The Star Treatment: Estimating the Impact of Star Ratings on Medicare Advantage Enrollments

Michael Darden*

Department of Economics

Tulane University

Ian M. McCarthy†

Department of Economics

Emory University

May 2014

Abstract

The Centers for Medicare and Medicaid Services (CMS) has calculated and disseminated an overall contract quality star rating system (from one to five stars) for all Medicare Advantage (MA) contracts since 2009. In this paper, we study the effect of CMS-reported star ratings on MA plan enrollment. We formulate a discrete choice demand model for differentiated MA plans and estimate the model with market-level plan enrollment data. We identify separate enrollment effects for each star level using a regression discontinuity research design that exploits plausibly random variation around star thresholds. The results suggest that the 2009 published star ratings directed beneficiaries away from low rated plans more than actively toward high rated plans. When we repeat the analysis for 2010 published quality stars, we find no significant

*206 Tilton Memorial Hall, Tulane University, New Orleans, LA 70115. E-mail: mdarden1@tulane.edu

†1602 Fishburne Dr., Rich Building Room 306, Atlanta, Georgia 30322. E-mail: ianmccarthy.econ@gmail.com

effects. We present suggestive evidence that supply-side responses to the star rating system may explain the one-time enrollment response to CMS-published quality stars.

JEL Classification: C11; C21; D21; D43; M31

Keywords: Medicare Advantage, CMS Star Ratings, Health Plan Quality, Health Plan Choice, Regression Discontinuity

1 Introduction

Since the Medicare Modernization Act of 2003, the role of Medicare Advantage (MA) plans in the provision of health insurance to Medicare beneficiaries has grown substantially. From 2003 to 2012, the share of Medicare eligible individuals in a MA health plan increased from 13.7% to 27%.[1] Alongside the increasing role of MA plans, the Centers for Medicare and Medicaid Services (CMS) has undergone a significant effort to better inform the Medicare population of MA quality. In 2007, CMS introduced a five-star rating system that provided a rating of one to five stars to each MA *contract* -- a private organization that administers potentially many differentiated plans -- in each of five quality domains.[2] In 2009, CMS began aggregating the domain level quality scores into an overall star rating for each MA contract that was made available to beneficiaries via the Medicare Plan Finder website and 1-800-MEDICARE. Finally, beginning in the 2012 enrollment period, contracts were incentivized to receive a high rating through rating-dependent reimbursement and bonus schemes. Given the substantial efforts involved in calculating and disseminating the star ratings, the new incentives on plans to achieve a high rating, and the growing role of the MA program in the provision of health insurance to Medicare beneficiaries, understanding the impact of star ratings on MA enrollment is critical to future Medicare policy.

While disseminating quality information may theoretically improve market outcomes (Akerlof, 1970), surprisingly few studies have empirically examined the impact of publicly available quality measures on Medicare health plan choice. A recent exception is Reid *et al.* (2013), who study the impact of star ratings on MA plan choice using individual-level CMS data. However, the analysis in Reid *et al.* (2013) does not differentiate between overall quality beliefs and *CMS-reported* quality information. Alternatively, Dafny & Dranove (2008) separately identify the effect of CMS-reported quality from other sources of quality information on plan enrollment. They find that the publication of MA quality information in 2001 did affect enrollment, but published CMS quality information was less important than other sources of quality information available. How-

[1] Kaiser Family Foundation MA Update, available at www.kff.org/medicare/upload/8323.pdf.

[2] For example, one domain on which contracts were rated was ``Helping You Stay Healthy."

ever, the quality information available during the study time frame was limited to the 2000-2001 ``Medicare & You'' CMS booklet, which published a variety of disaggregated quality metrics (e.g., the percentage of women receiving a mammogram) rather than the star rating system currently in place. The private provision of Medicare health insurance has also grown significantly since the time frame studied in Dafny & Dranove (2008).

In this paper, we analyze the effect of the marginal quality star on MA plan enrollment. We formulate a discrete choice demand model for differentiated MA plans, and we estimate our model with data on plan enrollment, plan characteristics, and market-area information from 2009 and 2010. Our empirical model is consistent with a nested logit demand model of differentiated products (Berry, 1994). We define a nesting structure that allows for flexible substitution patterns between plans by defining nests on the basis of prescription drug coverage and positive monthly premiums.[3] Our dependent variable is the plan's log market share relative to the log market share of traditional Medicare fee-for-service (FFS), which serves as a common outside option for all Medicare eligibles in all markets.

We contribute to the literature on the enrollment effects of CMS quality reporting in three important ways. First, within a theoretically grounded model of plan choice, we identify the effect of quality star reporting on a plan's log relative market share with a novel regression discontinuity (RD) design that exploits plausibly random variation around star thresholds. We replicate CMS's underlying, continuous contract quality score and examine log relative share changes within a bandwidth around each star threshold. Importantly, our research design allows us to differentiate the effect of *CMS-reported* quality information on enrollment from other sources of quality infor-

[3]Dafny & Dranove (2008) only nest MA plans, and Town & Liu (2003) delineate contracts by whether the contract offers prescription drug coverage. In 2009, only 58% of MA plans charged a positive monthly premium, and we provide evidence that substitution patterns may differ within, as compared to across, these plans.

mation and prior quality beliefs.[4] Ramanarayanan & Snyder (2012) adopt a similar RD approach in their study of the impact of CMS rating systems in the dialysis industry.

Second, we extend the current literature, namely Dafny & Dranove (2008) and Town & Liu (2003), by modeling enrollment at the plan level. Due to data limitations, those authors were forced to aggregate enrollment data to the contract level and thus not able to model how plan characteristic variation within MA contracts may influence plan enrollment. Although CMS assigns star ratings and grants permission to enroll Medicare beneficiaries at the contract level, a specific contract will generally contain several differentiated plans. While all plans operating under a given contract receive the same quality score, these plans are differentiated on the basis of plan attributes such as premium and prescription drug coverage. By estimating our model of plan choice at the plan level, we more precisely estimate the effect of quality reporting on plan enrollment and avoid the plan characteristic aggregation assumptions of other papers.

Finally, while Reid *et al.* (2013) assume a linear effect of quality stars on enrollment, we separately examine enrollment patterns around different star thresholds. Our analysis therefore captures the potentially different enrollment effects of the marginal quality star across the CMS-reported quality distribution.[5] Furthermore, we show that treating star ratings as exogenous may understate the marginal star enrollment effect in 2009, particularly at the low end of the star rating distribution.

We find generally positive and significant effects of CMS-reported stars on plan enrollment in 2009, with larger effects for the lower star-rated contracts. Specifically, plan market share is found to increase by 4.75 percentage points relative to traditional FFS for plans with 3 stars relative to 2.5 stars (a 1.5 standard deviation increase in relative market share). This effect drops to 2.74

[4]This distinction is a common concern in the literature on the effects of information campaigns. For example, Dranove & Sfekas (2008), in studying the effect of hospital report cards on hospital market share, use the example that a positive quality report is unlikely to increase the market share of the Mayo Clinic.

[5]Other papers that investigate the effects of quality information dissemination also find differential effects at different points in the quality distribution. For example, see Scanlon *et al.* (2002) on health plan demand and Dranove & Sfekas (2008) on hospital demand.

percentage points (a 0.9 standard deviation increase) for plans with 4 stars relative to 3.5 stars. In context, a 4.75 percentage point increase in relative market share is equivalent to 131 additional enrollments among smaller counties and 1,353 additional enrollments among larger counties.[6] Our results are robust to different bandwidth and nesting specifications.

When we repeat the analysis for 2010, we find smaller effects, none of which are statistically significant. Although our market level data are not sufficiently rich to explain *why* the quality star system yielded enrollment shifts in 2009 and not in 2010, we present suggestive evidence for two supply-side responses to the 2009 ratings that may explain our observed differences in results. First, we observe a disproportionate percentage of plans from low star-rated contracts in 2009 exiting the market between 2009 and 2010, suggesting that lower star-rated contracts that are observed in both years dropped plans at a greater rate than higher star-rated contracts. Second, our results suggest that marginally higher rated contracts in 2009 raised premiums from 2009 to 2010, potentially pricing out any quality advantages.

Our results have implications within the CMS rating program and potentially broader implications for future rating programs. Within the current program, our results indicate that the rating program has not meaningfully driven beneficiaries toward the highest rated plans. The Patient Protection and Affordable Care Act (ACA) now ties bonus and other incentive payments with the CMS star ratings; however, our results suggest that the incentive to increase quality stars for enrollment purposes is relatively minimal at the high end of the star rating distribution, making the CMS bonus program the predominant incentive in place for contracts to achieve the highest possible star ratings. Considering the design of future rating programs (e.g., rating plans on the insurance exchanges), our results also highlight potential unintended consequences via supply-side responses from one year to the next.

[6]For purposes of interpretation, we define a small county as any county in the bottom 50th percentile of Medicare eligible population (2,760 eligibles on average), and conversely for large counties (28,505 eligibles on average). Our market share effects in 2009 are similar in magnitude to those of Dafny & Dranove (2008); however, our effects are more concentrated at the lower end of the star rating distribution.

This paper proceeds as follows. We summarize the background of the MA program in Section 2, including the recent star rating program. Details of the star rating program and calculations are presented in Appendices A-D. Section 3 reviews the current literature on the role of quality and quality reporting on health plan choice. Section 4 presents our proposed methodology and relevant econometric background. Section 5 summarizes the data used in our analysis. Section 6 presents the results of our analysis along with a series of robustness checks and an investigation into the differences in our 2009 and 2010 results, and Section 7 concludes.

2 Medicare Advantage

The Balanced Budget Act of 1997 (BBA) introduced private health insurance options known as Medicare + Choice plans (M+C, or Medicare Part C), which included traditional HMOs but allowed additional plan types including Preferred Provider Organizations (PPOs), Provider-Sponsored Organizations (PSOs), Private FFS (PFFS) plans, and Medical Savings Accounts (MSAs). The introduction of Medicare Part C generated an influx of supplemental care options for the Medicare population. Medicare Part C was then revised in 2003 under the Medicare Modernization Act and renamed Medicare Advantage (MA), which allowed additional plan types such as Regional PPOs and Special Needs Plans (SNPs) as well as prescription drug coverage. MA plans, like the preceding Medicare Part C plans, are provided by private companies who contract with CMS yearly. Contracts are approved at the county level, and there can be multiple individual plans provided under each contract. As such, only those residents living in the county in which the contract applies have access to the plans under that contract.

