Generalized Roy Model and Treatment Effects

Christopher Taber

Wisconsin

February 10, 2010

Up to this point we have considered two models

- In the Roy model there is a lot of heterogeneity, but people purely maximize income
- In the Compensating Differentials model workers care about more than income, but no heterogeneity in ability

The generalized Roy model allows for heterogeneity in ability, but also allows people to care about other things

The Generalized Roy Model

In particular we relax the assumption that choice only depends on income.

Let U_{fi} and U_{ri} be the utility that individual i would receive from being a fisherman or a hunter respectively where for $j \in \{f, r\}$,

$$U_{ji} = Y_{ji} + \varphi_j(Z_i, X_{0i}) + \nu_{ji}.$$

The variable Z_i allows for the fact that there may be other variables that effect the taste for hunting versus fishing directly, but not affect wages in either sector.

Workers choose to fish when

$$U_{fi} > U_{ri}.$$

We continue to assume that

$$
\begin{aligned}
Y_{fi} &= g_f(X_{fi}, X_{0i}) + \varepsilon_{fi} \\
Y_{hi} &= g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}.
\end{aligned}
$$

Lets think about identification of this model using the same 4 steps we used for the pure Roy model

Step 1

This is identical to what we did before, notice that

$$
\begin{aligned}
0 \leq & U_{fi} - U_{hi} \\
= & Y_{fi} + \varphi_f(Z_i, X_{0i}) + \nu_{fi} - (Y_{hi} + \varphi_h(Z_i, X_{0i}) + \nu_{hi}) \\
= & g_f(X_{fi}, X_{0i}) + \varphi_f(Z_i, X_{0i}) - g_h(X_{hi}, X_{0i}) - \varphi_h(X_{hi}, X_{0i}) \\
& + \varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}
\end{aligned}
$$

We can use the same normalization as before to say $J_i = f$ when

$$\varphi(Z_i, X_i) - \nu_i \geq 0$$

and ν_i has a uniform distribution

Step 2

Getting g_f is exactly the same as before

For example

$$\begin{aligned}\lim_{\varphi(z,x)\to 1} & E(Y_i \mid X_i = x, Z_i = z, J_i = f) \\ &= g_f(x_f, x_0) + \lim_{\varphi(z,x)\to 1} E(\varepsilon_{fi} \mid X_i = x, Z_i = z, J_i = f) \\ &= g_f(x_f, x_0)\end{aligned}$$

Getting g_h is analogous, we just need to send probability of hunting to 1 rather than fishing

Step 3

Again almost completely analogous to the case before since

$$U_{fi} - U_{ri} = g_f(X_{fi}, X_{0i}) + \varphi_f(Z_i, X_{0i}) - g_h(X_{hi}, X_{0i}) - \varphi_h(Z_i, X_{0i}) \\ + \varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}$$

We need exclusion restrictions in X_{fi} **or** X_{hi} since under the assumption that the median of $\varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}$, means that if

$$Pr(J_i = f \mid X_i = x, Z_i = z) = 0.5$$

then

$$\varphi_f(z, x_0) - \varphi_h(z, x_0) = g_h(x_h, x_0) - g_f(x_f, x_0)$$

Step 4

Finally we need to identify the distribution of the error terms.

Recall that the model is

$$U_{fi} - U_{ri} = \varphi(Z_i, X_i) - \nu_i$$

$$Y_{fi} = g_f(X_{fi}, X_{0i}) + \varepsilon_{fi}$$

$$Y_{hi} = g_h(X_{hi}, X_{0i}) + \varepsilon_{hi}$$

We can use an argument identical to the standard Roy model

$$\begin{aligned} Pr(J_i = f, Y_{fi} \le y \mid (Z_i, X_i) = (z, x)) \\ &= Pr(\nu_i \le \varphi(z, x), g_f(x_f, x_0) + \varepsilon_{fi} \le y) \\ &= G_{\nu,\varepsilon_f}\left(\varphi(z, x), y - g_f(x_f, x_0)\right). \end{aligned}$$

where G_{v,ε_f} is the joint distribution of $(\varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}, \varepsilon_{fi})$.

Using an analogous argument we can get the joint distribution of

$(\varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}, \varepsilon_{hi})$

However, this is all that can be identified.

