Economics 250a
Lecture 4, Labor Supply Continued
Outline

1. Simple labor supply models with corrections for non-participation (PS #2)
2. Other approaches to kinked budget sets - discretization
3. Dynamic labor supply - introduction and overview

Readings

Thomas A. Mroz. "The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions." Econometrica, 55 (4) (July, 1987), pp. 765-799.

Arthur van Soest. "Structural Models of Family Labor Supply" A Discrete Choice Approach." Journal of Human Resources 30(1), Winter 1995.

David Card. "Intertemporal Labor Supply: An Assessment". In Christopher Sims, editor, Advances in Econometrics, Sixth World Congresss (volume 2). Cambridge University Press, 1994 (nber version available on my web site)

1. Labor Supply and Corrections for Non-Participation

We briefly discussed corner solutions in Lecture 2. We return to this issue using the setup in Mroz's study (which you will be replicating in the next problem set). We are trying to model (say) married women's labor supply, and we are prepared to take their partners' earnings as exogenous. The setup includes a Stern labor supply function and a standard human capital (Mincer) model for log wages:

$$\begin{aligned} h_i^* &= a_0 + a_1 \log w_i + a_2 y_i + a_3 Z_i + u_{1i} \\ \log w_i &= c Z_i + u_{2i} \end{aligned} \quad (1)$$

where Z_i is a vector of background variables (including characteristics of person i, the wife in a 2-person family, as well as the characteristics of her partner and her family). The parameters of interest are a_1 and a_2. Person i is observed working hours h_i^* (or possibly $h_i^* + \epsilon_i$, where ϵ_i, is a measurement error) if

$$h_i^* \geq h_i^{\min}$$

A pure reservation wage model assumes $h_i^{\min} = 0$, which, with the Stern labor supply function, gives rise to a Tobit model. A more general fixed cost model assumes

$$h_i^{\min} = b_0 + b_1 Z_i + u_{3i}$$

which allows the lowest-hour job to depend on person-specific characteristics and a stochastic term. Working occurs if

$$a_0 + a_1 \log w_i + a_2 y_i + a_3 Z_i + u_{1i} \geq b_0 + b_1 Z_i + u_{3i}$$

$$\begin{aligned} \text{or } \xi_i &= u_{1i} + a_1 u_{2i} - u_{3i} \\ &> -(a_0 - b_0) - a_2 y_i - (a_3 + a_1 c - b_1) Z_i. \end{aligned} \quad (2)$$

If we assume that $(u_{1i}, u_{2i}, u_{3i}) \sim N(0, \Sigma)$ we get a very convenient setup. Equation (2) yields a simple probit model for participation with latent normal error ξ_i. With 3 underlying error terms, this error is arbitrarily correlated with the error in the labor supply equation and with the error in the wage equation. (On the other hand, if $u_{3i} = 0$ we have only 2 independent errors). We can estimate both the wage model and the hours equation by 2-step Heckit, or IV-Heckit.

The Heckit approach is a particular case of the so-called control function (CF) approach to censored data. The CF idea is to recognize that in the censored sample, the distribution of the error term is different for people with higher and lower probabilities of being in the sample. Heckman's original derivation assumed normality, which we will start with.

A key fact is that if $(z_1, z_2) \sim N(0, \Sigma)$, the distribution of z_1 *conditional on the event that* $z_2 > c$, has a convenient functional form. Specifically, for joint normals we can always write:

$$z_1 = r_{1,2} z_2 + \nu$$

where $r_{1,2} \equiv cov[z_1, z_2]/var[z_2] = \rho\sigma_1/\sigma_2$ is the regression coefficient of z_1 on z_2 and ν is a normal variate with mean 0 that is independent of z_2. Thus

