

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

Xiaodong Gong (NATSEM/IGPA, U. of Canberra and Crawford ANU)

and

Jiti Gao (Monash University)

July 2015 (Seminar at Crawford School)

Motivation

The paper is motivated by solving an empirical problem: estimating the effect of Medicare Levy Surcharge (MLS) on the take-up of Private Health Insurances (PHI) from contaminated data.

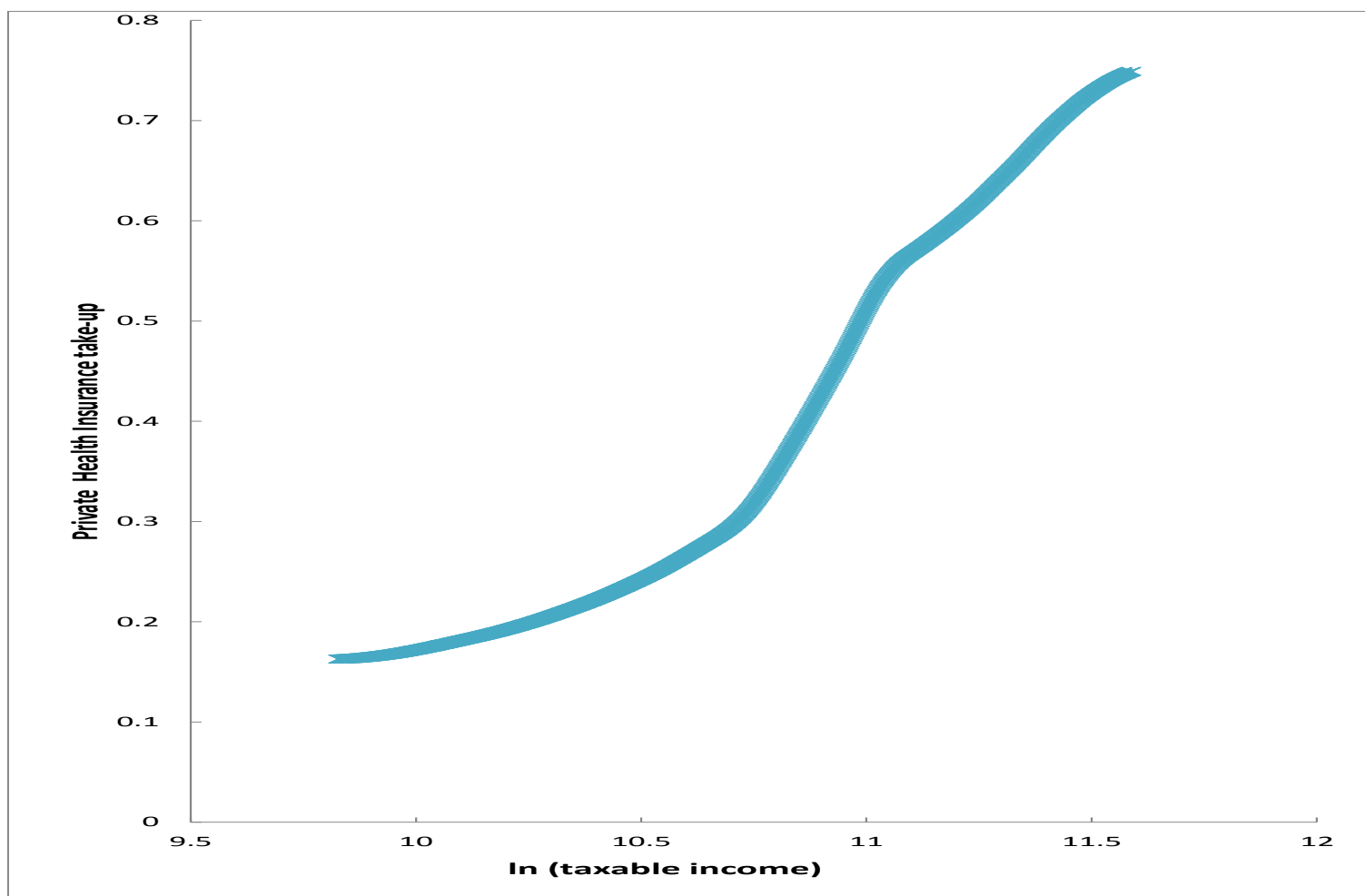


Figure 1. PHI take-up against observed \ln taxable income

Two issues:

- Function has discontinuous point(s) (probably at unknown locations);
- Data are contaminated by measurement errors.