By choosing an MA plan, beneficiaries no longer receive the traditional benefits of Medicare FFS but must still enroll in Medicare Parts A and B and pay the Part B premium. CMS requires that MA plans offer at least what the beneficiary could receive from Medicare FFS. Most MA contracts are also required to offer at least one plan that includes prescription drug coverage. MA plans are therefore a form of voluntary supplemental coverage for the Medicare population. Other avenues of supplemental coverage available for Medicare beneficiaries include Medicaid, Medigap, and employer-provided insurance.

For a given calendar year, beneficiaries can enroll in MA plans during an open enrollment period from November 1 through December 31 of the prior year. During this time, eligibles aged 65 years or older without end stage renal disease can choose any plan available in their area. Limited enrollment continues through June 30 of the current year, during which time eligibles may only enroll in an MA plan if the plan is currently accepting new members. Beneficiaries can also switch plans during the limited enrollment period, but only one switch is allowed per year.

As of September 2012, MA plans provided coverage to 13.7 million beneficiaries (approximately 27% of the total Medicare population) based on a total of 2,074 MA plans across the country, with an average of 20 plans available to a given beneficiary.[7] This is a substantial increase in plans from 2002, where just 452 total plans based on 174 contracts were available across the U.S., covering a total of 5.3 million Medicare beneficiaries.

Although existing plans offer a variety of optional benefits, the abundance of MA plan choices can be overwhelming for much of the Medicare population. In a study commissioned by the AARP, Hibbard *et al.* (1998) found that most beneficiaries do not fully understand the differences between the standard Medicare FFS and Medicare Part C. Atherly (2001), Frank (2004), and Heiss *et al.* (2006), among others, have also found strong survey evidence that the choice set is overly complicated and that beneficiaries often do not make fully informed decisions.

CMS has historically recognized the potential difficulty for Medicare beneficiaries to make informed decisions regarding their health plans. For example, under the BBA, CMS was mandated to provide health plan quality information to Medicare beneficiaries. Among other things, CMS responded with a ``Medicare & You'' handbook mailed to all Medicare eligibles at the start of the open enrollment period. This handbook was first circulated to eligibles in November 1999 (relevant for coverage in 2000) and included a variety of measures intended to better inform ben-

[7]Kaiser Family Foundation MA 2013 Spotlight, available at www.kaiserfamilyfoundation.files.wordpress.com/2013/01/8388.pdf.

eficiaries of the quality of plans available in their respective geographic areas.

Building upon these prior quality reporting initiatives, CMS launched a star rating program in 2007 by which contracts are rated from one to five stars based on several measures covering five domains: 1) ``helping you stay healthy"; 2) ``getting care from your doctors and specialists"; 3) ``getting timely information and care from your health plan"; 4) ``managing chronic conditions"; and 5) ``your rights to appeal." Since 2007, the individual measures and domains have changed nearly every year. For example, CMS expanded the ``rights to appeal" domain in 2010 to include measures on complaints and the number of beneficiaries leaving the plan, among others. Also in 2010, the ``timely information" domain was replaced by ``customer service." Appendix A details the individual measures and domains for 2009.

In each year, individual measures are quantified and given a star rating based on the criteria published by CMS. For example, in 2008, the ``helping you stay healthy" domain included 8 individual measures. One such individual measure was ``breast cancer screening," measured as the percent of female MA enrollees ages 50 to 69 who had a mammogram during the measurement year or the year prior to the measurement year. The contract received a star rating for this measure based on its percentile rank, with 1-star, 2-star, 3-star, 4-star, and 5-star ratings based on the 15th, 35th, 65th, 85th, and 100th percentile, respectively. Data for this measure were collected from the Healthcare Effectiveness Data and Information Set (HEDIS), previously known as the Health Plan Employer Data and Information Set.

Star ratings are calculated for each individual measure based on data collected from a variety of sources, including HEDIS, the Consumer Assessment of Healthcare Providers and Systems (CAHPS), the Health Outcomes Survey (HOS), the Independent Review Entity (IRE), the Complaints Tracking Module (CTM), and CMS administrative data. These star ratings are then averaged to generate a star rating for each domain. Beginning in 2009, the star ratings of individual measures were also averaged to generate an overall star rating for the contract (after additional adjustments by CMS intended to reward consistency across individual metrics), rounded to the nearest half-star. The star rating for each domain and each individual metric are still available on the plan finder website, but are not provided as part of the high level summary and comparison of each plan. Appendix B details the star rating calculations for 2009 and provides example calculations for five contracts.

3 Evidence on the Effects of Quality Reporting

The current study concerns the impact of star ratings on MA enrollments and relates generally to the literature on quality ratings and health plan choice (Beaulieu, 2002; Chernew *et al.*, 1998; Dafny & Dranove, 2008; Farley *et al.*, 2002; Jin & Sorensen, 2006; Scanlon *et al.*, 2002; Spranca *et al.*, 2000; Wedig & Tai-Seale, 2002).[8] Existing studies in this area have relied on a variety of techniques and datasets in order to more accurately estimate the effect of quality information on a consumer's health plan choice.

In a laboratory setting, Spranca *et al.* (2000) found that consumers with access to quality ratings were more likely to choose higher rated but less expensive, less comprehensive plans. Conversely, subjects without access to quality ratings were more likely to choose more expensive plans offering more comprehensive coverage. These results suggest an important role for quality ratings - namely, improved matching between consumers and plans. The results also suggest that, in the absence of quality ratings, consumers may rely more heavily on decision rules to aid in their choice of health plans. Screening rules have been the subject of recent studies in health plan choice (Li & Trivedi, 2012; Schram & Sonnemans, 2011); however, such studies generally do not consider alternative decision rules among those with and without access to plan quality information.

Studies based on observational data have focused largely on the private insurance market. Beaulieu (2002), for example, analyzed plan choice and switching behavior using data on Harvard University health insurance plans from 1995 through 1997. Comparing switching behaviors

[8]See Dranove & Jin (2010) for a thorough review of the broader literature on the effects of quality disclosure on consumer choice and seller behaviors.

from 1995 to 1996 (when health plan quality information was not made available to enrollees) to switching behavior in 1996 to 1997 (when quality information was available), she found that enrollees responded (albeit modestly) to quality reports by switching away from lesser quality plans and concluded that quality reports provided additional information beyond what consumers independently obtained from experience. She also found evidence that older employees used quality information differently than younger employees.

Scanlon *et al.* (2002) analyzed the impact of quality ratings on health plan choice for General Motors' non-union employees in 1997. Aggregating data to the market share level, the authors found that individuals avoided plans with several below average ratings; however, the analysis could not identify the impact of specific health quality domains on choice (e.g., the impact of patient satisfaction ratings versus medical/surgical care). Alternatively, Chernew *et al.* (1998) modeled the impact of quality ratings in a Bayesian learning framework, where quality ratings served as a signal with which employees updated their prior beliefs. Experience was therefore implicitly accounted for in employees' prior beliefs. The authors found that quality ratings had a modest but significant impact on health plan choice. The authors also found evidence of heterogeneity across dimensions in which health plans were rated, with medical/surgical ratings having the largest impact.

Wedig & Tai-Seale (2002) analyzed the impact of report card ratings distributed to federal employees under the Federal Employees Health Benefit Plan (FEHBP). Comparing health plan choices in 1995 (prior to report cards being issued to enrollees) to choices in 1996 (after report cards were made available), the authors found that report card ratings on quality of care had a significant impact on health plan choice, particularly on new employees, while report card ratings on plan coverage impacted existing employees and showed no significant effect for new employee plan choice. The study also found that report cards increase the price elasticity of demand, indicating that report card ratings also impact how enrollees value other plan attributes.

Using FEHBP data from 1998 to 1999, Jin & Sorensen (2006) studied the impact of NCQA ratings on health plan choice among retirees and surviving family members of federal employees. The dataset included two types of plans: 1) plans that allowed quality ratings to be distributed to enrollees; and 2) plans for which ratings were calculated but were not made public. The data and methodology therefore allowed the analysis to distinguish between quality information provided by health plan ratings versus other quality signals known to consumers but not observed by the researcher. The authors found that NCQA ratings had an important impact on individuals selecting a plan for the first time in their geographic area and those who switched plans from 1998 to 1999.

Dafny & Dranove (2008) provide one of the few studies to date on the impact of health plan ratings for Medicare beneficiaries. The authors analyzed the impact of health plan ratings published in the 2000 and 2001 ``Medicare & You'' booklets on Medicare HMO enrollees from 1994 to 2002. Comparing plan switching behaviors after 2000 to switching behaviors prior to 2000, the authors found little evidence that the ``Medicare & You'' booklets offered any additional quality information beyond existing ratings from *U.S. News & World Report*, word of mouth, or prior experience. Of what relatively small effects emerged from the ``Medicare & You'' booklets, the authors identified consumer satisfaction scores as the only metric impacting beneficiaries' switching behaviors. The authors concluded that public quality reporting initiatives offered relatively little information not already revealed through market-based learning (e.g., grading systems from other sources, personal experience, and word of mouth).

Most recently, Reid *et al.* (2013) studied the impact of star ratings on MA plan choice using individual-level plan choice data from CMS. Similar to Wedig & Tai-Seale (2002) and Jin & Sorensen (2006), the authors found that star ratings had the largest positive effect on first-time MA enrollees. Importantly, the analysis in Reid *et al.* (2013) did not differentiate between plan quality versus *reported* quality information. This is a critical area of distinction, as reflected by previous efforts in the literature to parse out these different effects (Scanlon *et al.*, 2002; Wedig & Tai-Seale, 2002; Jin & Sorensen, 2006; Dafny & Dranove, 2008). Our paper follows this area of research in attempting to quantify the role of star ratings as a reporting mechanism rather than as an overall indicator of quality. Given the variety of avenues by which beneficiaries can gauge a plan's quality, the value of star ratings solely as a reporting mechanism remains a critical and policy-relevant question that is unaddressed in the current literature.

