Identification of The Roy Model

Christopher Taber

Wisconsin

January 24, 2010

Outline

Heckman and Honore

Simultaneous Equations Model

Parametric Roy Model

Nonparametric Identification of Roy Model

Outline

Heckman and Honore

Simultaneous Equations Model

Parametric Roy Model

Nonparametric Identification of Roy Model

How do we think about estimating this model?

This is discussed in Heckman and Honore (EMA, 1990)

We will follow the discussion in French and Taber fairly closely

Lets think about how we would estimate the model

Suppose we have data on Occupation and Wages from a cross section

with

$$\begin{aligned} W_F &= \pi_F F \\ W_R &= \pi_R R \end{aligned}$$

Can we identify $G(F, R)$- the joint distribution of F and R?

First a normalization is in order.

We can redefine the units of F and R arbitrarily

Lets normalize

$$\pi_F = \pi_R = 1$$

This still isn't enough in general

From the data we can observe

$$G(R \mid R > F)$$
$$G(F \mid R \leq F)$$

Lets think more generally about what identification means

Why is thinking about nonparametric identification useful?

- Speaking for myself, I think it is. I always begin a research project by thinking about nonparametric identification.
- Literature on nonparametric identification not particularly highly cited-particularly by labor economists
- At the same time this literature has had a huge impact on the field. A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days-presumably because of nonparametric identification
- It is useful for telling you what questions the data can possibly answer. If what you are interested is not nonparametrically identified, it is not obvious you should proceed with what you are doing

Definition of Identification

We follow Matzkin's (2007) formal definition of identification and follow her notation exactly

Let ς represent a model or data generating process. It is essentially a combination of parameters, functions, and distribution functions where S is the space of functions that lies.

As an example consider the semiparametric regression model

$$Y_i = X_i'\beta + \varepsilon_i$$

with

$$E(\varepsilon_i \mid X_i) = 0$$

In this case $\varsigma = (\beta, F_{X,\varepsilon})$ where $F_{X,\varepsilon}$ is the joint distribution between X_i and ε_i

S is the set of permissible β and $F_{X,\varepsilon}$

The data we can potentially observe is the full joint distribution of (Y_i, X_i)

Define

$$\Gamma_{Y,X}(\psi, S) = \{F_{Y,X}(\cdot;\varsigma) \mid \varsigma \in S \text{ and } \Psi(\varsigma) = \psi\}.$$

$\psi^* \in \Omega$ is identified in the model S if for any $\psi \in \Omega$,

$$\left[\Gamma_{Y,X}(\psi, S) \cap \Gamma_{Y,X}(\psi^*, S)\right] = \emptyset$$

So what the heck does that mean?

Basically $\Psi(\varsigma)$. Measures some feature of the model.

Interesting examples in our case are:

- $\Psi(\varsigma) = \varsigma$
- $\Psi(\varsigma) = \beta$
- $\Psi(\varsigma) = \beta_1$

From this, $\Gamma_{Y,X}(\psi, S)$ is the set of possible data distributions that are consistent with the model and a given value $\Psi(\varsigma) = \psi$

$\psi^* \in \Omega$ is identified when there is no other value of ψ that is consistent with the joint distribution of the data

To see an example consider the regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + \varepsilon_i$$

One key aspect of the data is

$$E(Y_i \mid X_i) = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i}$$

As long as there is not perfect multicollinearity, for any β^* and β, we can find values of (X_{1i}, X_{2i}) so that

$$x'\beta^* \neq x'\beta$$

Thus two different values of β lead to different joint distribution of (X_i, Y_i) (i.e. $F_{Y,X}(\cdot; \beta, F_{X,\varepsilon})$) thus the model is identified

However next consider the case in which

$$X_{1i} + X_{2i} = 1.$$

For a particular value of $\beta^* = (\beta_0^*, \beta_1^*, \beta_2^*)$ take some other value $\beta = (\beta_0^* + \beta_2^*, \beta_1^* - \beta_2^*, 0)$. Then for any value of the data

$$\begin{aligned} E(Y_i \mid X_i = x) &= \beta_0^* + \beta_1^* X_{1i} + \beta_2^* X_{2i} \\ &= \beta_0^* + \beta_1^* X_{1i} + \beta_2^* (1 - X_{2i}) \\ &= \beta_0^* + \beta_2^* + (\beta_1^* - \beta_2^*) X_{1i} \\ &= \beta_0 + \beta_1 X_{1i}. \end{aligned}$$