The problem

- **The Model:**

$$Y_i = g(X_i) + \eta_i, \quad (1)$$

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

- **Observed sample:** $\{W_i, Y_i\}, i = 1, \dots, n$

$$W_i = X_i + \epsilon_i, \quad (2)$$

where ϵ_i has a known distribution.

The problem (2)

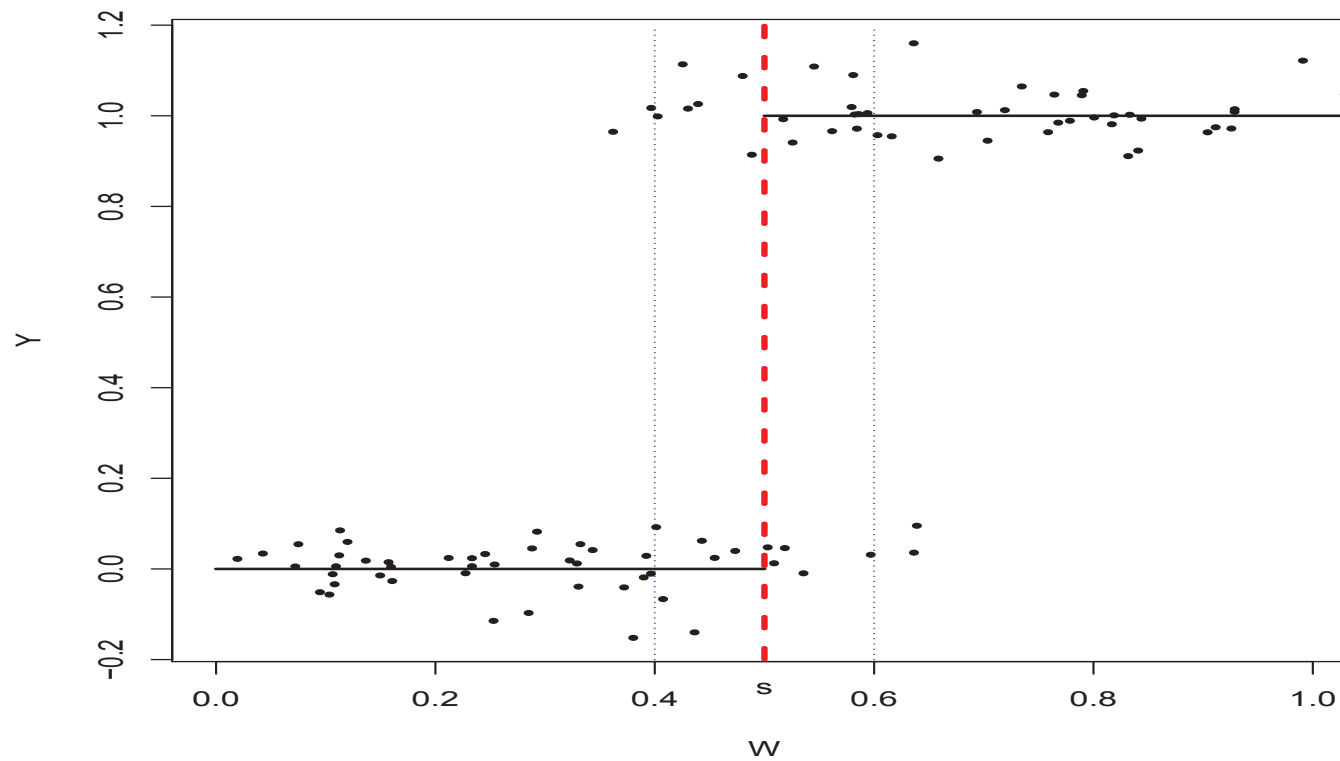


Figure 2. Illustration of the problem

Aims

- To solve our empirical problem; and in doing so,
- to develop a difference de-convolution kernel estimator of functions with discontinuities at unknown locations using contaminated data.

