Lecture 3: Estimation and Structural Identification

Quantitative Economics

Seyed Ali Madanizadeh and Hosein Joshaghani

Sharif University of Technology

March 2022

Estimation

Estimation

- So how do we estimate the model and do policy analysis? There are really 3 different approaches:
    1. Estimate full structural model (and thus data generating process) and simulate policy effect
    2. Estimate reduced form of data generating process and simulate policy effect
    3. Try to estimate policy directly without estimating full DGP
- By far the most common is the first so we will focus on that.

Estimation

There are a few examples of the third:

- Heckman and Vytlacil in a series of papers show how to use local instrumentalist variables to estimate policy relevant treatment effects.
- Sufficient statistics can be used to identify some policy effects. Raj Chetty has many such papers.
- Taber and Ichimura (2002)

Estimation Methods

There are really two basic ways of estimating the data generation process:

1. **Maximum (Simulated) Likelihood**
2. Simulation Methods
   - **Simulated Method of Moments (SMM)**
   - **Indirect Inference**

Maximum Likelihood: Review

- So far, we have used ML for estimating Logit and Probit models.
- Recall that, for some random variable Y_i let $f(Y, \theta)$ be the density of Y if it is generated by a model with parameter θ
- The likelihood function just writes the function the other way: '

$$\mathcal{L}(\theta|Y) = f(Y, \theta)$$

Let θ_0 represent the true parameter

Maximum Likelihood: Review

- The key result is this:

$$
\begin{aligned}
E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right] &= \int \frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)} f(Y;\theta_0)dY \\
&= \int \frac{f(Y,\theta)}{f(Y,\theta_0)} f(Y;\theta_0)dY \\
&= \int f(Y,\theta)dY \\
&= 1
\end{aligned}
$$

because $f(Y,\theta)$ is a density.

Maximum Likelihood: Review

- The key result is this:

$$
\begin{aligned}
E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right] &= \int \frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)} f(Y;\theta_0)dY \\
&= \int \frac{f(Y,\theta)}{f(Y,\theta_0)} f(Y;\theta_0)dY \\
&= \int f(Y,\theta)dY \\
&= 1
\end{aligned}
$$

because $f(Y,\theta)$ is a density.

Maximum Likelihood: Review

- The key result is this:

$$
\begin{aligned}
E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right] &= \int \frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)} f(Y;\theta_0)dY \\
&= \int \frac{f(Y,\theta)}{f(Y,\theta_0)} f(Y;\theta_0)dY \\
&= \int f(Y,\theta)dY \\
&= 1
\end{aligned}
$$

because $f(Y,\theta)$ is a density.

Maximum Likelihood: Review

- The key result is this:

$$
\begin{aligned}
E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right] =& \int \frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)} f(Y;\theta_0)dY \\
=& \int \frac{f(Y,\theta)}{f(Y,\theta_0)} f(Y;\theta_0)dY \\
=& \int f(Y,\theta)dY \\
=& 1
\end{aligned}
$$

because $f(Y,\theta)$ is a density.

Maximum Likelihood: Review

- We use Jensen's inequality which implies that for any random variable X_i, the fact that log is concave implies that:

$$E(\log(X_i)) \leq \log(E(X_i))$$

- We apply this with

$$X_i = \frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}$$

Maximum Likelihood: Review

- We use Jensen's inequality which implies that for any random variable X_i, the fact that log is concave implies that:

$$E(\log(X_i)) \leq \log(E(X_i))$$

- We apply this with

$$X_i = \frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}$$

Maximum Likelihood: Review

- Therefore

$$E\left[\log\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right]\right] \leq \log\left[E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right]\right]$$

$$E[log(\mathcal{L}(\theta|Y))] - E[\log(\mathcal{L}(\theta_0|Y))] \leq \log(1)$$

- in other words:

$$E[log(\mathcal{L}(\theta|Y))] < E[\log(\mathcal{L}(\theta_0|Y))]$$

- Thus we know that the true value of θ maximizes expected value of the log likelihood function, $E[\log(\mathcal{L}(\theta|Y))]$

Maximum Likelihood: Review

- Therefore

$$E\left[\log\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right]\right] \leq \log\left[E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right]\right]$$

$$E[log(\mathcal{L}(\theta|Y))] - E[\log(\mathcal{L}(\theta_0|Y))] \leq \log(1)$$

- in other words:

$$E[log(\mathcal{L}(\theta|Y))] < E[\log(\mathcal{L}(\theta_0|Y))]$$

- Thus we know that the true value of θ maximizes expected value of the log likelihood function, $E[\log(\mathcal{L}(\theta|Y))]$

Maximum Likelihood: Review

- Therefore

$$E\left[\log\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right]\right] \leq \log\left[E\left[\frac{\mathcal{L}(\theta|Y)}{\mathcal{L}(\theta_0|Y)}\right]\right]$$

$$E[log(\mathcal{L}(\theta|Y))] - E[\log(\mathcal{L}(\theta_0|Y))] \leq \log(1)$$

- in other words:

$$E[log(\mathcal{L}(\theta|Y))] < E[\log(\mathcal{L}(\theta_0|Y))]$$

- Thus we know that the true value of θ maximizes expected value of the log likelihood function, $E[\log(\mathcal{L}(\theta|Y))]$

Maximum Likelihood: Review

- Maximum likelihood estimator is just the sample analogue of this
- Choose $\hat{\theta}$ as the argument that maximizes

$$\frac{1}{N}\sum_{i=1}^{N}\log(\mathcal{L}(\theta|Y_i))$$

- The most important result for MLE is that it is efficient.
- In particular no alternative estimator can have a lower asymptotic variance.
- Therefore, ideally we like to estimate parameters with MLE. What makes it hard and sometimes infeasible?

Maximum Likelihood: Review

- Maximum likelihood estimator is just the sample analogue of this
- Choose $\hat{\theta}$ as the argument that maximizes

$$\frac{1}{N}\sum_{i=1}^{N}\log(\mathcal{L}(\theta|Y_i))$$

- The most important result for MLE is that it is efficient.
- In particular no alternative estimator can have a lower asymptotic variance.
- Therefore, ideally we like to estimate parameters with MLE. What makes it hard and sometimes infeasible?

Likelihood for the Roy Model

- Let's assume that

$$\begin{bmatrix} \epsilon_{fi} \\ \epsilon_{hi} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{ff} & \sigma_{fh} \\ \sigma_{fh} & \sigma_{hh} \end{bmatrix}\right)$$

- For a fisherman we observe whether you fish, Y_i:

$$\epsilon_{hi} - \epsilon_{fi} < g_f(x_f, x_0) - g_h(x_h, x_0)$$

- and their wage W_i

$$W_i = g_f(X_{fi}, X_{0i}) + \epsilon_{fi}$$

- so we know that $\epsilon_{hi} < W_i - g_h(X_{hi}, X_{0i})$
- Then, the likelihood for each observation is:

$$\int_{-\infty}^{W_i - g_h(X_{hi}, X_{0i})} \phi\left(W_i - g_f(X_{hi}, X_{0i}); \Sigma\right) d\epsilon_{hi}$$

Likelihood for the Roy Model

- Let's assume that

$$\begin{bmatrix}\epsilon_{fi}\\ \epsilon_{hi}\end{bmatrix} \sim N\left(0, \begin{bmatrix}\sigma_{ff} & \sigma_{fh}\\ \sigma_{fh} & \sigma_{hh}\end{bmatrix}\right)$$

- For a fisherman we observe whether you fish, Y_i:

$$\epsilon_{hi} - \epsilon_{fi} < g_f(x_f, x_0) - g_h(x_h, x_0)$$

- and their wage W_i

$$W_i = g_f(X_{fi}, X_{0i}) + \epsilon_{fi}$$

- so we know that $\epsilon_{hi} < W_i - g_h(X_{hi}, X_{0i})$
- Then, the likelihood for each observation is:

$$\int_{-\infty}^{W_i - g_h(X_{hi}, X_{0i})} \phi\big(W_i - g_f(X_{hi}, X_{0i}); \Sigma\big) d\epsilon_{hi}$$

Likelihood for the Roy Model

- Let's assume that

$$\begin{bmatrix} \epsilon_{fi} \\ \epsilon_{hi} \end{bmatrix} \sim N\left(0, \begin{bmatrix} \sigma_{ff} & \sigma_{fh} \\ \sigma_{fh} & \sigma_{hh} \end{bmatrix}\right)$$

- For a fisherman we observe whether you fish, Y_i:

$$\epsilon_{hi} - \epsilon_{fi} < g_f(x_f, x_0) - g_h(x_h, x_0)$$

- and their wage W_i

$$W_i = g_f(X_{fi}, X_{0i}) + \epsilon_{fi}$$

- so we know that $\epsilon_{hi} < W_i - g_h(X_{hi}, X_{0i})$
- Then, the likelihood for each observation is:

$$\int_{-\infty}^{W_i - g_h(X_{hi}, X_{0i})} \phi\big(W_i - g_f(X_{hi}, X_{0i}); \Sigma\big) d\epsilon_{hi}$$

Likelihood for the Roy Model

- We get an analogous expression for hunters
- then the log Likelihood is

$$\frac{1}{N}\sum_{i=1}^{N}\Bigg[F_i \log\left(\int_{-\infty}^{W_i - g_h(X_{hi},X_{0i})} \phi\big(W_i - g_f(X_{fi}, X_{0i}); \Sigma\big)\, d\epsilon_{hi}\right)$$
$$+ (1 - F_i)\log\left(\int_{-\infty}^{W_i - g_f(X_{fi},X_{0i})} \phi\big(W_i - g_h(X_{fi}, X_{0i}); \Sigma\big)\, d\epsilon_{fi}\right)\Bigg]$$

Generalized Method of Moments

- Another way to estimate such a model is by GMM, simulated method of moments, or indirect inference.
- I am not sure these terms mean the same thing to everyone, so I will say what I mean by them but recognize it might mean different things to different people.

True Data Generating Process

- Lets continue to assume that the econometrician observes (Y_i, X_i) which are i.i.d. and both X_i and Y_i are potentially large dimensional.
- We defined the data generating process in the following general way

$$
\begin{aligned}
X_i \sim & H(X_i) \\
u_i \sim & F(u_i, \theta) \\
Y_i = & y_0(X_i, u_i, \theta)
\end{aligned}
$$

GMM

- The standard GMM model would come up with a set of moments

$$m(X_i, Y_i, \theta)$$

- for which

$$E[m(X_i, Y_i, \theta_0)] = 0$$

- the sample analogue comes from recognizing that

$$\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0) \approx 0$$

GMM

- The standard GMM model would come up with a set of moments

$$m(X_i, Y_i, \theta)$$

- for which

$$E[m(X_i, Y_i, \theta_0)] = 0$$

- the sample analogue comes from recognizing that

$$\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0) \approx 0$$

GMM: Minimization Problem

- But more generally we are overidentified so we choose $\hat{\theta}$ to minimize

$$\left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0)\right]' W' \left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0)\right]$$

- where W can be any arbitrary positive definite matrix. In fact any such matrix will produce a consistent and asymptotically normal GMM estimator, the only difference will be in the asymptotic variance of that estimator.
- But, It can be shown that taking $W = \Omega^{-1}$ will result in the most efficient estimator in the class of all asymptotically normal estimators, where

$$\Omega = E[m(X_i, Y_i, \theta_0)' m(X_i, Y_i, \theta_0)]$$

GMM: Minimization Problem

- But more generally we are overidentified so we choose $\hat{\theta}$ to minimize

$$\left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0)\right]' W' \left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0)\right]$$

- where W can be any arbitrary positive definite matrix. In fact any such matrix will produce a consistent and asymptotically normal GMM estimator, the only difference will be in the asymptotic variance of that estimator.
- But, It can be shown that taking $W = \Omega^{-1}$ will result in the most efficient estimator in the class of all asymptotically normal estimators, where

$$\Omega = E[m(X_i, Y_i, \theta_0)' m(X_i, Y_i, \theta_0)]$$

GMM: Minimization Problem

- But more generally we are overidentified so we choose $\hat{\theta}$ to minimize

$$\left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0)\right]' W' \left[\frac{1}{N}\sum_{i=1}^{N} m(X_i, Y_i, \theta_0)\right]$$

- where W can be any arbitrary positive definite matrix. In fact any such matrix will produce a consistent and asymptotically normal GMM estimator, the only difference will be in the asymptotic variance of that estimator.
- But, It can be shown that taking $W = \Omega^{-1}$ will result in the most efficient estimator in the class of all asymptotically normal estimators, where

$$\Omega = E[m(X_i, Y_i, \theta_0)' m(X_i, Y_i, \theta_0)]$$

Relationship between GMM and MLE

- Actually in one way you can think of MLE as a special case of GMM
- We showed above that

$$\theta_0 = \arg\max_{\theta} \left[E(\log(\mathcal{L}(\theta | Y_i))) \right]$$

- but as long as everything is well behaved this means that

$$E\left[\frac{\partial \mathcal{L}(\theta | Y_i)) \log}{\partial \theta} \right] = 0$$

- We can use this as a moment condition
- The one very important caveat is that this is only true if the log likelihood function is strictly concave (recall problem set 4)
- Otherwise there might be multiple solutions to the first order conditions, but only one actual maximum to the likelihood function

Relationship between GMM and MLE

- Actually in one way you can think of MLE as a special case of GMM
- We showed above that

$$\theta_0 = \arg\max_{\theta} \left[E(\log(\mathcal{L}(\theta | Y_i))) \right]$$

- but as long as everything is well behaved this means that

$$E\left[\frac{\partial \mathcal{L}(\theta | Y_i)) \log}{\partial \theta} \right] = 0$$

- We can use this as a moment condition
- The one very important caveat is that this is only true if the log likelihood function is strictly concave (recall problem set 4)
- Otherwise there might be multiple solutions to the first order conditions, but only one actual maximum to the likelihood function

Relationship between GMM and MLE

- Actually in one way you can think of MLE as a special case of GMM
- We showed above that

$$\theta_0 = \arg\max_{\theta} \left[E(\log(\mathcal{L}(\theta|Y_i))) \right]$$

- but as long as everything is well behaved this means that

$$E\left[\frac{\partial \mathcal{L}(\theta|Y_i)) \log}{\partial \theta} \right] = 0$$

- We can use this as a moment condition
- The one very important caveat is that this is only true if the log likelihood function is strictly concave (recall problem set 4)
- Otherwise there might be multiple solutions to the first order conditions, but only one actual maximum to the likelihood function

Simulated Method of Moments

- The classic reference is "A Method of Simulated Moments of Estimation of Discrete Response Models Without Numerical Integration" McFadden, EMA, 1989
- However, we will present it in a different way
- Take any function of the data that you like say $g(Y_i, X_i)$ (where the dimension of g is often large)

GMM $\to$ SMM

- Then notice that since y_0 and $F(u, \theta_0)$ represent the data generating process, then

$$E[g(Y_i, X_i)] = \int\int g(y_0(X, u; \theta_0), X_i) dF(u, \theta_0) dH(X)$$

- So this means that we can do GMM with

$$m(X_i, Y_i, \theta) = E[g(Y_i, X_i)] - \int\int g(y_0(X, u; \theta), X_i) dF(u, \theta) dH(X)$$

- So what? Is this a progress?