Thus the model is not identified because $F_{Y,X}(\cdot; \beta^*, F_{X,\varepsilon})$ is identical to $F_{Y,X}(\cdot; \beta, F_{X,\varepsilon})$

Lets use this to think about the Roy model

Outline

Heckman and Honore

Simultaneous Equations Model

Parametric Roy Model

Nonparametric Identification of Roy Model

Before studying the Roy model I want to take a bit of a detour and talk about identification in the simultaneous equations model

$$
\begin{aligned}
Y_{1i} &= \alpha_1 Y_{2i} + X_i' \beta_1 + Z_i' \gamma + u_{1i} \\
Y_{2i} &= \alpha_2 Y_{1i} + X_i' \beta_2 + W' \delta + u_{2i}
\end{aligned}
$$

These are called the "structural equations"

Note the difference between X_i, Z_i, and W_i in that we restrict what can affect what.

We assume that

$$
\begin{aligned}
E(u_{1i} \mid X_i, Z_i, W_i) = 0 \\
E(u_{2i} \mid X_i, Z_i, W_i) = 0
\end{aligned}
$$

but notice that almost for sure Y_{2i} is correlated with u_{1i} because u_{1i} influences Y_{2i} through Y_{1i}

It is useful to calculate the "reduced form" for $E(u_{1i} \mid X_i, Z_i, W_i)$, namely

$$\begin{aligned}
Y_{1i} &= \alpha_1 Y_{2i} + X_i'\beta_1 + Z_i'\gamma + u_{1i} \\
&= \alpha_1 \left[\alpha_2 Y_{1i} + X_i'\beta_2 + W'\delta + u_{2i}\right] + X_i'\beta_1 + Z_i'\gamma + u_{1i} \\
&= \alpha_1\alpha_2 Y_{1i} + X_i'\left[\alpha_1\beta_2 + \beta_1\right] + W'\alpha_1\delta + Z_i'\gamma + (\alpha_1 u_{2i} + u_{1i}) \\
&= X_i'\frac{\alpha_1\beta_2 + \beta_1}{1 - \alpha_1\alpha_2} + W'\frac{\alpha_1\delta}{1 - \alpha_1\alpha_2} + Z_i'\frac{\gamma}{1 - \alpha_1\alpha_2} + \frac{\alpha_1 u_{2i} + u_{1i}}{1 - \alpha_1\alpha_2} \\
&= X_i'\beta_1^* + W_i'\delta_1^* + Z_i'\gamma_1^* + u_{1i}^*
\end{aligned}$$

where

$$\begin{aligned}
\beta_1^* &= \frac{\alpha_1\beta_2 + \beta_1}{1 - \alpha_1\alpha_2} \\
\delta_1^* &= \frac{\alpha_1\delta}{1 - \alpha_1\alpha_2} \\
\gamma_1^* &= \frac{\gamma}{1 - \alpha_1\alpha_2}
\end{aligned}$$

Note that $E(u^*_{1i} \mid X_i, Z_i, W_i) = 0$, so one can identify β^*_1, δ^*_1, and γ^*_1.

This is called the “reduced form” equations.

Note that the parameters here are not the fundamental structural parameters, but the are a known function of these parameters

This model is identified if we have “exclusion restrictions”

That is we can identify α_2 as long as we have some Z_i

I want to show this three different ways

Method 1:

The model is completely symmetric which means we can identify the reduced form model for Y_{2i} which we can write as

$$Y_{2i}^* = X_i'\beta_2^* + W_i'\delta_2^* + Z_i'\gamma_2^* + u_{2i}^*$$

where

$$\beta_2^* = \frac{\alpha_2\beta_1 + \beta_2}{1 - \alpha_1\alpha_2}$$
$$\delta_2^* = \frac{\delta}{1 - \alpha_1\alpha_2}$$
$$\gamma_2^* = \frac{\alpha_2\gamma}{1 - \alpha_1\alpha_2}$$

Notice then that

$$\frac{\gamma_2^*}{\gamma_1^*} = \alpha_2$$

Since this is the coefficient on Z_i, you need a Z_i to do this.

