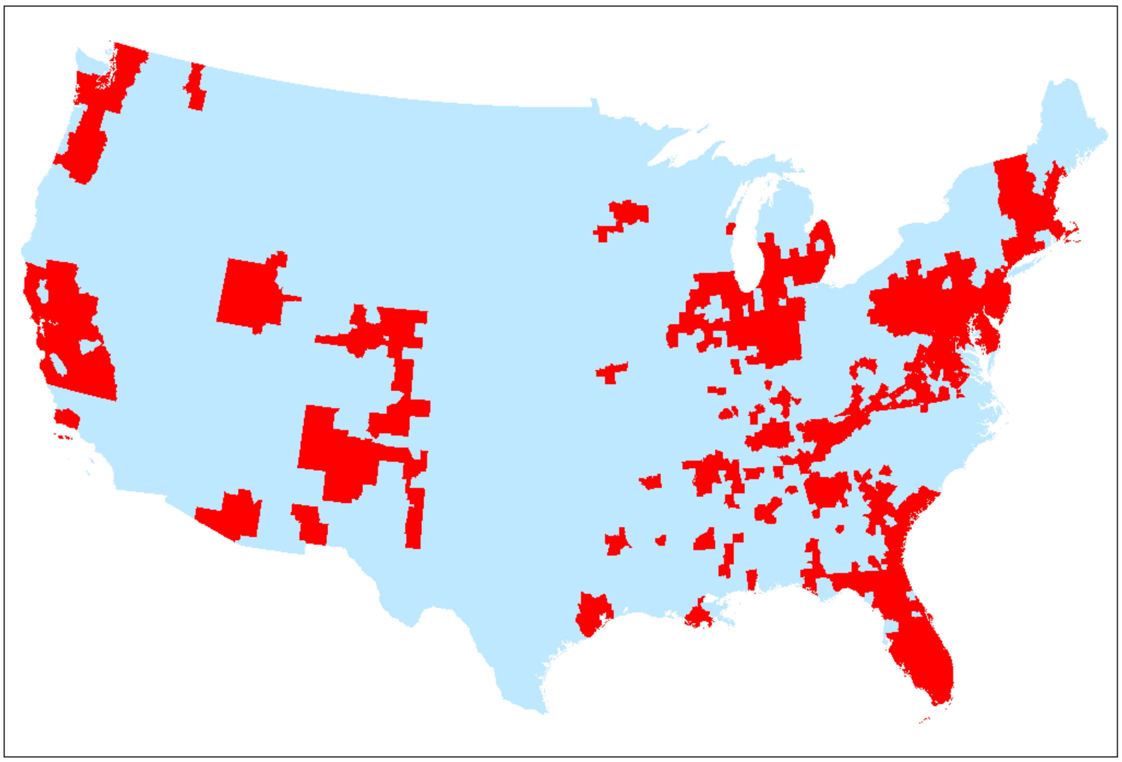
2.5.1 Horizontal acquisitions

Of the 15,357 CUIDs that were acquired by one of the large firms during the study period, 190 were considered horizontal purchases. These 190 switches were part of 74



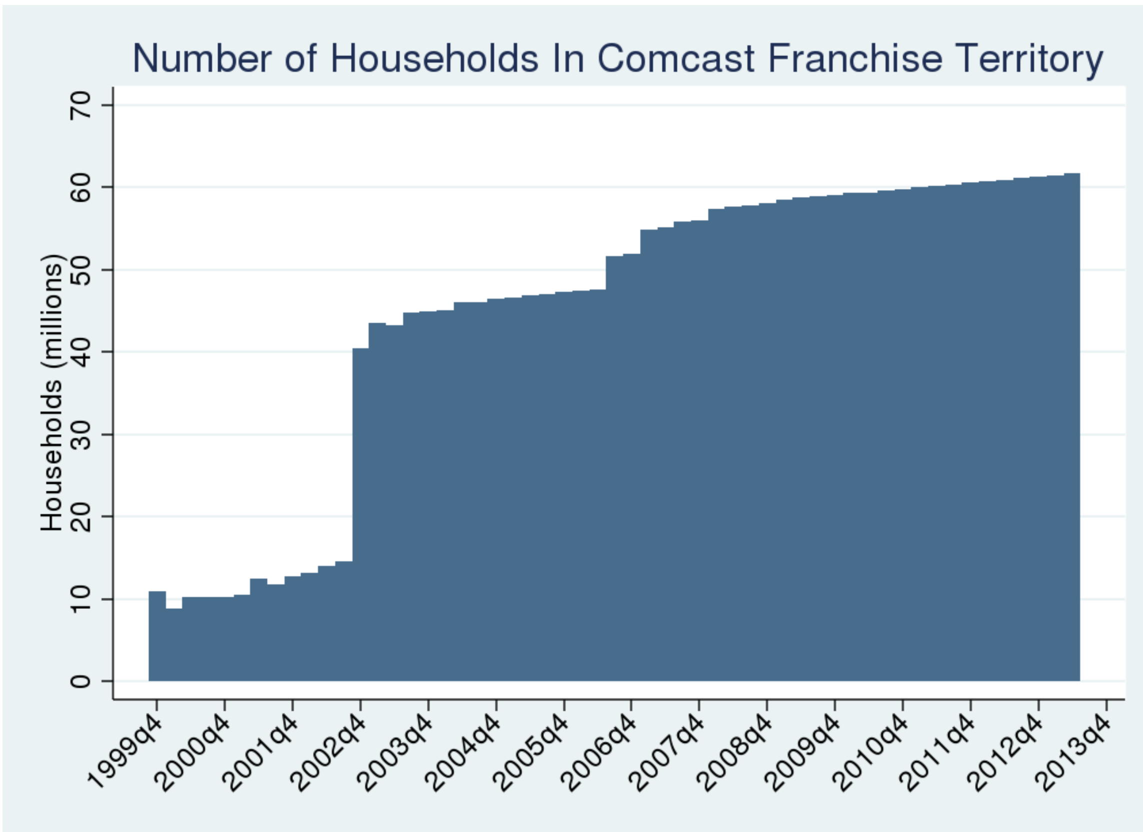
[p]

Figure 2.5: Map of Comcast's holdings by county in 2003. Counties are red if Comcast serves at least one community in the county.