4 Methodology

Following Berry (1994), and the subsequent, voluminous literature on demand estimation, we estimate a discrete choice model in which a Medicare eligible individual maximizes her utility over a menu of Medicare options available in her market area. The set of differentiated products from which an individual may select is specific to an individual's area of residence. In all markets, an individual may opt for traditional Medicare FFS, which we define as the outside option $j = 0$. Alternatively, an individual in market area m may select a contract(plan), $j(p)$, from the set $j \in \{0, 1(p) \ldots, J^m(p)\}$.[9] Due to data limitations, previous studies of Medicare plan choice with market-level data aggregated the decision model to the contract level, including Dafny & Dranove (2008) and Town & Liu (2003); however, there are often different plans within a contract that are differentiated on the basis of premiums, deductibles, and prescription drug benefits. Indeed, Dafny & Dranove (2008) and Town & Liu (2003) use the lowest premium plan within a contract as the contract premium. Given data on plan characteristics, we are able to identify a richer model of plan choice.

Let the utility of individual i from selecting Medicare option $j(p)$ in market area m be given as

$$U_{ij(p)m} = \delta_{j(p)m} + \xi_{j(p)m} + \zeta_{ig} + (1-\sigma)\epsilon_{ij(p)m}, \tag{1}$$

where $\delta_{j(p)m}$ and $\xi_{j(p)m}$ represent the mean level of utility derived from observed and unobserved contract-plan-market area characteristics, respectively. Following the nested logit structure of Berry (1994), we partition the set of Medicare options into groups, indexed by g, as follows:

[9] There is variation in the number of plans offered by each contract, but we omit this notation for brevity.

1) MA plans that offer prescription drug coverage (MA-PD plans) with no monthly premiums; 2) MA-PD plans with positive monthly premiums; 3) MA plans that do not offer prescription drug coverage (MA-Only plans) with no monthly premiums; 4) MA-Only plans with positive monthly premiums; and 5) traditional FFS ($g = j = 0$). In addition to the i.i.d. extreme value error $\epsilon_{ij(p)m}$, individual preferences are allowed to vary through group dummies ζ_{ig}. This nested logit structure relaxes the independence of irrelevant alternatives assumption and allows for differential substitution patterns between nests. The nesting parameter, σ, captures the within-group correlation of utility levels.

To capture the effect of quality reporting on enrollment, we specify mean utility as follows:

$$\delta_{j(p)m} = q_j \gamma + X_{j(p)m}\beta - \alpha F_p,$$

where q_j denotes the quality of contract j and $X_{j(p)m}$ are contract, plan, and market-level observed characteristics, including a dummy variable for Part D participation and market level data on the age, race, and education profile of a given county.[10] If the plan charges a monthly premium, F, then α captures the price effect.

Berry (1994) shows how to consistently estimate the parameters of utility function (1) by integrating out the individual level variation in preferences. If we assume that $\epsilon_{ij(p)m}$ follows a multivariate extreme value distribution, then from Cardell (1997), it follows that $\zeta_{ig} + (1-\sigma)\epsilon_{ij(p)m}$ is also an extreme value random variable. The relative probability that an individual in market area m will select option $j(p)$, as compared to Medicare FFS, therefore has the following closed-form:

$$ln(P_{j(p)m}) - ln(P_{0m}) = \delta_{j(p)m} + \sigma ln(P_{j(p)m|g}) + \xi_{j(p)m}.$$

Here, $P_{j(p)m|g}$ is the conditional probability of an individual enrolling in option $j(p)$ within group g. Applying market share data as empirical estimates of the probabilities yields our final estimation

[10] Our specification makes the simplifying assumption that individuals derive utility directly from quality ratings.

equation

$$ln(S_{j(p)m}) - ln(S_{0m}) = \delta_{j(p)m} + \sigma ln(S_{j(p)m|g}) + \xi_{j(p)m}, \tag{2}$$

where $S_{j(p)m}$ denotes the share of individuals (relative to all Medicare eligibles) enrolling in option $j(p)$ in market area m, $S_{j(p)m|g}$ denotes the within-group market share of option $j(p)$, and $\xi_{j(p)m}$ denotes the mean utility derived from unobserved plan characteristics.

Most studies of differentiated products recognize that observed product characteristics are likely to be correlated with the mean utility derived from unobserved product and market characteristics, $\xi_{j(p)m}$. Indeed, the standard approach to recovering unbiased estimates treats observed product characteristics, $X_{j(p)m}$, as exogenous after product fixed effects and instruments for prices with those of other markets (Town & Liu, 2003); however, we cannot use a fixed effects approach at the contract level because quality information is only disseminated at the contract level. Moreover, as reflected in Dafny & Dranove (2008) and others, a beneficiary's knowledge of contract quality is made up of several individual components. Our interest lies in the specific informational value of the CMS star rating program, separate from other sources of quality information available to beneficiaries. Therefore, the overall quality component of an individual's utility, q_j, can be partitioned into a prior quality component and a potentially informative signal in the form of CMS star ratings, which are likely to be correlated.

Exploiting the nature of the CMS rating system, we propose an alternative methodology for the estimation of the effect of CMS star ratings based on a regression discontinuity (RD) design.[11] The RD design is intended to simulate a randomized trial by limiting the analysis to a window around threshold values. As discussed in Lee & Lemieux (2010), the RD design more closely resembles a randomized trial than some of the more common empirical approaches, including instrumental variables and fixed effects estimation. In our setting of plan choice and contract quality, the RD

[11] The RD design was first introduced by Thistlethwaite & Campbell (1960) but has only recently gained popularity in the economics literature, spurred in large part by Imbens & Lemieux (2008). Ramanarayanan & Snyder (2012) also provide a recent application of RD design to the estimation of the effect of quality ratings in the dialysis industry.

design focuses the analysis on contracts where assignment into treatment (e.g., a half-point improvement in star rating) is arguably random. In other words, a contract with an overall summary score of 2.26 is essentially identical in overall quality to a contract with a summary score of 2.24, even though the first contract will receive a star rating of 2.5 while the second contract will receive a star rating of 2.0. The two contracts therefore differ only in terms of *CMS-reported* quality.

Following Imbens & Lemieux (2008), our methodology proceeds by first restricting the sample to a pre-specified bandwidth, h, around each of four star rating thresholds. For example, to analyze the impact of improving from 2.0 to 2.5 stars, the sample is restricted to contracts with summary scores of 2.25 $\pm h$, where contracts with scores less than 2.25 were rounded to 2.0 stars and contracts with scores of 2.25 or greater were rounded to 2.5 stars. We then define R_j as the underlying summary score, $\hat{R}$ as the threshold summary score at which a new star rating is achieved (e.g., $\hat{R} = 2.25$ when considering the 2.5 star rating), and $\tilde{R}_j = R_j - \hat{R}$, which represents the amount of improvement necessary to achieve an incremental improvement in rating. To implement our approach, we specify the quality component of utility, q_j, as follows:

$$q_j = \gamma_1 + \gamma_2 \times \mathbf{I}\left(R_j > \hat{R}\right) + \sum_{d=1}^{D} \left[\gamma_3 \times \tilde{R}_j^d + \gamma_4 \times \mathbf{I}\left(R_j > \hat{R}\right) \times \tilde{R}_j^d\right], \tag{3}$$

where interest ultimately lies in γ_2 and where D denotes the order of the polynomial. Setting $D = 1$ therefore amounts to a local linear regression within the specified bandwidth around each star rating threshold. Incorporating equation (3) with the mean utility specification yields the final regression equation,

$$\begin{aligned} ln(S_{j(p)m}) - ln(S_{0m}) &= \gamma_1 + \gamma_2 \times \mathbf{I}\left(R_j > \hat{R}\right) + \textstyle\sum_{d=1}^{D} \left[\gamma_3 \times \tilde{R}_j^d + \gamma_4 \times \mathbf{I}\left(R_j > \hat{R}\right) \times \tilde{R}_j^d\right] \\ &\quad + X_{j(p)m}\beta - \alpha F_p + \sigma ln(S_{j(p)m|g}) + \xi_{j(p)m}. \end{aligned} \tag{4}$$

Our application of the RD design differs from that of the standard RD application in one critical way. Typically, the assignment variable is measured at the same level as the outcome variable so that plausibly random variation around the discontinuity will explain subsequent variation in outcomes. In our application, however, quality is reported as a contract-level characteristic while market shares are at the plan level. Multiple plans operating under the same contract will therefore have different market shares with no variation in CMS-reported quality.

Moreover, even with purely randomized quality scores at the contract level, we would not expect to see randomization in plan or county-level variables, just as one would not expect to see randomization in household characteristics in a survey with randomly selected states or counties. Because of this institutional fact, plan and county level covariates do not merely serve to reduce the variation in coefficient estimates as in the case with the standard RD design (Imbens & Lemieux, 2008; Lee & Lemieux, 2010). Instead, covariates come directly from a utility-maximizing model of health plan choice and serve the traditional regression purpose to control for plan and county level characteristics that likely influence a plan's market share. Specific covariates included in our analysis are county level demographic data (unemployment rate, population over 65, population over 85, percent female, percent black, percent white, percent of college graduates, and a dummy for southern counties), plan premium, drug copays, and dummy variables for whether the plan offers prescription drug coverage, whether the plan charges a positive monthly premium, and plan type (HMO, PPO, or PFFS).

Finally, if the star ratings have an impact on overall market share, then the within-group shares will be endogeneous and tend to bias the estimate of γ_2. We therefore estimate equation 4 using two-stage least squares (within the bandwidth around each star rating), using age of the contract and the number of within-group plans as instruments for the within-group share term, $S_{j(p)m|g}$, the latter of which has also been used in Town & Liu (2003) and Dafny & Dranove (2008).