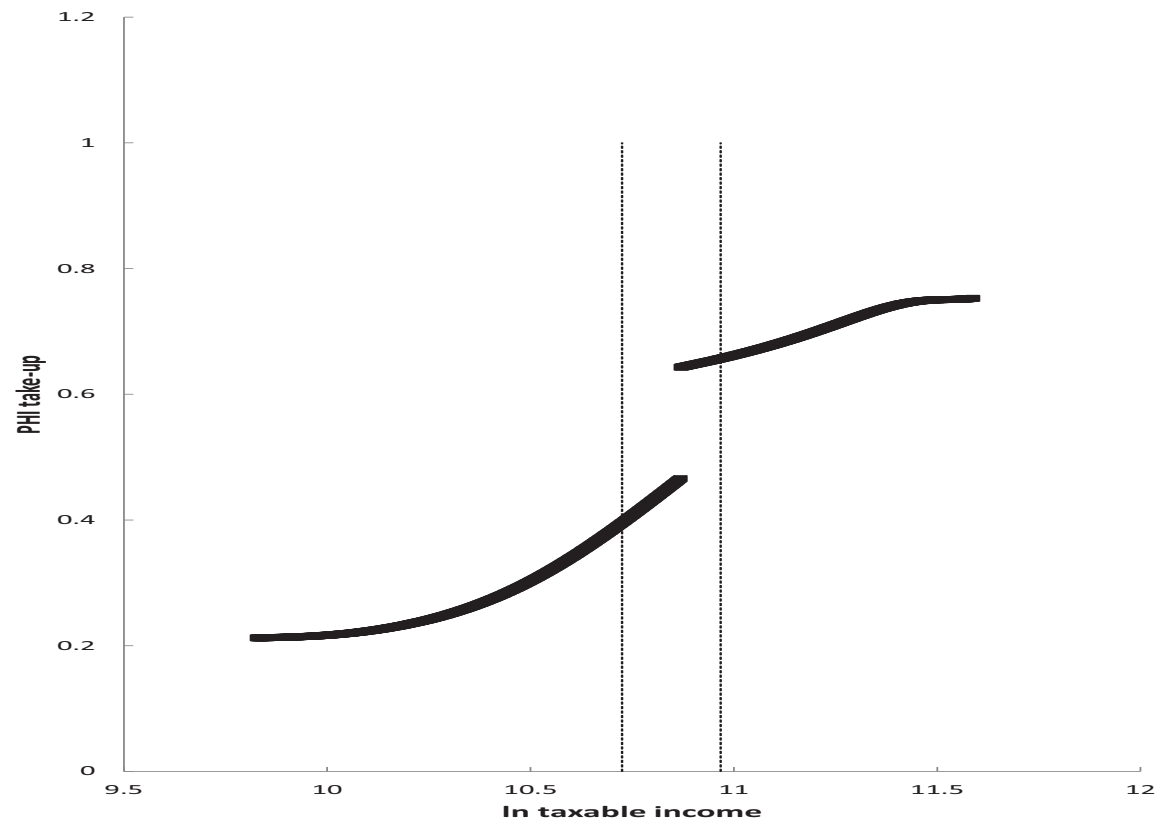


Figure 3. Function estimates with 95% confidence bands for PHI against \ln taxable income

Structure of the talk

1. Existing estimators of function discontinuities with error-free data and of continuous functions with contaminated data;
2. Our approach when the distribution of ϵ is known;
3. Some Monte Carlo results;
4. Application: estimating the effect of MLS on the take-up of PHI using contaminated data;
5. An alternative approach when the distribution of ϵ is unknown;
6. Conclusions.

Estimating functional discontinuities with error-free data

When X is observed, the unknown regression function with a discontinuity is usually estimated in two stages:

- 1) The discontinuity and its location are detected and estimated using a range of estimators proposed in a few related literatures such as change-point detection, edge detection and image reconstruction (see Qiu, 2005);
- 2) The regression curve is estimated separately for each side of the estimated discontinuity point.

Difference kernel Estimators of functional discontinuities

Diagnostic function:

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

where $\hat{g}_+(x) = \sum Y_i K_r(\frac{X_i - x}{h}) / \sum K_r(\frac{X_i - x}{h})$; and $K_r(\cdot)$, a one-sided kernel function.

Estimators of the location and size of the discontinuity:

$$\hat{s} = \arg \max_{x \in [S_0, S_1]} \hat{d}_k(x, h), \quad (4)$$

$$\hat{D} = \hat{d}_k(\hat{s}, h). \quad (5)$$

Estimating continuous functions with error-ridden regressors

With measurement errors, a continuous g can be estimated with alternative methods including the so-called deconvolution kernel estimator proposed by Fan and Truong (1993).

This estimator makes use of the property that the Fourier Transform of the convolution of two distributions is the product of Fourier Transforms of the two distributions.