$$\begin{aligned} E[z_1|z_2 &> c] = r_{1,2} E[z_2|z_2 > c] \\ &= r_{1,2}\sigma_2 E[\frac{z_2}{\sigma_2}|\frac{z_2}{\sigma_2} > \frac{c}{\sigma_2}] \\ &= \rho\sigma_1 \frac{\phi\left(\frac{c}{\sigma_2}\right)}{1 - \Phi\left(\frac{c}{\sigma_2}\right)} = \rho\sigma_1 \frac{\phi\left(\frac{-c}{\sigma_2}\right)}{\Phi\left(\frac{-c}{\sigma_2}\right)} = \rho\sigma_1 \frac{\phi(\Phi^{-1}(p(\frac{-c}{\sigma_2}))}{p(\frac{-c}{\sigma_2})} \equiv \rho\sigma_1 \lambda(p(\frac{-c}{\sigma_2})) \end{aligned}$$

where $p(\frac{-c}{\sigma_2}) = \Phi\left(\frac{-c}{\sigma_2}\right)$ and we have used the result that for a standard normal variate ω, $E(\omega|\omega > a) = \phi(a)/[1 - \Phi(a)]$. Look at this carefully: it says that

$$E[z_1|z_2 > c] = k \cdot \lambda(p(\frac{-c}{\sigma_2}))$$

where p is the probability of selection for the observation in question, $k = \rho\sigma_1$ is a constant that has the same sign as $\rho = correl(z_1, z_2)$ and $\lambda(p) = \frac{\phi(\Phi^{-1}(p))}{p}$ is a known, decreasing function of p, with $\lambda(p) > 0$, $\lambda(p) \to \infty$ as $p \to 0$, and $\lambda(p) \to 0$ as $p \to 1$ (see the Figure at the end of the lecture). In our application $z_2 > c$ is the condition for working (with probability p) and we will have 2 $z_1's$: hours and wages, neither of which is observed for non-workers.

In a 2-stage Heckit you first estimate the parameters of a reduced form probit model that determines whether the outcome (or outcomes) of interest are observed. You then take the parameters of this model, use them to predict p for each observation in the observed data set, and form $\lambda(p)$. In the observed sample

$$\begin{aligned} E[h_i|w_i, y_i, Z_i, working] &= a_0 + a_1 \log w_i + a_2 y_i + a_3 Z_i \\ &\quad + E[u_{1i}|\xi_i > -(a_0 - b_0) - a_2 y_i - (a_3 + a_1 c - b_1)Z_i] \\ \Rightarrow h_i &= a_0 + a_1 \log w_i + a_2 y_i + a_3 Z_i + k_1 \lambda(p_i) + \nu_{1i} \end{aligned}$$

and

$$\begin{aligned} E[log\, w_i|y_i, Z_i, working] &= cZ_i + E[u_{2i}|\xi_i > -(a_0 - b_0) - a_2 y_i - (a_3 + a_1 c - b_1)Z_i] \\ \Rightarrow log\, w_i &= cZ_i + k_2 \lambda(p_i) + \nu_{2i} \end{aligned}$$

where (v_{i1}, v_{2i}) are "good" errors (orthogonal to the included covariates). If we are worried that $\log w_i$ or y_i are endogenous (or measured with a lot of error) then we have to apply IV-Heckit. Step 1 = probit for participation; step 2 = form $\lambda(p)$; step 3: estimate first stage model for wages; step 4: estimate the labor supply model including $\lambda(p)$ and instrumented values for wages. (Note: you should normally NOT do the 4th step yourself - use the 2SLS command). Another way to proceed would be to substitute $\log w_i = cZ_i + u_{2i}$ into the hours equation, yielding

$$h_i^* = a_0 + (a_1 c + a_3)Z_i + a_2 y_i + u_{1i} + a_1 u_{2i}]$$

which is only observed for participants. So we get a selection-corrected reduced form hours model

$$h_i = a_0 + (a_1 c + a_3)Z_i + a_2 y_i + k_3 \lambda(p) + \nu_{3i}$$

We can estimate the parameters c using a selection-corrected wage model, and the parameters $(a_1 c + a_3)$ and a_2 using a selection-corrected hours model. We can then do method of moments to infer a_1.