5 Data

To estimate equation 4, we collect data on market shares, contract/plan characteristics, and market area characteristics from several publicly available sources from June 2008 through December

2011. The unit of observation is the contract/plan/county/month/year. However, because our identification strategy exploits plausibly random variation that is cross-sectional, we do not exploit the panel nature of the contract/plan records. Instead, consistent with the MA open enrollment period, we model the mean market share over a year as a function of CMS-reported contract quality during the previous open enrollment period. We report summary statistics and results for 2009 and 2010 enrollments based on quality information that was first reported in the previous October.

Market share data are constructed from the Medicare Service Area files, which list all approved MA contracts within a county/month/year.[12] To these records, we merge enrollment and plan information at the contract/plan level from the MA enrollment files. We also merge county level MA penetration information to control for the prevalence of MA enrollment. Next, we merge quality information at the contract/year level, which includes an overall summary star measure, star ratings for different domains of quality (e.g., helping you stay healthy), as well as star ratings and continuous summary scores for each individual metric (e.g., percentage of women receiving breast cancer screening and an associated star rating). As discussed above and in Appendix B, data are not available for the overall continuous summary score (i.e., the score rounded to generate an overall star rating), but we are able to replicate this variable by aggregating the specific quality measures as CMS does. Finally, we merge plan premium information at the contract/plan/county/year level, and county level census demographic and socioeconomic information from the American Community Survey.

There exists very little variation in contract/plan enrollment across months in the same calendar year. We therefore take the average enrollment over a year to construct market share variables rather than arbitrarily choosing a given month. After constructing the relevant market share variables from mean enrollment, we drop all observations with missing quality information or with

[12]As our base, we use the Service Area files because the CMS enrollment files include those that move and keep their MA coverage despite the fact that a particular MA contract may not be approved in the new market area, and thus, not part of an potential enrollee's choice set. Data are available for download at www.cms.gov. See Appendix C for a detailed discussion of our dataset and specific links.

overall star ratings which could not be replicated (6 contracts in 2009 and 7 contracts in 2010).[13] We also drop all contract/plans with missing premium information. Dropping these observations does not change our main findings relative to assuming a zero premium. Finally, we focus on HMO, PPO, and PFFS plans, and we drop all special needs and employer specific plans. For 2009 and 2010, we have 20,374 and 17,130 contract/plan/county observations, respectively.[14]

Table 1 provides summary statistics by year for the final datasets used in the analysis. The mean contract/plan enrollment is 292 enrollees in 2009, and the mean overall market share is 1% of the eligible market. 42% of plans have a zero premium, and 72% offer some form of Part D drug coverage in 2009. The quality distribution in 2009 is roughly mound-shaped with three stars as the modal contract quality star value.

Table 1

The MA market appears to consolidate between 2009 and 2010. While the share of MA enrollees relative to Medicare eligible individuals increases from 16.33% to 17.69%, the mean number of contract/plans in a county decreases from 7.12 to 6.26. Meanwhile, the average enrollment increases by 22.8% to 359 enrollees per contract/plan. While no 1.5-star contracts from 2009 remain in our 2010 sample, the quality distribution does not appreciably shift between 2009 and 2010.[15]

[13]Many contracts were too new to be evaluated on all quality dimensions, and thus were omitted from Medicare's quality reporting. Of the ratings that were accurately replicated, we estimate contracts with summary score values marginally above or below the relevant thresholds. Our summary score calculations are therefore not systematically biased in one direction or the other, and it is unlikely that our results are biased by the handful of contracts for which star ratings were not replicated.

[14]See Appendix C for further details on our sample construction.

[15]We do not consider a panel of contract/plans precisely because of institutional changes in how quality is measured and reported. Our RD design also does not require panel data for identification of the effects of interest.

6 Results

First, we estimate standard OLS and 2SLS regressions based on the full sample of plans, with results summarized in Table 2. Note that the star rating coefficients in Table 2 are interpreted relative to all plans with less than a 2.5 rating. For the 2SLS results in 2009, individual tests of each star rating coefficient relative to the next lowest rating reveals no significant effects from 2.5 to 3 stars or from 3 to 3.5 stars, with a significant effect of 0.61 (1.32 percentage point increase in relative market share) when moving from 3.5 to 4 stars.[16] For the 2SLS results in 2010, the only significant effects emerged when moving from 2 to 2.5 stars or from 2.5 to 3 stars. Importantly, the estimates of the quality coefficient from these regressions represent the impact of overall plan quality (as measured by the star ratings) and not specifically the impact of CMS-reported quality from the CMS star rating program.

Table 2

Table 3 presents the estimated star rating coefficients from 2009 data using local linear regressions.[17] The results show a positive and significant effect of star ratings at the 2.5, 3, and 4-star thresholds. The largest effect comes from a 3-star rating relative to a 2.5-star rating (2.75 threshold) which, based on the local linear regression, suggests an increase in a plan's market share relative to Medicare FFS of 4.75 percentage points. The effect drops to a 2.74 percentage point increase in relative market share for 4-star contracts relative to 3.5-star contracts. By contrast, the two-stage least squares effects in Table 2 suggest no significant effects for 3 relative to 2.5 star contracts and a 1.32 percentage point ($p<0.01$) increase for 4 relative to 3.5 star contracts. These results indicate that the effects vary across the quality distribution, but also that there may exist a

[16] Conversion from coefficients to estimated effects on relative market shares follows from Kennedy (1981): $\hat{g} = \left(e^{\hat{\gamma}_2 - 0.5\hat{V}_{\gamma_2}} - 1\right)$, where $\hat{V}_{\gamma_2}$ denotes the estimated variance of the γ_2 coefficient.

[17] Our preferred specification is linear in R_j for small sample concerns. Higher order polynomial specifications yield large standard errors and are subject to overfitting problems.

negative correlation between CMS star ratings and enrollees' prior beliefs in 2009, particularly at the lower end of the star rating distribution. Additionally, our estimate of the nesting parameter, σ, on the within-group log market share ranges between 0.17 and 0.36, suggesting a small but statistically significant correlation in mean utility within groups. The magnitude of the estimated nesting parameter σ is similar to that of the full-sample 2SLS results in Table 2.

Table 3

Results for an identical analysis based on 2010 MA share data are presented in Table 4. Counter to the results in 2009, our 2010 analysis reveals no significant effects of CMS-reported star ratings on relative market shares. We discuss these differences in more detail in Section 6.2; however, before we formally compare the 2009 and 2010 results, we first assess the robustness and sensitivity of the 2009 and 2010 results individually.

Table 4

6.1 Robustness and Sensitivity Analysis

Several empirical questions must also be addressed regarding the appropriateness of the analysis and robustness of the results above. These questions can be generally grouped into three categories: 1) the ability of contracts to manipulate the continuous quality score; 2) bandwidth selection; and 3) the definition of nests used in the discrete choice framework. We discuss each of these areas in more detail below.

First, MA contracts may behave strategically by improving their summary scores just beyond the relevant threshold values. If this were the case, we would expect to see large jumps in the empirical CDF of the summary scores at the threshold values of 2.25, 2.75, 3.25, etc. Figure 1 presents the empirical CDFs for the 2009 and 2010 summary scores, with different CDFs for each nest. As is clear from the figure, there are no apparent jumps in the CDF at the star rating thresholds, indicating no evidence of ``gaming" on behalf of the payers. Differences in CDFs between plans with and without monthly premiums also tends to support our use of positive and $0 premium plans as separate nests.

Figure 1

To more formally assess potential manipulation of the running variable, we adopt the density test proposed by McCrary (2008). This is a test of the continuity of the running variable at the threshold values. As McCrary (2008) points out, this test is most appropriate in situations where manipulation of the running variable is monotonic. This is a reasonable assumption in our case since (if sufficient manipulation is possible) contracts would only have an incentive to increase their continuous quality score. The estimated discontinuity and standard errors at each threshold point are summarized in Table 5, and presented graphically in Figure 2. As is clear from the estimated size of each discontinuity relative to its standard error, there is no evidence that plans can sufficiently manipulate the continuous quality score.

Table 5 and Figure 2

The institutional details behind the star rating program would also tend to refute the claim that contracts manipulate their quality score. This is because the star ratings are calculated based on data two or three years prior to the current enrollment period. The underlying data for the star ratings were therefore predetermined as of the introduction of the star rating program in the respective enrollment year. Similarly, due to the timing of CMS quality reporting and the open enrollment period, contracts would not have the opportunity to manipulate other observable plan characteristics in response to their same-year star ratings. Our application is therefore devoid of many of the supply-side issues raised by Werner *et al.* (2012) in their study of nursing home care, wherein the authors noted that suppliers may respond to low quality scores by changing other product characteristics. One exception to this issue concerns potential adjustments from 2009 to 2010, which we discuss in more detail in Section 6.2

As an additional investigation of firms' ability to manipulate the summary score, we assessed the extent to which threshold star ratings are significant predictors of observed plan characteristics included in our estimation. To do so, we estimated separate RD regressions with plan characteristics as the dependent variable. Insignificant star rating coefficients would tend to support the argument that star ratings are randomly assigned within a bandwidth. Since star ratings are assigned at the contract level, we aggregated plan characteristics to the contract level by taking the mean values across plans within a contract. Results are summarized in Table 6 and consistently show no significant differences in characteristics below versus above the respective star rating thresholds.[18]

Table 6

The choice of bandwidth is another common area of concern in the RD literature. To assess the sensitivity of our results to the choice of bandwidth, we replicated the local linear regression approach from Tables 3 and 4 for alternative bandwidths ranging from 0.095 to 0.50 in increments of 0.005. The results are illustrated in Figure 3, where each graph presents the estimated star-rating coefficient, $\hat{\gamma}_2$, along with the upper and lower 95% confidence bounds. In 2009, for increasingly large bandwidths, the effect at 2.5 and 3.0 stars diminishes in magnitude and becomes insignificant, while the estimated effect at 3.5 stars is persistently insignificant, and the effect at 4.0 stars is remains significant with a magnitude of between 1 and 1.5. Within the narrower bandwidths (less than 0.15), the 2009 panel is consistent with the results from Table 3 in that the effect of star ratings is highest among the lower threshold values. As the bandwidth enlarges, the role of existing quality

[18]We also follow Dafny & Dranove (2008) and Town & Liu (2003) and consider the minimum premium across plans operating under a given contract, with no qualitative change in results from Table 6.

information (as opposed to CMS-reported quality) would be expected to have a larger influence. As such, the results largely collapse to those in Table 2 at larger bandwidths, which explains the persistent effects at the 4.0 star threshold relative to the 3.5 star threshold. Similarly, the effects in the 2010 panel are consistently insignificant and tend to support the robustness of the results in Table 4 with regard to bandwidth selection.

