



MONASH
BUSINESS
SCHOOL

ISSN 1440-771X

Department of Econometrics and Business Statistics

<http://business.monash.edu/econometrics-and-business-statistics/research/publications>

**Nonparametric Kernel Estimation of the Impact
of Tax Policy on the Demand for Private Health
Insurance in Australia**

Xiaodong Gong and Jiti Gao

April 2017

Working Paper 07/17

Nonparametric Kernel Estimation of the Impact of Tax Policy on the Demand for Private Health Insurance in Australia

Xiaodong Gong and Jiti Gao*

April 2017

Abstract

This paper is motivated by our attempt to answer an empirical question: how is private health insurance take-up in Australia affected by the income threshold at which the Medicare Levy Surcharge (MLS) kicks in? We propose a new difference de-convolution kernel estimator for the location and size of regression discontinuities. We also propose a bootstrapping procedure for estimating confidence bands for the estimated discontinuity. Performance of the estimator is evaluated by Monte Carlo simulations before it is applied to estimating the effect of the income threshold of Medicare Levy Surcharge on the take-up of private health insurance in Australia using contaminated data.

JEL Classification: C13; C14; C29; I13

Keywords: De-convolution kernel estimator; Regression discontinuity; Error-in-variables; Demand for private health insurance.

*Xiaodong Gong is an Associate Professor at NATSEM of IGPA, University of Canberra, e-mail: xiaodong.gong@canberra.edu.au. Jiti Gao is a Professor at Department of Econometrics and Business Statistics, Monash University, email: jiti.gao@monash.edu. We thank Professor Farshid Vahid and the participants of ‘Monash Workshop on Nonparametrics, Time Series and Panel Data’ and of various conferences, for their comments and suggestions. The authors of this paper acknowledge the Australian Research Council Discovery Grants Program support under Grant numbers: DP130104229 and DP150101012.

1 Introduction

This paper is motivated by our attempt to answer an empirical question: how is private health insurance (PHI) take-up in Australia affected by the income threshold at which the Medicare Levy Surcharge (MLS) kicks in?

In Australia, individuals are liable for MLS, which is 1 per cent of their taxable income, if they do not take-up PHI and their taxable income is above certain threshold. This is in addition to the normal Medicare Levy. For example, for single individuals without a child, the threshold was \$50,000 per annum in the 2003-04 financial year. The purpose of this policy measure was, by encouraging people to join the private health system, to reduce the burden of the public medical system and provide more funding to it. It is expected that this may generate a discontinuity (or a jump) in the take-up rate of the PHI at around the threshold. It is of particular interest to estimate the size of the discontinuity. The size of this discontinuity can be seen as the effect of this policy measure under the standard assumption that individuals from both sides of the discontinuity point in the income distribution are valid treated and control groups. The size of the discontinuity also provides some idea on the price elasticity for the demand of PHI as this policy measure provides a rare exogenous variation in the after-tax price of PHI faced by the individuals.

With suitable data, the size of the discontinuity could be estimated with the standard methods used in economics (see for example, Hahn et al. (2001), Porter (2003), and Lee and Lemieux (2010)). However, the only data that are available to us are contaminated. The data for our empirical case are drawn from a ‘1% Sample Unit Record File of Individual Tax Returns’ for the 2003-04 financial year¹ provided by the Australian Tax Office (ATO). Out of privacy considerations, ATO ‘perturbed’ the data by adding measurement errors to the income variables. ATO does provide some aspects of the error distribution. Together with some feature of the tax deduction rules, the distribution of the measurement error can be estimated for certain groups of people. ATO also left out the number of children from the data set, which determines the threshold where MLS applies together with income. Answering this seemingly straightforward question appears to be quite a challenge.

To illustrate the problem, suppose the relationship between the explained variable (the take-up of PHI in our case) Y_i and the regressor (the taxable income in our case) X_i of individual i is given by

$$Y_i = g(X_i) + \eta_i, \tag{1}$$

where $g(\cdot)$ is a continuous function except that it has a discontinuity at the location s with the size of the discontinuity $D = g(s+) - g(s-)$ unknown; the location of

¹Australian Financial year is from 1 July to 30 June.

the discontinuity s is either known or unknown (in our case, it is known); and the error term η_i is uncorrelated with X_i .

Suppose also that we only have a sample of n observations: $\chi = \{Z_i, Y_i\}$, $i = 1, \dots, n$, where Z_i is an error-ridden regressor variable

$$Z_i = X_i + \epsilon_i, \tag{2}$$

where ϵ_i is the measurement error with a known distribution and uncorrelated with X_i and η_i . The aim is to recover the size of the discontinuity D .

A vast literature provides consistent estimators for either the cases when there is no discontinuity ($D = 0$) or the cases without measurement errors ϵ in the regressor, but we find no satisfactory, ready-to-use methods in the literature to deal with the complicated case when both of them are present.

Without measurement errors in the regressor ($\epsilon_i \equiv 0$), the unknown regression function with a discontinuity (or a finite number of discontinuities) at known locations is usually estimated by fitting smooth curves to the left and right of the discontinuity using traditional nonparametric techniques.²

If the locations are unknown, they have to be detected first. This can be done using a range of estimators proposed in a few related literatures such as change-point detection, edge detection and image reconstruction (see Qiu (2005) for a review of these techniques). A range of kernel based estimators are available, for example, Müller (1992); Hall and Titterton (1992); Wu and Chu (1993a,b,c); Gijbels et al. (1999); Gijbels and Goderniaux (2004); and Gijbels et al. (2004), among others.³ Most of these estimators use some form of the first-order gradients of the function to diagnose and estimate the discontinuity. For example, the one proposed by Müller (1992) explores the difference of the two one-sided Nadaraya-Watson estimators; while Gijbels et al. (1999) use the first order derivatives g' as the diagnostic function.

Without discontinuity ($D = 0$) but with measurement errors, conventional parametric and nonparametric regression techniques to recover the unknown function g are no longer valid. A continuous g can be estimated with alternative methods including SIMEX method (see Stefanski and Cook (1995)) and the so-called de-convolution kernel estimator proposed by Fan and Truong (1993) (see Delaigle and Meister (2007) for more references and Carroll et al. (2006) for an extensive literature review). Again, we focus on the de-convolution kernel-based estimator of Fan and Truong (1993). The estimator, which is a transformation of

²Alternatively, Kang et al. (2000) propose to estimate g using adjusted data $Y_i - \hat{D}$ in error-free cases.

³Other types of estimators such as local polynomial, spline-based, or wavelet-based are also studied extensively. See Gijbels and Goderniaux (2004) for a list of references.

traditional kernel estimators using Fourier inversion, is closely related to the de-convolution kernel density estimators such as Carroll and Hall (1988), Fan (1991) and Stefanski and Carroll (1990). We will describe the estimator briefly in the next section.

Estimators for density with discontinuities in presence of measurement errors have been discussed in the literature. For example, Delaigle and Gijbels (2006b) and Delaigle and Gijbels (2006a) propose an estimator based upon the first-order derivatives of the de-convolution kernel density estimator. Yet, to our knowledge, similar estimators in the context of regressions are rare in the literature. Kang et al. (2015) is the only other such paper known to us. In Kang et al. (2015), we tried a ‘one-step-right’ estimator for our case. Noting that a conventional kernel estimator would be biased by the measurement error in the regressor, they first obtain a kernel estimator that is one bandwidth away from the point for estimation. If the point for estimation is the discontinuity point, this estimator will be affected much less by the measurement error. And then, using this estimator as a benchmark, they modify the conventional kernel estimator by penalising those observations that differ from this benchmark. In that method, the distribution of the measurement error is assumed to be unknown.

In this paper, we propose a new de-convolution based estimator when there are discontinuities in the regression function and the regressor is only observed with measurement errors. This is for the cases when we have information on the distribution of the measurement errors. The estimator is adapted from the change-point estimator in error-free cases.⁴ The size of the discontinuity is estimated by the differences of two ‘one-sided’ de-convolution kernel estimators. We put quotation marks around the word one-sided because in presence of measurement errors, we cannot observe precisely which observations are from the left or right side. The idea is simply to construct the ‘one-sided’ kernels by weighting observations with the probabilities of them being on one side of the point at which the function is estimated. The performance of the estimator is examined using Monte Carlo simulations. When the exact location of the discontinuities are unknown, the differences of two ‘one-sided’ de-convolution kernel estimators can be used as the diagnostic function for detecting the discontinuities.

Performance of our estimator, as other de-convolution estimators or those for estimating discontinuity in the error-free cases, depends heavily on the choice of the smoothing parameters, or the bandwidths as we call them in this paper.⁵ We use the bootstrap estimator as discussed in Delaigle and Gijbels (2004b,a) to choose

⁴See for example Müller (1992), Qiu (1991), Qiu et al. (1991), and Wu and Chu (1993a). Also see Qiu (2005) for a review of the methods.

