寄付申告と弾力性について

1　申告データを用いるときの問題点

1.1　寄付申告のコストと申告寄付額について

所得 y_i に対する課税スケジュールを $T(\cdot)$ とすると、寄付申告をしなかったときの納税額は $T(y_i)$、申告をしたときの納税額は $T(y_i, g_i)$ となる。このような設定のもと、申告を行うときに 1 をとり、行わないときに 0 をとる変数を D_i とすると

$$D_i = \begin{cases} 1 \text{ if } T(y_i) - T(y_i, g_i) \geq C_i \\ 0 \text{ if } T(y_i) - T(y_i, g_i) < C_i \end{cases} \tag{1}$$

となることがわかる。g_i, C_i を固定したもとで y_i と D_i の関係を考えると、累進課税制度より、y_i が大きくなると限界税率が増加し、所与の $g_i > 0$ に対して $T(y_i) - T(y_i, g_i)$ が増加する。

そのため、高い限界税率に直面するもとでは寄付申告を積極的に行うことが考えられる。

またこのことから、y_i が高い高所得者のほうが寄付申告を積極的に行うこともわかる。これは、実際には寄付を行っていても、高所得者のほうが寄付申告を積極的に行い、低所得者はあまり寄付申告をしないことを示唆する。

1.2　推定上の問題点

真の寄付額を g_i、サーベイデータが捉える寄付額を g_i^S、申告データが捉える寄付額を g_i^I とする。

サーベイデータも申告データも真の寄付額とは一致せず、過大ないし過小な寄付額を報告している可能性があるため、これを捉える変数として f_i^S, f_i^I を考え、$g_i = f_i^S g_i^S$、$g_i = f_i^S g_i^S$ となるとする。

真の寄付額をつかった推計モデルは

$$\ln g_i = \alpha + \beta_\tau \ln(1 - \tau_i) + \beta_y \ln y_i + \epsilon_i \tag{2}$$

となり、β_τ, β_y を正しく推定することが目的となる。

$g_i = f_i^S g_i^S$ と $g_i = f_i^I g_i^I$ を使うと、サーベイデータを使った場合、推計モデルは

$$\begin{aligned} \ln g_i^S &= \alpha + \beta_\tau \ln(1 - \tau_i) + \beta_y \ln y_i + \ln f_i^S + \epsilon_i \\ &= \alpha + \beta_\tau \ln(1 - \tau_i) + \beta_y \ln y_i + e_i \end{aligned} \tag{3}$$

となり、申告データを使った場合、推計モデルは

$$\begin{aligned} \ln g_i^I &= \alpha + \beta_\tau \ln(1 - \tau_i) + \beta_y \ln y_i + \ln f_i^I + \epsilon_i \\ &= \alpha + \beta_\tau \ln(1 - \tau_i) + \beta_y \ln y_i + u_i \end{aligned} \tag{4}$$

となる。

このとき、(3) が正しい推定を行うためには、誤差項 e_i の平均独立が言える必要があり、$\ln f_i^S$ が $\ln(1-\tau_i)$ や $\ln y_i$ と相関していないことが求められる。

同様に (4) が正しい推定を行うためには、誤差項 u_i の平均独立が言える必要があり、$\ln f_i^I$ が $\ln(1-\tau_i)$ や $\ln y_i$ と相関していないことが求められる。

しかしながら、実際に寄付を行っていても、高い限界税率に直面する者や高所得者は寄付申告をより多く行い、低所得者は寄付申告をあまり行わないということが示唆されるため、

$$corr(\ln f_i^I, \ln(1-\tau_i)) < 0$$
$$corr(\ln f_i^I, \ln y_i) > 0$$

が考えられる。そのため推定結果について $\beta_\tau < \hat{\beta_\tau}^I, \beta_y > \hat{\beta_y}^I$ となるバイアスが発生すると予想される[*1]。

他方でサーベイデータについてはこのようなバイアスがシステマティックに起こるとは考えにくい。そのため

$$\ln g_i^S = \alpha + \beta_\tau \ln(1-\tau_i) + \beta_y \ln y_i + e_i$$

を分析することによって、正しい推定結果が得られると考えられる。

1.3 今後考えるべき点

- Intensive Margin については上の議論で成り立つが、予想されるバイアスの向きがいまの結果と異なる。(サーベイデータを使った推定のほうが小さい係数値が算出される。上の議論は絶対値の話を全くしていないことに注意。)
- Extensive Margin については同様の議論が成り立つとは思うのだが、もう少し考える必要。

[*1] この計算には $\ln y$ と $\ln(1-\tau_i)$ が相関しないという仮定のもと、$\hat{\beta_\tau} = \beta_\tau + \frac{Cov(\ln(1-\tau), f_i)}{V(\ln(1-\tau_i))}$ がいえること (Wooldridge, p.81 参照) を使っている。