Universität Trier

Using quantile regression for optimal risk adjustment

Normann Lorenz

Research Papers in Economics
No. 11/14

Using quantile regression for optimal risk adjustment

Normann Lorenz*

May 15, 2014

Abstract

This paper analyzes optimal risk adjustment for direct risk selection (DRS). Integrating insurers activities for risk selection into a discrete choice model of individuals' health insurance choice shows that DRS has the structure of a contest. For the contest success function used in most of the contest literature, optimal transfers for a risk adjustment scheme have to be determined by means of a restricted quantile regression, irrespective of whether insurers primarily engage in positive DRS (attracting low risks) or negative DRS (repelling high risks). This is at odds with the common practice of determining transfers by means of a least squares regression. However, this common practice can be rationalized within a discrete choice model for a new class of contest success functions, but only if positive and negative DRS are equally important; if not, optimal transfers have to be calculated from a restricted asymmetric least squares regression. Using data from a German and a Swiss health insurer, we find considerable differences between the three types of regressions. Optimal transfers therefore critically depend on which contest success function represents insurers' incentives for DRS and whether positive and negative DRS are equally important or not. Results from the two data sets indicate that if a regulator does not know which case applies, transfers should rather be calculated by means of a quantile than a least squares regression.

JEL-**classification:** I 13, I 18, L 13.

Keywords: Risk selection, risk adjustment, discrete choice, contest, quantile regression.

*Universität Trier, Universitätsring 15, 54286 Trier, Germany; e-mail: lorenzn@uni-trier.de; phone: 0049-651-2012624. I thank Friedrich Breyer and Mathias Kifmann for helpful comments and suggestions.

1 Introduction

In many health insurance markets, insurers are not allowed to differentiate their premiums according to individuals' expected cost; instead, they have to charge a uniform premium to all risk types. In such a setting, insurers will make (expected) profits with some individuals, and losses with others. This creates incentives to attract profitable and repel unprofitable individuals. If insurers act on these incentives, they are said to be engaged in risk selection.[1]

Two forms of risk selection can be distinguished: direct and indirect risk selection.[2] With indirect risk selection (IRS), insurers influence who wants to join them by designing their benefit package in a way so that it is attractive for low risks but not for high risks. This can be achieved with, e.g., scrupulous utilization reviews, a physician network with only a small number of specialists or by not covering certain services primarily needed by the high risks. With direct risk selection (DRS), insurers use measures unrelated to the benefit package like selective advertising, offering discounts for fitness club memberships or by 'losing' applications of unprofitable individuals. Since DRS is targeted at a specific risk type, insurers have to know whether a particular individual (or group of individuals) is of below or above average risk. Examples for such groups are certain age brackets or individuals living in high cost areas.

In some settings, like the U.S. Medicare Advantage program, where insurers are allowed to differ in their physician networks or drug formularies, incentives to distort the benefit package can be severe and are at least of similar importance as are incentives for DRS.[3] However, in the European context, where the benefit package is usually heavily regulated, the scope for IRS is rather limited; here, insurers who try to influence the risk structure of their insured will be primarily engaged in DRS.[4]

A regulator can reduce the incentives for both DRS and IRS by implementing a risk adjustment scheme, i.e., by setting transfers insurers receive or have to pay depending on the risk structure of their insured.[5] There are several ways to organize these payments, but effectively, each insurer has to pay a uniform risk adjustment fee for each insured equal to the average cost of all insured in the respective health insurance market, and in return receives an individual specific transfer for each insured (depending on the signals of the insured). In most risk adjustment schemes, these transfers to insurers equal the predicted values of a regression of actual cost on a set of variables like age, gender and morbidity. However, even sophisticated risk adjustment schemes only reduce, but do not eliminate incentives for risk selection.[6]

There is a huge literature on how the regression models used for risk adjustment can be improved so that more accurate cost predictions (and therefore transfers) can be determined.

[1] See van de Ven and Ellis (2000).

[2] See Breyer et al. (2011).

[3] For incentives to distort the benefit package, see Frank et al. (2000), Cao and McGuire (2003) and Ellis and McGuire (2007); for the profitability of DRS in the U.S. setting, see Shen and Ellis (2002).

[4] See, e.g., the Special Issue on 'Risk adjustment in Europe' in *Health Policy* (Chernichovsky and van de Ven 2003).

[5] See Zweifel et al. (2009), chapter 7.

[6] See van Veen et al. (2014).

The criterion to choose among competing models has almost always been the R^2, the explained part of the variance. The larger the R^2, the closer the transfers are to actual cost, and the lower the incentives for risk selection should be.

There is a small literature that deviates from this statistical approach; instead, it explicitly analyzes insurers' incentives to engage in risk selection and determines the optimal transfers as a solution to such an incentive problem. This approach has been termed 'optimal risk adjustment' by Glazer and McGuire (2000). They have shown that a regulator can increase the effectiveness of a risk adjustment scheme by distorting the transfers as calculated with conventional, regression based risk adjustment: if the signals used as explanatory variables in the regulator's regression are less than perfectly correlated with risk type, there has to be overpayment for signals which indicate high risk, and underpayment for signals which indicate low risk.

So far, the optimal risk adjustment literature has been exclusively concerned with IRS, i.e., with insurers' incentives to distort the benefit package.[7] This is suitable for all settings where insurers can influence at least some aspect of the benefit package (as in the Medicare Advantage program in the U.S.), but in the European setting with its heavily regulated benefit packages, DRS is the more severe problem.

This study analyzes optimal risk adjustment for DRS and shows that – as in the case of IRS – a regulator can in general increase the effectiveness of a risk adjustment scheme by deviating from the transfers as calculated from a least squares regression that maximizes the R^2. Integrating insurers' activities to risk select in a discrete choice model of individuals' health insurance choice, we first derive that DRS has the structure of a contest. We show that for the Tullock-contest success function – the contest success function employed in the vast majority of all models in the contest literature – maximizing the R^2 does not minimize insurers' incentives for DRS; rather, the correct objective is to minimize the mean absolute deviation, MAD.[8] This is achieved by using a restricted quantile regression instead of a least squares regression. We show that such a quantile regression is optimal regardless of whether insurers are primarily (or exclusively) engaged in positive DRS (attracting profitable individuals) or in negative DRS (repelling unprofitable individuals).

Since almost all risk adjustment schemes calculate transfers from a least squares regression, we proceed by asking whether this common practice can be rationalized within a discrete choice model, i.e., for a different contest success function (csf) than the Tullock-csf. We find that such a csf exists (although it has not been employed in the contest literature so far). However, contrary to the Tullock-csf, for this csf, the least squares regression is only optimal for the symmetric contest where positive and negative DRS are equally important. In the asymmetric case, where insurers focus on one of the two types of DRS, transfers have to be determined by means of a restricted asymmetric least squares regression.

Which of the three regression models – the least squares, the asymmetric least squares or the quantile regression – should be used by the regulator therefore depends on which of the two contest success functions applies and, if it is the second one, whether both forms

[7] See, e.g., Glazer and McGuire (2002), Jack (2006), Bijlsma et al. (2011) and McGuire et al. (2013).

[8] The Tullock-contest success function has been introduced by Tullock (1980). For an overview of the contest literature see, e.g., Konrad (2009) and Congleton et al. (2008).

of DRS are equally important or not. If there was data on insurers' expenditures for risk selection, this could be easily be determined, but we are not aware that such data exists.[9]

In the empirical part of this paper, using data from a German sickness fund and from a Swiss health insurer, we therefore compare the performance of the three types of regressions and find the following two asymmetries: When the asymmetric least squares regression is optimal, it is only somewhat more successful than the quantile regression, but when it is not optimal, it can perform much worse than the quantile regression; in some settings, it even increases incentives for risk selection. We find the same asymmetry when comparing the least squares and the quantile regression: the least squares regression never performs substantially better, but sometimes considerably worse than the quantile regression. If these asymmetries were also found in other data sets, the quantile regression could be considered a valuable alternative to the least squares regression for calculating the transfers of a risk adjustment scheme.

The remainder of this paper is organized as follows. In Section 2, we derive the contest structure of DRS within a discrete choice model. In Section 3, we show that for the Tullock-csf, a regulator should not maximize the R^2, but minimize the MAD by using a restricted quantile regression. The R^2 criterion is rationalized in Section 4. We illustrate the difference between the least squares and the quantile regression with a simple example in Section 5 and show that these differences can be substantial for real data in Section 6. Section 7 concludes.