⁵Discussions can be found in Delaigle and Gijbels (2004b,a) for de-convolution estimators and in Gijbels and Goderniaux (2004) for regression discontinuity in error-free cases.

the bandwidths. Confidence bands are obtained from bootstrapped samples.

We then use this estimator to estimate the take-up of PHI by single males as a function of taxable income. This is to get a more homogeneous sample in which the income threshold where the MLS applies lies at \$50,000. This is because the threshold where the MLS applies also depends upon number of children lived with the individual. For each child living with the individual, the threshold is lifted by \$3,000. For example, if the individual lived with one child, the threshold would be \$53,000. Restricting the sample to single males minimises the number of households with dependent children.⁶

The rest of the paper is organised as follows. In Section 2, we describe our new point estimator, preceded by a brief summary of the error-free kernel estimator for regression discontinuity and the de-convolution estimator for unknown regression functions from both of which our new estimator is adapted. In Section 2.4 we discuss the issue of bandwidth selection and the bootstrapping procedure for selecting the bandwidth and estimating the confidence bands. Results of Monte Carlo simulations are presented in Section 3. In Section 4 we estimate the effect of MLS on the take-up of PHI in Australian using this estimator. Section 5 gives some concluding remarks.

2 A de-convolution Estimator for regression discontinuity

We first summarise briefly the two estimators that our proposed estimator is based upon and then build up the estimator for regression discontinuity with an error-ridden regressor.

2.1 Difference kernel estimator for regression discontinuity with an error-free regressor

Suppose we wish to estimate the discontinuity in g using n observations $\{X_i, Y_i\}, i = 1, 2, \dots, n$ generated from Equation (1). The idea of this type of change-point estimator is to base inference for change-points on differences of right- and left-sided kernel estimates:

$$\hat{d}_k(x, h) = \hat{g}_+(x) - \hat{g}_-(x), \quad (3)$$

⁶The sample used in Kang et al. (2015) also includes single women, which may be a likely reason why the estimated discontinuity location appears to be different from \$50,000, the threshold for single persons.

where

$$\hat{g}_+(x) = \sum Y_i K_r \left(\frac{X_i - x}{h} \right) / \sum K_r \left(\frac{X_i - x}{h} \right)$$

is the right-side Nadaraya-Watson estimator in which $K_r(\cdot)$ is a kernel function defined over $[0, 1]$ (the right-hand side of x) which otherwise satisfies the usual conditions for a kernel function; and h is the bandwidth. Hence, $\hat{g}_+(x)$ is a weighted average of the observations in the right-sided neighborhood $[x, x + h]$. $\hat{g}_-(x)$ and $K_l(\cdot)$ are similarly defined (thus, the support for $K_l(\cdot)$ is $[-1, 0]$). It can be shown that $\hat{d}_k(x, h)$ is a consistent estimator of $g_+(x) - g_-(x)$ for all continuous and discontinuous points. When the location of the discontinuity is known, the size of the discontinuity is estimated by

$$\hat{D} = \hat{d}_k(s, h). \quad (4)$$

When the location is unknown, both the location and the size of the discontinuity are consistently estimated by

$$\hat{s} = \arg \max_{x \in [S_0, S_1]} \hat{d}_k(x, h), \quad (5)$$

and

$$\hat{D} = \hat{d}_k(\hat{s}, h). \quad (6)$$

Qiu et al. (1991) label it as the ‘difference kernel estimator’. Also see Qiu (2005) for more details of this estimator. When the locations of discontinuity are unknown, other forms of first-order derivative estimators can also be used in place of $\hat{d}_k(x, h)$ as the diagnostic function such as the first-order kernel estimator and difference local polynomial estimators.⁷ Once the location of the discontinuity is identified, the continuous parts of the function on each side of the discontinuity can be recovered separately with nonparametric techniques using observations of that side only.

2.2 De-convolution kernel estimator for continuous regressions with an error-ridden regressor

Suppose, we wish to estimate g in Equation (1) but with $D = 0$ (that is, without discontinuity), and instead of X_i we only observe n observations $\{Z_i, Y_i\}$, $i = 1, 2, \dots, n$ where Z_i is generated from Equation (2). In this case, the traditional

⁷Gijbels et al. (1999) noted that identifying the discontinuity by simply searching the diagnostic function for its maximum might be problematic. In this paper, we forego the discussion on this potential identification issue and leave it for future work.

nonparametric techniques have to be modified to do the job. Fan and Truong (1993) propose the following de-convolution estimator:

$$\hat{g}^d(x) = \sum Y_j K^* \left(\frac{x - Z_j}{h_d} \right) / \sum K^* \left(\frac{x - Z_j}{h_d} \right) \quad (7)$$

$$= \frac{1}{nh_d} \sum Y_j K^* \left(\frac{x - Z_j}{h_d} \right) / \hat{f}_n(x) \quad (8)$$

where K^* , is the de-convoluted kernel, given by,

$$K^*(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp^{-itu} \Phi_K(t) / \Phi_\epsilon(t/h_d) dt, \quad (9)$$

with Φ_L being the characteristic function of variable L ; h_d the bandwidth; and $\hat{f}_n(x)$ the de-convolution kernel estimator of the density of x . This estimator makes use of the property that the Fourier Transform of the convolution of two distributions is the product of those of the two distributions. It has been shown that it is consistent under mild regularity conditions. See Fan and Truong (1993) for more details of the estimator.

Unlike in the error-free cases, performance of the estimator depends upon the choice of the kernel functions that satisfy certain regularity conditions. Following the discussion in Delaigle and Hall (2006), we use the second-order kernel in this paper. It is defined as $K_2(x) = 48 \cos x \{1 - 15x^{-2}\} / (\pi x^4) - 144 \sin x \{2 - 5x^{-2}\} / (\pi x^5)$ with characteristic function given by $(1 - t^2)^3 I_{[-1,1]}(t)$.

2.3 A difference de-convolution kernel estimator for regression discontinuity with an error-ridden regressor

When both regression discontinuity and errors in the regressor are present, it becomes more complicated.

We need to combine these two estimators and find an analog of the difference kernel estimator in the de-convolution context.⁸ The key issue is how to construct the one-sided de-convolution kernel.

Our idea is quite simple. Since we do not observe the exact locations of the observations in x , the one-sided kernels cannot be constructed from the observations

⁸Of course one could use the first derivative of the de-convolution kernel estimator as the diagnostic function, but we find it does not perform as well as the difference de-convolution kernel estimator we are proposing, perhaps because of an even slower convergence of the derivative estimator.

from that side. However, with the knowledge of the error distribution, the conditional distribution $X_i|Z_i$ can be estimated so that the probability for the true regressor of each observation to be in each side of a particular estimated point can be calculated. Then, the ‘one-sided’ kernel estimator can be constructed by weighting the standard de-convolution kernel estimator (7) with the probability for the observation to be on that side of the point at which the function is calculated.

Suppose again that we wish to identify $D \neq 0$ in Equation (1) but only n observations of $\{Z_i, Y_i\}$ observed. Assume that x has a smooth density function $f_X(x)$ which is non-zero almost everywhere over its support; and that the error ϵ has a density function f_ϵ with a characteristic function $\Phi_\epsilon(t) \neq 0$. We construct the ‘one-sided’ de-convolution kernel estimators of g as follows.

The right-sided kernel estimator is given by

$$\hat{g}_r^d(x) = \sum Y_j K_r^*\left(\frac{x - Z_j}{\tilde{h}_d}\right) / \sum K_r^*\left(\frac{x - Z_j}{\tilde{h}_d}\right), \quad (10)$$

where

$$K_r^*(u) = \hat{w}_i^r(x) K^*(u), \quad (11)$$

$K^*(u)$ is the same as in (9) and the weight $\hat{w}_i^r(x)$ is given by

$$\hat{w}_i^r(x) = \frac{\int_{-\infty}^{Z_i-x} \hat{f}_X(Z_i - \epsilon) f_\epsilon(\epsilon) d\epsilon}{\int_{-\infty}^{+\infty} \hat{f}_X(Z_i - \epsilon) f_\epsilon(\epsilon) d\epsilon} \quad (12)$$

$$= \int_{-\infty}^{Z_i-x} \frac{\hat{f}_X(Z_i - \epsilon) f_\epsilon(\epsilon)}{\hat{f}_Z(Z_i)} d\epsilon \quad (13)$$

$$= \hat{Pr}ob\{X_i > x | Z_i\}, \quad (14)$$

where f_ϵ is assumed to be known and $\hat{f}_X(\cdot)$ can be calculated using the standard de-convolution kernel density estimator (8).