In particular the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ is not identified.

The problem is that we never observe the same person as both a hunter and a fisherman

Even with Normal error terms the covariance between ε_{fi} and ε_{hi} is not identified

In the Roy model it was identified because from the choice/wage equation we identified the joint distribution of $(\varepsilon_{fi} - \varepsilon_{hi}, \varepsilon_{hi})$ from which you can get the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$

Clearly one can not do this from the joint distribution of $(\varepsilon_{fi} + \nu_{fi} - \varepsilon_{hi} - \nu_{hi}, \varepsilon_{hi})$.

Examples

There are a ton of examples of this model in labor economics:

1. Labor Supply
2. Occupational choice
3. Schooling
4. Job Training
5. Migration
6. General treatment effects

Treatment Effects

There is a very large literature on the estimation of treatment effects.

We don't want to discuss full literature, but want to talk about how it fits into our framework

Assume the data is generated according to generalized Roy model

Define

$$\pi_i = Y_{fi} - Y_{hi}.$$

and think about identification of Average Treatment Effect

$$\begin{aligned} ATE &\equiv E(\pi_i) \\ &= E(Y_{fi}) - E(Y_{hi}). \end{aligned}$$

We focus on the "reduced form" selection model which here we define as $J_i = f$ when

$$\varphi(Z_i) + \nu_i \geq 0$$

With the one additional assumption that expectations are finite, it is trivial to show that if the Generalized Roy model is identified, the ATE is identified.

It is weaker in that you only need an exclusion restriction in the selection equation, not in the outcome equation

It is really just "identification at infinity" (+ and -) with an exclusion restriction

$$\lim_{\varphi(z)\to 1} E(Y_i \mid J_i = f, Z_i = z) - \lim_{\varphi(z)\to 0} E(Y_i \mid J_i = j, Z_i = z)$$
$$= E(Y_{fi}) - E(Y_{hi})$$
$$= ATE$$

For the most part the goal of different approaches in the treatment effect literature is to relax these assumptions in one way or another

Some focus on relaxing the support conditions

others focus on relaxing the exclusion restrictions.

First focus on relaxing the support conditions (and note that the authors don't necessarily sell it this way)

Instrumental Variables

Before thinking about Local Average Treatment Effects, it is useful to think about instrumental variables

Define

$$Y_i = \begin{cases} Y_{fi} & \text{if } J_i = f \\ Y_{ri} & \text{if } J_i = h \end{cases},$$

and letting D_{fi} be a dummy variable indicating whether $J_i = f$.

Let μ_h be the mean of Y_{hi} so $Y_{hi} = \mu_h + \varepsilon_{hi}$ then notice that

$$Y_i = Y_{hi} + D_{fi}[Y_{fi} - Y_{hi}]$$
$$= \mu_h + \pi_i D_{fi} + \varepsilon_{hi}$$

Assume that Z_i is correlated with D_{fi} but not with u_i.

Lets abstract from other regressors

IV yields

$$
\begin{aligned}
\text{plim}\widehat{\beta}_1 &= \frac{Cov(Z_i, Y_i)}{Cov(Z_i, D_{fi})} \\
&= \frac{Cov(Z_i, \mu_h + \pi_i D_{fi} + \varepsilon_{hi})}{Cov(Z_i, D_{fi})} \\
&= \frac{Cov(Z_i, \mu_h)}{Cov(Z_i, D_{fi})} + \frac{Cov(Z_i, \pi_i D_{fi})}{Cov(Z_i, D_{fi})} + \frac{Cov(Z_i, \varepsilon_{hi})}{Cov(Z_i, D_{fi})} \\
&= \frac{Cov(Z_i, \pi_i D_{fi})}{Cov(Z_i, D_{fi})}.
\end{aligned}
$$

In the case in which treatment effects are constant so that $\pi_i = \pi_0$ for everyone

$$\text{plim}\widehat{\beta}_1 = \frac{Cov(Z_i, \pi_0 D_{fi})}{Cov(Z_i, D_{fi})}$$
$$= \pi_0$$

However, more generally IV does not converge to the Average treatment effect

Local Average Treatment Effects

Imbens and Angrist (1994) consider the case in which there are not constant treatment effects

