Empirical Exercise on Structural Estimation

Econ 350: The University of Chicago

This Draft: December 28, 2013

Abstract

This document briefly describes what are structural models in Economics. It distinguishes structural and non-structural estimation as approaches that enable to answer different research questions. Also, it discusses parametric and non-parametric assumptions as auxiliary means that help to recover parameters that enable to answer these research questions. The document builds on Keane et al. (2011). The difference between this document and that paper is that this document intends to be more concrete and schematic about how the modeling approaches behave, it skips the least algebra steps in all the derivations, and it provides practical examples. Also, it provides extra exercises that teach basic tools of Computational Econometrics.

1 Introduction

The objective of this document is to build on Keane et al. (2011) and clarify Structural Estimation methods. The focus is on the estimation of Discrete Choice Dynamic Programming (DCDP) models. In particular, the idea is to (i) illustrate the basic ideas and concepts; (ii) provide examples.

1.1 DCDP as an Extension of the Static Discrete Choice Framework

DCDP models are a natural generalization of static discrete choice models. They share the latent variable specification. To illustrate this, consider a binary choice model in which an economic agent, i, makes a choice at each discrete period t. $\mathcal{I}$ indexes individuals and $\mathcal{T}$ time. She has two alternatives: $d_{it} \in \{0, 1\}$. A latent variable, v^*_{it}, which is the difference in the expected payoffs between the choices $d_{it} = 1$ and $d_{it} = 0$, determines the outcome. Specifically, if v^*_{it} is greater than certain threshold, the agent chooses $d_{it} = 1$. Without loss of generality, the threshold is normalized to zero. Thus, $d_{it} = 1$ iff $v_{it}^* \geq 0$ and $d_{it} = 0$ otherwise.

Exercise 1.1 *(Identification of the Probit Model) Model a bivariate, static, discrete choice through a Probit model. The convention is that in this model the unobserved variable is normally distributed. (i) Show that you can normalize the threshold that defines the agent's decision without loss of generality; (ii) why is the normalization without loss of generality?; (iii) what other normalization can you make without loss of generality in your Probit model?; (iv) what does this normalization implies with respect to the distribution of the unobserved variable? (v) are you able to identify all the parameters of the model?; (vi) how does identification and the normalizations relate to each other? Hint: think of the scalar and spatial identification issues that the structure of a Probit model generates.*

Answer:

See the separate handout.

Exercise 1.2 *(Computational Econometrics: Warm-up) Solve the exercise "Warmup.pdf" posted on the web site. The objective of this is for you to go from the basics (i.e., installing Python in your computer and setting up the function maximizer) to an exercise in which you can maximize a likelihood function.*

Answer:

See the separate handout.

Exercise 1.3 *(Estimation of the Probit Model) From Exercise 1.1 you have clear the setup of the Probit model. Make sure you know what the correct parametric assumptions are and what can you identify in the model. Simulate all the data necessary to estimate a Probit model. The instructions are loose in purpose because we want to evaluate your ability to create the data and estimate the model from scratch. Hint: simulate a single independent variable. This is sufficient for the purposes of this exercise.*

Answer:

See the separate handout.

In general, the latent variable is a function of three variables: (i) $\tilde{D_{it}}$, a vector of the history of past choices (i.e., $d_{i\tau}, \tau = 0, \ldots, t-1$); (ii) $\tilde{X_{it}}$, a vector of contemporaneous and lagged values of J variables (i.e., $X_{ij\tau}, j = 1, \ldots, J; \tau = 0, \ldots, t-1$); (iii) $\tilde{\epsilon_{it}}$, a vector of contemporaneous and lagged unobserved variables (i.e., $\epsilon_{i\tau}, \tau = 0, \ldots, t-1$). Thus, the general decision rule of the agent is:

$$d_{it} = \begin{cases} 1 \text{ if } v_{it}^{*}\left(\tilde{D_{it}}, \tilde{X_{it}}, \tilde{\epsilon_{it}}\right) \geq 0 \\ 0 \text{ if } v_{it}^{*}\left(\tilde{D_{it}}, \tilde{X_{it}}, \tilde{\epsilon_{it}}\right) < 0. \end{cases} \tag{1}$$

Any binary choice model is a special cases of this formulation, no matter if they are static or dynamic. The model is dynamic if agents are forward looking and either $v_{it}^{*}(\cdot)$ contains past choices, $\tilde{D_{it}}$, or unobserved variables in $\tilde{\epsilon_{it}}$ that are serially correlated. The model is static if (i) agents are myopic so that even when they accumulate information on past decisions or past unobserved variables they do not take them into account; (ii) agents are forward looking but there is no link between present and past decisions and unobserved variables.

Remark 1.4 *In this context, forward looking refers to the behavior in which agents consider how their present decisions affect their future welfare. The exact way in which they form the expectations on how their welfare is affected is a modeling decision that the researcher makes.*

Exercise 1.5 *The last paragraphs clarify that there is a general framework to think of either static or dynamic binary choice models. Argue that this can be generalized for multivariate models. Write down a general framework for the multivariate case that encompasses static and dynamic models. Specify conditions under which the model is either static or dynamic.*

Answer:

In the case of the binary choice, $v_{it}^{*}\left(\tilde{D_{it}}, \tilde{X_{it}}, \tilde{\epsilon_{it}}\right)$ is a function of the utility individual i has at time t in each of the states. Often, $v(\cdot)$, is actually the utility in one state less the utility in the other. The rule is simple. In a multiple choice framework a bit more notation is necessary. The choice set, $\mathcal{S}$, has a cardinality greater or equal than 3, $\#\mathcal{S} \geq 3$. Let $U_{it}(s)\left(\tilde{D_{it}}, \tilde{X_{it}}, \tilde{\epsilon_{it}}\right)$ be the utility of *individual i at time t in state (i.e., when she chooses) s, $\forall s \in \mathcal{S}$. Then,*

$$D_{it}(j) = \begin{cases} 1: & \underset{s\in\mathcal{S}}{\text{argmax}} \ \{U_{it}(s)\} = j \\ 0: & \textit{otherwise.} \end{cases} \tag{2}$$

with $\sum_{s\in\mathcal{S}} D_{it}(s) = 1$. $\tilde{D_{it}}$ is still the set that encodes all the past choices individuals make before t,. This model encompasses both the static and the dynamic cases. The model is dynamic if there is a link between between t and $t-1$ in any of the variables $\tilde{D_{it}}, \tilde{X_{it}}, \tilde{\epsilon_{it}}$.

This document follows Keane et al. (2011) and argues that there are three broad research goals in the estimation of DCDP models:

1. Test a prediction of the theory: how an observed variable in v_{it}^* affects d_{it}.
2. Determine the effect of an endogenous shift: how a change in $\tilde{D_{it}}$ or $\tilde{X_{it}}$ affects d_{it}.
3. Determine the effect of an exogenous shift: how a change in something not in $\tilde{D_{it}}$ or $\tilde{X_{it}}$ affects d_{it}.

