Bounding the Effect of Private Health Insurance on Dental Care Utilisation

Chuhui Li, D.S. Poskitt, and Xueyan Zhao

Department of Econometrics and Business Statistics
Monash University, Australia

January 29, 2016

Abstract

Recent developments in the literature of partial identification have significant implications for the econometric estimation of important policy effects. In the case of health economics, it is often of interest to estimate the effect of a binary policy treatment variable on a binary outcome variable where both may be driven by common observable and unobservable factors. A typical approach for health economists is to assume a parametric model, such as a bivariate probit, together with the use of instrumental variables to achieve point identification. Partial identification analysis of such problems allows for less restrictive assumptions for the underlying data generating process (DGP) in empirical applications, and the estimated bounds for policy measures evaluated under this framework offer more robust measures for policy impacts.

This paper applies the partial analysis approach to a health economics application. We estimate the bounds for average treatment effect (ATE) of private health insurance status on dental service utilisation under a partial identification framework, using data from the Australian National Health Survey. Four bounds from the literature under varying DGP assumptions and their 95% confidence regions are estimated. We show that the resulted confidence bounds for the ATE are much wider than the confidence intervals using a bivariate probit. We found that the bounds based on Chesher (2010) and Shaikh and Vytlacil (2011) have reasonably narrow widths to be informative. We also estimate the bounds for different types of individuals in the population, and we find that the width of the bounds can be very different for different sub-populations. We compare several estimation methods including parametric, non-parametric and semi-parametric smoothing estimators.

Keywords: partial identification, binary models, treatment effect, bounds, health economics.

JEL codes: C31, C35, C36, I13.

1 Introduction

We examine the impact of private health insurance (PHI) status on individuals' utilisation of dental care services in Australia under a partial identification framework using several bounds under alternative data generating assumptions. The role of private health insurance in the utilisation of health care services has long been of interest to health economists in many countries (see for example Manning et al. 1987; Cameron et al. 1988; Holly et al. 1998; Propper 2000; Savage and Wright 2003; Jones et al. 2006 and Kreider et al., 2014). This relationship is particularly of interest in Australia due to its unique mixed system of private and public hospital care.

Whilst all Australians enjoy a universal hospital cover (Medicare), around 50% of the population elect to pay for PHI in order to reduce waiting time for elective surgeries, to access more personalised care in private hospitals, and to receive subsidised dental and other auxiliary health services. To ease the mounting pressure to the public health system, the Australian government has introduced a series of financial incentives and penalties in the past decades to encourage PHI uptake, including an up to 30% rebate for the PHI premium from the public purse, an income tax surcharge (Medicare Levy) for high income earners without PHI, and Lifetime Community Rating to give incentives for people joining at a younger age (see Hall et al., 1999; Srivastava and Zhao, 2008). However, these policy changes also raised heated debates about opportunity cost for the public funds and equity issues (Duckett and Jackson, 2000; Wilcox, 2001). The two tier system creates significant inequality in health care access between the insured and uninsured, and long waiting time in public hospitals is a frequent topic of media coverage. In particular, inequality in dental care access is a major concern as dental service is not covered by Medicare (Hopkins et al., 2013).

The estimation of treatment effect of PHI on health care utilisation is challenged by the issue of identification due to potential non-random selection of treatment. Unobservable factors such as risk attitude and health status that influence the choice of taking up PHI also likely to influence an individual's decision to seek health care services, rendering the treatment variable being endogenous. Existing studies either ignored the potential endogeneity of treatment (Propper, 2000), or relying on the assumption of a parametric model, such as a bivariate probit, together with the use of instrumental variables to achieve point identification. For example, Holly et al. (1998), Jones et al. 2006 and Hopkins et al. (2013) used a bivariate probit model, whilst Harmon and Nolan (2001) and Höfter (2006) used a two-step IV Probit estimator. However, the assumptions of a threshold crossing rule for the binary variables, a separable error structure and a particular parametric functional form for the error terms are unlikely to be true of real data. Partial identification analysis of such problems allows for less restrictive assumptions for the underlying data generating process in empirical applications. And the estimated bounds for the impact of PHI evaluated under this framework offer more robust measures for evaluating policy impacts.

This paper estimates the bounds for average treatment effect (ATE) of private health insurance status on dental service utilisation under a partial identification framework, using data from the Australian National Health Survey. We apply the bounds derived in Chesher (2010) which do not assume an additive error structure or any particular functional form for the error distribution, and which have not been applied in empirical applications. We also estimate three other sets of ATE bounds based on Manski (1990); Manski and Pepper (2000) and Shaikh and Vytlacil (2011) under weaker or stronger assumptions than those in Chesher (2010). We use the PHI prices, hospital service co-payment and state level hospital bed density as instrumental variables (IVs) to shrink the bounds. We show that the resulted 95% confidence regions for these ATE bounds are much wider than the 95% confidence intervals using a particular parametric DGP bivariate probit. We found that the bounds based on Manski (1990) and Manski and Pepper (2000) are less useful in this application, whilst with tighter assumptions bounds based on Chesher (2010) and Shaikh and Vytlacil (2011) have reasonably narrow widths to be informative. We control for a large number of covariates and bounds for different sub-populations are estimated and are graphically illustrated. We find that the width of the bounds are very different for different sub-populations. The paper focuses on practical issues in applying bound analysis to real data, and compares several estimation methods including parametric, non-parametric and semi-parametric smoothing estimators. Whilst raw non-parametric estimation becomes infeasible for certain population types due to limited sample size in empirical applications, we find that parametric and QMLE semi-parametric estimators are convenient estimators and they are equally good. Whilst there have been many empirical applications using non-parametric bounds following Manski (1990), Manski and Pepper (2000) and Kreider and Pepper (2007) (Gerfin and Schellhorn, 2006; Gundersen and Kreider, 2008, 2009; Gundersen et al., 2011; Kreider et al., 2012, and Kreider et al., 2014), empirical application of tighter bounds based on Chesher (2010) and Shaikh and Vytlacil (2011) are limited Bhattacharya et al. (2012). This is the first partial identification analysis of ATE of health insurance on healthcare utilisation in the Australian context.

Section 2 presents some details of four bounds for the treatment effect from the literature for binary models. Section 3 introduces some background information on the private health insurance and the Australian health system. Section 4 details the data and Section 5 introduces the estimation methods. Section 6 presents some results of the bound of the treatment effect of private health insurance on dental service utilisation. Section 7 summarises the paper.

2 Partial identification of *ATE* for binary outcome models

Following the potential outcome framework in Heckman (1990) and Angrist and Pischke (2008), let Y_i represent the outcome observed for individual i and let $D_i = \{0, 1\}$ denote an indicator variable that designates the status of a treatment. Let Y_{1i} denote the outcome when individual i is treated ($D_i = 1$) and Y_{0i} the outcome when untreated ($D_i = 0$). The basic quantity of interest in most policy analysis and program evaluation problems is the treatment effect.

- The causal effect of the treatment (treatment effect) for individual i is given by $TE_i = [Y_{1i} - Y_{0i}]$.

- The average causal effect of the treatment is $ATE = \mathbb{E}[Y_1 - Y_0]$.

We wish to estimate *ATE*, but we never observe Y_{0i} and Y_{1i} for the same individual at the same time. How then can *ATE* be identified, if at all?

2.1 Bound in Chesher (2010)

Chesher (2005) provided conditions under which there is interval identification of the features of a structural function that depends on a discrete endogenous variable and is nonseparable in latent variates, and he showed in Chesher (2010) (See also Chesher, 2007) that single equation instrumental variable models for discrete outcomes are not point but set identified for the structural functions that deliver the values of the discrete outcome. To be specific, following Chesher (2010), consider a model for a scalar binary outcome variable Y of the following form:

Assumption CS 1 $Y = h(\mathbf{X},\ D,\ U)$ (a structural equation) where $\mathbf{X}$ denotes a vector of exogenous explanatory variables, D is a binary endogenous regressor, and the structure function h is weakly monotonic in U and normalized so that the marginal distribution of $U \in (0, 1)$ is uniform.

Suppose also that the following assumption is satisfied.

Assumption CS 2 There exists a vector of instruments $\mathbf{Z}$ such that the probability $P[U \leq \tau | \mathbf{Z} = \mathbf{z}] = \tau$ for all $\tau \in (0, 1)$ and all $\mathbf{z} \in \Omega_Z$.

Since Y is binary $h(\mathbf{x}, d, u)$ must be a non-decreasing step function, and recognizing that Y is a Bernoulli random variable, we find that $h(\mathbf{x}, d, u)$ is characterized by a probability threshold function $p(d, \mathbf{x})$ such that

$$Y = h(D, \mathbf{X}, U) = \begin{cases} 0, & 0 < U \leq p(D, \mathbf{X}) \\ 1, & p(D, \mathbf{X}) < U \leq 1 \end{cases}, \quad U \perp Z, \quad \mathbf{z} \in \Omega_Z, \quad U \sim Unif(0, 1) \tag{1}$$

where, without loss of generality, the support of Y and the "dummy" treatment variable D can be taken as $\{0, 1\}$. (The practical example we use here is where Y indicates if an individual has visited a dentist in the last year, $\mathbf{X}$ measures various socioeconomic and demographic factors, and D denotes whether the individual possess health insurance.)

From the definition of this model, the probability of the binary outcome taking the value 1 is $P(Y = 1|D, \mathbf{X}) = 1 - p(D, \mathbf{X})$. Thus, by definition, the *ATE* conditional on the exogenous variables $\mathbf{X}$ is $\mathbb{E}[Y_{1i}|\mathbf{X} = \mathbf{x}_i, D_i = 1] - \mathbb{E}[Y_{0i}|\mathbf{X} = \mathbf{x}_i, D_i = 0]$, which equals

$$\begin{aligned} P(Y_{1i} = 1|\mathbf{X} = \mathbf{x}_i, D_i = 1) - P(Y_{0i} = 1|\mathbf{X} = \mathbf{x}_i, D_i = 0) &= [1 - p(1, \mathbf{x}_i)] - [1 - p(0, \mathbf{x}_i)] \\ &= p(0, \mathbf{x}_i) - p(1, \mathbf{x}_i)\,. \end{aligned}$$

Under the weak monotonicity condition embodied in the above model an admissible structure delivers a conditional distribution for $(Y, \mathbf{X})$ given $\mathbf{Z}$, and distinct structures can deliver identical distributions of $(Y, \mathbf{X})$ given $\mathbf{Z}$ for all $\mathbf{z} \in \Omega_Z$. This can happen because variations in the threshold function $p(d, \mathbf{x})$ can be offset by altering the sensitivity of the distribution functions to variations in $\mathbf{x}$ so that the joint distribution function of Y and $\mathbf{X}$ conditional on $\mathbf{Z}$ is left unchanged. Such structures are observationally equivalent and the model is set, not point, identifying for the *ATE* because within a set of admissible observationally equivalent structures there can be more than one distinct structural function.

