Labor Supply D. Card, Sept 2008

1. Preliminary issues (compensated, Frisch, Marshallian responses)

2. non-participation in a single-agent setting

3. models with multiple solutions – the “old” view

4. Ransom’s model of family labor supply

5. models with multiple solutions – the “new” view

1. Preliminary issues

Consider the canonical consumer choice problem with an interior maximum:

Max $U(x)$ s.t. $p \cdot x = y$

f.o.c. $DU(x) - \lambda p = 0$

$$p \cdot x - y = 0$$

s.o.c. $v' D^2U(x)\, v \leq 0$ for all $v \neq 0$ s.t. $p \cdot v = 0$.

Solution functions:

$$x = x(p, y)$$

$$\lambda = \lambda(p, y)$$

Differentiate the f.o.c., letting $S = D^2U(x)$ denote the matrix of 2nd partials:

$$\begin{array}{cc} S & p \\ p' & 0 \end{array} \quad \begin{array}{cc} \partial x/\partial p' & \partial x/'\partial y \\ -\partial\lambda/\partial p' & -\partial\lambda/\partial y \end{array} \qquad \begin{array}{cc} \lambda I & 0 \\ -x' & 1 \end{array}$$

Using the partitioned inverse formula: $\partial\lambda/\partial y = [\, p'S^{-1} p \,]^{-1} < 0$ if U is concave. For simple static choice problems we wouldn’t put restrictions on the concavity of U (only quasi-concavity). For intertemporal problems, or problems with uncertainty, we do. Note that $\partial\lambda/\partial y$ is related to the degree of risk aversion, and also to the intertemporal substitution elasticity of consumption demand (Robert Hall, *JPE*, 1988)

Further application of partitioned inverse formula yields:

$\partial x/\partial y \;=\; \partial\lambda/\partial y \,\cdot\, S^{-1}\, p$

$\partial x/\partial p' \;=\; \lambda\, S^{-1} - \lambda\, [\partial\lambda/\partial y]^{-1}\, [\, \partial x/\partial y \,\cdot\, \partial x'/\partial y \,] \quad - \quad \partial x/\partial y \cdot x'$

The third term of the second line is what we conventionally call the "income effect". The usual "substitution effect" is represented by the <u>first two terms</u>. Specifically, define the Hicksian compensated demand function $x^c(p,u) \equiv x(p, e(p,u))$, where $e(p,u)$ is the expenditure function. From the Slutsky equation:

(S) $\quad \partial x^c(p, u)/\partial p' \;=\; \partial x(p, y)/\partial p' \;+\; \partial x/\partial y \cdot x'$,

and hence:

$$\partial x^c(p, u)/\partial p' \;=\; \lambda\, S^{-1} - \lambda\, [\partial\lambda/\partial y]^{-1}\, [\, \partial x/\partial y \,\cdot\, \partial x'/\partial y \,]$$

Looking back at the f.o.c, define the **Frisch demands** $x^F(p, \lambda)$ as the solution functions:

(F) $\quad DU(\, x^F(p, \lambda)\,) = \lambda\, p.$

This is the set of demands the consumer would have, facing prices p and a given marginal utility of income λ. In simple intertemporal choice problems $U = \sum_t \beta^t\, v(x_t)$, where x_t is consumption in period t, with discounted price p_t, and the Frisch demands satisfy

$$\beta^t\, v'(\, x_t^F(p, \lambda)\,) \;=\; \lambda\, p_t \;,$$

which is the traditional first order condition for optimal intertemporal consumption with a fixed lifetime budge constraint. Differentiating (F) w.r.t. p′ and using $S=D^2U(x)$,

$$S\;\; \partial x^F(p, \lambda)/\partial p' = \lambda \;,$$

$\Longrightarrow \quad \partial x^F(p, \lambda)/\partial p' \;=\; \lambda\, S^{-1}$, and thus we can write

$$\partial x^c(p, u)/\partial p' \;=\; \partial x^F(p, \lambda)/\partial p' \;-\; \lambda\, [\partial\lambda/\partial y]^{-1}\, [\, \partial x/\partial y \,\cdot\, \partial x'/\partial y \,]$$

For the i,j element we have

(*) $\quad \partial x^c_i\, (p, u)/\partial p_j \;=\; \partial x^F_i\, (p, \lambda)/\partial p_j \;-\; \lambda\, [\partial\lambda/\partial y]^{-1}\;\; \partial x_i/\partial y \,\cdot\, \partial x_j/\partial y$.