When ϵ has a discrete distribution, the integrals are replaced by sums and the calculations can be simpler. For example, suppose ϵ takes K values ϵ_k ($k = 1, \dots, K$), with p_k the probability associating with each point, then the weights can be calculated relatively straightforwardly as

$$\hat{w}_i^r(x) = \frac{\sum_{Z_i - \epsilon_j > x} \hat{f}_X(Z_i - \epsilon_j) * p_j}{\sum_{\text{all } s} \hat{f}_X(Z_i - \epsilon_s) p_s}.$$

The left-sided kernel estimator is similarly defined as

$$\hat{g}_l^d(x) = \sum Y_j K_l^*\left(\frac{x - Z_j}{\tilde{h}_d}\right) / \sum K_l^*\left(\frac{x - Z_j}{\tilde{h}_d}\right), \quad (15)$$

with

$$K_l^*(u) = \hat{w}_i^l(x)K^*(u),$$

where $\hat{w}_i^l(x) = 1 - \hat{w}_i^r(x)$. Note that the bandwidth for the one-sided kernel estimators may be different from h_d , the one for estimating the density function.

2.3.1 When the location of the discontinuity is known

With $\hat{g}_r^d(\cdot)$ and $\hat{g}_l^d(\cdot)$ in hands, an estimator of the size of discontinuity at a known point s can be defined in the similar fashion of (3) as:

$$\hat{d}_d(s, \tilde{h}_d) = \hat{g}_r^d(s) - \hat{g}_l^d(s) \quad (16)$$

As is usually the case (and also to construct the confidence intervals for the discontinuity when the location is unknown), the calculation is done conditional on a set of equally spaced design points over a closed interval of X . Let x_j denote such a design point where $j = 1, \dots, S$ is the index of the design points.

2.3.2 When the location of the discontinuity is unknown

If the location of the discontinuity s is unknown, the function $\hat{d}_d(\cdot, \tilde{h}_d)$ is used as the diagnostic function. The location and the size of the discontinuity can be identified by searching the maximum of $\hat{d}_d(x, \tilde{h}_d)$ over $[S_0, S_1]$:

$$\hat{s}_d = \arg \max_{x_j \in [S_0, S_1]} \hat{d}_d(x, \tilde{h}_d), \quad (17)$$

which corresponds to an index \hat{j}_s ; and

$$\hat{D}_d = \hat{d}_d(\hat{s}_d). \quad (18)$$

We follow the convention of Qiu et al. (1991) and call it as the ‘difference de-convolution kernel estimator’.

Once s is estimated, g can be estimated separately for points at each side of the discontinuity: $x < \hat{s}_d$ and $x > \hat{s}_d$, respectively. Here we need to use the same trick of weighting kernel of each observation i by $\hat{w}_i^r(\hat{s}_d)$ and $\hat{w}_i^l(\hat{s}_d)$ to estimate $\hat{g}_{sr}^d \equiv \hat{g}^d(x|x > \hat{s}_d)$ and $\hat{g}_{sl}^d \equiv \hat{g}^d(x|x < \hat{s}_d)$, respectively. Note that the weights are relative to the discontinuity point so that \hat{g}_{sr}^d and \hat{g}_{sl}^d are still ‘two-sided’ kernel estimators in the sense that observations from both sides of x are used. The bandwidths used for \hat{g}_{sr}^d and \hat{g}_{sl}^d , denoted as h_{dr} and h_{dl} , can be different from h_d , \tilde{h}_d , and from each other.

The procedure can be summarised as follows:

- Step 1: Estimate \hat{f}_X using (8) from which $w_i^r(x)$ can be calculated using (12);
- Step 2: Construct $\hat{g}_r^d(x)$ and $\hat{g}_l^d(x)$ using (10) and (15);
- Step 3: If s is known, \hat{D}_d can be estimated using (18); if s is unknown, estimate \hat{s}_d and \hat{D}_d using (17) and (18), respectively;
- Step 4: Obtain \hat{g}_{lr}^d , the estimates of g , by estimating \hat{g}_{sr}^d and \hat{g}_{sl}^d around \hat{s}_d separately using de-convolution kernels weighted by $\hat{w}_i^r(\hat{s}_d)$ and $\hat{w}_i^l(\hat{s}_d)$, respectively.

2.3.3 Choosing the bandwidths

As mentioned earlier, selecting appropriate bandwidths (or smoothing parameters) is crucial for the performance of de-convolution estimators. This is one of the main drawbacks of this kind of estimators. Various methods have been proposed for a continuous function. See for example, Delaigle and Gijbels (2004a). and Delaigle and Gijbels (2004b). We choose the bandwidths using a bootstrapping procedure for a continuous function which minimise the Asymptotic Mean Integrated Square of Error (AMISE), as those proposed by Delaigle and Gijbels (2004a).

2.4 Estimating the confidence bands of the discontinuity when the location is unknown

The confidence bands of the discontinuity location is obtained using a bootstrapping procedure. We draw with replacements R bootstrap samples from the original dataset. For each of these R samples, we obtain estimated discontinuity location \hat{s}^r , $r = 1, \dots, R$. A confidence interval of s is then constructed from the empirical distribution of $\{\hat{s}^1, \dots, \hat{s}^R\}$. Specifically, for a given significance level $\alpha \in (0, 1)$, a $100(1 - \alpha)\%$ confidence interval for s is defined to be $(\hat{s}_{R,\alpha/2}, \hat{s}_{R,1-\alpha/2})$, where $\hat{s}_{R,\alpha/2}$ and $\hat{s}_{R,1-\alpha/2}$ are the $(\alpha/2)100\%$ -th and $(1 - \alpha/2)100\%$ -th percentiles of the bootstrapped distribution of $\{\hat{s}^1, \dots, \hat{s}^R\}$.

3 Monte Carlo Simulations

To evaluate the performance of the estimator, we conduct Monte Carlo simulations for a setting $\{y, x, z\}$ that is close to our empirical study. Specifically, the dependent variable y is a binary variable with the probability of $y = 1$ is given by

a logistic function where

$$\begin{aligned} Prob\{y = 1|x\} &= \frac{\exp(x/4 + 0.8I(x > 1.0))}{\exp(x/4 + 0.8I(x > 1.0)) + 1}, \\ x &\sim N(0, 1) \\ Z &= x + \epsilon \end{aligned}$$

where ϵ follows a two-point distribution. This setting is similar to many empirical cases encountered in economics including ours. We chose a normally distributed x that has a variable density over its support and a discontinuity at a point that has different density to its left and right.

The two-point distribution of the error term simplifies the calculation a lot. First of all, for this distribution, the characteristic function Φ_ϵ in (9) is simply $\cos(t\sigma_\epsilon/h)$ ⁹. Secondly, the weights can be calculated as

$$\hat{w}_i^r(x) = \begin{cases} 0, & \text{if } Z_i < x - \sigma_\epsilon; \\ \frac{\hat{f}_X(Z_i + \sigma_\epsilon)}{\hat{f}_X(Z_i - \sigma_\epsilon) + \hat{f}_X(Z_i + \sigma_\epsilon)}, & \text{if } Z_i \in [x - \sigma_\epsilon, x + \sigma_\epsilon]; \\ 1, & \text{if } Z_i > x + \sigma_\epsilon; \end{cases}$$

Thirdly, to calculate the function to the two sides of the estimated discontinuity point, \hat{g}_{sr}^d and \hat{g}_{sl}^d , the sample can be partitioned into the one that $Z_i > \hat{s}_d - \sigma_\epsilon$ and $Z_i < \hat{s}_d + \sigma_\epsilon$. The observations within the neighbourhood of $[\hat{s}_d - \sigma_\epsilon, \hat{s}_d + \sigma_\epsilon]$ are then weighted using $\hat{w}_i^r(\hat{s}_d)$ which is again easy to calculate.