Let

$$\begin{aligned} f_0(\mathbf{x}, \mathbf{z}) &\equiv P[Y = 0 | \mathbf{X} = \mathbf{x}, D = 0, \mathbf{Z} = \mathbf{z}], \\ f_1(\mathbf{x}, \mathbf{z}) &\equiv P[Y = 0 | \mathbf{X} = \mathbf{x}, D = 1, \mathbf{Z} = \mathbf{z}], \\ g_0(\mathbf{x}, \mathbf{z}) &\equiv P[D = 0 | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}], \\ g_1(\mathbf{x}, \mathbf{z}) &\equiv P[D = 1 | \mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}]. \end{aligned} \tag{2}$$

denote the stated conditional probabilities. Then the following inequalities for $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ can be derived (Chesher, 2010):

$$\begin{aligned} p(0,\mathbf{x}) < p(1,\mathbf{x}): \quad & f_0(\mathbf{x},\mathbf{z})g_0(\mathbf{x},\mathbf{z}) \le p(0,\mathbf{x}) \le f_0(\mathbf{x},\mathbf{z})g_0(\mathbf{x},\mathbf{z}) + f_1(\mathbf{x},\mathbf{z})g_1(\mathbf{x},\mathbf{z}) \\ & \le p(1,\mathbf{x}) \le g_0(\mathbf{x},\mathbf{z}) + f_1(\mathbf{x},\mathbf{z})g_1(\mathbf{x},\mathbf{z}), \\ p(0,\mathbf{x}) \ge p(1,\mathbf{x}): \quad & f_1(\mathbf{x},\mathbf{z})g_1(\mathbf{x},\mathbf{z}) \le p(1,\mathbf{x}) \le f_0(\mathbf{x},\mathbf{z})g_0(\mathbf{x},\mathbf{z}) + f_1(\mathbf{x},\mathbf{z})g_1(\mathbf{x},\mathbf{z}) \\ & \le p(0,\mathbf{x}) \le g_1(\mathbf{x},\mathbf{z}) + f_0(\mathbf{x},\mathbf{z})g_0(\mathbf{x},\mathbf{z}). \end{aligned} \tag{3}$$

As $\mathbf{Z}$ is independent of U the bounds on $p(1, \mathbf{x})$ and $p(0, \mathbf{x})$ can be tightened by taking the intersection of the intervals in (3) for different values of $\mathbf{z} \in \Omega_{\mathbf{Z}}$. As a result, the *ATE* for an individual with characteristics $\mathbf{x}$,

$$ATE(\mathbf{x}) = \mathbb{E}[Y_1 - Y_0 | \mathbf{X} = \mathbf{x}] = p(0, \mathbf{x}) - p(1, \mathbf{x}),$$

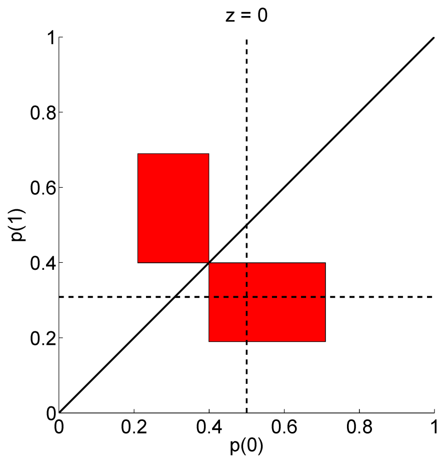
can be bounded by the interval

$$\begin{aligned} &\Big[\sup_{\mathbf{z}\in\Omega_Z} P(Y=0, D=0|\mathbf{x},\mathbf{z}) - \inf_{\mathbf{z}\in\Omega_Z} \{P(D=0|\mathbf{x},\mathbf{z}) + P(Y=0,D=1|\mathbf{x},\mathbf{z})\}, \\ &\inf_{\mathbf{z}\in\Omega_Z} \{P(Y=0|\mathbf{x},\mathbf{z})\} - \sup_{\mathbf{z}\in\Omega_Z} \{P(Y=0|\mathbf{x},\mathbf{z})\} \Big] \end{aligned} \tag{4}$$

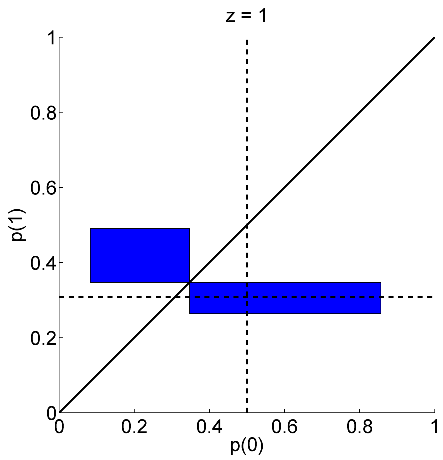
when $p(0, \mathbf{x}) \ge p(1, \mathbf{x})$, and when $p(0, \mathbf{x}) < p(1, \mathbf{x})$

$$\begin{aligned} &\Big[\sup_{\mathbf{z}\in\Omega_Z} \{P(Y=0|\mathbf{x},\mathbf{z})\} - \inf_{\mathbf{z}\in\Omega_Z} \{P(Y=0|\mathbf{x},\mathbf{z})\}, \\ &\inf_{\mathbf{z}\in\Omega_Z} \{P(D=1|\mathbf{x},\mathbf{z}) + P(Y=0,D=0|\mathbf{x},\mathbf{z})\} - \sup_{\mathbf{z}\in\Omega_Z} P(Y=0,D=1|\mathbf{x},\mathbf{z}) \Big]. \end{aligned} \tag{5}$$

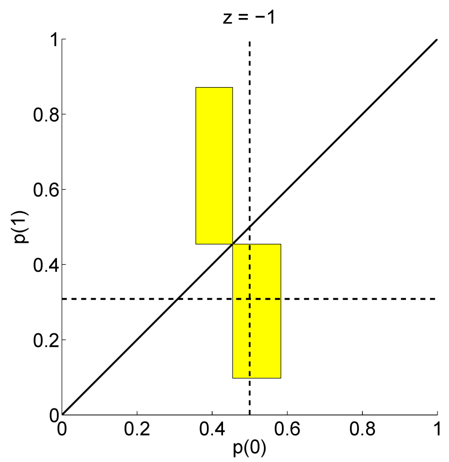
The intervals or bounds in (4) and (5) define identified sets for the *ATE*. Thus we find that even if the conditional probabilities in (2) were known (given to us by some oracle), the *ATE* would only be partially identified.



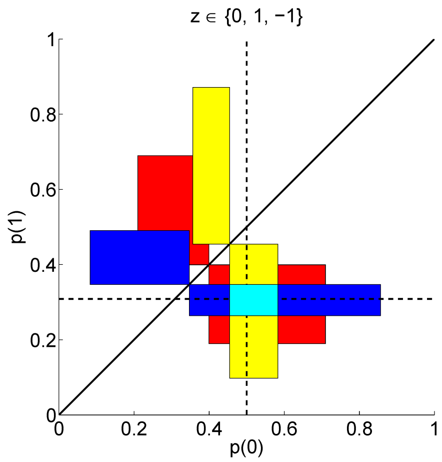
(a)



(b)



(c)



(d)

Figure 1. Illustration of the identified sets of $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ with respect to variation of Z. For either case of $p(0, \mathbf{x}) > p(1, \mathbf{x})$ and $p(0, \mathbf{x}) < p(1, \mathbf{x})$ (separated by the half-line), there is an identified set of $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ for each value of z. As z varies, the location and area of the identified sets of $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ will vary accordingly. The final identified set for $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ is the intersection of all these sets, which is the shaded area in the graph.

Figure 1 illustrates an example (Chesher, 2010) of partial identification of threshold functions $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ in the structural model (1). For either case of $p(0, \mathbf{x}) > p(1, \mathbf{x})$ and $p(0, \mathbf{x}) < p(1, \mathbf{x})$ (separated by the half-line), there is an identified set of $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ for each value of z. As z varies, the location and area of the identified sets of $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$ will vary accordingly, as shown in Figure 1a through Figure 1c. These identified sets are intersected in Figure 1d to obtain the final identified sets for $p(0, \mathbf{x})$ and $p(1, \mathbf{x})$, and hence the identified set for the *ATE*.

2.2 Bound in Shaikh and Vytlacil (2011)

Shaikh and Vytlacil (2011) imposed the weakly separable model

$$
\begin{aligned}
Y &= I\{\nu_1(D, X) \geq \varepsilon_1\} \\
D &= I\{\nu_2(Z) \geq \varepsilon_2\}
\end{aligned}
\tag{6}
$$

Here, Y denotes the observed binary outcome of interest, D denotes the observed binary endogenous regressor, X and Z are observed random vectors, and ε_1 and ε_2 are unobserved random variables. The identifying assumptions they introduced are listed below:

Assumption SV 1 $(X, Z) \perp\!\!\!\perp (\varepsilon_1, \varepsilon_2)$.

Assumption SV 2 The distribution of $(\varepsilon_1, \varepsilon_2)$ has strictly positive density with respect to Lebesgue measure on $\mathbf{R}^2$.

Assumption SV 3 The support of the distribution of (X, Z), supp(X, Z), is compact, and supp$(X, Z) =$ supp$(X) \times$ supp(Z).

Assumption SV 4 The functions $\nu_2(Z)|X$ is nondegenerate.

Assumption SV 5 The functions $\nu_1(\cdot)$ and $\nu_2(\cdot)$ are continuous.

Assumption SV 1 is a critical independence condition stating that the observed covariates are independent of the unobserved error terms. Assumption SV 2 needs no explanation. Assumption SV 3 implies that for all $\mathbf{x}, \mathbf{x}' \in$ supp(X) and $\mathbf{z} \in$ supp(Z), there must exist $\mathbf{z}' \in$ supp(Z) such that $p(\mathbf{x}, \mathbf{z}) = p(\mathbf{x}', \mathbf{z}')$. This is a "perfect matching" condition which may not be realized in many applications. We can try to search numerically in the support of Z (supp(Z)) to find a $\mathbf{z}'$ such that $|p(\mathbf{x}, \mathbf{z}) - p(\mathbf{x}', \mathbf{z}')| < \tau$, where τ is an appropriately defined tolerance, but this type of search for the "nearest neighbour" can be computationally intensive and time consuming. Assumption SV 4 is basically an exclusion restriction requiring that there is at least one variable in Z that is not a component of X, while SV 5 ensures that certain supremums and infimums can be obtained.