2 Direct risk selection

2.1 Profits and losses for insurers with a risk adjustment scheme

For direct risk selection to occur, insurers have to be able to classify individuals into different risk types according to some signals which are informative about expected cost. Denote the set of signals insurers observe by S^H, and let S_i^H be the set of signals for an individual i.[10] Using these signals, insurers' expectation of individual i's cost is given by

$$c_i^H = \mathrm{E}[c|S_i^H]. \tag{1}$$

If the regulator was able to observe the full set of signals, S^H, he could infer insurers' cost predictions c^H, (where c^H is the vector of all cost predictions for the I individuals in the respective health insurance market). Setting the transfers of the risk adjustment scheme, c^R, equal to these cost predictions, insurers' expected profits (or losses) for each individual would be equal to zero and the incentives for risk selection eliminated completely.

There are two reasons why, in general, the transfers c^R will differ from insurers' cost predictions c^H. First, the regulator may not be able to observe the full set of signals used by

[9] E.g., Newhouse (2002, p. 176) notes: "…the data suggest that cost is not a bar to some selection. I have no evidence, however, on the cost of engaging in selection."

[10] For all variables the superscript H will be used for the health insurer and R for the regulator.

the insurers.[11] Secondly, the regulator may not want to use the full set of signals because some of the signals can be influenced by the insurers. One example for such a signal is prior year expenditures. Using this variable increases the explained part of the variance by several percentage points, but is in fact just retrospective partial cost reimbursement, which reduces incentives for cost efficiency.[12]

For these two reasons, the regulator will base his cost predictions (and thus his transfers) only on a subset $S^R \subset S^H$ of the variables used by the insurers. Therefore, transfers will be calculated as

$$c_i^R = h(S_i^R) \tag{2}$$

and differ from the insurers' cost predictions c_i^H. Denote this difference by $D_i = c_i^R - c_i^H$, and the vector of all these differences by D. If $D_i > 0$, insurers expect a profit, if $D_i < 0$, they expect a loss. Insurers will act on these incentives and try to increase the probability of being chosen by individuals with $D_i > 0$ (positive DRS), and reduce the probability of being chosen by those with $D_i < 0$ (negative DRS).

We now show how these activities can be integrated in a discrete choice model of individuals' health insurance choice. We integrate DRS in a model of imperfect competition because DRS seems incompatible with perfect competition, where individuals are perfectly informed about the benefit packages offered and the premiums charged, and always choose the insurer who offers the best benefit package-premium combination.

2.2 The discrete choice model without DRS

There are J insurers j, each offering a benefit package-premium combination which, for individual i, yields utility V_i^j. Individual i's decision of which insurer to choose, however, not only depends on which insurer offers the highest utility V_i^j, but also on some other factors like location or which insurer was recommended by family and friends. In a discrete choice model, the influence of these other factors is captured by an individual- and insurer-specific utility component ε_i^j. Without DRS, individual i's utility when choosing an insurer j is therefore given by

$$u_i^j = V_i^j + \varepsilon_i^j. \tag{3}$$

Assuming ε_i^j to be i.i.d. extreme value, the logit model arises. Specifying the variance of ε_i^j as $Var(\varepsilon_i^j) = \sigma^2 \frac{\pi^2}{6}$, the probability of individual i choosing a particular insurer k is given by[13]

$$Prob(i \text{ chooses } k) = Prob(V_i^k + \varepsilon_i^k > V_i^l + \varepsilon_i^l \; \forall \, l \neq k) = \frac{e^{\frac{V_i^k}{\sigma}}}{\sum_{j=1}^{J} e^{\frac{V_i^j}{\sigma}}}. \tag{4}$$

[11] E.g., in Germany, the regulator does not know the zip code of the insured, a variable readily observable for insurers who can use this information to identify high and low cost areas.

[12] See van de Ven and Ellis (2000, p. 805). See also Schokkaert and van de Voorde (2004); their examples of variables insurers can and should influence (and which should therefore not be used for risk adjustment) are the lifestyle of the insured or their propensity to go immediately to a specialist; accordingly, they denote these variables as responsibility variables.

[13] See Train (2009, p. 40).

2.3 The discrete choice model with DRS

Positive DRS is an activity insurers are engaged in which increases the probability of being chosen by a particular individual (or group of individuals); of course, any such activity will also generate some cost.[14] We denote the cost insurer j incurs by a_j, and model the increase in the probability of being chosen to stem from an increase in the utility u_i^j (as perceived by individual i) by $g(a_j)$, where g is increasing and concave. With positive DRS, individual i's perceived utility when choosing insurer j therefore is

$$u_i^j = V_i^j + g(a_j) + \varepsilon_i^j. \tag{5}$$

Because we consider a setting where the benefit package is regulated and is thus identical for all insurers (so that the premium does not differ between insurers as well), V_i^j is the same for all insurers (but will be different for different individuals). Therefore, the probability that individual i chooses insurer k if there is positive DRS is given by

$$\begin{aligned} Prob(i \text{ chooses } k) &= Prob(V_i^k + g(a_k) + \varepsilon_i^k > V_i^l + g(a_l) + \varepsilon_i^l \;\forall\, l \neq k) \\ &= Prob(g(a_k) + \varepsilon_i^k > g(a_l) + \varepsilon_i^l \;\forall\, l \neq k) \\ &= \frac{e^{\frac{g(a_k)}{\sigma}}}{\sum_j e^{\frac{g(a_j)}{\sigma}}}. \end{aligned} \tag{6}$$

Likewise, negative DRS is an activity insurers are engaged in which decreases the probability of being chosen.[15] We denote the cost of negative DRS by b_j and the utility decrease as perceived by individual i by $f(b_j)$, where f is increasing and concave, so that individual i's utility when choosing insurer j is

$$u_i = V_i^j - f(b_j) + \varepsilon_i^j. \tag{7}$$

In this case, the probability that individual i chooses a particular insurer k is given by

$$Prob(i \text{ chooses } k) = \frac{e^{\frac{-f(b_k)}{\sigma}}}{\sum_j e^{\frac{-f(b_j)}{\sigma}}}. \tag{8}$$

This shows that DRS has the structure of a contest (with (6) and (8) as the contest success functions): There are several agents competing for a rent by spending money to increase the probability of receiving the rent: With positive DRS, insurers compete for individuals with $D_i > 0$; with negative DRS, they compete for the rent of not having to bear the loss associated with individuals for which $D_i < 0$.

The money spent in this risk selection contest is completely wasteful, and the insured do not want insurers to waste that money: It does not improve the quality of the benefit package but

[14] Examples for positive DRS are selective advertising or offering discounts for fitness club memberships.

[15] Activities falling into this category are that insurers require additional (unnecessary) paper work or involve the high risk individuals in phone calls in which they try to persuade these individuals to choose a different insurer. In fact, after a German sickness fund operating mainly in high cost areas went bankrupt in 2011, members of this fund, who then applied at other funds received phone calls in which some of the insurers told them that they could not continue their drug therapy or disease management program should they not choose a different insurer; see, e.g., Spiegel (2011).

eventually has to be borne by the insured whose premiums are increased. It is exactly this waste of money which constitutes the welfare loss of DRS.[16] The objective of the regulator therefore has to be to minimize insurers' investments in this risk selection contest. In the following two sections we show that the solution to this objective critically depends on which particular csf (within the class of contest success functions as given by (6) and (8)) best represents insurers' incentives for risk selection.

3 The MAD-criterion

3.1 Derivation of the MAD-criterion for the Tullock-csf

The contest success function used in the vast majority of all contest models is the Tullock-csf.[17] The general csf derived in the previous section encompasses the Tullock-csf, since, with $g(a) = \gamma \ln(a)$ and setting $\frac{\gamma}{\sigma} = m$,

$$Prob(i \text{ chooses } k \text{ with positive DRS}) = \frac{e^{\frac{\gamma \ln(a_k)}{\sigma}}}{\sum_j e^{\frac{\gamma \ln(a_j)}{\sigma}}} = \frac{(a_k)^{\frac{\gamma}{\sigma}}}{\sum_j (a_j)^{\frac{\gamma}{\sigma}}} = \frac{(a_k)^m}{\sum_j (a_j)^m}. \tag{9}$$

Likewise, if for the case of a negative rent we assume $f(b) = \delta \ln(b)$ and set $\frac{\delta}{\sigma} = n$, we have

$$Prob(i \text{ chooses } k \text{ with negative DRS}) = \frac{(b_k)^{-n}}{\sum_j (b_j)^{-n}}. \tag{10}$$

Employing the Tullock-csf as given in (9), expected profit for insurer k in the contest for a positive rent D_i is given by

$$\pi_k = \frac{(a_k)^m}{\sum_j (a_j)^m} D_i - a_k. \tag{11}$$

Solving the set of FOCs for all insurers yields the well known result that the equilibrium level of investments is[18]

$$a_j^* = \frac{(J-1)m}{J^2} D_i. \tag{12}$$

Therefore, the sum of investments of all insurers for a positive rent D_i equals

$$SI_i^+ = \sum_j a_j^* = \frac{(J-1)m}{J} D_i. \tag{13}$$

For a negative rent, using the contest success function as given in (10), insurer k's objective reads as

$$\pi_k = \frac{(b_k)^{-n}}{\sum_j (b_j)^{-n}} D_i - b_k. \tag{14}$$

[16] In Lorenz (2014b), it has been shown that in some settings it is the insurers who have to bear the cost of DRS. Nevertheless, the cost of DRS is a welfare loss in such a setting as well.

