Chapter 4. Dynamic Discrete Choice Models II: Conditional Choice Probability (CCP) Estimation

Advanced Econometric Methods II
Barcelona GSE

Introduction

Motivation

Full solution techniques are **computationally challenging**.

CCP estimation: avoid solving the DP in each iteration of estimation

Advantages:

- (Less efficient but) faster.
- More transparent about sources of variation that identify the parameters of the model.
- Robustness checks.
- Expand the set of problems that can be handled (e.g. dynamic games or non-stationary environments).

In a nutshell

Build on the seminal work of **Hotz and Miller (1993)**.

Main idea: individual choices (reflected in their CCPs) contain rich information on expectations about future outcomes.

There exists a **mapping** between conditional value functions $v_{jt}(\boldsymbol{x}_t)$ and CCPs $p_t(\boldsymbol{x}_t)$ that, in general, can be inverted.

Representing the mapping as a function of the CCPs + **nonparametric estimation** of the CCPs $\Rightarrow$ **no need to solve** for value functions.

Conditional value function representation

The Rust example

Let's first start with the representation in the Rust example.

Taking replacement as the base category, the CCPs are:

$$p_1(x) = \frac{1}{1 + e^{v_0(x) - v_1(x)}} \quad \text{and} \quad p_0(x) = \frac{e^{v_0(x) - v_1(x)}}{1 + e^{v_0(x) - v_1(x)}}.$$

Given this:

$$-\ln p_1(x) = \ln\left(1 + e^{v_0(x) - v_1(x)}\right) = \ln\left(\sum_{h \in \mathcal{D}} \exp\{v_h(x) - v_1(x)\}\right).$$

Rewriting the fixed point equation:

$$v_j(x_t) = u_j(x_t) + \beta \sum_{x \in X} \ln\left(\sum_{h \in \mathcal{D}} \exp\{v_h(x)\}\right) F^j_{x,x_t} + \beta\gamma$$

$$= u_j(x_t) + \beta \sum_{x \in X} \ln e^{v_1(x)} \left(\sum_{h \in \mathcal{D}} \exp\{v_h(x) - v_1(x)\}\right) F^j_{x,x_t} + \beta\gamma$$

$$= u_j(x_t) + \beta \sum_{x \in X} \left(v_1(x) - \ln p_1(x)\right) F^j_{x,x_t} + \beta\gamma$$

$$= u_j(x_t) + \beta v_1(0) - \beta \sum_{x \in X} \ln p_1(x) F^j_{x,x_t} + \beta\gamma.$$

The Rust example (cont'd)

Noting that $\sum_{x \in X} \ln p_1(x) F^1_{x,x_t} = \ln p_1(0)$, as F^1_{x,x_t} is **degenerate**:

$$v_0(x_t) - v_1(x_t) = u_0(x_t) - u_1(x_t) + \beta \left(\ln p_1(0) - \sum_{x \in X} \ln p_1(x) F^0_{x,x_t} \right)$$

$$= \theta_R - \theta_M x + \beta \left(\ln p_1(0) - \sum_{x \in X} \ln p_1(x) F^0_{x,x_t} \right).$$

We can obtain **non-parametric estimates** of $p_1(x)$ from the data, for instance from relative frequencies:

$$\hat{p}_1(x) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{t=1}^{T} d_{1it}\, \mathbb{1}\{x_{it} = x\}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{t=1}^{T} \mathbb{1}\{x_{it} = x\}}.$$

Replace $p_1(x)$ by $\hat{p}_1(x) \Rightarrow$ replace $v_0(x) - v_1(x)$ into **CCPs equation** $\Rightarrow$ straightforward **binary logit** (F^0_{x,x_t} estimated in a **first stage**).