GMM $\to$ SMM

- Then notice that since y_0 and $F(u, \theta_0)$ represent the data generating process, then

$$E[g(Y_i, X_i)] = \int \int g(y_0(X, u; \theta_0), X_i) dF(u, \theta_0) dH(X)$$

- So this means that we can do GMM with

$$m(X_i, Y_i, \theta) = E[g(Y_i, X_i)] - \int \int g(y_0(X, u; \theta), X_i) dF(u, \theta) dH(X)$$

- So what? Is this a progress?

Cool SMM

- Here is where things get pretty cool
- for each θ, what you need to compute (simulate) is reduced to

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) - \frac{1}{R}\sum_{r=1}^{R} g\left(x_r, y_0(x_r, u_r, \theta)\right)$$

- Why? what is so cool about this?
- The nice thing about this is that we didn't need R to be large for every N, we only needed R to be large for the one integral.
- For MLE we had to approximate the integral well for every single observation
- 1 Million observations and 1 day each simulation $\rightarrow$ 27.4 centuries (1 Hour $\rightarrow$ 114 years)!
- Notice that at the true value the estimator is approximately

$$E[g(Y_i, X_i)] - \int\int g(y_0(X, u; \theta), X_i) dF(u, \theta) dH(X) = 0$$

Cool SMM

- Here is where things get pretty cool
- for each θ, what you need to compute (simulate) is reduced to

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) - \frac{1}{R}\sum_{r=1}^{R} g\left(x_r, y_0(x_r, u_r, \theta)\right)$$

- Why? what is so cool about this?
- The nice thing about this is that we didn't need R to be large for every N, we only needed R to be large for the one integral.
- For MLE we had to approximate the integral well for every single observation
- 1 Million observations and 1 day each simulation $\rightarrow$ 27.4 centuries (1 Hour $\rightarrow$ 114 years)!
- Notice that at the true value the estimator is approximately

$$E[g(Y_i, X_i)] - \int\int g(y_0(X, u; \theta), X_i) dF(u, \theta) dH(X) = 0$$

Cool SMM

- Here is where things get pretty cool
- for each θ, what you need to compute (simulate) is reduced to

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) - \frac{1}{R}\sum_{r=1}^{R} g\left(x_r, y_0(x_r, u_r, \theta)\right)$$

- Why? what is so cool about this?
- The nice thing about this is that we didn't need R to be large for every N, we only needed R to be large for the one integral.
- For MLE we had to approximate the integral well for every single observation
- 1 Million observations and 1 day each simulation $\rightarrow$ 27.4 centuries (1 Hour $\rightarrow$ 114 years)!
- Notice that at the true value the estimator is approximately

$$E[g(Y_i, X_i)] - \int\int g(y_0(X, u; \theta), X_i) dF(u, \theta) dH(X) = 0$$

Cool SMM

- Here is where things get pretty cool
- for each θ, what you need to compute (simulate) is reduced to

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) - \frac{1}{R}\sum_{r=1}^{R} g\Big(x_r, y_0(x_r, u_r, \theta)\Big)$$

- Why? what is so cool about this?
- The nice thing about this is that we didn't need R to be large for every N, we only needed R to be large for the one integral.
- For MLE we had to approximate the integral well for every single observation
- 1 Million observations and 1 day each simulation $\rightarrow$ 27.4 centuries (1 Hour $\rightarrow$ 114 years)!
- Notice that at the true value the estimator is approximately

$$E[g(Y_i, X_i)] - \int\int g(y_0(X, u; \theta), X_i) dF(u, \theta) dH(X) = 0$$

Indirect Inference

- The classic reference here is "Indirect Inference" Gourieroux, Monrort, and Renault, Journal of Applied Econometrics, (1993)
- Again we will think about this in a different way then them
- Think about the intuition for the SMM estimator

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) \approx \frac{1}{R}\sum_{r=1}^{R} g\Big(x_r, y_0(x_r, u_r, \theta_0)\Big)$$

- If I have the right data generating model taking the mean of the simulated data should give me the same answer as taking the mean of the actual data
- Under what conditions this can go wrong?

Indirect Inference

- The classic reference here is "Indirect Inference" Gourieroux, Monrort, and Renault, Journal of Applied Econometrics, (1993)
- Again we will think about this in a different way then them
- Think about the intuition for the SMM estimator

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) \approx \frac{1}{R}\sum_{r=1}^{R} g\Big(x_r, y_0(x_r, u_r, \theta_0)\Big)$$

- If I have the right data generating model taking the mean of the simulated data should give me the same answer as taking the mean of the actual data
- Under what conditions this can go wrong?

Indirect Inference

- The classic reference here is "Indirect Inference" Gourieroux, Monrort, and Renault, Journal of Applied Econometrics, (1993)
- Again we will think about this in a different way then them
- Think about the intuition for the SMM estimator

$$\frac{1}{N}\sum_{i=1}^{N} g(X_i, Y_i) \approx \frac{1}{R}\sum_{r=1}^{R} g\left(x_r, y_0(x_r, u_r, \theta_0)\right)$$

- If I have the right data generating model taking the mean of the simulated data should give me the same answer as taking the mean of the actual data
- Under what conditions this can go wrong?

Indirect Inference

- But we can generalize that idea
- If I have the right data generating model, if I use the true parameter value, the simulated data should look the same as the actual data
- That means whatever the heck I do to the real data, if I do exactly the same thing to the simulated data I should get the same answer

Indirect Inference

- But we can generalize that idea
- If I have the right data generating model, if I use the true parameter value, the simulated data should look the same as the actual data
- That means whatever the heck I do to the real data, if I do exactly the same thing to the simulated data I should get the same answer

Indirect Inference

- Estimate auxiliary parameter $\hat{\beta}$ using some estimation scheme in real data
- for any particular value of θ:
  1. Simulate data using data generation process:

$$y_0(x, u, \theta), H(X), G(u, \theta)$$

  2. Estimate $\hat{B}(\theta)$ using exactly the same estimation scheme on simulated data
- Then choose θ to minimize:

$$\left(\hat{B}(\theta) - \hat{\beta}\right)' \Omega \left(\hat{B}(\theta) - \hat{\beta}\right)$$

- This is consistent because

$$\hat{B}(\theta_0) - \hat{\beta} \xrightarrow{p} 0$$

Indirect Inference

- Estimate auxiliary parameter $\hat{\beta}$ using some estimation scheme in real data
- for any particular value of θ:
  1. Simulate data using data generation process:

$$y_0(x, u, \theta), H(X), G(u, \theta)$$

  2. Estimate $\hat{B}(\theta)$ using exactly the same estimation scheme on simulated data

- Then choose θ to minimize:

$$\left(\hat{B}(\theta) - \hat{\beta}\right)' \Omega \left(\hat{B}(\theta) - \hat{\beta}\right)$$

- This is consistent because

$$\hat{B}(\theta_0) - \hat{\beta} \xrightarrow{p} 0$$

Indirect Inference

- Estimate auxiliary parameter $\hat{\beta}$ using some estimation scheme in real data
- for any particular value of θ:
  1. Simulate data using data generation process:

$$y_0(x,u,\theta), H(X), G(u,\theta)$$

  2. Estimate $\hat{B}(\theta)$ using exactly the same estimation scheme on simulated data
- Then choose θ to minimize:

$$\left(\hat{B}(\theta)-\hat{\beta}\right)'\Omega\left(\hat{B}(\theta)-\hat{\beta}\right)$$

- This is consistent because

$$\hat{B}(\theta_0)-\hat{\beta}\xrightarrow{p}0$$

Indirect Inference

- Estimate auxiliary parameter $\hat{\beta}$ using some estimation scheme in real data
- for any particular value of θ:
  1. Simulate data using data generation process:

$$y_0(x, u, \theta), H(X), G(u, \theta)$$

  2. Estimate $\hat{B}(\theta)$ using exactly the same estimation scheme on simulated data
- Then choose θ to minimize:

$$\left(\hat{B}(\theta) - \hat{\beta}\right)' \Omega \left(\hat{B}(\theta) - \hat{\beta}\right)$$

- This is consistent because

$$\hat{B}(\theta_0) - \hat{\beta} \xrightarrow{p} 0$$

Indirect Inference

- The most important thing: this can be misspecified, it doesn't have to estimate a true causal parameter
- Creates a nice connection with reduced form stuff, we can use 2SLS or Diff in Diff as auxiliary parameters and it is clear where identification comes from
- Can think of the analogue to the forecasting out of the sample
- we use Indirect Inference to extend the convincing identification scheme into a structural framework

Examples of $\hat{\beta}$

- Moments
- Regression models
- Misspecified MLE
- Misspecified GMM
- IV
- Difference in Differences
- Regression Discontinuity
- Even Randomized Control Files

Maximum Likelihood versus Indirect Inference

- MLE is efficient
- Indirect inference you pick auxiliary model
- Which is better is not obvious.
- Picking auxiliary model is somewhat arbitrary, but you can pick what you want the data to fit.
- MLE essentially picks the moments that are most efficient-a statistical criterion

Maximum Likelihood versus Indirect Inference

- Indirect inference is often computationally easier because of the simulation approximation of integrals
- With confidential data, Indirect Inference often is easier because only need to use the actual data to get $\hat{\beta}$
- A drawback of simulation estimators is that they often lead to non-smooth objective functions
- Indirect inference preserves some of the advantages of design-based estimation
- Map from data to parameters is more transparent
- Becomes like the forecasting experiment where we are forecasting out of the range of the data

Maximum Likelihood versus Indirect Inference

- Indirect inference is often computationally easier because of the simulation approximation of integrals
- With confidential data, Indirect Inference often is easier because only need to use the actual data to get $\hat{\beta}$
- A drawback of simulation estimators is that they often lead to non-smooth objective functions
- Indirect inference preserves some of the advantages of design-based estimation
- Map from data to parameters is more transparent
- Becomes like the forecasting experiment where we are forecasting out of the range of the data

Maximum Likelihood versus Indirect Inference

- Indirect inference is often computationally easier because of the simulation approximation of integrals
- With confidential data, Indirect Inference often is easier because only need to use the actual data to get $\hat{\beta}$
- A drawback of simulation estimators is that they often lead to non-smooth objective functions
- Indirect inference preserves some of the advantages of design-based estimation
- Map from data to parameters is more transparent
- Becomes like the forecasting experiment where we are forecasting out of the range of the data

Conclusion

- Structural Estimation is one very powerful method for better understanding the economy and the real world
- For whatever reason, we don't see much research in Iran using this method.
- This course is designed to make the change in empirical research.
- Almost surly you will forget this material, unless you actually use these methods in your actual research!
- don't forget to work on problem set 5 and 6!
- Heads up: In pset 5 you should write a research proposal ...

Conclusion

- Structural Estimation is one very powerful method for better understanding the economy and the real world
- For whatever reason, we don't see much research in Iran using this method.
- This course is designed to make the change in empirical research.
- Almost surly you will forget this material, unless you actually use these methods in your actual research!
- don't forget to work on problem set 5 and 6!
- Heads up: In pset 5 you should write a research proposal ...

Conclusion

- Structural Estimation is one very powerful method for better understanding the economy and the real world
- For whatever reason, we don't see much research in Iran using this method.
- This course is designed to make the change in empirical research.
- Almost surly you will forget this material, unless you actually use these methods in your actual research!
- don't forget to work on problem set 5 and 6!
- Heads up: In pset 5 you should write a research proposal ...

Conclusion

- Structural Estimation is one very powerful method for better understanding the economy and the real world
- For whatever reason, we don't see much research in Iran using this method.
- This course is designed to make the change in empirical research.
- Almost surly you will forget this material, unless you actually use these methods in your actual research!
- don't forget to work on problem set 5 and 6!
- Heads up: In pset 5 you should write a research proposal ...

Cooley and Prescott (1976)

- In many instances economic theory suggests that relationships will change over time. Lucas, for example, has shown that econometric models, as they are now structured, are inappropriate tools for long-term policy evaluation precisely because they assume a stable structure. The structure of an *econometric model* represents the optimal decision rules of economic agents. From dynamic economic theory we know that optimal decision rules vary systematically with changes in the structure of series relevant to the decision makers.
- It follows that changes in policy will systematically alter the structure of the series being forecasted by decision makers, and, therefore, the behavioral relationships as well.

Hansen and Heckman (1996)

- Kydland and Prescott are to be praised for taking the general equilibrium analysis of Shoven and Whalley one step further by using stochastic general equilibrium as a framework for understanding macroeconomics
- While Kydland and Prescott advocate the use of "well-tested theories" in their essay, they never move beyond this slogan, and they do not justify their claim of fulfilling this criterion in their own research. "Well tested" must mean more than "familiar" or "widely accepted" or "agreed on by convention," if it is to mean anything!
- Their suggestion that we "calibrate the model" is similarly vague.
- ... On the other hand, Kydland and Prescott never provide a coherent framework for extracting parameters from microeconomic data.

Hansen and Heckman (1996)

- Kydland and Prescott are to be praised for taking the general equilibrium analysis of Shoven and Whalley one step further by using stochastic general equilibrium as a framework for understanding macroeconomics
- While Kydland and Prescott advocate the use of "well-tested theories" in their essay, they never move beyond this slogan, and they do not justify their claim of fulfilling this criterion in their own research. "Well tested" must mean more than "familiar" or "widely accepted" or "agreed on by convention," if it is to mean anything!
- Their suggestion that we "calibrate the model" is similarly vague.
- ... On the other hand, Kydland and Prescott never provide a coherent framework for extracting parameters from microeconomic data.

Hansen and Heckman (1996)

- Kydland and Prescott are to be praised for taking the general equilibrium analysis of Shoven and Whalley one step further by using stochastic general equilibrium as a framework for understanding macroeconomics
- While Kydland and Prescott advocate the use of "well-tested theories" in their essay, they never move beyond this slogan, and they do not justify their claim of fulfilling this criterion in their own research. "Well tested" must mean more than "familiar" or "widely accepted" or "agreed on by convention," if it is to mean anything!
- Their suggestion that we "calibrate the model" is similarly vague.
- ... On the other hand, Kydland and Prescott never provide a coherent framework for extracting parameters from microeconomic data.