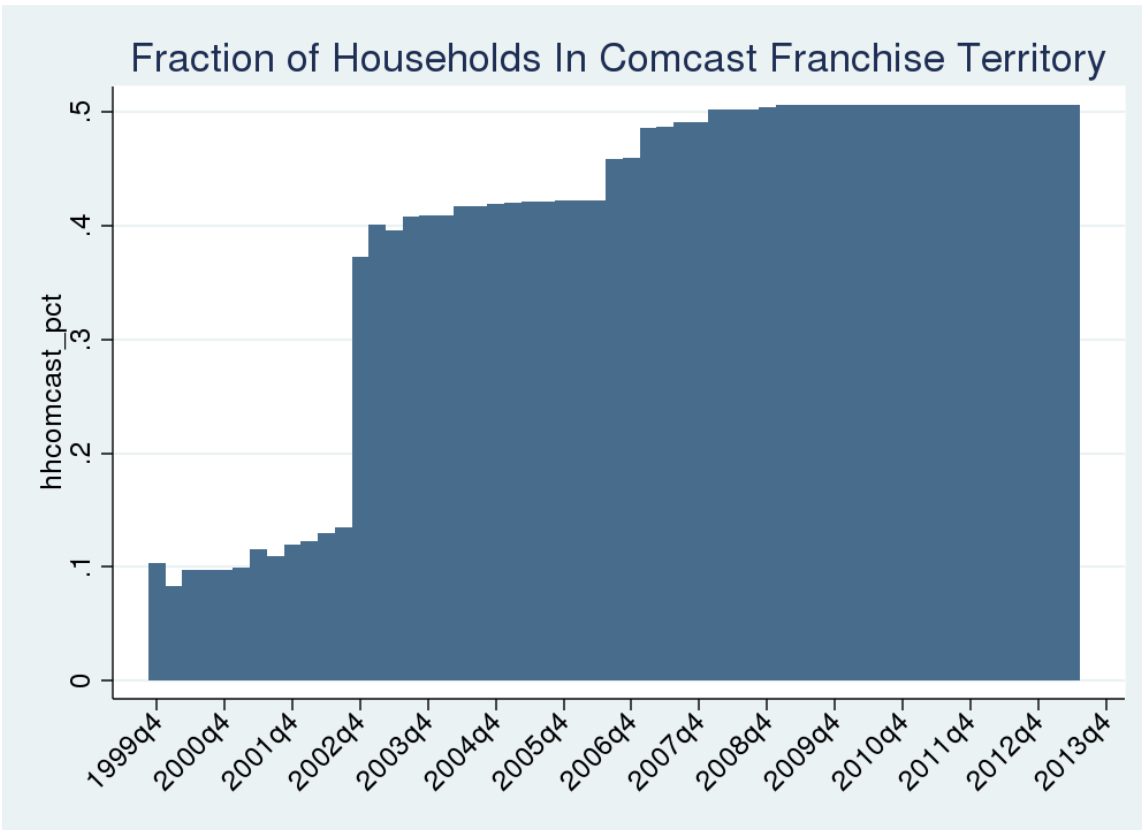
[p]

Figure 2.6: Map of Comcast's holdings by county in 2010. Counties are red if Comcast serves at least one community in the county.



[p]

Figure 2.7: The number of households within Comcast's franchise territory (as identified through our PSID/Census match process) has increased steadily throughout our study period. The large jumps in 2002 and 2006 are the result of the AT&T Broadband and Adelphia acquisitions, respectively. Quarterly household counts are imputed using 2010 Census levels and 2000-2010 growth rates by county.



[p]

Figure 2.8: The percentage of households within Comcast's franchise territory (as identified through our PSID/Census match process) has increased steadily throughout our study period. The large jumps in 2002 and 2006 are the result of the Adelphia and Susquehanna acquisitions, respectively. Quarterly household counts and percentages are imputed using 2010 Census levels and 2000-2010 growth rates by county.

distinct acquisition events – 10% of the total number of events seen.

Within the 74 events that included a horizontal component, the median percentage of CUIDs involved in the purchase that were considered horizontal was 12.5%. The mean was 33.2%. Several small purchases that consisted of a completely horizontal takeover contributed significantly to this mean – these tended to be municipality-run networks that were sold.

Of the 23 acquisitions with more than 50% of the CUIDs considered horizontal, the median number of households involved was 24,504, implying that many of these purchases required disclosure and scrutiny under HSR.

Number of mergers	712
Median CUIDs per merger	3
Average CUIDs per merger	21.6
Std. Dev. CUIDs per merger	82.6
Median mean distance to nearest owned CUID	.267
Average mean distance to nearest owned CUID	4.66
Std. Dev. of minimum distance to nearest owned CUID	29.98
Total CUIDs acquired	15,357
CUIDs missing population data	3,894
Median population per merger	31,123
Average population per merger	540,510.4
Std. Dev population per merger	2,861,357
Median households per merger	12,637
Average households per merger	201,099.4
Std. Dev households per merger	1,029,455

Table 2.2: Acquisition summary. Note: Household statistics include missing data for some rural CUIDs.

	Comcast	**Time Warner**	**Charter**	**Cox**	**AT&T**	**Adelphia**
Num. of mergers	307	157	139	34	26	49
Median CUIDs	3	3	3	3	4	4
Average CUIDs	22	23	22	19	22	13
Std. dev CUIDs	95	79	84	39	52	26
Median mean dist	0.20	0.17	0.35	1.27	3.13	0.79
Average mean dist	1.12	2.9	1.33	10.2	42.7	18.2
Std. dev mean dist	7.70	13.73	3.35	31.73	116.05	59.86
Total CUIDs	6,849	3,612	3,066	628	566	636
Missing pop data	1,830	814	900	69	108	173
Median pop	53,284	24,541	13,680	30,980	99,166	65,201
Average pop	555,237	940,561	219,075	329,859	450,493	272,204
Std dev pop	2,426,169	4,953,978	877,393	611,282	787,075	512,692
Median HH	20,898	8,759	5,633	12,446	37,691	26,114
Average HH	213,551	333,154	83,246	127,692	172,818	100,223
Std dev HH	934,400	1,718,539	323,470	236,221	302,860	183,336

Table 2.3: Acqusition summary by acquiring firm.

2.6 Suggestive Evidence

The data presented in the previous section lend themselves to two clear hypotheses:

1. Outside of true horizontal purchases, Hart-Scott-Rodino has little effect on merger strategy.

2. Firms place a high value on "near-horizontal" or highly-clustered acquisitions.

Both of these hypotheses are testable. First, if HSR filing rules place a major burden on transactions over a certain size, large firms should be less willing to pursue those transactions, relative to the opportunities available in the marketplace. Second, if firms value clustered systems, they should be more willing to pursue those transactions relative to the available opportunities.