Method 2:

Define

$$\widehat{Y_{1i}} \equiv X_i'\beta_1^* + W_i'\delta_1^* + Z_i'\gamma_1^*$$

we have shown that this is identified

Now notice that

$$\begin{aligned} Y_{2i} &= \alpha_2 Y_{1i} + X_i'\beta_2 + W_i'\delta + u_{2i} \\ &= \alpha_2 \left[\widehat{Y}_{1i} + u_{1i}^*\right] + X_i'\beta_2 + W_i'\delta + u_{2i} \\ &= \alpha_2 \widehat{Y}_{1i} + X_i'\beta_2 + W_i'\delta + (\alpha_2 u_{1i}^* + u_{2i}) \end{aligned}$$

One can get a consistent estimate of α_2 by regressing Y_{2i} on X_i, W_i and $\widehat{Y}_{1i}$.

To do this one needs to be able to vary $\widehat{Y}_{1i}$ separately from X_i and W_i which can only be done if there is a Z_i

This is the intuition behind two staged least squares

Method 3:

Just think about the IV problem.

$$Y_{2i} = \alpha_2 Y_{1i} + X_i'\beta_2 + W'\delta + u_{2i}$$

One can use Z_i as an instrument for Y_{1i} because

- It is correlated with Y_{1i} since it directly influences it
- It is uncorrelated with u_{2i}

so the model is identified if we have an instrument

Estimation

If we think about the empirical implementation of these, they are all identical in the "just identified" case

That is if Z_i is one dimensional you will get numerically the same answer

When Z_i is more than one dimensional, one would need to think about how to implement method 1 or method 3

Outline

Heckman and Honore

Simultaneous Equations Model

Parametric Roy Model

Nonparametric Identification of Roy Model

Lets return to the Roy model

Before thinking about nonparametric identification, lets think about parametric estimation

If you understand that, it will turn out that the nonparametric identification is analogous.

French and Taber focus on the labor supply case, and we will as well

That is let

$$Y_{fi} = X_{0i}'\gamma_{0f} + X_{fi}'\gamma_{ff} + \varepsilon_{fi}$$
$$Y_{hi} = X_{0i}'\gamma_{0h} + X_{hi}'\gamma_{hh} + \varepsilon_{hi}$$
$$\begin{bmatrix} \varepsilon_{fi} \\ \varepsilon_{hi} \end{bmatrix} = N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma_f^2 & \sigma_{fh} \\ \sigma_{fh} & \sigma_h^2 \end{bmatrix}\right).$$

But we will take Y_{fi} to be market production and Y_{hi} to be market production so the individual works if $Y_{fi} > Y_{hi}$

The econometrician gets to observe whether the individual works, and if they work they observe the wage

The key distinction between this and the more general Roy model is that the econometrician does not observe Y_{hi} for people who do not work (which seems reasonable in the labor supply problem)

We can estimate this model in 4 steps:

Step 1: Estimation of Choice Model

The probability of choosing is:

$$
\begin{aligned}
\Pr(J_i = f \mid X_i = x) &= \Pr(Y_{fi} > Y_{hi} \mid X_i = x) \\
&= \Pr\left(x_0'\gamma_{0f} + x_f'\gamma_{ff} + \varepsilon_{fi} > x_0'\gamma_{0h} + x_h'\gamma_{hh} + \varepsilon_{hi}\right) \\
&= \Pr\left(x_0'(\gamma_{0f} - \gamma_{0h}) + x_f'\gamma_{ff} - x_h'\gamma_{hh} > \varepsilon_{hi} - \varepsilon_{fi}\right) \\
&= \Phi\left(\frac{x_0'(\gamma_{0f} - \gamma_{0h}) + x_f'\gamma_{ff} - x_h'\gamma_{hh}}{\sigma^*}\right) \\
&= \Phi\left(x'\gamma^*\right)
\end{aligned}
$$

where Φ is the cdf of a standard normal, σ^* is the standard deviation of $(\varepsilon_{hi} - \varepsilon_{fi})$ (recall that if $\varepsilon_{fi}, \varepsilon_{hi}$ normal, then $(\varepsilon_{hi} - \varepsilon_{fi})$ normal) and

$$
\gamma^* \equiv \left(\frac{\gamma_{0f} - \gamma_{0h}}{\sigma^*}, \frac{\gamma_{ff}}{\sigma^*}, \frac{-\gamma_{hh}}{\sigma^*}\right).
$$

From the choice model alone we can only identify γ^* a

This is referred to as the "reduced form probit"

can be estimated by maximum likelihood as a probit model

Let $\widehat{\gamma^*}$ represent the estimated parameter.

