ノンパラメトリック計量経済分析：労働経済学への応用

川口　大司

一橋大学大学院経済学研究科

2014 年 6 月 14 日
日本経済学会チュートリアルセッション

労働組合参加の賃金分布への影響

Hara and Kawaguchi (2008)

同一属性を持つ労働者は労働組合に加入することでどれくらい賃金が上がるか。労働組合加入者と非加入者の観察不能な異質性は無視。

労働組合員、非労働組合員の賃金分布

- $f(w|union = 1)$ と $f(w|union = 0)$

選択すべきもの

- カーネル関数
- バンド幅

Stata のコマンド

- `kdensity`
- `dfl` インストールされていなければ `net search dfl` でインストールする。

データセット

- 日本版総合的社会調査 (JGSS) 2000-2003
- 男性、20 − 60 歳
- N=1,865, 組合員=642, 非組合員=1,223
- 時間当たり賃金、労働組合への加入の有無

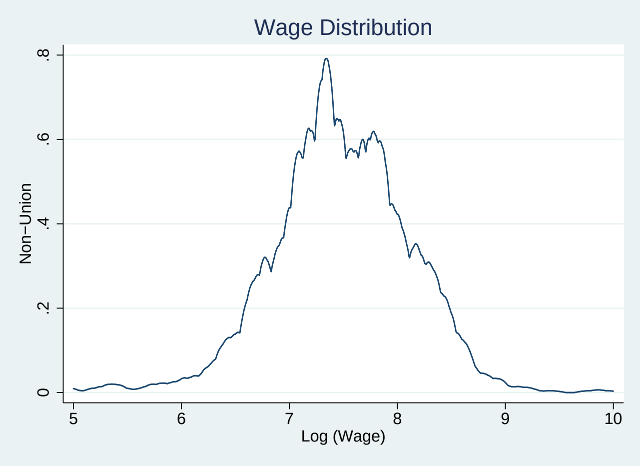
エパネチニコフカーネル、バンド幅デフォルト (0.130)


```
kdensity lnwage if !union
```

エパネチニコフカーネル、バンド幅=0.260


```
kdensity lnwage if !union, bwidth(0.260)
```

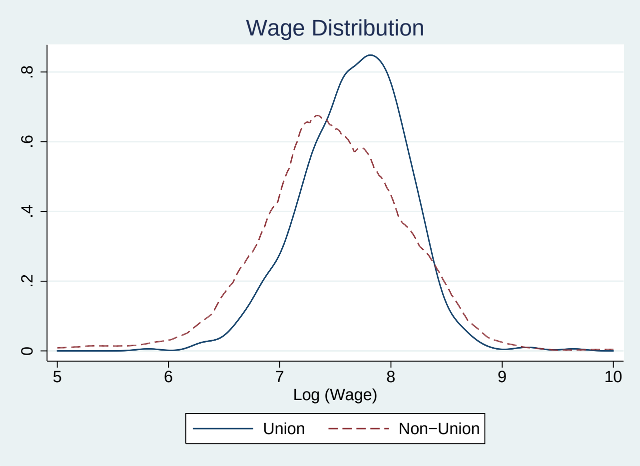
エパネチニコフカーネル、バンド幅=0.065




```
kdensity lnwage if !union, bwidth(0.065)
```

組合員と非組合員

エパネチニコフカーネル、バンド幅デフォルト




```
kdensity lnwage if !union, gaussian gen(facnu) at(disc) nograph
kdensity lnwage if union, gaussian gen(facu) at(disc) nograph
graph twoway line facu facnu cfac disc, title(Wage Distribution) xtitle(Log (Wage))
```

組合員と非組合員の属性の違い

	非組合員	組合員
教育年数	13.03	13.32
経験年数	23.20	20.18
勤続年数	12.11	15.19
結婚	0.75	0.77
子供一人	0.14	0.17
子供二人	0.35	0.37
子供三人以上	0.18	0.13

Table : 組合員と非組合員の属性の違い

組合員が非組合員になった時の仮想的賃金密度関数

組合員と非組合員の賃金密度関数

- $f(w|u=1) = \int f_1(w|x)g(x|u=1)dx$
  $f_1(w|x)$ は組合員の賃金構造。$g(x|u=1)$ は組合員の属性ベクトル（教育年数など 7 変数）の密度関数。
- $f(w|u=0) = \int f_0(w|x)g(x|u=0)dx$
  $f_1(w|x)$ 非組合員の賃金構造

賃金構造が非組合員のものに従う組合員の仮想的賃金密度

- $\int f_0(w|x)g(x|u=1)dx$
- 属性ベクトル x は 7 つ変数を含んでいるので $f_0(w|x)$ や $g(x|u=1)$ の推定は高次元ノンパラメトリック推定となり、現実的には困難。

