Alberto Holly, University of Lausanne

Abstract: The issue of under- and over-compensation in the context of health insurance risk-adjustment models are well known phenomena that have attracted a growing interest in the recent years. One possible explanation for these phenomena is that the risk adjustment models estimated by OLS do not consider two well-known characteristics of health expenditure distributions, namely their positive skewness and long tail. However, this does not seem to be an issue for very large datasets, such as those used in risk adjustment models estimated at population level. If we accept this result, it is natural to ask what alternative explanations there might be for the persistence of the under- or overcompensation phenomena in risk adjustment models estimated with an extremely large sample. In this paper, we explore the hypothesis that the persistence of these phenomena could be due to the failure of risk adjustment models to take observable or unobservable heterogeneities into account. This is in particular the case of models estimated by OLS which assume that the regression coefficients are constant across the population. We have thus turned to econometric models of healthcare expenditure that allow for the relationship between the response and the covariates to vary directly with the levels of the response, accounting in this way for heterogeneity of effects. In a first exploration we have selected two models of particular interest, namely Fnite mixture (FM) and quantile regression (QR) models. We estimated these models with a small patient classification system (PCS) that we developed in a previous project using data for Switzerland. It is important to note that in this PCS, the risk adjuster variables generate a partition of the expenditure data, with each individual in the sample belonging to one and only one of the risk groups. This property is consistent with that of most PCSs used in risk adjustment models. Although FM models, in their latent class interpretation, are not feasible for risk adjustment mechanisms, our interest lies in the fact they provide a possible justification to the presence of under- and over-compensation phenomena even if all the risk adjuster variables are available and the dataset is extremely large. These phenomena would be due to a specification error regarding the conditional mean when unobserved heterogeneity is ignored. Our estimate of the distribution of healthcare expenditures with our database illustrates the value of this method. On the other hand, as shown by Lorenz (2017), quantile regression could be used in the context of risk adjustment mechanisms. We applied this method to estimate the quantile regression of expenditures at several quantile levels to determine whether and how the e§ects of the risk adjustment variables change as a function of quantile level. We then considered the estimation of a quantile risk adjustment model subject to the balanced budget constraint as in Lorenz (2017). We demonstrate that in any model where the PCS generates a partition of the explained variable, as in our case, the estimates of the constrained quantile regression model are easily obtained from the quantiles of the distributions of health expenditures in the risk groups. This result makes it also very easy to compare the coefficients of the constrained quantile regression with those obtained with OLS. Our estimation illustrates this general result. It leads to a form of arbitrage, the e§ects of which must be considered in terms of risk selective behavior of health insurers. At the end of this exercise, it seems that taking heterogeneity in the data into account might lead to interesting considerations in the context of risk adjustment models.