The objective is to answer these questions *caeteris paribus*. *Caeteris paribus* in this context not only means that the "rest" of the variables are held fixed. It implies that the the unobserved variables are also held fixed and that their joint distribution is not altered.[1] Different modeling decisions and estimation approaches enable to attain these research goals to different extents (see Section 2).

2 Model Classifications, Estimation Strategies, and Research Goals

In this section we consider the static model in Keane et al. (2011) to illustrate how different modeling approaches and estimation strategies enable to attain the research goals in Section 1.1.

[1]See Heckman and Pinto (2013) for a discussion on what "fixing" in Economics means and how it differs from "conditioning" in Statistics.

2.1 Woman's Labor Force Participation

Consider the following model of the labor force participation of a married woman. The model is unitary and the couple's i utility at time t is

$$U_{it} = U\left(c_{it}, 1 - d_{it}; n_{it}(1 - d_{it}), \kappa_{it}(1 - d_{it}), \epsilon_{it}(1 - d_{it})\right) \tag{3}$$

where c_{it} is consumption, d_{it} is an indicator of the woman's labor supply (1 if she works and 0 if she does not), n_{it} is the number of young children that the couple has, κ_{it} and ϵ_{it} are observed and unobserved factors that shift the couple's valuation of home production. Actually, t corresponds to the couple's marriage duration. The utility function satisfies standard concavity and Inada conditions.

The wife receives a wage offer w_{it} in each period t and the husband, who works every period, receives an income y_{it}. If the wife works, the family needs to pay child care, π for each child in each period. Hence, the budget constraint is

$$c_{it} = y_{it} + w_{it}d_{it} - \pi n_{it}d_{it}. \tag{4}$$

In this simple model, a wage function determines the wage offer that women receive:

$$w_{it} = w(z_{it}, \eta_{it}) \tag{5}$$

where z_{it} are observed and η_{it} unobserved factors. By assumption, ϵ_{it}, η_{it} are not serially correlated between each other.

Exercise 2.1 *Is this model static or dynamic?*

Answer:

This model is static. There is no link between t and $t-1$ in any of the variables (withing or across).

Exercise 2.2 *What are the variables that you expect to find in z_{it}?*

Answer:

Education and experience are two typical examples of variables in z_{it}.

Exercise 2.3 *Why is no serial correlation between ϵ_{it}, η_{it} a relevant assumption? Is it a technical or an economic assumption? Is it realistic? Hint: go ahead and answer the reminder of the exercise and then come back to this question.*

Answer:

This assumption is technical. Note that no serial correlation between ϵ_{it}, η_{it} is one of the reasons why the model is static. Whether this is realistic or not depends on the context but of course it is natural to think that shocks people receive are correlated over time.

This structure, actually, is enough to describe the problem through a decision rule that a latent variable dictates, as in (1). Specifically, substitute (4),(5) into (3) and note that

$$d_{it} = \begin{cases} 1 \text{ if } v_{it}^{*}\left(y_{it}, z_{it}, n_{it}, \kappa_{it}, \epsilon_{it}, \eta_{it}\right) \; \geq 0 \\ 0 \text{ if } v_{it}^{*}\left(y_{it}, z_{it}, n_{it}, \kappa_{it}, \epsilon_{it}, \eta_{it}\right) < 0 \end{cases} \tag{6}$$

where $v_{it}^{*}\left(y_{it}, z_{it}, n_{it}, \kappa_{it}, \epsilon_{it}, \eta_{it}\right) \equiv U_{it}^{1} - U_{it}^{0}$ and

$$U_{it}^{1} = U(y_{it} + w_{it}(z_{it}, \eta_{it}) - \pi n_{it}, 0) \tag{7}$$

$$U_{it}^{0} = U(y_{it}, 1; n_{it}, \kappa_{it}, \epsilon_{it}). \tag{8}$$

Definition 2.4 *(The State Space)*

1. *Household State Space:* $\Omega_{it} = \{y_{it}, z_{it}, n_{it}, \kappa_{it}, \epsilon_{it}, \eta_{it}\}$.

2. *Observed Household State Space:* $\Omega_{it}^{-} = \{y_{it}, z_{it}, n_{it}, \kappa_{it}\}$.

3. *The set of values of the unobserved variables that makes a household with observed state space Ω_{it}^{-} choose $d_{it} = 1$:* $S\left(\Omega_{it}^{-}\right) = \{\epsilon_{it}, \eta_{it} : v^{*}\left(\epsilon_{it}, \eta_{it}; \Omega_{it}^{-}\right) \geq 0\}$.

This enables to write

$$\begin{aligned} \Pr\left(d_{it} = 1 | \Omega_{it}^{-}\right) &= \int\limits_{S\left(\Omega_{it}^{-}\right)} dF_{\epsilon,\eta|y,\kappa,z,n} \\ &= G\left(y_{it}, z_{it}, n_{it}, \kappa_{it}\right). \end{aligned} \tag{9}$$

Obviously, $\Pr\left(d_{it}=0|\Omega_{it}^{-}\right) = 1-\Pr\left(d_{it}=1|\Omega_{it}^{-}\right)$. The main components of $G(y,\kappa,z,n)$ are $U(\cdot), w(\cdot), F_{\epsilon,\eta,y,\kappa,z,n}$, which conform the *structure* or the *set of primitives* of the model. Consider the following definitions of estimation approaches and auxiliary assumptions.

Definition 2.5 *(Estimation Approaches)*

1. *Structural (S): it recovers some or all of the parameters of that define the structure of the model.*

2. *Non-Structural (NS): it recovers* $G(\cdot)$.

Definition 2.6 *(Auxiliary Assumptions for Identification)*

1. *Parametric (P): assumes parametric forms about the structure of the model or about* $G(\cdot)$.

2. *Non-Parametric (NP): it does not impose parametric forms on either the structure or* $G(\cdot)$.

The combination of the initial approaches and the two auxiliary assumptions for identifications leads to a total of four possible estimation approaches: (i) S-P; (ii) S-NP; (iii) NS-P; (iv) NS-NP. The relevant question to ask is which of the estimation approaches enable to attain the research goals in Section 1.1.

Exercise 2.7 *(The Joint Distribution of Observed and Unobserved Variables) Give a sufficient condition on the joint distribution of observed and unobserved variables to attain each of the research goals in Section 1.1. Hint: Think if you can ask the questions implied by the research goals without an assumption about the relation between the unobserved variables that affect preferences and the wage function and the observed variables. Then, make a simple assumption about the joint distribution of the observed and unobserved variables.*

Answer:

It is necessary *to make an assumption of independence between the unobserved variables,* ϵ_{it}, η_{it}, *and the observed variables,* $y_{it}, z_{it}, n_{it}, \kappa_{it}$. *Otherwise, variation in* $y_{it}, z_{it}, n_{it}, \kappa_{it}$ *either across individuals in* t *or within individuals in any period* $t,\ldots,T$, *causes* d_{it} *through two channels: (i) direct effect on preferences and wages; (ii) indirect effects on preferences and wages through the effect that unobserved variables have on this. A natural step is to shut down the second channel and a* sufficient *condition is full independence between* ϵ_{it}, η_{it} *and* $y_{it}, z_{it}, n_{it}, \kappa_{it}$: $F_{\epsilon,\eta}|y,z,n,\kappa = F_{\epsilon,\eta}$.