Let $P(Z) = \Pr(D = 1|Z)$. From Assumption SV 1 we have that $\Pr(Y_1 = 1|X) = \Pr(Y_1 = 1|X, P(Z))$ and $\Pr(D = 1|X, P(Z)) = P(Z)$. Using P in place of $P(Z)$ then gives us

$$
\begin{aligned}
&\Pr(Y_1 = 1|X, P) \\
&= \Pr(D = 1, Y_1 = 1|X, P) + \Pr(D = 0, Y_1 = 1|X, P) \\
&= \Pr(D = 1, Y = 1|X, P) + (1 - P)\Pr(Y_1 = 1|X, P, D = 0)
\end{aligned}
$$

$$
\begin{aligned}
&\Pr(Y_0 = 1|X, P(Z)) \\
&= \Pr(D = 0, Y_0 = 1|X, P) + \Pr(D = 1, Y_0 = 1|X, P) \\
&= \Pr(D = 0, Y = 1|X, P) + P\Pr(Y_0 = 1|X, P, D = 1)
\end{aligned}
$$

Now suppose Y and D are determined by the thresh-hold crossing model in (6) and that assumptions SV 1 to SV 5 hold. Let

$$
\begin{aligned}
&h(\mathbf{x}, \mathbf{x}', p, p') \\
&= \left\{\Pr(D = 1, Y = 1|X = \mathbf{x}', P = p) - \Pr(D = 1, Y = 1|X = \mathbf{x}', P = p')\right\} \\
&\quad - \left\{\Pr(D = 0, Y = 1|X = \mathbf{x}, P = p') - \Pr(D = 0, Y = 1|X = \mathbf{x}, P = p)\right\}
\end{aligned}
$$

Then, Shaikh and Vytlacil (2011) proved that whenever all the above conditional probabilities are well defined, we have for $p > p'$ that $h(\mathbf{x}, \mathbf{x}', p, p')$ and $\nu_1(1, \mathbf{x}') - \nu_1(0, \mathbf{x})$ share the same sign. In particular, the sign of $h(\mathbf{x}, \mathbf{x}', p, p')$ does not depend on p or p' provided $p > p'$.

Define

$$
H(\mathbf{x}, \mathbf{x}') = \mathbb{E}[h(\mathbf{x}, \mathbf{x}', p, p')|p > p']
$$

in which $h(\mathbf{x}, \mathbf{x}', p, p') = 0$ whenever it is not well defined. Let $X_0^+(\mathbf{x}) = \{\mathbf{x}' : H(\mathbf{x}, \mathbf{x}') \geq 0\}$, $X_0^-(\mathbf{x}) = \{\mathbf{x}' : H(\mathbf{x}, \mathbf{x}') \leq 0\}$, $X_1^+(\mathbf{x}) = \{\mathbf{x}' : H(\mathbf{x}', \mathbf{x}) \geq 0\}$, $X_1^-(\mathbf{x}) = \{\mathbf{x}' : H(\mathbf{x}', \mathbf{x}) \leq 0\}$, where $H(\mathbf{x}, \mathbf{x}')$ is defined above if $h(\mathbf{x}, \mathbf{x}', p, p')$ is well defined for some $p > p'$, and with each set understood to be empty if $h(\mathbf{x}, \mathbf{x}', p, p')$ is not well defined for any $p > p'$. Then we have the following properties:

1. If Assumption SV 1 and SV 2 hold, then $\Pr(Y_d = 1|X = \mathbf{x}) \in [L_d(\mathbf{x}), U_d(\mathbf{x})]$ for $d \in \{0, 1\}$, and

$$
\begin{aligned}
ATE &= \{\Pr(Y_1 = 1|X = \mathbf{x}) - \Pr(Y_0 = 1|X = \mathbf{x})\} \\
&\in [L_1(x) - U_0(\mathbf{x}), U_1(\mathbf{x}) - L_0(\mathbf{x})] \qquad (7)
\end{aligned}
$$

   where

$$
\begin{aligned}
L_0(\mathbf{x}) &= \sup_p \left\{\Pr(D = 0, Y = 1|X = \mathbf{x}, P = p) + \sup_{\mathbf{x}' \in X_0^-(\mathbf{x})} \{\Pr(D = 1, Y = 1|X = \mathbf{x}', P = p)\}\right\} \\
U_0(\mathbf{x}) &= \inf_p \left\{\Pr(D = 0, Y = 1|X = \mathbf{x}, P = p) + p \inf_{\mathbf{x}' \in X_0^+(\mathbf{x})} \{\Pr(Y = 1|D = 1, X = \mathbf{x}', P = p)\}\right\} \\
L_1(\mathbf{x}) &= \sup_p \left\{\Pr(D = 1, Y = 1|X = \mathbf{x}, P = p) + \sup_{\mathbf{x}' \in X_1^+(\mathbf{x})} \{\Pr(D = 0, Y = 1|X = \mathbf{x}', P = p)\}\right\} \\
U_1(\mathbf{x}) &= \inf_p \left\{\Pr(D = 1, Y = 1|X = \mathbf{x}, P = p) + (1 - p) \inf_{\mathbf{x}' \in X_1^-(\mathbf{x})} \{\Pr(Y = 1|D = 0, X = \mathbf{x}', P = p)\}\right\}
\end{aligned}
$$

2. If Assumption SV 1 to SV 4 hold and $\text{supp}(P \times X) = \text{supp}(P) \times \text{supp}(X)$, then the above bounds are sharp.

The proof of the above properties is long and complicated, see Shaikh and Vytlacil (2011) for details.

2.3 Bound in Manski (1990)

Prior to the technical developments presented in Chesher (2007, 2010), and Shaikh and Vytlacil (2011), Manski (1989, 1990) derived an estimable bound for $\mathbb{E}[Y|\mathbf{X}=\mathbf{x}]$ based only on the assumption that the distribution of Y was concentrated in a given interval $[K_{0x}, K_{1x}]$. When Y is binary, Y is definitionally bounded with $K_{0x}=0$ and $K_{1x}=1$. Without imposing any additional assumptions, we have

$$\begin{aligned}\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}] = \mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}, D=1]\, P(D=1|\mathbf{X}=\mathbf{x})\\ + \mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}, D=0]\, P(D=0|\mathbf{X}=\mathbf{x})\end{aligned} \tag{8}$$

Since $\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}, D=0]$ is not observed from the sample, $\mathbb{E}[Y_1=1|\mathbf{X}=\mathbf{x}, D=0]$ is not identified. However, when Y is binary, $\mathbb{E}[Y_1=1|\mathbf{X}=\mathbf{x}, D=0]$ is bounded between 0 and 1. As a result,

$$P(Y_1=1, D=1|\mathbf{X}=\mathbf{x}) \le \mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}] \le P(Y_1=1, D=1|\mathbf{X}=\mathbf{x}) + P(D=0|\mathbf{X}=\mathbf{x})$$

Similarly we can derive bounds on $\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x}]$

$$P(Y_0=1, D=0|\mathbf{X}=\mathbf{x}) \le \mathbb{E}[Y_0|\mathbf{X}=\mathbf{x}] \le P(Y_0=1, D=0|\mathbf{X}=\mathbf{x}) + P(D=1|\mathbf{X}=\mathbf{x})$$

The above two inequalities on $\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}]$ and $\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x}]$ can be denoted as

$$m_1(\mathbf{x},1)\, p(\mathbf{x}) \le p_1(\mathbf{x}) \le m_1(\mathbf{x},1)\, p(\mathbf{x}) + (1-p(\mathbf{x})) \tag{9}$$

$$(1-p(\mathbf{x}))\, m_0(\mathbf{x},0) \le p_0(\mathbf{x}) \le (1-p(\mathbf{x}))\, m_0(\mathbf{x},0) + p(\mathbf{x}) \tag{10}$$

where

$$\begin{aligned} m_0(\mathbf{x},d) &= \mathbb{E}[Y_0|\mathbf{X}=\mathbf{x}, D=d]\,,\\ m_1(\mathbf{x},d) &= \mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}, D=d]\,,\\ p(\mathbf{x}) &= P(D=1|\mathbf{X}=\mathbf{x})\,. \end{aligned}$$

As a result, the average treatment effect is bounded as

$$\begin{aligned} &(m_1(\mathbf{x},1)+m_0(\mathbf{x},0)-1)\, p(\mathbf{x}) - m_0(\mathbf{x},0)\\ &\le\ \mathbb{E}(Y_1-Y_0|\mathbf{X}=\mathbf{x})\ \le\\ &(m_1(\mathbf{x},1)+m_0(\mathbf{x},0)-1)\, p(\mathbf{x}) - m_0(\mathbf{x},0)+1 \end{aligned} \tag{11}$$

The width of this bound is 1 and it is called the worst case scenario. Optionally, there may be a set of instrumental variables $\mathbf{Z}$ such that for each $\mathbf{x}\in\mathbf{X}$ and $\mathbf{z}\in\mathbf{Z}$, the following conditional independence restriction is satisfied

$$\begin{aligned}\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}] = \mathbb{E}[Y_0|\mathbf{X}=\mathbf{x}]\\ \mathbb{E}[Y_1|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}] = \mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}]\end{aligned} \tag{12}$$

Using the conditional independence in (12), and extending the inequalities and equalities in (8) to (11) to the situation when there are instrumental variables $\mathbf{Z}$, Manski (1989, 1990) showed that by intersecting the bounds of $\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x}]$ and $\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x}]$ across the support of the instruments, the identified set of *ATE* could be tightened to

$$\left[\sup_{\mathbf{z}\in\Omega_{\mathbf{Z}}}\{m_1(\mathbf{x},\mathbf{z},1)\,p(\mathbf{x},\mathbf{z})\} - \inf_{\mathbf{z}\in\Omega_{\mathbf{Z}}}\{(1-p(\mathbf{x},\mathbf{z}))\,m_0(\mathbf{x},\mathbf{z},0)+p(\mathbf{x},\mathbf{z})\}\ ,\right.$$
$$\left.\inf_{\mathbf{z}\in\Omega_{\mathbf{Z}}}\{m_1(\mathbf{x},\mathbf{z},1)\,p(\mathbf{x},\mathbf{z})+(1-p(\mathbf{x},\mathbf{z}))\} - \sup_{\mathbf{z}\in\Omega_{\mathbf{Z}}}\{(1-p(\mathbf{x},\mathbf{z}))\,m_0(\mathbf{x},\mathbf{z},0)\}\right]$$

which is

$$\left[\sup_{\mathbf{z}\in\Omega_{\mathbf{Z}}}P(Y=1,D=1|\mathbf{x},\mathbf{z}) - \inf_{\mathbf{z}\in\Omega_{\mathbf{Z}}}\{P(Y=1,D=0|\mathbf{x},\mathbf{z})+P(D=1|\mathbf{x},\mathbf{z})\}\ ,\right.$$
$$\left.\inf_{\mathbf{z}\in\Omega_{\mathbf{Z}}}\{P(Y=1,D=1|\mathbf{x},\mathbf{z})+P(D=0|\mathbf{x},\mathbf{z})\} - \sup_{\mathbf{z}\in\Omega_{\mathbf{Z}}}P(Y=1,D=0|\mathbf{x},\mathbf{z})\right] \tag{13}$$

2.4 Bound in Manski and Pepper (2000)

In order to further tighten down the identified set of *ATE* given in expression (13), Manski and Pepper (2000) introduce the following assumptions:

Assumption MP 1 (Monotone Treatment Response) $Y_{1i} \geq Y_{0i}$ for each individual i.

Assumption MP 2 (Monotone Treatment Selection) $\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x},D=d]$ and $\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x},D=\mathrm{d}]$ are weakly increasing in the realized treatment d.

Assumption MP 3 (Monotone Instrumental Variable) Let $\Omega_{\mathbf{Z}}$ be an ordered set, $\mathbf{z}_1\in\Omega_{\mathbf{Z}}$, $\mathbf{z}_1\in\Omega_{\mathbf{Z}}$ and $\mathbf{z}_1\preceq\mathbf{z}_2$. $\mathbf{Z}$ is a monotone instrumental variable in the sense of mean-monotonicity if $\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}_1]\leq\mathbb{E}[Y_0|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}_2]$ and $\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}_1]\leq\mathbb{E}[Y_1|\mathbf{X}=\mathbf{x},\mathbf{Z}=\mathbf{z}_2]$ for all $\mathbf{x}$.