Holland 1986

- The goal of the structural econometrics literature, like the goal of all science, is to understand the causal mechanisms producing effects so that one can use empirical versions of models to forecast the effects of interventions never previously experienced, to calculate a variety of policy counterfactuals and to use theory to guide choices of estimators to interpret evidence and to cumulate evidence across studies.
- These activities require models for understanding "causes of effects" in contrast to the program evaluation literature that focuses only on the "effects of causes"

Holland 1986

- The goal of the structural econometrics literature, like the goal of all science, is to understand the causal mechanisms producing effects so that one can use empirical versions of models to forecast the effects of interventions never previously experienced, to calculate a variety of policy counterfactuals and to use theory to guide choices of estimators to interpret evidence and to cumulate evidence across studies.
- These activities require models for understanding "causes of effects" in contrast to the program evaluation literature that focuses only on the "effects of causes"

Recourses

- This lecture note and slides are mainly based on presentation of Chris Taber at the university of Chicago.
- Micheal Keane also presented practical notes about Structural estimation at the University of Chicago.
- Also lecture notes of Tony Whited from Ross School of Business at the University of Michigan is widely incorporated.
- Lecture notes of Lucian Taylor at Wharton school in UPenn is also used.
- More reading:
  1. French and Taber (2011) Handbook chapter of Labor Economics
  2. Heckman and Honroe (1991) Econometrica
  3. problem set 5 and 6!
  4. If you want to learn more read papers that used structural estimation such as Rust (1987), Lucian Taylor (2010 and 2013) etc.

Recourses

- This lecture note and slides are mainly based on presentation of Chris Taber at the university of Chicago.
- Micheal Keane also presented practical notes about Structural estimation at the University of Chicago.
- Also lecture notes of Tony Whited from Ross School of Business at the University of Michigan is widely incorporated.
- Lecture notes of Lucian Taylor at Wharton school in UPenn is also used.
- More reading:
  1. French and Taber (2011) Handbook chapter of Labor Economics
  2. Heckman and Honroe (1991) Econometrica
  3. problem set 5 and 6!
  4. If you want to learn more read papers that used structural estimation such as Rust (1987), Lucian Taylor (2010 and 2013) etc.

Road map

- Examples
    1. Supply and Demand
    2. The Roy Model
- Structural and Reduced Form Models
- Identification
    1. Identification of Supply and Demand
    2. Identification of The Roy Model
- Estimation
    1. Maximum Simulated Likelihood (MSL)
    2. General Method of Moments (GMM)
    3. Simulated Method of Moments (SMM)
    4. Indirect Inference

Examples

1. Supply and Demand
2. The Roy Model

Example 1: Supply and Demand

Consider the classic simultaneous equations model in a policy regime with no taxes:

- Supply Curve:

$$Q_t = \alpha_s P_t + X_t' \beta_s + Z_{st}' \gamma_s + u_t$$

- Demand Curve:

$$Q_t = \alpha_d P_t + X_t' \beta_d + Z_{dt}' \gamma_d + \nu_t$$

- We can solve for prices and quantities as:

$$P_t = \frac{X_t'(\beta_d - \beta_s) + Z_{dt}' \gamma_d - Z_{st}' \gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d} \tag{1}$$

$$Q_t = \frac{\alpha_s (X_t' \beta_d + Z_{dt}' \gamma_d + \nu_t) - \alpha_d (X_t' \beta_s + Z_{st}' \gamma_s + u_t)}{\alpha_s - \alpha_d} \tag{2}$$

Example 1: Supply and Demand

Consider the classic simultaneous equations model in a policy regime with no taxes:

- Supply Curve:

$$Q_t = \alpha_s P_t + X_t' \beta_s + Z_{st}' \gamma_s + u_t$$

- Demand Curve:

$$Q_t = \alpha_d P_t + X_t' \beta_d + Z_{dt}' \gamma_d + \nu_t$$

- We can solve for prices and quantities as:

$$P_t = \frac{X_t'(\beta_d - \beta_s) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d} \quad (1)$$

$$Q_t = \frac{\alpha_s(X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t) - \alpha_d(X_t'\beta_s + Z_{st}'\gamma_s + u_t)}{\alpha_s - \alpha_d} \quad (2)$$

Example 1: Supply and Demand

Now suppose we want to introduce a tax on this good, imposed on consumers. So now

$$Q_t = \alpha_d(1+\tau)P_t + X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t$$

Then the new equilibrium is:

$$P_t = \frac{X_t'(\beta_d - \beta_s) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d(1+\tau)}$$

$$Q_t = \frac{\alpha_s(X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t) - \alpha_d(1+\tau)(X_t'\beta_s + Z_{st}'\gamma_s + u_t)}{\alpha_s - \alpha_d(1+\tau)}$$

Assumption: taking the model seriously $\rightarrow$ all of the parameters are policy invariant

Example 1: Supply and Demand

Now suppose we want to introduce a tax on this good, imposed on consumers. So now

$$Q_t = \alpha_d(1+\tau)P_t + X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t$$

Then the new equilibrium is:

$$P_t = \frac{X_t'(\beta_d - \beta_s) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d(1+\tau)}$$
$$Q_t = \frac{\alpha_s(X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t) - \alpha_d(1+\tau)(X_t'\beta_s + Z_{st}'\gamma_s + u_t)}{\alpha_s - \alpha_d(1+\tau)}$$

Assumption: taking the model seriously $\rightarrow$ all of the parameters are policy invariant

Example 2: The Roy Model

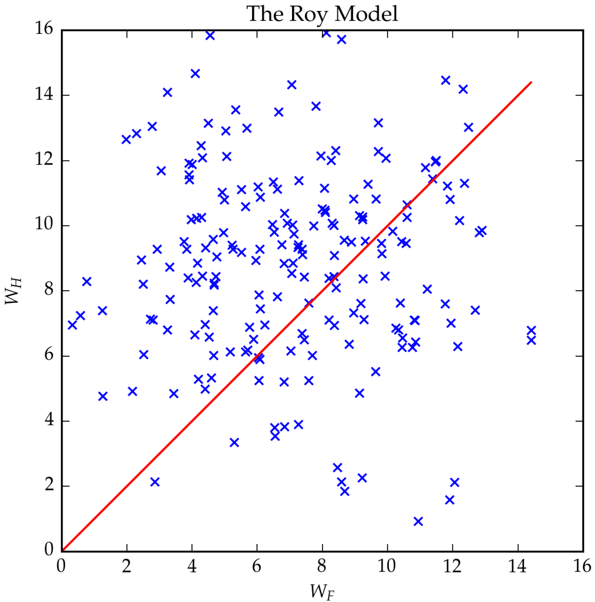
- Why the Roy Model? Simple intuition, hard identification, lots of applications
- Labor Market is a village
- There are two occupations
  1. Hunter
  2. Fisherman
- Fish and Rabbits are completely homogeneous
- No uncertainty in number you catch

Example 2: The Roy Model

- Wages are:

$$W_f = \pi_F F$$
$$W_H = \pi_H R$$

- Notice that we have imposed no structure on F and R yet.
- We can't say much without imposing the structure.


The Roy Model
W_H
W_F
0
2
4
6
8
10
12
14
16

Example 2: The Roy Model

- Once we know the model we could think of several different policies
- One is suppose we impose a minimum wage $\bar{w}$ in the fishing sector but not in the hunting sector?
- What will this due to earnings inequality?
- Anyone who with $W_F < \bar{w}$ will no longer be employed in the fishing sector and must now hunt where they earn lower wages.
- Inequality will like rise and we can determine the magnitude from the model

Example 2: The Roy Model

- Once we know the model we could think of several different policies
- One is suppose we impose a minimum wage $\bar{w}$ in the fishing sector but not in the hunting sector?
- What will this due to earnings inequality?
- Anyone who with $W_F < \bar{w}$ will no longer be employed in the fishing sector and must now hunt where they earn lower wages.
- Inequality will like rise and we can determine the magnitude from the model

Other Examples of the Roy Model

- Effects of Affordable Care Act on labor market outcomes (Aizawa and Fang, 2015)
- Tuition Subsidies on Health (Heckman, Humphries, and Veramundi, 2015)
- Effects of extending lenth of payment for college loan programs on college enrollment (Li, 2015)
- Peer effects of school vouchers on public school students (Altonji, Huang, and Taber, 2015)
- Tax credits versus income support (Blundell, Costa Dias, Meghir, and Shaw, 2015)
- Effects of border tightening on the U.S. government budget constraints (Nakajima, 2015)
- Welfare effects of alternative designs of school choice programs (Calsamiglia, Fu, and Guell, 2014)

Structural and Reduced Form Models

What Does Structural Mean?

- It makes no sense to say "structural model."
- All economic models are "structural."
- Usually when people say "structural model," they really mean "dynamic model."
- It makes a lot of sense to talk about "structural" versus "reduced-form" estimation.

What Does Structural Mean?

No obvious answer, it means different things to different people!
at least 4 definitions.

1. Historical: The structural parameters in a simultaneous equations model
2. Estimation of preference and technology parameters in a maximizing model (perhaps combined with some specification of markets)
3. Parameters are policy invariant
4. Structural as opposed to Statistical

What Does Structural Mean?

No obvious answer, it means different things to different people!
at least 4 definitions.

1. Historical: The structural parameters in a simultaneous equations model
2. Estimation of preference and technology parameters in a maximizing model (perhaps combined with some specification of markets)
3. Parameters are policy invariant
4. Structural as opposed to Statistical

What Does Structural Mean?

No obvious answer, it means different things to different people!
at least 4 definitions.

1. Historical: The structural parameters in a simultaneous equations model
2. Estimation of preference and technology parameters in a maximizing model (perhaps combined with some specification of markets)
3. Parameters are policy invariant
4. Structural as opposed to Statistical

What Does Structural Mean?

No obvious answer, it means different things to different people!
at least 4 definitions.

1. Historical: The structural parameters in a simultaneous equations model
2. Estimation of preference and technology parameters in a maximizing model (perhaps combined with some specification of markets)
3. Parameters are policy invariant
4. Structural as opposed to Statistical

What Does Structural Mean?

No obvious answer, it means different things to different people!
at least 4 definitions.

1. Historical: The structural parameters in a simultaneous equations model
2. Estimation of preference and technology parameters in a maximizing model (perhaps combined with some specification of markets)
3. Parameters are policy invariant
4. Structural as opposed to Statistical

What Does Structural Mean?

- Structural as opposed to Statistical : A statistical model describes the relation between two or more random variables:

$$y = X\beta + \epsilon$$

- An economic model starts with assumptions about:
  - agents' preferences,
  - constraints,
  - firms' production functions,
  - some notion of equilibrium, etc.
- Then it makes predictions about the relation between observable, often endogenous variables.

What Does Structural Mean?

- Structural as opposed to Statistical : A statistical model describes the relation between two or more random variables:

$$y = X\beta + \epsilon$$

- An economic model starts with assumptions about:
  - agents' preferences,
  - constraints,
  - firms' production functions,
  - some notion of equilibrium, etc.
- Then it makes predictions about the relation between observable, often endogenous variables.

What Does Reduced-Form Mean?

Now for many people it essentially means anything that is not structural

- Preferred definition: reduced form parameters are a known function of underlying structural parameters.
- fits classic Simultaneous Equation definition
- might not be invertible (say without an instrument)
- for something to be reduced form according to this definition you need to write down a structural model
- this actually has content-you can sometimes use reduced form models to simulate a policy that has never been implemented (as often reduced form parameters are structural in the sense that they are policy invariant)

What Does Reduced-Form Mean?

Now for many people it essentially means anything that is not structural

- Preferred definition: reduced form parameters are a known function of underlying structural parameters.
- fits classic Simultaneous Equation definition
- might not be invertible (say without an instrument)
- for something to be reduced form according to this definition you need to write down a structural model
- this actually has content-you can sometimes use reduced form models to simulate a policy that has never been implemented (as often reduced form parameters are structural in the sense that they are policy invariant)

what does reduced form mean?

- For many other researchers reduced form means:
  - What is the (causal) effect of X on Y?
- While structural means:
  - Why does X affect Y?
  - What are the magnitudes of the parameters?
  - How well does theory line up with the data?
  - How would the world look if one of the parameters (counterfactually) changed?
  - What would happen if you (counterfactually) shocked the system?

"Structural" versus "Design-Based"

- the fact that there are advantages and disadvantages makes them complements rather than substitutes
- Avoid structural estimation if there is designed based solution for your question
- In practice, good research uses both modeling
- There are very very few (if any?) structural paper about Iran being explicit about identification and estimation!

"Structural" versus "Design-Based"

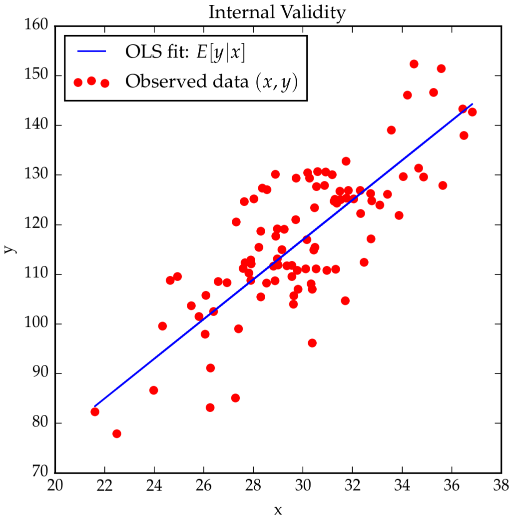
- the fact that there are advantages and disadvantages makes them complements rather than substitutes
- Avoid structural estimation if there is designed based solution for your question
- In practice, good research uses both modeling
- There are very very few (if any?) structural paper about Iran being explicit about identification and estimation!

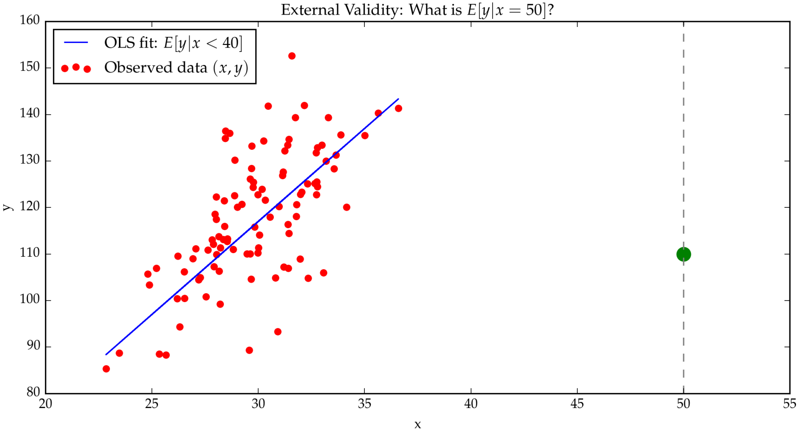
"Structural" versus "Design-Based"

- the fact that there are advantages and disadvantages makes them complements rather than substitutes
- Avoid structural estimation if there is designed based solution for your question
- In practice, good research uses both modeling
- There are very very few (if any?) structural paper about Iran being explicit about identification and estimation!