[17] The Tullock-csf has been employed for the analysis of activities as diverse as rent-seeking (Lockard and Tullock 2001), political campaigns (Skaperdas and Grofman 1995) or sports (Szymanski 2003).

[18] See, e.g., Nitzan (1994).

The equilibrium level of each insurer's investment is

$$b_j^* = \frac{(J-1)(-n)}{J^2} D_i, \tag{15}$$

so that the sum of investments of all insurers is given by

$$SI_i^- = \sum_j b_j^* = \frac{(J-1)n}{J} |D_i|. \tag{16}$$

Comparing the sum of investments for a positive and a negative rent of equal absolute value shows that insurers invest more for a positive rent if $m > n$: If investments for a positive rent are more effective (in the sense that $g'(a) > f'(b) \; \forall \, a = b$), incentives for positive DRS are higher than incentives for negative DRS. Accordingly, the regulator should focus on reducing positive rents. In the following, we will first assume that $n = m$, so that positive and negative DRS are equally important. We will consider the more realistic case that $m \neq n$ in Section 3.3.

If $m = n$, the total sum of investments, TSI, of all insurers for all rents D_i is given by

$$TSI = \sum_{i^+} SI_i^+ + \sum_{i^-} SI_i^- = \sum_{i=1}^{I} \frac{(J-1)m}{J} |D_i| = \frac{(J-1)mI}{J} MAD. \tag{17}$$

For the symmetric case of $m = n$, we can therefore state the following proposition:

Proposition 1. *In a symmetric contest with a Tullock-contest success function, the mean absolute deviation is the correct measure for insurers' incentives for risk selection.*

For the Tullock-csf, the total sum of investments is proportional to the MAD. This implies that the regulator can minimize the welfare loss caused by DRS by minimizing the MAD. However, in the next section, we show that such a risk adjustment scheme is in general not feasible, so that the MAD-criterion has to be qualified.

3.2 The estimation method for the MAD-criterion

Let X be an $I \times T$-matrix, containing in each column t one of the T variables used by the regulator (including the constant), and denote the i'th row of X by x_i'. Then the regulator's objective is given by

$$\min_{\beta} \sum_{i=1}^{I} |c_i^H - x_i'\beta|. \tag{18}$$

Using the estimated coefficients, $\hat{\beta}$, cost predictions are given by $c^R = X\hat{\beta}$. However, the regulator cannot use these cost predictions as the transfers in a risk adjustment scheme because the sum of these predictions in general does not equal the sum of c^H. This can most easily be seen if X only consists of a constant, in which case the optimization problem becomes

$$\min_{\beta_1} \sum_{i=1}^{I} |c_i^H - \beta_1|. \tag{19}$$

The solution to this minimization problem is the median: $\hat{\beta}_1 = c^{median}$. Since health care expenditures are usually skewed to the right, the median will be below average cost. If the regulator used this estimate and set transfers equal to the median, he would minimize incentives for risk selection, but insurers would make a loss. The solution which minimizes the MAD is therefore not feasible for risk adjustment.

To ensure a balanced budget for the risk adjustment scheme so that insurers neither make profits nor losses due to the transfers set by the regulator, the following constraint has to be satisfied:

$$\sum_i c_i^R = \sum_i c_i^H. \tag{20}$$

Therefore, the full optimization problem of the regulator is given by

$$\min_{\beta} \sum_i |c_i^H - x_i'\beta| \quad \text{s.t.} \quad \sum_i (c_i^H - x_i'\beta) = 0, \tag{21}$$

which is equivalent to the optimization problem of a restricted quantile regression. For estimation purposes, it is convenient to reformulate it as an unrestricted optimization problem. Expressing constraint (20) as

$$\bar{x}'\beta = \bar{c}^H, \tag{22}$$

where the bar represents the mean, solving for the last element of β, i.e., β_T, and plugging into (18) yields

$$\min_{\beta_1,\ldots,\beta_{T-1}} \sum_i |(c_i^H - \frac{\bar{c}^H}{\bar{x}^T} x_i^T) - \sum_{t=1}^{T-1} (x_i^t - \frac{\bar{x}^t}{\bar{x}^T} x_i^T)\beta_t|. \tag{23}$$

This is the optimization problem of an unrestricted quantile regression for the 0.5-quantile with $c_i^H - (\bar{c}^H/\bar{x}^T)x_i^T$ as the dependent and $x_i^t - (\bar{x}^t/\bar{x}^T)x_i^T$ as the explanatory variables.[19] Having estimated $(\hat{\beta}_1, \ldots, \hat{\beta}_{T-1})$, one can determine $\hat{\beta}_T$ using (22).

3.3 Asymmetric investments

So far we have derived the estimation method which minimizes insurers' investments for the symmetric case of $m = n$. We now consider the more realistic case that $m \neq n$. In this case, the total sum of investments is given by

$$\begin{aligned} TSI &= \sum_i \left(\frac{(J-1)m}{J} D_i \mathbb{1}_{(D_i>0)} - \frac{(J-1)n}{J} D_i \mathbb{1}_{(D_i\leq 0)} \right) \\ &= \frac{(J-1)(m+n)}{J} \sum_i \left(\frac{n}{m+n}(-D_i)\mathbb{1}_{(D_i\leq 0)} - \frac{m}{m+n}(-D_i)\mathbb{1}_{(D_i>0)} \right), \end{aligned} \tag{24}$$

[19] It can be solved using the simplex algorithm, see Barrodale and Roberts (1974). For large data sets, alternatives to the simplex algorithm like the interior point method have to be used. Most statistical software packages have implemented different algorithms for quantile regression.

where $\mathbb{1}_{(\cdot)}$ is the indicator function. Substituting $\alpha = \frac{n}{m+n}$ and $D_i = x_i'\beta - c_i^H$ shows that the total sum of investments can be minimized by solving

$$\min_{\beta} \sum_i \left((c_i^H - x_i'\beta)\alpha \mathbb{1}_{(c_i^H - x_i'\beta \geq 0)} - (c_i^H - x_i'\beta)(1-\alpha)\mathbb{1}_{(c_i^H - x_i'\beta < 0)} \right). \tag{25}$$

This is the optimization problem of a general quantile regression for the α-quantile.[20]

If $m < n$, insurers invest less for a positive rent than for a negative rent of equal absolute value. The regulator should therefore put less emphasis on reducing positive rents than on reducing the absolute value of the negative rents. Since rents are defined as $D_i = c_i^R - c_i^H = x_i'\hat{\beta} - c_i^H$, while residuals are defined as $e_i = c_i^H - x_i'\hat{\beta}$, a positive rent corresponds to a negative residual, and vice versa. Putting more emphasis on negative rents therefore requires putting more emphasis on positive residuals in the regression. This is exactly what is achieved by a quantile regression for an α-quantile with $\alpha > 0.5$; the larger α, the higher the weight on positive residuals. With $\alpha = 1$, only positive residuals, i.e., only negative rents are considered, while for $\alpha = 0$, negative rents are ignored. $\alpha = 1$ therefore captures the case that insurers are only engaged in negative DRS, and $\alpha = 0$ that they are only engaged in positive DRS.

The solution to (25) will in general depend on α, i.e., $\hat{\beta} = \hat{\beta}(\alpha)$. Therefore, to determine the optimal transfers $c^R(\alpha)$, it seems necessary to know m and n (or $\frac{m}{n}$, to be more precise). If this was correct, the optimal transfers could not be calculated because it is unlikely that a regulator could infer these parameters (with reasonable precision). However, the solution to (25) which satisfies the balanced budget constraint does not depend on α, so that we can state the following proposition:

Proposition 2. *For the Tullock-csf, the optimal transfers which minimize insurers' investments in the risk selection contest are independent of whether insurers are primarily engaged in positive or negative DRS.*

Proof: See Appendix A.1.

This result can most easily be understood by noting the following implication of the constraint: If, e.g., $\alpha > \frac{1}{2}$, the regulator will want to put more emphasis on reducing positive residuals than on reducing the absolute value of negative ones. In the unrestricted quantile regression this can be achieved by reducing the sum of positive residuals and increasing the sum of negative ones. In the restricted regression this is not feasible, because the sum of positive residuals always has to equal the sum of negative residuals so that the balanced budget constraint is satisfied. The optimal transfers can therefore be determined without knowing whether insurers invest more for a positive or a negative rent of equal absolute value.

[20] See Koenker (2005).