Figure 3

Finally, our delineation of nests according to premium and Part D participation is generally supported from Figure 1, which reveals relatively clear differences in the distribution of star ratings across these four groups. However, previous work from Town & Liu (2003) and others tends to delineate plans by MA-PD versus MA-Only, with no hardened differentiation based on premiums. We therefore replicated our analysis using Part D participation as our sole nesting variable. Results from these regressions are summarized in Tables 7 and 8 for 2009 and 2010, respectively, where the estimated coefficients on the star rating are similar to the results based on our original nests in Tables 3 and 4.

Tables 7 and 8

6.2 Comparison of 2009 and 2010

The OLS and 2SLS results in Table 2 suggest that overall contract quality plays an important role in plan choice in both 2009 and 2010. However, the results from the RD design differ substantially in 2010 compared to 2009, with no significant enrollment effects from CMS-reported star ratings estimated in 2010. As indicated by the robustness and sensitivity analysis in Section 6.1, these differences in the 2009 and 2010 RD results do not appear to be driven by our empirical specification. Moreover, the institutional features of the star rating program do not generally allow for manipulation of the contract star rating in the short term because individual quality metrics are collected over a two-year period prior to the year of aggregation and dissemination. Still, a closer inspection of the 2009 and 2010 star ratings reveals two potential supply-side responses that may explain the observed differences in results.

Although a given insurer's ability to actively change their star rating from 2009 to 2010 is limited, a contract may have responded to the 2009 ratings by changing its plan offerings for 2010. For example, a contract may operate a variety of plans in 2009, but upon receiving a low star rating, the contract may reduce the number of plans under that contract in 2010.[19] There is empirical evidence in our data suggesting that this is the case. As summarized in Table 9, 98% of the plans in contracts with a 1.5 star rating in 2009 were dropped in 2010. Similarly, 50% of the plans in a 2-star contract in 2009 were dropped in 2010. Meanwhile, less than 20% of plans in a contract rated 3 stars or above were dropped, and only 3% of the plans from a 4.5-rated contract were dropped.

Table 9

To assess the impact of star ratings on plan exit more formally, we estimate a probit regression model of an indicator for plan exit in 2010 as a function of dummies for the 2009 star rating and the same county and plan level characteristics included in all prior regressions.[20] Probit coefficents are reported in the second to last column of Table 9, and the associated marginal effects suggest that plans from a contract with a 3-star rating or higher were between 31% and 45% less likely to be dropped in 2010 than plans with a rating of less than 2.5 stars.

In addition to changing the composition of the plans offered, a contract may have responded to the 2009 publication of quality stars by increasing (decreasing) plan premiums if the contract received a high (low) star rating. If consumers are more sensitive to premium changes than to reported quality changes, then we may observe outflows of enrollees from higher rated plans between

[19]Especially if the contract can identify a specific plan for which ratings are likely to be low.
[20]The sample includes all 20,374 plan/contract/county observations in 2009.

2009 and 2010. To investigate this possibility, we estimate a regression of 2010 plan premium as a function of its 2009 value, 2009 star ratings, and 2009 county/plan characteristics. The results are summarized in the final column of Table 9 and indicate that 2.5, 3, and 3.5-star plans had a significant increase in premium from 2009 to 2010.[21] Given the existing literature concerning the role of premium on plan choice (see Town & Liu (2003) and others), the results suggest insurers may have priced themselves out of any quality advantages realized from their 2009 star ratings.

Table 10 reports results from the same two dependent variables -- plan exit and 2010 plan premium, but within a bandwidth of 0.125 around the 2009 star thresholds. While underpowered, the results in Table 10 suggest that contract/plan observations just above the star thresholds for 2.5 and 3.5 stars in 2009 may have substantially increased premiums (43% and 27% of a 2010 premium standard deviation, respectively), and contract/plan observations just below some thresholds may have disproportionately exited the market. A full analysis of the causal impact of the 2009 rating on plan exit and adjustments to plan characteristics is beyond the scope of the paper, but the results in Tables 9 and 10 are suggestive of potential supply-side adjustments to the 2009 star ratings that may explain the lack of any significant, observed effects of quality stars in 2010. However, we stress that the results of this analysis are suggestive at best, and we cannot confidently identify either response given the market-level nature of our data and the short two-year time frame. Ultimately, it could be that the changes to the 2009 to 2010 rating system, in which some domains changed or were removed/replaced altogether, sufficiently altered the rating system such that comparisons across years are not meaningful.

Table 10

[21]The sample includes all plans with a positive premium in both 2009 and 2010, $n = 7,483$.

7 Discussion

This paper analyzes the impact of the CMS star rating program on MA enrollments, focusing particularly on the impact of CMS-reported star ratings rather than the more general impact of plan quality. The analysis adopts a standard discrete-choice framework, estimated with market-level enrollment data as initially proposed in Berry (1994) and subsequently widely implemented in the literature on demand estimation. Our analysis also exploits the inherent sharp RD design of the star rating program, wherein the reported star rating calculated by CMS derives from a continuous summary score rounded to the nearest half-star value.

There are two important points that support our adoption of the RD design based on the overall star rating. First, the *overall* star rating in 2009 and 2010 is that which is easily observed by beneficiaries. Although enrollees can access the domain-level star ratings with some additional effort, a beneficiary does not generally have easy access to the individual metrics such as generosity of breast cancer screening. Second, our design only requires that plans within a bandwidth have comparable *overall* quality -- not that the plans are identical in every dimension. In particular, our design allows for two plans to have similar or even identical summary scores but very different individual metrics. If a plan's overall rating as reported by CMS is important for plan choice, our RD design will appropriately estimate the effect of a marginal improvement in star rating by comparing plans with comparable overall quality scores but plausibly randomly assigned star ratings. To the extent that beneficiaries rely more on domain-level ratings and not on the overall rating, our RD design is still sufficient to conclude that the star rating program had a small effect.

Our results indicate that the 2009 published star ratings directed beneficiaries away from low rated plans more than actively toward high rated plans, suggesting that low star plans may have a relatively larger incentive to increase *CMS-reported* quality compared to higher rated plans. Meanwhile, when we extended the analysis to the 2010 published star ratings, we found no significant effects at any of the star rating thresholds. A closer inspection of plan behaviors from 2009 to 2010 suggests that lower rated plans may have disproportionately exited the market from 2009 to

2010, while marginally higher rated plans in 2009 increased premiums in 2010. These supply-side responses may explain at least some portion of the differences between our 2009 and 2010 results.

Our results have implications within the CMS rating program and potentially broader implications for future rating programs. Within the current program, our results indicate that the rating program has not meaningfully driven beneficiaries toward the highest rated plans. Recent reforms that have partially tied MA contract reimbursement and bonuses to CMS-assessed quality provide clear incentives for contracts to improve quality, but our results suggest that any additional incentives generated by enrollment effects are relatively minimal. Considering the design of future rating programs (e.g., rating plans on the insurance exchanges), our results also highlight potential unintended consequences via supply-side responses from one year to the next. Given the increasing prevalence of rating programs in healthcare, more precisely measuring the supply-side response to published rating programs is an important area of future research.

References

Akerlof, George A. 1970. The Market for Lemons: Quality Uncertainty and the Market Mechanism. *Quarterly Journal of Economics*, **84**(3), 488--500.

Atherly, A. 2001. Supplemental insurance: Medicare's accidental stepchild. *Medical Care Research and Review*, **58**(2), 131--161.

Beaulieu, N.D. 2002. Quality information and consumer health plan choices. *Journal of Health Economics*, **21**(1), 43--63.

Berry, Steven T. 1994. Estimating discrete-choice models of product differentiation. *The RAND Journal of Economics*, 242--262.

Cardell, N.S. 1997. Variance components structures for the extreme-value and logistic distributions with application to models of heterogeneity. *Econometric Theory*, **13**(02), 185--213.

Chernew, M., Scanlon, D.P., *et al.* 1998. Health plan report cards and insurance choice. *Inquiry: a journal of medical care organization, provision and financing*, **35**(1), 9.

Dafny, L., & Dranove, D. 2008. Do report cards tell consumers anything they don't already know? The case of Medicare HMOs. *The Rand journal of economics*, **39**(3), 790--821.

Dranove, David, & Jin, Ginger Zhe. 2010. Quality Disclosure and Certification: Theory and Practice. *Journal of Economic Literature*, **48**(4), 935--963.

Dranove, David, & Sfekas, Andrew. 2008. Start spreading the news: a structural estimate of the effects of New York hospital report cards. *Journal of Health Economics*, **27**(5), 1201--1207.

Farley, D.O., Short, P.F., Elliott, M.N., Kanouse, D.E., Brown, J.A., & Hays, R.D. 2002. Effects of CAHPS health plan performance information on plan choices by New Jersey Medicaid beneficiaries. *Health Services Research*, **37**(4), 985--1007.

Frank, R.G. 2004. *Behavioral economics and health economics*. Tech. rept. National Bureau of Economic Research.

Heiss, F., McFadden, D., & Winter, J. 2006. Who failed to enroll in Medicare Part D, and why? Early results. *Health Affairs*, **25**(5), w344--w354.

Hibbard, J.H., Jewett, J.J., Engelmann, S., & Tusler, M. 1998. Can Medicare beneficiaries make informed choices? *Health Affairs*, **17**(6), 181--193.

Imbens, G.W., & Lemieux, T. 2008. Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, **142**(2), 615--635.