General representation

The general idea is:

$$p_{Jt}(\boldsymbol{x}_t) \equiv \int d^*_{Jt}(\boldsymbol{x}_t, \boldsymbol{\varepsilon}_t) dF_\varepsilon(\boldsymbol{\varepsilon}_t)$$

$$= \int_{-\infty}^{\infty} \int_{-\infty}^{\varepsilon_{Jt} + v_{Jt}(\boldsymbol{x}_t) - v_{J-1t}(\boldsymbol{x}_t)} \dots \int_{-\infty}^{\varepsilon_{Jt} + v_{Jt}(\boldsymbol{x}_t) - v_{1t}(\boldsymbol{x}_t)} dF_\varepsilon(\boldsymbol{\varepsilon}_t)$$

$$\equiv Q_J(v_{Jt}(\boldsymbol{x}_t) - v_{1t}(\boldsymbol{x}_t), \dots, v_{Jt}(\boldsymbol{x}_t) - v_{J-1t}(\boldsymbol{x}_t)).$$

The main **theorem** in Hotz and Miller (1993) specifies that if we define $\boldsymbol{p}_t$ such that $p_{jt} > 0 \; \forall j$, there exists a real-valued function $\psi_j(\boldsymbol{p})$ for every $j \in \mathcal{D}$ such that:

$$\psi_j(\boldsymbol{p}_t(\boldsymbol{x}_t)) \equiv V_t(\boldsymbol{x}_t) - v_{jt}(\boldsymbol{x}_t) \Leftrightarrow V_t(\boldsymbol{x}_t) \equiv v_{jt}(\boldsymbol{x}_t) + \psi_j(\boldsymbol{p}_t(\boldsymbol{x}_t)).$$

Notice the connection with **selection models**.

This **formulation** is from Arcidiacono and Miller (2011), in the original, $\psi_j(\boldsymbol{p}) \equiv Q_j^{-1}(\boldsymbol{p})$ is a function of differences in v_js.

In our example, $\psi_j(\boldsymbol{p}(x_t)) = -\ln p_j(x_t) + \gamma$.

General representation

Rewrite DP into a known function of data, parameters, and CCPs:

$$v_{jt}(\boldsymbol{x}_t) = u_{jt}(\boldsymbol{x}_t) + \beta \int V_{t+1}(\boldsymbol{x}_{t+1}) dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j)$$

$$= u_{jt}(\boldsymbol{x}_t) + \beta \int \left[v_{kt+1}(\boldsymbol{x}_{t+1}) + \psi_k(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1})) \right] dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j).$$

Repeat this procedure **ad infinitum**, substituting $v_{kt+1}(\boldsymbol{x}_{t+1})$ using the above expression.

This would lead to a representation that would **not require solving** for the value functions, provided we have **nonparametric estimates** of the CCPs.

One can **choose to which alternative** k to make the future term relative $\Rightarrow$ future utility terms across two choices to be **differenced out**.

Terminal/renewal action

In the Rust example, we only needed **one-period-ahead** CCPs.

This is the case, in general, for problems with:

- **Terminal action**: no further choices are made following R.
- **Renewal action**: state variables are reset after R (Rust).

Trick: substitute $V_{t+1}(\boldsymbol{x}_{t+1}) = v_{Rt+1}(\boldsymbol{x}_{t+1}) + \psi_R(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1}))$.

Terminal action: $v_{Rt+1}(\boldsymbol{x}_{t+1})$ doesn't include **continuation value**.

Renewal action: see next slide.

Renewal action

This is the case of the **Rust** example.

By definition, the distribution of $\boldsymbol{x}_{t+2}$ **after choosing renewal** in $t+1$ is the same regardless of initial choice d_t or d_t':

$$\int f_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, d_t) f_x(\boldsymbol{x}_{t+2}|\boldsymbol{x}_{t+1}, R) d\boldsymbol{x}_{t+1} = \int f_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, d_t') f_x(\boldsymbol{x}_{t+2}|\boldsymbol{x}_{t+1}, R) d\boldsymbol{x}_{t+1}.$$