This model enables to illustrate how different estimation approaches help to attain the different research goals in Section 1.1. Consider the following examples:

1. Goal 1: from (3) note that an increase in wage increases the utility the household has if the woman works and does not affect the utility the household has when the woman does not work. Then, a test of the theory is to analyze if the probability of woman's employment is increasing in wage.

2. Goal 2: take the derivative of $G(\cdot)$, the probability of woman's employment, with respect to any of the state variables.

3. Goal 3: take the derivative of $G(\cdot)$, the probability of woman's employment, with respect to a variable that is outside the model (e.g., π, the per-child cost of child-care).

Exercise 2.8 *(Estimation Approaches and Research Goals) What research goals can you attain with the estimation approaches NP-NS, P-NS, P-NS. Be as formal as possible. Hint: think of the different effects that you are able to identify.*

- *NP-NS: the base of this approach is a non-parametric estimate of $G(\cdot)$.*
  - *Goal 1: the typical example of an exercise towards Goal 1 is to inspect how a change in the wage affects the woman's labor force participation decision. An important feature of this model is that wages are observed iff the woman participates of the labor market. This shapes the possibility to answer research questions related to Goal 1. This exercise requires variation in w_{it} that is independent to variation in other variables that determine participation. Put differently, the exercise requires a variable in z_{it} which is not in κ_{it} (an exclusion restriction). Also, it is necessary to sign the effect of the exclusion restriction on the wage. This follows because w_{it} enters indirectly into the latent function, through z_{it}. Again, the reason of this is that the wages are not observed for everyone.*
  - *Goal 2: it is possible to obtain a non-parametric estimate of $G(\cdot)$ (think of a empirical c.d.f.). This enables to test the change of any of the variables on participation* within the range of the data. *Since the estimate is non-parametric, it is impossible to make out-of-sample tests.*

- *Goal 3: a natural research question towards this goal is to study what happens when π changes. It is impossible to identify π from $G(\cdot)$, however. Note that $G_n = \pi G_{n\pi}$. Thus, it is not possible to identify π from $G_{n\pi}$ and carry on with the exercise.*

- *P-NS: the base of this approach is a parametric assumption about $G(\cdot)$.*
  - *Goal 1: by the same reasoning as in the NP-NS approach this requires an exclusion restriction. This is simply because a parametric assumption does not imply that the effects through z_{it} and κ_{it} without an exclusion restriction.*
  - *Goal 2: the parametric assumption on $G(\cdot)$ enables to test the change of any of the variables on participation* within and outside the range of the data.
  - *Goal 3: again, it is impossible to identify π from $G(\cdot)$.*
- *NP-S: this approach is not feasible because it implies to identify $U(\cdot), w(\cdot), F$ without any parametric or auxiliary assumptions and the wages for the women who do not work are not observed for everyone. When the wages are observed for everyone the structure is reduced and some normalizations enable to identify it (see Matzkin, 1993).*
- *P-S: see Exercise 2.17.*

2.2 Estimation of a Parametric, Structural Model

In this exercise you will take a S-P approach to estimate the model in Section 2.1. Of course, there are many variations of parametric assumptions that you can impose. We guide you and you estimate. Various exercises lead to the final answer.

Assumption 2.9 *(Utility and Wage Functions and the Joint Distribution of Unobserved Variables) The utility function is:*

$$U_{it} = c_{it} + \alpha_{it}(1 - d_{it}) \tag{10}$$

where $\alpha_{it} = \beta_\kappa \kappa_{it} + \beta_n n_{it} + \epsilon_{it}$ and β_κ, β_n are scalars. The wage function is:

$$w_{it} = z_{it}\gamma + \eta_{it}. \tag{11}$$

The distribution of unobserved variables is

$$f\left(\epsilon_{it},\eta_{it}\right)\sim\mathcal{N}\left(0,\Lambda\right) \tag{12}$$

where $\begin{pmatrix} \sigma_{\epsilon}^{2} & \cdot \\ \sigma_{\epsilon,\eta} & \sigma_{\eta}^{2} \end{pmatrix}$.

Exercise 2.10 *(Wage Normality) Is it odd to model the shock to wages as normal? Why is it useful?*

Answer:

It is odd because wages are positive values. If $z_i\gamma$ is large enough this causes no negative wages. Modeling the wage shocks through a distribution with positive support is more natural but normality makes the model algebraically tractable.

Exercise 2.11 *(The State Space) Define Ω_{it} and Ω_{it}^{-} for this problem.*

Answer:

(Same as before) $\Omega_{it}=\{y_{it},z_{it},n_{it},\kappa_{it},\epsilon_{it},\eta_{it}\}$. $\Omega_{it}^{-}=\{y_{it},z_{it},n_{it},\kappa_{it}\}$.

Exercise 2.12 *(Latent Variable Function) Use Assumption 2.9 to write down the latent variable function. First define U_{it}^{1} and U_{it}^{0}. Your latent function should be a function of $\xi_{it}\equiv\eta_{it}-\epsilon_{it}$ and $\xi_{it}^{*}\left(\Omega_{it}^{-}\right)\equiv z_{it}\gamma-(\pi\beta_n)-\kappa_{it}\beta_{\kappa}$. Use this notation for the rest of the problem.*

Answer:

$$\begin{aligned}
U_{it}^{1} &= c_{it}\\
&= y_{it}+w_{it}-\pi n_{it}\\
&= y_{it}+z_{it}\gamma+\eta_{it}-\pi n_{it} && (13)\\
U_{it}^{0} &= y_{it}+\beta_{\kappa}\kappa_{it}+\beta_{n}n_{it}+\epsilon_{it} && (14)\\
U_{it}^{1}-U_{it}^{0} &\equiv v_{it}^{*}\left(y_{it},z_{it},n_{it},\kappa_{it},\epsilon_{it},\eta_{it}\right)\\
&= z_{it}\gamma-\left(\pi+\beta_{n}\right)n_{it}-\beta_{\kappa}\kappa_{it}+\eta_{it}-\epsilon_{it}\\
&\equiv \xi_{it}^{*}\left(\Omega_{it}^{-}\right)+\xi_{it}. && (15)
\end{aligned}$$

Exercise 2.13 *(Individual and Sample Likelihood Function) Write down the individual likelihood that individual i at time t contributes to the sample likelihood function. Write down the sample likelihood function.*