"Structural" versus "Design-Based"

Structural	Design-Based
Better on External Validity	Better on Internal Validity
Map from parameters to implications clearer	Map from data to parameters more transparent
Formalizes conditions for external validity	Requires fewer assumptions
Forces one to think about where data comes from	Might come from somewhere else
Easier to interpret what parameters mean	Estimates more credible


Internal Validity
OLS fit: $E[y|x]$
Observed data (x, y)
160
150
140
130
120
110
100
90
80
70
y
20
22
24
26
28
30
32
34
36
38
x


External Validity: What is $E[y|x = 50]$?
OLS fit: $E[y|x < 40]$
Observed data (x, y)
y
x
160
150
140
130
120
110
100
90
80
20
25
30
35
40
45
50
55

"Calibration" versus "Structural Estimation"

- Calibration:
  - Take many parameter values from other papers
  - Usually have more parameters than moments $\rightarrow$ model isn't identified $\rightarrow$ can't put standard errors on parameters
  - Mainly a theoretical exercise
  - Never, ever, say that in front of James Heckman!
- Structural estimation:
  - Infer parameter values from the data
  - Get standard errors for parameters
  - An empirical exercise

Chris Taber: To me, calibration is structural estimation without identification and standard error!

"Calibration" versus "Structural Estimation"

- Calibration:
  - Take many parameter values from other papers
  - Usually have more parameters than moments $\rightarrow$ model isn't identified $\rightarrow$ can't put standard errors on parameters
  - Mainly a theoretical exercise
  - Never, ever, say that in front of James Heckman!
- Structural estimation:
  - Infer parameter values from the data
  - Get standard errors for parameters
  - An empirical exercise

Chris Taber: To me, calibration is structural estimation without identification and standard error!

"Calibration" versus "Structural Estimation"

- Calibration:
  - Take many parameter values from other papers
  - Usually have more parameters than moments $\rightarrow$ model isn't identified $\rightarrow$ can't put standard errors on parameters
  - Mainly a theoretical exercise
  - Never, ever, say that in front of James Heckman!
- Structural estimation:
  - Infer parameter values from the data
  - Get standard errors for parameters
  - An empirical exercise

Chris Taber: To me, calibration is structural estimation without identification and standard error!

Hansen and Heckman (1996)

- What credibility should we attach to numbers produced from their [Kydland and Prescott] "computational experiments," and why should we use their "calibrated models" as a basis for serious quantitative policy evaluation? The essay by Kydland and Prescott begs these fundamental questions.
- The deliberately limited use of available information in such computational experiments runs the danger of making many economic models with very different welfare implications compatible with the evidence.
- We suggest that Kydland and Prescott's account of the availability and value of micro estimates for macro models is dramatically overstated. There is no filing cabinet full of robust micro estimates ready to use in calibrating dynamic stochastic general equilibrium models.

Hansen and Heckman (1996)

- What credibility should we attach to numbers produced from their [Kydland and Prescott] "computational experiments," and why should we use their "calibrated models" as a basis for serious quantitative policy evaluation? The essay by Kydland and Prescott begs these fundamental questions.
- The deliberately limited use of available information in such computational experiments runs the danger of making many economic models with very different welfare implications compatible with the evidence.
- We suggest that Kydland and Prescott's account of the availability and value of micro estimates for macro models is dramatically overstated. There is no filing cabinet full of robust micro estimates ready to use in calibrating dynamic stochastic general equilibrium models.

Hansen and Heckman (1996)

- What credibility should we attach to numbers produced from their [Kydland and Prescott] "computational experiments," and why should we use their "calibrated models" as a basis for serious quantitative policy evaluation? The essay by Kydland and Prescott begs these fundamental questions.
- The deliberately limited use of available information in such computational experiments runs the danger of making many economic models with very different welfare implications compatible with the evidence.
- We suggest that Kydland and Prescott's account of the availability and value of micro estimates for macro models is dramatically overstated. There is no filing cabinet full of robust micro estimates ready to use in calibrating dynamic stochastic general equilibrium models.

Hansen and Heckman (1996)

Calibration versus Estimation

- A novel feature of the real business cycle research program is its endorsement of "calibration" as an alternative to "estimation."
- However, the distinction drawn between calibrating and estimating the parameters of a model is artificial at best.
- Moreover, the justification for what is called "calibration" is vague and confusing.
- Since the Kydland-Prescott essay is vague about the operating principles of calibration, we turn elsewhere for specificity. For instance, in a recent description of the use of numerical models in the earth sciences, Oreskes, Shrader-Frechette and Belitz (1994, pp. 642, 643) describe calibration as follows:

Hansen and Heckman (1996)

Calibration versus Estimation

- A novel feature of the real business cycle research program is its endorsement of "calibration" as an alternative to "estimation."
- However, the distinction drawn between calibrating and estimating the parameters of a model is artificial at best.
- Moreover, the justification for what is called "calibration" is vague and confusing.
- Since the Kydland-Prescott essay is vague about the operating principles of calibration, we turn elsewhere for specificity. For instance, in a recent description of the use of numerical models in the earth sciences, Oreskes, Shrader-Frechette and Belitz (1994, pp. 642, 643) describe calibration as follows:

Hansen and Heckman (1996)

Calibration versus Estimation

- *In earth sciences, the modeler is commonly faced with the inverse problem: The distribution of the dependent variable (for example, the hydraulic head) is the most well known aspect of the system; the distribution of the independent variable is the least well known. The process of tuning the model: that is, the manipulation of the independent variables to obtain a match between the observed and simulated distribution or distributions of a dependent variable or variables, is known as calibration.*
- *Some hydrologists have suggested a two-step calibration scheme in which the available dependent data set is divided into two parts. In the first step, the independent parameters of the model are adjusted to reproduce the first part of the data. Then in the second step the model is run and the results are compared with the second part of the data. In this scheme, the first step is labeled "calibration" and the second step is labeled "verification."*
- Econometricians refer to the first stage as *estimation* and the second stage as *testing*.

Hansen and Heckman (1996)

Calibration versus Estimation

- *In earth sciences, the modeler is commonly faced with the inverse problem: The distribution of the dependent variable (for example, the hydraulic head) is the most well known aspect of the system; the distribution of the independent variable is the least well known. The process of tuning the model: that is, the manipulation of the independent variables to obtain a match between the observed and simulated distribution or distributions of a dependent variable or variables, is known as calibration.*
- *Some hydrologists have suggested a two-step calibration scheme in which the available dependent data set is divided into two parts. In the first step, the independent parameters of the model are adjusted to reproduce the first part of the data. Then in the second step the model is run and the results are compared with the second part of the data. In this scheme, the first step is labeled "calibration" and the second step is labeled "verification."*
- Econometricians refer to the first stage as *estimation* and the second stage as *testing*.

Hansen and Heckman (1996)

Calibration versus Estimation

- In their proposed paradigm for empirical research, correlations are to be saved and used to test models, but are not to be used as a source of information about parameter values.
- It has become commonplace in the real business cycle research program to match the steady-state implications of models to time series averages.
- To an outsider, this looks remarkably like a way of doing estimation without accounting for sampling error in the sample means.

Hansen and Heckman (1996): External Validity

- It can be very misleading to plug microeconometric parameter estimates into a macroeconomic model when the economic environments for the two models are fundamentally different.
- In fact, many of the micro studies that the "calibrators" draw upon do not estimate the parameters required by the models being simulated.

Hansen and Heckman (1996): External Validity

- It can be very misleading to plug microeconometric parameter estimates into a macroeconomic model when the economic environments for the two models are fundamentally different.
- In fact, many of the micro studies that the "calibrators" draw upon do not estimate the parameters required by the models being simulated.

"Calibration" versus "Structural Estimation"

Both "Calibration" and "Structural Estimation":

- Can assess how well model fits the data, however, no statistical tests with calibration
- Can use model to ask counterfactual questions:
  - What would happen if we shocked this variable?
  - How would world look like if we changed this parameter's value?

(Possible) Steps for Writing Structural Paper

We expect you to write a paper for this course which hopefully will lead to your dissertation. Here are the steps:

1. Identify the policy question to be answered
2. Write down a model that can simulate policy
3. Understanding how the model works
4. Think about identification/data (with the goal being the policy counterfactual)
5. Estimate the model
6. Test for External Validity
7. Simulate the policy counterfactual

Other reasons to write structural models

- Further evaluation of an established policy: we might want to know welfare effect
- Basic Research: we want to understand the world better
  - Use data to help understand model
  - Use model to help understand data (use structural model as a lens)

Identification

Identification: Data Generating Process

- Define the data generating process in the following general way

$$
\begin{aligned}
X_i \sim & H(X_i) \\
u_i \sim & F(u_i, \theta) \\
Y_i = & y_0(X_i, u_i, \theta)
\end{aligned}
$$

- The observed data is (Y_i, X_i) with u_i unobserved
- We know the model up to parameter θ
- To think of this as non-parametric we can think of θ is infinite dimensional

$$\theta = (\theta_1, F(.))$$

- To simulate a policy counterfactual your policy needs to be a known manipulation of this structural model (i.e. $\pi(\theta)$)

Identification: Data Generating Process

- Define the data generating process in the following general way

$$
\begin{aligned}
X_i \sim & H(X_i) \\
u_i \sim & F(u_i, \theta) \\
Y_i = & y_0(X_i, u_i, \theta)
\end{aligned}
$$

- The observed data is (Y_i, X_i) with u_i unobserved
- We know the model up to parameter θ
- To think of this as non-parametric we can think of θ is infinite dimensional

$$\theta = (\theta_1, F(.))$$

- To simulate a policy counterfactual your policy needs to be a known manipulation of this structural model (i.e. $\pi(\theta)$)

Data Generating Process: Supply and Demand

$$
\begin{aligned}
Y_t =& (P_t, Q_t) \\
X_t =& (X_t, Z_{dt}, Z_{st}) \\
u_t =& (u_t, \nu_t) \\
\theta =& (\alpha_d, \alpha_s, \beta_d, \beta_s, \gamma_d, \gamma_d, G(u, \nu))
\end{aligned}
$$

$$
y_0(X_i, u_i; \theta) = \begin{bmatrix} \frac{X_t'(\beta_d - \beta_s) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d} \\ \frac{\alpha_s(X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t) - \alpha_d(X_t'\beta_s + Z_{st}'\gamma_s + u_t)}{\alpha_s - \alpha_d} \end{bmatrix}
$$

Data Generating Process: The Roy Model

- For the Roy Model we need to add some more structure to go from an economic model into an econometric model .
- This means writing down the full data generation model.
- First a normalization is in order. We can redefine the units of F and H arbitrarily. Lets normalize

$$\pi_F = \pi_H = 1$$

- We consider the model

$$W_{fi} = g_f(X_{fi}, X_{0i}) + \epsilon_{fi}$$
$$W_{hi} = g_h(X_{hi}, X_{0i}) + \epsilon_{hi}$$

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

Data Generating Process: The Roy Model

- For the Roy Model we need to add some more structure to go from an economic model into an econometric model .
- This means writing down the full data generation model.
- First a normalization is in order. We can redefine the units of F and H arbitrarily. Lets normalize

$$\pi_F = \pi_H = 1$$

- We consider the model

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

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

Data Generating Process: The Roy Model

- We can observe F_i and W_i:

$$W_i = F_i W_{fi} + (1 - F_i) W_{hi}$$

- Thus for the Roy model

$$\begin{aligned}
Y_i =& (F_i, W_i) \\
X_i =& (X_{0i}, X_{fi}, X_{hi}) \\
u_t =& (\epsilon_{fi}, \epsilon_{hi}) \\
\theta =& (g_f, g_h, G) \\
y_0(X_i, u_i; \theta) =& \begin{bmatrix} \mathrm{I}\left(g_f(X_{fi}, X_{0i}) + \epsilon_{fi} > g_h(X_{hi}, X_{0i}) + \epsilon_{hi} \right) \\ \max\left(g_f(X_{fi}, X_{0i}) + \epsilon_{fi}, g_h(X_{hi}, X_{0i}) + \epsilon_{hi} \right) \end{bmatrix}
\end{aligned}$$

Data Generating Process: The Roy Model

- We can observe F_i and W_i:

$$W_i = F_i W_{fi} + (1 - F_i) W_{hi}$$

- Thus for the Roy model

$$\begin{aligned}
Y_i =& (F_i, W_i) \\
X_i =& (X_{0i}, X_{fi}, X_{hi}) \\
u_t =& (\epsilon_{fi}, \epsilon_{hi}) \\
\theta =& (g_f, g_h, G) \\
y_0(X_i, u_i; \theta) =& \begin{bmatrix} \mathrm{I}\left(g_f(X_{fi}, X_{0i}) + \epsilon_{fi} > g_h(X_{hi}, X_{0i}) + \epsilon_{hi} \right) \\ \max\left(g_f(X_{fi}, X_{0i}) + \epsilon_{fi}, g_h(X_{hi}, X_{0i}) + \epsilon_{hi} \right) \end{bmatrix}
\end{aligned}$$

Identification?

- Another term that means different things to different people
- This is what it means to me:

Identification?

- Another term that means different things to different people
- This is what it means to me:

Identification?

Identification?

- I want to think about it in an econometric way
- This will all be about the Population
- In thinking about identification we will completely ignore sampling issues.
- The model is identified if there is a unique θ that could have generated the population distribution of the observable data (X_i, Y_i).

Identification?

- I want to think about it in an econometric way
- This will all be about the Population
- In thinking about identification we will completely ignore sampling issues.
- The model is identified if there is a unique θ that could have generated the population distribution of the observable data (X_i, Y_i).

Identification?

- A bit more formally, let Θ be the parameter space of θ and let θ_0 be the true value
- If there is some other $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$ for which the joint distribution of (X_i, Y_i) when generated by θ_1 is identical to the joint distribution of (X_i, Y_i) when generated by θ_0 then θ is not identified.
- If there is no such θ_1 then θ is identified.

Identification?

- A bit more formally, let Θ be the parameter space of θ and let θ_0 be the true value
- If there is some other $\theta_1 \in \Theta$ with $\theta_1 \neq \theta_0$ for which the joint distribution of (X_i, Y_i) when generated by θ_1 is identical to the joint distribution of (X_i, Y_i) when generated by θ_0 then θ is not identified.
- If there is no such θ_1 then θ is identified.