To test these hypotheses, we ran a simple exercise. For each year in our study period, we created a list of cable systems the large firms could have acquired based on the ownership records.

We then used a simple logistic regression to estimate the probability of a successful acquisition event based on the size of the acquisition and the percentage of the potential purchase's horizontality based upon the acquiring firm's presence in the communities involved at the time of the purchase. We added a dummy variable representing the necessity of Hart-Scott-Rodino disclosure, as well as year dummy variables to reflect changing macroeconomic conditions.

The most important decision in the execution of this exercise is the selection of the decision set available to the firms. The main decisions essentially boil down to the following questions:

1. Can firms partially acquire firms? How do we determine the possible subsets?

2. What level of horizontality is allowed?

3. Can firms acquire other large firms?

4. Should potential targets acquired by other large firms be included?

The first question essentially defines the cardinality of the set. Though partial acquisitions do occur, they are relatively rare. Additionally, many partial acquisitions lead to further transactions with the same target later in the study period – meaning the "cumulative size" portion of the HSR rules applies. For this reason, we opt to model acquisitions as absolute: you either buy the whole company, or you buy nothing.[6]

Since we observe several truly horizontal purchases in the data, we allow any level of horizontality in our potential purchase set. Additionally, since the only "purchase" of a large firm (Adelphia purchased by Comcast and Time Warner) was the result of a bankruptcy process, we do not allow the large firms to acquire each other.

The final question is also the most vexing. Unfortunately, we have no data covering behind-the-scenes overtures and negotiations, so we are unable to observe (for instance) targets of mutual interest, bidding wars, and other types of strategic activity. Therefore, we estimate the model with several variants of the data representing alternative answers to this question.

The first variant treats all large firms as members of a hypothetical larger firm we call the "megafirm." In this variant, we calculate the distance variables according to the nearest distance to any cable system owned by any of the megafirm's "subsidiaries." In the second variant, we estimate separate models for the individual firms but exclude any company acquired by other firms from the set of potential purchases available to the firm in question. This assumes any negotiation process acts as a truth-telling device and large firms with the highest internal valuation always have the first option to purchase small concerns. In essence, if Firm B acquires Firm C,

[6] An alternative interpretation of this assumption is: you either execute a transaction with the firm or not.

that event is viewed as evidence that Firm C was never truly an option for Firm A. In the third variant, we estimate separate models for each of the major firms and allow them the possibility of acquiring any other firm in the market. This assumes the negotiation process may break down and firms may end up acquiring a target despite a different firm's higher valuation. Furthermore, if Firm B acquires only some CUID operations of Firm C, then we reason that this subset of Firm C's CUID's was also a potential purchase by Firm A.

The results for the "megafirm" specification are shown in Table 2.4. Parameter estimates for the second and third variations are shown in Table 2.5 and Table 2.6 respectively. For clarity, we discuss the results related to each of the hypotheses in separate subsections.

2.6.1 Does Hart-Scott-Rodino have an effect?

Across our specifications, a couple of patterns emerge. First, the HSR disclosure flag on its own has a positive coefficient and is highly significant. This implies that firms aren't dissuaded from pursuing large acquisitions by the HSR rules alone. However, when interacted with the horizontal flag, HSR disclosure has a negative effect, though the effect is much less significant. While we refuse to believe regulators do not scrutinize large mergers with a strong horizontal component, this suggests such scrutiny is not particularly burdensome, particularly compared with the benefits of horizontality as measured by the horizontal flag on its own.

2.6.2 How important is clustering?

Though the minimum distance parameter is not significant in any estimation apart from for AT&T Broadband, the parameter is negative in every specification estimated. This suggests that while firms pursue purchases that are located close to their current

	(1)
	acquired
Mean Distance	-0.00673
	(0.00826)
Horizontal Flag	0.388
	(0.259)
HSR Flag	1.270***
	(0.0919)
HSR * Horizontal	-0.641*
	(0.317)
[1em] Num. Households	0.000000133*
	(6.47e-08)
Year Dummies	Yes
Constant	-4.665***
	(0.167)
N	39813

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.4: Parameter estimates for the “megafirm” specification of our ‘potential merger’ exercise.

	(1) Comcast	(2) AT&T Broadband	(3) Cox	(4) Time Warner	(5) Charter
Mean Distance	-0.00219	-0.00832**	-0.0119	-0.00686	-0.0617
	(0.00206)	(0.00295)	(0.00839)	(0.00545)	(0.0379)
Horizontal Flag	0.663*	-9.431	1.880*	1.661***	1.802***
	(0.277)	(666.0)	(0.758)	(0.287)	(0.274)
HSR Flag	1.545***	1.340**	1.304**	0.870***	0.194
	(0.146)	(0.455)	(0.402)	(0.222)	(0.297)
HSR * Horizontal	-0.604	10.77	-1.026	-0.692	-0.427
	(0.330)	(666.0)	(1.021)	(0.395)	(0.484)
Num. Households	0.000000111	3.29e-08	-0.000000114	3.55e-08	-0.000000101
	(8.66e-08)	(0.000000369)	(0.000000443)	(0.000000132)	(0.000000308)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Constant	-6.542***	-6.623***	-6.483***	-6.224***	-5.206***
	(0.386)	(0.532)	(0.457)	(0.364)	(0.265)
N	33443	5926	27391	39813	36636

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.5: Parameter estimates for our ‘potential merger’ exercise assuming large firms were able to buy any small firm.

	(1) Comcast	(2) AT&T Broadband	(3) Cox	(4) Time Warner	(5) Charter
Mean Distance	-0.00227	-0.00829**	-0.0118	-0.00685	-0.0622
	(0.00209)	(0.00292)	(0.00836)	(0.00544)	(0.0380)
Horizontal Flag	0.654*	-9.435	1.868*	1.653***	1.795***
	(0.277)	(669.9)	(0.758)	(0.287)	(0.274)
HSR Flag	1.574***	1.398**	1.332***	0.919***	0.240
	(0.146)	(0.456)	(0.402)	(0.222)	(0.297)
HSR * Horizontal	-0.619	10.78	-1.026	-0.721	-0.435
	(0.330)	(669.9)	(1.021)	(0.395)	(0.484)
Num. Households	0.000000114	5.65e-08	-0.000000111	4.27e-08	-8.55e-08
	(8.61e-08)	(0.000000348)	(0.000000441)	(0.000000131)	(0.000000306)
Year Dummies	Yes	Yes	Yes	Yes	Yes
Constant	-6.534***	-6.582***	-6.477***	-6.224***	-5.203***
	(0.386)	(0.531)	(0.457)	(0.364)	(0.265)
N	33095	5723	26904	39320	36121

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.6: Parameter estimates for our ‘potential merger’ exercise assuming small firms bought by other large firms were unavailable.

holdings, it is not an overwhelming factor in their decision. Alternatively, given the relative crudity of our distance measure, it is possible our model is insufficiently nuanced to capture the true value. An ideal measure of distance would combine a concept of adjacency and the amount of right-of-way needed to combine physical systems.