Answer (assume that the data is strongly balanced, i.e. observe all the individuals in the same periods):

The likelihood of individual i at time t is

$$\begin{aligned} L_{it}\left(\theta|\Omega_{it}^{-}\right) &= \Pr\left(d_{it}=1, w_{it}|\Omega_{it}^{-}\right)^{d_{it}} \times \Pr\left(d_{it}=0|\Omega_{it}^{-}\right)^{1-d_{it}} \\ &= \Pr\left(\xi_{it} \geq -\xi_{it}^{*}\left(\Omega_{it}^{-}\right), \eta_{it}=w_{it}-z_{it}\gamma\right)^{d_{it}} \\ &\times \Pr\left(\xi_{it} < -\xi_{it}^{*}\left(\Omega_{it}^{-}\right)\right)^{1-d_{it}} \end{aligned} \tag{16}$$

... and the sample likelihood is

$$L\left(\theta|\Omega_{it}^{-}\right) = \prod_{i\in\mathcal{I}}\prod_{t\in\mathcal{T}} L_{it}\left(\theta|\Omega_{it}^{-}\right). \tag{17}$$

where θ is the vector of estimands (see Exercise 2.14). Importantly, note that $\Pr\left(d_{it}=1,\eta_{it}|\cdot\right)^{d_{it}}$ is not a probability but a mixture of the c.d.f. (of the decision, d) and the p.d.f. (of the wage, w_{it}). Further clarification is necessary. The likelihood for the household in which woman do not work is

$$\begin{aligned} L_{it}^{0} &= \Pr\left(d_{it}=0|z_{it},n_{it},\kappa_{it}\right) \\ &= \Pr\left(\xi_{it} \leq -\xi_{it}^{*}|\cdot\right) \\ &= \Pr\left(\frac{\xi_{it}}{\sigma_{\xi}} \leq -\frac{\xi_{it}^{*}}{\sigma_{\xi}}|\cdot\right) \\ &= \Phi\left(-\frac{\xi_{it}^{*}}{\sigma_{\xi}}\right). \end{aligned} \tag{18}$$

As usual, Φ and ϕ are the c.d.f. and p.d.f. of a univariate normal standard distribution, respectively.

If the woman works:

$$
\begin{aligned}
L_{it}^{1} &= \Pr\left(d_{it}=1, w_{it}|z_{it}, n_{it}, \kappa_{it}\right) \\
&= \Pr\left(\xi_{it} \geq -\xi_{it}^{*}, \eta_{it}=w_{it}-z_{it}\gamma|\cdot\right) \\
&= \int\limits_{-\xi_{it}^{*}}^{\infty} f\left(\xi_{it}, \eta_{it}\right) d\xi_{it} \\
&= \int\limits_{-\xi_{it}^{*}}^{\infty} f\left(\xi_{it}|\eta_{it}\right) f\left(\eta_{it}\right) d\xi_{it} \\
&= f\left(\eta_{it}\right) \int\limits_{-\xi_{it}^{*}}^{\infty} f\left(\xi_{it}|\eta_{it}\right) d\xi_{it} \\
&= \frac{1}{\sigma_{\eta}} \phi\left(\frac{\eta_{it}}{\sigma_{\eta}}\right) \int\limits_{-\xi_{it}^{*}}^{\infty} f\left(\xi_{it}|\eta_{it}\right) d\xi_{it} \mathbf{??}.
\end{aligned}
\tag{19}
$$

where the last line follows from Subsection A.1 in Appendix A. From A.2 in Appendix A note that

$$
\xi_{it}|\eta_{it} \overset{iid}{\sim} \mathcal{N}\left(\mu_{\xi.\eta}, \sigma_{\xi.\eta}^{2}\right) \tag{20}
$$

where $\mu_{\xi.\eta} \equiv \frac{\sigma_{\xi,\eta}\eta_{it}}{\sigma_{\eta}^{2}}$ *and* $\sigma_{\xi.\eta}^{2} = \sigma_{\xi}^{2} - \frac{\sigma_{\xi,\eta}^{2}}{\sigma_{\eta}^{2}}$. *This suffices to calculate the value of the integral in (**??**) (it is simply equal to* $\Phi\left(\frac{\xi_{it}^{*}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right)$*). For completeness, consider the following*

$$
\int\limits_{-\xi_{it}^{*}}^{\infty} f\left(\xi_{it}|\eta_{it}\right) d\xi_{it} = \frac{1}{\sigma_{\xi.\eta}} \int\limits_{-\left(\frac{\xi_{it}^{*}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right)}^{\infty} \phi\left(\frac{\xi_{it}-\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right) d\xi_{it}. \tag{21}
$$

To simplify this further, let $u_{it} \equiv \frac{\xi_{it}-\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}$ *and note that* $\frac{du_{it}}{d\xi_{it}} = \frac{1}{\sigma_{\xi.\eta}}$*. Thus,*

$$\begin{aligned}
\frac{1}{\sigma_{\xi.\eta}} \int_{-\left(\frac{\xi^*_{it}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right)}^{\infty} \phi\left(\frac{\xi_{it}-\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right) d\xi_{it} &= \int_{-\left(\frac{\xi^*_{it}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right)}^{\infty} \phi(u_{it}) du_{it} \\
&= 1 - \Phi\left(-\frac{\xi^*_{it}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right) \\
&= \Phi\left(\frac{\xi^*_{it}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right). \qquad (22)
\end{aligned}$$

Therefore:

$$L^1_{it} = \frac{1}{\sigma_\eta}\phi\left(\frac{\eta_{it}}{\sigma_\eta}\right) \times \Phi\left(\frac{\xi^*_{it}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right). \qquad (23)$$

so that:

$$L_{it}\left(\theta|\Omega^-_{it}\right) = \left[\frac{1}{\sigma_\eta}\phi\left(\frac{\eta_{it}}{\sigma_\eta}\right) \times \Phi\left(\frac{\xi^*_{it}+\mu_{\xi.\eta}}{\sigma_{\xi.\eta}}\right)\right]^{d_{it}} \left[\Phi\left(-\frac{\xi^*_{it}}{\sigma_\xi}\right)\right]^{1-d_{it}}. \qquad (24)$$

Exercise 2.14 *(Estimands and Identification) What is the set of parameters that you want to estimate? Are all of these parameters identified? Hint: read Heckman (1979).*

Answer:

The estimands are $\beta_k, \beta_n, \gamma, \pi, \sigma^2_\epsilon, \sigma^2_\eta, \sigma_{\epsilon,\eta}$*. It is impossible to identify* β_n *from* π*. It is only possible to identify* $\beta_n + \pi$*. For the rest of the parameters Heckman (1979) shows identification. Joint normality enables to identify* $\gamma, \sigma^2_\eta, \frac{\sigma^2_\eta - \sigma_{\epsilon,\eta}}{\sigma_\xi}$*. The data on work choices identifies* $\frac{\gamma}{\sigma_\xi}$ *and* $\frac{\beta}{\sigma_\xi}$*. The identification of* σ_ξ *requires a variable in* z_{it} *which is not in* κ_{it}*. This identification provides the identification of* $\sigma^2_{\epsilon,\eta}$*.*