Identification of Simultaneous Equation Model

$$Q_t = \alpha_s P_t + X_t' \beta_s + Z_{st}' \gamma_s + u_t$$

$$Q_t = \alpha_d P_t + X_t' \beta_d + Z_{dt}' \gamma_d + \nu_t$$

- Here the hard part is going to be identifying α_s and α_d. Why?
- given $(\alpha_s, \alpha_d, \beta_s, \beta_d, \gamma_s, \gamma_d)$ getting the joint distribution of the error terms is trivial
- Since this is symmetric, let's focus on identification of α_s
- We will also use the assumptions

$$E(u_t | X_t, Z_{dt}, Z_{st}) = 0$$

$$E(\nu_t | X_t, Z_{dt}, Z_{st}) = 0$$

Identification of Simultaneous Equation Model

- Can we just run a regression of Q_t on P_t and Z_t to estimate α_d and γ?
- Think about the "reduced form equation"

$$P_t = \frac{X_t'(\beta_s - \beta_d) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d(1+\tau)}$$

- since ν_t is a direct determinant of P_t, P_t is correlated with ν_t so OLS is not consistent
- So is α_s identified? The key will be Z_{dt}
- We will think about this in three different ways. The first two will involve the use of the "reduced form"

Identification of Simultaneous Equation Model

- Can we just run a regression of Q_t on P_t and Z_t to estimate α_d and γ?
- Think about the "reduced form equation"

$$P_t = \frac{X_t'(\beta_s - \beta_d) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d(1+\tau)}$$

- since ν_t is a direct determinant of P_t, P_t is correlated with ν_t so OLS is not consistent
- So is α_s identified? The key will be Z_{dt}
- We will think about this in three different ways. The first two will involve the use of the "reduced form"

Identification of Simultaneous Equation Model

- Can we just run a regression of Q_t on P_t and Z_t to estimate α_d and γ?
- Think about the "reduced form equation"

$$P_t = \frac{X_t'(\beta_s - \beta_d) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d(1+\tau)}$$

- since ν_t is a direct determinant of P_t, P_t is correlated with ν_t so OLS is not consistent
- So is α_s identified? The key will be Z_{dt}
- We will think about this in three different ways. The first two will involve the use of the "reduced form"

Identification of Simultaneous Equation Model: Method 1

Let's define δ^*_{px}, δ^*_{pd}, δ^*_{ps}, and ν^*_p implicitly as:

$$P_t = \frac{X_t'(\beta_s - \beta_d) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d}$$
$$\equiv X_t\delta^*_{px} + Z_{dt}'\delta^*_{pd} + Z_{st}'\delta^*_{ps} + \nu^*_p$$

where

$$\delta^*_{px} \equiv \frac{\beta_s - \beta_d}{\alpha_s - \alpha_d}$$
$$\delta^*_{pd} \equiv \frac{\gamma_d}{\alpha_s - \alpha_d}$$
$$\delta^*_{ps} \equiv \frac{\gamma_s}{\alpha_s - \alpha_d}$$
$$\nu^*_p \equiv \frac{\nu_t - u_t}{\alpha_s - \alpha_d}$$

Which parameters are identified?

Identification of Simultaneous Equation Model: Method 1

Let's define δ^*_{px}, δ^*_{pd}, δ^*_{ps}, and ν^*_p implicitly as:

$$
\begin{aligned}
P_t =& \frac{X_t'(\beta_s - \beta_d) + Z_{dt}'\gamma_d - Z_{st}'\gamma_s + \nu_t - u_t}{\alpha_s - \alpha_d} \\
\equiv& X_t\delta^*_{px} + Z_{dt}'\delta^*_{pd} + Z_{st}'\delta^*_{ps} + \nu^*_p
\end{aligned}
$$

where

$$
\begin{aligned}
\delta^*_{px} \equiv& \frac{\beta_s - \beta_d}{\alpha_s - \alpha_d} \\
\delta^*_{pd} \equiv& \frac{\gamma_d}{\alpha_s - \alpha_d} \\
\delta^*_{ps} \equiv& \frac{\gamma_s}{\alpha_s - \alpha_d} \\
\nu^*_p \equiv& \frac{\nu_t - u_t}{\alpha_s - \alpha_d}
\end{aligned}
$$

Which parameters are identified?

Identification of Simultaneous Equation Model: Method 1

- Note that $E(\nu_t^*|X_t, Z_{dt}, Z_{st}) = 0$, so one can identify $\delta_p^* = (\delta_{px}^*, \delta_{pd}^*, \delta_{ps}^*)$ by regressing P_t on $W_t = (X_t, Z_{dt}, Z_{st})$.
- That is:

$$
\begin{aligned}
E[W_t'W_t]^{-1}E[W_tP_t] =& E[W_t'W_t]^{-1}E[W_t(W_t'\delta_p^* + \nu_p^*)] \\
=& E[W_t'W_t]^{-1}E[W_tW_t'\delta_p^*] + E[W_t'W_t]^{-1}E[W_t\nu_p^*] \\
=& \delta_p^*
\end{aligned}
$$

- Notice that we used $E[W_t\nu_t^*] = 0$. Why?

Identification of Simultaneous Equation Model: Method 1

- Note that $E(\nu_t^*|X_t, Z_{dt}, Z_{st}) = 0$, so one can identify $\delta_p^* = (\delta_{px}^*, \delta_{pd}^*, \delta_{ps}^*)$ by regressing P_t on $W_t = (X_t, Z_{dt}, Z_{st})$.
- That is:

$$
\begin{aligned}
E[W_t' W_t]^{-1} E[W_t P_t] =& E[W_t' W_t]^{-1} E[W_t (W_t' \delta_p^* + \nu_p^*)] \\
=& E[W_t' W_t]^{-1} E[W_t W_t' \delta_p^*] + E[W_t' W_t]^{-1} E[W_t \nu_p^*] \\
=& \delta_p^*
\end{aligned}
$$

- Notice that we used $E[W_t \nu_t^*] = 0$. Why?

Identification of Simultaneous Equation Model: Method 1

- This is called the "reduced form" equation for P_t
- Note that the parameters here are not the fundamental structural parameters themselves, but they are a known function of these parameters
- To me this is the classic definition of reduced form (you need to have a structural model)
- How is this useful for identifying the model?

Identification of Simultaneous Equation Model: Method 1

We can also solve for the reduced form for Q_t

$$
\begin{aligned}
Q_t = & \frac{\alpha_s(X_t'\beta_d + Z_{dt}'\gamma_d + \nu_t) - \alpha_d(X_t'\beta_s + Z_{st}'\gamma_s + u_t)}{\alpha_s - \alpha_d} \\
\equiv & X_t\delta_{qx}^* + Z_{dt}'\delta_{qd}^* + Z_{st}'\delta_{qs}^* + \nu_q^*
\end{aligned}
$$

where

$$
\begin{aligned}
\delta_{qx}^* \equiv & \frac{\alpha_s\beta_s - \alpha_d\beta_d}{\alpha_s - \alpha_d} \\
\delta_{qd}^* \equiv & \frac{\alpha_s\gamma_d}{\alpha_s - \alpha_d} \\
\delta_{qs}^* \equiv & \frac{-\alpha_d\gamma_s}{\alpha_s - \alpha_d} \\
\nu_p^* \equiv & \frac{\alpha_s\nu_t - \alpha_d u_t}{\alpha_s - \alpha_d}
\end{aligned}
$$

Which parameters are identified?

Identification of Simultaneous Equation Model: Method 1

- Like the other reduced form, we can identify δ_q^* by regressing Q_t on $W_t = (X_t, Z_{dt}, Z_{st})$.
- That is:

$$
\begin{aligned}
E[W_t' W_t]^{-1} E[W_t Q_t] &= E[W_t' W_t]^{-1} E[W_t (W_t' \delta_q^* + \nu_q^*)] \\
&= E[W_t' W_t]^{-1} E[W_t W_t' \delta_p^*] + E[W_t' W_t]^{-1} E[W_t \nu_q^*)] \\
&= \delta_q^*
\end{aligned}
$$

Identification of Simultaneous Equation Model: Method 1

- 6 equations in 6 unknowns $(\alpha_s, \alpha_d, \beta_s, \beta_d, \gamma_s, \gamma_s)$
- That is:

$$\begin{bmatrix} \delta^*_{px} \\ \delta^*_{pd} \\ \delta^*_{ps} \end{bmatrix} = \begin{bmatrix} \frac{\beta_s - \beta_d}{\alpha_s - \alpha_d} \\ \frac{\gamma_d}{\alpha_s - \alpha_d} \\ \frac{\gamma_s}{\alpha_s - \alpha_d} \end{bmatrix} \text{ and } \begin{bmatrix} \delta^*_{qx} \\ \delta^*_{qd} \\ \delta^*_{qs} \end{bmatrix} = \begin{bmatrix} \frac{\alpha_s \beta_s - \alpha_d \beta_d}{\alpha_s - \alpha_d} \\ \frac{\alpha_s \gamma_d}{\alpha_s - \alpha_d} \\ \frac{-\alpha_d \gamma_s}{\alpha_s - \alpha_d} \end{bmatrix}$$

- Can you identify α_s? What implicit and explicit conditions do you need? exclusion restriction ...
- What about α_d?

Identification of Simultaneous Equation Model: Method 1

- 6 equations in 6 unknowns $(\alpha_s, \alpha_d, \beta_s, \beta_d, \gamma_s, \gamma_s)$
- That is:

$$\begin{bmatrix} \delta^*_{px} \\ \delta^*_{pd} \\ \delta^*_{ps} \end{bmatrix} = \begin{bmatrix} \frac{\beta_s - \beta_d}{\alpha_s - \alpha_d} \\ \frac{\gamma_d}{\alpha_s - \alpha_d} \\ \frac{\gamma_s}{\alpha_s - \alpha_d} \end{bmatrix} \text{ and } \begin{bmatrix} \delta^*_{qx} \\ \delta^*_{qd} \\ \delta^*_{qs} \end{bmatrix} = \begin{bmatrix} \frac{\alpha_s \beta_s - \alpha_d \beta_d}{\alpha_s - \alpha_d} \\ \frac{\alpha_s \gamma_d}{\alpha_s - \alpha_d} \\ \frac{-\alpha_d \gamma_s}{\alpha_s - \alpha_d} \end{bmatrix}$$

- Can you identify α_s? What implicit and explicit conditions do you need? exclusion restriction ...
- What about α_d?

Identification of Simultaneous Equation Model: Method 1

- Notice that
$$\frac{\delta^*_{qd}}{\delta^*_{pd}} = \alpha_s$$
- So we can identify α_s simply by taking the ratio of the reduced form coefficients on Z_{dt}
- Intuition:
$$\frac{dQ^s_t(P_t)}{dZ_{dt}} = \frac{\partial dQ^s_t(P_t)}{\partial P_t}\frac{\partial P_t}{\partial Z_{dt}}$$
- **Exclusion Restriction:** If Z_{dt} affects Q_t in any way other than altering the price through the demand curve then this won't work
- Notice that the same argument will work for α_d with the Z_{st} coefficients.

Identification of Simultaneous Equation Model: Method 1

- Notice that
$$\frac{\delta^*_{qd}}{\delta^*_{pd}} = \alpha_s$$
- So we can identify α_s simply by taking the ratio of the reduced form coefficients on Z_{dt}
- Intuition:
$$\frac{dQ^s_t(P_t)}{dZ_{dt}} = \frac{\partial dQ^s_t(P_t)}{\partial P_t}\frac{\partial P_t}{\partial Z_{dt}}$$
- **Exclusion Restriction:** If Z_{dt} affects Q_t in any way other than altering the price through the demand curve then this won't work
- Notice that the same argument will work for α_d with the Z_{st} coefficients.

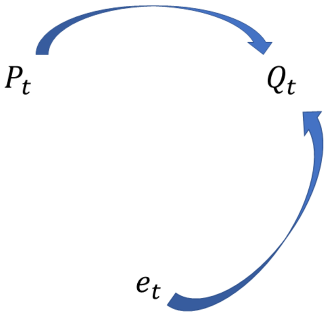
Identification of Simultaneous Equation Model: Method 1

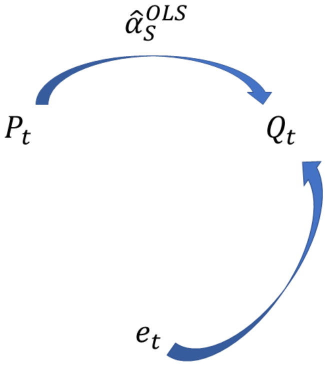
- Notice that
$$\frac{\delta^*_{qd}}{\delta^*_{pd}} = \alpha_s$$
- So we can identify α_s simply by taking the ratio of the reduced form coefficients on Z_{dt}
- Intuition:
$$\frac{dQ^s_t(P_t)}{dZ_{dt}} = \frac{\partial dQ^s_t(P_t)}{\partial P_t} \frac{\partial P_t}{\partial Z_{dt}}$$
- **Exclusion Restriction:** If Z_{dt} affects Q_t in any way other than altering the price through the demand curve then this won't work
- Notice that the same argument will work for α_d with the Z_{st} coefficients.