The consider a simple version of the model in which Z_i takes on 2 values, call them 0 and 1 for simplicty and without loss of generality asssume that $Pr(D_{fi} = 1 \mid Z_i = 1) > Pr(D_{fi} = 1 \mid Z_i = 0)$

There are 4 different types of people those who fish when:

1. $Z_i = 1, Z_i = 0$
2. never
3. $Z_i = 1$ only
4. $Z_i = 0$ only

Imbens and Angrist's monotonicity rules out 4 as a possibility

This is guaranteed in our index model where $J_i = f$ when

$$\varphi(Z_i) + \nu_i \geq 0$$

since we have normalized $\varphi(1) > \varphi(0)$

Imbens and Angrist do not use an index model, but since we are I will use it

Note that

$$\widehat{\beta}_1 \xrightarrow{p} \frac{Cov(Z_i, D_{fi}\pi_i)}{Cov(Z_i, D_{fi})}$$

$$= \frac{E(\pi_i D_{fi} Z_i) - E\left(\pi_i D_{fi}\right) E\left(Z_i\right)}{E(D_{fi} Z_i) - E\left(D_{fi}\right) E\left(Z_i\right)}$$

Let P_h denote the probability that $Z_i = 1$. Lets look at the pieces first the numerator

$$
\begin{aligned}
& E(\pi_i D_{fi} Z_i) - E(\pi_i D_{fi}) E(Z_i) \\
= & P_h E(\pi_i D_{fi} \mid Z_i = 1) - E(\pi_i D_{fi}) P_h \\
= & P_h E(\pi_i D_{fi} \mid Z_i = 1) \\
& - [P_h E(\pi_i D_{fi} \mid Z_i = 1) + (1 - P_h) E(\pi_i, D_{fi} \mid Z_i = 0)] P_h \\
= & P_h(1 - P_h) [E(\pi_i D_{fi} \mid Z_i = 1) - E(\pi_i D_{fi} \mid Z_i = 0)] \\
= & P_h(1 - P_h) E(\pi_i \mid \varphi(0) < \nu_i \leq \varphi(1)) Pr(\varphi(0) < \nu_i \leq \varphi(1))
\end{aligned}
$$

where the key simplification comes from the fact that

$$
\begin{aligned}
& E(\pi_i D_{fi} \mid Z_i = 1) \\
=& E\left(\pi_i 1\left(\nu_i \leq \varphi(1)\right)\right) \\
=& E\left(\pi_i \left[1\left(\nu_i \leq \varphi(0)\right) + 1\left(\varphi(0) < \nu_i \leq \varphi(1)\right)\right]\right) \\
=& E(\pi_i D_{fi} \mid Z_i = 0) \\
& \quad + E(\pi_i \mid \varphi(0) < \nu_i \leq \varphi\left(1\right)) Pr(\varphi\left(0\right) < \nu_i \leq \varphi\left(0\right)).
\end{aligned}
$$

Next consider the denominator

$$
\begin{aligned}
& E(D_{fi} Z_i) - E\left(D_{fi}\right) E\left(Z_i\right) \\
=& P_h E(D_{fi} \mid Z_i = 1) - E\left(D_{fi}\right) P_h \\
=& P_h E(D_{fi} \mid Z_i = 1) \\
& \quad - \left[P_h E(D_{fi} \mid Z_i = 1) + \left(1 - P_h\right) E(D_{fi} \mid Z_i = 0)\right] P_h \\
=& P_h(1 - P_h) \left[E(D_{fi} \mid Z_i = 1) - E(D_{fi} \mid Z_i = 0)\right] \\
=& P_h(1 - P_h) Pr(\varphi\left(0\right) < \nu_i \leq \varphi\left(1\right))
\end{aligned}
$$

Thus

$$\begin{aligned}
\widehat{\beta}_1 \xrightarrow{p} & \frac{E(\pi_i D_{fi} Z_i) - E\left(\pi_i D_{fi}\right) E\left(Z_i\right)}{E(D_{fi} Z_i) - E\left(D_{fi}\right) E\left(Z_i\right)} \\
&= \frac{P_h(1-P_h) E(\pi_i \mid \varphi(0) < \nu_i \leq \varphi(1)) Pr(\varphi(0) < \nu_i \leq \varphi(1))}{P_h(1-P_h) Pr(\varphi(0) < \nu_i \leq \varphi(1))} \\
&= E(\pi_i \mid \varphi(0) < \nu_i \leq \varphi(1))
\end{aligned}$$

They call this the local average treatment effect

Lets think about how this relates to identification at infinite and support considerations

Instead of binary suppose that Z_i had larger support.