Step 2: Estimating the Wage Equation

This is essentially the second stage of a Heckman two step. To review the idea behind that, let

$$\varepsilon_i^* = \frac{\varepsilon_{hi} - \varepsilon_{fi}}{\sigma^*}$$

Then consider the regression

$$\varepsilon_{fi} = \tau \varepsilon_i^* + \zeta_i$$

where $cov\left(\varepsilon_i^*, \zeta_i\right) = 0$ (by definition of regression) and thus:

$$\begin{aligned}
\tau &= \frac{cov\left(\varepsilon_{fi}, \varepsilon_i^*\right)}{var\left(\varepsilon_i^*\right)} \\
&= E\left[\varepsilon_{fi}\left(\frac{\varepsilon_{fi} - \varepsilon_{fi}}{\sigma^*}\right)\right] \\
&= \frac{\sigma_f^2 - \sigma_{fh}}{\sigma^*}
\end{aligned}$$

Now notice that

$$\begin{aligned}
E\left(Y_i \mid J_i = f, X_i = x\right) &= x_0'\gamma_{0f} + x_f'\gamma_{ff} + E\left(\varepsilon_{fi} \mid J_i = f, X_i = x\right) \\
&= x_0'\gamma_{0f} + x_f'\gamma_{ff} + E\left(\tau\varepsilon_{fi}^* + \zeta_i \mid \varepsilon_i^* > x'\gamma^*\right) \\
&= x_0'\gamma_{0f} + x_f'\gamma_{ff} + \tau E\left(\varepsilon_{fi}^* \mid \varepsilon_i^* > x'\gamma^*\right) \\
&= x_0'\gamma_{0f} + x_f'\gamma_{ff} + \tau\lambda\left(x'\gamma^*\right)
\end{aligned}$$

where $\lambda\left(x'\gamma^*\right) = \frac{\phi(x'\gamma^*)}{(1-\Phi(x'\gamma^*))}$ is the inverse Mills ratio.

OLS of Y_i on X_{0i}, X_{fi}, and $\lambda\left(X_i'\widehat{\gamma^*}\right)$ gives consistent estimates of γ_{0f}, γ_{ff}, and τ

Since λ is a nonlinear function we don't have to have an exclusion restriction

Step 3: The Structural Probit

Our next goal is to estimate γ_{0h} and γ_{hh}. Note that at this point we have shown how to obtain consistent estimates of

$$\gamma^* \equiv \left(\frac{\gamma_{0f} - \gamma_{0h}}{\sigma^*}, \frac{\gamma_{ff}}{\sigma^*}, \frac{-\gamma_{hh}}{\sigma^*} \right)$$

But from the Heckman Two step we got a consistent estimates of γ_{0f} and γ_{ff}

Thus analogous to Method 1 above, as long as we have an exclusion restriction X_{fi} we can identify σ^*

Once we have σ^* it is easy to see how to identify γ_{hh} and γ_{0h}

In terms of estimation of these objects the typical way is like the second step described above.

We can estimate the "structural probit:"

$$Pr(J_i = f \mid X_i = x) = \Phi\left(\frac{1}{\sigma^*}\left(x_0'\gamma_{0f} + x_f'\gamma_{ff}\right) - x_0'\frac{\gamma_{0h}}{\sigma^*} - x_h'\frac{\gamma_{hh}}{\sigma^*}\right). \tag{1}$$

That is one just runs a probit of J_i on $\left(X_{0i}'\widehat{\gamma_{0f}} + X_{fi}'\widehat{\gamma_{ff}}\right)$, X_{0i}, and X_{hi}.