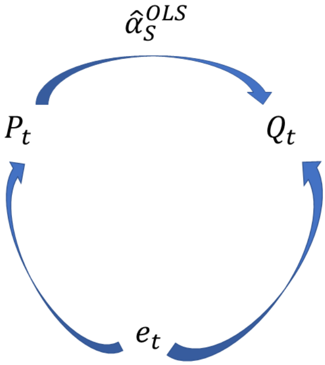
Identification of Simultaneous Equation Model: Method 1

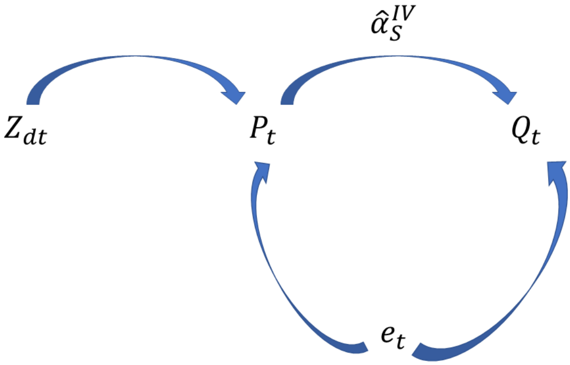
- Z_{dt} is called **Instrument Variable**.
- IV estimator is

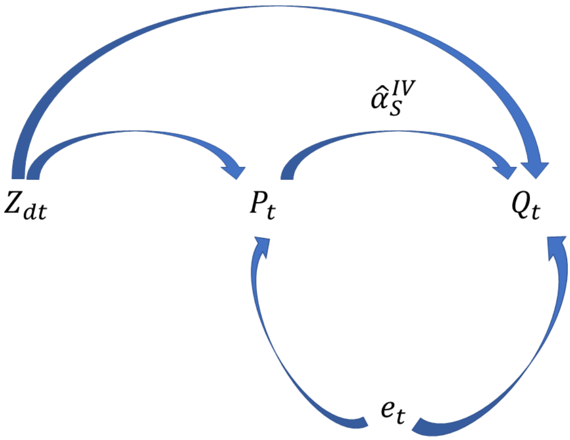
$$\hat{\alpha}_s^{IV} = \frac{\delta_{qd}^*}{\delta_{pd}^*} = \frac{\text{Reduced Form}}{\text{First Stage}}$$

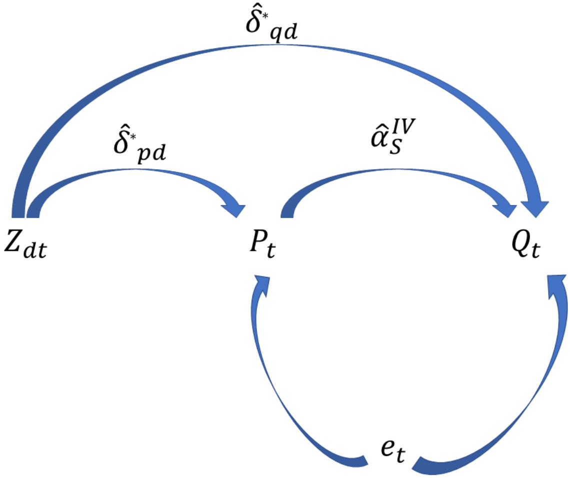

P_t
Q_t
e_t

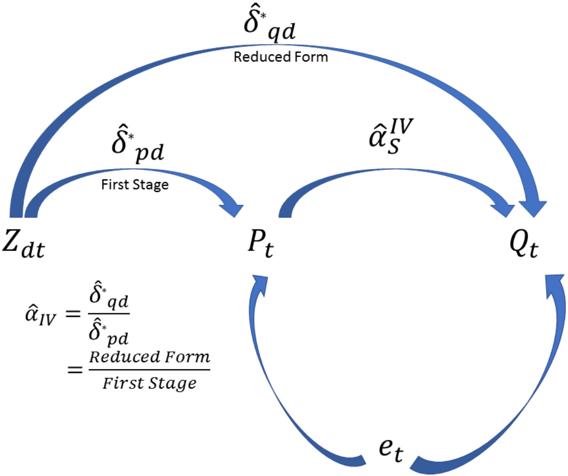

$\hat{\alpha}_S^{OLS}$
P_t
Q_t
e_t


$\hat{\alpha}_S^{OLS}$
P_t
Q_t
e_t


$\hat{\alpha}_S^{IV}$
Z_{dt}
P_t
Q_t
e_t


$\hat{\alpha}_S^{IV}$
Z_{dt}
P_t
Q_t
e_t


$\hat{\delta}^*_{qd}$
$\hat{\delta}^*_{pd}$
$\hat{\alpha}_S^{IV}$
Z_{dt}
P_t
Q_t
e_t



$$\hat{\alpha}_{IV} = \frac{\hat{\delta}^*{}_{qd}}{\hat{\delta}^*{}_{pd}} = \frac{\text{Reduced Form}}{\text{First Stage}}$$

Identification of Simultaneous Equation Model: Method 2

Define

$$P_t^* = X_t \delta_{px}^* + Z'_{dt} \delta_{pd}^* + Z'_{st} \delta_{ps}^*$$

so

$$P_t = P_t^* + \nu_p^*$$

- This is identified since $\delta_p^* = (\delta_{px}^*, \delta_{pd}^*, \delta_{ps}^*)$ is identified.
- Now notice that:

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t' \beta_s + Z'_{st} \gamma_s + u_t \\ =& \alpha_s (P_t^* + \nu_p^*) + X_t' \beta_s + Z'_{st} \gamma_s + u_t \\ =& \alpha_s P_t^* + X_t' \beta_s + Z'_{st} \gamma_s + u_t + \alpha_s \nu_p^* \end{aligned}$$

- Is α_s now identified? Why?

Identification of Simultaneous Equation Model: Method 2

Define

$$P_t^* = X_t \delta_{px}^* + Z_{dt}' \delta_{pd}^* + Z_{st}' \delta_{ps}^*$$

so

$$P_t = P_t^* + \nu_p^*$$

- This is identified since $\delta_p^* = (\delta_{px}^*, \delta_{pd}^*, \delta_{ps}^*)$ is identified.
- Now notice that:

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t' \beta_s + Z_{st}' \gamma_s + u_t \\ =& \alpha_s (P_t^* + \nu_p^*) + X_t' \beta_s + Z_{st}' \gamma_s + u_t \\ =& \alpha_s P_t^* + X_t' \beta_s + Z_{st}' \gamma_s + u_t + \alpha_s \nu_p^* \end{aligned}$$

- Is α_s now identified? Why?

Identification of Simultaneous Equation Model: Method 2

Define

$$P_t^* = X_t \delta_{px}^* + Z_{dt}' \delta_{pd}^* + Z_{st}' \delta_{ps}^*$$

so

$$P_t = P_t^* + \nu_p^*$$

- This is identified since $\delta_p^* = (\delta_{px}^*, \delta_{pd}^*, \delta_{ps}^*)$ is identified.
- Now notice that:

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t' \beta_s + Z_{st}' \gamma_s + u_t \\ =& \alpha_s (P_t^* + \nu_p^*) + X_t' \beta_s + Z_{st}' \gamma_s + u_t \\ =& \alpha_s P_t^* + X_t' \beta_s + Z_{st}' \gamma_s + u_t + \alpha_s \nu_p^* \end{aligned}$$

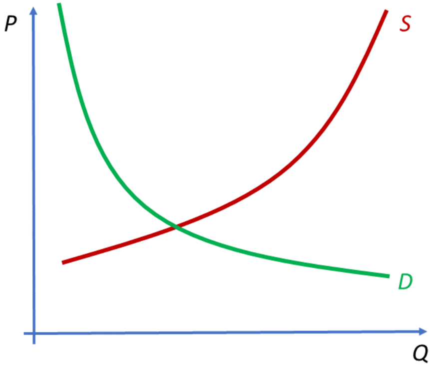
- Is α_s now identified? Why?

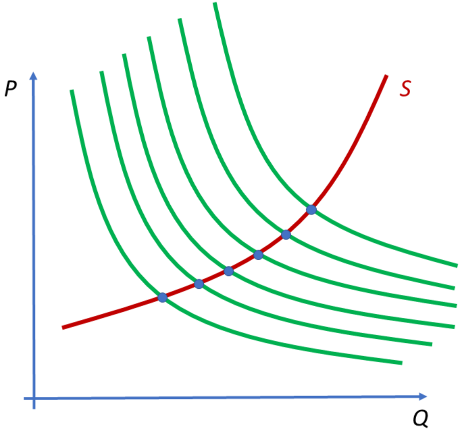
Identification of Simultaneous Equation Model: Method 2

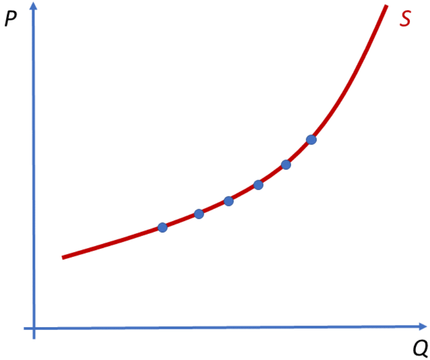
One could get a consistent estimate of α_s by regressing Q_t on $W_t^* = (P_t^*, X_t, Z_{st})$. That is:

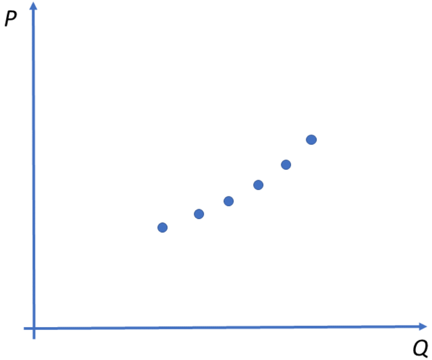
$$\begin{aligned}
E[W_t^* W_t^{*'}]^{-1} E[W_t^{*'} P_t] =& E[W_t^* W_t^{*'}]^{-1} E\left[W_t^* \left(W_t^{*'} \begin{bmatrix}\alpha_s \\ \beta_s \\ \gamma_s\end{bmatrix} + u_t + \alpha_s \nu_{pt}^*\right)\right] \\
=& \begin{bmatrix}\alpha_s \\ \beta_s \\ \gamma_s\end{bmatrix} + E[W_t^* W_t^{*'}]^{-1} E[W_t^* (u_t + \alpha_s \nu_{pt}^*)] \\
=& \begin{bmatrix}\alpha_s \\ \beta_s \\ \gamma_s\end{bmatrix}
\end{aligned}$$

This is the so called two step IV estimation.


P
S
D
Q


P
S
Q


P
S
Q


P
Q

Identification of Simultaneous Equation Model: Method 3

For any random variable Y_t define

$$\tilde{Y}_t = Y_t - E(Y_t|X_t, Z_{st})$$

Since

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t'\beta_s + Z_{st}'\gamma_s + u_t \\ E(Q_t|X_t, Z_{st}) =& \alpha_s E(P_t|X_t, Z_{st}) + X_t'\beta_s + Z_{st}'\gamma_s \end{aligned}$$

then

$$\tilde{Q}_t = \alpha_s \tilde{P}_t + u_t$$

Therefore

$$cov(\tilde{Z}_{dt}, \tilde{Q}_t) = \alpha_s cov(\tilde{Z}_{dt}, \tilde{P}_t) + cov(\tilde{Z}_{dt}, u_t)$$

Identification of Simultaneous Equation Model: Method 3

For any random variable Y_t define

$$\tilde{Y}_t = Y_t - E(Y_t|X_t, Z_{st})$$

Since

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t'\beta_s + Z_{st}'\gamma_s + u_t \\ E(Q_t|X_t, Z_{st}) =& \alpha_s E(P_t|X_t, Z_{st}) + X_t'\beta_s + Z_{st}'\gamma_s \end{aligned}$$

then

$$\tilde{Q}_t = \alpha_s \tilde{P}_t + u_t$$

Therefore

$$cov(\tilde{Z}_{dt}, \tilde{Q}_t) = \alpha_s cov(\tilde{Z}_{dt}, \tilde{P}_t) + cov(\tilde{Z}_{dt}, u_t)$$

Identification of Simultaneous Equation Model: Method 3

For any random variable Y_t define

$$\tilde{Y}_t = Y_t - E(Y_t|X_t, Z_{st})$$

Since

$$\begin{aligned} Q_t =& \alpha_s P_t + X_t'\beta_s + Z_{st}'\gamma_s + u_t \\ E(Q_t|X_t, Z_{st}) =& \alpha_s E(P_t|X_t, Z_{st}) + X_t'\beta_s + Z_{st}'\gamma_s \end{aligned}$$

then

$$\tilde{Q}_t = \alpha_s \tilde{P}_t + u_t$$

Therefore

$$cov(\tilde{Z}_{dt}, \tilde{Q}_t) = \alpha_s cov(\tilde{Z}_{dt}, \tilde{P}_t) + cov(\tilde{Z}_{dt}, u_t)$$

Identification of Simultaneous Equation Model: Method 3

Exclusion Restriction:

$$cov(\tilde{Z}_{dt}, u_t) = 0$$

Then

$$\alpha_s = \frac{cov(\tilde{Z}_{dt}, \tilde{Q}_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)} = \frac{\text{reduced form}}{\text{first stage}}$$

Let's think about the denominator (first stage), $cov(\tilde{Z}_{dt}, \tilde{P}_t)$.

$$\begin{aligned} P_t =& X_t\delta^*_{px} + Z'_{dt}\delta^*_{pd} + Z'_{st}\delta^*_{ps} + \nu^*_p \\ E(P_t|X_t, Z_{st}) =& X_t\delta^*_{px} + E(Z'_{dt}|X_t, Z_{st})\delta^*_{pd} + Z'_{st}\delta^*_{ps} + \nu^*_p \\ \tilde{P}_t =& \tilde{Z}_{dt}\delta^*_{pd} + \nu^*_{pt} \end{aligned}$$

Thus

$$cov(\tilde{Z}_{dt}, \tilde{P}_t) = \delta^*_{pd} var(\tilde{Z}_{dt})$$

which will be non zero if $\delta^*_{pd} \neq 0$ and $var(\tilde{Z}_{dt}) \neq 0$

Identification of Simultaneous Equation Model: Method 3

Exclusion Restriction:

$$cov(\tilde{Z}_{dt}, u_t) = 0$$

Then

$$\alpha_s = \frac{cov(\tilde{Z}_{dt}, \tilde{Q}_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)} = \frac{\text{reduced form}}{\text{first stage}}$$

Let's think about the denominator (first stage), $cov(\tilde{Z}_{dt}, \tilde{P}_t)$.

$$\begin{aligned} P_t =& X_t\delta^*_{px} + Z'_{dt}\delta^*_{pd} + Z'_{st}\delta^*_{ps} + \nu^*_p \\ E(P_t|X_t, Z_{st}) =& X_t\delta^*_{px} + E(Z'_{dt}|X_t, Z_{st})\delta^*_{pd} + Z'_{st}\delta^*_{ps} + \nu^*_p \\ \tilde{P}_t =& \tilde{Z}_{dt}\delta^*_{pd} + \nu^*_{pt} \end{aligned}$$

Thus

$$cov(\tilde{Z}_{dt}, \tilde{P}_t) = \delta^*_{pd} var(\tilde{Z}_{dt})$$

which will be non zero if $\delta^*_{pd} \neq 0$ and $var(\tilde{Z}_{dt}) \neq 0$

Identification of Simultaneous Equation Model: Method 3

Exclusion Restriction:

$$cov(\tilde{Z}_{dt}, u_t) = 0$$

Then

$$\alpha_s = \frac{cov(\tilde{Z}_{dt}, \tilde{Q}_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)} = \frac{\text{reduced form}}{\text{first stage}}$$

Let's think about the denominator (first stage), $cov(\tilde{Z}_{dt}, \tilde{P}_t)$.

$$\begin{aligned}
P_t =& X_t \delta^*_{px} + Z'_{dt}\delta^*_{pd} + Z'_{st}\delta^*_{ps} + \nu^*_p \\
E(P_t|X_t, Z_{st}) =& X_t \delta^*_{px} + E(Z'_{dt}|X_t, Z_{st})\delta^*_{pd} + Z'_{st}\delta^*_{ps} + \nu^*_p \\
\tilde{P}_t =& \tilde{Z}_{dt}\delta^*_{pd} + \nu^*_{pt}
\end{aligned}$$

Thus

$$cov(\tilde{Z}_{dt}, \tilde{P}_t) = \delta^*_{pd} var(\tilde{Z}_{dt})$$

which will be non zero if $\delta^*_{pd} \neq 0$ and $var(\tilde{Z}_{dt}) \neq 0$

Identification of Simultaneous Equation Model: Method 3

One can see the importance of the **Exclusion Restriction** assumption: $cov(\tilde{Z}_{dt}, u_t) = 0$

$$\hat{\alpha}_s^{IV} = \frac{cov(\tilde{Z}_{dt}, \tilde{Q}_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)}$$
$$= \alpha_s + \frac{cov(\tilde{Z}_{dt}, u_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)}$$

Good IV: In order for the model to be consistent you need:

1. **Exclusion Restriction:** $cov(\tilde{Z}_{dt}, u_t) = 0$
2. **First Stage:** $cov(\tilde{Z}_{dt}, \tilde{P}_t) \neq 0$

But more generally for the asymptotic bias to be small you want $cov(\tilde{Z}_{dt}, u_t)$ to be small and $|cov(\tilde{Z}_{dt}, \tilde{P}_t)|$ to be large.