An even more complex situation arises when we want to instrument y_i (either because of measurement error, or a concern that some of the variables that drive y_i also affect hours choices. This would lead to IV-Heckit with 2 endogenous variables

Questions for discussion:

1) We clearly need instruments for the wage (given division bias). What are plausible candidates?

2) We also need variables that move $\lambda(p)$ around – so variables that determine the probability of participation but do not directly affect wages and hours. Thoughts?

Non-normal alternatives:

In general, for any pair of random variables (u, v) with some continuous density $f(u, v)$ and marginal cdf $F_v(v)$:

$$\begin{aligned} E[u|v > c] &= \frac{\int_{-\infty}^{\infty}\int_{c}^{\infty} uf(u,v)dvdu}{\int_{-\infty}^{\infty}\int_{c}^{\infty} f(u,v)dvdu} \\ &= \frac{1}{P(v>c)}\int_{-\infty}^{\infty} uf(u)\left(\int_{c}^{\infty} f(v|u)dv\right)du \\ &= \frac{1}{P(v>c)}\int_{-\infty}^{\infty} u\,P(v>c|u)du \\ &= G(c) \text{ for some function } G \\ &= G(F_v^{-1}(p)) = H(p) \end{aligned}$$

where $p = Prob(v > c)$. So any two realizations of u from a censored sample where the censoring probabilities are the same have the same expected value. In general then, you can think of controlling for some flexible function of p in the censored regression model. If $y = x\beta + \epsilon$ then

$$E[y|x, v > c] = x\beta + H(p).$$

for some $H(p) = E[\epsilon|censoring\,event\,with\,prob.\,p]$. This idea can be generalized to more complex models for the "conditioning event".

2. Other approaches to kinked budget sets - discretization

Kinked budget sets and non-participation can be handled in a unified way by thinking of the budget set as consisting of discrete points. For example, we could discretize hours choices into 1 hour bins, or 5 hour bins, or even 3 choices: h=0, h=20, h=40 (non-work, part-time, and full-time). This is arguably the state of the art for static labor supply.

The simplest implementation is to pick a parametric utility function $u(x, h; \theta, z, \epsilon)$, where θ represent parameters, z represent observable characteristics, and ϵ represents unobserved heterogeneity. Then assume some simple way of classifying observed choices into the discrete options $\{h_0 = 0, h_1, h_2, ...h_J\}$, and make a set of assumptions that allow you to associate an earnings amount $\{e_1, e_2, ...e_J\}$ with each choice. If the worker faces a fixed wage w, for example, then you could assume $e_j = wh_j$. More generally, the earnings amounts e_j can incorporate taxes and transfers, a premium or discount for part-time work, a fixed cost of working, etc). A worker with heterogeneity (z, ϵ) evaluates $u(y + e_j\ h_j; \theta, z, \epsilon)$ for each option and chooses the highest utility. In practice there are a lot of delicate issues. The most serious problem is that now we have to come up with values for (h_j, e_j) for J choices, at most 1 of which is actually observed.

A second issue is how to incorporate random taste variation in the ϵ vector. A starting point for this line of work is usually a multinomial logit model

$$u(y+e_j,h_j;\theta,z,\epsilon)=v(y+e_j,h_j;\theta,z)+\epsilon_j$$

where ϵ_j are extreme value type 1 errors, and v is some convenient functional form. If the wage w was known for each person this would lead to a simple MNL choice model

$$\Pr(choice\ j\mid w,z)=\frac{\exp(v(y+e_j(w),h_j;\theta,z))}{\sum_k v(y+e_k(w),h_k;\theta,z)}$$