Exercise 2.15 *(Simulation) Simulate a strongly balanced data set with* $n = 1000$ *observations and* $T = 3$*. Use the following parameters:* $\beta_\kappa = 0.5, \beta_n = 0.2, \sigma_\epsilon = 1, \pi = 0.2, \gamma_1 = 0.8, \sigma_\eta = 0.2, \sigma_{\epsilon\eta} = 0.3$*. Assume that* $y_{it} \overset{iid}{\sim} \mathcal{U}(0, 10)$*.* κ_{it}*,* z_{it} *and* n_{it} *are time invariant. In particular,* $\kappa_i, z_i \overset{iid}{\sim} \mathcal{U}(0, 5)$*, and* n_i *follows a discrete uniform distribution and* $n_i \in \{0, 1, 2, 3\}$*. Use the Numpy random package in Python and set the seed to zero.*

Answer:

See file "womanssimulation.py".

Exercise 2.16 *(Estimation) Estimate the parameters of the model by ML. Compare your results with the parameters in Exercise 2.15. Hint: if the BFGS algorithm does not work use the Nelder-Mead* *algorithm.*

Answer:

See file "womansestimation.py".

Exercise 2.17 *What research goals are you able to attain with this approach?*

Answer:

- *Goal 1: As discussed above, you need an exclusion restriction. In the simulation exercise you actually have it because κ_{it}, z_{it} are generated independently.*
- *Goal 2: These approach enables to determine the effect on the the woman's labor force participation in any of the variables within and outside of sample. This follows because it recovers the structure of the model and it has specific functional forms.*
- *Goal 3: Of course! This is one of the appealing features of the P-S approach. This is usually referred as counter-factual analysis. We explain a counter-factual exercise in the TA session.*

Exercise 2.18 *Think of three policy questions that you can address with each of estimation of the model you just did. You need to link each question to each research goal in Section 1.1.*

Answer:

There are various examples. Let's discuss this in the TA session.

Exercise 2.19 *What are you able to learn from each estimation approach? Is any estimation approach better than the other? Why?*

Answer:

This follows from the answer to Exercise 2.8. There are no "best approaches". It all depends on the question you want to ask, the data availability, and the assumptions that are plausible in each scenario.

3 Discrete Choice Dynamic Programming

Consider the dynamic version of the model in Section 2.1. The utility function, the budget constraint, and the distribution of the unobserved variables are the same. The dynamics of the model come through the wage process. In particular, w_{it} increases with work experience, h_{it}. This equals the total number of periods that the woman in household i accumulates in all the periods previous to t:

$$h_{it} = \sum_{\tau=1}^{t-1} d_{i\tau}, \tag{25}$$

where $h_{i1} = 0$ for simplicity. The wage function is

$$w_{it} = z_i\gamma_1 + \gamma_2 h_{it} + \eta_{it}. \tag{26}$$

For simplicity, the variables κ_{it}, n_{it}, z_{it} are non-stochastic and time invariant.

Exercise 3.1 *(Dynamic Programming Set-up) Write down the household's problem for each period t. Let δ be the discount factor, Ω_{it} the state space, and Ω_{it}^{-} the observed state space. Write down the elements in $\Omega_{it}, \Omega_{it}^{-}$.*

Answer:

The problem is

$$\max_{d_{it}} \mathbb{E}\left\{ \sum_{\tau=t}^{T} \delta^{\tau-t} \left[U_{i\tau}^{1} d_{i\tau} + U_{i\tau}^{0} \left(1 - d_{i\tau}\right) \right] \middle| \Omega_{it} \right\} \tag{27}$$

where

$$\begin{aligned} U_{i\tau}^{1} &= y_{i\tau} + z_i\gamma_1 + \gamma_2 h_{i\tau} + \eta_{i\tau} - \pi n_i && (28)\\ U_{i\tau}^{0} &= y_{i\tau} + \beta_\kappa \kappa_i + \beta_n n_i + \epsilon_{i\tau} && (29)\\ \Omega_{it} &= \{y_{it}, z_i, n_i, \kappa_i, \epsilon_{it}, \eta_{it}, h_{it}\} && (30)\\ \Omega_{it}^{-} &= \{y_{it}, z_i, n_i, \kappa_i, h_{it}\} && (31) \end{aligned}$$

Exercise 3.2 *(Bellman Equation) Write down the recursive formulation of the household's problem.*

Answer:

Write down the value function as the maximization over the two alternative-specific value functions:

$$V_t\left(\Omega_{it}\right) = \max_{d_{it}\in\{0,1\}} \left\{ V_t^{0}\left(\Omega_{it}\right), V_t^{1}\left(\Omega_{it}\right) \right\} \tag{32}$$

where

$$V_t^k = \begin{cases} U_{it}^k(\Omega_{it}) + \delta \mathbb{E}\left[V_{t+1}(\Omega_{i,t+1}) | \Omega_{it}, d_{it} = k\right], & \textit{for } t < T \\ U_{iT}^k(\Omega_{iT}), & \textit{for } t = T \end{cases} \tag{33}$$

Exercise 3.3 *(Solution) Solve the dynamic problem of the household for for an arbitrary time horizon, T. Hint: use a backward recursion.*

Solution: (Omit the subscript i for simplicity. For the final period, $t = T$, the value functions for $d_T = 1$ and $d_T = 0$ are, respectively:

$$V_T^1(\Omega_T) = U_T^1(\Omega_T) = y_T + z\gamma_1 + h_T\gamma_2 + \eta_T - \pi n \tag{34}$$

$$V_T^0(\Omega_T) = U_T^0(\Omega_T) = y_T + \beta_\kappa \kappa + \beta_n n + \epsilon_T. \tag{35}$$

Let:

$$W^1(z, h_T, n) \equiv z\gamma_1 + h_T\gamma_2 - \pi n \tag{36}$$

$$W^0(\kappa, n) \equiv \beta_\kappa \kappa + \beta_n n. \tag{37}$$

Thus,

$$V_T^1(\Omega_T) = y_T + W^1(z, h_T, n) + \eta_T \tag{38}$$

$$V_T^0(\Omega_T) = y_T + W^0(\kappa, n) + \epsilon_T. \tag{39}$$

This implies

$$V_T^1 - V_T^0 = W^1(z, h_T, n) + \eta_T - W^0(\kappa, n) - \epsilon_T. \tag{40}$$

Define

$$\xi_T \equiv \eta_T - \epsilon_T \tag{41}$$

$$\xi_T^*(z, h_T, n, \kappa) \equiv W^1(z, h_T, n) - W^0(\kappa, n) \tag{42}$$

and note that ξ_T^* does not depend on y_T because this cancels out when in the subtraction of W^0 from W^1. Hence, $d_T = 1$ iff $\xi_T > -\xi_T^*(z, h_T, n, \kappa)$.