The Monotone Treatment Response (MTR) assumption means that, *ceteris paribus*, the response varies monotonically with treatment. Under the MTR assumption, Manski shows the sharp bounds on $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$

$$\mathbb{E}[Y|\mathbf{X}]\leq\mathbb{E}[Y_1|\mathbf{X}]\leq\mathbb{E}[Y|\mathbf{X},D=1]P(D=1|\mathbf{X})+P(D=0|\mathbf{X})$$
$$\mathbb{E}[Y|\mathbf{X},D=0]P(D=0|\mathbf{X})\leq\mathbb{E}[Y_1|\mathbf{X}]\leq\mathbb{E}[Y|\mathbf{X}]$$

The Monotone Treatment Selection (MTS) assumption implies that people with treatment have weakly higher mean response function than people without treatment. Under the MTR and MTS assumptions, the bounds on $\mathbb{E}[Y_1]$ and $\mathbb{E}[Y_0]$ are

$$\begin{aligned}\mathbb{E}[Y|\mathbf{X}]\leq\mathbb{E}[Y_1|\mathbf{X}]\leq\mathbb{E}[Y|\mathbf{X},D=1]\\ \mathbb{E}[Y|\mathbf{X},D=0]\leq\mathbb{E}[Y_0|\mathbf{X}]\leq\mathbb{E}[Y|\mathbf{X}]\end{aligned} \tag{14}$$

From above, MTR and MTS assumptions reduce the treatment effect bound to $[0,\mathbb{E}(Y|\mathbf{X},D=1)-\mathbb{E}(Y|\mathbf{X},D=0)]$, where $\mathbb{E}[Y|\mathbf{X},D=1]-\mathbb{E}[Y|\mathbf{X},D=0]$ is the Ordinary Least Square estimator.

The MTS Assumption MP 2 is a special case of the Monotone Instrumental Variable (MIV) Assumption MP 3 where the IV coincides the treatment. If we only have one instrumental variable Z, the MIV assumption can easily be translated to bound estimator; however, it becomes more complicated when we have a vector of instrumental variables **Z**. We can slightly modify the MIV assumption from conditioning on instrumental variable Z to conditioning on propensity score $P(D = 1|X, Z)$ and derive the expression of the bound estimator accordingly.

Together with the MIV assumption, the bounds on potential outcomes Y_1 and Y_0 are

$$\begin{aligned} \sum_{\mathbf{z}_1 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_1) \left\{ \sup_{\mathbf{z} \preceq \mathbf{z}_1} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}] \right\} \ \leq \ \mathbb{E}[Y_1|\mathbf{X} = \mathbf{x}] \ \leq \\ \sum_{\mathbf{z}_2 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_2) \left\{ \inf_{\mathbf{z} \succeq \mathbf{z}_2} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, D = 1, \mathbf{Z} = \mathbf{z}] \right\} \\ \sum_{\mathbf{z}_1 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_1) \left\{ \sup_{\mathbf{z} \preceq \mathbf{z}_1} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, D = 0, \mathbf{Z} = \mathbf{z}] \right\} \ \leq \ \mathbb{E}[Y_0|\mathbf{X} = \mathbf{x}] \ \leq \\ \sum_{\mathbf{z}_2 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_2) \left\{ \inf_{\mathbf{z} \succeq \mathbf{z}_2} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}] \right\} \end{aligned} \tag{15}$$

As a result, the *ATE* is bounded by

$$\begin{aligned} \Bigg[\sum_{\mathbf{z}_1 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_1) \left\{ \sup_{\mathbf{z} \preceq \mathbf{z}_1} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}] \right\} - \sum_{\mathbf{z}_2 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_2) \left\{ \inf_{\mathbf{z} \succeq \mathbf{z}_2} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, \mathbf{Z} = \mathbf{z}] \right\}, \\ \sum_{\mathbf{z}_2 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_2) \left\{ \inf_{\mathbf{z} \succeq \mathbf{z}_2} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, D = 1, \mathbf{Z} = \mathbf{z}] \right\} - \sum_{\mathbf{z}_1 \in \Omega_{\mathbf{Z}}} \Pr(\mathbf{Z} = \mathbf{z}_1) \left\{ \sup_{\mathbf{z} \preceq \mathbf{z}_1} \mathbb{E}[Y|\mathbf{X} = \mathbf{x}, D = 0, \mathbf{Z} = \mathbf{z}] \right\} \Bigg] \end{aligned} \tag{16}$$

By comparing (14) and (15) with (9) and (10), we can see that bounds under the MTS, MTR, and MIV assumptions are much narrower than Manski (1990).

3 Background on PHI and health service utilisation

3.1 PHI and Australian health system

Australia has a mixed public and private financing and provision of health care. The national health care system, Medicare, provides universal access to subsidised or free out-of-hospital medical services, free public hospital medical services, and subsidised pharmaceuticals. It is supplemented by a sizeable private health insurance sector. As in many developed countries, the rising cost of health services and the growing demands placed upon the health system driven by ageing population and technology advances are of great concern to the government.

The major instrument adopted to divert demand from the public to the private sector was a set of policy initiatives aimed at increasing the participation in private health insurance. In attempt to improve private health insurance coverage, the commonwealth government introduced several policies to encourage its uptake. The taxation system encourages middle to high income earners to take out private health insurance. While most taxpayers pay a 1.5 % Medicare Levy, an additional Medicare Levy surcharge of between 1 % and 1.5 % is payable by those taxpayers who earn more than 88000 and do not have private insurance. The Commonwealth also provides a rebate to help people meet the cost of purchasing PHI. The size of the rebate varies with age and is currently means-tested. The Lifetime Health Cover Loading is a regulatory stick designed to discourage people from delaying purchase of PHI by allowing funds to vary the premiums of individuals above 30 according to the age of entry into the fund. Lifetime Health Cover (LHC) is a Government initiative designed to encourage people to take out hospital insurance earlier in life and to maintain their cover. These policies take some of the demand away from the public health system by encouraging those who can afford it to use the private hospital system. The policy objectives of these changes are to increase the share of private health insurance, thereby lessening the burden on the public hospital system.

The inter-relationship between private health insurance cover and health service utilisation and expenditure is complex. Under Medicare, all Australian residents are entitled to free public hospital treatment, regardless of whether they have private health insurance. Private health insurance covers treatment in a private hospital or as a private patient in a public hospital, allowing choice of doctor and potentially shorter waiting times for some procedures. There are generally significant out-of-pocket costs associated with private treatment, creating disincentives to use private treatment even for those with private health insurance. Tax penalties for higher income individuals without private health insurance and a non-means tested 30 % rebate on private health insurance premium mean that for high income individuals the net premium can be negative. Thus, private health insurance purchase may not be related to expectations of use of hospital services, particularly private hospital services. It will be informative to policy makers to study the effect of private health insurance status and health services utilisation.

3.2 PHI and dental service utilisation

The Australian health system has evolved into a complex mix of private/public service provision and funding, involving the three tiers of government and the private sector. The system receives public funding, but the costs are also borne in part by private individuals supported via the private health insurance sector.

Unlike other health services, dental health services in Australia have not been generally covered by the Medicare system that provides universal coverage for other medical services. Oral diseases, unlike medical diseases, are largely predictable and as such do not have the essential characteristics of an insurable risk. On the other hand, A universal dental care scheme is simply not affordable because of the degree of unmet need for dental care in the community. Medicare is already under severe financial strain and the addition of a comprehensive universal dental scheme would simply lead to total collapse unless significant increases in the Medicare Levy were to be introduced.

Consequently, individual Australians usually finance their own dental health services. The vast majority of expenditure on dental services in Australia is therefore borne by individuals, while government funding for dental services remains low in comparison. According to the 2010–11 Australian Institute of Health and Welfare report, in 2010–11 the largest source of funds for dental expenditure was individuals, paying directly out-of-pocket for 4564 million on dental services (58.1 % of total dental costs). Health insurance funds provided a further 14.3 %. Australian government premium rebates accounted for 6.7 %, other government contributed 20.5 % of total expenditure (11.6 % Australian government direct outlay and 8.9 % from state and local governments). The other 0.4 % is provided by other insurance such as Compulsory Third Party.

Although Australians primarily bear the cost of their own dental health services, it is becoming increasingly accepted that there is a public benefit from some level of public funding to support them in meeting these costs. Poor dental health is associated with a range of serious health conditions such as poor nutrition, cardiovascular disease, stroke and diabetes that can place other burdens on the health system (Buhlin et al., 2002; Joshipura et al., 1996). In addition, some argue that treating the funding of basic dental services differently to other medical services is contrary to the view, expressed by the World Health Organization, that oral health is integral to overall health and an important part of primary health care.

In Australia, individuals who hold private health insurance (PHI) that covers dental services may enjoy subsidized dental treatments and the benefit of much shorter waiting times than those in the public system. According to the 2010–11 Australian Institute of Health and Welfare report, in 2010, over half (53.8 %) of all people over the age of 5 reported having some level of dental insurance. Adults aged 45 to 64 had significantly higher rates of dental insurance and those aged 65 and over had significantly lower rates of insurance than other age groups. Approximately three-quarters (76.3 %) of dentate adults in the $100000 and over per year annual household income group had some level of dental insurance, compared with less than a third of adults in the bottom three income groups (from 27.3 % in the $12000–20000 income group to 29.8 % in the $20000–30000 income group).

However, access to affordable private dental care remains elusive for many. Private dental fees have increased at rates substantially higher than the Consumer Price Index and other health services. In 2010–11, more than one-quarter of people aged 5 or older (28.2 %) avoided or delayed visiting a dentist due to cost. This ranged from almost 14 % of children aged 5–14 to 37 % for adults aged 25–44. Nearly one-fifth (18.3 %) of people aged 5 and over indicated that cost prevented them from receiving recommended dental treatment. Long waiting lists for public dental services and other emerging evidence showing that poorer dental health is associated with lower socioeconomic status has led to calls for increasing Commonwealth funding for dental services. Thus, there is a need for empirical studies on the impact of PHI, socioeconomic factors and other individual factors on dental services utilisation to inform policies on national dental care and funding.

A significant amount of previous research examines the relationship between private health insurance and utilisation of hospital services or general health services (Savage and Wright, 2003; Srivastava and Zhao, 2008; Cheng and Vahid, 2010; Cameron et al., 1988; Cutler and Reber, 1998; Hofter, 2006), and the endogeneity of PHI in general health care utilisation has been addressed in the literature. However, only a few studies investigate the impact of ancillary coverage on dental care utilisation in Australia (Spencer, 2004; Brennan et al., 2008; Luzzi and Spencer, 2009; Hopkins et al., 2013). The endogeneity issue of PHI has not been fully investigated in the case of dental service utilisation. In this paper, we apply the partial identification to analyze the average treatment effect of private health insurance status on three types of health service utilisation.