Jin, G.Z., & Sorensen, A.T. 2006. Information and consumer choice: the value of publicized health plan ratings. *Journal of Health Economics*, **25**(2), 248--275.

Kennedy, Peter E. 1981. Estimation with Correctly Interpreted Dummy Variables in Semilogarithmic Equations. *American Economic Review*, **71**(4).

Lee, David S, & Lemieux, Thomas. 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature*, **48**, 281--355.

Li, Q., & Trivedi, P.K. 2012. Medicare Health Plan Choices of the Elderly: A Choice-With-Screening Model. *Journal of Business and Economic Statistics*, 1--13.

McCrary, Justin. 2008. Manipulation of the running variable in the regression discontinuity design: A density test. *Journal of Econometrics*, **142**(2), 698--714.

Ramanarayanan, Subramaniam, & Snyder, Jason. 2012. Information Disclosure and Firm Performance: Evidence from the Dialysis Industry. *Available at SSRN 2152796*.

Reid, Rachel O, Deb, Partha, Howell, Benjamin L, & Shrank, William H. 2013. Association Between Medicare Advantage Plan Star Ratings and EnrollmentStar Ratings for Medicare Advantage Plan. *JAMA*, **309**(3), 267--274.

Scanlon, D.P., Chernew, M., McLaughlin, C., & Solon, G. 2002. The impact of health plan report cards on managed care enrollment. *Journal of Health Economics*, **21**(1), 19--41.

Schram, A., & Sonnemans, J. 2011. How individuals choose health insurance: An experimental analysis. *European Economic Review*.

Spranca, M., Kanouse, D.E., Elliott, M., Short, P.F., Farley, D.O., & Hays, R.D. 2000. Do consumer reports of health plan quality affect health plan selection? *Health Services Research*, **35**(5 Pt 1), 933.

Thistlethwaite, D.L., & Campbell, D.T. 1960. Regression-discontinuity analysis: An alternative to the ex post facto experiment. *Journal of Educational Psychology*, **51**(6), 309.

Town, Robert, & Liu, Su. 2003. The welfare impact of Medicare HMOs. *RAND Journal of Economics*, 719--736.

Wedig, G.J., & Tai-Seale, M. 2002. The effect of report cards on consumer choice in the health insurance market. *Journal of Health Economics*, **21**(6), 1031--1048.

Werner, Rachel M., Norton, Edward C., Konetzka, R. Tamara, & Polsky, Daniel. 2012. Do consumers respond to publicly reported quality information? Evidence from nursing homes. *Journal of Health Economics*, **31**(1), 50--61.

Tables and Figures

Table 1: **Summary Statistics**

	2009 Mean (S.D.)	2010 Mean (S.D.)
Overall Quality Stars %. n=246 and 289		
1.5	2.03	
2.0	10.16	4.15
2.5	24.39	24.57
3.0	30.08	30.80
3.5	19.92	21.11
4.0	10.98	11.76
4.5	2.44	6.57
5.0		1.04
Contract/Plan/County. n=20,374 and 17,130		
Enrollment	291.96 (1,422.73)	358.52 (1,584.49)
Overall Market Share, %	1.01 (1.97)	1.26 (2.14)
Share within MA, %	6.94 (10.29)	6.58 (9.08)
Premium =$0, %	42.05	37.38
Monthly Premium, $/Month \| Prem.$>$ $0	35.97 (42.96)	43.63 (50.64)
Drug Coverage, %	72.86	72.01
HMO, %	25.99	35.34
PPO, %	22.81	37.52
County Characteristics. n=2,861 and 2,735		
MA Penetration	16.33 (11.00)	17.69 (12.12)
Mean Number of Plans	7.12 (5.76)	6.26 (5.41)
Population >65 in 1,000s	13.30 (36.36)	14.10 (37.75)
Population >85 in 1,000s	1.86 (5.33)	2.00 (6.65)
% Unemployed	5.90 (2.02)	9.38 (3.08)
% White	86.36 (15.35)	86.18 (15.20)
% Black	9.48 (14.46)	9.74 (14.51)
% Female	50.29 (1.99)	50.28 (2.07)
% College Grad.	18.75 (8.58)	18.81 (8.68)
% South	43.27 (49.55)	44.68 (49.73)

Table 2: **OLS and 2SLS Regression Results for MA Shares**[a]

	2009 MA Shares		2010 MA Shares	
	OLS	**2SLS**	**OLS**	**2SLS**
Star Value				
2.5	0.10	0.26*	0.43***	0.44***
	(0.11)	(0.15)	(0.16)	(0.10)
3.0	0.22	0.46**	0.60***	0.72***
	(0.15)	(0.19)	(0.16)	(0.11)
3.5	0.56***	0.74***	0.79***	0.90***
	(0.19)	(0.24)	(0.17)	(0.14)
4.0	1.07***	1.35***	0.96***	1.02***
	(0.24)	(0.31)	(0.22)	(0.21)
4.5	0.54*	0.64*	1.08***	1.18***
	(0.31)	(0.37)	(0.21)	(0.22)
5.0	-	-	0.59	0.87**
	-	-	(0.44)	(0.37)
Plan Characteristics				
HMO	0.12	-0.01	-0.04	-0.11
	(0.13)	(0.17)	(0.11)	(0.15)
PPO	-0.17	-0.26	-0.26***	-0.28***
	(0.13)	(0.17)	(0.08)	(0.11)
Premium = \$ 0	-0.32*	-0.27	-0.11	0.06
	(0.19)	(0.26)	(0.08)	(0.10)
Premium	-0.00	-0.01	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Offers Part D	0.99***	0.85***	1.53***	1.28***
	(0.20)	(0.19)	(0.12)	(0.12)
Part D Deductible	-0.00	-0.00	-0.00***	-0.00***
	(0.00)	(0.00)	(0.00)	(0.00)
Nest Share	0.64***	0.23***	0.66***	0.35***
	(0.03)	(0.04)	(0.03)	(0.04)
County Characteristics				
Over 65	-0.08***	-0.07***	-0.05**	-0.04
	(0.02)	(0.03)	(0.02)	(0.02)
Over 85	0.52***	0.31**	0.30**	0.15
	(0.10)	(0.15)	(0.14)	(0.15)
Unemployment Rate	-0.01	-0.04***	0.01	-0.00
	(0.01)	(0.01)	(0.01)	(0.01)
% White	0.00***	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
% Black	0.00	0.00	-0.01**	-0.01**
	(0.00)	(0.00)	(0.00)	(0.00)
% Female	-0.02***	-0.05***	-0.02***	-0.04***
	(0.01)	(0.01)	(0.01)	(0.01)
% College Graduate	-0.02***	-0.03***	-0.01***	-0.02***
	(0.00)	(0.00)	(0.00)	(0.00)
South Region	-0.10	-0.11	-0.17***	-0.20***
	(0.07)	(0.08)	(0.06)	(0.07)
Constant	-3.42***	-1.95***	-3.83***	-2.79***
	(0.38)	(0.40)	(0.53)	(0.55)

[a]Results based on OLS and 2SLS regressions estimated on the full sample of plans, with standard errors in parentheses and robust to heterogeneities across plans and clustering at the contract level. Omitted star category is less than 2.5 stars. For 2SLS results, within-group shares were instrumented with the age of the contract and the number of plans in the same nest. * p<0.1. ** p<0.05. *** p<0.01.

Table 3: **RD Regression Results for 2009 MA Shares**[a]

Ind. Var.	2.25	2.75	3.25	3.75
Star Value				
Star	1.18**	1.77**	0.46	1.08**
	(0.46)	(0.90)	(0.47)	(0.43)
Plan Characteristics				
HMO	0.96**	0.33	0.73	0.59**
	(0.43)	(0.49)	(0.58)	(0.27)
PPO	1.15***	-0.31	-	-
	(0.41)	(0.31)	-	-
Premium = $ 0	0.16	-0.39	0.04	-0.73**
	(0.16)	(0.40)	(0.40)	(0.31)
Premium	0.00	-0.00	0.00	0.00
	(0.00)	(0.01)	(0.00)	(0.00)
Offers Part D	0.23	1.61***	1.54***	0.68***
	(0.14)	(0.25)	(0.41)	(0.26)
Part D Deductible	0.00	-0.01***	0.00	0.00**
	(0.00)	(0.00)	(0.00)	(0.00)
County Characteristics				
Over 65	-0.06	-0.18**	-0.11	0.05
	(0.05)	(0.08)	(0.13)	(0.09)
Over 85	-0.04	0.83**	0.64	-0.22
	(0.25)	(0.41)	(0.84)	(0.59)
Unemployment Rate	-0.01	-0.04	-0.03	-0.10
	(0.02)	(0.02)	(0.12)	(0.07)
% White	0.01	-0.02***	0.02	-0.04**
	(0.00)	(0.01)	(0.02)	(0.02)
% Black	0.01***	-0.02***	0.02	-0.06***
	(0.00)	(0.01)	(0.02)	(0.02)
% Female	-0.05***	-0.07***	-0.09	-0.02
	(0.01)	(0.01)	(0.06)	(0.04)
% College Graduate	-0.03***	-0.03***	-0.00	-0.02**
	(0.00)	(0.01)	(0.01)	(0.01)
South Region	-0.40***	-0.18	-1.34***	0.34
	(0.12)	(0.15)	(0.51)	(0.31)
Nest Share	0.17**	0.18*	0.21	0.36**
	(0.08)	(0.09)	(0.18)	(0.18)
Constant	-4.09***	-1.37	-3.47	0.86
	(0.68)	(1.14)	(3.05)	(2.80)
Contracts	22	33	20	27
Observations	4,315	2,063	701	864

[a]Results based on 2SLS with linear RD approach and a bandwidth of $h = 0.125$. Standard errors in parentheses are robust to heterogeneities across plans and clustering at the contract levels. Results were excluded for the 1.5 and 4.5 star ratings due to an insufficient number of contracts on the lower and upper ends of the 1.75 and 4.25 thresholds, respectively. Within-group shares were instrumented with the age of the contract and the number of plans in the same nest. * p<0.1. ** p<0.05. *** p<0.01.