Identification of Simultaneous Equation Model: Method 3

One can see the importance of the **Exclusion Restriction** assumption: $cov(\tilde{Z}_{dt}, u_t) = 0$

$$\begin{aligned}\hat{\alpha}_s^{IV} =& \frac{cov(\tilde{Z}_{dt}, \tilde{Q}_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)} \\ =& \alpha_s + \frac{cov(\tilde{Z}_{dt}, u_t)}{cov(\tilde{Z}_{dt}, \tilde{P}_t)}\end{aligned}$$

Good IV: In order for the model to be consistent you need:

1. **Exclusion Restriction:** $cov(\tilde{Z}_{dt}, u_t) = 0$
2. **First Stage:** $cov(\tilde{Z}_{dt}, \tilde{P}_t) \neq 0$

But more generally for the asymptotic bias to be small you want $cov(\tilde{Z}_{dt}, u_t)$ to be small and $|cov(\tilde{Z}_{dt}, \tilde{P}_t)|$ to be large.

Identification of The Roy Model

- Let's think about the Roy Model.
- This is discussed in Heckman and Honore (EMA, 1990)
- We will follow the discussion in French and Taber, Handbook of Labor Economics, (2011)
- we will review identification only non-parametrically in this lecture note.
- In problem set 5, you will study the identification parametrically
- While the model is about the simplest in the world, identification is difficult and non-trivial.

Why is nonparametric identification useful?

- Chris Taber: "I always begin a research project by thinking about nonparametric identification."
- Literature on nonparametric identification not particularly highly cited
- At the same time this literature has had a huge impact on empirical work in practice.
- A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days because of nonparametric identification
- It is also 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

Why is nonparametric identification useful?

- Chris Taber: "I always begin a research project by thinking about nonparametric identification."
- Literature on nonparametric identification not particularly highly cited
- At the same time this literature has had a huge impact on empirical work in practice.
- A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days because of nonparametric identification
- It is also 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

Why is nonparametric identification useful?

- Chris Taber: "I always begin a research project by thinking about nonparametric identification."
- Literature on nonparametric identification not particularly highly cited
- At the same time this literature has had a huge impact on empirical work in practice.
- A Heckman two step model without an exclusion restriction is often viewed as highly problematic these days because of nonparametric identification
- It is also 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

Non-Parametric Identification of The Roy Model

- Recall the Roy model

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

- where the joint distribution of $(\epsilon_{fi}, \epsilon_{hi})$ is G.
- in this case $\theta = (g_f, g_h, G)$
- Of course, we don't care about fishing and hunting! Depending on the context, you can change these two.
- Let's think of Fishing as jobs in the formal sector and Hunting as jobs in informal sector.
- Another usage: Fishermen $\rightarrow$ Employed, Hunters $\rightarrow$ Unemployed

Non-Parametric Identification of The Roy Model

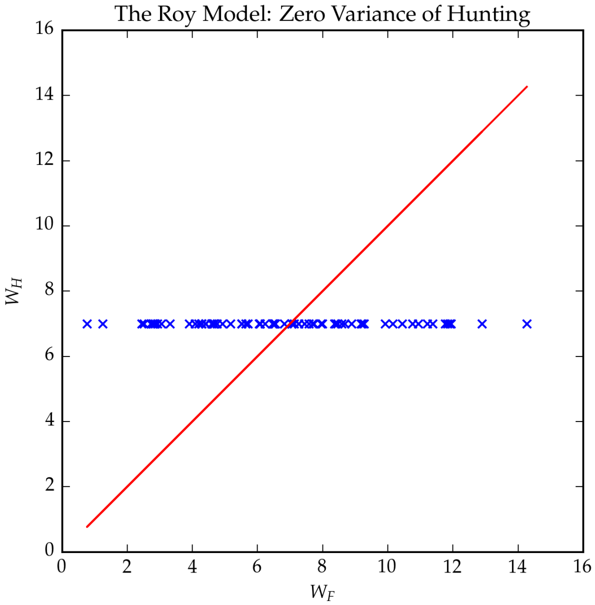
- Recall the Roy model

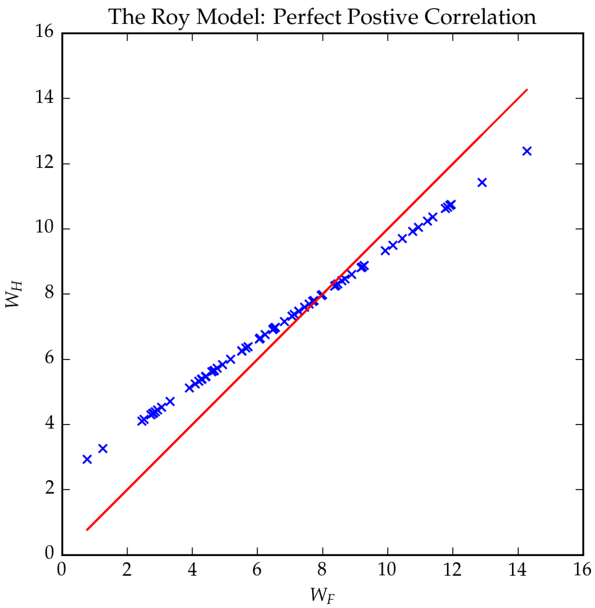
$$W_{fi} = g_f(X_{fi}, X_{0i}) + \epsilon_{fi}$$
$$W_{hi} = g_h(X_{hi}, X_{0i}) + \epsilon_{hi}$$

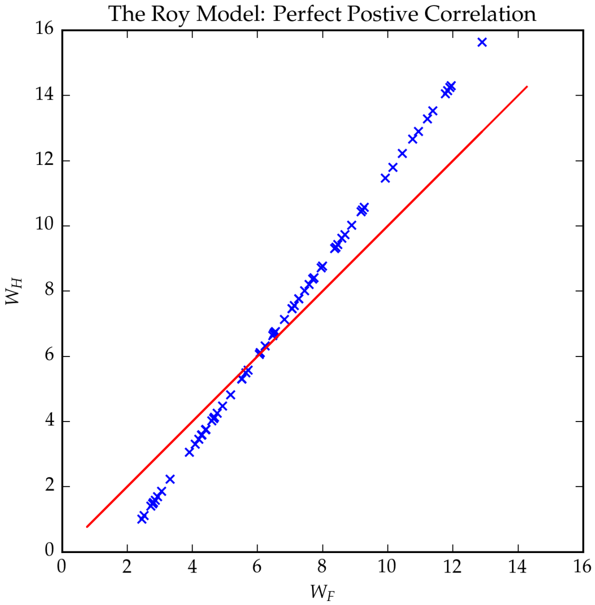
- where the joint distribution of $(\epsilon_{fi}, \epsilon_{hi})$ is G.
- in this case $\theta = (g_f, g_h, G)$
- Of course, we don't care about fishing and hunting! Depending on the context, you can change these two.
- Let's think of Fishing as jobs in the formal sector and Hunting as jobs in informal sector.
- Another usage: Fishermen $\rightarrow$ Employed, Hunters $\rightarrow$ Unemployed

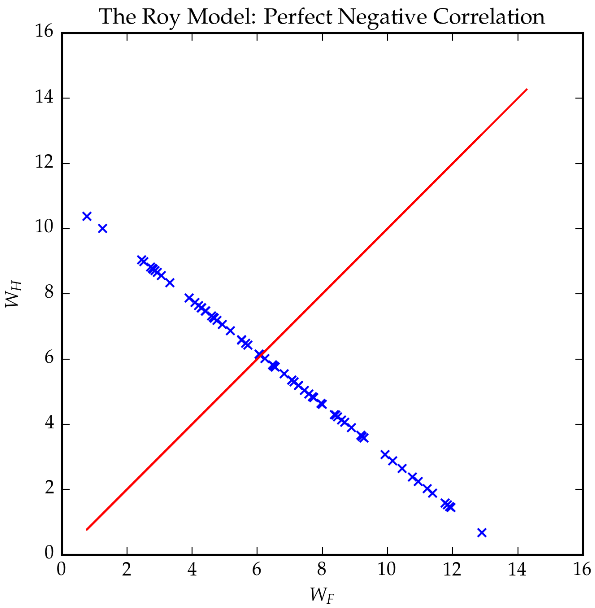
Does G Matter?

- You may ask does G matter? YES!
- think about extereme cases:
  1. $var(\epsilon_{fi}) = 0$ (Reservation Wage models)
  2. perfect small positive correlation
  3. perfect large positive correlation
  4. perfect negative correlation


The Roy Model: Zero Variance of Hunting
16
14
12
10
8
6
4
2
0
W_H
0
2
4
6
8
10
12
14
16
W_F


The Roy Model: Perfect Postive Correlation
W_H
W_F
0
2
4
6
8
10
12
14
16


The Roy Model: Perfect Postive Correlation
W_H
W_F
0
2
4
6
8
10
12
14
16



The Roy Model: Perfect Negative Correlation
16
14
12
10
8
6
4
2
0
W_H
0
2
4
6
8
10
12
14
16
W_F

Assumptions

- $(\epsilon_{fi}, \epsilon_{hi})$ is independent of $X_i = (X_{0i}, X_{fi}, X_{hi})$
- Normalize $E(\epsilon_{fi}) = 0$
- Normalize the median of $\epsilon_{fi} - \epsilon_{hi}$ to zero.
- Exclusion Restriction:

$$supp\Big(g_f(X_{fi}, x_0), g_h(X_{hi}, x_{0i})\Big) = \mathbf{R}^2$$

  for all $x_0 \in supp(X_{0i})$
- Heckman and Honroe (1991) suggest a four step identification procedure.

Step 1: Identification of Reduced Form 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

$$\begin{aligned}\Pr(F_i = 1 | X_i = x) =& \Pr(\epsilon_{hi} - \epsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0)) \\ =& G_{h-f}(g^*(x))\end{aligned}$$

- where G_{h-f} is the distribution function for $\epsilon_{hi} - \epsilon_{fi}$ and

$$g^*(x) = g_f(x_f, x_0) - g_h(x_h, x_0)$$

- We can not separate g^* from G_{h-f} , but we can identify this as a function of x

Step 1: Identification of Reduced Form 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

$$\begin{aligned}\Pr(F_i = 1 | X_i = x) =& \Pr(\epsilon_{hi} - \epsilon_{fi} \leq g_f(x_f, x_0) - g_h(x_h, x_0)) \\ =& G_{h-f}(g^*(x))\end{aligned}$$

- where G_{h-f} is the distribution function for $\epsilon_{hi} - \epsilon_{fi}$ and

$$g^*(x) = g_f(x_f, x_0) - g_h(x_h, x_0)$$

- We can not separate g^* from G_{h-f} , but we can identify this as a function of x

Step 1: Identification of Reduced Form Choice Model

- We also know that for any two values x^a and x^b, if

$$\Pr(F_i = 1 | X_i = x^a) = \Pr(F_i = 1 | X_i = x^b)$$

- Then

$$g^*(x^a) = g^*(x^b)$$

Step 1: Parametric Identification

- Assume g_f and g_h are linear function of observables X_i,

$$g_f(X_{fi}, X_{0i}) = X_{fi}'\alpha_f + X_{0i}\beta_f$$

$$g_h(X_{hi}, X_{0i}) = X_{hi}'\alpha_h + X_{0i}\beta_h$$

- Moreover assume that $G_{h-f} \sim N(0, \Sigma)$
- Then what would step 1 look like?
- see problem set 5 for this parametric identification ...

Step 2: Identification of the Wage Equation $g_f(X_i)$

- Next consider identification of g_f. This is basically the standard selection problem:
  1. You can observe wage offer for only employed workers
  2. You observe only wage of immigrants
  3. You only observe answers of respondents who choose to answer a survey ...
  4. In the Roy model, we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1)$. In particular we can identify

$$E(W_i|X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{hi}|\epsilon_{fi} - \epsilon_{hi} < g^*(x))$$

- Heckman (1979) proposed a two step method to identify wages when we have selection problem.
- His paper has 24,505 citation and buys him a Nobel prize in 2000.
- This step is essentially the second stage of a Heckman (1979) two step.

Step 2: Identification of the Wage Equation $g_f(X_i)$

- Next consider identification of g_f. This is basically the standard selection problem:
    1. You can observe wage offer for only employed workers
    2. You observe only wage of immigrants
    3. You only observe answers of respondents who choose to answer a survey ...
    4. In the Roy model, we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1)$. In particular we can identify

$$E(W_i | X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{hi} | \epsilon_{fi} - \epsilon_{hi} < g^*(x))$$

- Heckman (1979) proposed a two step method to identify wages when we have selection problem.
- His paper has 24,505 citation and buys him a Nobel prize in 2000.
- This step is essentially the second stage of a Heckman (1979) two step.

Step 2: Identification of the Wage Equation $g_f(X_i)$

- Next consider identification of g_f. This is basically the standard selection problem:
  1. You can observe wage offer for only employed workers
  2. You observe only wage of immigrants
  3. You only observe answers of respondents who choose to answer a survey ...
  4. In the Roy model, we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1)$. In particular we can identify

$$E(W_i | X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{hi} | \epsilon_{fi} - \epsilon_{hi} < g^*(x))$$

- Heckman (1979) proposed a two step method to identify wages when we have selection problem.
- His paper has 24,505 citation and buys him a Nobel prize in 2000.
- This step is essentially the second stage of a Heckman (1979) two step.

