International Journal of Industrial Organization 37 (2014) 178–191

Contents lists available at ScienceDirect

International Journal of Industrial Organization

journal homepage: www.elsevier.com/locate/ijio

Estimating demand elasticities using nonlinear pricing☆

Christina M. Dalton *

Department of Economics, 204B Kirby Hall, Wake Forest University, Winston-Salem, NC 27109, United States

ARTICLE INFO

Article history:
Received 14 January 2013
Received in revised form 21 June 2014
Accepted 25 August 2014
Available online 16 September 2014

JEL classification:
D40
I11
C14

Keywords:
Elasticity
Nonlinear
Health insurance
Moral hazard

ABSTRACT

Nonlinear pricing is prevalent in industries such as health care, public utilities, and telecommunications. However, this pricing scheme introduces bias into estimating elasticities for welfare analysis or policy changes. I develop a local elasticity estimation method that uses nonlinear price schedules to isolate consumers' expenditure choices from selection and simultaneity biases. This method improves over previous approaches by using commonly-available observational data and requiring only a single general monotonicity assumption. Using claims-level data on health insurance with two nonlinearities, I am able to measure two separate elasticities, and find that elasticity declines from −0.26 to −0.09 by the second nonlinearity. These estimates are then used to calculate moral hazard deadweight loss. This method enables estimation of many policies with nonlinear pricing which previous tools could not address.

1. Introduction

Demand elasticities are important to policy makers for designing cost-sharing and calculating welfare in sectors such as health insurance, public utilities, and telecommunications. However, pricing is commonly nonlinear in these sectors, for example in deductibles in health insurance, tiered pricing in public utilities, and contracts with usage allowances in telecommunications.[1] Nonlinear pricing contributes to efficient plan design, but complicates estimation of elasticities for several reasons. First, the price a consumer faces is a function of quantity; consumers must pass a certain level of spending to reach a new price level. Second, selection bias occurs when an unobservable factor, such as health status or preferences for high versus low data use, pushes a consumer above or below the nonlinearity. Using observable variables such as age to proxy may not reduce bias, since unobservable health status is likely correlated with age. It is difficult to get rid of this selection bias without experimental data or an exogenous shock, both of which are empirically rare.

In this paper, I present a method to calculate elasticity in the presence of nonlinear pricing in consumer contracts. This method uses the nonlinearity itself to control for bias by taking advantage of the discontinuous change in price across the nonlinearity, while controlling for the underlying distribution of individual unobserved characteristics. This method has very general data requirements and uses only one minimally restrictive assumption: that the expenditure of interest must be increasing in the unobserved preference characteristic. I then apply this method to a private health insurance claims-level dataset with two nonlinearities. In addition to providing an updated health expenditure demand elasticity, my results also are novel because I am able to estimate elasticities at different points on the same demand curve. Identification uses the nonparametric estimation framework of Matzkin (2003). This method uses the same key insight as Bajari et al. (2010), but here I focus on individual consumer contracts rather than provider contracts. Consumers have less precise control over health expenditures given health status than Bajari et al. (2010) find using provider charges by hospitals over a diversity of expense categories. Besides the novel setting of consumer contracts, applying this method to health insurance contracts estimates health expenditure elasticities which are used to design contracts and make welfare predictions of insurance expansions. This paper is able to measure elasticities in two separate regions, which is informative since demand for health care likely changes along its typically skewed spending distribution.

The goal of the method is to generate local elasticities within a contract with nonlinear pricing. The method is aimed at policy applications such as understanding consumer behavior in certain regions, or how

☆ Particular thanks to Patrick Bajari and Robert Town. I also thank W. David Bradford, Meghan Busse, Meg Henkel, Thomas Holmes, Ginger Jin, Willard Manning, Amil Petrin, Connan Snider, Kevin Wiseman, and Thomas Youle for helpful comments.

* Tel.: +1 336 758 4495.
E-mail address: marshcl@wfu.edu.

[1] See for example, Reiss and White (2002), Herriges and King (1994), Maddock et al. (1992) in electricity, Szabo (2010) and Diakite et al. (2009) in water and development, and Grubb and Osborne (2012), Reiss and White (2006), Seim and Viard (2011), and Huang (2008) in cell phone markets.

http://dx.doi.org/10.1016/j.ijindorg.2014.08.007
0167-7187/

changing pricing schedules might affect the distribution of spending, given a particular consumer contract design.

The "gold standard" of elasticity estimation is experimental data. The best example in the health industry is the RAND Health Insurance Experiment (HIE), which began in 1971 and was conducted over 15 years (Manning et al., 1987; Newhouse, 1993). The RAND HIE avoided selection bias by randomizing patients into health plans' pricing schedules. While excellent for reducing selection bias, experimental data is extremely costly in terms of both time and money and is difficult to replicate. In addition, the results from the HIE best apply to the same population type and insurance framework of the HIE. This estimation method can be used on more specific populations of interest to policy makers or on new insurance structures. Since the RAND HIE, exogenous shocks or natural experiments have been used to control for simultaneity and selection bias. Cherkin et al. (1989) use the introduction of office visit copayments for government employees to create a quasi-experimental price change with which to measure elasticity. Selby et al. (1996) use a similar technique taking advantage of a copayment introduction for emergency room visits in a large HMO. In measuring price response more generally, Doyle and Almond (2011) find a substantial increase in mother's length of stay due to better insurance coverage around a policy treatment for children born just before and just after midnight. These natural experiments are difficult for policy makers to use regularly, however, because they rely on unique exogenous changes.