A very simple example of this approach is as follows. First, assume that each person faces a parameteric gross market wage w, and that the hours choices are some fixed set, e.g., $\{0,5,10,15....\}$. Knowing the tax schedule, find $\{e_j(w)\}$, the net earnings at each h_j for a person with gross wage w. Since the wage is not observed for non-workers (and may be measured with error), assume a DGP for wages that implies a density $f(w|z)$, and treat w as an endogenous outcome. Then the likelihood for the observed workers is of the form:

$$\Pr(choice\ j,w\mid z)=\Pr(choice\ j\mid w,z)\times f(w\mid z).$$

The likelihood for non-workers is

$$\Pr(nonwork\mid z)=\int\Pr(choice\ j=0\mid w,z)\times f(w\mid z)dw$$

An example of this approach is van Soest (1995). A more recent variant is:

Arthur van Soest, Marcel Das and Xiaodong Gong, "A Structural Labour Supply Model with Flexible Preferences". Journal of Econometrics 107 (2002), pp. 345-374.

3. Intertemporal Labor Supply

Intertemporal labor supply (ils) models are used to analyze responses to variation over time in wages or income. There are three main dimensions of intertemporal labor supply:

(1) the pure lifecycle dimension. Wages have a hump-shaped pattern over the lifecycle. Do people respond to this variation?

(2) the macro dimension. Hours of work vary over the business cycle. Can we interpret this as an endogenous response to wages?

(3) the idiosyncratic dimension. A person may have temporarily higher wages in some period. How does he/she respond?

Denote periods (e.g., years) by t. Consumption in period t is c_t, hours of work are h_t, the wage is w_t. Individuals have an additively separable intertemporal utility function

$$u(c_t,h_t;a_t)+\beta u(c_{t+1},h_{t+1};a_{t+1})+\beta^2u(c_{t+2},h_{t+2};a_{t+2})+...$$

where a_t is a preference shock in period t. The intertemporal budget contraint is

$$A_{t+1} = (1 + r_t)(A_t + y_t + w_t h_t - c_t)$$

where A_t represent real assets in period t, and r_t is the real interest rate from t to $t+1$. The Bellman equation is

$$V_t(A_t) = \max_{c_t, h_t} u(c_t, h_t; a_t) + \beta E_t[V_{t+1}((1 + r_t)(A_t + y_t + w_t h_t - c_t))]$$

The f.o.c. are

$$\begin{aligned} u_c(c_t, h_t; a_t) - \beta(1 + r_t)E_t[V'_{t+1}(.)] &= 0 \\ u_h(c_t, h_t; a_t) + \beta w_t(1 + r_t)E_t[V'_{t+1}(.)] &= 0 \end{aligned}$$

Define

$$\lambda_t \equiv V'_t(A_t) = \beta(1 + r_t)E_t[V'_{t+1}(.)] = \beta(1 + r_t)E_t[\lambda_{t+1}],$$

where the second equality is from the envelope theorem and the third just uses the definition of λ_{t+1}. We can rewrite the f.o.c. as

$$\begin{aligned} u_c(c_t, h_t; a_t) &= \lambda_t \\ u_h(c_t, h_t; a_t) &= -w_t \lambda_t \end{aligned}$$

implying a tangency condition (or within-period efficiency condition)

$$\frac{-u_h(c_t, h_t; a_t)}{u_c(c_t, h_t; a_t)} = w_t$$

Note that the budget constraint within the period is

$$c_t = w_t h_t + y_t - [\frac{A_{t+1}}{1 + r_t} - A_t]$$

So the within-period choices are the same as a person with the same utility function would make in a static problem with non-labor income

$$\begin{aligned} & y_t - S_t, \quad \text{where} \\ S_t &= \frac{A_{t+1}}{1 + r_t} - A_t \end{aligned}$$

The within-period f.o.c. define Frisch demands

$$\begin{aligned} c_t &= c^F(w_t, \lambda_t, a_t) \\ h_t &= h^F(w_t, \lambda_t, a_t) \end{aligned}$$

These are also called the lambda-constant or intertemporal consumption and hours choices. There are 4 elasticities

$$\frac{\partial \log h^F}{\partial \log w}, \frac{\partial \log h^F}{\partial \log \lambda}, \frac{\partial \log c^F}{\partial \log w}, \frac{\partial \log c^F}{\partial \log \lambda}.$$