For $t = T - 1$

$$V_{T-1}^1(\Omega_{T-1}) = U_{T-1}^1 + \delta \mathbb{E}\{V_T(\Omega_T)|\Omega_{T-1}, d_{T-1} = 1\} \tag{43}$$

$$V_{T-1}^0(\Omega_{T-1}) = U_{T-1}^0 + \delta \mathbb{E}\{V_T(\Omega_T)|\Omega_{T-1}, d_{T-1} = 0\}. \tag{44}$$

...so that it is necessary to calculate $\mathbb{E}\{V_T(\Omega_T)|\Omega_{T-1}, d_{T-1}\}$:

$$\begin{aligned}
\mathbb{E}\{V_T(\Omega_T)|\Omega_{T-1}, d_{T-1}\} &= \mathbb{E}\left\{\max\left\{V_T^1(\Omega_T), V_T^0(\Omega_T)\right\}\middle|\Omega_{T-1}, d_{T-1}\right\} \\
&= \mathbb{E}\left\{V_T^1(\Omega_T)\middle|\Omega_{T-1}, d_{T-1}, V_T^1(\Omega_T) \geq V_T^0(\Omega_T)\right\} \\
&\quad \times \Pr\left(V_T^1(\Omega_T) \geq V_T^0(\Omega_T)\middle|\Omega_{T-1}, d_{T-1}\right) \\
&\quad + \mathbb{E}\left\{V_T^0(\Omega_T)\middle|\Omega_{T-1}, d_{T-1}, V_T^1(\Omega_T) < V_T^0(\Omega_T)\right\} \\
&\quad \times \Pr\left(V_T^1(\Omega_T) < V_T^0(\Omega_T)\middle|\Omega_{T-1}, d_{T-1}\right)
\end{aligned} \tag{45}$$

where

$$\begin{aligned}
\Pr\left(V_T^1(\Omega_T) \geq V_T^0(\Omega_T)\middle|\Omega_{T-1}, d_{T-1}\right) &= \Pr\left(\xi_T \geq -\xi_T^*(z, h_{T-1} + d_{T-1}, n, \kappa)\right) \\
&= 1 - \Phi\left(-\frac{\xi_T^*(z, h_{T-1} + d_{T-1}, n, \kappa)}{\sigma_\xi}\right) \\
&= \Phi\left(\frac{\xi_T^*(z, h_{T-1} + d_{T-1}, n, \kappa)}{\sigma_\xi}\right)
\end{aligned} \tag{46}$$

$$\begin{aligned}
\mathbb{E}\left\{V_T^1(\Omega_T)\middle|\Omega_{T-1}, d_{T-1}, V_T^1(\Omega_T) \geq V_T^0(\Omega_T)\right\} &= \mathbb{E}[y_T|\Omega_{T-1}] + W^1(z, h_{T-1} + d_{T-1}, n) \\
&\quad + \mathbb{E}\{\eta_T|\xi_T \geq -\xi_T^*(z, h_{T-1} + d_{T-1}, n, \kappa)\} \\
&= \mathbb{E}[y_T|\Omega_{T-1}] + W^1(z, h_{T-1} + d_{T-1}, n) \\
&\quad + \frac{\sigma_{\eta\xi}}{\sigma_\xi} \frac{\phi\left(-\frac{\xi_T^*(z, h_{T-1}+d_{T-1}, n, \kappa)}{\sigma_\xi}\right)}{1 - \Phi\left(-\frac{\xi_T^*(z, h_{T-1}+d_{T-1}, n, \kappa)}{\sigma_\xi}\right)}
\end{aligned} \tag{47}$$

$$\begin{aligned}
\mathbb{E}\left\{V_T^0\left(\Omega_T\right)\middle|\Omega_{T-1}, d_{T-1}, V_T^1\left(\Omega_T\right) < V_T^0\left(\Omega_T\right)\right\} &= \mathbb{E}[y_T|\Omega_{T-1}] + W^0\left(\kappa, n\right) \\
&\quad + \mathbb{E}\left\{\epsilon_T|\xi_T < -\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)\right\} \\
&= \mathbb{E}[y_T|\Omega_{T-1}] + W^0\left(\kappa, n\right) \\
&\quad - \frac{\sigma_{\epsilon\xi}}{\sigma_\xi} \frac{\phi\left(-\frac{\xi_T^*(z, h_{T-1}+d_{T-1}, n, \kappa)}{\sigma_\xi}\right)}{\Phi\left(-\frac{\xi_T^*(z, h_{T-1}+d_{T-1}, n, \kappa)}{\sigma_\xi}\right)}. \qquad (48)
\end{aligned}$$

Then,

$$\begin{aligned}
\mathbb{E}\left\{V_T\left(\Omega_T\right)|\Omega_{T-1}, d_{T-1}\right\} &= \mathbb{E}[y_T|\Omega_{T-1}] + \Phi\left(\frac{\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)}{\sigma_\xi}\right) \times W^1\left(z, h_{T-1} + d_{T-1}, n\right) \\
&\quad + \left(1 - \Phi\left(\frac{\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)}{\sigma_\xi}\right)\right) \times W^0\left(\kappa, n\right) \\
&\quad + \sigma_\xi \times \phi\left(\frac{-\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)}{\sigma_\xi}\right). \qquad (49)
\end{aligned}$$

Let

$$\begin{aligned}
\mathrm{Emax}_T\left(z, h_{T-1} + d_{T-1}, n, \kappa\right) &\equiv \Phi\left(\frac{\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)}{\sigma_\xi}\right) \times W^1\left(z, h_{T-1} + d_{T-1}, n\right) \qquad (50) \\
&\quad + \left(1 - \Phi\left(\frac{\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)}{\sigma_\xi}\right)\right) \times W^0\left(\kappa, n\right) \qquad (51) \\
&\quad + \sigma_\xi \times \phi\left(\frac{-\xi_T^*\left(z, h_{T-1} + d_{T-1}, n, \kappa\right)}{\sigma_\xi}\right) \qquad (52)
\end{aligned}$$

. . . so that

$$\mathbb{E}\left\{V_T\left(\Omega_T\right)|\Omega_{T-1}, d_{T-1}\right\} = \mathbb{E}[y_T|\Omega_{T-1}] + \mathrm{Emax}_T\left(z, h_{T-1} + d_{T-1}, n, \kappa\right).$$

This enables to write

$$\begin{aligned}
V_{T-1}^1 - V_{T-1}^0 &= W^1\left(z, h_{T-1}, n\right) - W^0\left(\kappa, n\right) + \eta_{T-1} - \epsilon_{T-1} \\
&\quad + \delta[\mathrm{Emax}_T\left(z, h_{T-1} + 1, n, \kappa\right) - \mathrm{Emax}_T\left(z, h_{T-1}, n, \kappa\right)] \qquad (53)
\end{aligned}$$

and $d_{T-1} = 1$ iff $\xi_{T-1} > -\xi^*_{T-1}(z, h_{T-1}, n, \kappa)$ where

$$\begin{aligned}
\xi^*_{T-1}(z, h_{T-1}, n, \kappa) \equiv\; & W^1(z, h_{T-1}, n) - W^0(\kappa, n) \\
& + \delta[\mathrm{Emax}_T(z, h_{T-1}+1, n, \kappa) - \mathrm{Emax}_T(z, h_{T-1}, n, \kappa)]. \qquad (54)
\end{aligned}$$

Importantly, $\mathbb{E}[y_T|\Omega_{T-1}]$ and y_{T-1} cancel out and do not appear in ξ^*_{T-1}.