2.7 Conclusion and future work

Although many have tried to measure the effectiveness of U.S. merger policy in an empirical way, these attempts have largely been stymied by the problem of sample size.Clougherty and Seldeslachts (2011); Carlton (2009). This project has attempted to cast the problem into the context of a specific industry, cable television service, in order to achieve enough variation to provide an empirically robust answer.

The results of our simple 'potential acquisition' exercise suggest policy may be too focused on particular types of acquisitions without considering the industry at large. In particular, it is not difficult to imagine that regulators in 1999 may have rejected a proposal to combine the cable television access of 50% of U.S. households into a single company.[7] Yet this is precisely what has occurred.[8]

While this paper lays out the acquisition history and strategy of the largest players in the cable provider market, it cannot fully answer questions about the effectiveness of U.S. merger policy. To that end, we have developed a structural model of firm acquisition to produce a truly robust and coherent quantitative look at both the effect of HSR and the benefits of clustering without the cavalcade of assumptions we have used in our 'potential acquisition' exercise.

[7]If this thought experiment does not convince you, consider a proposal to combine cable television, internet, and voice services for 50% of American households into a single company that *also* controls a quarter of the broadcast television market.

[8]To be clear, we are not making any claims about consumer or firm welfare through this period of consolidation. Rather, we believe regulators may have opted for additional scrutiny.

The next steps in the execution of this agenda include incorporating of Cable Factbook data into our acquisition dataset, followed by estimating our structural model using the techniques established in PPHI. With our structural model estimated, we can investigate several counterfactuals, including different regulatory regimes for acquisitions and alternative distance integration costs.

2.8 Appendix: Data details

Our main sources of data are the U.S. Census Bureau and the Cable Operations and Licensing System (COALS), operated by the FCC. We also obtained information on Early Terminations from the Federal Trade Commission and supplemented our procedures with several additional sources. This appendix gives details of our various data collection and processing procedures.

2.8.1 Early Terminations

The Federal Trade Commission maintains lists of all early terminations granted each week under the Hart-Scott-Rodino Act.[9] We manually searched these lists for events that included the large firms we were concerned with.

2.8.2 Comcast Letters

As part of a public comment period on proposed ownership rules in the cable industry, Comcast voluntarily submitted quarterly letters detailing their acquisition activity to the FCC, which subsequently published them on their website. We collected all of the letters available.

[9] Available at http://www.ftc.gov/bc/earlyterm

2.8.3 Geographic Data

We obtained population data at the Census Place[10] level from the National Historical Geographic Information System for Census 2000 and 2010 and augmented this data with 2010 data directly from the Census Bureau. For a city that crosses county lines, population counts are available for each "county-part" of the city while household counts are only available for the city as a whole. We imputed 2010 household counts for multi-county cities by taking the city-wide ratio of households to population and multiplying it by the population of each "county-part." Population and household counts were also available for the balance of counties (or other civil divisions) that are unincorporated – similar to the FCC community classifications described below.

We estimated the 2010 household counts for unincorporated communities by using a simple linear regression of household count on total population interacted with state dummies for all communities for which household data was available. We then used the growth rates of household counts by county from 2000-2010 to impute CUID household counts from 2000-2010.

Finally, we incorporated latitude and longitude data from the State Gazetteer prepared by the United States Board on Geographic Names,[11] matching by place name and the Census' internal unique identifiers. Where exact matches weren't available, we used the geographical centroid of the containing county or township.[12] Additionally, several manual links were made to account for changes in the definitions of certain political units (i.e. changes in county and city boundaries) throughout the country during our study period.

[10] This includes Census Designated Places

[11] Available at http://geonames.usgs.gov/domestic/fips55codedef.html

[12] This ensures every CUID can be included in distance calculations.

2.8.4 COALS

Overview of COALS and FCC identifiers

COALS consists of a database of cable system information, with a publicly accessible front end, as well as secured-access options for cable systems owners and administrators.[13]

Cable systems regulated by the FCC (and collected in COALS) are identified through Physical System Identification numbers (PSIDs) and the communities they service are identified through Community Unit Identification numbers (CUIDs). In towns where more than one physical system operates, multiple CUIDs are created. Additional CUIDs may also be created when towns cross county lines. For example, the city of Minneapolis, Minnesota, which is currently served by Comcast, is assigned a single CUID, MN0180. That CUID is "owned" by PSID 011339, which serves the greater Twin Cities area. On the other hand, Kansas City, Missouri, which spans four separate counties, is host to five separate CUIDs serviced by three PSIDs representing Comcast, Time Warner, and Surewest. The presence of two CUIDs with identical community names does not necessarily imply true overbuild; many of these cases occur in large geographic areas, such as the non-incorporated portions of counties.

CUIDs may also represent unincorporated areas and communities at a variety of scales. At the low end of the spectrum, a single CUID may represent a single 'private' settlement such as an apartment complex or hotel. A CUID may be created for an unincorporated community regardless of Census status. A single CUID may also be used to represent the 'balance' of a county: the total area of that county not included in any incorporated city contained within that county. Table 2.7 shows the distribution of CUIDs by FCC community type classification.