Finally, the residuals can now be calculated as

$$\begin{aligned} \tilde{\eta}_i &= Y_i - \{\hat{g}_{lr}^d(Z_i - \sigma_\epsilon)Prob\{x = Z_i - \sigma_\epsilon|Z_i\} + \hat{g}_{lr}^d(Z_i + \sigma_\epsilon)Prob\{x = Z_i + \sigma_\epsilon|Z_i\}\}. \\ &= Y_i - \left\{ \frac{\hat{g}_{lr}^d(Z_i - \sigma_\epsilon)\hat{f}_X(Z_i - \sigma_\epsilon)}{\hat{f}_X(Z_i - \sigma_\epsilon) + \hat{f}_X(Z_i + \sigma_\epsilon)} + \frac{\hat{g}_{lr}^d(Z_i + \sigma_\epsilon)\hat{f}_X(Z_i + \sigma_\epsilon)}{\hat{f}_X(Z_i - \sigma_\epsilon) + \hat{f}_X(Z_i + \sigma_\epsilon)} \right\}. \\ &= Y_i - \{\hat{g}_{lr}^d(Z_i - \sigma_\epsilon)\hat{w}_i^l(Z_i - \sigma_\epsilon) + \hat{g}_{lr}^d(Z_i + \sigma_\epsilon)\hat{w}_i^r(Z_i + \sigma_\epsilon)\}. \end{aligned}$$

We conducted a few sets of simulations with two different sample sizes and with σ_ϵ taking two different values, both with the location of the discontinuity assumed to be known and unknown. In Table 1, we present the Mean Integrated Squares of Errors (MISE) of 200 samples of these simulations. And in Figures 1 and 2, where

⁹This characteristic function contains isolated points over its support at $(2k + 1)/2\pi, k = \dots, -1, 0, 1, \dots$, when the bandwidth h becomes too small relative to σ_ϵ . In such cases, deconvolution estimators can still be consistent with some treatment at those points. This appears not an issue when σ_ϵ is small relative to the standard deviation of the true regressor, which is the case for most of the empirical applications.

we plot the distributions of the location and the size of the discontinuity. The bandwidth \tilde{h}_d is chosen by the method described above. The results show that the estimates are centered around their true values, and the estimator performs reasonably well for samples of 3,000 observations. But the MISE's for a sample of 1,000 observations can still be quite substantial. This means that the convergence is quite slow and sample size needs to be quite large for the estimator to work. This is one of the main drawbacks of this estimator. Nevertheless, it is expected because both the noise in the structure equation and the measurement errors contribute to the estimation errors. In addition, when the location is unknown, the MISE's of the location are also affected by the variance of the measurement errors. When the variance of the measurement errors become larger, the location is less accurately estimated. It can be seen that estimated locations are estimated more accurately than the size. This is also because of the noise in the function of the simulations. Figure 3 presents the estimates, from a sample of 3,000 observations, of the function and the discontinuity location, together with its confidence intervals. For this sample, the confidence bands of the location is [0.947, 1.281], and that of the size is [0.132, 0.229].

[Table 1 goes here]
[Figure 1 goes here]
[Figure 2 goes here]
[Figure 3 goes here]

4 Empirical Application

4.1 Background

Since the introduction of Medicare in 1984, private health insurance coverage in Australia declined steadily towards the end of 1990s (when the take-up was only 31 per cent) until a series of policies were introduced. 1) In 1997, the Private Health Insurance Incentives Scheme (PHIIS) was introduced, which imposes a the MLS (a tax levy) on high-income taxpayers who do not have private insurance and provides a means-tested subsidy schedule for low-income earners who purchase; 2) In 1999, a 30% tax rebate on private insurance premium was introduced for all PHI policies and the means-tested component under PHIIS was replaced; and 3) in 2000, Life Time Health Cover (LHC), a system of entry-age ratings in which a premium surcharge of 2 percent is charged for every year that the initial purchase is delayed after age 30. Between 1997-1998 and 2007-2008, the threshold of taxable income at which MLS is payable was \$50,000 for singles without children and combined \$100,000 for couples. For each dependent child, in the household, the threshold

increases by \$3,000. After these measures, private health insurance is taken-up by around 45 per cent of Australians (see Palangkaraya et al. (2009)). Impacts of some of these policy measures, e.g., LHC, have been studied in a few studies, including Butler (2002), Frech et al. (2003), Palangkaraya and Yong (2005), Palangkaraya et al. (2009), Ellis and Savage (2008) and Stavrunova and Yerokhin (2014). For example, using a structural model, Ellis and Savage (2008) investigate the impact of the series of reforms. They conclude that the major driver of the increased enrollment as a response to a deadline and advertising blitz, rather than a pure price response'. Nevertheless, with the exception of the effect of Stavrunova and Yerokhin (2014), MLS is hardly studied separately. Stavrunova and Yerokhin (2014) examine the impact of MLS on the take up of private health insurance by exploring the relationship between income level and private health insurance coverage. The data they used were aggregate but not individual data, although also from the same source. They conclude that for the year of 2007-2008, the impact of MLS at the threshold was 6.5 percentage points.

Estimating such an effect is not only interesting for the sake of evaluating this particular policy, but also for informing the value of and the demand for PHI's. There is a large literature using tax changes (either over time or cross individuals) as a source of variation in the after-tax price of health insurance to make inferences on the demand of PHI. However, with a few exceptions such as Ellis and Savage (2008) and Stavrunova and Yerokhin (2014), most of the studies are for the US employer provided health insurance. See Gruber and Poterba (1994), Finkelstein (2002), Rodriguez and Stoyanova (2004), Buchmueller et al. (2011), for a few examples. Rarely is the case, though, that the tax-changes could be argued as exogenous. Discontinuities caused by policy design such as the MLS in Australia have been argued to be exogenous locally for the individuals around it, and are explored popularly in the literature in similar contexts (see the review by Lee and Lemieux (2010)).

4.2 Perturbed data and the error distribution

The data we used for the empirical study are drawn from a confidentialised '1% Sample Unit Record File of Individual Income Tax Returns' for the 2003-04 financial year developed by the Australian Tax Office (ATO) for research purposes. The file contains just over 109,000 records of individual tax returns and detailed information on income from various sources; different types of tax deductions; taxable income; and of course the take-up of PHI by the individuals. It also contains a limited number of demographic variables including gender, age group, and marital status. Unfortunately, the number of dependent children is not included in the sample.

For our purpose, we focus only on single males who are between 20 and 69 years of age so that most of them face the same \$50,000 threshold. To minimise the number of income sources/deduction sources so that we can have enough knowledge of the error distribution, the sample is restricted further according to the following criteria: 1) Only those who have positive earnings as the only sources of income are selected; 2) Individuals whose taxable income is not positive (which means their total tax deductions are no less than their earnings) are dropped; and 3) We further drop individuals whose non work related deductions form significant part of their taxable income—specifically, we drop those individuals whose work related deductions are less than 90 percent of earnings when the total deductions are more than 10 percent of earnings; whose total deductions are over 50 percent of earnings; or whose total deductions are all non-work related and the total deductions are over 10 percent of their earnings.

It will become clear below that restricting the sample to individuals with earnings as the only source of income and those with work-related deductions as the main source of deduction is to ensure only one of the added errors is important in the sample so that we can reasonably identify the distribution of the error. The final sample for analysis consists of 4,357 individuals. We summarise the sample statistics in Table 2. The table shows that on average, about 30 percent of the singles took up PHI in 2003-04. Looking into the sample more closely, we find that the difference between individuals below the observed \$50,000 taxable income mark and those above is large. The take-up rate for the former group is about 26 percent, and about 66 percent for the latter. Of course the difference may not only be due to the MLS. In Figure 4, we estimate the PHI take-up against the observed \ln taxable income. The figure gives a rough idea how the take-up increases with income and that the increase is the fastest in the neighbourhood between around \$50,000 and \$60,000 (≈ 10.8 - 11.0 in the \ln scale).

[Table 2 goes here]

[Figure 4 goes here]

As a method of confidentialisation, ATO ‘perturbed’ the income variables and the deductions but provided some information on the way the data are perturbed: a several random numbers within a specified range for each individual are generated, which are converted into a rate (equal probability of being positive or negative) which is applied to the various components of the tax return. These rates are applied to the components in a way to try to maintain relationships with similar items. This is achieved by grouping the components into three broad categories: work or employment related income and deductions; investment income and deductions; and business and other income and deductions.

To implement our procedure, we need to know the error distribution reasonably

well. From the description of the perturbing procedure, we know that the three errors take the form of two-point discrete distributions with similar variances (the rates); and that the same error is added to income and deductions in the same category. The way we restrict the sample allows us to limit the influence of the two non-work-related errors so that they can be ignored. More specifically, suppose $\{X^1, r^1, \epsilon^1\}$, $\{X^2, r^2, \epsilon^2\}$, and $\{X^3, r^3, \epsilon^3\}$, are income, deductions, and the error in each income category; and the total taxable income $X = (X^1 - r^1) + (X^2 - r^2) + (X^3 - r^3)$.