Fan and Truong (1993) de-convolution estimator:

$$\hat{g}^d(x) = \sum Y_j K^*\left(\frac{x - W_j}{h_d}\right) / \sum K^*\left(\frac{x - W_j}{h_d}\right) \quad (6)$$

$$= \frac{1}{nh_d} \sum Y_j K^*\left(\frac{x - W_j}{h_d}\right) / \hat{f}_n(x) \quad (7)$$

where

$$K^*(u) = \frac{1}{2\pi} \int_{-\infty}^{+\infty} \exp^{-itu} \Phi_K(t) / \Phi_\epsilon(t/h_d) dt, \quad (8)$$

A difference de-convolution kernel estimator

Diagnostic function:

$$\hat{d}_d(x, \tilde{h}_d) = \hat{g}_r^d(x) - \hat{g}_l^d(x) \quad (9)$$

where

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

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

$$\hat{w}_i^r(x) = P\hat{r}ob\{X_i > x | W_i\} \quad (12)$$

If the location s is known, estimators of the size of the discontinuity:

$$\hat{D}_d = \hat{d}_d(s, \tilde{h}_d). \quad (13)$$

If the location is unknown, estimators of the location and size of the discontinuity:

$$\hat{s}_d = \arg \max_{x_j \in [S_0, S_1]} \hat{d}_d(x, \tilde{h}_d), \quad (14)$$

$$\hat{D}_d = \hat{d}_d(\hat{s}_d, \tilde{h}_d). \quad (15)$$

Recover the unknown function

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.

The same trick of weighting kernels of each observation i by $\hat{w}_i^r(\hat{s}_d)$ and $\hat{w}_i^l(\hat{s}_d)$ is needed, 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.

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.

The procedure can be summarised as follows:

Step 1: Estimate \hat{f}_X using \hat{f}_n in (7) from which $w_i^r(x)$ can be calculated;

Step 2: Construct $\hat{g}_r^d(x)$ (and $\hat{g}_l^d(x)$) using (10);

Step 3: Estimate \hat{s}_d and \hat{D}_d using (14) and (15), 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.

Bandwidth selection and band estimation

- Bandwidths are chosen 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 (Annals of the Institute of Statistical Mathematics, 2004).
- The confidence bands of the unknown 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 $\{\hat{s}^1, \dots, \hat{s}^R\}$.

Monte Carlo Simulations

The setting:

$$y \in \{0, 1\}$$

with

$$\text{Prob}\{y = 1|x\} = \frac{\exp(x/4 + 0.8I(x > 1.0))}{\exp(x/4 + 0.8I(x > 1.0)) + 1},$$

$$x \sim N(0, 1)$$

$$W = x + \epsilon$$

and ϵ follows a two-point distribution with $\sigma_\epsilon = .1$ or $.2$.

Results of Monte Carlo Simulations

Table 1. MISE of 200 Monte Carlo Simulations
(True Parameters: $x = 1$ and $d \approx .179$)

σ	n	Known location		Unknown location		
		Size	Function	Location	Size	Function
0.1	1,000	0.0037	0.0018	0.0355	0.0022	0.0018
0.2	1,000	0.0024	0.0012	0.0757	0.0013	0.0019
0.1	3,000	0.0013	0.0007	0.0110	0.0006	0.0008