4 Data

The data for this study are from the Australian National Health Survey (NHS) 2004/5. It was conducted by the Australian Bureau of Statistics (ABS) with a representative sample of 19501 private dwellings across Australia. Within each sampled household a random sub-sample of usual residents was selected for inclusion in the survey comprising one adult (18 years of age and over) and one child (under age of 18 years). A total of 25906 respondent records (19501 adult records and 6405 child records) are included in the survey.

The main objectives of the NHS surveys are to obtain information on a range of health-related issues in Australia and to monitor trends in health over time. The survey collected information on the health status of the population, including long term medical conditions; health-related behaviors, such as smoking, exercise and alcohol consumption; use of health services such as consultations with doctors and dentists, and hospital visits; private health insurance coverage; and demographic and socioeconomic characteristics. Respondents are asked whether they are covered by private health insurance, and if so, what type of cover they possess — ancillary cover only, hospital cover only, both ancillary and hospital cover, or none. Since our measure of health care utilisation is dental, the relevant insurance measure is ancillary cover. Accordingly, we classify status of PHI for all those individuals as having private ancillary insurance if they responded as having either private ancillary insurance only or having both private ancillary and hospital cover. Those who claim to have only hospital cover, or no private insurance at all, are classified as not having private ancillary insurance. When respondents were unsure of their private insurance status, the corresponding values were classified as missing.

In our study, we only use the data for the 19501 adults. After observations with missing information are deleted, the remaining sample consists of 17187 observations (7970 males and 9217 females). Other explanatory variables include socio-demographic variables such as age, income, marital status, education, and lifestyle variables such as smoking, drinking, and exercising, etc. Table 1 presents basic descriptive statistics for this sample, weighted by the person weights provided in the survey.

According to different time since last consulted a dentist or dental professional, individual's dental services utilisation is classified into two categories according to dental visits during last 12 months. See Table 1 for detailed summary statistics. $Y_i = 1$ stands for individual i had dental visits during last 12 months. D_i represents the status of PHI, which is an endogenous variable. $D_i = 1$ means that individual i has PHI and $D_i = 0$ otherwise. Table 1 reveals that nearly 44 % of the sample had private hospital insurance, and about 41 % people visited a dental professional in the last 12 months.

Table 2 presents some selected sample proportions of dental services utilisation conditional on private health insurance status and demographic and socioeconomic factors. These proportions give some general information about the effect of individual factors on the decision regarding dental service utilisation.

Based on NHS 2004/5, 43.7 % of Australian adults used dental services in the last year, 40.9 % had ancillary type PHI. Individuals most likely to use dental services are those who have ancillary type insurance. Among them, 43 % did not use dental service in the last year. This proportion is 13.3 % lower than that for the whole sample, which 56.3 % of individuals use dental services in the last 12 months.

Moreover, it is shown that 53.4 % of females had never visited dentists or any other dental professionals for the last year, while more than 59.6 % of males had not had any dental service in the last year or had never used a dental service. It is also found that married persons are more likely to use dental services compared to those who are not married. Country of birth also has an impact on dental services utilisation decisions. Individuals born outside Australia are more likely to utilize dental services. Persons with higher education degrees use dental services more than others. For example, only about 42 % of those with a higher degree had not had any dental service in the last year, while about 55 %, 54 % and 65 %, respectively, of those with diploma, year 12, and less than year 12 education were in this category.

It is worthwhile to note that the observed sample proportions presented above can only indicate how frequencies of dental services utilisation vary with different types of PHI, as well as cross different socioeconomic and demographic groups, but cannot isolate the effects of individual attributes which are always correlated. For instance, individuals with a higher degree are more likely to utilize dental services. However, people with higher education level are also more likely to have higher income, which makes them more likely to have PHI. In such cases, the high observed probabilities of visiting dentists more frequently for those with higher degrees may not necessarily represent the impact of higher education on dental service utilisation. To be informative to policy changes on private health insurance, we need further investigation and data analysis.

5 Bound Estimation Methods

The different bounds described in Section 2 rely on the imposition of different constraints, but they are all characterised as functions of various conditional probabilities that are used to derive the bounds. The estimation of the *ATE* bounds obviously, therefore, turns on the estimation of those conditional probabilities and we will compare the performance of four approaches to estimating the conditional probabilities

Table 1. Definitions and summary statistics for all variables

Variable	Definition	Mean	Std Dev
D	1 if have ancillary type private health insurance, 0 otherwise	.409	.492
Y	1 if visited a dental professional within last 12 month, 0 otherwise	.437	.496
AGE18	1 if aged from 18 to 19, 0 otherwise	.021	.143
AGE24	1 if aged from 20 to 24, 0 otherwise	.070	.255
AGE29	1 if aged from 25 to 29, 0 otherwise	.079	.270
AGE34	1 if aged from 30 to 34, 0 otherwise	.101	.301
AGE39	1 if aged from 35 to 39, 0 otherwise	.104	.306
AGE44	1 if aged from 40 to 44, 0 otherwise	.106	.308
AGE49	1 if aged from 45 to 49, 0 otherwise	.094	.292
AGE54	1 if aged from 50 to 54, 0 otherwise	.082	.274
AGE59	1 if aged from 55 to 59, 0 otherwise	.078	.268
AGE64	1 if aged from 60 to 64, 0 otherwise	.067	.251
AGE69	1 if aged from 65 to 69, 0 otherwise	.055	.228
AGE74	1 if aged from 70 to 74, 0 otherwise	.053	.224
AGE75	1 if aged 75 or more (reference group)	.089	.285
MALE	1 if male, 0 otherwise	.464	.499
MARRIED	1 if married, 0 otherwise	.504	.500
AUS	1 if born in Australia, 0 therwise	.743	.437
ENGCOUN	1 if born in major speaking-English countries, 0 otherwise	.120	.325
OTHERCOU	1 if born in other places (reference group)	.137	.344
HIGHDEG	1 if have higher education level, 0 other wise	.177	.381
DIPLOMA	1 if have diploma, 0 otherwise	.350	.477
YEAR12	1 if have year 12 education, 0 otherwise	.129	.335
LESSYR12	1 if not finish year 12 education (reference group)	.344	.475
INCDECH1	1 if income falls into the 1st decile, 0 otherwise (reference group)	.065	.246
INCDECH2	1 if income falls into the 2nd decile, 0 otherwise	.085	.279
INCDECH3	1 if income falls into the 3rd decile, 0 otherwise	.135	.342
INCDECH4	1 if income falls into the 4th decile, 0 otherwise	.111	.314
INCDECH5	1 if income falls into the 5th decile, 0 otherwise	.105	.306
INCDECH6	1 if income falls into the 6th decile, 0 otherwise	.099	.298
INCDECH7	1 if income falls into the 7th decile, 0 otherwise	.083	.276
INCDECH8	1 if income falls into the 8th decile, 0 otherwise	.100	.300
INCDECH9	1 if income falls into the 9th decile, 0 otherwise	.121	.326
INCDECH10	1 if income falls into the 10th decile, 0 otherwise	.097	.296
EXCELLENT	1 if excellent self-rated health, 0 otherwise	.185	.388
VERYGOOD	1 if verygood self-rated health, 0 otherwise	.346	.476
GOOD	1 if good self-rated health, 0 otherwise	.287	.452
FAIR	1 if fair self-rated health, 0 otherwise	.130	.336
DUMCSM	1 if smoke currently, 0 otherwise	.243	.429
DUMEXREG	1 if smoke regularly before, 0 otherwise	.315	.465
NOSM	1 if never smoke regularly (reference group)	.442	.497
LTC	1 if have long term condition, 0 otherwise	.894	.309
HIGHA	1 if high risk drinking, 0 other wise	.057	.232
MEDA	1 if medium risk drinking, 0 other wise	.080	.272
LOWA	1 if low risk drinking, 0 other wise	.489	.500
NOLWA	1 if no alcohol last week, 0 other wise	.297	.457
NEVERA	1 if never drinking (reference group)	.077	.267
EXH	1 if have high level exercise, 0 otherwise	.055	.228
EXM	1 if have medium level exercise, 0 otherwise	.239	.427
EXL	1 if have low level exercise, 0 otherwise	.366	.482
EXN	1 if no exercise (reference group)	.340	.474
BED	Beds available per 1000 population in public hospitals, at state level	2.864	.842
COPAY	Average gap payment for in-hospital medical services	18.095	9.913
HPRICE	Hospital type private insurance price	10.013	7.566
GPRICE	Ancillary type private insurance price	7.014	1.890

Table 2. Observed probabilities of dental services utilisation conditional on PHI status and other individual factors (%)

Individual factors		**Last dental visit** less than 1 year	more than 1 year
PHI status	ancillary insurance	57.00	43.00
	no ancillary insurance	34.47	65.53
Gender	male	40.35	59.65
	female	46.56	53.44
Marital status	married	46.15	53.85
	single	41.17	58.83
Country of birth	Australia	42.62	57.38
	main English speaking country	48.67	51.33
	other	45.05	54.95
Education	high degree	57.53	42.47
	diploma	44.66	55.34
	year 12	46.12	53.88
	less than year 12	34.66	65.34

Source: NHS 2004/5

and hence the *ATE* and its bounds:

1. Parametric model estimation via the commonly employed recursive bivariate probit (RBVP) model. The RBVP model is specified by the equations

$$\begin{cases} \mathrm{d}^* = \mathbf{x}'\boldsymbol{\beta}_\mathrm{d} + \mathbf{z}'\,\boldsymbol{\delta} + \eta, & \mathrm{d} = 1 \text{ if } \mathrm{d}^* > 0, 0 \text{ otherwise;} \\ \mathrm{y}^* = \mathbf{x}'\boldsymbol{\beta}_\mathrm{y} + \mathrm{d}\,\alpha + \varepsilon, & \mathrm{y} = 1 \text{ if } \mathrm{y}^* > 0, 0 \text{ otherwise,} \end{cases}$$

where $\mathbf{X}$ is a vector of exogenous variables and $\mathbf{Z}$ is a vector of instrumental variables satisfying $(\varepsilon, \eta) \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}$ were ε and η are standard normal random variables with correlation ρ. This is a linear index threshold crossing model with both the discrete outcome variable and the endogenous dummy variable being determined by a threshold equation and linear index distribution. In order to link the RBVP model to the structural model let $U = \Phi(\varepsilon)$ where $\Phi(\cdot)$ denotes the standard normal CDF. Then $U \in (0, 1)$ and the marginal distribution of U is uniform. Denoting the indicator function by $\chi[\cdot]$, it follows that the structural function

$$h(\mathrm{d}, \mathbf{x}, \mathrm{u}) = \chi[\mathrm{u} \geq \Phi(-\mathbf{x}'\boldsymbol{\beta}_\mathrm{y} - \mathrm{d}\,\alpha)]$$

is such that h is weakly monotonic in U and $P(U \leq \tau|\mathbf{Z} = \mathbf{z}) = \tau$ for all $\tau \in (0, 1)$ and all $\mathbf{z} \in \Omega_\mathbf{Z}$ since the independence assumption implies that $U \perp\!\!\!\perp \mathbf{Z}|\mathbf{X}$. The assumptions of the structural equation model are therefore satisfied and

$$\mathrm{y} = h(\mathrm{d}, \mathbf{x}, \mathrm{u}) = \begin{cases} 0, & \text{if } 0 < \mathrm{u} \leq p(\mathrm{d}, \mathbf{x}) \\ 1, & \text{if } p(\mathrm{d}, \mathbf{x}) < \mathrm{u} \leq 1\,, \end{cases} \tag{17}$$

where the threshold function is given by $p(\mathrm{d}, \mathbf{x}) = \Phi(\mathbf{x}'\boldsymbol{\beta}_\mathrm{y} + \mathrm{d}\,\alpha)$. Thus the RBVP model corresponds to a structural equation model as in (1) augmented with parametric assumptions. Similarly, it is not difficult to show that the RBVP model represents a special case of the weakly separable model in (6).