With earnings as the only source of income in this sample, the observed taxable income is then

$$\begin{aligned} Z &= (1 + \epsilon^1)(X^1 - r^1) - (1 + \epsilon^2)r^2 - (1 + \epsilon^3)r^3 \\ &= (1 + \epsilon^1)(X^1 - r^1 - r^2 - r^3) + (\epsilon^1 - \epsilon^2)r^2 + (\epsilon^1 - \epsilon^3)r^3 \end{aligned}$$

As shown in Table 2, the non-work related deductions consist of less than half percent of the gross earnings. Together with the fact that ϵ 's are small, the second and the third terms in Z can be ignored so that $Z \approx (1 + \epsilon^1)(X^1 - r^1 - r^2 - r^3)$. Taking log, we got

$$\ln Z \approx \ln(1 + \epsilon^1) + \ln \overbrace{(X^1 - r^1 - r^2 - r^3)}^X \approx \epsilon^1 + \ln X.$$

Now all that we need is to find out the variance σ_ϵ^2 to identify the error distribution. To do so, we explore a built-in feature of the tax law related to the deductions: if the total deductions an individual claims in their tax return are \$300 or less, no receipt is required to be kept. We suspect that this would cause a spike in the density of deductions around \$300. When the measurement error is added, the spike would split into two symmetric ones around the \$300 mark. If this is the case, from the distance of the two spikes, we would be able to estimate σ_ϵ . In Figure 5, we plot the density functions of the deductions for individuals with work-related deductions only (no other type of deductions), estimated using various bandwidths,. Indeed, symmetric around the \$300 mark (≈ 5.7 in the ln scale), except when the bandwidth becomes too large and the curve is over-smoothed, there are two spikes approximately of the same height with a distance of about 15 percent. A second pair of such spikes symmetric to the \$150 mark can also be seen, again with a distance of around 15 percent. We do not know the reasons for the second pairs of spikes but they seem to be reassuring. We thus can reasonably assume that $\sigma_\epsilon \approx .075$.

[Figure 5 goes here]

4.3 Estimation and the results

The estimated results for three different bandwidths are summarised in Table 3. Using the ‘optimal’ bandwidth, the size of the discontinuity at \$50,000 (≈ 10.82 in log term) is estimated to be about 0.223 with a 95 percent confidence interval of [.18, .27]. This means that at the income level of \$50,000, the take-up of PHI was increased by about 22 percentage points due to the \$500 MLS at this income level. In other words, these 22 per cent of individuals took up PHI to avoid the \$500 tax, or needed to be compensated up to this amount to take up the PHI. This also implies a negative price elasticity of PHI demand since the jump in the take-up can be seen as a response to a price discount in the premium. As one would expect, the estimates increase with the bandwidth, but are reasonably robust. Our estimated effect is larger than what Stavrunova and Yerokhin (2014) estimated. The difference could be due to a couple of reasons. They use aggregate level data which give rates of private health insurance coverage at various income intervals, while we use individual level data but with measurement errors. They argue that taxpayers may manipulate their income and take it into account using predicted private health insurance coverage at the threshold.

[Table 3 goes here]

To check that there is indeed a discontinuity at the threshold, we estimate the model again pretending the location of the discontinuity were unknown. We restrict the possible discontinuity to be between \$45,000 and \$70,000, which means $[S_0, S_1] \approx [10.22, 11.23]$ in the ln scale. The results are summarised in Figures 6 and 7, as well as in Table 4. In Figure 6, we present the two one-sided kernels together, their differences, together with the final two-sided kernel estimates of the function. In particular, the differences of the two one-sided kernels are the diagnostic function $\hat{d}_d(\cdot, \tilde{h}_d)$ defined by Equation (16). The discontinuity point is estimated by its maximum at 10.869 ($\approx 52, 523$) (indicated by the pole). As shown in Figure 7 and in Table 4, the 95 per cent confidence intervals of the discontinuity location constructed using the bootstrapping methods include the true threshold, indicating an insignificant difference between the estimated and the true location. Moreover, the estimated size of the discontinuity is also very close with each other at these two estimated locations. It worth noting that the location estimate is very robust.

[Table 4 goes here]

[Figure 6 goes here]

[Figure 7 goes here]

Apparently, the estimation would have been simpler if the function g were assumed to be known, but we do not know the functional form a priori. In Figure 8, we plot the nonparametric estimates of the function, together with the their

fitted values using linear and quadratic models. What we can see is that the linear model would not fit the pattern.

[Figure 8 goes here]

5 Conclusions

In this paper, we provide a workable solution for estimating unknown functions with a finite number of discontinuity points using contaminated data. The new de-convolution based regression estimator is adapted from the change-point estimator in error-free cases. The idea is to construct the ‘one-sided’ kernels by weighting each observation with the probability of it being on one side of the point to be estimated at; and use the difference of the two ‘one-sided’ de-convolution kernel estimators as the diagnostic function for detecting the discontinuity.

Performance of our estimator, as other de-convolution estimators or those for estimating discontinuity in the error-free cases, depends heavily on the choice of the bandwidths. We propose a bootstrapping procedure for bandwidth selection and interval estimation for the error-in-variable case by adapting the procedure in Gijbels and Goderniaux (2004) and Gijbels et al. (2004) for the error-free case.

The performance of the estimator is examined using Monte Carlo simulations. The results show that the estimator performs reasonably well, but the convergence is quite slow, which is true for this type of estimators in general.

As an application, we use this estimator to estimate the take-up of PHI by single individuals as a function of taxable income, which is expected to have a discontinuous point generated by the MLS policy. We find that at least at the income level of \$50,000, MLS has brought extra 22 percent take-up of private health insurance in Australia. This implies that these people need compensation of up to \$500 to take up the PHI and indicates that the demand for PHI in Australia responds negatively to price changes of the premium. We verified the results by estimating the location and the size simultaneously. We find that neither the difference between the estimated and the true threshold nor that between the two estimated sizes of discontinuity is significant. Our study also illustrates a way to verify whether the estimated discontinuity is indeed the true one in regression discontinuity exercises.

Left for future work are 1) deriving the asymptotic properties of the estimator; 2) discussion of the identification issues of the discontinuity points; 3) improving ways to find optimal bandwidths; and 4) exploring the possibility of including the Loader (1996) type of procedures to refine the estimator. More generally, it is also worthwhile to extend it to semi-parametric settings to allow for other control variables in the model.

Figures

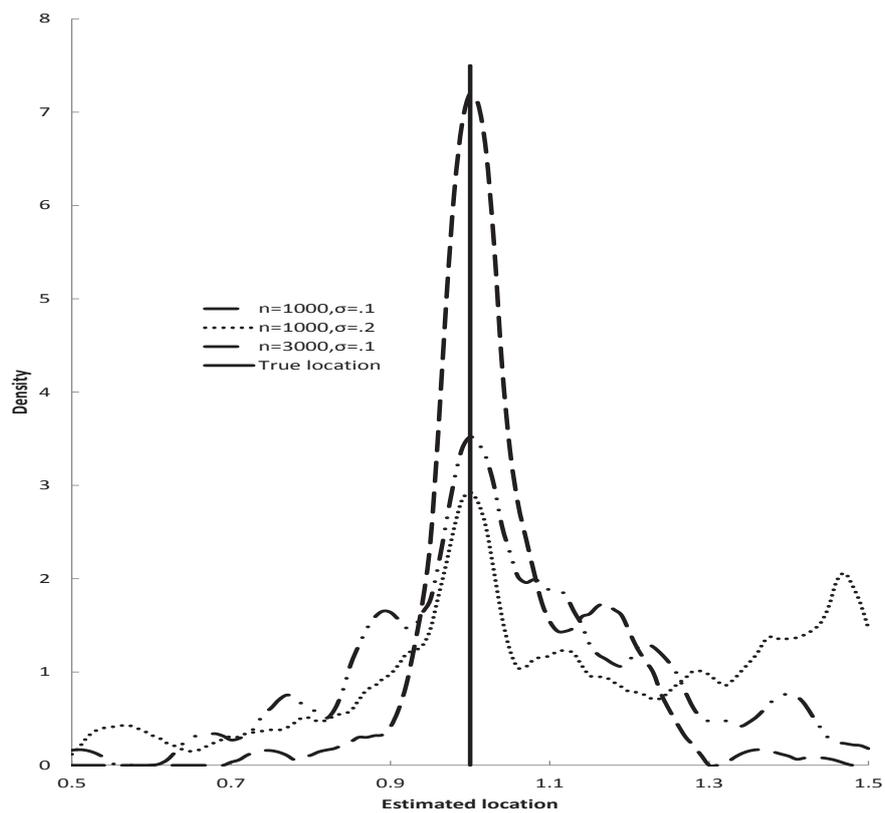


Figure 1. The distribution of the estimated discontinuity location (200 Monte Carlos)
(True location is at 1.)