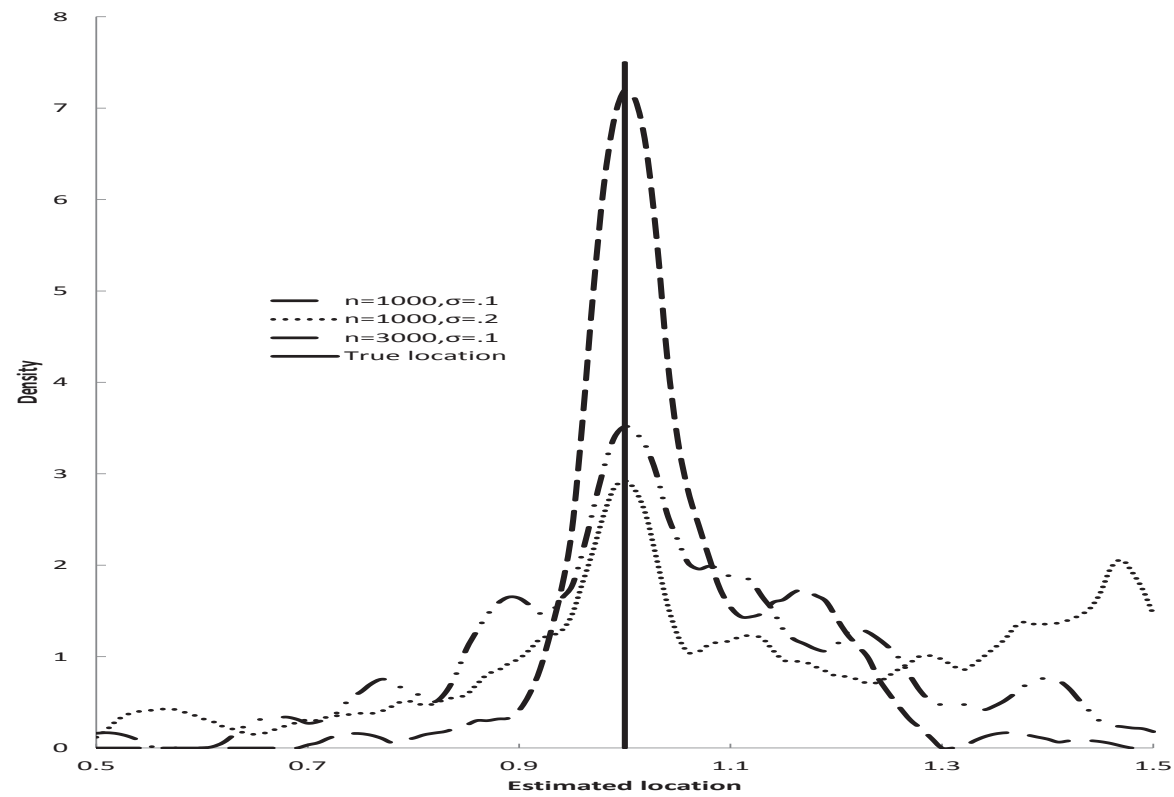


Figure 4. The distribution of the estimated discontinuity location
(200 Monte Carlos)

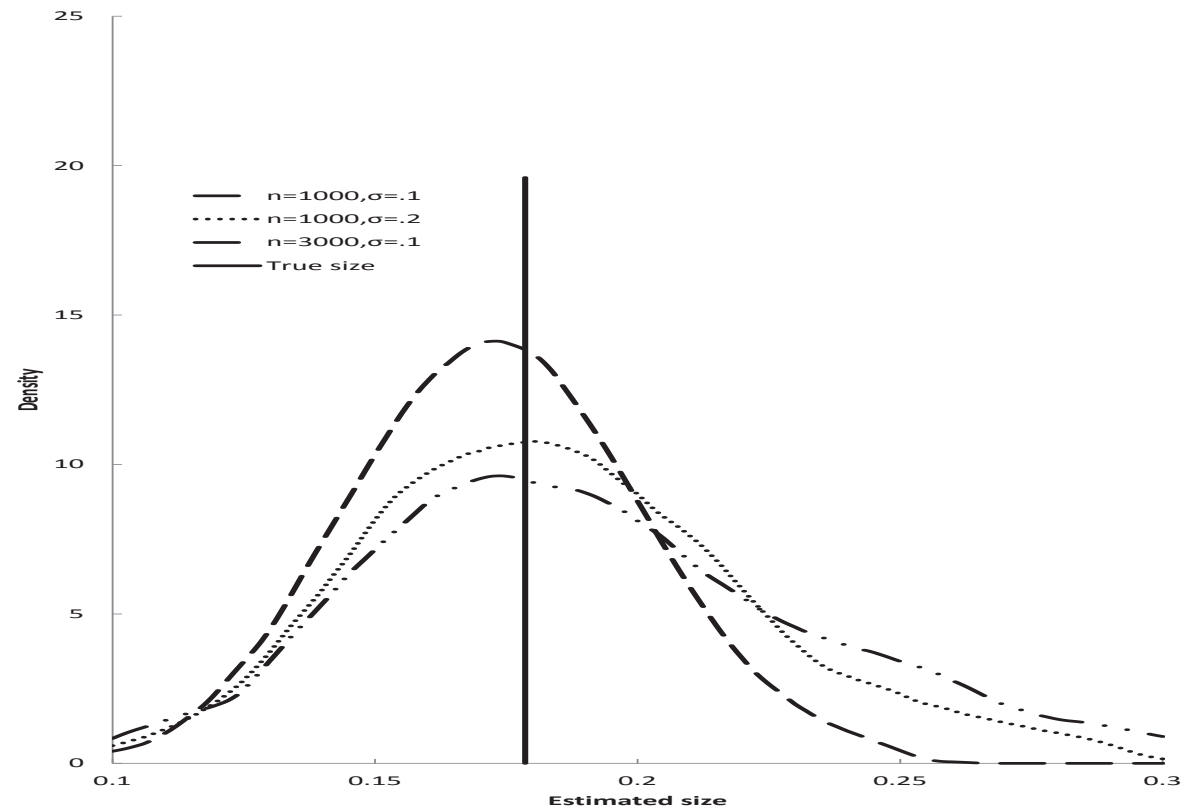


Figure 5. The distribution of the estimated discontinuity size (200 Monte Carlos)