For the purposes of this paper we view the RBVP model as representing a convenient approximation to the true data generating process that is used as a means of generating a quasi (Gaussian) maximum likelihood estimator (QMLE).

2. Semi-parametric model estimation using local likelihood techniques for fitting generalized linear models and regression equations to evaluate what we will term a local QMLE (See Tibshirani and Hastie, 1987; Fan et al., 1998; Kauermann and Opsomer, 2003). This is a natural adaptation of the QMLE given that we do not believe that the RBVP model represents the truth but is being used as an approximation.

3. Naive estimates obtained by using observed relative frequencies and empirical distributions to calculate sample analogues of the required conditional probabilities and expectations.

4. Non-parametric smoothing estimates derived from adaptations of kernel smoothing methods for mixed data type, as studied in Li and Racine (2003), see also Li and Racine (2011). This estimator is motivated by the fact, illustrated below, that naive estimates are often not feasible because the data is not sufficiently rich and there are not "enough" observations across all possible subcategories.

6 Results

Table 3 presents the QMLE coefficient estimates of all the parameters in RBVP model. The highly significant correlation coefficient between dental services utilization and PHI provides evidence for the endogeneity of PHI status, and in a conventional analysis would justify the joint estimation of dental service utilization and PHI status via the RBVP model. Although the magnitudes of the regressor coefficients are not directly interpretable in terms of the marginal effects of the associated variables, they do indicate the direction of the impact on the latent variables and thus the effect of a unit change in individual explanatory factors on the probabilities of dental services utilization and PHI status. The results in Table 3 show that the direction of the impact of each of the exogenous variables on dental service utilization is the same as would be expected, and, in particular, the coefficient of PHI status in the equation for dental services utilization is positive and significant, indicating that having PHI directly increases the latent variable of dental services utilization, making the individual more likely to access dental services.

To construct QMLE estimates of the Chesher bound for the *ATE* (Chesher, 2010) we use the QMLE parameter estimates to derive the conditional probabilities in (2), and hence the bounds in (4) and (5). Similarly, estimates of the Shaikh and Vytlacil (2011), Manski (1990) and Manski and Pepper (2000) bounds can be calculated by evaluating expressions (7), (13) and (16), respectively, using the QMLE estimates.

Table 3. Coefficient estimates of RBVP model: dental services utilisation

	$P(Y=1)$				$P(D=1)$		
Dental	COEF	STD.E	T-STAT	PHI	COEF	STD.E	T-STAT
CONST	−.666	.070	−9.48	CONST	−.715	.099	−7.23
AGE18	.078	.078	1.00	AGE18	.402	.110	3.67
AGE24	.078	.053	1.46	AGE24	.038	.078	.49
AGE29	.134	.053	2.55	AGE29	−.282	.066	−4.25
AGE34	.060	.050	1.21	AGE34	−.184	.063	−2.92
AGE39	.099	.048	2.05	AGE39	−.051	.061	−.85
AGE44	.139	.048	2.93	AGE44	.022	.060	.36
AGE49	.180	.049	3.69	AGE49	.042	.058	.72
AGE54	.135	.051	2.66	AGE54	.175	.056	3.13
AGE59	.135	.052	2.61	AGE59	.268	.054	4.92
AGE64	.084	.052	1.60	AGE64	.262	.055	4.76
AGE69	.104	.055	1.90	AGE69	.253	.057	4.41
AGE74	.012	.054	.23	AGE74	.236	.058	4.06
MALE	−.153	.023	−6.58	MALE	−.139	.029	−4.80
MARRIED	−.107	.024	−4.53	MARRIED	.344	.023	15.30
AUS	−.170	.030	−5.75	AUS	.224	.032	7.05
ENGCOUN	−.020	.038	−.52	ENGCOUN	.086	.042	2.07
HIGHDEG	.190	.038	4.98	HIGHDEG	.369	.034	10.72
DIPLOMA	.109	.026	4.18	DIPLOMA	.168	.027	6.35
YEAR12	.134	.036	3.74	YEAR12	.282	.036	7.85
INCDECH2	.107	.054	1.98	INCDECH2	−.542	.055	−9.87
INCDECH3	.046	.051	.90	INCDECH3	−.502	.052	−9.65
INCDECH4	.095	.049	1.95	INCDECH4	−.258	.051	−5.07
INCDECH5	.038	.048	.79	INCDECH5	−.011	.050	−.22
INCDECH6	.037	.049	.77	INCDECH6	.021	.051	.41
INCDECH7	.021	.051	.41	INCDECH7	.123	.053	2.29
INCDECH8	−.036	.050	−.72	INCDECH8	.272	.052	5.25
INCDECH9	−.035	.052	−.67	INCDECH9	.504	.051	9.85
INCDECH10	−.109	.060	−1.83	INCDECH10	.788	.055	14.24
EXCELLEN	.043	.035	1.22	EXCELLEN	.175	.037	4.67
VERYGOOD	.001	.030	.03	VERYGOOD	.151	.033	4.64
GOOD	−.008	.030	−.28	GOOD	.125	.033	3.85
DUMCSM	.040	.029	1.38	DUMCSM	−.398	.029	−13.92
DUMEXREG	.054	.023	2.38	DUMEXREG	−.127	.025	−5.15
LTC	−.015	.034	−.45	LTC	.193	.036	5.40
ANCPHI	1.325	.089	14.90				
				HIGHA	.057	.058	1.00
				MEDA	.234	.052	4.51
ATE:				LOWA	.180	.041	4.43
Point Estimate	.4839			NOLWA	.056	.041	1.37
95 % CI	[.4147, .5382]			EXH	.307	.046	6.63
				EXM	.182	.027	6.70
				EXL	.139	.024	5.80
				BED	−.016	.012	−1.29
				COPAY	−.008	.001	−7.70
				HPRICE	−.007	.003	−2.73
				GPRICE	−.030	.008	−3.99
				RHO	−.571	.004	−151.89

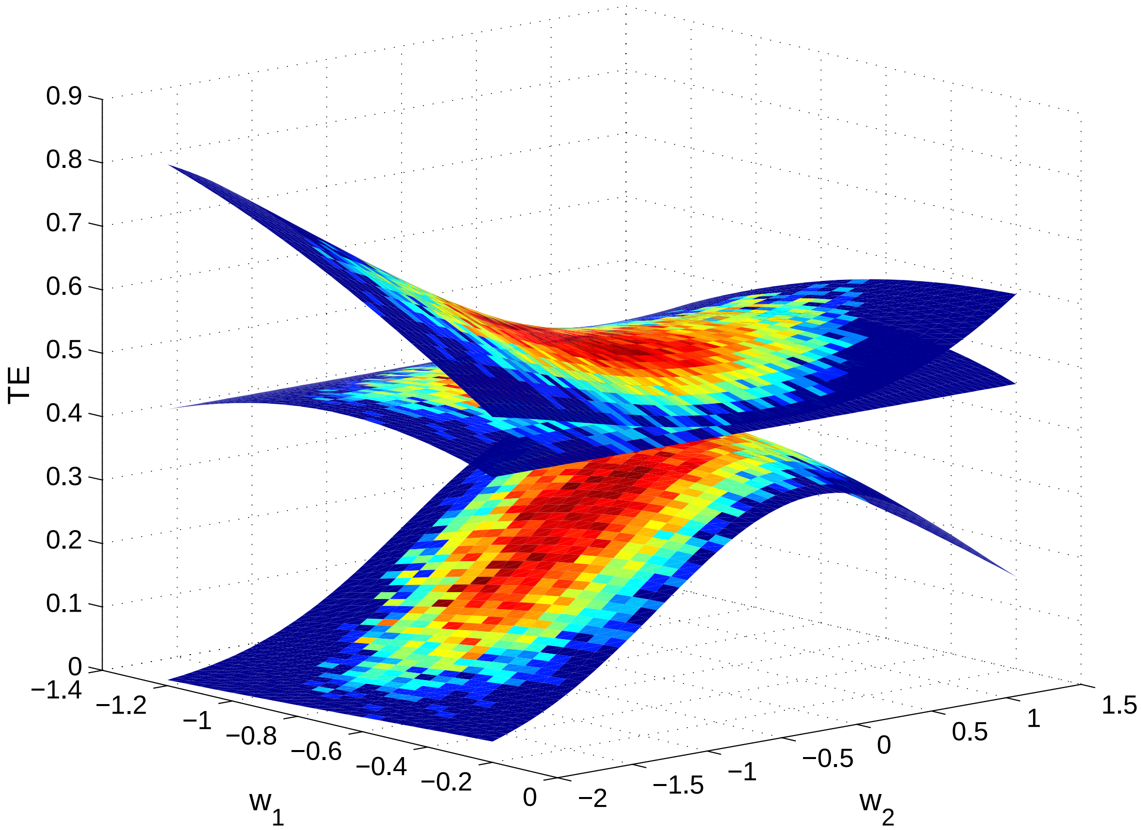
Table 4. QMLE estimates of four *ATE* bounds (weighted average over the sample).

	Manski	Manski and Pepper	Chesher	Shaikh and Vytlacil
ATE	$[-.3382, .5380]$	$[0, .1291]$	$[.2488, .5380]$	$[.4630, .5463]$

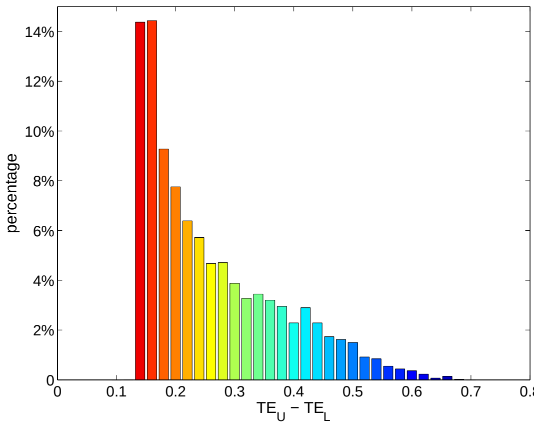
We first present the weighted averaged *ATE* bounds over all types of individuals in the sample. The four bounds, so calculated, are given in Table 4. The Manski, and Manski and Pepper bounds appear not to be particularly useful in this application, the first being too wide and the second incorrectly placed. The Chesher, and Shaikh and Vytlacil bounds are appropriately located and both are reasonably short, although the Shaikh and Vytlacil bound is narrower than the Chesher bound due to the imposition of additional constraints such as "perfect matching". Since we wish to examine the practical implication of the results on the partial identification of the *ATE* under as weak a set of assumptions as possible, in what follows we will focus on the Chesher bounds.