The response of hours to wages holding constant λ is called the intertemporal substitution elasticity. Recall there are also compensated and uncompensated responses

$$\frac{\partial \log h^c(w,u)}{\partial \log w}, \frac{\partial \log h(w,y)}{\partial \log w}$$

We showed in the first lecture that $\frac{\partial \log h^F}{\partial \log w} \geq \frac{\partial \log h^c(w,u)}{\partial \log w} \geq 0$. Looking at the f.o.c. we get

$$\begin{pmatrix} dc \\ dh \end{pmatrix} = \begin{bmatrix} U_{cc} & U_{ch} \\ U_{hc} & U_{hh} \end{bmatrix}^{-1} \begin{bmatrix} 1 & 0 \\ -w & -\lambda \end{bmatrix}$$

So

$$\frac{\partial h^F}{\partial w} = \frac{1}{U_{cc}U_{hh} - U_{ch}^2} \begin{vmatrix} U_{cc} & 0 \\ U_{hc} & -\lambda \end{vmatrix} = \frac{-\lambda U_{cc}}{\Delta}$$

The other derivatives are

$$\begin{aligned} \frac{\partial h^F}{\partial \lambda} &= \frac{-wU_{cc} - U_{hc}}{\Delta} \\ \frac{\partial c^F}{\partial \lambda} &= \frac{wU_{ch} + U_{hh}}{\Delta} \\ \frac{\partial c^F}{\partial w} &= \frac{\lambda U_{ch}}{\Delta} \end{aligned}$$

Using these relations we can show that

$$\frac{\partial \log h^F}{\partial \log w} = \frac{\partial \log h^F}{\partial \log \lambda} + \frac{c}{wh} \frac{\partial \log c^F}{\partial \log w}$$

So in the separable case ($U_{ch} = 0$), we get that the $\frac{\partial \log h^F}{\partial \log w} = \frac{\partial \log h^F}{\partial \log \lambda}$. More generally if $c \approx wh$ this shows that hours have to vary more than consumption in response to expected wage variation. So an agent should increase savings when the wage is high.

EXERCISE: Consider an agent with a given per-period utility function $u(c,h)$ who solves a 1-period optimization:

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

Using Cramer's rule, find $\frac{\partial h}{\partial w}, \frac{\partial h}{\partial y}$. Show that

$$\frac{\partial h^c}{\partial w} = \frac{-\lambda}{w^2 U_{cc} + 2wU_{ch} + U_{hh}}$$

and that

$$\frac{1}{\frac{\partial h^c}{\partial w}} - \frac{1}{\frac{\partial h^f}{\partial w}} \geq 0.$$

What is the intertemporal substitution elasticity? Start with the case where $U(c,h) = \phi(c) - \psi(h)$. Then $u_c = \lambda_t$ means that $c_t = c^* = const.$ (holding constant a_t). This was Friedman's assumption in the PIH model. In this case, as wages vary along an expected profile (so $\lambda_t = E_{t-1}[\lambda_t]$) the agent finds an hours choice such that

$$\frac{-u_h(c^*, h_t)}{u_c(c^*, h_t)} = w_t$$

This is illustrated in Figure 4.2 at the end.

Note that utility $U(c,h) = \phi(c^*) - \psi(h_t)$ is *lower* in high wage-periods, but the person also gains savings,which have marginal value $\lambda_t = \phi'(c^*)$: so flow utility is

$$\phi(c^*) - \psi(h_t) + \lambda_t(w_t h_t - c^*).$$

In the non-separable case we have a locus of points (c,h) s.t. $U_c(c,h) = \lambda$. As wages vary along an expected trajectory the agent finds a point on this locus with $\frac{-u_h(c,h)}{u_c(c,h)} = w$. Note that (ignoring preference changes over time):

$$U_c(c^F(w_t,\lambda), h^F(w_t,\lambda)) = \lambda$$

implying that

$$\frac{\partial c^F(w_t,\lambda)}{\partial w} = -\frac{U_{ch}}{U_{cc}} \frac{\partial h^F(w_t,\lambda)}{\partial w}$$

and since $\frac{\partial h^F(w_t,\lambda)}{\partial w} > 0$, and $U_{cc} < 0$, the agent will consume more or less in high-wage periods depending on the sign of U_{ch}. It is commonly asserted that $U_{ch} > 0$. In this case the locus is negatively sloped, as shown in Figure 4.3.