Note that if D^2U is diagonal (all second cross-partials=0), as would be true in the case where $U = \sum_t \beta^t v(x_t)$, then S^{-1} is diagonal, so for i≠j the compensated derivatives are proportional to the income effects. Also, (*) implies a nice link between the compensated own elasticity, the Frish elasticity, and a combination of income effects and the elasticity of λ w.r.t. (basically the degree of risk aversion).

<u>Applications to intertemporal labor supply (or, how do we know what we're estimating?)</u>

Consider an application of the previous framework to an intertemporal planning problem, when preferences are additively separable across periods. Let c_t denote consumption in period t, h_t denote hours of work in period t, w_t = wage (measured as of period t). Assuming a fixed real interest rate r, the choice problem is

$$\text{Max} \quad \sum_t \beta^t v(c_t, h_t) \quad \text{s.t.}$$

$$\sum_t (1+r)^{-t} [\ c_t - w_t h_t)\ ] = A_0 + \sum_t (1+r)^{-t}\ y_t \quad \text{where } y_t = \text{nonlabor income.}$$

f.o.c.

$$\beta^t v_c(c_t, h_t) - \lambda (1+r)^{-t} = 0$$

$$\beta^t v_h(c_t, \ell_t) + w_t \lambda (1+r)^{-t} = 0$$

Note we get a "within-period tangency" condition

$$-v_h(c_t, h_t) / v_c(c_t, h_t) = w_t$$

and an intertemporal allocation condition on the m.u. of consumption:

$$v_c(c_t, h_t) = \lambda \ \beta^t (1+r)^{-t} \quad => \quad c_t = w_t h_t - s_t \text{ for some optimal "savings".}$$

Now, consider the same consumer who has no access to credit markets. This consumer's optimum would be characterized by:

$$-v_h(c_t, h_t) / v_c(c_t, h_t) = w_t$$

$$c_t = w_t h_t + y_t .$$

The difference is in what level of income is "brought in" or "taken out" of the period. In the perfect foresight intertemporal case the consumer saves or borrows s_t, where s_t serves to keep the MU of consumption on the "right trajectory". The Frisch supply function incorporates this optimal adjustment. If on the other hand, s_t was adjusted to keep $v(c_t, h_t) = v^0$, we'd be observing a Hicksian labor supply function. Finally, if we compare labor supply choices taking $y_t + s_t$ as given, we are observing a Marshallian (uncompensated) within-period labor supply function.

For more on the relation between risk aversion and the behavioral responses to labor supply, see R. Chetty, "A New Method for Estimating Risk Aversion", *AER*, 96 (Dec. 2006): 1821-1834.

2. Non-participation – single agent setting

For the remainder of this lecture we will focus on static labor supply models. One interpretation of this approach is that we have in mind comparisons between people who face "permanent" differences in their wages (and non-labor income opportunities). Chetty (2006) has a nice 2-stage budgeting representation. Suppose the agent's problem is

$$\max\ U(c_1, c_2, \ldots c_T, h_1, h_2, \ldots h_T)$$

$$\text{s.t.}\quad p_1 c_1 + p_2 c_2 + \ldots + p_T c_T = y + w_1 h_1 + w_2 h_2 + \ldots + w_T h_T$$

$$= y + w(\theta_1 h_1 + \theta_2 h_2 + \ldots + \theta_T h_T)$$

Here, p_t is the present value of the price of consumption in period t (e.g. $p_t = (1+r)^{-t}$) and $w_t = w\,\theta_t$ is the present value of wages in period t (so w is a 'scale factor' that blows up wages at all ages, and θ_t represents the combination of the discount factor and any transitory deviation in wages in period t). The problem can be recast as