Again for identification we need an X_{fi}

Step 4: Identification of the Variance Covariance Matrix of the Residuals

Lastly, we identify all the components of Σ, $(\sigma_f^2, \sigma_h^2, \sigma_{fh})$ as follows. First, we have identified $(\sigma^*)^2 = \sigma_f^2 + \sigma_h^2 - \sigma_{fh}^2$. Second, we have identified $\tau = \frac{\sigma_f^2 - \sigma_{fh}}{\sigma^*}$. This gives us two equations in three parameters. We can obtain the final equation by using the variance of the residual in the selection model since as Heckman and Honore point out

$$\sigma_f^2 = Var(Y_i \mid J_i = f, X_i = x) - \tau^2 \left(\lambda(x'\gamma^*)x'\gamma^* - \lambda^2(x'\gamma^*) \right)$$

$$\sigma_{fh} = \sigma_f^2 - \tau\sigma^*$$

$$\sigma_h^2 = \sigma^{*2} - \sigma_f^2 + 2\sigma_{fh}.$$

Outline

Heckman and Honore

Simultaneous Equations Model

Parametric Roy Model

Nonparametric Identification of Roy Model

Nonparametric Identification

Next we consider nonparametric identification of the Roy model

We consider the model

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

where the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ is G.

Assumption

$X_i = (X_{0i}, X_{fi}, X_{hi})$ can be written as $(X_{0i}^c, X_{0i}^d, X_{fi}^c, X_{fi}^d, X_{hi}^c, X_{hi}^d)$ where the elements of $(X_{0i}^c, X_{fi}^c, X_{hi}^c)$ are continuously distributed (no point has positive mass), and $(X_{0i}^d, X_{fi}^d, X_{hi}^d)$ is distributed discretely (all support points have positive mass).

Assumption

For any $(x_0^d, x_f^d, x_h^d) \in supp(X_{0i}^d, X_{fi}^d, X_{hi}^d)$, $g_f(x_f^c, x_f^d, x_0^c, x_0^d)$ *and* $g_h(x_h^c, x_h^d, x_0^c, x_0^d)$ *are almost surely continuous across* $x^c \in supp(X_i^c \mid X_i^d = x^d)$.

Assumption

$(\varepsilon_{fi}, \varepsilon_{hi})$ *is continuously distributed with distribution function G, support* $\mathbb{R}^2$*, and is independent of* X_i*.*

Assumption

$supp(g_f(X_{fi}, x_0), g_h(X_{hi}, x_0)) = \mathbb{R}^2$ *for all* $x_0 \in supp(X_{0i})$*.*

Assumption

The marginal distributions of ε_{fi} *and* $\varepsilon_{fi} - \varepsilon_{hi}$ *have medians equal to zero.*

Theorem

If $(J_i \in \{f, h\}$, Y_{fi} *if* $J_i = f$, $X_i)$ *are all observed and generated under the Roy model, under these assumptions,* g_f, g_h, *and* G *are identified on a set* $\mathcal{X}^*$ *that has measure 1.*

Step 1: Identification of Choice Model

This part is well known in a number of papers (Manski and Matzkin being the main contributors) We can write the model as

$$\begin{aligned} Pr(J_i = f \mid X_i = x) &= Pr(\varepsilon_{ih} - \varepsilon_{if} \leq g_f(x_f, x_0) - g_h(x_h, x_0)) \\ &= G_{h-f}(g^*(x)), \end{aligned}$$

where G_{h-f} is the distribution function for $\varepsilon_{ih} - \varepsilon_{if}$ and $g^*(x) \equiv g_f(x_f, x_0) - g_h(x_h, x_0)$.

Given data only on choices, the model is only identified up to a monotonic transformation. Let M be any strictly increasing function, then

$$g^*(X_i) \geq \varepsilon_{ih} - \varepsilon_{if}$$

if and only if

$$M(g^*(X_i)) \geq M(\varepsilon_{ih} - \varepsilon_{if}).$$

A very convenient normalization is to choose the uniform distribution for $\varepsilon_{ih} - \varepsilon_{if}$.

Note that for any random variable ε with cdf F,

$$\begin{aligned} F(x) &\equiv Pr(\varepsilon \leq x) \\ &= Pr(F(\varepsilon) \leq F(x)) \end{aligned}$$

Thus $F(\varepsilon)$ has a uniform distribution.