[13] COALS is available at https://apps.fcc.gov/coals/

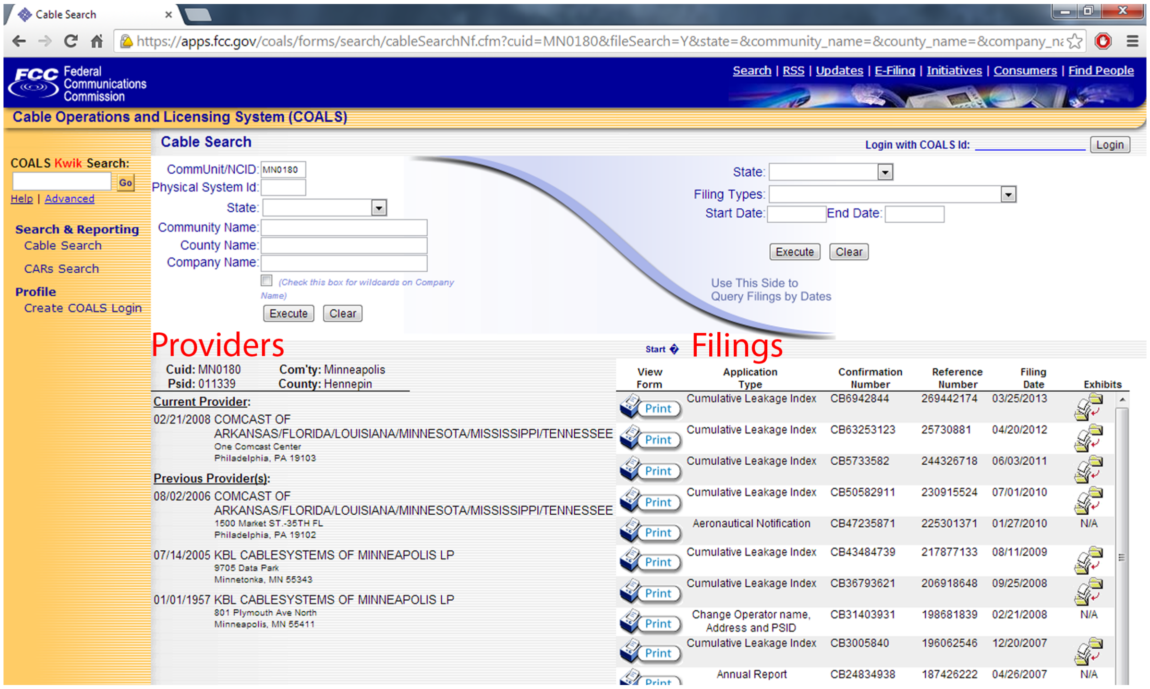
[p] 

Figure 2.9: A screenshot of the COALS page for the cable system in Minneapolis Minnesota, with emphasis on the providers and filings information we scraped.

Data Collection

Our data collection process begins with a exhaustive list of every CUID in the United States, taken from an FCC-provided current-status digest.[14] This CUID list is used as the input to a Python script which opens the public COALS page, parses the source HTML, and saves relevant information on providers and filings.[15] The primary output of this script is a dataset of every CUID/provider combination in the COALS system.

Merging COALS and Census Data

With our geographic data and CUID data collected at the finest levels possible, we use a "specific-to-general" process to combine the data. We map the Census Place classi-

[14]Available at http://www.fcc.gov/mb/vax/registeredcuid.xls
[15]See figure Figure 2.9 for an example CUID shown in COALS.

fications to the FCC CUID classifications according to Table 2.8. We then match the community type and the community, county, and state names as closely as possible. An overview of the match quality is tabulated in Table 2.9. Of the 45,146 CUIDs in the FCC file, we match 31,598 to Census locations. Of those 31,598 matches, 5,517 are unincorporated communities and therefore use imputed household data. Though all major cities match successfully, the CUID file contains many unmatched entries. While some of the unmatched CUIDs consist of individual housing developments or government facilities, most are unincorporated communities or areas which do not qualify as a Census Designated Place.

2.8.5 Data cleaning

The first step in our analysis is a manual cleaning process focusing on the 9,506 unique legal entities that control CUIDs at various points in time throughout our raw dataset. The vast majority of the changes come from either missing address information or typographical errors in the legal name or address.[16] Many additional changes are made through the identification of franchised or otherwise split legal entities which are in fact owned by a single corporation. These entities were identified either through analysis of their names or publicly available business databases maintained by Business Week and Funding Universe.[17] We also used SEC filings to identify lists of subsidiaries in 2000.[18]The result of this process is a mapping that links each of the 9,506 "raw" legal entities to one of "cleaned" 3,889 entities. These cleaned entities are then merged back into the original providers dataset.

With the legal entities cleaned, it is now the case that several "switches" in a single CUID may now actually be multiple entries of the same parent company. We

[16]See figure Figure 2.10 for examples of these two cases.
[17]Figure Figure 2.11 has examples of this sort of cleaning.
[18]Comcast: http://www.cmcsa.com/secfiling.cfm?filingID=950159-00-66

Municipality Type	**CUIDs**
Incorporated Borough	1,733
Incorporated City	10,873
Incorporated Town	8,878
Incorporated Village	4,211
Privately owned settlement	1,072
State or Federal Reservation	440
Unincorporated area adjacent to incorporated community	1,478
Unincorporated area commonly known as	5,809
Unincorporated unnamed area within a County or Parish	4,211
Grand Total	45,146

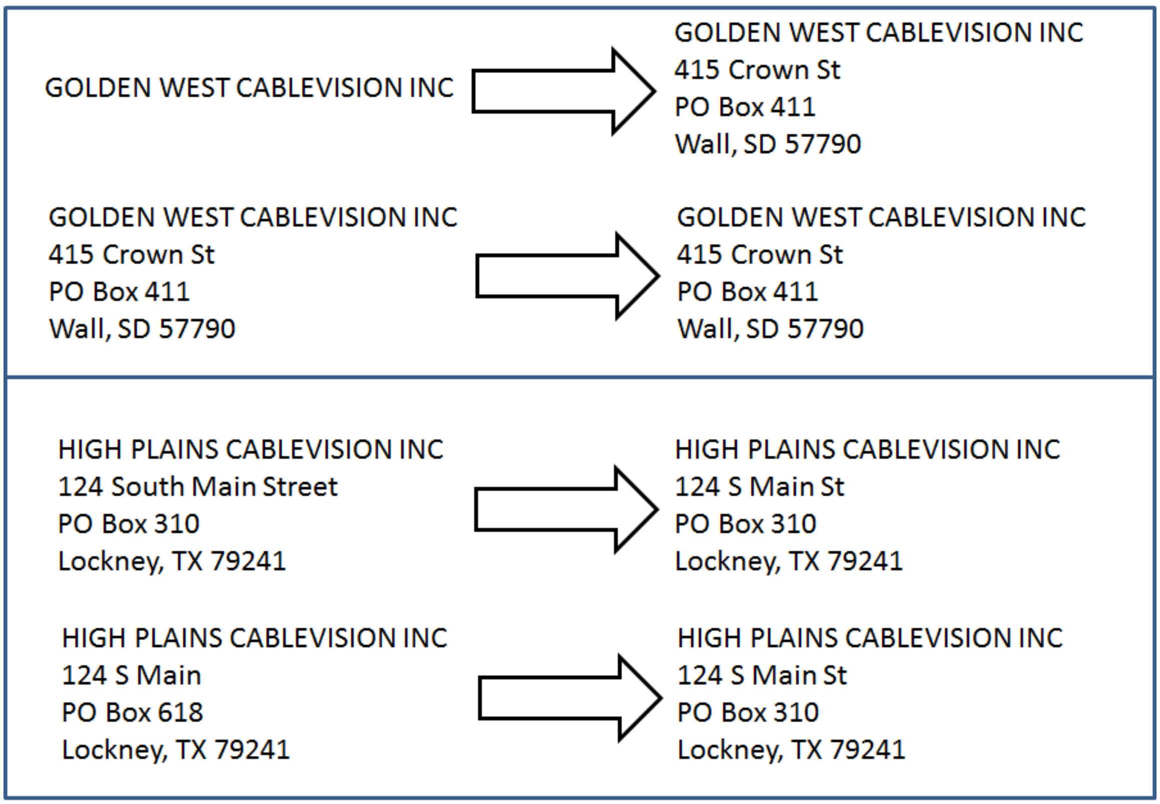
Table 2.7: CUID types identified by the FCC