As a result:

$$\begin{aligned} v_{jt}(\boldsymbol{x}_t) &= u_{jt}(\boldsymbol{x}_t) + \beta \int \left[v_{Rt+1}(\boldsymbol{x}_{t+1}) + \psi_R(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1})) \right] dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j) \\ &= u_{jt}(\boldsymbol{x}_t) + \beta \int \left[\begin{array}{l} u_{Rt+1}(\boldsymbol{x}_{t+1}) \\ +\beta \int V_{t+2}(\boldsymbol{x}_{t+2}) dF_x(\boldsymbol{x}_{t+2}|\boldsymbol{x}_{t+1}, R) \\ +\psi_R(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1})) \end{array} \right] dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j) \\ &= u_{jt}(\boldsymbol{x}_t) + \beta \int \left[u_{Rt+1}(\boldsymbol{x}_{t+1}) + \psi_R(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1})) \right] dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j) \\ &\quad + \beta^2 \int \int V_{t+2}(\boldsymbol{x}_{t+2}) f_x(\boldsymbol{x}_{t+2}|\boldsymbol{x}_{t+1}, R) f_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j) d\boldsymbol{x}_{t+2} d\boldsymbol{x}_{t+1}. \end{aligned}$$

Renewal action (cont'd)

Given the renewal action, the last term is equal **regardless of** j.

Let alternative 1 denote the base category:

$$p_{jt}(\boldsymbol{x}_t) = \frac{\exp(v_{jt}(\boldsymbol{x}_t) - v_{1t}(\boldsymbol{x}_t))}{\sum_{h\in\mathcal{D}} \exp(v_{ht}(\boldsymbol{x}_t) - v_{1t}(\boldsymbol{x}_t))}.$$

We only need **one-period-ahead** CCPs:

$$\begin{aligned} & v_{jt}(\boldsymbol{x}_t) - v_{1t}(\boldsymbol{x}_t) \\ & \quad = (u_{jt}(\boldsymbol{x}_t) - u_{1t}(\boldsymbol{x}_t)) + \beta \left\{ \begin{array}{c} \int \left[u_{Rt+1}(\boldsymbol{x}_{t+1}) + \psi_R(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1}))\right] dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j) \\ \int \left[u_{Rt+1}(\boldsymbol{x}_{t+1}) + \psi_R(\boldsymbol{p}_{t+1}(\boldsymbol{x}_{t+1}))\right] dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, 1) \end{array} \right\}, \end{aligned}$$

for every $j \in \mathcal{D}$.

Occupational choice example

The ideas above for renewal are generalizable to a larger class of problems $\Rightarrow$ **ρ-periods-ahead** CCPs needed.

Altuğ and Miller (1998)/Arcidiacono and Miller (2011): **finite dependence**.

Occupational choice example:

- Very simplified version of Keane and Wolpin (1997).
- Stay home $d_t = H$ or work $d_t = W$.
- Observable state variable: experience (degenerate and does not depreciate).

Finite dependence: experience two periods ahead is the same if I choose work today and stay home next period than if I stay home today and work next period.

Occupational choice example (cont'd)

Therefore:

$$v_{Wt}(x_t) = u_{Wt}(x_t) + \beta \left[u_{Ht+1}(x_t + 1) + \psi_H(\boldsymbol{p}_{t+1}(x_t + 1)) \right] + \beta^2 V_{t+2}(x_t + 1).$$

$$v_{Ht}(x_t) = u_{Ht}(x_t) + \beta \left[u_{Wt+1}(x_t) + \psi_W(\boldsymbol{p}_{t+1}(x_t)) \right] + \beta^2 V_{t+2}(x_t + 1).$$

The CCPs depend on the **difference** between the two conditional value functions, which can be expressed as:

$$v_{Wt}(x_t) - v_{Ht}(x_t) = u_{Wt}(x_t) - u_{Ht}(x_t) + \beta \begin{bmatrix} u_{Ht+1}(x_t + 1) - u_{Wt+1}(x_t) \\ \psi_H(\boldsymbol{p}_{t+1}(x_t + 1)) - \psi_W(\boldsymbol{p}_{t+1}(x_t)) \end{bmatrix},$$

and only depends on the ρ**-periods ahead** CCPs ($\rho = 1$ in this case).

General representation

The above example exhibits **exact** finite dependence (sufficient condition).

The necessary condition requires it in **expectation**.