Finally, at $t = T - j$ for j = ,

$$\begin{aligned}
V^1_{T-2}(\Omega_{T-2}) &= U^1_{T-2} + \delta\mathbb{E}\{V_{T-1}(\Omega_{T-1})|\Omega_{T-2}, d_{T-2} = 1\} \\
V^0_{T-2}(\Omega_{T-2}) &= U^0_{T-2} + \delta\mathbb{E}\{V_{T-1}(\Omega_{T-1})|\Omega_{T-2}, d_{T-2} = 0\}
\end{aligned}$$

To calculate $\mathbb{E}\{V_{T-1}(\Omega_{T-1})|\Omega_{T-2}, d_{T-2}\}$, we have:

$$\begin{aligned}
\mathbb{E}\{V_{T-1}(\Omega_{T-1})|\Omega_{T-2}, d_{T-2}\} &= \mathbb{E}\left\{\max\left\{V^1_{T-1}(\Omega_{T-1}), V^0_{T-1}(\Omega_{T-1})\right\}\big|\Omega_{T-2}, d_{T-2}\right\} \\
&= \mathbb{E}\left\{V^1_{T-1}(\Omega_{T-1})\big|\Omega_{T-2}, d_{T-2}, V^1_{T-1}(\Omega_{T-1}) \geq V^0_{T-1}(\Omega_{T-1})\right\} \\
&\quad \times \Pr\left(V^1_{T-1}(\Omega_{T-1}) \geq V^0_{T-1}(\Omega_{T-1})\big|\Omega_{T-2}, d_{T-2}\right) \\
&\quad + \mathbb{E}\left\{V^0_{T-1}(\Omega_{T-1})\big|\Omega_{T-2}, d_{T-2}, V^1_{T-1}(\Omega_{T-1}) < V^0_{T-1}(\Omega_{T-1})\right\} \\
&\quad \times \Pr\left(V^1_{T-1}(\Omega_{T-1}) < V^0_{T-1}(\Omega_{T-1})\big|\Omega_{T-2}, d_{T-2}\right)
\end{aligned}$$

in which we have:

$$\begin{aligned}
\Pr\left(V^1_{T-1}(\Omega_{T-1}) \geq V^0_{T-1}(\Omega_{T-1})\big|\Omega_{T-2}, d_{T-2}\right) &= \Pr\left(\xi_{T-1} > -\xi^*_{T-1}(z, h_{T-2}+d_{T-2}, n, \kappa)\right) \\
&= 1 - \Phi\left(-\frac{\xi^*_{T-1}(z, h_{T-2}+d_{T-2}, n, \kappa)}{\sigma_\xi}\right) \\
&= \Phi\left(\frac{\xi^*_{T-1}(z, h_{T-2}+d_{T-2}, n, \kappa)}{\sigma_\xi}\right)
\end{aligned}$$

$$
\begin{aligned}
\mathbb{E}\left\{V^1_{T-1}\left(\Omega_{T-1}\right)\middle|\Omega_{T-2},d_{T-2},V^1_{T-1}\left(\Omega_{T-1}\right)\geq V^0_{T-1}\left(\Omega_{T-1}\right)\right\} \quad = \quad & \mathbb{E}[y_{T-1}|\Omega_{T-2}]+W^1\left(z,h_{T-2}+d_{T-2},n\right) \\
& +\delta\mathbb{E}[y_T|\Omega_{T-2}] \\
& +\delta\,\mathrm{Emax}_T\left(z,h_{T-2}+d_{T-2}+1,n,\kappa\right) \\
& +\frac{\sigma_{\eta\xi}}{\sigma_\xi}\frac{\phi\left(-\frac{\xi^*_{T-1}(z,h_{T-2}+d_{T-2},n,\kappa)}{\sigma_\xi}\right)}{1-\Phi\left(-\frac{\xi^*_{T-1}(z,h_{T-2}+d_{T-2},n,\kappa)}{\sigma_\xi}\right)}
\end{aligned}
$$

$$
\begin{aligned}
\mathbb{E}\left\{V^0_{T-1}\left(\Omega_{T-1}\right)|\Omega_{T-2},d_{T-2},V^1_{T-1}\left(\Omega_{T-1}\right)< V^0_{T-1}\left(\Omega_{T-1}\right)\right\} \quad = \quad & \mathbb{E}[y_{T-1}|\Omega_{T-2}]+W^0\left(\kappa,n\right) \\
& +\delta\mathbb{E}[y_T|\Omega_{T-2}] \\
& +\delta\,\mathrm{Emax}_T\left(z,h_{T-2}+d_{T-2},n,\kappa\right) \\
& -\frac{\sigma_{\epsilon\xi}}{\sigma_\xi}\frac{\phi\left(-\frac{\xi^*_{T-1}(z,h_{T-2}+d_{T-2},n,\kappa)}{\sigma_\xi}\right)}{\Phi\left(-\frac{\xi^*_{T-1}(z,h_{T-2}+d_{T-2},n,\kappa)}{\sigma_\xi}\right)}
\end{aligned}
$$

So,

$$
\begin{aligned}
\mathbb{E}\left\{V_{T-1}\left(\Omega_{T-1}\right)|\Omega_{T-2},d_{T-2}\right\} \quad = \quad & \mathbb{E}[y_{T-1}|\Omega_{T-2}]+\delta\mathbb{E}[y_T|\Omega_{T-2}] \\
& +\Big(W^1\left(z,h_{T-2}+d_{T-2},n\right)+\delta\,\mathrm{Emax}_T\left(z,h_{T-2}+d_{T-2}+1,n,\kappa\right)\Big) \\
& \times\Phi\left(\frac{\xi^*_{T-1}\left(z,h_{T-2}+d_{T-2},n,\kappa\right)}{\sigma_\xi}\right) \\
& +\Big(W^0\left(\kappa,n\right)+\delta\,\mathrm{Emax}_T\left(z,h_{T-2}+d_{T-2},n,\kappa\right)\Big) \\
& \times\left(1-\Phi\left(\frac{\xi^*_{T-1}\left(z,h_{T-2}+d_{T-2},n,\kappa\right)}{\sigma_\xi}\right)\right) \\
& +\sigma_\xi\phi\left(-\frac{\xi^*_{T-1}\left(z,h_{T-2}+d_{T-2},n,\kappa\right)}{\sigma_\xi}\right)
\end{aligned}
$$