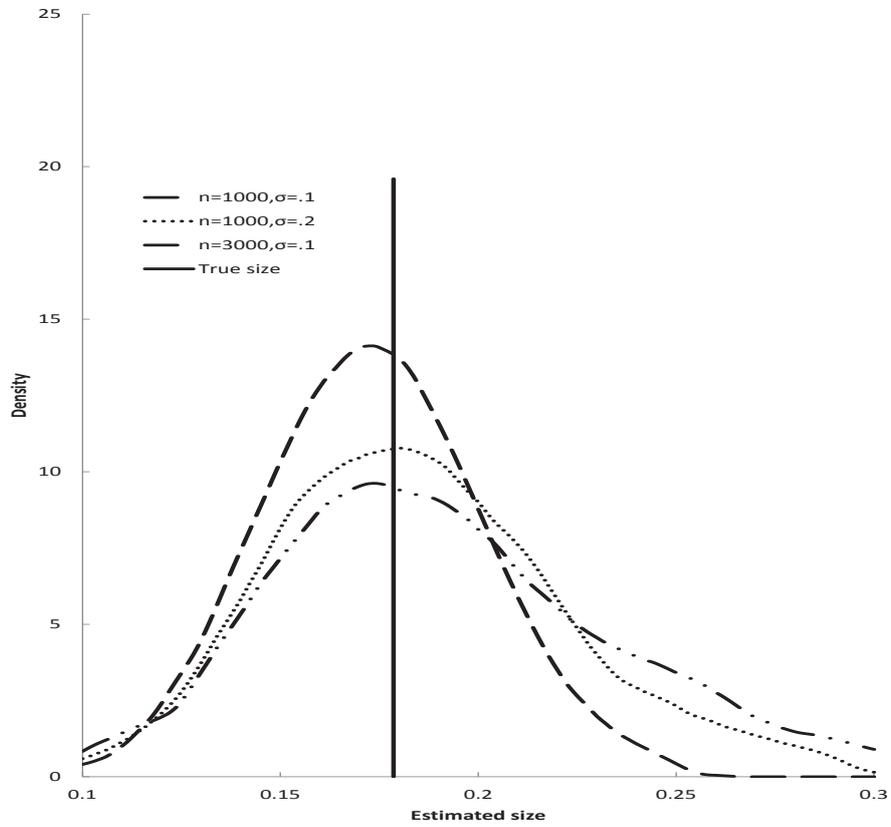


Figure 2. The distribution of the estimated discontinuity size (200 Monte Carlos)
(True size $\approx .179$)

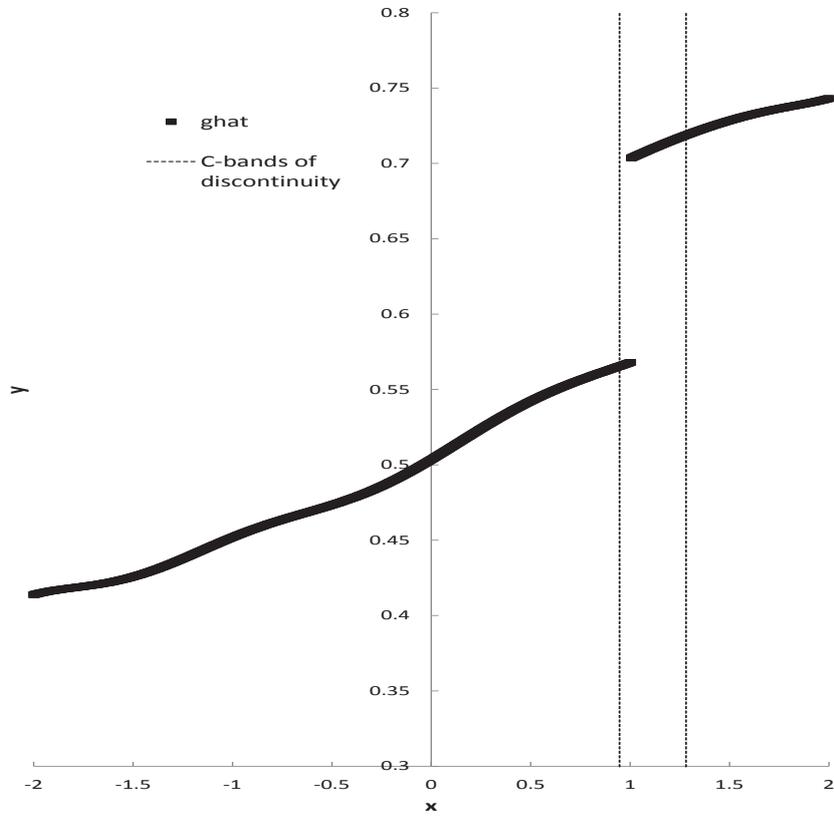


Figure 3. Estimated unknown function and its confidence bands of a simulated sample (Sample size =3,000)

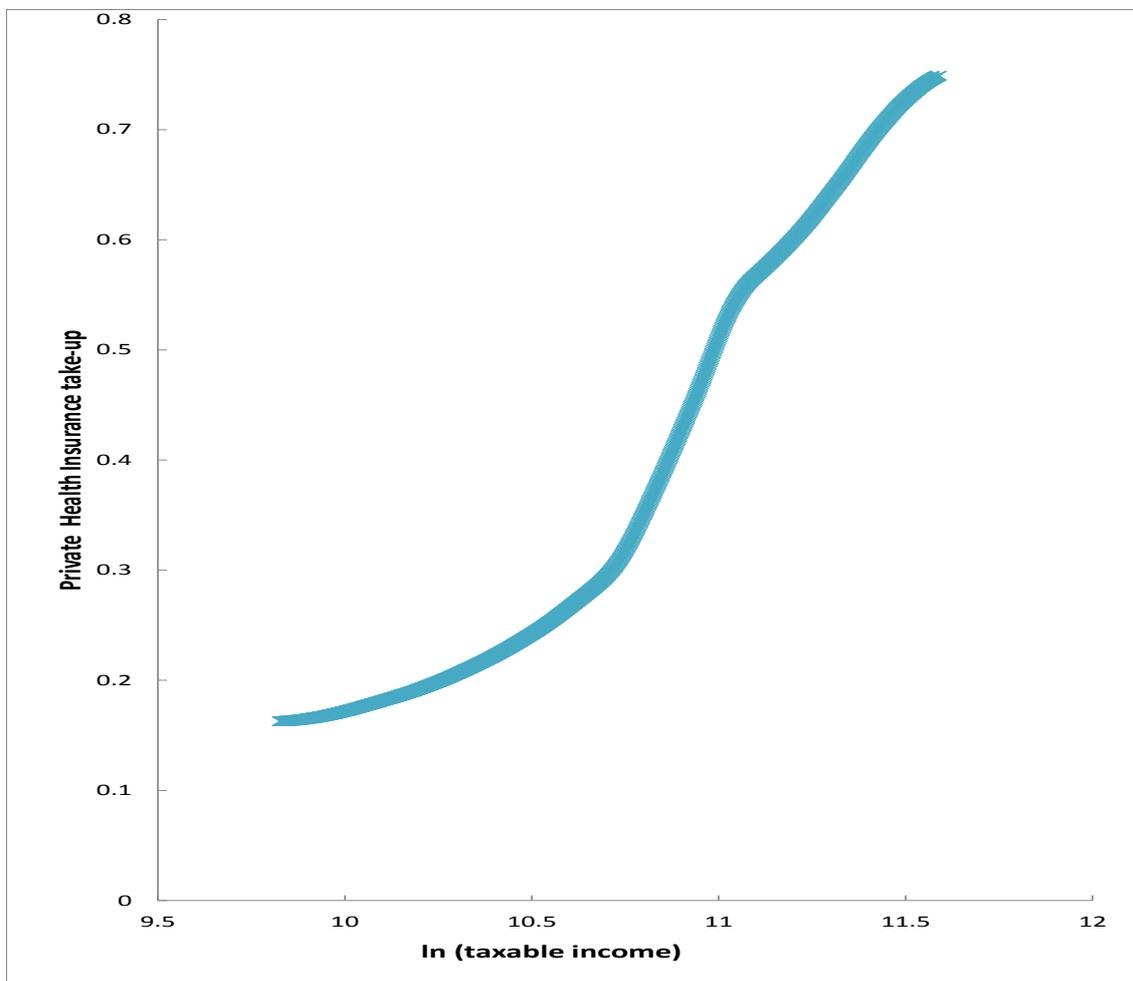


Figure 4. Estimated PHI Take-up against observed \ln taxable income (calculated with Quartic kernel; $h = .153$)