Let $\{d'_t, d'_{t+1}, ..., d'_{t+\rho}\}$ define a **sequence of decisions** from t to $t+\rho$ (no need to be optimal).

For each $\tau \in \{t, ..., t+\rho\}$, denote $\kappa'_\tau(\boldsymbol{x}_{\tau+1}|\boldsymbol{x}_t)$ as the **cumulative probability of being in state $\boldsymbol{x}_{\tau+1}$** given $\{d'_t, d'_{t+1}, ..., d'_{t+\rho}\}$, recursively defined as:

$$\kappa'_\tau(\boldsymbol{x}_{\tau+1}|\boldsymbol{x}_t) \equiv \begin{cases} f(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, d'_t) & \text{if } \tau = t \\ \int f(\boldsymbol{x}_{\tau+1}|\boldsymbol{x}_\tau, d'_\tau)\kappa'_{\tau-1}(\boldsymbol{x}_\tau|\boldsymbol{x}_t)d\boldsymbol{x}_\tau & \text{if } \tau > t. \end{cases}$$

General representation

Using this expression, we can **telescope** $v_{jt}(\boldsymbol{x}_t)$ ρ-periods ahead to obtain:

$$v_{\{d'_t\}t}(\boldsymbol{x}_t) = u_{\{d'_t\}t}(\boldsymbol{x}_t) + \sum_{\tau=t+1}^{t+\rho} \beta^{\tau-t} \int \left[u_{\{d'_\tau\}\tau}(\boldsymbol{x}_\tau) + \psi_{\{d'_\tau\}}(\boldsymbol{p}_\tau(\boldsymbol{x}_\tau))\right] \kappa'_{\tau-1}(\boldsymbol{x}_\tau|\boldsymbol{x}_t)d\boldsymbol{x}_\tau$$
$$+ \beta^{\rho+1} \int V_{t+\rho+1}(\boldsymbol{x}_{t+\rho+1})\kappa'_{t+\rho}(\boldsymbol{x}_{t+\rho+1}|\boldsymbol{x}_t)d\boldsymbol{x}_{t+\rho+1}.$$

Now we can define an alternative sequence of decisions $\{d''_t, d''_{t+1}, ..., d''_{t+\rho}\}$ that lead the individual to the **same state in expectation**, or, equivalently, such that:

$$\kappa'_{t+\rho}(\boldsymbol{x}_{t+\rho+1}|\boldsymbol{x}_t) = \kappa''_{t+\rho}(\boldsymbol{x}_{t+\rho+1}|\boldsymbol{x}_t) \text{ for all } \boldsymbol{x}_{t+\rho+1}.$$

Clearly, in the difference $v_{\{d''_t\}t}(\boldsymbol{x}_t) - v_{\{d'_t\}t}(\boldsymbol{x}_t)$, the **last term** of the top expression **cancels**, and the resulting expression only depends on flow payoffs and CCPs up to period $t+\rho$.

Infinite-horizon stationary settings

CCP representation is still very straightforward in **stationary infinite-horizon settings** in the absence of finite dependence.

Matrix representation by Aguirregabiria and Mira (2002).

Define $\varepsilon^*_{jt}(\boldsymbol{x}_t) \equiv \mathbb{E}[\varepsilon_{jt}|d^*_t = j, \boldsymbol{x}_t]$, where d^*_t denotes the optimal choice.
The **ex-ante** value function can be written as:

$$\begin{aligned} V(\boldsymbol{x}_t) &= \mathbb{E}[\max_d\{v_j(\boldsymbol{x}_t) + \varepsilon_{jt}\}] \\ &= \sum_{j\in\mathcal{D}} p(j|\boldsymbol{x}_t)\left[v_j(\boldsymbol{x}_t) + \varepsilon^*_{jt}(\boldsymbol{x}_t)\right] \\ &= \sum_{j\in\mathcal{D}} p(j|\boldsymbol{x}_t)\left[u_j(\boldsymbol{x}_t) + \beta\sum_{\boldsymbol{x}\in\mathcal{X}} V(\boldsymbol{x}) f(\boldsymbol{x}|\boldsymbol{x}_t, j) + \varepsilon^*_{jt}(\boldsymbol{x}_t)\right], \end{aligned}$$

where $\mathcal{X}$ denotes the set that includes the X possible values that $\boldsymbol{x}_{t+1}$ can take (we do not use subscript in $V(\cdot)$ to reflect the infinite-horizon stationary nature of the setting).