This is a really nice normalization, we let M be G_{h-r} and define

$$\begin{aligned}\widehat{\varepsilon}_i &= G_{h-r}(\varepsilon_{ih} - \varepsilon_{if}) \\ \widehat{g}^*(x)) &= G_{h-r}(g^*(x))\end{aligned}$$

Then

$$\begin{aligned}Pr(J_i = f \mid X_i - x) &= Pr(\varepsilon_{ih} - \varepsilon_{if} < g^*(x)) \\ &= Pr(G_{h-r}(\varepsilon_{ih} - \varepsilon_{if}) < G_{h-r}g^*(x)) \\ &= Pr(\widehat{\varepsilon}_i < \widehat{g}^*(x)) \\ &= \widehat{g}^*(x).\end{aligned}$$

Thus we have thus established that we can write the model as $J_i = f$ if and only if $\widehat{g}^*(X_i) > \widehat{\varepsilon}_i$ where $\widehat{\varepsilon}_i$ is uniform $(0, 1)$ and that $\widehat{g}^*$ is identified.

Step 2: Identification of the Wage Equation g_f

Next consider identification of $g_{f.}$ This is basically the standard selection problem.

Notice that we can identify the distribution of Y_f conditional on $(X_i = x, J_i = f.)$

In particular we can identify

$$\begin{aligned} Med(Y_i \mid X_i = x, J_i = f) = & g_f(x_f, x_0) \\ & + Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x)). \end{aligned}$$

An exclusion restriction is key, we need a variable x_h that allows us move (x_f, x_0) holding $\widehat{g}^*(x)$ and thus $Med(\varepsilon_{fi} \mid X_i = x, \widehat{\varepsilon}_i < \widehat{g}^*(x))$ fixed.

By holding $\widehat{g}^*(x)$ fixed and varying (x_h, x_0) we can identify g_h

Identification at Infinity

What about the location?

Notice that

$$\begin{aligned}
&\lim_{\widehat{g}^*(x)\to 1} Med(Y_f \mid X_i = x, J = f) \\
&\qquad = g_f(x_f, x_0) + \lim_{\widehat{g}^*(x)\to 1} Med(\varepsilon_{fi} \mid \widehat{\varepsilon}_i < \widehat{g}^*(x)) \\
&\qquad = g_f(x_f, x_0) + Med(\varepsilon_f \mid \widehat{\varepsilon} < 1) \\
&\qquad = g_f(x_f, x_0) + Med(\varepsilon_f) \\
&\qquad = g_f(x_f, x_0).
\end{aligned}$$

Thus we are done.

Another important point we want to make is that the model is not identified without identification at infinity.

To see why suppose that $\widehat{g}^*(X_{fi}, X_{ri}, X_{0i}))$ is bounded from above at g^u then if $\widehat{\varepsilon}_i > g^u$, $J_i = r$. Thus the data is completely uninformative about

$$E(Y_{fi} \mid \widehat{\varepsilon}_i > g^u)$$

so the model is not identified.

Parametric assumptions on the distribution of the error term is an alternative.

Really this is the same point as in the regression example we talk about to undergraduates-you can not predict outside the range of the data.

Whether it is a big deal or not depends on the question of interest.

Step 3: Identification of g_h

What will be crucial is the other exclusion restriction (i.e. X_{fi}).

For any (x_r, x_0) we want to find an x_f so that

$$Pr(J_i = f \mid X_i = (x_f, x_r, x_0)) = 0.5.$$

Assumption 4 guarantees that we can do this.

This means that

$$0.5 = \Pr\left(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0)\right).$$

But the fact that $\varepsilon_{hi} - \varepsilon_{fi}$ has median zero implies that

$$g_h(x_h, x_0) = g_f(x_f, x_0).$$

Since

g_f is identified, clearly g_h is identified from this expression.

Step 4: Identification of G

To identify the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$ note that from the data one can observe

$$\begin{aligned} & \Pr(J_i = f, \log(Y_{fi}) < s \mid X_i = x) \\ = \; & \Pr(g_h(x_h, x_0) + \varepsilon_{hi} \leq g_h(x_h, x_0) + \varepsilon_{hi}, g_f(x_f, x_0) + \varepsilon_{fi} \leq s) \\ = \; & \Pr(\varepsilon_{hi} - \varepsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0), \varepsilon_{fi} \leq s - g_f(x_f, x_0)) \end{aligned}$$

which is the cumulative distribution function of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$ evaluated at the point $(g_f(x_f, x_0) - g_r(x_r, x_0), s - g_f(x_f, x_0))$

Thus we know the joint distribution of $(\varepsilon_{hi} - \varepsilon_{fi}, \varepsilon_{fi})$

from this we can get the joint distribution of $(\varepsilon_{fi}, \varepsilon_{hi})$.