DiNardo, Fortin and Lemieux (1996)

組合員が非組合員になった時の仮想的賃金密度関数

- $$\int f_0(w|x)g(x|u=1)dx$$
$$= \int f_0(w|x)h(x)g(x|u=0)dx,$$

$$h(x) = \frac{g(x|u=1)}{g(x|u=0)} = \frac{p(u=1|x)g(x)/p(u=1)}{p(u=0|x)g(x)/p(u=0)}$$
$$= \frac{p(u=1|x)}{1-p(u=1|x)} \frac{1-p(u=1)}{p(u=1)}.$$

- 組合員である確率 $p(u=1|x)$ をプロビット、ロジットで推定。
- この表現は非組合員の密度関数の加重推定だけを行うことになるため、高次元のノンパラメトリック推定は必要としない。

DFL の Stata を用いた実行

- 組合員である確率を x で説明

```
logit union education experience experience2
jobten jobten2 marry ...
predict propunion
```

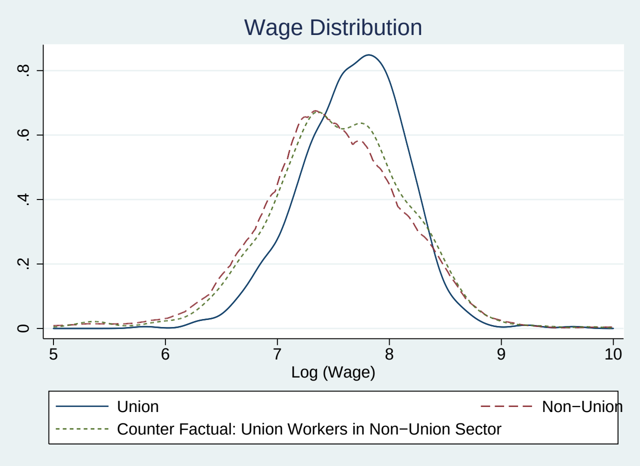
- ウエイトの計算

```
egen punion=mean(union)
gen
theta=propunion/(1-propunion)*(1-punion)/punion
```

- ウエイトつきのカーネル密度推定

```
kdensity lnwage if !union [aw=theta]
```

労働組合の賃金への影響
DFL 推定の結果



回帰切断法 (Regression Discontinuity Design) における局所線形回帰モデルの利用例

- 法定労働時間が労働時間に与えた影響の推定

$$h = \beta_0 + \beta_1 \bar{h} + covariates + u$$

- 法定労働時間の内生性
  もとより労働時間が短い国、時期、産業、事業所規模で法定労働時間が短くなる。
- $\bar{h}$ と u は正の相関関係を持つため、OLS 推定量 $\hat{\beta}_1$ には上方バイアスがかかる。
- 法定労働時間から実際の労働時間への因果関係がなくても、相関関係は生じてしまう。

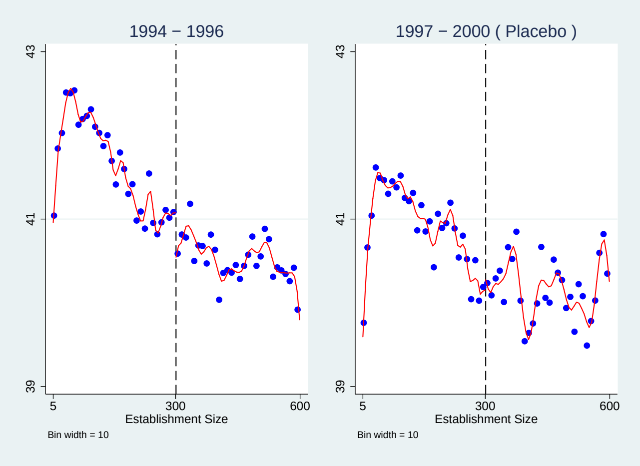
Kawaguchi, Naito and Yokoyama (2014)

- 事業所規模によって異なる法定労働時間
  1994 年から 1997 年の労働基準法による製造業の法定労働時間が事業所規模 300 人以下で 44 時間、事業所規模 301 人以上で 40 時間であったことに注目。
  300 人の事業所と 301 人の事業所の労働時間の非連続性を政策効果として捉える。
- モデルと仮定

$$h = \beta_0 + \beta_1 1(z \geq 301) + f(z) + u$$

  $E(u|z)$ と $f(z)$ は $z = 300$ において連続。
- データ
  賃金構造基本統計調査 1994-1997 と 1998-2000(プラシボ)