4 The R^2-criterion

4.1 Rationalization of the R^2 criterion

The preceding section showed that in a contest model with the Tullock-csf, the MAD is the correct measure for insurers' incentives for risk selection. However, in almost all risk adjustment schemes transfers are determined by means of a least squares regression, which maximizes the R^2. The R^2 is also the criterion that is used to choose among competing models (e.g. if there are different ways to employ diagnostic information).[21] Maximizing the R^2, the explained part of the variance, is equivalent to minimizing the unexplained part of the variance. Because the unexplained part of the variance is just the sum of squared deviations, it is obvious that the R^2 criterion applies if insurers' investments are proportional to the square of the rents.

If for the general contest success functions (6) and (8) we assume $g(a) = \gamma\sqrt{a}$ and $f(b) = \delta\sqrt{b}$, and set $\frac{\gamma}{\sigma} = m$ and $\frac{\delta}{\sigma} = n$, we arrive at the csf as given in (27) and (28) below. With these contest success functions, the solutions to the insurers' objectives as stated in (11) and (14) are

$$a_j^* = \frac{m^2(J-1)^2}{4J^4} D_i^2 \text{ for } D_i > 0 \quad \text{and} \quad b_j^* = \frac{n^2(J-1)^2}{4J^4} D_i^2 \text{ for } D_i < 0. \tag{26}$$

As is apparent, these investments are proportional to the square of the rent.[22] Because insurers' investments for positive and negative rents differ if $m \neq n$, for the R^2-criterion, which puts equal weight on positive and negative deviations, we have to have $m = n$. We can therefore state the following proposition:

Proposition 3. *The R^2-criterion can be rationalized in a contest model for the contest success functions*

$$Prob(i \text{ chooses } k \text{ with positive DRS}) = \frac{e^{m\sqrt{a_k}}}{\sum_j e^{m\sqrt{a_j}}} \tag{27}$$

and

$$Prob(i \text{ chooses } k \text{ with negative DRS}) = \frac{e^{-n\sqrt{b_k}}}{\sum_j e^{-n\sqrt{b_j}}} \tag{28}$$

with $m = n$.

If these two contest success functions apply and $m = n$, insurers' incentives for risk selection are minimized using the least squares regression. Because for any least squares regression the sum of the residuals equals zero by definition, condition (22) is always satisfied and does not have to be stated as an explicit constraint.

[21] Part of the risk adjustment literature deals explicitly with determining the maximum R^2 that can be achieved by different regression models, see van de Ven and Ellis (2000), Section 3.2.6.

[22] In Lorenz (2014a) it has been shown that the csf as given in (27) is the only one within the class of contest success functions given by (6) for which investments in equilibrium are proportional to the square of the rent. There it has also been shown that for the existence of an equilibrium in pure strategies with two players, $0 < m < 5.49D^{-\frac{1}{2}}$ has to be satisfied; the same inequalities hold for n with D replaced by $|D|$.

4.2 The estimation method for asymmetric investments

We now consider the case that $m \neq n$. If $m < n$, insurers' investments for a positive rent are smaller than for a negative rent of equal absolute value. This requires the regulator to put less emphasis on reducing positive rents than on reducing the absolute value of negative rents. This is achieved by employing an asymmetric least squares regression (also termed expectile regression).[23] Following Schnabel (2011), in the remainder of this paper we will refer to this regression as the LAWS (least asymmetrically weighted squares) regression.

Because for the LAWS regression (as for the quantile regression) the sum of residuals does in general not equal zero, for $m \neq n$ we have to explicitly state the balanced budget constraint (22). Therefore, to minimize insurers' investments, the regulator has to solve

$$\min_{\beta} \sum_i \left((c_i^H - x_i'\beta)^2 \alpha \mathbb{1}_{(c_i^H - x_i'\beta \geq 0)} + (c_i^H - x_i'\beta)^2 (1-\alpha) \mathbb{1}_{(c_i^H - x_i'\beta < 0)} \right) \quad \text{s.t. } \bar{x}'\beta = \bar{c}^H, \tag{29}$$

where $\alpha = \frac{n^2}{m^2+n^2}$. We will refer to this regression as the RLAWS (restricted LAWS) regression. Employing the same transformation of variables as used in (23) for the quantile regression, (29) can be reformulated as an unrestricted asymmetric least squares regression with $c_i^H - (\bar{c}^H/\bar{x}^T)x_i^T$ as the dependent and $x_i^t - (\bar{x}^t/\bar{x}^T)x_i^T$ as the explanatory variables.[24]

Unlike with the quantile regression, the balanced budget constraint does not imply that the solution to (29) is independent of α. Therefore, if insurers' investments are proportional to the square of the rent, the optimal transfers are not independent of whether insurers invest more for a positive or for a negative rent of equal absolute value. As we show in Section 6, the effectiveness of the transfers calculated form the different regression models crucially depends on which of the two cases applies. Although (as we already argued in Section 3.3) it will be difficult for a regulator to infer $\frac{m}{n}$ exactly, he may nevertheless know whether positive or negative DRS is the more severe problem in the health insurance market he is responsible for and, accordingly, choose a different regression model for the two cases.

5 Comparison of the LS, the RLAWS and the RQ regression

Although the difference between the LS, the quantile and the LAWS regression could be considered straightforward, the difference for the restricted versions (RQ and RLAWS) may not be immediately obvious. We therefore illustrate this difference with a simple example.

There are ten individuals which can be distinguished according to a dummy variable x_1, say, gender. Five of the individuals are male ($x_1 = 0$), five are female ($x_1 = 1$). Cost predictions of insurers for these individuals are as given in Figure 1, where each dot represents an

[23] See Newey and Powell (1987).

[24] This regression can be estimated using, e.g., the expectreg-package of the statistical software R, see Sobotka et al. (2014). The expectreg-package does not allow estimating a model without an intercept, as is necessary with these transformed variables. I thank Jan Pablo Burgard for adapting the expectreg.ls-function so that a model without an intercept could be estimated.

individual. Both cost distributions have the same mean and are skewed to the right, but the skewness is higher for males than for females.

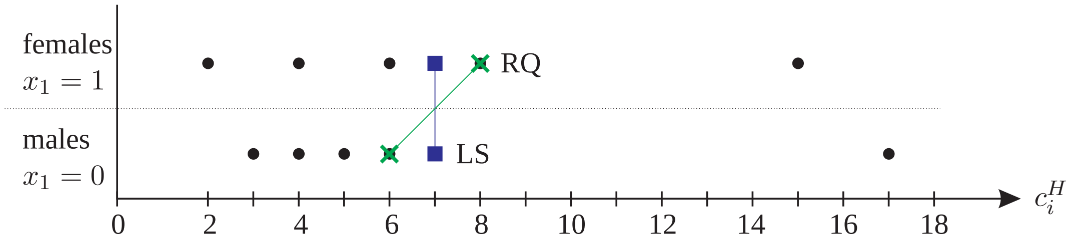


Figure 1: Difference between the LS and the RQ regression for discrete cost distributions

In the LS regression for $c_i^H = \beta_0 + \beta_1 x_{1i} + \eta_i$, coefficient β_0 will be estimated as the mean cost of males, and β_1 as the difference of the mean cost of females compared to males. Since average cost is 7 for both groups, $\hat{\beta}_0 = 7$ and $\hat{\beta}_1 = 0$, so that the regulator's cost predictions and transfers are $c^R(males) = c^R(females) = 7$, as indicated by the blue squares in Figure 1.

In this example, the number of individuals below (and above) the two cost predictions c^R from the LS regression is different for the two groups. This allows the regulator to reduce the sum (or mean) of absolute deviations by deviating from these cost predictions. However, any other pair of cost predictions just as well has to satisfy the balanced budget constraint. Reducing $\hat{\beta}_0$ by some $\Delta\hat{\beta}$ therefore requires increasing $\hat{\beta}_1$ by $2\Delta\hat{\beta}$. In this example, such a reduction of $\hat{\beta}_0$ accompanied by the respective increase of $\hat{\beta}_1$ will indeed reduce the sum of absolute deviations: First, the decrease of $\hat{\beta}_0$ reduces the absolute value of the negative residuals for four individuals, while the increase of $\hat{\beta}_1$ increases it for only three individuals; this reduces the sum of absolute deviations by $\Delta\hat{\beta}$. Secondly, the decrease of $\hat{\beta}_0$ increases one positive residual, while the increase of $\hat{\beta}_1$ reduces two positive residuals; this again implies a reduction of the sum of absolute deviations by $\Delta\hat{\beta}$.