Table 4: **RD Regression Results for 2010 MA Shares**[a]

Ind. Var.	2.25	2.75	3.25	3.75
Star Value				
Star	-0.55	0.19	0.70	0.11
	(0.69)	(0.42)	(0.46)	(0.35)
Plan Characteristics				
HMO	-0.41	0.35	0.82	1.94***
	(0.56)	(0.30)	(0.52)	(0.45)
PPO	-	0.14	0.25	1.39***
	-	(0.24)	(0.44)	(0.44)
Premium = 0	-1.04	-0.20*	0.35	0.17
	(0.65)	(0.12)	(0.23)	(0.20)
Premium	-0.03***	-0.00***	-0.00	-0.00
	(0.01)	(0.00)	(0.00)	(0.00)
Offers Part D	1.68***	0.84***	1.82***	1.40***
	(0.24)	(0.14)	(0.19)	(0.27)
Part D Deductible	-0.00*	-0.00	-0.00	0.00
	(0.00)	(0.00)	(0.00)	(0.00)
County Characteristics				
Over 65	0.09**	-0.14**	-0.03	-0.24*
	(0.04)	(0.06)	(0.04)	(0.13)
Over 85	-0.69**	0.62*	0.07	1.52*
	(0.28)	(0.33)	(0.29)	(0.85)
Unemployment Rate	-0.10*	0.03***	0.08**	-0.09
	(0.06)	(0.01)	(0.03)	(0.06)
% White	-0.01	-0.00	0.01	0.01
	(0.02)	(0.00)	(0.01)	(0.02)
% Black	-0.01	-0.00	0.00	0.01
	(0.02)	(0.00)	(0.01)	(0.02)
% Female	0.02	-0.03***	-0.06*	-0.01
	(0.05)	(0.01)	(0.03)	(0.02)
% College Graduate	-0.05***	-0.02***	0.00	-0.02**
	(0.01)	(0.00)	(0.01)	(0.01)
South Region	0.06	-0.35***	-0.30	0.06
	(0.30)	(0.10)	(0.23)	(0.52)
Nest Share	0.03	0.39***	0.32***	0.13
	(0.09)	(0.06)	(0.07)	(0.17)
Constant	-2.57	-2.48***	-4.29**	-5.91***
	(2.73)	(0.74)	(1.70)	(2.01)
Contracts	22	48	45	35
Observations	436	5,246	1,737	1,497

[a]Results based on 2SLS with linear RD approach and a bandwidth of $h = 0.125$. Standard errors in parentheses are robust to heterogeneities across plans and clustering at the contract levels. Results were excluded for the 1.5 and 4.5 star ratings due to an insufficient number of contracts on the lower and upper ends of the 1.75 and 4.25 thresholds, respectively. Within-group shares were instrumented with the age of the contract and the number of plans in the same nest. * p<0.1. ** p<0.05. *** p<0.01.

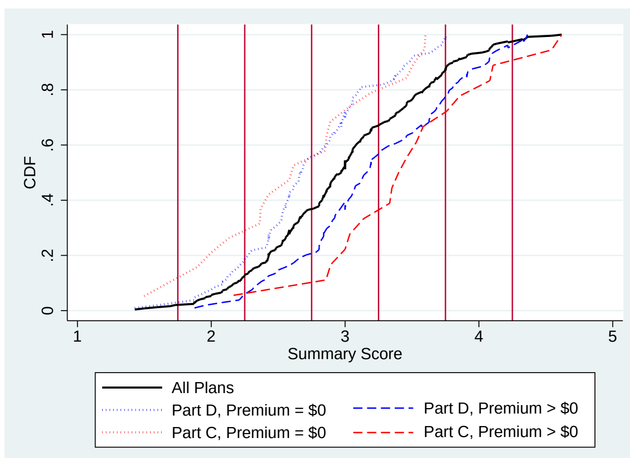
Table 5: **McCrary (2008) Density Test**[a]

	Threshold Values			
	2.25	2.75	3.25	3.75
2009				
Discontinuity	-0.46	0.51	-0.69	0.38
St. Err.	(0.92)	(1.29)	(1.25)	(0.84)
2010				
Discontinuity	1.39	-0.61	0.58	-0.41
St. Err.	(1.89)	(0.61)	(0.80)	(0.89)

[a]Table reports the estimated jump in the density of the continuous quality score at each of star four thresholds for both 2009 and 2010. Standard errors are in parentheses. * p<0.1. ** p<0.05. *** p<0.01.

Figure 1: Empirical CDF of Summary Scores

2009 CDF



2010 CDF

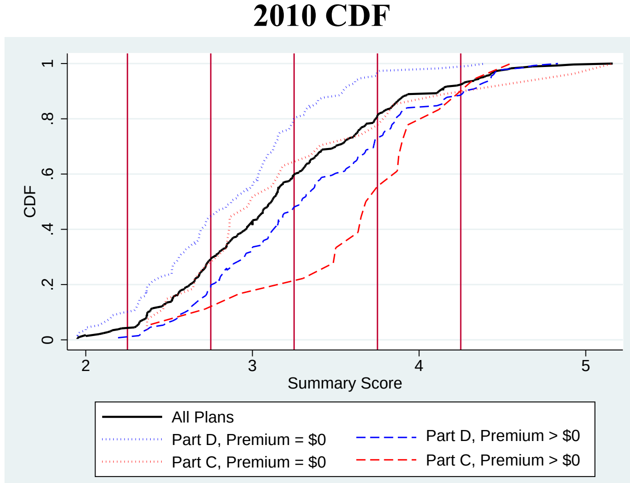


Table 6: **Test of Plan Characteristics within Relevant Bandwidths**[a]

	Threshold Values			
	2.25	2.75	3.25	3.75
	2009 Plan Characteristics			
Premium[b]	-29.84	20.08	6.48	-92.71
$\mid$ Premium > 0	(26.53)	(23.72)	(48.64)	(74.31)
Premium $= 0$	1.46	0.90	-1.73	1.00
	(1.29)	(1.39)	(1.35)	(0.90)
HMO	0.42	-1.02	1.62	-1.70**
	(1.09)	(1.21)	(1.20)	(0.84)
PPO	-0.61	1.27	-1.62	1.70**
	(1.11)	(1.22)	(1.20)	(0.84)
Part D Deductible	-	-0.44	-0.10	-0.84
	-	(1.30)	(1.28)	(0.96)
	2010 Plan Characteristics			
Premium	-14.17	-8.15	-9.44	44.91
$\mid$ Premium > 0	(20.94)	(23.81)	(29.00)	(54.97)
Premium $= 0$	-	-0.23	-0.90	0.12
	-	(0.74)	(0.90)	(0.78)
HMO	-	0.39	-1.37	0.39
	-	(0.76)	(0.89)	(0.82)
PPO	-	-0.17	0.78	-0.59
	-	(0.77)	(0.91)	(0.84)
Part D Deductible	-	-0.24	0.86	0.81
	-	(0.82)	(0.95)	(0.84)

[a] Results based on OLS regressions with RD approach and a bandwidth of $h = 0.125$. Robust standard errors in parenthesis. Results were excluded for the 1.5 and 4.5 star ratings due to an insufficient number of contracts on the lower and upper ends of the 1.75 and 4.25 thresholds, respectively. Regressions estimated at the contract level, with dependent variables measured as the average value of each plan characteristic by contract. Part D deductible measured only among plans offering Part D, and similarly the plan premium is measured only among plans charging a positive premium. * p<0.1. ** p<0.05.

[b] The magnitude of the premium coefficient at the 3.75 threshold is driven largely by a single contract with a high premium and a summary score of 3.74. Removing this contract reduces the coefficient to -9.65 and a standard error of 21.61.

Table 7: **RD Regression Results for 2009 MA Shares with Alternative Nests**[a]

	Threshold Values			
	2.25	2.75	3.25	3.75
$\hat{\gamma}_2$	0.99**	1.50**	0.48	0.87*
S.E.	(0.42)	(0.72)	(0.46)	(0.48)
$\hat{\sigma}$	0.25**	0.30**	0.18	0.37
S.E.	(0.07)	(0.08)	0.22	(0.23)
Contracts	22	33	20	27
Observations	4,315	2,063	701	864

[a]Results based on 2SLS regressions with RD approach and a bandwidth of $h = 0.125$. Standard errors in parenthesis are robust to heterogeneities across plans and clustering at the contract level. Results were excluded for the 1.5 and 4.5 star ratings due to an insufficient number of contracts on the lower and upper ends of the 1.75 and 4.25 thresholds, respectively. Nests were delineated by plans with or without prescription drug coverage. Within-group shares were instrumented with the age of the contract and the number of plans in the same nest. * p<0.1. ** p<0.05.