Next, we examine the *ATE* bounds for the heterogenous population types, as the bounds can be very different for different types of the sub-populations. Figure 2 graphs the QMLE point estimates of the treatment effect (*TE*) and the Chesher bounds as a function of the linear indices $w_1 = \beta_y' \mathbf{x}$ and $w_2 = \beta_d' \mathbf{x}$. The surface colours in Figure 2a indicate the relative frequency of occurrence of those $\mathbf{x}$ values in the sample that generate a given linear index (w_1, w_2) coordinate pair, the colour code corresponding to that used in Figure 2b, which depicts the bound interval length distribution, and Figure 2c, which plots the bound length as a function of (w_1, w_2). From Figure 2a we can see that the *TE* and bound estimates are always positive, although the actual values vary substantially with the changes in w_1 and w_2 induced by the variations in $\mathbf{X}$. Figure 2b and Figure 2c indicate that the shorter treatment effect bounds are associated with $\mathbf{x}$ values that generate high frequency (w_1, w_2) coordinate pairs, with more than 70% of individual treatment effect bounds being shorter than 0.3, and less than 15% of bounds being longer than 0.4.

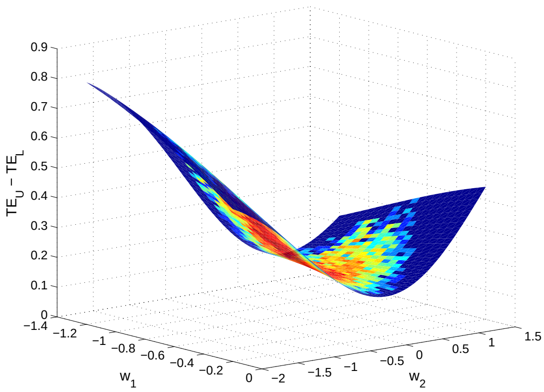
Another way of looking at the heterogenous *ATE* bounds is to focus on the effect of individual regressors. The Chesher bounds for the average treatment effect for the whole sample and the different sub-samples based on individual X variables are shown in Table 5. The first feature of note is that all the average treatment effects are positive ($p(0, \mathbf{x}) > p(1, \mathbf{x})$) for the whole sample and all the different sub-samples of interest. Nevertheless, there are some interesting contrasts between different sub-populations. The widest *ATE* bound among the different age bands is the one for those falling in the above 75 age group, $[.1719, .5808]$. It is much wider than $[.2791, .5221]$, the bound for middle aged people between 55 and 59. People in the 10th income decile have an *ATE* bound of $[.3133, .5055]$, while that for the 3rd income decile it is $[.1541, .5866]$, more than twice as wide as the *ATE* bound for the 10th income decile. Note that the *ATE* bound for the 10th income decile is the narrowest among all sub-populations, and that for the 3rd income decile is the widest. The *ATE* bound for people with higher degrees is $[.3058, .5105]$, which is



(a) Estimated (TE), and upper and lower bounds.



(b) Interval length distribution.



(c) Interval length surface.

Figure 2. QMLE treatment effect (TE), and upper and lower bounds, for each $\mathbf{X}$.

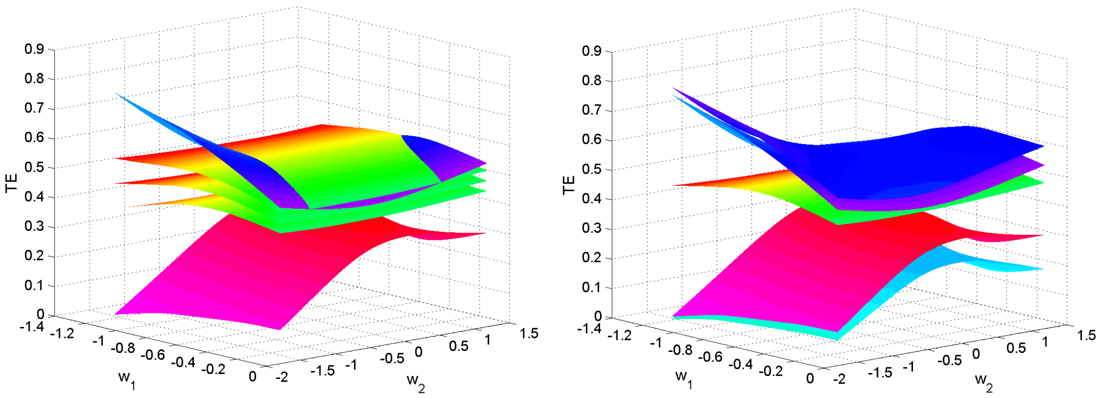
about half the width of the *ATE* bound for people who did not finish year 12 [.1972, .5658].

Table 5. QMLE *ATE* point and bound estimates: dental service utilisation

X	Sub-sample %	*ATE*	95% CI for *ATE*	*ATE* Bound	95% CI for *ATE* Bound
ALL	100	.4834	[.4147, .5382]	[.2488, .5380]	[.1899, .5690]
AGE18	2.1	.4853	[.4179, .5415]	[.2722, .5289]	[.2067, .5657]
AGE24	7.0	.4848	[.4159, .5403]	[.2363, .5436]	[.1775, .5725]
AGE29	7.9	.4843	[.4171, .5369]	[.2265, .5463]	[.1774, .5753]
AGE34	10.0	.4827	[.4140, .5386]	[.2443, .5413]	[.1870, .5719]
AGE39	10.4	.4832	[.4140, .5378]	[.2621, .5317]	[.2000, .5647]
AGE44	10.6	.4834	[.4161, .5378]	[.2704, .5269]	[.2045, .5611]
AGE49	9.4	.4836	[.4167, .5381]	[.2730, .5248]	[.2046, .5583]
AGE54	8.2	.4835	[.4165, .5379]	[.2782, .5235]	[.2057, .5584]
AGE59	7.8	.4833	[.4150, .5372]	[.2791, .5221]	[.2078, .5580]
AGE64	6.7	.4832	[.4156, .5382]	[.2558, .5341]	[.1978, .5675]
AGE69	5.5	.4842	[.4167, .5389]	[.2425, .5385]	[.1882, .5693]
AGE74	5.3	.4822	[.4123, .5393]	[.2218, .5516]	[.1750, .5823]
AGE75	8.9	.4822	[.4115, .5387]	[.1719, .5808]	[.1375, .6016]
INCDECH1	6.5	.4841	[.4150, .5396]	[.2781, .5259]	[.2160, .5601]
INCDECH2	8.5	.4838	[.4160, .5393]	[.1741, .5714]	[.1378, .5918]
INCDECH3	13.5	.4836	[.4116, .5397]	[.1541, .5866]	[.1247, .6058]
INCDECH4	11.1	.4840	[.4170, .5381]	[.2136, .5488]	[.1714, .5764]
INCDECH5	10.5	.4843	[.4165, .5392]	[.2565, .5330]	[.1997, .5667]
INCDECH6	9.9	.4841	[.4152, .5395]	[.2555, .5337]	[.1986, .5677]
INCDECH7	8.3	.4842	[.4163, .5398]	[.2666, .5299]	[.2073, .5619]
INCDECH8	10.0	.4830	[.4143, .5389]	[.2915, .5198]	[.2225, .5566]
INCDECH9	12.1	.4829	[.4141, .5380]	[.3124, .5112]	[.2267, .5507]
INCDECH10	9.7	.4802	[.4134, .5319]	[.3133, .5055]	[.2154, .5425]
MALE	46.4	.4819	[.4124, .5364]	[.2448, .5425]	[.1864, .5719]
FEMALE	53.6	.4847	[.4167, .5397]	[.2522, .5341]	[.1929, .5651]
MARRIED	50.4	.4824	[.4135, .5369]	[.2764, .5258]	[.2079, .5602]
UNMARRIE	49.6	.4844	[.4158, .5395]	[.2208, .5503]	[.1706, .5777]
AUS	74.3	.4835	[.4142, .5383]	[.2505, .5394]	[.1913, .5702]
ENGCOUN	12.0	.4833	[.4172, .5381]	[.2550, .5299]	[.1954, .5618]
OTHECOUN	13.7	.4828	[.4147, .5363]	[.2340, .5372]	[.1772, .5668]
HIGHDEG	17.7	.4834	[.4156, .5373]	[.3058, .5105]	[.2174, .5494]
DIPLOMA	35.0	.4840	[.4159, .5382]	[.2633, .5293]	[.2030, .5620]
YEAR12	12.9	.4841	[.4171, .5387]	[.2690, .5250]	[.2041, .5586]
LESSYR12	34.4	.4825	[.4119, .5384]	[.1972, .5658]	[.1583, .5893]
EXCELLEN	18.5	.4839	[.4163, .5381]	[.2795, .5224]	[.2074, .5577]
VERYGOOD	34.5	.4834	[.4149, .5379]	[.2662, .5296]	[.2029, .5626]
GOOD	28.7	.4831	[.4143, .5383]	[.2451, .5396]	[.1886, .5686]
FAIR	12.9	.4833	[.4128, .5385]	[.1977, .5634]	[.1568, .5880]
DUMCSM	24.3	.4836	[.4149, .5383]	[.1932, .5640]	[.1516, .5876]
DUMEXREG	31.5	.4834	[.4147, .5382]	[.2567, .5342]	[.1968, .5650]
NOSM	44.2	.4833	[.4145, .5381]	[.2736, .5264]	[.2077, .5600]
LTC	89.4	.4834	[.4146, .5381]	[.2504, .5374]	[.1913, .5685]

To allow for sampling errors in the estimation of the bounds, we further estimate the 95 % confidence region for Chesher bounds. In order to compare the 95 % confidence interval of the point estimated ATE and the Chesher bound and its 95 % confidence region, we plot all these measures as functions of indices w_1 and w_2 in Figure 3a and Figure 3b. The blue and pink surfaces in Fig. 3a are the upper and lower bound of the ATE, while the three surfaces in the middle are the point estimate of ATE and its 95 %

confidence interval. We can see that the lower 95 % confidence interval always lie above the lower ATE bound, however, for some values of w_1 and w_2, the upper 95 % confidence interval will go above the upper bound of the ATE. Though in Fig. 3b the 95 % confidence region of the ATE bound will always contain the 95 % confidence interval of the point estimate. Any notion that conventional inferential procedures can accommodate partial identification of the *ATE* by simply allowing for parameter uncertainty in the estimation of the *ATE* is clearly not upheld.



(a) Confidence intervals for point identified ATE (b) Confidence region for the partially identified ATE

Figure 3. Confidence intervals of point identified ATE and confidence regions of partially identified ATE, and their upper and lower bounds

In light of the variation in the *ATE* bounds seen above, it is perhaps of interest to investigate how this translates into different types of individuals. Table 6 describes the *ATE* bounds for three types of individual in the sample. The first type of individual is an Australian born married female in her early 50's. She has

Table 6. *ATE* bounds for three types of individual.