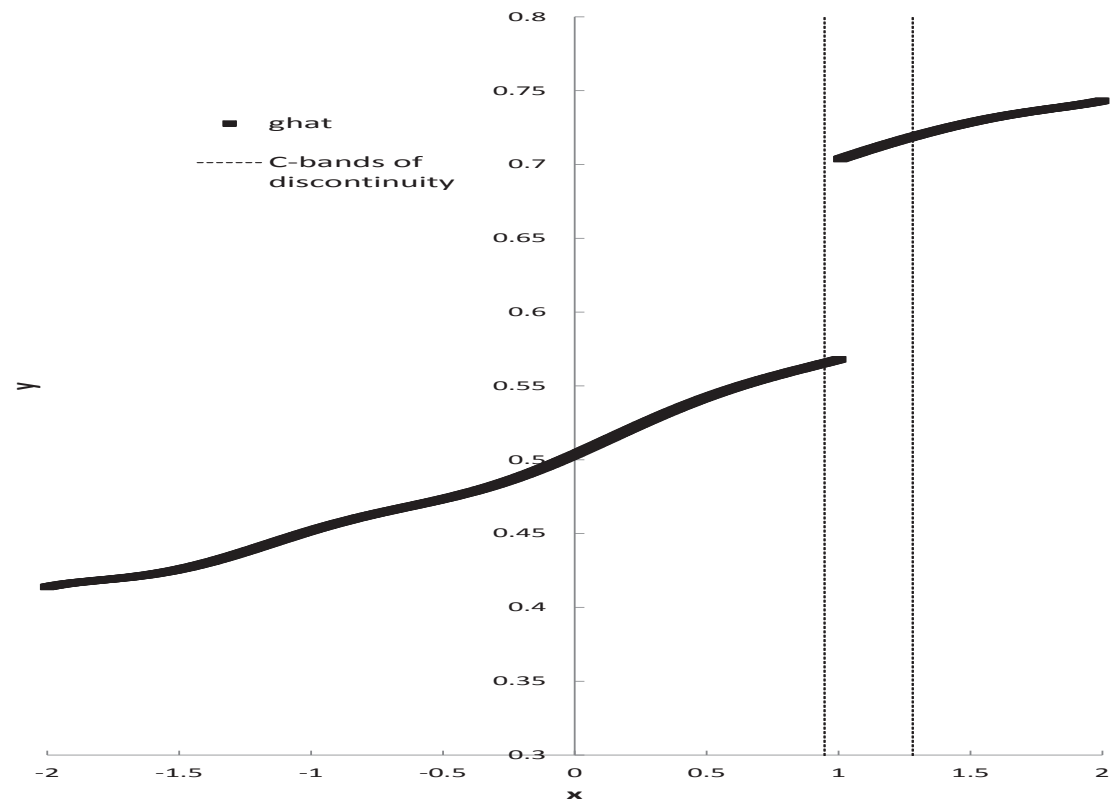


Figure 6. Estimated unknown function and its confidence bands of a simulated sample (Sample size =3,000)

Empirical application—Perturbed data

- 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 our purpose, we focus only on single males who are between 20 and 69 years of age so that they all 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;
- The final sample for analysis consists of 4,357 individuals.