Reducing $\hat{\beta}_0$ (accompanied by the increase of $\hat{\beta}_1$) will reduce the sum of absolute deviations as long as the number of males with cost below $c^R(males)$ is larger than the number of females with cost below $c^R(females)$. The RQ regression therefore sets the cost predictions at the same quantile of the two cost distributions and chooses the quantile that satisfies the balanced budget constraint; (in this example, it is the 80%-quantile, see the green crosses in Figure 1).[25] The same applies for continuous cost distributions, and also for a continuous explanatory variable.[26]

Comparing the residuals of the LS and the RQ regression for the example in Figure 1 shows that the RQ regression reduces the absolute value of a relatively large number of negative

[25]If the number of individuals differs for the two groups, the argument has to be slightly altered to take this into account, but the result that the same quantile is chosen for both distributions still holds.

[26]If expected cost is linear in the continuous explanatory variable and, e.g., the skewness of the distribution (conditional on this variable) is decreasing in this variable, then the coefficient for this variable will be larger for the RQ than for the LS regression (so that cost predictions are reduced for small values of this variable and increased for large values of this variable).

residuals which are close to zero and increases a relatively small number of very large positive residuals. This is a general pattern of the RQ regression; we return to this in Section 6.4, where we show the residuals for real data.

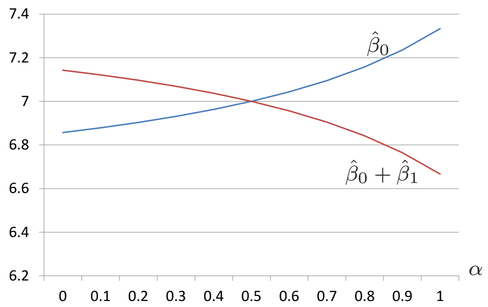


Figure 2: Coefficients for the RLAWS regression for the example of Figure 1

We finally consider the RLAWS regression. Results for this regression can be found in Figure 2, where we plot $\hat{\beta}_0$ and $\hat{\beta}_0 + \hat{\beta}_1$, i.e., $c^R(males)$ and $c^R(females)$, for different levels of α. For $\alpha = 0.5$, the RLAWS regression puts equal weight on positive and negative (squared) residuals and is therefore identical to the LS regression (so $\hat{\beta}_0 = 7$ and $\hat{\beta}_1 = 0$). For $\alpha < 0.5$, there is less weight on positive residuals. Because positive residuals (negative rents) are relatively large (compared to the absolute values of the negative residuals), this implies putting less weight on large residuals. This is similar to the quantile regression, which, compared to the LS regression, also puts less weight on large residuals (by weighting each residual by one instead of the absolute value of the residual itself as with the LS regression). Reducing α below 0.5 in the RLAWS regression is therefore a step into the direction of the RQ regression.[27]

6 Empirical analysis

In this section, we show that the differences between the LS, the RLAWS and the RQ regression discussed in the previous section can – in some cases – be substantial for real data.

6.1 Data

We present results for two different data sets: one of a German sickness fund covering the years 1998 to 2006, the other of a Swiss health insurer, covering the years 1997 to 1999. These panel data sets contain information on age, gender, cost, hospitalization, number of months insured and whether the individual died. There is no information on morbidity, a variable now used in several risk adjustment schemes. We can therefore not determine

[27]For $\alpha = 0$, the RLAWS can deviate from the LS regression to a greater or lesser extent than the RQ regression.

by how much the three types of regressions differ for the typical set of variables currently used for risk adjustment. Whether the results would be more pronounced with such a full set of variables is hard to tell a priori and depends on whether the cost distributions for the morbidity groups (defined, e.g., by DCGs) differ to a greater or lesser extent (in their skewness) than for, e.g., gender.

For both data sets, we show the results for the most recent year.[28] Because we use prior year expenditures to determine the cost prediction of the insurer, we only use those observations which are observable in both the last and the second to last year. We do not drop individuals which are only observable for part of the year; instead, their cost is annualized and in all regressions these observations are weighted by the fraction of the year they are observable.[29]

Table 1: Descriptive Statistics for the Swiss and German data set

Variable	Swiss (1999)	German (2006)
Male (in %)	45.8	50.8
Mean Age	41.3	31.6
Hospitalization in prior year (in %)	10.4	8.0
Number of observations	147,306	109,208

Descriptive statistics for the two data sets can be found in Table 1. The Swiss data set contains all expenditures covered by the health insurer, while in the German data set individuals' expenditures on ambulatory care are missing which at that time were covered by a uniform fee per capita insurers paid to the Association of SHI Physicians.

6.2 Prediction measures

With these two data sets, we determine the coefficients and the cost predictions for the LS, the RLAWS and the RQ regression. We then compare by how much these cost predictions, when used as transfers, reduce insurers' incentives for risk selection. To do so, we employ the appropriate (prediction) measures for the different settings we considered: Investments are either proportional to the square or the absolute value of the rent and are either equal or different for positive and negative rents of equal absolute value. For an overview of the four settings and the corresponding measures, see Table 2.

If insurers' investments are proportional to the square of the rents and equal for positive and negative rents of equal absolute value ($m = n$), the R^2 is the correct measure of insurers' incentives for risk selection. For $m \neq n$, it is the 'asymmetric version' of the R^2, given by

$$asymR^2(\alpha) = 1 - \frac{\sum_i \left((c_i^H - c_i^R)^2 \alpha \mathbb{1}_{(c_i^H - c_i^R \geq 0)} + (c_i^H - c_i^R)^2 (1-\alpha) \mathbb{1}_{(c_i^H - c_i^R < 0)} \right)}{\sum_i \left((c_i^H - \bar{c})^2 \alpha \mathbb{1}_{(c_i^H - \bar{c} \geq 0)} + (c_i^H - \bar{c})^2 (1-\alpha) \mathbb{1}_{(c_i^H - \bar{c} < 0)} \right)}.$$

[28] Results are very similar for the other years.

[29] Although in all regressions all observations are weighted by the fraction of the year they are observable, we do not alter the terminology used so far, i.e., we do not refer to these regressions as WLS, WRQ or WRLAWS.

Table 2: Insurers' investments, optimal regression model and corresponding prediction measure

Insurers' investments proportional to:	Symmetry of investments ($m = n$)	Optimal regression	Corresponding prediction measure
square of the rent	yes	LS	R^2
square of the rent	no	RLAWS	$asymR^2(\alpha)$
absolute value of the rent	yes	RQ	CPM
absolute value of the rent	no	RQ	CPM

If insurers' investments are proportional to the absolute value of the rent and $m = n$, the correct criterion is the MAD. To have a measure that is comparable to the R^2, we use Cumming's prediction measure, representing the explained part of the MAD:[30]

$$CPM = 1 - \frac{MAD(model)}{MAD(no\ model)} = 1 - \frac{\sum_i |c_i^H - c_i^R|}{\sum_i |c_i^H - \bar{c}|}.$$

Like with the R^2, we could, as a fourth measure, introduce an asymmetric version of the CPM, the $asymCPM(\alpha)$. However, because of the balanced budget constraint, the sum of the positive residuals always equals the absolute of the sum of the negative residuals, so any asymmetric weighting of positive and negative residuals does not alter the result, i.e., $CPM = asymCPM(\alpha)\ \forall\, \alpha$. Therefore, the CPM is the correct measure if insurers' investments are proportional to the absolute value of the rents, independent of whether insurers are primarily (or only) engaged in positive or negative DRS.

Like the R^2, the $asymR^2$ and the CPM are normalized to the unit interval. Therefore, with $c_i^R = \bar{c}^H$, all three measures assume the value zero, and with $c_i^R = c_i^H$, they all assume the value one.

6.3 Choice of the dependent variable

We present results for two different dependent variables: actual cost and a cost prediction of the insurer. In most risk adjustment schemes, the regulator uses actual cost as the explanatory variable in the regression. However, as shown in Section 2, the objective of the regulator has to be to minimize the difference between his transfers and insurers' cost predictions, not actual cost.

We therefore first present the results using insurers' cost predictions c^H as the dependent variable. We determine c^H as the predicted values from a (weighted) least squares regression using as explanatory variables age, age^2, age^3, hospitalization in the prior year, prior year expenditures and prior year expenditures squared, these six variables interacted with

[30] See Cumming et al. (2002).

a male dummy, and the male dummy itself.[31] This regression uses all the variables available in the two data sets. With better data, certainly more precise cost predictions could be calculated.[32]

If there are some signals used by insurers to determine their cost predictions which cannot be observed by the regulator even with such a better data set, the regulator cannot infer these predictions. This is of no consequence for the LS regression because, by the Frisch-Waugh-Lovell-theorem, the estimated coefficients from the regression with actual cost as the dependent variable are identical to the coefficients from the regression with insurers' cost prediction as the dependent variable.[33] The FWL-theorem, however, does not apply to the quantile and the LAWS regression, because these regressions are not orthogonal projections. In these settings, the optimal transfers which minimize incentives for risk selection simply cannot be calculated.