$$\max u(c,h) \quad \text{s.t.} \quad c = y + wh$$

where $u(c,h) \equiv \max U(c_1, c_2, \ldots c_T, h_1, h_2, \ldots h_T)$

$$\text{s.t.} \quad c = p_1 c_1 + p_2 c_2 + \ldots + p_T c_T$$

$$h = \theta_1 h_1 + \theta_2 h_2 + \ldots + \theta_T h_T$$

Here c and h are "lifecycle aggregates" (we can rescale them to represent annualized consumption and hours of work). If the main variation of interest is in y and w, and NOT intertemporal variation in θ_t, then it makes sense to think of the analysis as a static model of choice over the lifecycle aggregates. Note (using the envelop property of the value function) that the marginal utility of wealth in the second stage problem is the multiplier on the lifecycle consumption allocation constraint (i.e.

$u_c(c,h) = \lambda = \partial U/\partial c_t / p_t$ for all t .

Ignoring non-participation, the standard way to proceed in static labor supply studies is to adopt a convenient parametric functional form, e.g. my personal favorite:

(1) $h_i = X_i \beta + \gamma \log w_i + \theta y_i + \varepsilon_i$.

In a stunning feat of reverse engineering (integration) Stern (1985) showed there is a utility function $v(c, h; \tau)$ that generates a labor supply function $h = \tau + \gamma \log w + \theta y$. Let $\tau_i = X_i \beta + \varepsilon_i$ be the combination of observed and unobserved heterogeneity. Assuming a distribution for ε_i, we have derived a d.g.p. for the data h_i, conditional on $(X_i, \log w_i, y_i)$. We probably don't want to assume that $\log w_i$ is orthogonal to ε_i , so we might have to augment the model with

(2) $\log w_i = Z_i \pi + \eta_i$.

Combining this with (1), we have

(3) $$h_i = X_i \beta + Z_i \gamma\pi + \theta y_i + \zeta_i, \quad \text{where} \quad \zeta_i = \varepsilon_i + \gamma \eta_i .$$

Assuming normal errors gives a likelihood for (h_i, log w_i) conditional on (X_i, Z_i, y_i).

If there is a significant fraction of non-workers in the sample of interest, the "cheap fix" is to impute a wage (using a selection-corrected regression model fit to the observed wage data to simulate wages for those who don't work, possibly with multiple draws for the error term). and then convert (1) into a Tobit model. A more attractive idea (Heckman, 1974) is to define i's reservation wage r_i such that:

$$0 = X_i \beta + \gamma \log r_i + \theta y_i + \varepsilon_i , \quad \text{i.e.} \ \log r_i = -X_i \beta/\gamma - \theta/\gamma \, y_i - 1/\gamma \, \varepsilon_i .$$

When log $w_i \leq$ log r_i the person doesn't work (and we observe h_i=0). When log w_i > log r_i we observe wages and positive hours. The likelihood consists of two parts:

- a probability for nonworking = $p(\zeta_i \leq -X_i \beta - Z_i \gamma\pi - \theta y_i)$

- a density for ($\zeta_i = h_i - X_i \beta - Z_i \gamma\pi - \theta y_i$; $\eta_i = \log w_i - Z_i \pi$)

  conditional on $\zeta_i > -X_i \beta - Z_i \gamma\pi - \theta y_i$.

It's a good exercise to work out the correct likelihood.

3. Models with multiple solutions – the "old" view

In the next section we will be discussing Ransom's family labor supply model. A key preliminary issue is the possibility of multiple equilibria that arises in multi-variate discrete choice settings. The simplest model that illustrates the issues is a bivariate probit. Define y_1^* and y_2^* as latent random variables and y_1 and y_2 as the associated indicator functions: $y_1 = 1(y_1^* \geq 0)$, $y_2 = 1(y_2^* \geq 0)$. Think of y_1 and y_2 as dummies for whether family members 1 and 2 work. The simplest possible model is

$$y_1^* = \beta_1 y_2 + \gamma_1 X_1 + v_1 ; \qquad y_2^* = \beta_2 y_1 + \gamma_2 X_2 + v_2 .$$