Dealing with Uncertainty

What happens if there are unexpected changes? Let's log-linearize the f.o.c.:

$$\log h_t = a_t^{'} + \eta \log w_t + \delta \log \lambda_t$$

And in the previous period

$$\log h_{t-1} = a_{t-1}^{'} + \eta \log w_{t-1} + \delta \log \lambda_{t-1}$$

Differencing we get

$$\Delta \log h_t = \log h_t - \log h_{t-1} = \Delta a_t^{'} + \eta \Delta \log w_t + \delta(\log \lambda_t - \log \lambda_{t-1})$$

But from the f.o.c recall

$$\lambda_{t-1} = \beta(1+r) E_{t-1}[\lambda_t]$$

so

$$\log \lambda_{t-1} = \log[\beta(1+r)] + \log E_{t-1}[\lambda_t]$$

Now let

$$\log E_{t-1}[\lambda_t] = E_{t-1} \log[\lambda_t] + \phi_t \quad \text{(see note ⧫)}$$

and define the innovation in the log marginal utility of income as ξ_t where:

$$\log \lambda_t = E_{t-1} \log[\lambda_t] + \xi_t$$

Combining all these terms we get

$$\begin{aligned} \log \lambda_{t-1} &= \log[\beta(1+r)] + \log E_{t-1}[\lambda_t] \\ &= \log[\beta(1+r)] + E_{t-1} \log[\lambda_t] + \phi_t \\ &= \log[\beta(1+r)] + \log \lambda_t - \xi_t + \phi_t \end{aligned}$$

or

$$\log \lambda_t = \log \lambda_{t-1} - \log[\beta(1+r)] - \phi_t + \xi_t$$

So

$$\Delta \log h_t == \Delta a_t^{'} + \eta \Delta \log w_t + \delta \xi_t - \delta(\phi_t + \log[\beta(1+r)]$$

This says that the change in hours depends on:

the change in tastes $\Delta a_t^{'}$
the change in wages (with an elasticity η)
the innovation in the log MU of income (with an elasticity $\delta \geq 0$)
two terms which are constants

An *expected* change in wages leaves λ_t unchanged and leads to a change in hours $\eta \Delta \log w$. An *unexpected* change, however, has wealth effects: an unexpected rise in w_t will lower $\log \lambda_t$, especially to the extent it is expected to persist. Empirically this means that it is very important to know whether wage innovations are mainly transitory or mainly persistent. The supply response to a transitory wage change is necessarily bigger (unless leisure is an inferior good).

♦Aside

what is the relationship of $E[\log x]$ and $\log E[x]$? One easy case is when x is log-normally distributed: $\log x \sim N(\mu, \sigma^2)$. (Wages are approximately log-normal so this is often a useful case for labor economics). In the log-normal case:

$$E[x] = \exp(\mu + \frac{1}{2}\sigma^2)$$

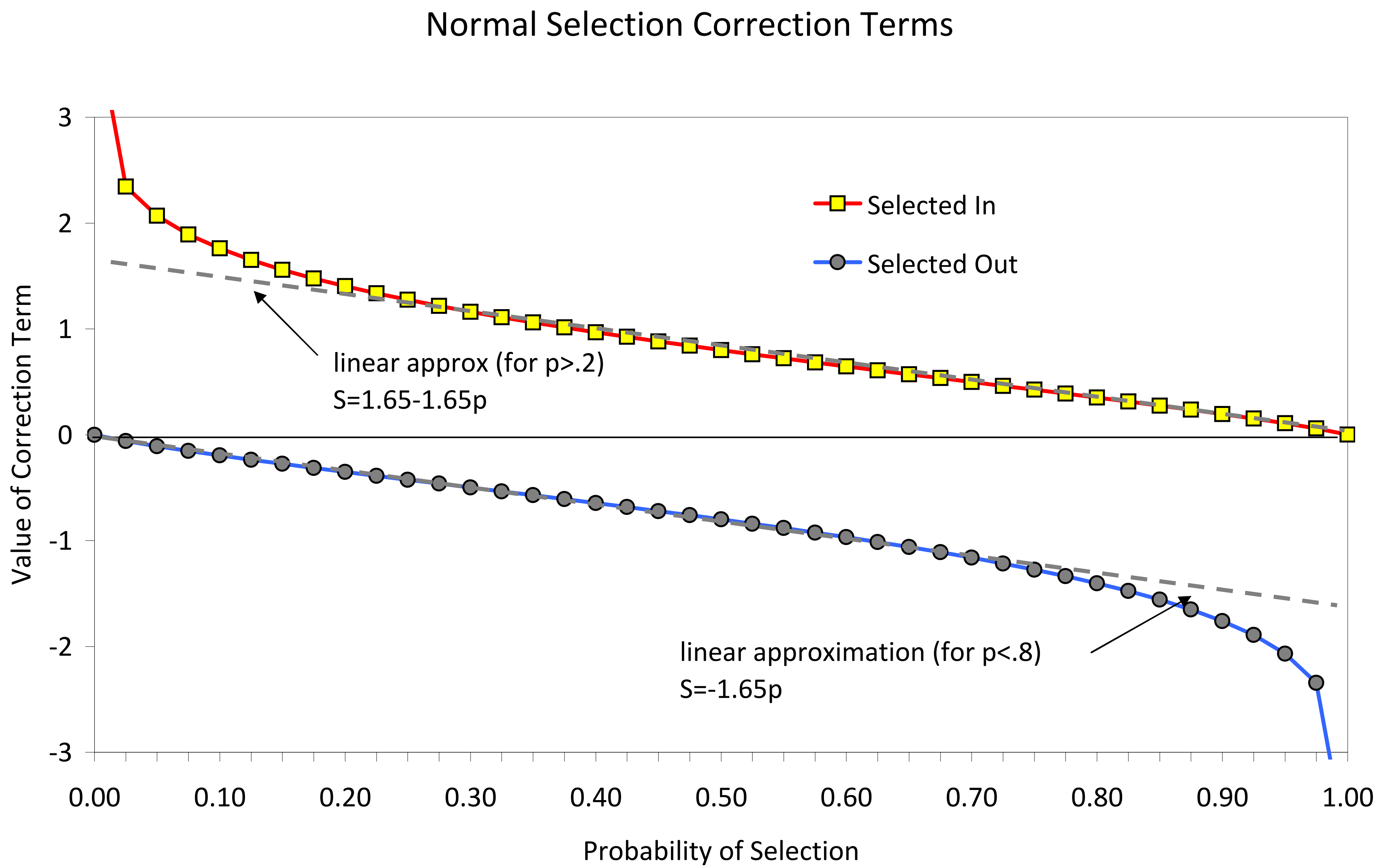
so

$$\log E[x] = \mu + \frac{1}{2}\sigma^2 = E[\log x] + \frac{1}{2}var[\log x]$$

In the labor supply setting, then, if λ_t was log-normal we would have that

$$\phi_t = \frac{1}{2}var_{t-1}[\log \lambda_t]$$

which could well vary over time or across people.