6.4 Results for insurers' cost prediction as the dependent variable

We first compare the coefficients of the three types of regressions for the simplest case of only one explanatory variable (see Table 3): In Model 1, we use gender, in Model 2, age (as a continuous variable) and in Model 3, a dummy variable for hospitalization in the prior year. In all three cases, the coefficients for the RQ regression differ markedly from those of the LS regression. E.g., for hospitalization in the prior year, transfers increase by 1,350 € for the LS regression, but only by 1,011 € for the RQ regression. It can also be seen that the RLAWS regression with $\alpha < 0.5$ is always a step into the direction of the quantile regression, and away from it for $\alpha > 0.5$.

Table 3: Estimated coefficients for the LS, the RLAWS and the RQ regression; German data set; dependent variable: cost prediction of insurer c^H; explanatory variable: Model 1: male dummy; Model 2: age; Model 3: hospitalization in the prior year

	Model 1		Model 2		Model 3	
	intercept	male	intercept	age	intercept	hosp_{t-1}
RLAWS ($\alpha = 0.6$)	606.41	-86.67	-7.21	18.00	430.83	1639.74
LS	613.45	-100.53	42.02	16.45	454.03	1350.47
RLAWS ($\alpha = 0.4$)	619.16	-111.76	95.75	14.75	473.19	1111.58
RQ	652.58	-177.51	203.38	11.35	481.26	1011.07

[31] The R^2 for these regressions is 19.6% for the German and 46.6% for the Swiss data set. The R^2 for the Swiss data set seems very high; however a similar figure has been reported by Beck (2004). The high value is mostly due to the fact that insurers only have to pay 50% of inpatient bills in Switzerland, which have a particularly high variance.

[32] Such a better data set may have to be collected by the regulator at some cost. Stam et al. (2010) analyze such a setting where a regulator incurs some cost to collect a small data set with more variables than usually observable to him.

[33] See Davidson and MacKinnon (2004).

However, more important than these differences in the coefficients are the differences in the cost predictions derived from these coefficients as evaluated by the prediction measures. We begin by comparing the R^2 and the CPM for the LS and the RQ regression (for different models with different explanatory variables), see Table 4.

Table 4: Predictive performance of different regression models; German data set; dependent variable: cost prediction of insurer c^H

Model	constant	male	age	age · male	hosp_{t-1}	hosp_{t-1} · male	age^2, age^2 · male, age^3, age^3 · male	least squares regression (LS)	restricted quantile regression (RQ)	least squares regression (LS)	restricted quantile regression (RQ)	$R^2(LS) - R^2(RQ)$	$CPM(RQ) - CPM(LS)$
								R^2		CPM		ΔR^2	ΔCPM
1	✓	✓						0.18	0.08	1.43	1.75	0.11	0.31
2	✓		✓					5.38	4.86	8.52	13.24	0.52	4.72
3	✓	✓	✓					5.58	5.06	10.00	13.60	0.52	3.60
4	✓	✓	✓	✓				5.60	5.05	9.97	14.62	0.55	4.64
5	✓	✓	✓	✓	✓			14.39	13.07	16.84	24.81	1.33	7.97
6	✓	✓	✓	✓	✓	✓		14.96	13.75	17.92	25.02	1.21	7.11
7	✓	✓	✓	✓	✓	✓	✓	19.05	18.72	36.21	38.07	0.33	1.86

By definition, the LS regression always performs better than the RQ regression for the R^2 criterion, while the RQ regression always achieves a higher CPM. For all models, we find that the R^2 is not much higher for the LS than for the RQ regression (see the second to last column of Table 4). This is different for the CPM, which for some models is considerably higher for the RQ than for the LS regression (see the last column of Table 4). The results are similar for the Swiss data set (see the middle part of Table 5 in Appendix A.2), but less pronounced.[34] In both data sets, we therefore find that erroneously using the RQ regression when the LS regression should be used seems less problematic than vice versa, i.e., using the LS regression when the RQ regression should be used.

For one of the models (Model 3 with age and gender as the explanatory variables) we give a more detailed picture of the difference between the LS and the RQ regression by comparing the distributions of the residuals. In Figure 3, we plot the negative of these residuals, i.e., insurers' rents, for all percentiles of the distribution.[35]

[34]It seems reasonable to assume that this is due to the fact that expenditures in the Swiss data set are less skewed, since health insurers cover only half of inpatient cost, see Beck (2004).

[35]The curves are thus the inverse of the distribution functions.

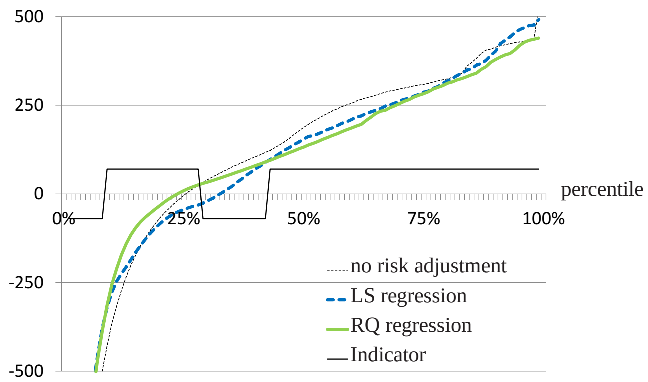
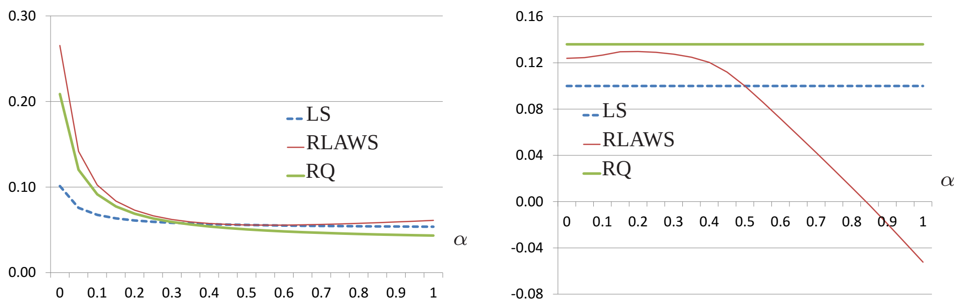


Figure 3: Insurers' rents without risk adjustment and for the LS and the RQ regression. German data set; Model 3; dependent variable: cost prediction of insurer c^H

The thin dotted curve shows the distribution of insurers' rents without risk adjustment (i.e., for $c_i^R = \bar{c}^H$), plotted in ascending order. As is to be expected, the number of negative rents (about 25% of all observations) is much smaller than the number of positive rents, and the absolute value of the largest negative rents is much larger than the largest positive rents. The blue dotted curve shows insurers' rents (i.e., the negative of the residuals) for the LS regression, and the green curve for the RQ regression. The black line represents an indicator function: if it is above zero, the absolute value of the rent is smaller for the RQ regression; if it is below zero, it is smaller for the LS regression. As can be seen from this indicator function, the LS regression yields smaller absolute values of the rents for the largest negative rents (up to the eighth percentile) and for the 28. to 42. percentile. For all the other percentiles (9 to 28 and 43 to 100), the absolute values of the rents are smaller for the RQ regression. As already discussed in Section 5, this is a general feature of the RQ compared to the LS regression: a small number of very large positive residuals (negative rents) are increased, while a large number of the remaining residuals are reduced.[36]

So far we have compared the results for the symmetric contest in which insurers invest the same amount for positive and negative rents of equal absolute value. We now consider the asymmetric case. Figure 4 shows the results, again for Model 3 with age and gender as the explanatory variables. The results for the $asymR^2$ for different levels of α can be found in Figure 4(a). (Recall that for low levels of α, there is a small weight on positive residuals, i.e., on negative rents; a low level of α therefore captures the case that insurers are primarily engaged in positive DRS.) Of course, the RLAWS regression performs best for the $asymR^2$ criterion (except for $\alpha = 0.5$, when it is identical to the LS regression). For $\alpha > 0.5$, the RLAWS and the LS regression perform somewhat better than RQ regression, which does not put as much weight on the few very large negative rents. On the other hand, for low levels of α, the RQ performs much better than the LS regression.

[36] van Barneveld et al. (2000) have suggested to also consider a 'modified version' of the MAD where all deviations below a certain threshold (e.g. 100 €) are ignored. As is to be expected from what we just derived for the residuals, for this measure, the advantage of the RQ over the LS regression is even larger than for the 'regular' MAD (or the CPM), because the RQ regression results in a considerably larger share of residuals below the threshold.