Table 8: **RD Regression Results for 2010 MA Shares with Alternative Nests**[a]

	Threshold Values			
	2.25	2.75	3.25	3.75
$\hat{\gamma}_2$	-0.57	0.30	0.64	0.13
S.E.	(0.74)	(0.36)	(0.40)	(0.36)
$\hat{\sigma}$	0.00	0.38***	0.42***	0.10
S.E.	(0.17)	(0.08)	0.07	(0.12)
Contracts	22	48	45	35
Observations	436	5,246	1,737	1,497

[a]Results based on 2SLS regressions with RD approach and a bandwidth of $h = 0.125$. Standard errors in parenthesis are robust to heterogeneities across plans and clustering at the contract level. Results were excluded for the 1.5 and 4.5 star ratings due to an insufficient number of contracts on the lower and upper ends of the 1.75 and 4.25 thresholds, respectively. Nests were delineated by plans with or without prescription drug coverage. Within-group shares were instrumented with the age of the contract and the number of plans in the same nest. * p<0.1. ** p<0.05. *** p<0.01

Table 9: **Plans from 2009 to 2010 by Star Rating**[a]

2009 Rating	Count	Percent	% Dropped for 2010	2009 Premium| >0 $n = 11,806$	2010 Premium| >0 $n = 7,483$	2010 Plan Exit $n = 20,374$	2010 Premium $n = 7,483$
1.5	1,526	7.49	97.8%	$48.23 (24.78)	37.17 (4.59)	.	.
2.0	4,435	21.77	49.8%	41.92 (29.25)	29.98 (19.86)	.	.
2.5	7,320	35.93	58.1%	57.77 (37.82)	65.10 (28.77)	0.34	26.34***
						(0.44)	(9.36)
3.0	3,905	19.17	20.4%	59.08 (32.99)	81.76 (40.98)	-0.95**	20.60**
						(0.39)	(8.63)
3.5	1,791	8.79	16.2%	77.09 (44.37)	99.89 (64.48)	-0.90***	28.04***
						(0.35)	(10.01)
4.0	947	4.65	20.0%	85.91 (48.06)	97.51 (57.88)	-0.95***	15.72
						(0.42)	(10.14)
4.5	450	2.21	3.3%	80.55 (42.21)	91.52 (50.48)	-2.16***	16.92
						(0.48)	(10.29)
Total	20,374	100		62.08 (39.36)	73.94 (48.58)		

[a]Data based on all 20,374 contract/county/plan observations in 2009 and their associated star ratings. Of these contract/county/plan combinations, 11,128 remained in 2010, reflective of plans being dropped in some counties or within some contracts as well as contracts being dropped entirely. The second to last column reports results from a probit regression of plan exit in 2010 on the full sample from 2009. The last column reports results from a regression of the 2010 premium conditional upon a positive premium in 2010. Plan exit and 2010 premium are also modeled as a function of 2009 star plan/contract/county characteristics. Clustered standard errors in parenthesis. * p<0.1. ** p<0.05. *** p<0.01.

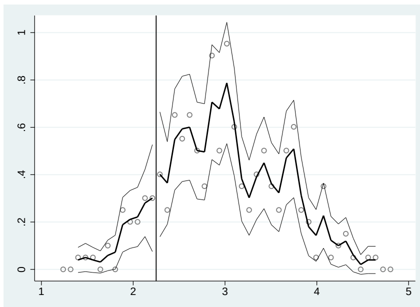
Table 10: **Regression Discontinuity Estimates - 2009 to 2010**[a]

	Threshold Values			
	2.25	2.75	3.25	3.75
Plan Exit $\hat{\gamma}_2$	-0.97***	-0.07	0.34***	-0.35*
S.E.	(0.20)	(0.16)	(0.11)	(0.19)
Contracts	16	31	18	864
Observations	4,315	2,063	701	26
2010 Premium $\hat{\gamma}_2$	67.18***	13.82	40.76***	24.92
S.E.	(20.67)	(27.69)	(13.08)	(20.33)
Contracts	3	9	9	14
Observations	268	1,245	502	512

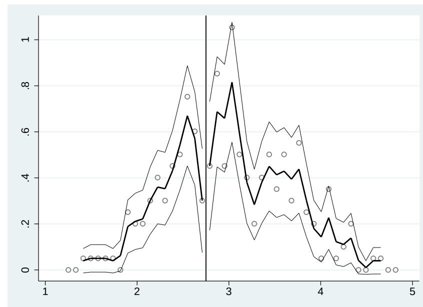
[a]The top panel reports results from a linear probability model, within our preferred 0.125 bandwidth, of plan exit in 2010. The bottom panel reports results of an OLS regression, within the 0.125 bandwidth, of 2010 premium conditional upon a positive premium in both 2009 and 2010. Both models are a function of the relevant 2009 quality star and additional independent variables from Tables 3 and 4. * p<0.1. ** p<0.05. *** p< 0.01.

Figure 2: McCrary (2008) Discontinuity Test at Thresholds 2.25-3.75

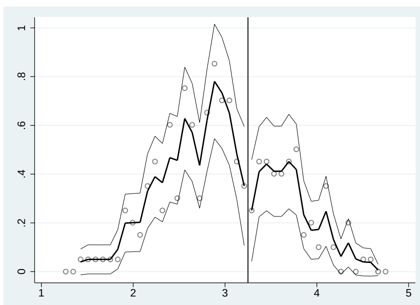
2009 Densities



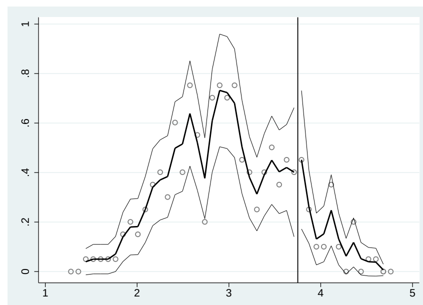
a. 2.25



b. 2.75

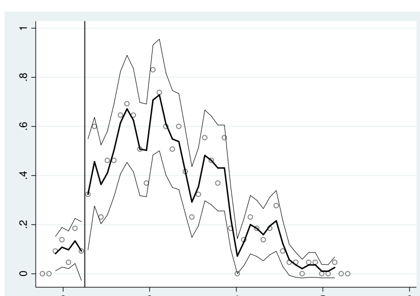


c. 3.25

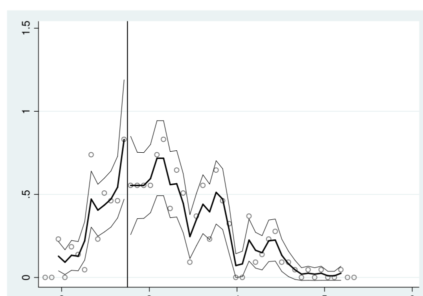


d. 3.75

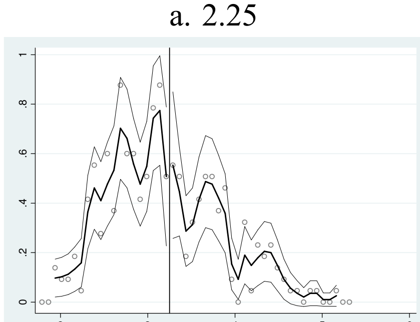
2010 Densities



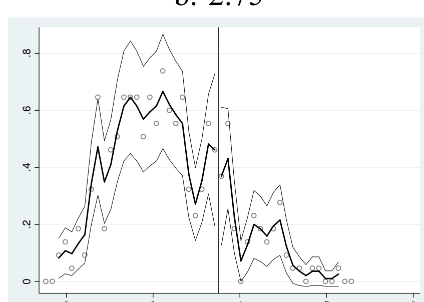
a. 2.25



b. 2.75



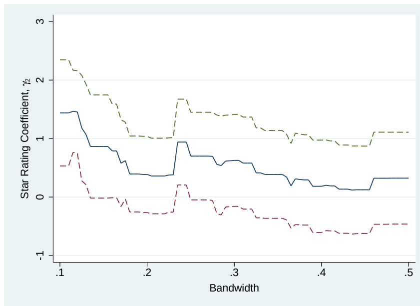
c. 3.25



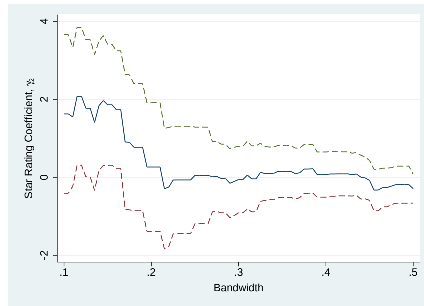
d. 3.75

Figure 3: Star Effect at Varying Bandwidths Around Stars 2.5, 3, 3.5 and 4

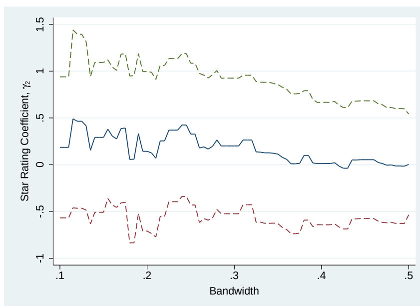
2009 Effects



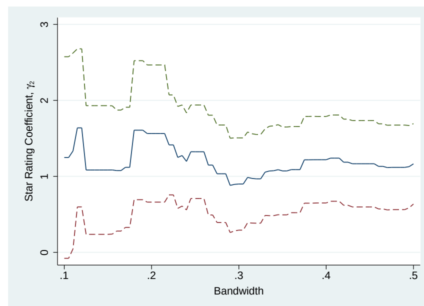
a. 2.5



b. 3.0

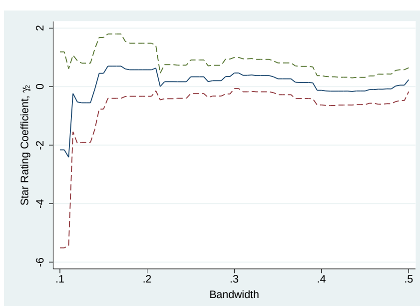


c. 3.5

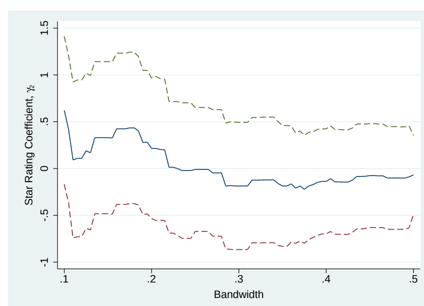


d. 4.0

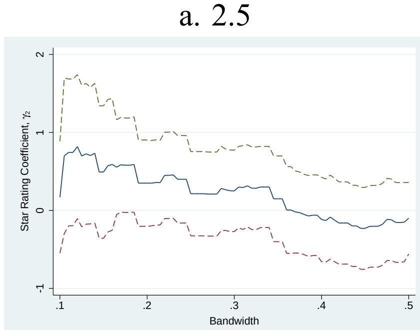
2010 Effects



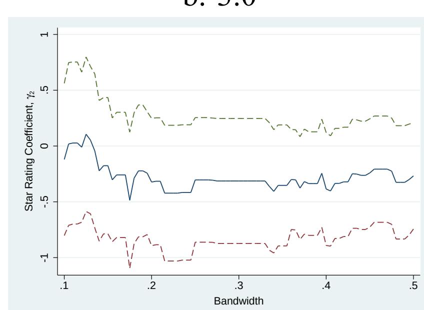
a. 2.5



b. 3.0



c. 3.5



d. 4.0