製造業労働者の週当たり労働時間



1994-1996 は 300 人以下 44 時間、301 人以上 40 時間

1997-2000 はすべての事業所で 40 時間（零細事業所のぞく）

局所線形回帰とバンド幅の選択

- 事業所規模 300 人と 301 人のギャップをどう推定するか?

$$h = \beta_0 + \beta_1 1(z \geq 301)$$

$$+\beta_2(z-300) \times 1(z \leq 300) + \beta_3(z-301) \times 1(z \geq 301) + u$$

  を $z \in [300-b, 301+b]$ を使って推定。
- b をバンド幅とする長方形カーネル (rectangular kernel, $1(-1 \leq (z-300)/b < 0)$ or $1(0 < (z-301)/b \leq 1))$ を使った局所線形回帰モデル。$\hat{\beta}_1$ がギャップの推定値。
- 短いバンド幅　例　$b = 0.001$
  事業所規模 300 人と事業所規模 301 人の労働時間の平均値の差。バイアスは小さいが、サンプルサイズが小さく不精確。
- 長いバンド幅　例　$b = 100$
  事業所規模と平均労働時間の関係について線形関数が良い近似ならば、サンプルサイズが大きく推定は精確。線形関数が良い近似でなければ、バイアスが大きくなる。

バンド幅とギャップの推定値

- 推定結果

バンド幅	20	30	40	50	60
ギャップ	-0.862	-0.834	-0.829	-0.645	-0.432
	(0.185)	(0.152)	(0.133)	(0.119)	(0.108)
バンド幅	70	80	90	100	
ギャップ	-0.284	-0.318	-0.403	-0.410	
	(0.101)	(0.094)	(0.089)	(0.085)	

Table : バンド幅とギャップの推定値

- ギャップの推定値はバンド幅によって変動する。
- ギャップ推定値の標準誤差はバンド幅が広がると縮小する。

クロスバリデーションによる最適バンド幅の選択

- 事業所規模別の労働時間予測

$$\hat{h}(z;b) = \begin{cases} \hat{\beta}_0 + \hat{\beta}_2(z-300) & (300-b \leq z < 300) \\ \hat{\beta}_3 + \hat{\beta}_4(z-301) & (301 < z \leq 301+b) \end{cases}$$

- バンド幅ごとの残差二乗和

$$CV(b) = \sum_{z=\underline{z}}^{\overline{z}} (h(z) - \hat{h}(z;b))^2$$

- 評価する範囲

  $\underline{z}$: 300 人以下事業所の中位数

  $\overline{z}$: 301 人以上事業所の中位数

- 最適バンド幅

$$b^* = \arg\min_{b \geq 2} CV(b)$$

最適バンド幅とギャップ推定値

- 推定結果

最適バンド幅	最適バンドでのギャップ
92	-0.404
	(0.089)

Table : 最適バンド幅とギャップ推定値

- ありうる批判
  $\underline{z}$ から $\bar{z}$ までの広い範囲の関数のフィットの良さでバンド幅を選んでいる。しかし、本当に知りたいのは $z = 300$ or $z = 301$ における最適バンド幅である。
- Imbens and Kalyanaraman (2012) や Arai and Ichimura (2014) といった解析的手法の利用。Stata における `rd` コマンドで Imbens and Kalyanaraman (2012) は求められるが、私たちの応用例では狭いバンド幅が選ばれがちであった。

参考文献

- Yoichi Arai and Hidehiko Ichimura, 2014, Simultaneous Selection of Optimal Bandwidths for the Sharp Regression Discontinuity Estimator, CIRJE F-Series CIRJE-F-927, CIRJE, Faculty of Economics, University of Tokyo.
- John DiNardo, Nicole M Fortin and Thomas Lemieux, 1996, Labor Market Institutions and the Distribution of Wages, 1973-1992: A Semiparametric Approach, *Econometrica*, Vol. 64, Issue. 5, pp. 1001-44.
- Hiromi Hara and Daiji Kawaguchi, 2008, The Union Wage Effect in Japan. *Industrial Relations*, Vol. 47, Issue 4, pp. 569-590.
- Guido Imbens and Karthik Kalyanaraman, 2012, Optimal Bandwidth Choice for the Regression Discontinuity Estimator, *Review of Economic Studies*, Vol. 79, Issue 3, pp. 933-959.
- Daiji Kawaguchi, Hisahiro Naito and Izumi Yokoyama, 2014, Assessing the Effectiveness of Reducing Standard Hours: Regression Discontinuity Evidence from Japan, Mimeograph, Hitotsubashi University.
- David S. Lee and Thomas Lemieux, 2010, Regression Discontinuity Designs in Economics, *Journal of Economic Literature*, Vol. 48, No. 2, pp. 281-355.