$P(1,1) = P(v_1 > -\gamma_1 X_1 - \beta_1 , v_2 > -\gamma_2 X_2 - \beta_2)$

$P(0,0) = P(v_1 \leq -\gamma_1 X_1 , v_2 \leq -\gamma_2 X_2)$

$P(1,0) = P(v_1 > -\gamma_1 X_1, v_2 \leq -\gamma_2 X_2 - \beta_2)$

$P(0,1) = P(v_1 \leq -\gamma_1 X_1 - \beta_1 , v_2 \leq -\gamma_2 X_2)$

Draw (v_1 , v_2) space, partitioning the area into the regions that map to the 4 outcomes:

There is a problem: the dark area maps into multiple outcomes so there is not a unique mapping from the v's to the y's. In the old (pre-2000) literature, this was called an "incoherency" problem. We are familiar with this problem from simple 2x2 games, where under some conditions there are 2 equilibria: 1 and 2 both work, or 1 and 2 both do not work (if the β's are positive). (In other settings where the β's are negative, there are also 2 equilibria: 1 works and 2 doesn't, or vice versa). Note that the dark area collapses if either $\beta_1=0$ or $\beta_2=0$. Until recently, it was taken for granted that in a bivariate probit, you had to assume 1-way feedback only (ie, $\beta_1\beta_2=0$). This is equivalent to assuming one person is the Stackelberg leader – not so attractive in a family labor supply setting. We return to the "modern" approach later in the lecture.

<u>A related case</u>

Consider a mixed 2-equation model of the form:

$$y_1 = \beta_1 y_2 + \gamma_1 X_1 + v_1$$

$$y_2^* = \beta_2 y_1 + \gamma_2 X_2 + v_2 \ ,$$

$$y_2 = \max(y_2^*, 0) \qquad \text{(a Tobit-model)}$$

Note

$$\begin{aligned} y_2^* &= \beta_2(\beta_1 y_2 + \gamma_1 X_1 + v_1) + \gamma_2 X_2 + v_2 \\ &= \beta_1\beta_2 y_2 + z, \qquad \text{where} \quad z = \gamma_2 X_2 + \beta_2 \gamma_1 X_1 + \beta_2 v_1 + v_2 \ . \end{aligned}$$

Consider the mapping from z to y_2^* :

Case 1: $\beta_1\beta_2>0$. Case 2: $\beta_1\beta_2<0$.

If $\beta_1\beta_2<0$ there are multiple solutions for z<0 and no solution for z>0.

4. Ransom's model of family labor supply

Ransom considers a "unitary" model of family decision-making:

$$\max\ U(T_1-h_1, T_2-h_2, x) \quad \text{s.t.} \quad x = w_1h_1 + w_2h_2 + y\ .$$

Define $m_1(h_1, h_2) = -U_1(\quad) + w_1U_3$, $m_2(h_1, h_2) = -U_2(\quad) + w_2U_3$.

Two cases considered:

both work: $m_1(h_1, h_2) = 0$, $m_2(h_1, h_2) = 0$

1 works, 2 doesn't: $m_1(h_1, 0) = 0$, $m_2(h_1, 0) < 0$.

Ransom assumes U is quadratic in the 3 arguments => U_1, U_2, U_3 are linear in the arguments. Thus m_1 include a constant, linear terms in h_1, h_2, and x, and interactions of w_1 with h_1, h_2, and x. In addition, he assumes the constant contains an additive normal error component ε_1. Thus

$$m_1 = S_1(h_1 , h_2 ;\ w_1, w_2, y) + \varepsilon_1$$

for some function S_1 that is <u>linear</u> in h_1 and h_2 (though the coefficients depend on w_1,w_2 and y – see his equation (5a)). Likewise

$$m_2 = S_2(h_1 , h_2 ;\ w_1, w_2, y) + \varepsilon_2 .$$

Write $S_1(h_1 , h_2) = a_{10} + a_{11}h_1 + a_{12}h_2$, and $S_2(h_1 , h_2) = a_{20} + a_{21}h_1 + a_{22}h_2$. In the both-work regime we have $m_1 = 0$ and $m_2 = 0$, which can be re-cast as