Eichner (1998) and Kowalski (2010) create a natural experiment in the presence of a deductible when an unexpected injury exogenously pushes other non-injured family members into a different pricing zone. Using a two-period utility model, Duarte (2012) also uses an unforeseen accident instrument on Chilean data to reveal how elasticities vary by income and demographics. However, unexpected injury in a deductible structure is hard to replicate in pharmaceutical or public utilities data, for example. The method presented here is accessible to policy makers outside of the health plan family deductible, a useful tool given the prevalence of nonlinear pricing in many other sectors.

Previous methods also estimate one elasticity over the whole range of expenditures. In health expenditures especially, distributions are commonly skewed, with a large proportion of consumers spending small amounts and a long tail of high spending consumers. Tiered pricing structures are often created precisely because different groups of consumers exist. Telecommunications users who end up near usage allowance limits are using bandwidth differently than low bandwidth users, i.e. using email versus video streaming. Estimating one elasticity over an entire range may mask heterogeneity of elasticity values along the distribution. An advantage of this method to policy makers is that it provides a local estimate of elasticity around current pricing points—those very areas that policy makers and insurance administrators may be modifying.

The main intuition of this estimation uses the kink in the price schedule at the nonlinearity. Selection bias exists because agents on either side of the nonlinearity face different prices, but are also different on an unobservable dimension such as health status or preferences for bandwidth use. In this paper's setting of a deductible, patients who surpass the deductible face a lower price for care, but also likely had more health shocks. However, the marginal price of an additional unit of care remains constant on each side, but changes suddenly at the nonlinearity. Identification is off the fact that marginal price is constant within the estimation regions, but the distribution of health status changes along the estimation window. Using the differences in the density of final spending before and after the nonlinearity allows us to isolate the change in spending due only to prices.

Identification is based on Matzkin (2003). The only condition that must hold is that final expenditure is strictly increasing in the individual unobserved characteristics that induce expenditure. For example, if an individual has a higher preference for bandwidth use, his final expenditure on bandwidth usage will be higher than an individual with a lower preference for use. For health insurance, this unobserved characteristic measure will be able to capture a more general ranking of health than diagnosis codes or self-reported health status. The unobserved characteristics are essentially a latent error term. Given this condition and using Matzkin (2003) I am able to proxy the distribution of unobservable characteristics using the percentiles of final expenditures.

Given both final expenditures and the estimated relative values of the unobserved characteristics, the method uses local linear regression to measure how expenditure increases for an increase in the unobserved characteristic. I calculate this slope on each side of the nonlinearity. The final elasticity estimate is the difference between the two slopes as they approach the nonlinearity and the threshold enrollee, thus controlling for selection and simultaneity bias while isolating the response due solely to price. I then plug this price response into an elasticity formula which includes the price level at the nonlinearity to calculate final elasticity.

I apply this method to a detailed claims-level dataset for an employer-sponsored Consumer Driven Health Plan (CDHP). This plan was chosen because it has two nonlinear pricing points. Although baseline implementation of this method only requires individual-level final expenditure and the pricing structure associated with the expenditures, the greater detail in my data allows me to perform several robustness checks of the method with observable variables. I find elasticity estimates of −0.26 in lower expenditure ranges compared with −0.09 in higher spending ranges. These estimates are slightly above and below the RAND HIE estimate of −0.22, which was not a local estimate, but instead estimated over a broad range of spending. Previous literature uses elasticities as an indicator of moral hazard in insurance. I take my elasticity estimate one step further to measure moral hazard deadweight loss by calculating the counterfactual choices the elasticity predicts. The deadweight loss from full-coverage insurance is approximately 20% of final expenditures less than $1000.

This paper builds on the elasticity estimation literature in health, but also into a more general nonlinear estimation literature. Maximum likelihood approaches such as in Gary and Hausman (1978) and Hausman (1985) require specific distributional assumptions, whereas the method outlined here uses nonparametric estimation and requires only one strict monotonicity assumption. Other tax applications, such as Blomquist and Newey (2002) require substantial variation in prices across sample observations, which is less likely to hold for pricing in the sectors above than for taxes. Recent work by Saez (2010) and Chetty et al. (2013) also look at nonlinearities in the EITC tax code. Saez finds evidence consistent with changing labor hours in response to changes in the tax code, but finds that the most pronounced changes can be attributed to tax evasion. The method here is related, but has the advantage that the main condition of monotonicity links the outcome of interest and unobserved characteristics more flexibly, which allows for the lack of distinct bunching cited by Saez. Aron-Dine et al. (2012) also highlight the highly nonlinear environment of health insurance. The authors examine expenditure response to health insurance price within a year, to address the problem that a patient's price changes along his distribution of expenditure. This work highlights the difficulties of calculating an elasticity using only one price over a large range of values. This question of forward-looking or myopic behavior is not of first-order concern in this paper, however, because this method targets those just below or just above a deductible—individuals with relatively similar probabilities of reaching a post-deductible price. Those individuals well beyond a nonlinearity are not in the scope of this estimation method or local elasticity.

This paper has three contributions. First, I present a new method for measuring elasticities with minimal distributional or modeling assumptions. The method has commonly attainable data requirements and can be applied to consumer contracts. Second, this method is based on a common feature which previously introduced bias in estimation, but can now be used in a variety of sectors. Using nonlinearities means that this method is most useful for local elasticities along expenditure distributions. Finally, I use this method to estimate elasticities for an