Infinite-horizon stationary settings (cont'd)

Now we can express each of the components of the previous equation in matrix form:

$$\boldsymbol{V} \equiv \begin{bmatrix} V(\boldsymbol{x}^{(1)}) \\ \vdots \\ V(\boldsymbol{x}^{(X)}) \end{bmatrix}, \qquad \boldsymbol{u}_j \equiv \begin{bmatrix} u_j(\boldsymbol{x}^{(1)}) \\ \vdots \\ u_j(\boldsymbol{x}^{(X)}) \end{bmatrix}, \qquad \boldsymbol{\varepsilon}^*_{jt} \equiv \begin{bmatrix} \varepsilon^*_{jt}(\boldsymbol{x}^{(1)}) \\ \vdots \\ \varepsilon^*_{jt}(\boldsymbol{x}^{(X)}) \end{bmatrix}$$

$$\boldsymbol{p}_j \equiv \begin{bmatrix} p(j|\boldsymbol{x}^{(1)}) \\ \vdots \\ p(j|\boldsymbol{x}^{(X)}) \end{bmatrix}, \qquad F_j \equiv \begin{bmatrix} f(\boldsymbol{x}^{(1)}|\boldsymbol{x}^{(1)}, j) & \dots & f(\boldsymbol{x}^{(X)}|\boldsymbol{x}^{(1)}, j) \\ \vdots & \ddots & \vdots \\ f(\boldsymbol{x}^{(1)}|\boldsymbol{x}^{(X)}, j) & \dots & f(\boldsymbol{x}^{(X)}|\boldsymbol{x}^{(X)}, j) \end{bmatrix},$$

so that the ex-ante value function reads as:

$$\boldsymbol{V} = \sum_{j \in \mathcal{D}} \boldsymbol{p}_j \circ [\boldsymbol{u}_j + \beta F_j \boldsymbol{V} + \boldsymbol{\varepsilon}^*_{jt}],$$

where $\circ$ denotes the Hadamard product (or element-by-element multiplication).

Infinite-horizon stationary settings (cont'd)

Solving for $\boldsymbol{V}$ in the above expression yields:

$$\boldsymbol{V} = \left(I_X - \beta \sum_{j \in \mathcal{D}} \boldsymbol{p}_j \iota_X' \circ F_j \right)^{-1} \left(\sum_{j \in \mathcal{D}} \boldsymbol{p}_j \circ [\boldsymbol{u}_j + \boldsymbol{\varepsilon}_{jt}^*] \right),$$

where I_X denotes the size X identity matrix and ι_X denotes the size X vector of ones.

This matrix expression is **operational**, in practice, in any standard matrix software, and avoids the solution of the DP in estimation.

Furthermore, it is useful in the **estimation of dynamic games**, as we will see in Chapter 7.

Estimation Methods

Estimation

Estimation of the models we have seen so far, in which ε_t is i.i.d. across time, is in **two-stages**:

1. Estimate **CCPs** and **transition functions** for the state variables.

2. Form the value functions using the CCPs estimated in the first stage and estimate the **structural parameters**.

CCP estimators of the structural parameters are $\sqrt{N}$**-consistent** and **asymtotically normal** under standard regularity conditions.

Approximation error of the CCPs can introduce **small sample bias** in the structural parameter estimates $\Rightarrow$ we discuss refinements below.