(a) $asymR^2(\alpha)$ (b) CPM

Figure 4: $asymR^2(\alpha)$ and CPM for the LS, the RLAWS and the RQ regression. German data set; Model 3; dependent variable: cost prediction of insurer c^H

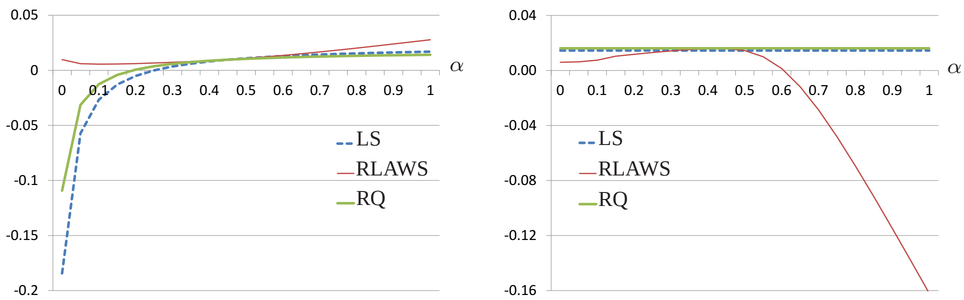
The results for the CPM are shown in Figure 4(b). Independent of the level of α, the highest CPM is of course achieved by the RQ regression; in this model, it is considerably higher than for the LS regression (see also the third line of Table 4). For $\alpha = 0.5$, the RLAWS equals the LS regression so that both achieve the same CPM. Being a step into the direction of the RQ regression for smaller levels of α, for $\alpha < 0.5$ the RLAWS performs somewhat better than the LS, but is still less successful than the RQ regression. For $\alpha > 0.5$, it performs considerably worse than the LS regression, and for $\alpha > 0.85$, it is even worse than setting a uniform transfer of $c_i^R = \bar{c}^H$. With α close to one, the RLAWS regression effectively aims to explain the few outliers with very high cost, and ignores the large number of small (negative) residuals which enter the CPM with equal weight. Therefore, the CPM achieved by the RLAWS regression drops drastically for high values of α.

In our data sets, we find the patterns shown in Figure 4(a) and (b) in all the models we estimated (i.e., for different sets of explanatory variables). If these patterns were also found in other data sets with more variables (especially morbidity variables now used in many risk adjustment schemes), one might draw the following conclusion: Since it is not clear whether insurers' investments are proportional to the absolute value or the square of the rent and whether insurers are primarily engaged in positive or negative DRS, it might be more appropriate to calculate transfers for a risk adjustment scheme by means of a restricted quantile instead of a least squares regression: If investments are proportional to the absolute value of the rent, it performs better for all levels of α; if investments are proportional to the square of the rent, the RQ regression performs considerably better than the LS regression for low levels of α and only somewhat worse for high levels of α. However, if the regulator is certain that insurers' investments are proportional to the square of the rent and that insurers are at least as much engaged in negative as they are in positive DRS, there is no need to change the common practice of using least squares regression for risk adjustment: In this case, it performs better than the RQ regression and only slightly worse than the RLAWS regression.

6.5 Results for actual cost as the dependent variable

Because in basically all risk adjustment schemes actual cost is used as the dependent variable in the regression and also to evaluate the performance of different models, we also present the results for actual cost. For the German data set, we only find very small differences in the R^2 and the CPM between the LS and the RQ regression (see the upper part of Table 5 in Appendix A.2); for the Swiss data set, we find results which are comparable to the regressions with insurers' cost predictions as the dependent variable (but somewhat less pronounced, see the middle and lower part of Table 5 in Appendix A.2).

In Figure 5 we replicate the results of Figure 4 for actual cost as the dependent variable. Again, the RQ regression performs considerably better than the LS regression according to the $asymR^2$-criterion for low levels of α, and only somewhat worse for high levels of α, while for the CPM criterion, the RLAWS regression performs much worse than the LS or the RQ regression for high levels of α. As is the case with insurers' cost prediction as the dependent variable, using the RQ regression is preferable to the LS regression unless insurers are primarily engaged in negative DRS and investments are proportional to the square of the rents.



(a) $asymR^2(\alpha)$ (b) $CPM(\alpha)$

Figure 5: $asymR^2(\alpha)$ and CPM for the LS, the RLAWS and the RQ regression. German data set; Model 3; dependent variable: actual cost

7 Conclusion

In this paper we have analyzed optimal risk adjustment for direct risk selection (DRS). Integrating insurers' activities for risk selection in a discrete choice model of individuals' health insurance choice shows that DRS has the structure of a contest. For the Tullock-contest success function used in most of the contest literature, optimal transfers have to be determined by means of a restricted quantile regression: This regression minimizes the mean absolute deviation conditional on satisfying the balanced budget constraint for the risk adjustment scheme. It is optimal regardless of whether insurers are primarily engaged in positive or negative DRS.

The common practice, however, is to use a least squares and not a quantile regression to determine the transfers. We have shown that the least squares regression can be rationalized in a discrete choice model for a new class of contest success functions. However, the least squares regression is only optimal if positive and negative DRS are equally important. If they are not, transfers have to be determined by means of a restricted asymmetric least squares regression.

In the empirical part of the paper, using data from a German sickness fund and a Swiss health insurer, we find considerable differences between the cost predictions of the three types of regressions. We also find an asymmetry in that the quantile regression never performs much worse than the least squares and the asymmetric least squares regression, but sometimes considerably better. If these results were also found in other data sets, in particular those containing information on morbidity now used in many risk adjustment schemes, a regulator who does not know which contest success function applies and whether positive or negative DRS is the more important problem in the health insurance market he is responsible for, might want to calculate transfers for the risk adjustment scheme by means of a restricted quantile instead of a least squares regression.

A Appendix

A.1 Proof of Proposition 2

Assume that β^* is the solution to (25) s.t. the balanced budget constraint (22) for $\alpha = \frac{1}{2}$, and $\widetilde{\beta} \neq \beta^*$ is the solution for $\alpha \neq \frac{1}{2}$. If X has full rank, then $X\beta^* \neq X\widetilde{\beta}$; this implies

$$c^H - X\beta^* = e(\beta^*) \neq e(\widetilde{\beta}) = c^H - X\widetilde{\beta}. \tag{30}$$

The vector of residuals, $e(\beta^*)$, satisfies the constraint that the sum of positive residuals, $\sum\limits_{i^+} e_i^+(\beta^*)$, equals minus the sum of negative residuals, $\sum\limits_{i^-} e_i^-(\beta^*)$. The same holds for $\widetilde{\beta}$.

The weighted sum of residuals, WSR, i.e., the value of (25), for $\alpha = \frac{1}{2}$ for the optimal β^* and the non-optimal $\widetilde{\beta}$ are

$$WSR(\beta^*|\alpha = \frac{1}{2}) \quad = \quad \frac{1}{2}\sum_{i^+} e_i^+(\beta^*) - \frac{1}{2}\sum_{i^-} e_i^-(\beta^*) = \sum_{i^+} e_i^+(\beta^*) \tag{31}$$

$$WSR(\widetilde{\beta}|\alpha = \frac{1}{2}) \quad = \quad \frac{1}{2}\sum_{i^+} e_i^+(\widetilde{\beta}) - \frac{1}{2}\sum_{i^-} e_i^-(\widetilde{\beta}) = \sum_{i^+} e_i^+(\widetilde{\beta}), \tag{32}$$

where the last equality in both equations holds because the constraint is satisfied.

If $\alpha \neq \frac{1}{2}$, the weighted sum of residuals for the optimal $\widetilde{\beta}$ and the non-optimal β^* are given by

$$WSR(\widetilde{\beta}|\alpha \neq \frac{1}{2}) \quad = \quad \alpha\sum_{i^+} e_i^+(\widetilde{\beta}) - (1-\alpha)\sum_{i^-} e_i^-(\widetilde{\beta}) = \sum_{i^+} e_i^+(\widetilde{\beta}) \tag{33}$$

$$WSR(\beta^*|\alpha \neq \frac{1}{2}) \quad = \quad \alpha\sum_{i^+} e_i^+(\beta^*) - (1-\alpha)\sum_{i^-} e_i^-(\beta^*) = \sum_{i^+} e_i^+(\beta^*). \tag{34}$$

Now, if $\sum_{i^+} e_i^+(\widetilde{\beta}) < \sum_{i^+} e_i^+(\beta^*)$, then β^* cannot have minimized $WSR(\beta|\alpha = \frac{1}{2})$. If, on the other hand, $\sum_{i^+} e_i^+(\beta^*) < \sum_{i^+} e_i^+(\widetilde{\beta})$, then $\widetilde{\beta}$ cannot have minimized $WSR(\beta|\alpha \neq \frac{1}{2})$. This implies that both optimization problems must yield the same residuals and therefore the same cost predictions, so $\widetilde{\beta} = \beta^*$.