CUID classification	**CDP classification**
Incorporated Borough	City
Incorporated City	City
Incorporated Town	Town
Incorporated Village	Town
Privately owned settlement	Private
State or Federal Reservation	Reservation
Unincorporated area adjacent to incorporated community	Balance
Unincorporated area commonly known as	CDP
Unincorporated unnamed area within a County or Parish	Balance

Table 2.8: Mapping CUID classifications to CDP classifications

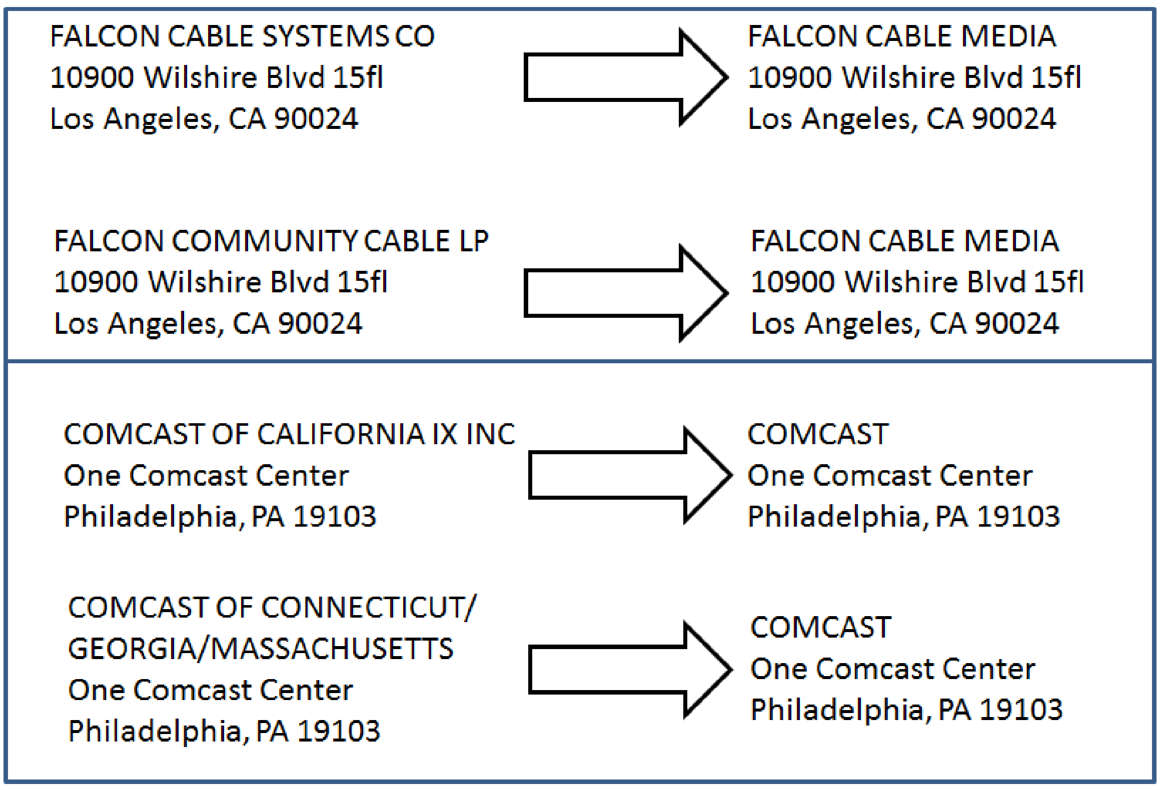
Match Type	**CUIDs**
Full (County, community type and name)	21,158
County and name	5,610
Type and name	2,210
Name only	2,620
Unmatched	13,548
Total	45,146

Table 2.9: Breakdown of CUID/Census match quality



[p]

Figure 2.10: Top: Some legal entity entries were missing address data. We filled in missing addresses using entries with identical names where available. Bottom: When multiple addresses were found (or when addresses had typos), we used the most-common entry for all identically named entities.



[p]

Figure 2.11: Top: Some legal entity differences came from subsidiaries with slightly different names. Bottom: Many cable operators operate through franchised or regionally-based subsidiaries.

perform a sifting procedure on the dataset to identify the earliest date a CUID was controlled by each of the legal entities which ever controlled the community during the period covered by COALS data. The result is a pared-down list of unique legal entities controlling CUIDs at different points in time.

We refactor this list into a set of switches, by combining multiple observations in our source data into a single observation for each switch containing information on the prior owner, the new owner, and the date of the switch. We group these switches by the two owners in question and the calendar quarter of the switch to identify mergers. These so-called "switch groups" represent the universe of possible merger events in our data.

These groups require additional manual cleaning. Although FCC rules require cable providers to inform the FCC of changes in the legal status of a CUID or cable system within 30 days of such a change,[19] we find several instances where the bulk of a change is consummated (according to the COALS providers data) on one day, and a few additional changes are made some days or months later. An example of this phenomenon is shown in Table 2.10. This process reduces the number of observed switch groups (and thus the number of mergers we report) from 896 to 713.