$$h_1 = - a_{10}/a_{11} – a_{12}/a_{11}\, h_2 + - \varepsilon_1/ a_{11} ,$$

$$h_2 = - a_{20}/a_{22} – a_{21}/a_{22}\, h_1 + - \varepsilon_2/ a_{22} ,$$

whereas in the case that 1 works and 2 doesn't we have

$$h_1 = - a_{10}/a_{11} – a_{12}/a_{11}\, h_2 + - \varepsilon_1/ a_{11} ,$$

$$h_2^* = - a_{20}/a_{22} – a_{21}/a_{22}\, h_1 + - \varepsilon_2/ a_{22} < 0.$$

This has the same structure as we discussed above, so for a unique solution we need that $1 - (a_{12}/a_{11})(a_{21}/a_{22}) > 0$. Ransom shows that this is satisfied if U is concave. In his related paper in the *J. Econometrics* he presents a generic proof, based on a result due to Amemiya. There is no "coherency" problem in a unitary decision maker model with well behaved preferences.

Ransom does not try to econometrically address the problem that when 2 doesn't work, w_2 is unobserved – instead he uses the "cheap fix" of imputing a wage for non-workers. He also ignores any measurement error or endogeneity in observed wages. In principle it is possible to expand his model to include 2 additional equations for wages, with a stochastic component that incorporates measurement error. A serious issue, however, is that the first order conditions from his specification are NOT very convenient to work with when observed wages include an error component.

Conditioning on wages and nonlabor income, in the “both work” regime we know $m_1=S_1+\varepsilon_1=0$ and $m_2=S_2+\varepsilon_2=0$. If $f(\varepsilon_1, \varepsilon_2; \Sigma)$ is the bivariate normal density (conditional on cov. Σ) the likelihood is $f(-S_1, -S_2; \Sigma)\cdot|J|$ where $J_{ij} = \partial S_i/\partial h_j = a_{ij}$ (J = the Jacobian). In the other regime we have $\varepsilon_1 = -S_1(h_1,0)$ and $\varepsilon_2 < -S_2(h_1,0)$, so the likelihood is

$$\varphi(-S_1; \sigma_1) \cdot | a_{11} | \cdot \Phi[(-S_2 - \rho\sigma_2/\sigma_1 S_1) / \sigma_2(1-\rho^2)^{1/2}] .$$

Extensions

Consider a two-person household where each person has a labor supply equation based on (1), and takes total non-labor income and the other’s earnings as given (non-cooperative household decision making):

$$h_{1i} = \gamma_1 \log w_{1i} + \theta_1 [y_i + w_{2i} h_{2i}] + \varepsilon_{1i} .$$

$$h_{2i}^* = \gamma_2 \log w_{2i} + \theta_2 [y_i + w_{2i} h_{2i}] + \varepsilon_{2i} \quad \text{with } h_{2i} = \max [0, h_{2i}^*]$$

We can rewrite the first two as

$$h_{1i} = \gamma_1 \log w_{1i} + \theta_1 y_i + \theta_1 w_{2i} h_{2i} + \varepsilon_{1i} .$$

$$h_{2i}^* = \gamma_2 \log w_{2i} + \theta_2 y_i + \theta_2 w_{2i} h_{2i} + \varepsilon_{2i}$$

From the general discussion above, this system will be “coherent” if $1- \theta_1 \theta_2 w_{2i} w_{1i} > 0$. Note that $\partial h_1/\partial y = \theta_1$, so $w_1\partial h_1/\partial y = w_1 \theta_1$ and likewise, $w_2\partial h_2/\partial y = w_2 \theta_2$. So the coherency condition is $w_1\partial h_1/\partial y \times w_2\partial h_2/\partial y < 1$. $w_1\partial h_1/\partial y$ is amount that #1’s earnings will fall if the family gets $1 of non-labor income. For men this is thought to be very small (0.05 - 0.15). For women it may be bigger but is probably less than 1. So the condition is probably OK. There could be a problem if there are minimum hours constraints (or fixed cost of working) so 2 has to work at least 20 hours/week, or not at all.