For any two points in the support $z^{\ell}$and z^h with $\varphi(z^h) > \varphi(z^{\ell})$ we can identify$E(\pi_i \mid \varphi(z^{\ell}) < \nu_i \leq \varphi\left(z^h\right))$

Thus if I can find points so that $\varphi(z^{\ell}) = 0$ and $\varphi(z^h) = 1$, then one can identify

$$E(\pi_i \mid 0 < \nu_i \leq 1) = ATE.$$

Clearly if the support of $\varphi(Z_i)$ is bounded away from 0 or from 1, then this does not work

Local Instrumental Variables and Marginal Treatment Effects

Heckman and Vytlacil (1999, 2001, 2005) construct a framework that is useful for constructing many types of treatment effects. They focus on the marginal treatment effect defined in our context as

$$\Delta^{MTE}(x, \nu) \equiv E(\pi_i \mid X_i = x, \nu_i = \nu).$$

To see where identification comes from, note that adding regressors to the argument above, it is easy to show that one can identify

$$E(\pi_i \mid \varphi(z^\ell, x) < \nu_i \leq \varphi\left(z^h, x\right), X_i = x)$$

for any z^ℓ and z^h.

From this note that

$$\begin{aligned}\lim_{\varphi(z^\ell,x)\uparrow\nu,\varphi(z^h,x)\downarrow\nu} & E(\pi_i \mid \varphi(z^\ell,x) < \nu_i \leq \varphi\left(z^h,x\right), X_i = x) \\ &= E(\pi_i \mid X_i = x, \nu_i = \nu) \\ &= \Delta^{MTE}(x,\nu)\end{aligned}$$

They show that one can use this to build a lot of treatment effects if they are identified including the ATE

$$ATE = \int \int_0^1 \Delta^{MTE}(x_f, x_r, x_0, \nu) d\nu d\mu(x_f, x_r, x_0).$$

Theyalso show the instrumental variables estimator can be written as

$$\int_0^1 \Delta^{MTE}(x, \nu) h_{IV}(x, \nu) d\nu$$

for some function h_{IV}that they can calculate

Further they show how to use this framework to think about policy evaluation.

Relaxing the Exclusion Restriction Assumption

We next think of relaxing the exclusion restrictions.

We know of two main nonparametric alternatives in this case

Selection only on Observables

Assumption

ν is independent of $(\varepsilon_f, \varepsilon_r)$

Interestingly this is still not enough if there are values of observable covariates (X_f, X_r, X_0) for which $Pr(J = f \mid X_f, X_r, X_0) = 1$ or $Pr(J = f \mid X_f, X_r, X_0) = 0$

Thus we need the additional assumption

Assumption

For almost all x in the support of X_i,

$$0 < Pr(J = f \mid X_i = x) < 1$$

Theorem

Under assumptions 1 and 2 the Average Treatment Effect is identified

Estimation in this case is relatively straightforward. One can use matching or regression analysis to estimate the average treatment effect.

Matching

The idea of matching with data with discrete support is relatively easy

1. Take any person in the population i
   1.1 If $J_i = f$ then find another individual ℓ with $J_\ell = h$, but $X_\ell = X_i$ and define
   $$\Delta_i = Y_i - Y_\ell$$
   1.2 If $J_i = h$ then find another individual ℓ with $J_\ell = j$, but $X_\ell = X_i$ and define
   $$\Delta_i = Y_\ell - Y_i$$
2. Estimate the average treatment effect as

$$\widehat{\Delta} = \frac{1}{N} \sum_{i=1}^{N} \Delta_i$$

This is difficult to do in practice for two reasons:

1. If X_i is continuous we can't match exactly
2. If X_i is very high dimensional, even with discrete data we couldn't match directly

Propensity Score Matching

Propensity score matching is a way of getting around this problem.