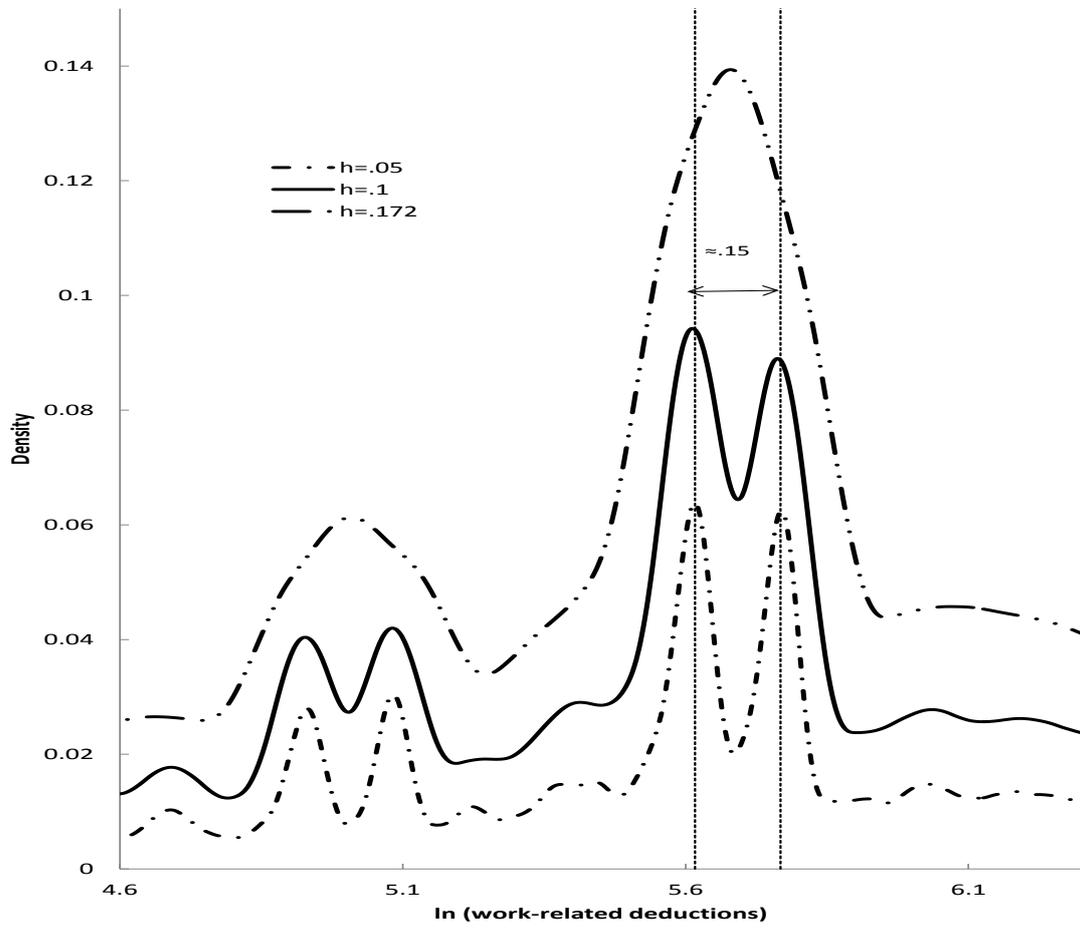


Figure 5. Estimated densities of observed work-related deductions using various bandwidths (calculated with Quartic kernel)

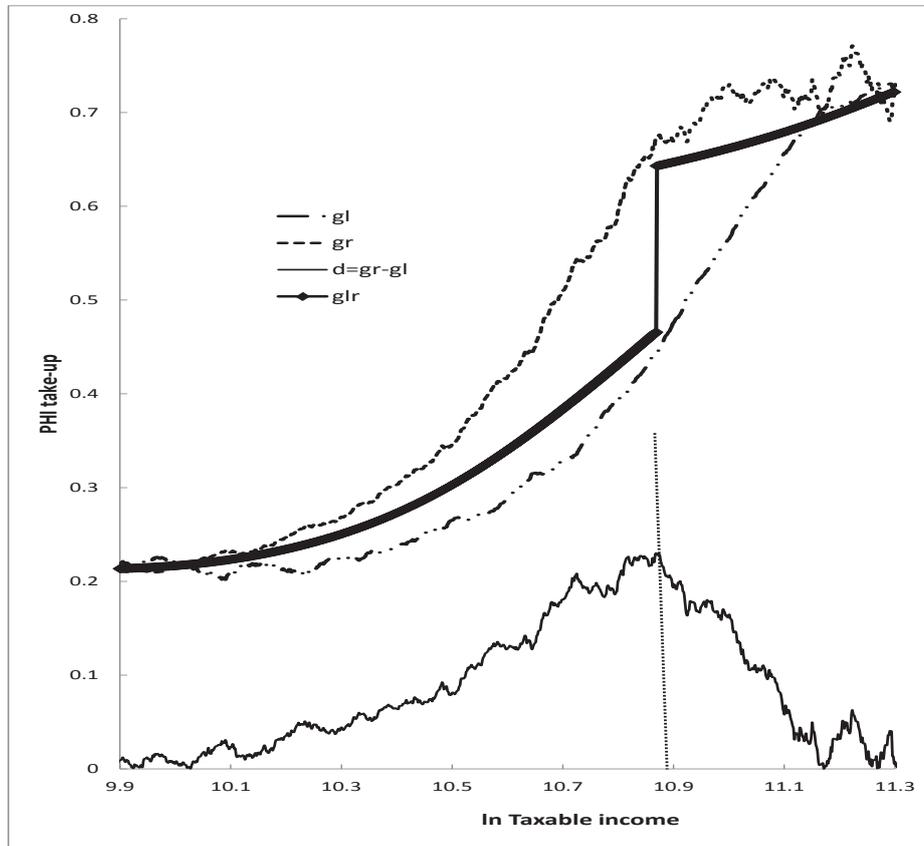


Figure 6. De-convolution kernel estimates for PHI against \ln taxable income (glr : two-sided; gl , gr : one-sided; $d = gr - gl$: the difference. $h_d = .0605$)

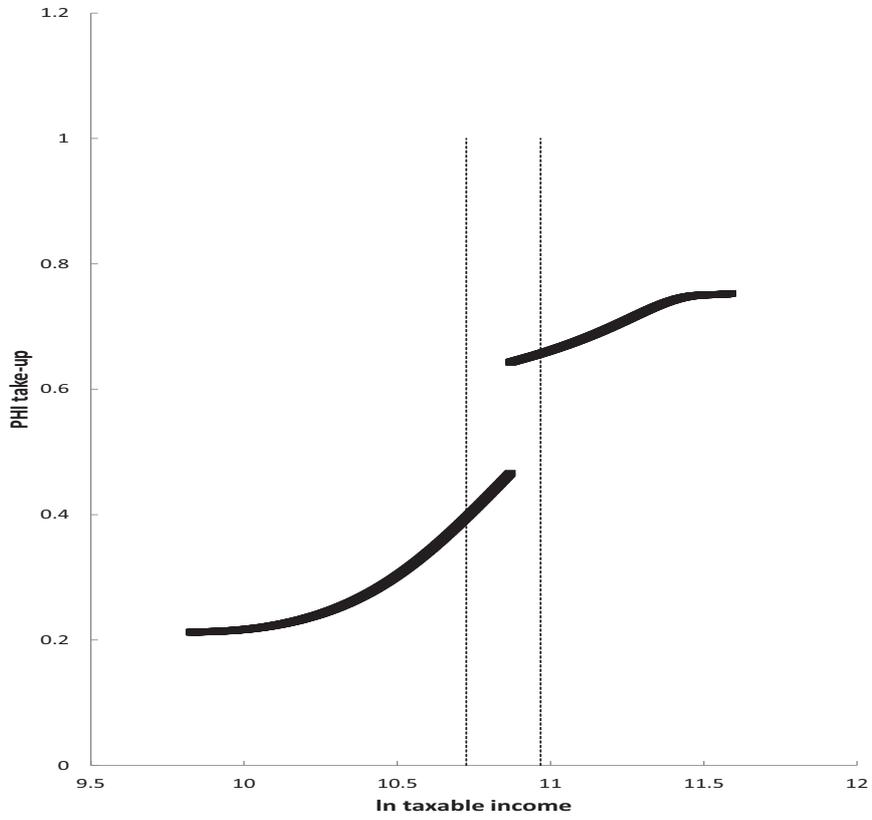


Figure 7. Function estimates with 95% confidence bands for PHI against \ln taxable income