Step 2: Identification of the Wage Equation $g_f(X_i)$

- Next consider identification of g_f. This is basically the standard selection problem:
  1. You can observe wage offer for only employed workers
  2. You observe only wage of immigrants
  3. You only observe answers of respondents who choose to answer a survey ...
  4. In the Roy model, we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1)$. In particular we can identify

$$E(W_i|X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{hi}|\epsilon_{fi} - \epsilon_{hi} < g^*(x))$$

- Heckman (1979) proposed a two step method to identify wages when we have selection problem.
- His paper has 24,505 citation and buys him a Nobel prize in 2000.
- This step is essentially the second stage of a Heckman (1979) two step.

Step 2: Identification of the Wage Equation $g_f(X_i)$

- Next consider identification of g_f. This is basically the standard selection problem:
  1. You can observe wage offer for only employed workers
  2. You observe only wage of immigrants
  3. You only observe answers of respondents who choose to answer a survey ...
  4. In the Roy model, we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1)$. In particular we can identify

$$E(W_i | X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{hi} | \epsilon_{fi} - \epsilon_{hi} < g^*(x))$$

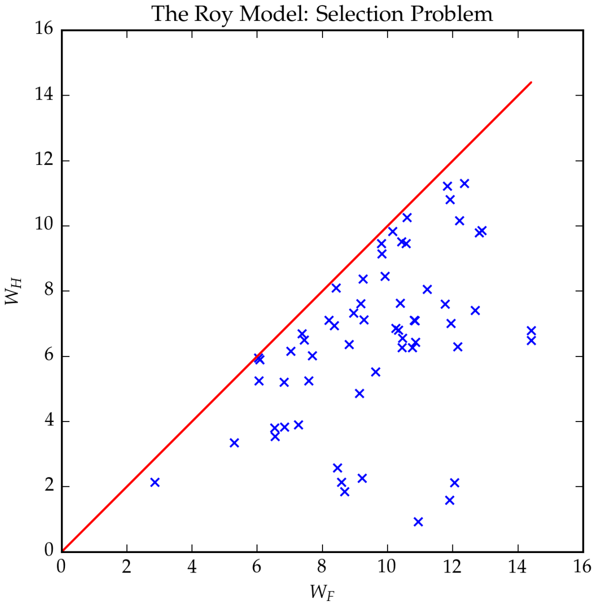
- Heckman (1979) proposed a two step method to identify wages when we have selection problem.
- His paper has 24,505 citation and buys him a Nobel prize in 2000.
- This step is essentially the second stage of a Heckman (1979) two step.

Step 2: Identification of the Wage Equation $g_f(X_i)$

- Next consider identification of g_f. This is basically the standard selection problem:
  1. You can observe wage offer for only employed workers
  2. You observe only wage of immigrants
  3. You only observe answers of respondents who choose to answer a survey ...
  4. In the Roy model, we can identify the distribution of W_{fi} conditional on $(X_i = x, F_i = 1)$. In particular we can identify

$$E(W_i | X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{hi} | \epsilon_{fi} - \epsilon_{hi} < g^*(x))$$

- Heckman (1979) proposed a two step method to identify wages when we have selection problem.
- His paper has 24,505 citation and buys him a Nobel prize in 2000.
- This step is essentially the second stage of a Heckman (1979) two step.



The Roy Model: Selection Problem
W_H
W_F
0
2
4
6
8
10
12
14
16

Step 2: Identification of the Wage Equation $g_f(X_i)$

- We want to identify $g_f(x_f, x_0)$ but we can only identify

$$E(W_i|X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x))$$

- Lets think about identifying g_f up to location.
- That is, for any (x_f^a, x_0^a) and (x_f^b, x_0^b) we want to identify

$$g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b)$$

Step 2: Identification of the Wage Equation $g_f(X_i)$

- We want to identify $g_f(x_f, x_0)$ but we can only identify

$$E(W_i|X_i = x, F_i = 1) = g_f(x_f, x_0) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x))$$

- Lets think about identifying g_f up to location.
- That is, for any (x_f^a, x_0^a) and (x_f^b, x_0^b) we want to identify

$$g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b)$$

Exclusion Restriction

- An exclusion restriction is key,
- Take x_h^b to be any number you want. From step 1 and from the support assumption we know that we can identify a x_h^a such that

$$\Pr\left(F_i = 1 | X_i = (x_0^a, x_h^a, x_f^a)\right) = \Pr\left(F_i = 1 | X_i = (x_0^b, x_h^b, x_f^b)\right)$$

which means that

$$g_f(x_f^a, x_0^a) - g_h(x_h^a, x_0^a) = g_f(x_f^b, x_0^b) - g_h(x_h^b, x_0^b)$$

Exclusion Restriction

- An exclusion restriction is key,
- Take x_h^b to be any number you want. From step 1 and from the support assumption we know that we can identify a x_h^a such that

$$\Pr\left(F_i = 1 | X_i = (x_0^a, x_h^a, x_f^a)\right) = \Pr\left(F_i = 1 | X_i = (x_0^b, x_h^b, x_f^b)\right)$$

which means that

$$g_f(x_f^a, x_0^a) - g_h(x_h^a, x_0^a) = g_f(x_f^b, x_0^b) - g_h(x_h^b, x_0^b)$$

Exclusion Restriction

But then

$$
\begin{aligned}
E(W_i|X_i = x^a, F_i = 1) =& g_f(x_f^a, x_0^a) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^a)) \\
E(W_i|X_i = x^b, F_i = 1) =& g_f(x_f^b, x_0^b) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^b))
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
E(W_i|X_i =& x^a, F_i = 1) - E(W_i|X_i = x^b, F_i = 1) \\
=& g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b) \\
&+ E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^a)) - E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^b)) \\
=& g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b)
\end{aligned}
$$

Exclusion Restriction

But then

$$\begin{aligned}E(W_i|X_i = x^a, F_i = 1) =& g_f(x_f^a, x_0^a) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^a))\\ E(W_i|X_i = x^b, F_i = 1) =& g_f(x_f^b, x_0^b) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^b))\end{aligned}$$

Therefore:

$$\begin{aligned}E(W_i|X_i =& x^a, F_i = 1) - E(W_i|X_i = x^b, F_i = 1)\\ =& g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b)\\ &+ E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^a)) - E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^b))\\ =& g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b)\end{aligned}$$

Exclusion Restriction

But then

$$
\begin{aligned}
E(W_i|X_i = x^a, F_i = 1) =& g_f(x_f^a, x_0^a) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^a)) \\
E(W_i|X_i = x^b, F_i = 1) =& g_f(x_f^b, x_0^b) + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^b))
\end{aligned}
$$

Therefore:

$$
\begin{aligned}
E(W_i|X_i =& x^a, F_i = 1) - E(W_i|X_i = x^b, F_i = 1) \\
=& g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b) \\
& + E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^a)) - E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x^b)) \\
=& g_f(x_f^a, x_0^a) - g_f(x_f^b, x_0^b)
\end{aligned}
$$

Step 2: Identification at Infinity

- What about the location?
- Notice that

$$
\begin{aligned}
&\lim_{\Pr(F_i=1|X_i=x)\to 1} E(W_i|X_i = x, F_i = 1)\\
&\quad = g_f(x_f, x_0) + \lim_{\Pr(F_i=1|X_i=x)\to 1} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x))\\
&\quad = g_f(x_f, x_0) + \lim_{g^*(x)\to\infty} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x))\\
&\quad = g_f(x_f, x_0) + E(\epsilon_{fi})\\
&\quad = g_f(x_f, x_0)
\end{aligned}
$$

- Thus we can identify the location of the wage equation by identification at infinity.

Step 2: Identification at Infinity

- What about the location?
- Notice that

$$
\begin{aligned}
&\lim_{\Pr(F_i=1|X_i=x)\to 1} E(W_i|X_i=x, F_i=1) \\
&= g_f(x_f, x_0) + \lim_{\Pr(F_i=1|X_i=x)\to 1} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + \lim_{g^*(x)\to\infty} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + E(\epsilon_{fi}) \\
&= g_f(x_f, x_0)
\end{aligned}
$$

- Thus we can identify the location of the wage equation by identification at infinity.

Step 2: Identification at Infinity

- What about the location?
- Notice that

$$\begin{aligned}
\lim_{\Pr(F_i=1|X_i=x)\to 1} E(W_i|X_i=x, F_i=1) \\
&= g_f(x_f, x_0) + \lim_{\Pr(F_i=1|X_i=x)\to 1} E(\epsilon_{fi}|\epsilon_{hi}-\epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + \lim_{g^*(x)\to\infty} E(\epsilon_{fi}|\epsilon_{hi}-\epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + E(\epsilon_{fi}) \\
&= g_f(x_f, x_0)
\end{aligned}$$

- Thus we can identify the location of the wage equation by identification at infinity.

Step 2: Identification at Infinity

- What about the location?
- Notice that

$$
\begin{aligned}
\lim_{\Pr(F_i=1|X_i=x)\to 1} & E(W_i|X_i = x, F_i = 1) \\
&= g_f(x_f, x_0) + \lim_{\Pr(F_i=1|X_i=x)\to 1} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + \lim_{g^*(x)\to\infty} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + E(\epsilon_{fi}) \\
&= g_f(x_f, x_0)
\end{aligned}
$$

- Thus we can identify the location of the wage equation by identification at infinity.

Step 2: Identification at Infinity

- What about the location?
- Notice that

$$
\begin{aligned}
&\lim_{\Pr(F_i=1|X_i=x)\to 1} E(W_i|X_i = x, F_i = 1) \\
&= g_f(x_f, x_0) + \lim_{\Pr(F_i=1|X_i=x)\to 1} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + \lim_{g^*(x)\to\infty} E(\epsilon_{fi}|\epsilon_{hi} - \epsilon_{fi} < g^*(x)) \\
&= g_f(x_f, x_0) + E(\epsilon_{fi}) \\
&= g_f(x_f, x_0)
\end{aligned}
$$

- Thus we can identify the location of the wage equation by identification at infinity.

Step 2: Identification at Infinity

- Identification at Infinity not only sufficient, but also necessary.
- The model is not identified without identification at infinity.
- To see why suppose that $g^*(x)$ is bounded from above at g^u such that if $\epsilon_{hi} - \epsilon_{fi} > g^u$ then $F_i = 0$.
- Thus the data is completely uninformative about

$$E(\epsilon_{fi} | \epsilon_{hi} - \epsilon_{fi} > g^u)$$

- so the model is not identified.
- Parametric assumptions on the distribution of the error term is an alternative.

Step 2: Identification at Infinity

- Identification at Infinity not only sufficient, but also necessary.
- The model is not identified without identification at infinity.
- To see why suppose that $g^*(x)$ is bounded from above at g^u such that if $\epsilon_{hi} - \epsilon_{fi} > g^u$ then $F_i = 0$.
- Thus the data is completely uninformative about

$$E(\epsilon_{fi} | \epsilon_{hi} - \epsilon_{fi} > g^u)$$

- so the model is not identified.
- Parametric assumptions on the distribution of the error term is an alternative.

Who cares about Location?

- Without our intercept we know something about wage variation within fishing
- However we can not compare the level of fishing to the level of hunting
- If our policy involves moving people from one to the other we need the intercepts

Who cares about Location?

- Without our intercept we know something about wage variation within fishing
- However we can not compare the level of fishing to the level of hunting
- If our policy involves moving people from one to the other we need the intercepts

Who cares about Location?

- Without our intercept we know something about wage variation within fishing
- However we can not compare the level of fishing to the level of hunting
- If our policy involves moving people from one to the other we need the intercepts

Step 3: Identification of g_h

- What will be crucial is the other exclusion restriction (i.e. X_{fi})
- For any (x_h, x_0) we want to find an x_f so that

$$\Pr(F_i = 1 | X_i = (x_f, x_h, x_0)) = 0.5$$

- this means that

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

- but the fact that $\epsilon_{hi} - \epsilon_{fi}$ 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 3: Identification of g_h

- What will be crucial is the other exclusion restriction (i.e. X_{fi})
- For any (x_h, x_0) we want to find an x_f so that

$$\Pr(F_i = 1 | X_i = (x_f, x_h, x_0)) = 0.5$$

- this means that

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

- but the fact that $\epsilon_{hi} - \epsilon_{fi}$ 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 3: Identification of g_h

- What will be crucial is the other exclusion restriction (i.e. X_{fi})
- For any (x_h, x_0) we want to find an x_f so that

$$\Pr(F_i = 1 | X_i = (x_f, x_h, x_0)) = 0.5$$

- this means that

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

- but the fact that $\epsilon_{hi} - \epsilon_{fi}$ 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

- Relatively straight forward given everything else
- To identify the joint distribution of $(\epsilon_{fi}, \epsilon_{hi})$ from the data one can observe $\Pr(F_i = 1, W_i < s | X_i = (x_f, x_h, x_0))$ which is :

$$= \Pr\left(g_f(x_f, x_0) + \epsilon_{fi} < s, g_h(x_h, x_0) + \epsilon_{hi} < g_f(x_f, x_0) + \epsilon_{fi}\right)$$
$$= \Pr\left(\epsilon_{fi} < s - g_f(x_f, x_0), \epsilon_{hi} - \epsilon_{fi} < g_f(x_f, x_0) - g_h(x_h, x_0)\right)$$

- which is the cumulative distribution function of $(\epsilon_{fi}, \epsilon_{hi} - \epsilon_{fi})$ evaluated at the point

$$\left(g_f(x_f, x_0) - g_h(x_h, x_0), s - g_f(x_f, x_0)\right)$$

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

Step 4: Identification of G

- Relatively straight forward given everything else
- To identify the joint distribution of $(\epsilon_{fi}, \epsilon_{hi})$ from the data one can observe $\Pr(F_i = 1, W_i < s | X_i = (x_f, x_h, x_0))$ which is :

$$= \Pr\left(g_f(x_f, x_0) + \epsilon_{fi} < s, g_h(x_h, x_0) + \epsilon_{hi} < g_f(x_f, x_0) + \epsilon_{fi}\right)$$
$$= \Pr\left(\epsilon_{fi} < s - g_f(x_f, x_0), \epsilon_{hi} - \epsilon_{fi} < g_f(x_f, x_0) - g_h(x_h, x_0)\right)$$

- which is the cumulative distribution function of $(\epsilon_{fi}, \epsilon_{hi} - \epsilon_{fi})$ evaluated at the point

$$\left(g_f(x_f, x_0) - g_h(x_h, x_0), s - g_f(x_f, x_0)\right)$$

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