A.2 Regression results

Table 5: Predictive performance of different regression models

Model	constant	male	age	age · male	hosp_{t-1}	hosp_{t-1} · male	age^2, age^2 · male, age^3, age^3 · male	least squares regression (LS)	restricted quantile regression (RQ)	least squares regression (LS)	restricted quantile regression (RQ)	$R^2(LS) - R^2(RQ)$	$CPM(RQ) - CPM(LS)$
								R^2		CPM		ΔR^2	ΔCPM
German data set; dependent variable: actual cost													
1	✓	✓						0.04	-0.01	0.16	0.22	0.04	0.06
2	✓		✓					1.06	1.03	1.39	1.49	0.03	0.10
3	✓	✓	✓					1.09	1.04	1.45	1.62	0.05	0.17
4	✓	✓	✓	✓				1.10	0.95	1.57	1.83	0.15	0.26
5	✓	✓	✓	✓	✓			2.82	2.44	3.49	4.11	0.38	0.63
6	✓	✓	✓	✓	✓	✓		2.93	2.53	3.51	4.13	0.41	0.62
7	✓	✓	✓	✓	✓	✓	✓	3.74	3.21	5.58	5.98	0.52	0.41
Swiss data set; dependent variable: cost prediction of insurer													
1	✓	✓						1.16	1.03	1.61	1.72	0.13	0.11
2	✓		✓					19.73	18.67	15.95	19.62	1.07	3.67
3	✓	✓	✓					20.09	18.99	16.78	19.96	1.11	3.18
4	✓	✓	✓	✓				20.34	19.39	17.08	20.48	0.95	3.40
5	✓	✓	✓	✓	✓			35.02	34.13	24.69	28.18	0.89	3.48
6	✓	✓	✓	✓	✓	✓		35.08	34.21	24.77	28.20	0.86	3.42
7	✓	✓	✓	✓	✓	✓	✓	39.68	39.53	33.50	33.96	0.15	0.47
Swiss data set; dependent variable: actual cost													
1	✓	✓						0.56	0.46	1.05	1.15	0.10	0.11
2	✓		✓					9.52	8.79	6.00	8.63	0.72	2.63
3	✓	✓	✓					9.69	8.93	6.36	8.90	0.76	2.54
4	✓	✓	✓	✓				9.81	9.09	6.80	9.29	0.71	2.49
5	✓	✓	✓	✓	✓			16.89	16.21	10.52	13.17	0.68	2.65
6	✓	✓	✓	✓	✓	✓		16.92	16.23	10.57	13.17	0.68	2.60
7	✓	✓	✓	✓	✓	✓	✓	19.14	18.72	15.19	15.73	0.41	0.54

References

BARRODALE, I., AND F. ROBERTS (1974): "Solution to an Overdetermined System of Equations in the l_1 Norm," *Communications of the ACM*, 17(6), 319–320.

BECK, K. (2004): *Risiko Krankenversicherung*. Haupt Verlag, Bern, Stuttgart und Wien.

BIJLSMA, M., J. BOONE, AND G. ZWART (2011): "Competition leverage: How the demand side affects optimal risk adjustment," *TILEC Discussion Paper*, 2011-039.

BREYER, F., M. K. BUNDORF, AND M. V. PAULY (2011): "Health Care Spending Risk, Health Insurance, and Payment to Health Plans," in *Handbook of Health Economics*, ed. by M. V. Pauly, T. G. McGuire, and P. P. Barros, vol. 2, pp. 691 – 762. Elsevier.

CAO, Z., AND T. G. MCGUIRE (2003): "Service-level selection by HMOs in Medicare," *Journal of Health Economics*, 22(6), 915–931.

CHERNICHOVSKY, D., AND W. P. VAN DE VEN (2003): "Risk adjustment in Europe," *Health Policy*, 65(1), 1–3.

CONGLETON, R. D., A. L. HILLMAN, AND K. A. KONRAD (2008): *40 Years of Research on Rent Seeking 1/2*. Springer, Berlin · Heidelberg.

CUMMING, R. B., D. KNUTSON, B. A. CAMERON, AND B. DERRICK (2002): "A comparative analysis of claims-based methods of health risk assessment for commercial populations," *A research study sponsored by the Society of Actuaries*.

DAVIDSON, R., AND J. G. MACKINNON (2004): *Econometric Theory and Methods*. Oxford University Press, New York and Oxford.

ELLIS, R. P., AND T. G. MCGUIRE (2007): "Predictability and Predictiveness of Health Care Spending," *Journal of Health Economics*, 26(1), 25–48.

FRANK, R. G., J. GLAZER, AND T. G. MCGUIRE (2000): "Measuring adverse selection in managed health care," *Journal of Health Economics*, 19(6), 829–854.

GLAZER, J., AND T. G. MCGUIRE (2000): "Optimal risk adjustment in markets with adverse selection: An application to managed care," *American Economic Review*, 90(4), 1055–1071.

——— (2002): "Setting health plan premiums to ensure efficient quality in health care: minimum variance optimal risk adjustment," *Journal of Public Economics*, 84(2), 153–173.

JACK, W. (2006): "Optimal risk adjustment with adverse selection and spatial competition," *Journal of Health Economics*, 25(5), 908–926.

KOENKER, R. (2005): *Quantile Regression*. Cambridge University Press, New York.

KONRAD, K. (2009): *Strategy and Dynamics in Contests*. Oxford University Press, Oxford.

LOCKARD, A. A., AND G. TULLOCK (2001): *Efficient rent-seeking*. Kluwer, Boston.

LORENZ, N. (2014a): "A contest success function with rent dependent dissipation rate," *Economics Bulletin*, forthcoming.

——— (2014b): "The interaction of direct and indirect risk selection," *Trier University Working Paper*, 12/14.

MCGUIRE, T. G., J. GLAZER, J. P. NEWHOUSE, S.-L. NORMAND, AND J. SHI (2013): "Integrating risk adjustment and enrollee premiums in health plan payment," *Journal of Health Economics*, 32(6), 1263–1277.

NEWEY, W. K., AND J. L. POWELL (1987): "Asymmetric least squares estimation and testing," *Econometrica*, 55(4), 819–847.

NEWHOUSE, J. P. (2002): *Pricing the priceless*. MIT Press, Cambridge and London.

NITZAN, S. (1994): "Modelling rent-seeking contests," *European Journal of Political Economy*, 10(1), 41–60.

SCHNABEL, S. (2011): *Expectile smoothing: new perspectives on asymmetric least squares. An application to life expectancy*. Proefschrift Universiteit Utrecht, Utrecht.

SCHOKKAERT, E., AND C. VAN DE VOORDE (2004): "Risk selection and the specification of the conventional risk adjustment formula," *Journal of Health Economics*, 23(6), 1237–1259.

SHEN, Y., AND R. P. ELLIS (2002): "How Profitable is Risk Selection? A Comparison of Four Risk Adjustment Models," *Health Economics*, 11(2), 165–174.

SKAPERDAS, S., AND B. GROFMAN (1995): "Modeling negative campaigning," *American Political Science Review*, 89(1), 49–61.

SOBOTKA, F., S. SCHNABEL, L. SCHULZE WALTRUP, P. EILERS, T. KNEIB, AND G. KAUERMANN (2014): "expectreg: Expectile and Quantile Regression," *R package version 0.39*.

SPIEGEL (2011): "Fragwürdige Beratung: Krankenversicherer wimmelt Senioren ab," *http://www.spiegel.de/wirtschaft/soziales/fragwuerdige-beratung-krankenversicherer-wimmelt-senioren-ab-a-761384.html*, May 9, 2011.

STAM, P. J. A., R. C. VAN VLIET, AND W. P. VAN DE VEN (2010): "A limited-sample benchmark approach to assess and improve the performance of risk equalization models," *Journal of Health Economics*, 29(3), 426–437.

SZYMANSKI, S. (2003): "The economic design of sporting contests," *Journal of Economic Literature*, 41, 1137–1187.

TRAIN, K. E. (2009): *Discrete Choice Methods with Simulation*. Cambridge University Press.

TULLOCK, G. (1980): “Efficient Rent Seeking,” in *Toward a theory of the Rent-Seeking Society*, ed. by J. Buchanan, R. Tollison, and G. Tullock. Texas A&M University Press, College Station.

VAN BARNEVELD, E. M., L. M. LAMERS, R. C. VAN VLIET, AND W. P. VAN DE VEN (2000): “Ignoring small predictable profits and losses: a new approach for measuring incentives for cream skimming,” *Health Care Management Science*, 3, 131–140.

VAN DE VEN, W. P., AND R. ELLIS (2000): “Risk adjustment in competitive health plan markets,” in *Handbook of Health Economics*, ed. by A. J. Culyer, and J. P. Newhouse, pp. 755–845. Elsevier, Amsterdam.

VAN VEEN, S. H. C. M., R. C. VAN KLEEF, W. P. M. M. VAN DE VEN, AND R. C. J. A. VAN VLIET (2014): “Improving the prediction model used in risk equalization: cost and diagnostic information from multiple prior years,” *European Journal of Health Economics*, forthcoming.

ZWEIFEL, P., F. BREYER, AND M. KIFMANN (2009): *Health Economics*. Springer.