As a check on our data cleaning procedures, we compare our final Comcast merger list (including dates) to the data we collected from the Comcast letters. We successfully match nearly all of the 119 reported Comcast acquisitions.[20]

To understand the geographic layout of the merger, we compare the distance of each CUID within a switch group with all of the CUIDs owned by the acquiring company at the time of the switch (excluding other CUIDs acquired within the same group). Distances are calculated from latitude/longitude data with the Equirectan-

[19] 47 C.F.R 76.1610, available at http://www.gpo.gov/fdsys/pkg/CFR-2010-title47-vol4/pdf/CFR-2010-title47-vol4-sec76-1610.pdf

[20] We believe our mismatches are due to differences in the names of entities as reported by Comcast and recorded in COALS.

gular Approximation which has high accuracy over the relatively short distances we observe.

2.8.6 Horizontal purchases

We distinguish between horizontal and conglomerate purchases with a simple process: For each CUID involved in the acquisition event, we examined the list of the acquiring company's existing properties at the time of the merger for an exact community name match. If a match is found, the CUID is flagged as a potential horizontal merger. Since we cannot confirm overbuild directly, we excluded those CUIDS which referred to townships or unincorporated areas of counties and parishes. It is unlikely that companies would pursue an overbuild strategy in these rural areas.

2.8.7 FCC Annual Report Data

To ground our subscription rate assumptions, we acquired all annual report data from 2002-2009 from the FCC. The FCC requires all cable systems with greater than 20,000 subscribers, as well as a random sample of smaller systems, to submit an annual report with details of their coverage, subscription rates, and offerings. These reports are filed at the Physical System level and are integrated into COALS upon submission. While this data is considered public, the FCC has agreed to an industry request to hold the report data for three years before release.

Unfortunately, due to the design of COALS, the annual report data does not contain any point-in-time geographic linkage information. In other words, we cannot identify which historical annual report corresponds to which CUIDs. Whenever a CUID is attached to a new PSID, it is immediately linked to all filings for that PSID and all previous linkages are destroyed. For example, Verizon registered a CUID for Medford, MA (CUID MA0484) in 2012 and attached it to their existing regional

PSID, 020666. COALS lists a 2008 annual report as a relevant filing for this CUID, despite the CUID's failure to exist in that year. Unfortunately, there does not seem to be a solution to this obstacle at this time.[21]

While we cannot precisely identify which physical systems controlled which CUIDs, we regress the number of subscribers on the number of households covered by the system interacted with year dummies. This regression captures the overall decline in cable subscription rates and is used to ground the value assumptions made in our potential merger exercise.

[21] We asked the FCC to release any geographic link data (beyond the "present-time view" available in COALS) they possess under the Freedom of Information Act. Mike Perko, the Chief of the FCC's Office of Communications and Industry Information, asserted no such information existed, and that storing such information was "not in the public interest." Since the lack of such information significantly reduces the usefulness of annual report data and hampers the FCC's ability to make informed decisions, we must disagree.

Date	**CUIDs**
January 15, 2008	364
April 25, 2008	2
August 1, 2008	2
Total	368

Table 2.10: An example of different dates within a "switch group." The event shown took place between Comcast and Insight Communications Co.

Bibliography

(2009). *Hart-Scott-Rodino Premerger Notification Program Introductory Guide*. Federal Trade Commission.

(2013). *Federal Register*. 78(No. 8; January 11), 2406.

Abaluck, J. and Gruber, J. (2013). Evolving choice inconsistencies in choice of prescription drug insurance. Working Paper 19163, National Bureau of Economic Research.

Aguirregabiria, V. and Mira, P. (2007). Sequential estimation of dynamic discrete games. *Econometrica*, 75(1):1–53.

Aizawa, N. and Kim, Y. S. (2013). Advertising Competition and Risk Selection in Health Insurance Markets: Evidence from Medicare Advantage. Job market paper.

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*, LIII(5):941–973.

Association, N. C. . T. (2013a). Industry data.

Association, N. C. . T. (2013b). Our story.

Bajari, P., Benkard, C. L., and Levin, J. (2007). Estimating dynamic models of imperfect competition. *Econometrica*, 75(5):1331–1370.

Berry, S. (1994). Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, pages 242–262.

Berry, S., Levinsohn, J., and Pakes, A. (1995). Automobile prices in market equilibrium. *Econometrica*, 63(4):pp. 841–890.

Berry, S., Levinsohn, J., and Pakes, A. (1999). Voluntary export restraints on automobiles: Evaluating a trade policy. *American Economic Review*, 89(3):400–430.

Brown, J., Duggan, M., Kuziemko, I., and Woolston, W. (2011). How does Risk Selection Respond to Risk Adjustment? Evidence from the Medicare Advantage Program. Working Paper 16977, National Bureau of Economic Research.

Carlton, D. W. (2009). How to measure the effectiveness of U.S. merger policy.

Cebul, R. D., Rebitzer, J. B., Taylor, L. J., and Votruba, M. E. (2011). Unhealthy insurance markets: Search frictions and the cost and quality of health insurance. *American Economic Review*, 101(5):1842–71.

Chamberlain, G. (1987). Asymptotic efficiency in estimation with conditional moment restrictions. *Journal of Econometrics*, 34(3):305–334.

Chernew, M. (2003). General equilibrium and marketability in the health care industry. In Hammer, P. J., Haas-Wilson, D., Peterson, M. A., and Sage, W. M., editors, *Uncertain times: Kenneth Arrow and the changing economics of health care*. Duke University Press.

Clougherty, J. A. and Seldeslachts, J. (2011). The deterrence effects of U.S. merger policy instruments. CEPR Discussion Papers 8482, C.E.P.R. Discussion Papers.

CMS (2014). Medicare Managed Care Manual. Technical report.

Collard-Wexler, A. (2013). Demand fluctuations in the ready-mix concrete industry. *Econometrica*, 81(3):1003–1037.

Commission, F. C. (2011). Telecommunications act of 1996.

Commission, F. C. (2012). Evolution of cable television.

Crawford, G. (2012). Accommodating endogenous product choices: A progress report. *International Journal of Industrial Organization*, 30:315–320.

Curto, V., Einav, L., Levin, J., and Bhattacharya, J. (2014). Can Health Insurance Competition Work? The U.S. Medicare Advantage Program. Working paper.