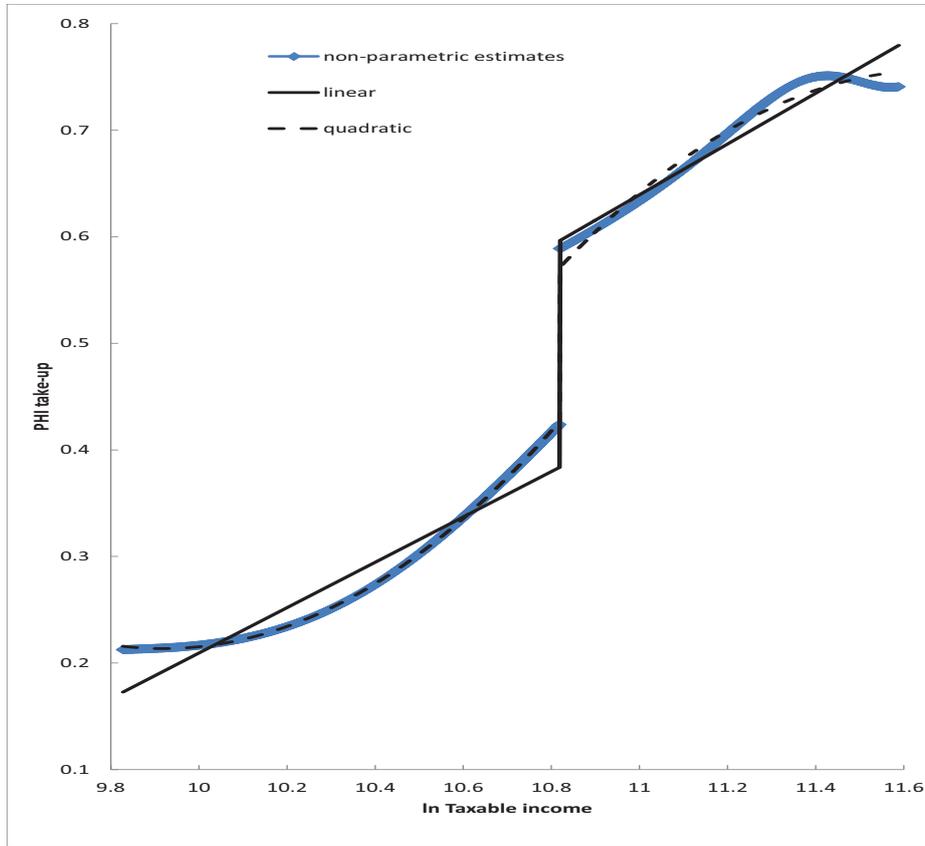


Figure 8. Non-parametric vs parametric estimates of the take-up function

Tables

Table 1. MISE of 200 Monte Carlo Simulations

(True Parameters: $x = 1$ and $d = .179$)

	Location is known		Location is unknown		
	Size	Function	Location	Size	Function
$\sigma = 0.1, n = 1,000$	0.0037	0.0018	0.0355	0.0022	0.0018
$\sigma = 0.2, n = 1,000$	0.0024	0.0012	0.0757	0.0013	0.0019
$\sigma = 0.1, n = 3,000$	0.0013	0.0007	0.0110	0.0006	0.0008

Table 2. Sample statistics

Variable	Mean	Std. Dev.
Dummy, with PHI	.302	
Gross earnings (\$k)	32.178	17.33
Total deductions (\$k)	1.094	1.71
Taxable income (\$k)	31.084	16.87
Work-related deductions (\$k)	.958	1.63
Work-related/Total deductions (%)	.809	.26
Nonwork related/Gross earnings (%)	.004	.01
Obs.		4,357

Table 3. Estimates of the MLS effect at the threshold

$h = .070$ (‘Optimal’)	0.223[0.18, 0.27]
$h = .077$	0.241[0.20, 0.28]
$h = .062$	0.199[0.15, 0.24]

In brackets are the bootstrapped 95% confidence intervals.

Table 4. Estimates of the MLS effect as if the threshold were unknown

	Location	Size
$h = .070$ (‘Optimal’)	10.869[10.73, 10.97]	.230[0.20, 0.28]
$h = .077$	10.869[10.73, 10.97]	.249[.22, .30]
$h = .062$	10.869[10.72, 10.99]	.209[0.19 0.26]

In brackets are the bootstrapped 95% confidence intervals.

References

Buchmueller, T., Didardo, J., and Valletta, R. (2011). The effect of an employer health insurance mandate on health insurance coverage and the demand for labor: Evidence from hawaii. *American Economic Journal Economic Policy*, 3(4):25–51.

- Butler, J. (2002). Policy change and private health insurance did the cheapest policy do the trick. *Australian Health Review*, 25(6):33–41.
- Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*, 83(404):1184–1186.
- Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models*. Boca Raton, FL: Chapman & Hall CRC, second edition.
- Delaigle, A. and Gijbels, I. (2004a). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, 56(1):19–47.
- Delaigle, A. and Gijbels, I. (2004b). Practical bandwidth selection in deconvolution kernel density estimation. *Computational statistics & data analysis*, 45(2):249–267.
- Delaigle, A. and Gijbels, I. (2006a). Data-driven boundary estimation in deconvolution problems. *Computational Statistics and Data Analysis*, 50(8):1965–1994.
- Delaigle, A. and Gijbels, I. (2006b). Estimation of boundary and discontinuity points in deconvolution problems. *Statistica Sinica*, 16(3):773–788.
- Delaigle, A. and Hall, P. (2006). On optimal kernel choice for deconvolution. *Statistics and Probability Letters*, 76(15):1594–1602.
- Delaigle, A. and Meister, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *Journal of the American Statistical Association*, 102(480):1416–1426.
- Ellis, R. P. and Savage, E. (2008). Run for cover now or later? the impact of premiums, threats and deadlines on private health insurance in australia. *International Journal of Health Care Finance and Economics*, 8(4):257–277.
- Fan, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *The Annals of Statistics*, 19(3):1257–1272.
- Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics*, 21(4):1900–1925.
- Finkelstein, A. (2002). The effect of tax subsidies to employer-provided supplementary health insurance: evidence from canada. *Journal of Public Economics*, 84(3):305–339.

- Frech, H., Hopkins, S., and MacDonald, G. (2003). The Australian private health insurance boom: Was it subsidies or liberalised regulation. *Economic papers*, 22(1):58–64.
- Gijbels, I. and Goderniaux, A.-C. (2004). Bandwidth selection for changepoint estimation in nonparametric regression. *Technometrics*, 46(1):76–86.
- Gijbels, I., Hall, P., and Kneip, A. (1999). On the estimation of jump points in smooth curves. *Annals of the Institute of Statistical Mathematics*, 51(2):231–251.
- Gijbels, I., Hall, P., and Kneip, A. (2004). Interval and band estimation for curves with jumps. *Journal of Applied Probability*, 41(Stochastic Methods and Their Applications):65–79.
- Gruber, J. and Poterba, D. (1994). Tax incentives and the decision to purchase health insurance: Evidence from the self-employed. *Quarterly Journal of Economics*, 123(2):831–862.
- Hahn, J., Todd, P., and Van de Klaauw, W. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.
- Hall, P. and Titterton, D. M. (1992). Edge preserving and peak-preserving smoothing. *Technometrics*, 34(4):429–440.
- Kang, K.-H., Koo, J.-Y., and Park, C.-W. (2000). Kernel estimation of discontinuous regression functions. *Statistics & probability letters*, 47(3):277–285.
- Kang, Y., Gong, X., Gao, J., and Qiu, P. (2015). Jump detection in generalized error-in-variables regression with an application to Australian health tax policies. *Annals of Applied Statistics*, forthcoming.
- Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.
- Loader, C. R. (1996). Change point estimation using nonparametric regression. *The Annals of Statistics*, 24(4):1667–1678.
- Müller, H.-G. (1992). Change-points in nonparametric regression analysis. *The Annals of Statistics*, 20(2):737–761.
- Palangkaraya, A. and Yong, J. (2005). Effects of recent carrot-and-stick policy initiatives on private health insurance coverage in Australia. *Economic Record*, 81(254):262 – 272.

- Palangkaraya, A., Yong, J., Webster, E., and Dawkins, P. (2009). The income distributive implications of recent private health insurance policy reforms in australia. *The European Journal of Health Economics*, 10(1).
- Porter, J. (2003). Estimation in the regression discontinuity model. Unpublished Manuscript, Department of Economics, University of Wisconsin at Madison (2003): 5-19.
- Qiu, P. (1991). Estimation of a kind of jump regression function. *Science and Mathematical Sciences*, 4(1):1-13.
- Qiu, P. (2005). *Image Processing and Jump Regression Analysis*. John Wiley & Sons, New York.
- Qiu, P., Asano, C., and Li, X. (1991). Estimation of jump regression functions. *Bulletin of Informatics and Cybernetics*, 24(3-4):197-212.
- Rodriguez, M. and Stoyanova, A. (2004). The effect of private insurance access on the choice of gp/specialist and public/private provider in spain. *Health Economics*, 13(7):689-703.
- Stavrunova, O. and Yerokhin, O. (2014). Tax incentives and the demand for private health insurance. *Journal of Health Economics*, 34:121-130.
- Stefanski, L. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21(2):169-184.
- Stefanski, L. A. and Cook, J. R. (1995). Simulation-extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*, 90(432):1247-1256.
- Wu, J. and Chu, C. (1993a). Kernel-type estimators of jump points and values of a regression function. *The Annals of Statistics*, 21(3):1545-1566.
- Wu, J. and Chu, C. (1993b). Modification for boundary effects and jump points in nonparametric regression. *Journal of Nonparametric Statistics*, 2(4):341-354.
- Wu, J. and Chu, C. (1993c). Nonparametric function estimation and bandwidth selection for discontinuous regression functions. *Statistica Sinica*, 3(2):557-576.