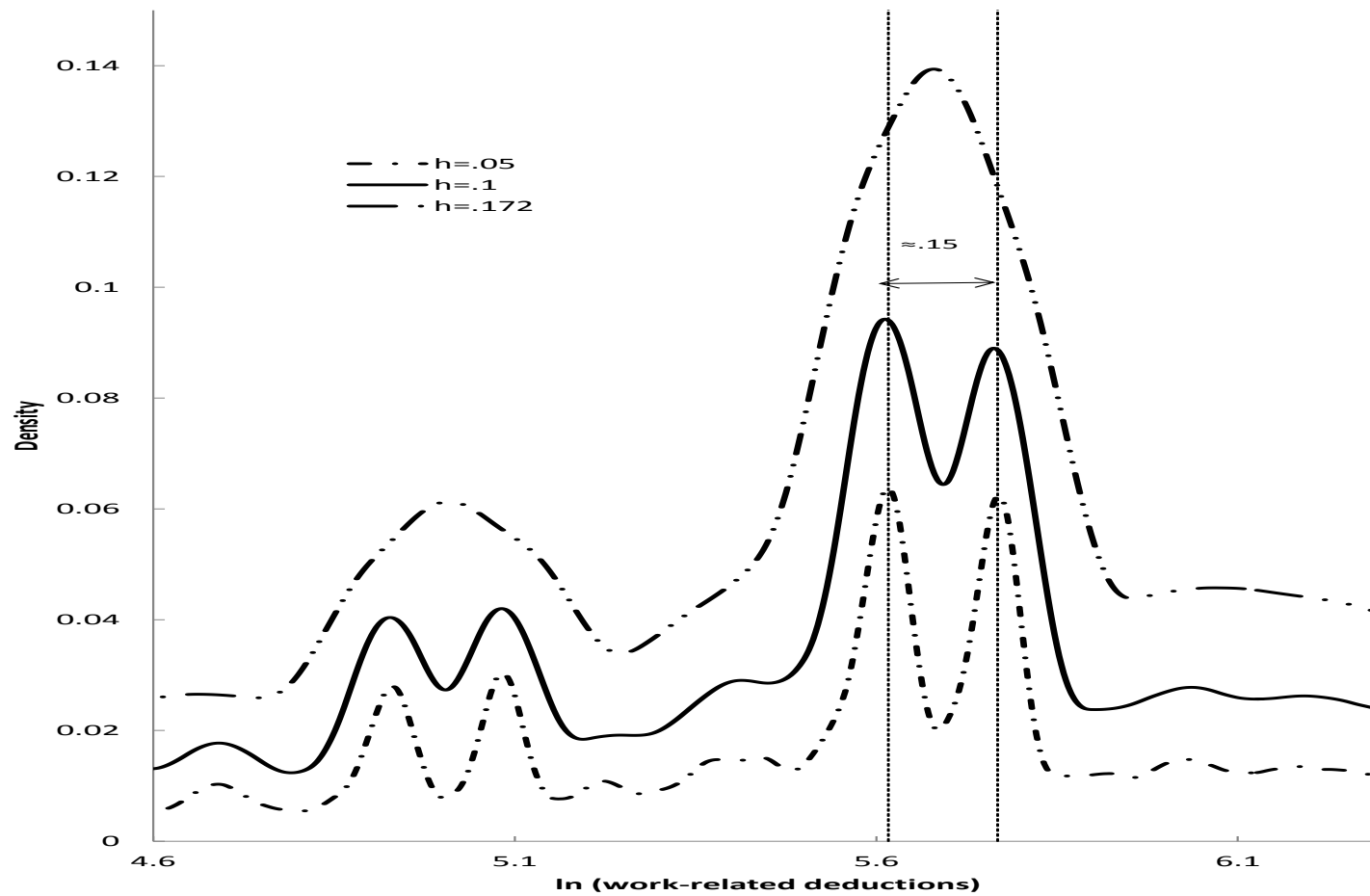


Figure 7. Estimated densities of observed work-related deductions

Empirical application—Sample statistics

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	

Empirical application—Results

Table 3. Estimates of the MLS effect at known 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.

Empirical application—Results

Table 4. Estimates of the MLS effect (unknown threshold)

	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.