Individual type	linear indices w_1	w_2	QMLE	Local QMLE	Raw nonparametric	Smoothed nonparametric
1	-0.756	0.362	[.2488, .5380]	[.2501, .5310]	N/A	[.2308, .5671]
2	-0.851	-1.589	[0.0, .7940]	[.0012, .7975]	N/A	[.0001, .8025]
3	-0.804	0.415	[.0422, .5750]	[.0649, .5802]	N/A	[.0103, .5997]

low educational status (did not finish year 12) but earns more than average (in the 7th income decile). She has never smoked, does not drink, but suffers from a long term medical condition (such as heart disease, diabetes, and so on). She is more likely to buy private health insurance than most other people but less likely to see a dentist. For such a person the width of the *ATE* bound is .28. The second type of individual is an Australian born unmarried man in his 30's. He also has low educational status, but earns less than average (in the 2nd income decile). He smokes and drinks heavily, but does not have any long term medical conditions. He is less likely to buy private health insurance and less likely to use dental services than most other people. For such an individual it is difficult to ascertain the *ATE*. Their *ATE* bound length is nearly .8, almost three times longer than that for the first type of person. The third type of individual is an Australian born married man in his 50's. He is highly educated and has an income in the top decile. He used to smoke, only drinks small amounts of alcohol and does not have any long term medical conditions. He is more likely to buy private health insurance but less likely to visit the dentist. For this third type of individual the *ATE* bound length is .53, falling roughly in the middle of that for the first and second type of individual.

Heretofore we have only discussed the (RBVP model) QMLE estimates. Table 6 also presents estimates of the *ATE* bounds obtained from the other three estimation methods outlined above. The most notable feature of Table 6 is that the raw nonparametric estimates are not available. This is because with this data set, as with much survey data, many variables of interest, even those that are essentially continuous, are measured on a discrete categorical scale (age broken down into 13 age bands for example). Consequently there are so many possible combinations of categories that cross tabulation cells required in the estimation of various conditional probabilities can be empty. For the NHS data we have an overall sample size of 17,187, but there are more than 24,960 possible cross tabulation cells and as a result there are insufficient observations to evaluate the raw nonparametric estimates for these three types of individual. Data paucity of this kind can be overcome by using the smoothed nonparametric estimator. The difference between the smoothed nonparametric estimates and the local QMLE estimates is quite small, however, and the latter are close to the (global) QMLE estimates. These relative differences between the QMLE, local QMLE and smoothed nonparametric estimates, which were also present in the analyses conducted above but were not shown here, corroborate evidence presented elsewhere (LPZ 2015a,b) that shows that the QMLE estimator is remarkably robust to model miss-specification and can indeed outperform nonparametric estimators across a broad range of data generating processes.

7 Conclusions

Policymakers in Australia have long been interested in the role of private health insurance (PHI) in relieving the burdens on its universally covered but crowded public health service system. Public health professionals are also concerned about inequality in health care access among individuals with and without private health cover. Various policies have been put in place over the past decades encouraging high income earners to buy PHI and penalising those who do not. Thus, evaluating the impact of government PHI policies and estimating the effect of PHI on the various health service utilisations has long been the focus of health economists.

We show in this paper that once due allowance has been made for the fact that the *ATE* is only partially identified, conventional analyses based on the RBVP model (say) become questionable. In particular, using standard confidence intervals to allow for parameter uncertainty will not be sufficient to encompass the lack of identification. Treating the RBVP model as an approximation to the true data generating process and using QMLE to estimate the *ATE* of PHI on dental service utilisation under a partial identification framework, we find that the resulting confidence bounds for the *ATE* are much wider than conventional confidence intervals. In this particular application the bounds based on Manski (1990) and Manski and Pepper (2000) are not useful, whilst the bounds based on Chesher (2010) and Shaikh and Vytlacil (2011) are informative.

For NHS survey data a finite sample with around 17000 observations is not enough to prevent data paucity making the use of raw nonparametric estimates unfeasible. Estimates of the *ATE* and *ATE* bounds based upon QMLE, QMLE semi-parametric smoothing or non-parametric smoothing estimators do not show substantial differences in the estimated values for a typical 'type' of individual. Estimated bounds for different types of individuals, and the width of the bounds, can be very different, however, as can the *ATE* and *ATE* bounds for different sub-samples. Our application indicates that estimating the *ATE* and *ATE* bounds with real data involves many more variables and categories, and a much higher computational cost, than the stylised settings used in many previous simulation studies. This suggests that the QMLE is a feasible choice of estimator for empirical applications with very large finite sample sizes (big data).

References

Angrist, J. D. and Pischke, J.-S. (2008). *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Bhattacharya, J., Shaikh, A. M., and Vytlacil, E. J. (2012). Treatment effect bounds: An application to swan–ganz catheterization. *Journal of Econometrics*, 168(2):223–243.

Brennan, D. S., Luzzi, L., and Roberts-Thomson, K. F. (2008). Dental service patterns among private and public adult patients in australia. *BMC Health Services Research*, 8(1):1–8.

Buhlin, K., Gustafsson, A., Håkansson, J., and Klinge, B. (2002). Oral health and cardiovascular disease in sweden. *Journal of Clinical Periodontology*, 29(3):254–259.

Cameron, A. C., Trivedi, P. K., Milne, F., and Piggott, J. (1988). A microeconometric model of the demand for health care and health insurance in australia. *Review of Economic Studies*, 55(1):85–106.

Cheng, T. C. and Vahid, F. (2010). Demand for hospital care and private health insurance in a mixed publicprivate system: empirical evidence using a simultaneous equation modeling approach. *Melbourne Institute Working Paper*, (10/25).

Chesher, A. (2005). Nonparametric identification under discrete variation. *Econometrica*, 73(5):1525–1550.

Chesher, A. (2007). Endogeneity and discrete outcomes. *CeMMAP Working Papers CWP 05/07*.

Chesher, A. (2010). Instrumental variable models for discrete outcomes. *Econometrica*, 78(2):575–601.

Cutler, D. M. and Reber, S. J. (1998). Paying for health insurance: the trade-off between competition and adverse selection. *The Quarterly Journal of Economics*, 113(2):433–466.

Duckett, S. J. and Jackson, T. J. (2000). The new health insurance rebate: an inefficient way of assisting public hospitals. *The Medical Journal of Australia*, 172(9):439–442.

Fan, J., Farmen, M., and Gijbels, I. (1998). Local maximum likelihood estimation and inference. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 60(3):591–608.

Gerfin, M. and Schellhorn, M. (2006). Nonparametric bounds on the effect of deductibles in health care insurance on doctor visits—Swiss evidence. *Health economics*, 15(9):1011–1020.

Gundersen, C. and Kreider, B. (2008). Food stamps and food insecurity: what can be learned in the presence of nonclassical measurement error? *Journal of Human Resources*, 43(2):352–382.

Gundersen, C. and Kreider, B. (2009). Bounding the effects of food insecurity on children's health outcomes. *Journal of Health Economics*, 28(5):971–983.

Gundersen, C., Kreider, B., and Pepper, J. (2011). Partially Identifying the Impact of the Supplemental Nutrition Assistance Program on Food Insecurity among Children: Addressing Endogeneity and Misreporting Using the SIPP. *Working Paper*.

Hall, J., de Abreu Lourenco, R., and Viney, R. (1999). Carrots and sticks–the fall and fall of private health insurance in australia. *Health Economics*, 8(8):653–660.

Harmon, C. and Nolan, B. (2001). Health insurance and health services utilization in ireland. *Health Economics*, 10(2):135–145.

Heckman, J. J. (1990). Varieties of selection bias. *The American Economic Review*, 80(2):313–318.

Höfter, R. H. (2006). Private health insurance and utilization of health services in chile. *Applied Economics*, 38(4):423–439.

Hofter, R. H. (2006). Private health insurance and utilization of health services in chile. *Applied Economics*, 38(4):423–439.

Holly, A., Gardiol, L., Domenighetti, G., and Bisig, B. (1998). An econometric model of health care utilization and health insurance in switzerland. *European Economic Review*, 42(3):513–522.

Hopkins, S., Kidd, M. P., and Ulker, A. (2013). Private health insurance status and utilisation of dental services in australia. *Economic Record*, 89(285):194–206.

Jones, A. M., Koolman, X., and Van Doorslaer, E. (2006). The impact of having supplementary private health insurance on the use of specialists. *Annales d'Economie et de Statistique*, pages 251–275.

Joshipura, K. J., Rimm, E. B., Douglass, C. W., Trichopoulos, D., Ascherio, A., and Willett, W. C. (1996). Poor oral health and coronary heart disease. *Journal of dental research*, 75(9):1631–1636.

Kauermann, G. and Opsomer, J. D. (2003). Local likelihood estimation in generalized additive models. *Scandinavian Journal of Statistics*, 30(2):317–337.

Kreider, B., Manski, R. J., Moeller, J., and Pepper, J. (2014). The effect of dental insurance on the use of dental care for older adults: A partial identification analysis. *Health Economics*.

Kreider, B. and Pepper, J. V. (2007). Disability and employment: reevaluating the evidence in light of reporting errors. *Journal of the American Statistical Association*, 102(478):432–441.

Kreider, B., Pepper, J. V., Gundersen, C., and Jolliffe, D. (2012). Identifying the effects of snap (food stamps) on child health outcomes when participation is endogenous and misreported. *Journal of the American Statistical Association*, 107(499):958–975.

Li, Q. and Racine, J. (2003). Nonparametric estimation of distributions with categorical and continuous data. *Journal of multivariate analysis*, 86(2):266–292.

Li, Q. and Racine, J. (2011). *Nonparametric econometrics: Theory and practice*. Princeton University Press.

Luzzi, L. and Spencer, A. J. (2009). Public dental service utilization among adults in south australia. *Australian Dental Journal*, 54(2):154–160.

Manning, W. G., Newhouse, J. P., Duan, N., Keeler, E. B., and Leibowitz, A. (1987). Health insurance and the demand for medical care: evidence from a randomized experiment. *The American Economic Review*, pages 251–277.

Manski, C. F. (1989). Anatomy of the selection problem. *The Journal of Human Resources*, 24(3):343–360.

Manski, C. F. (1990). Nonparametric bounds on treatment effects. *The American Economic Review*, 80(2):319–323.

Manski, C. F. and Pepper, J. V. (2000). Monotone instrumental variables: with an application to the returns to schooling. *Econometrica*, 68(4):997–1010.

Propper, C. (2000). The demand for private health care in the uk. *Journal of Health Economics*, 19(6):855–876.

Savage, E. and Wright, D. J. (2003). Moral hazard and adverse selection in australian private hospitals: 1989–1990. *Journal of Health Economics*, 22(3):331–359.

Shaikh, A. M. and Vytlacil, E. J. (2011). Partial identification in triangular systems of equations with binary dependent variables. *Econometrica*, 79(3):949–955.

Spencer, A. J. (2004). *Narrowing the inequality gap in oral health and dental care in Australia*. Australian Health Policy Institute at the University of Sydney.

Srivastava, P. and Zhao, X. (2008). Impact of private health insurance on the choice of public versus private hospital services. *Health, Econometrics and Data Group (HEDG) Working Papers*.

Tibshirani, R. and Hastie, T. (1987). Local likelihood estimation. *Journal of the American Statistical Association*, 82(398):559–567.

Wilcox, S. (2001). Promoting private health insurance in australia. *Health Affairs*, 20(3):152–161.