Define:

$$
\begin{aligned}
\mathrm{Emax}_{T-1}\left(z, h_{T-2}+d_{T-2}, n, \kappa\right) \equiv & \left(W^{1}\left(z, h_{T-2}+d_{T-2}, n\right)+\delta\,\mathrm{Emax}_{T}\left(z, h_{T-2}+d_{T-2}+1, n, \kappa\right)\right) \\
& \times \Phi\left(\frac{\xi_{T-1}^{*}\left(z, h_{T-2}+d_{T-2}, n, \kappa\right)}{\sigma_{\xi}}\right) \\
& +\left(W^{0}(\kappa, n)+\delta\,\mathrm{Emax}_{T}\left(z, h_{T-2}+d_{T-2}, n, \kappa\right)\right) \\
& \times\left(1-\Phi\left(\frac{\xi_{T-1}^{*}\left(z, h_{T-2}+d_{T-2}, n, \kappa\right)}{\sigma_{\xi}}\right)\right) \\
& +\sigma_{\xi} \phi\left(-\frac{\xi_{T-1}^{*}\left(z, h_{T-2}+d_{T-2}, n, \kappa\right)}{\sigma_{\xi}}\right)
\end{aligned}
$$

Then we have:

$$
\mathbb{E}\left\{V_{T-1}\left(\Omega_{T-1}\right) | \Omega_{T-2}, d_{T-2}\right\} = \mathbb{E}[y_{T-1}|\Omega_{T-2}]+\delta\mathbb{E}[y_{T}|\Omega_{T-2}]+\mathrm{Emax}_{T-1}\left(z, h_{T-2}+d_{T-2}, n, \kappa\right)
$$

Therefore, we have:

$$
\begin{aligned}
V_{T-2}^{1}-V_{T-2}^{0} = & W^{1}\left(z, h_{T-2}, n\right)-W^{0}(\kappa, n)+\eta_{T-2}-\epsilon_{T-2} \\
& +\delta[\mathrm{Emax}_{T-1}\left(z, h_{T-2}+1, n, \kappa\right) \\
& -\mathrm{Emax}_{T-1}\left(z, h_{T-2}, n, \kappa\right)]
\end{aligned}
$$

and $d_{T-2}=1$ iff:

$$
\xi_{T-2}>-\xi_{T-2}^{*}\left(z, h_{T-2}, n, \kappa\right)
$$

in which:

$$
\begin{aligned}
\xi_{T-2}^{*}\left(z, h_{T-2}, n, \kappa\right) = & W^{1}\left(z, h_{T-2}, n\right)-W^{0}(\kappa, n)+\eta_{T-2}-\epsilon_{T-2} \\
& +\delta[\mathrm{Emax}_{T-1}\left(z, h_{T-2}+1, n, \kappa\right) \\
& -\mathrm{Emax}_{T-1}\left(z, h_{T-2}, n, \kappa\right)]
\end{aligned}
$$

Again, ξ_{T-2}^{*} does not depend on y_{T-2}, $\mathbb{E}[y_{T-1}|\Omega_{T-2}]$, and $\mathbb{E}[y_{T}|\Omega_{T-2}]$.

3.1 Simulation and Estimation

Exercise 3.4 *(Likelihood) What is the individual likelihood of household i at time t? What is the sample likelihood across all periods?*

Answer:

The individual and sample likelihoods are the same as in the static case (see Section **??***). What changes are the definitions of w_{it}, ξ^*_{it}.*

Exercise 3.5 *(Simulation) Simulate a balanced data set with $n = 1000$ observations and $T = 6$. Use the same parameters as in the static model in Section 2. For the parameters that are exclusive of the dynamic model use the following: $\gamma_2 = 0.9, \delta = 0.85$. Set the experience of every woman to zero in $t = 1$. Save the data in a ".csv" file.*

Answer:

See "womansdsimulation.py".

Exercise 3.6 *(Estimation) Estimate the parameters of the model by ML. Compare your results with the parameters in Exercise 3.5.*

Answer:

See "womansdestimation.py"

References

Heckman, J. J. (1979). Sample Selection Bias as a Specification Error. *Econometrica: Journal of the econometric society*, 153–161.

Heckman, J. J. and R. Pinto (2013). Causal Analysis after Haavelmo. Technical report, National Bureau of Economic Research.

Keane, M. P., P. E. Todd, and K. I. Wolpin (2011). The Structural Estimation of Behavioral Models: Discrete Choice Dynamic Programming Methods and Applications. *Handbook of Labor Economics 4*, 331–461.

Matzkin, R. L. (1993). Non-parametric Identification and Estimation of Polychotomous Choice Models. *Journal of Econometrics 58*(1), 137–168.

A Appendix

A.1 The Univariate Normal Distribution

Let $x \sim \mathcal{N}\left(\mu, \sigma^2\right)$. Then

$$
\begin{aligned}
f(x) &= \frac{1}{\sigma}\frac{1}{2\pi}\exp\left(\frac{-1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right) \\
&= \frac{1}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right) \qquad (55)
\end{aligned}
$$

where $\phi(\cdot)$ is the p.d.f. of a univariate normal standard distribution.

A.2 The Conditional Normal Theorem

Consider two random vectors, $\mathbf{x_1}, \mathbf{x_2}$, which are jointly, normally distributed:

$$
\begin{bmatrix} \mathbf{x_1} \\ \mathbf{x_2} \end{bmatrix} \sim \mathcal{N}\left[\begin{pmatrix} \boldsymbol{\mu_1} \\ \boldsymbol{\mu_2} \end{pmatrix}, \begin{pmatrix} \boldsymbol{\Sigma_{11}} & \boldsymbol{\Sigma_{12}} \\ \boldsymbol{\Sigma_{21}} & \boldsymbol{\Sigma_{22}} \end{pmatrix}\right]. \qquad (56)
$$

Then

$$
\mathbf{x_1}|\mathbf{x_2} \sim \mathcal{N}(\boldsymbol{\mu_{1.2}}, \boldsymbol{\Sigma_{11.2}}) \qquad (57)
$$

where

$$
\begin{aligned}
\boldsymbol{\mu_{1.2}} &= \boldsymbol{\mu_1} + \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-\mathbf{1}}(\mathbf{x_2} - \boldsymbol{\mu_2}) \qquad (58) \\
\boldsymbol{\Sigma_{11.2}} &= \boldsymbol{\Sigma_{11}} - \boldsymbol{\Sigma_{12}}\boldsymbol{\Sigma_{22}}^{-\mathbf{1}}\boldsymbol{\Sigma_{21}} \qquad (59)
\end{aligned}
$$