The Modern Approach

A number of recent papers have considered estimation of models with non-unique equilibria – Imbens and Wooldridge (Chapter 9) provide an overview. The canonical model in the new literature is an entry game with 2 players. The underlying equations are expressions for the profits of firms 1 and 2 if the other is present or absent:

$$\pi_1 = a_1 + b_1 d_2 + \varepsilon_1$$

$$\pi_2 = a_2 + b_2 d_1 + \varepsilon_2$$

$$d_j = 1(\pi_j > 0)$$

Prob. of (0,0) = p(0,0) = $P(\varepsilon_1 \leq -a_1 ,\ \varepsilon_2 \leq -a_2)$

Prob. of (1,0) = p(1,0) = $P(\varepsilon_1 > -a_1 ,\ \varepsilon_2 \leq -a_2 - b_2)$

Prob. of (0,1) = p(0,1) = $P(\varepsilon_1 \leq -a_1 - b_1 ,\ \varepsilon_2 > -a_2)$

Prob. of (1,1) = p(1,1) = $P(\varepsilon_1 > -a_1 - b_1 ,\ \varepsilon_2 > -a_2 - b_2)$

Assuming $b_1<0$ and $b_2<0$ the mapping from $(\varepsilon_1 ,\varepsilon_2)$ space to (d_1 , d_2) space is:

These 4 areas add up to more than 1. The problem is the dark shaded area, which contributes to both p(0,1) and p(1,0). When d_1 and d_2 are negative, we can have a situation where either 1 enters and 2 does not, or 2 enters, and 1 does not. In the modern literature multiple equilibria is treated as a problem of "partial identification".

A very closely related application to labor economics is to retirement of older couples (Nicole Maestas, Rand Working Paper). In this case we might have d_1 and d_2 both positive (assuming spouses want to spend time together), which gives a slightly different picture.

To proceed, note that we can get expressions for p(0,0) and p(1,1). We can only get expressions for the <u>bounds</u> on p(0,1) and p(1,0), depending on how we allocate the dark area:

$$p(0,0) = P(\varepsilon_1 \le -a_1\ ,\ \ \varepsilon_2 \le -a_2\ )$$

$$p(1,1) = P(\varepsilon_1 > -a_1 - b_1\ ,\ \ \ \varepsilon_2 > -a_2 - b_2\ )$$

$$P(\varepsilon_1 > -a_1\ ,\ \ \varepsilon_2 \le -a_2 - b_2) - D \ \le\ p(1,0)\ \le\ P(\varepsilon_1 > -a_1\ ,\ \ \varepsilon_2 \le -a_2 - b_2)$$

$$P(\varepsilon_1 \le -a_1 - b_1\ ,\ \ \ \varepsilon_2 > -a_2) - D \ \le\ p(0,1) \le\ P(\varepsilon_1 \le -a_1 - b_1\ ,\ \ \ \varepsilon_2 > -a_2)$$

where

$$D = P(\ -a_1 < \varepsilon_1 \le -a_1 - b_1,\ \ \ -a_2 < \varepsilon_2 \le -a_2 - b_2) \quad \text{<the dark area in the graph>}$$

In general notation:

$$p(0,0) = \pi_{00}$$

$$p(1,1) = \pi_{11}$$

$$L_{10} \le p(1,0)\ \le\ U_{10}$$

$$L_{01} \le p(0,1)\ \le\ U_{01}$$

Consider the moment conditions formed from the data for each market (in which we observe the outcomes (d_1, d_2):

$$m_1 = \pi_{11} - d_1 d_2$$

$$m_2 = \pi_{00} - (1-d_1)(1-d_2)$$

$$m_3 = U_{10} - d_1 (1-d_2)$$

$$m_4 = d_1 (1-d_2) - L_{10}$$

$$m_5 = U_{01} - d_2 (1-d_1)$$

$$m_6 = d_2 (1-d_1) - L_{01}$$

We have the restriction from the model that $E[m(d_1, d_2) \mid \text{parms}] \geq 0$. So we want to find a set of parameter estimates such that the observed mean vector of moments satisfies the inequality restriction. This is a potentially solvable estimation problem (see WNiE, chapter 9).