Dagum, L. and Menon, R. (1998). OpenMP: an industry standard API for shared-memory programming. *Computational Science & Engineering, IEEE*, 5(1):46–55.

Duggan, M., Starc, A., and Vabson, B. (2014). Who benefits when the government pays more? pass-through in the medicare advantage program. Working Paper 19989, National Bureau of Economic Research.

Eisenmann, T. R. (2000). Cable TV: From community antennas to wired cities.

Emmons, W. M. and Prager, R. A. (1997). The effects of market structure and ownership on prices and service offerings in the U.S. cable television industry. *RAND Journal of Economics*, 28(4):732–750.

Enthoven, A. C. (1978). Consumer-Choice Health plan (second of two parts). A national-health-insurance proposal based on regulated competition in the private sector. *N. Engl. J. Med.*, 298(13):709–720. [DOI:10.1056/NEJM197803302981304] [PubMed:415241].

Eppig, F. and Chulis, G. S. (1997). Matching MCBS and Medicare Data: The Best of Both Worlds. *Health Care Financing Review*, 18(3):211–229.

Ericson, R. and Pakes, A. (1995). Markov-perfect industry dynamics: A framework for empirical work. *The Review of Economic Studies*, 62(1):53–82.

Farrell, J. and Klemperer, P. (2007). Chapter 31 coordination and lock-in: Competition with switching costs and network effects. volume 3 of *Handbook of Industrial Organization*, pages 1967 – 2072. Elsevier.

Gabel, J., Jajich-Toth, C., de Lissovoy, G., Rice, T., and Cohen, H. (1988). The changing world of group health insurance. *Health Affairs*, 7(3):48–65.

Gandhi, A., Lu, Z., and Shi, X. (2013). Estimating Demand for Differentiated Products with Error in Market Shares. Working paper.

Gaynor, M. and Town, R. J. (2011). Chapter nine - competition in health care markets. In Mark V. Pauly, T. G. M. and Barros, P. P., editors, *Handbook of Health Economics*, volume 2 of *Handbook of Health Economics*, pages 499 – 637. Elsevier.

Glied, S. (2000). Chapter 13 managed care. volume 1, Part A of *Handbook of Health Economics*, pages 707 – 753. Elsevier.

Goolsbee, A. and Petrin, A. (2004). The consumer gains from direct broadcast satellites and the competition with cable TV. *Econometrica*, 72(2):351–381.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. *American Economic Review*, 103(7):2643–82.

Ho, K., Hogan, J., and Morgan, F. S. (2014). The Impact of Consumer Inattention on Insurer Pricing in the Medicare Part D Program. Working paper.

Hotz, V. J. and Miller, R. A. (1993). Conditional choice probabilities and the estimation of dynamic models. *The Review of Economic Studies*, 60(3):497–529.

Kahan, W. (1965). Pracniques: Further remarks on reducing truncation errors. *Communications of the ACM*, 8(1):40.

Kaiser Family Foundation (2014). Medicare advantage fact sheet. Technical report.

Lustig, J. (2011). Measuring Welfare Losses from Adverse Selection and Imperfect Competition in Privatized Medicare. Working paper.

Markovich, M. (2003). *The Rise of HMOs*. PhD thesis, RAND Graduate School.

Miller, K., Petrin, A., Town, R., and Chernew, M. (2014). Health Plan Changes in Response to Medicare Advantage Payment Rates. Working paper.

Moon, M., Burke, S., DeArment, R., Feder, J., Helms, R., Hoffman, W., Maher, W., Phillps, M., Rappaport, A., Scanlon, W., Tallon, J., and Valdez, R. (2000). Financing Medicare's Future. Technical report, National Academy of Social Insurance.

Newey, W. K. and McFadden, D. (1994). Chapter 36 large sample estimation and hypothesis testing. volume 4 of *Handbook of Econometrics*, pages 2111 – 2245. Elsevier.

Newhouse, J. P., Price, M., McWilliams, J. M., Hsu, J., and McGuire, T. (2014). How much favorable selection is left in medicare advantage? Working Paper 20021, National Bureau of Economic Research.

Nielsen (2011). Nielsen estimates number of U.S. television homes to be 144.7 million.

Nyman, J. A., Barleen, N. A., Dowd, B. E., Russell, D. W., Coons, S. J., and Sullivan, P. W. (2007). Quality-of-life weights for the US population: self-reported health status and priority health conditions, by demographic characteristics. *Med Care*, 45(7):618–628.

of Broadcast Communications, T. M. (2013). Freeze of 1948.

Padilla, A. (2001). Cable 'overbuild' puts choice in hands of consumers. *L.A. Times*.

Pakes, A. and McGuire, P. (2001). Stochastic algorithms, symmetric markov perfect equilibrium, and the 'curse' of dimensionality. *Econometrica*, 69(5):1261–1281.

Pakes, A., Ostrovsky, M., and Berry, S. (2007). Simple estimators for the parameters of discrete dynamic games (with entry/exit examples). *The RAND Journal of Economics*, 38(2):373–399.

Pakes, A., Porter, J., Ho, K., and Ishii, J. (2006). Moment inequalities and their application. *Working Paper, Harvard University*.

Pear, R. and Bogdanich, W. (2003). Some Successful Models Ignored As Congress Works on Drug Bill. *The New York Times*.

Qi, S. (2013). The impact of advertising regulation on industry: the cigarette advertising ban of 1971. *The RAND Journal of Economics*, 44(2):215–248.

Rust, J. (1987). Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher. *Econometrica: Journal of the Econometric Society*, pages 999–1033.

Ryan, S. P. (2012). The costs of environmental regulation in a concentrated industry. *Econometrica*, 80(3):1019–1061.

Sanderson, C. (2010). Armadillo: An Open Source C++ Linear Algebra Library for Fast Prototyping and Computationally Intensive Experiments. Technical report, NICTA.

US Department of Health and Human Services, Health Sources and Services Administration, Bureau of Health Workforce, Rockville, MD (2013-2014). Area Health Resources Files. Available at `htpp://ahrf.hrsa.gov`.

Weintraub, G. Y., Benkard, C. L., and Van Roy, B. (2008). Markov perfect industry dynamics with many firms. *Econometrica*, 76(6):1375–1411.

Weintraub, G. Y., Benkard, C. L., and Van Roy, B. (2010). Computational methods for oblivious equilibrium. *Operations research*, 58(4-part-2):1247–1265.

Youle, T. (2014). How much did manipulation distort the libor? Working paper.