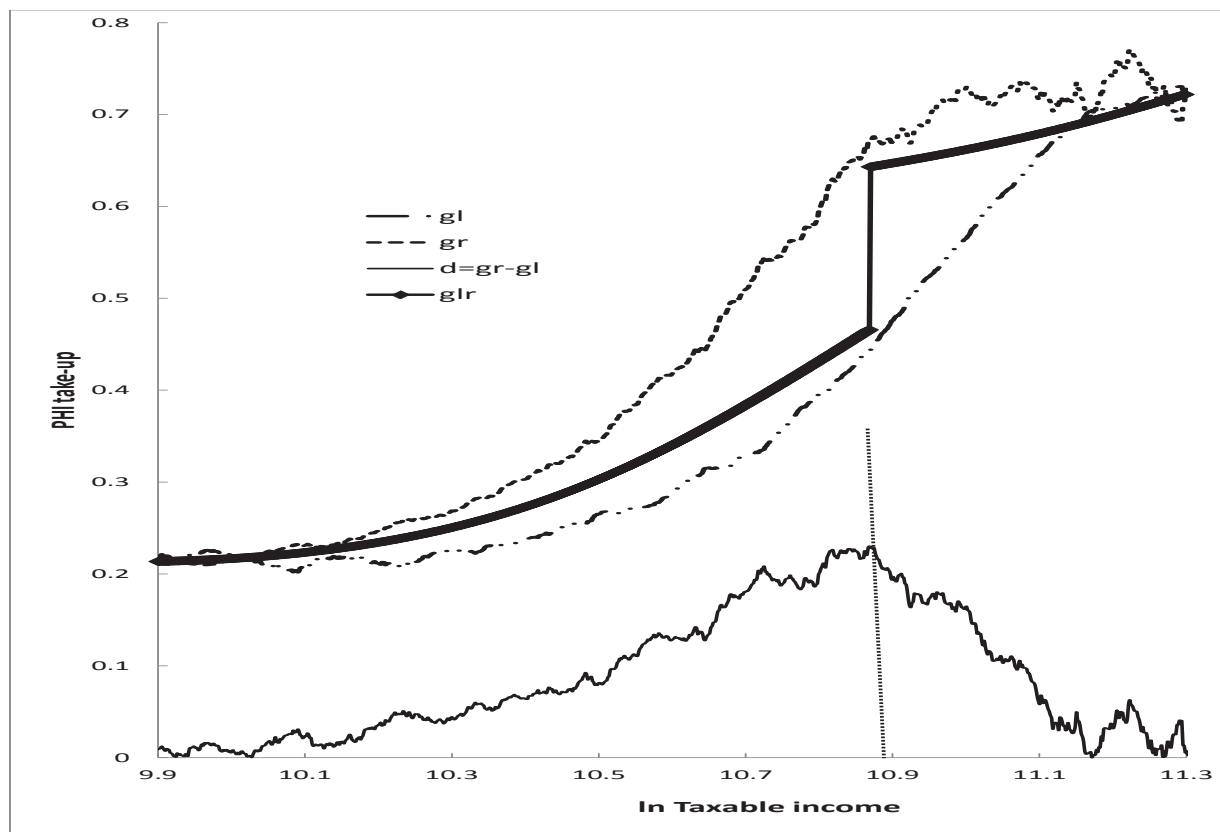


Figure 7. De-convolution kernel estimates for PHI against \ln taxable income (*glr*: two-sided; *gl*, *gr*: one-sided; $d = gr - gl$. $h_d = .070$)

**When the distribution of ϵ is unknown:
an alternative estimator (Kang, et al, 2015)**

Compare the conventional kernel estimator and a ‘one-step-right’ estimator:

$$\hat{g}_+(x) = \sum_{j=1}^n Y_j K_r \left(\frac{W_j - x}{h_n} \right) / \sum_{j=1}^n K_r \left(\frac{W_j - x}{h_n} \right)$$

$$\hat{g}_+^{n,r}(x) = \sum_{i=1}^n Y_i K_r \left(\frac{W_i - (x + h_n)}{h_n} \right) / \sum_{i=1}^n K_r \left(\frac{W_i - (x + h_n)}{h_n} \right), \quad (16)$$

A modified kernel estimator for $g_+(x)$

$$\hat{g}_+^n(x) = \frac{\sum_{i=1}^n Y_i K_r \left(\frac{W_i - x}{h_n} \right) K^q \left(\frac{|\hat{g}_+(W_i) - \hat{g}_+^{n,r}(x)|}{\rho_n} \right)}{\sum_{i=1}^n K_r \left(\frac{W_i - x}{h_n} \right) K^q \left(\frac{|\hat{g}_+(W_i) - \hat{g}_+^{n,r}(x)|}{\rho_n} \right)}, \quad (17)$$

where $\rho_n = \max_{x < W_i < x + h_n} |\hat{g}_+(W_i) - \hat{g}_+^{n,r}(x)|$;

and K^q (as K_r) is also a decreasing kernel function on $[0, 1]$.

A consistent estimator for s

- Under some regularity conditions, the estimator

$$\hat{s}_n = \arg \max_{x \in (2h_n, 1-2h_n)} \left| \hat{g}_+^n(x) - \hat{g}_-^n(x) \right|$$

is consistent: $|\hat{s}_n - s| = O(h_n)$,

and the asymptotic bias of this estimator is smaller than the estimator using conventional kernels.

- The corresponding jump magnitude D in $g(x)$ can be estimated by $\hat{D} = \hat{g}_+^n(\hat{s}_n) - \hat{g}_-^n(\hat{s}_n)$

Conclusions

- We propose a new devolution-kernel estimator for functions with discontinuities at unknown locations using contaminated data;
- The results of Monte Carlo simulations show that the estimator performs reasonably well;
- The estimator is used for estimating the effect of MLS on the take-up of PHI in Australia; A sizable MLS effect on the take-up of PHI is found at a location different from the threshold implied by the policy;
- If the distribution of the measurement error term is unknown, we propose a modified kernel estimator. Under certain regularity conditions, the estimator is consistent.