Normal Selection Correction Terms
Selected In
Selected Out
linear approx (for p>.2)
S=1.65-1.65p
linear approximation (for p<.8)
S=-1.65p
Value of Correction Term
Probability of Selection
3
2
1
0
-1
-2
-3
0.00
0.10
0.20
0.30
0.40
0.50
0.60
0.70
0.80
0.90
1.00

Figure 4.2: Intertemporal Labor Supply Choices: Separable Case

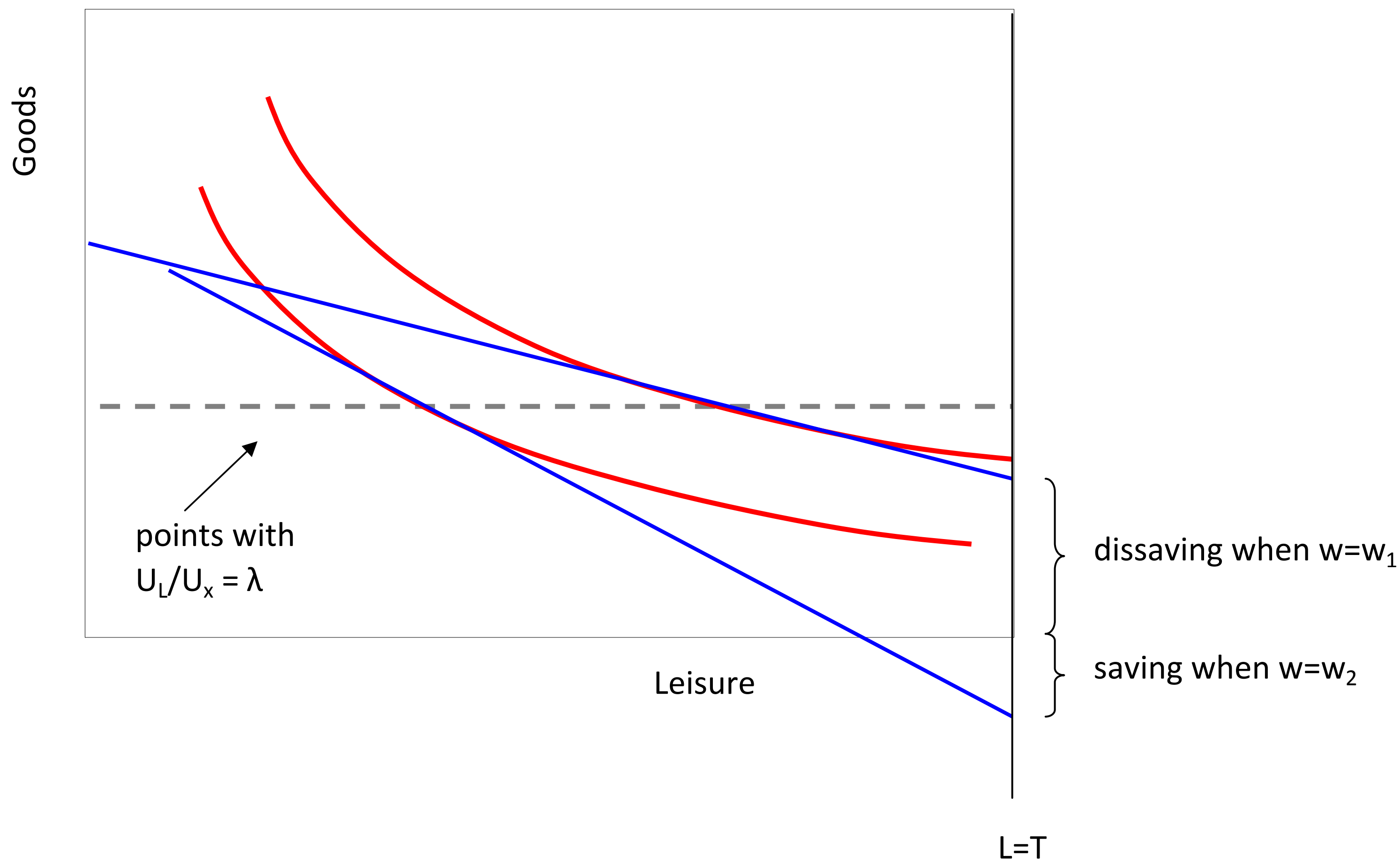


Figure 4.3: Intertemporal Labor Supply Choices: Non-separable Case