Rather than matching on the high dimensional X_i we can match on the low dimensional

$$P(X) = Pr(D_{fi} = 1 \mid X_i)$$

The reason why comes from Bayes Theorem

For a set $\mathcal{X}$,

$$
\begin{aligned}
&Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho, D_{fi} = 1) \\
&\quad = \frac{Pr(D_{fi} = 1 \mid X_i \in \mathcal{X}, P(X_i) = \rho) Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho)}{Pr(D_{fi} = 1 \mid P(X_i = \rho)} \\
&\quad = \frac{\rho Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho)}{\rho} \\
&\quad = Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho)
\end{aligned}
$$

$$
\begin{aligned}
&Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho, D_{fi} = 0) \\
&\quad = \frac{Pr(D_{fi} = 0 \mid X_i \in \mathcal{X}, P(X_i) = \rho) Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho)}{Pr(D_{fi} = 0 \mid P(X_i = \rho)} \\
&\quad = \frac{(1 - \rho)\, Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho)}{(1 - \rho)} \\
&\quad = Pr(X_i \in \mathcal{X} \mid P(X_i) = \rho)
\end{aligned}
$$

Thus if we condition on the propensity score, the distribution of X_i is identical for the controls and the treatments.

But since the error term is uncorrelated with X_i

$$
\begin{aligned}
&E(Y_i \mid D_{fi} = 1, P(X_i) = \rho) - E(Y_i \mid D_{fi} = 0, P(X_i) = \rho) \\
&\qquad = E(Y_i \mid D_{fi} = 1, P(X_i) = \rho) - E(Y_i \mid D_{fi} = 0, P(X_i) = \rho) \\
&\qquad = E(Y_{fi} - Y_{hi} \mid P(X_i) = \rho) \\
&\qquad = E(\pi_i \mid P(X_i) = \rho)
\end{aligned}
$$

This means that we can match on the propensity score rather than the full set of $X's$.

You still need to deal with the continuity problem, but there are quite a few ways of doing this.

Bounds on treatment Effects

Manski and others have focused on set identification rather than point identification.

That is even if I Can not point identify π_0 is, I can Identify a set in which π_0 lies

As an example consider the instrument case in which $\varphi(z^h)$ and $\varphi(z^\ell)$ represent the upper and lower bounds of the support of $\varphi(Z_i)$.

Further assume that the support of Y_{fi} and Y_{hi} are bounded above by y^h and from below by y^ℓ. Then notice that

$$E\left(Y_{fi}\right) = E(Y_{fi} \mid J_i = f, Z_i = z^h)Pr(J_i = f \mid Z = z^h) + E(Y_{fi} \mid \nu_i > \varphi(z^h))(1 - Pr(J_i = f \mid Z = z^h))$$

We know everything here but $E(Y_{fi} \mid \nu_i > \varphi(z^h))$ which was exactly what we said couldn't be identified without identification at infinity

However we do know something about this

We know that

$$y^{\ell} \leq E(Y_{fi} \mid \nu_i > \varphi(z^h)) \leq y^h$$

Using this, one can show that

$$
\begin{aligned}
&E(Y_{fi} \mid J_i = f, Z_i = z^h)P(z^h) + y^\ell(1 - P(z^h)) \\
&\qquad - E(Y_{hi} \mid J_i = h, Z_i = z^\ell)(1 - P(z^\ell)) + y^u P(z^\ell) \\
&\quad \leq ATE \leq \\
&E(Y_{fi} \mid J_i = f, Z_i = z^h)P(z^h) + y^u(1 - P(z^h)) \\
&\qquad - E(Y_{hi} \mid J_i = h, Z_i = z^\ell)(1 - P(z^\ell)) + y^\ell P(z^\ell).
\end{aligned}
$$

where

$$
P(z) = Pr(J_i = f \mid Z_i = z)
$$

Set Estimates of Treatment Effects

There are quite a few other ways to bound treatment effects

- No assumption bounds
- Montone Treatment Effects
- Monotone Treatment Response
- Montone Selection
- Use selection on observables to bound selection on unobservables