CCPs and transition functions

With **unlimited data**, both CCPs and transition functions can be estimated non-parametrically using simple bin estimators:

$$\hat{p}(d_t = d|\boldsymbol{x}_t = \boldsymbol{x}) = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{t=1}^{T} \mathbb{1}\{d_{it} = d\} \, \mathbb{1}\{\boldsymbol{x}_{it} = \boldsymbol{x}\}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{t=1}^{T} \mathbb{1}\{\boldsymbol{x}_{it} = \boldsymbol{x}\}}.$$

However, in reality **data limitations** and **continuous state variables** prevent us from doing it this way.

In this case, **smoothing** is needed, either through nonparametric kernels, basis functions, or flexibly specified logits or probits.

Estimating the structural parameters

Estimation under finite dependence, infinite-horizon stationary settings, or shortly-lived finite horizon settings is **straightforward**.

CCP representation together with **estimates of the CCPs** obtained in the first stage lead to simple expressions that can be included in **MLE** or **GMM**.

Maximum likelihood estimation in the **Rust example**:

$$\mathcal{L}_{\mathrm{N}} = \sum_{i=1}^{N} \sum_{t=1}^{T} d_{1it} \ln p_1(x_{it}) + (1 - d_{1it}) \ln[1 - p_1(x_{it})],$$

where:

$$p_1(x_{it}) = \frac{1}{1 + \exp\left\{\theta_R - \theta_M x_{it} + \beta\left(\ln \hat{p}_1(0) - \sum_{x \in X} \ln \hat{p}_1(x) \hat{F}^0_{x,x_t}\right)\right\}}.$$

Estimating the structural parameters (cont'd)

Alternatively, we can use a GMM (in this case simple regression) approach. Building on the inversion we did in the **Rust example**:

$$
\begin{aligned}
-\ln p_1(x_{it}) &= \ln\left[1+\exp(v_0(x_{it})-v_1(x_{it})\right] \\
&= \ln\left[1+\exp\left\{\theta_R-\theta_M x_{it}+\beta\left(\ln\hat{p}_1(0)-\sum_{x\in X}\ln\hat{p}_1(x)\hat{F}^0_{x,x_t}\right)\right\}\right].
\end{aligned}
$$

Rearranging the terms in the above expression yields:

$$
\ln\left(\frac{1-\hat{p}_1(x_{it})}{\hat{p}_1(x_{it})}\right)=\theta_R-\theta_M x_{it}+\beta\left(\ln\hat{p}_1(0)-\sum_{x\in X}\ln\hat{p}_1(x)\hat{F}^0_{x,x_t}\right).
$$

This is an **exact expression** with the population CCPs and transitions $\Rightarrow$ we introduce econometric error ($\hat{p}_1(\cdot)$ and $\hat{F}$ are estimated).

Orthogonality conditions $\Rightarrow$ GMM.

Llull (2018b): estimates of the CCPs obtained with a different dataset as an instrument (attenuation bias and correlated measurement error).

Forward simulation methods

Value functions can be **cumbersome** even with CCP representation.

Continuation value $\int V(\boldsymbol{x}_{t+1})dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j)$: integration over all possible **future states** of the world (paths) and all possible **choices** $\Rightarrow$ many paths for which we need to compute $u_j(\cdot) + \psi_j(\cdot)$.

$\Rightarrow$ Hotz, Miller, Sanders, and Smith (1994): **forward simulation**, based on the transition functions and, potentially, the CCPs.

Main idea: only compute them for a **finite number of draws** of these future paths for each individual. Then, when **averaging across individuals**, we obtain an approximation of the future expectation.

Two versions: either simulate choices and states or only the evolution of states for a given path of choices.

Aguirregabiria and Mira

Aguirregabiria and Mira (2002) propose a **nested pseudo-likelihood** algorithm that links both.

Inner loop: Hotz and Miller, starting from consistent estimates of CCPs, and then using as input CCPs from the outer loop.

Outer loop: With the parameter estimates from inner loop, compute CCPs, and iterate over them till convergence.

Repeat K times or until convergence in $\boldsymbol{\theta}$ and $\boldsymbol{p}$.

$K = 0 \Rightarrow$ Hotz-Miller.
$K \to \infty \Rightarrow$ Rust's NFXP.

Aguirregabiria and Mira (cont'd)

Aguirregabiria and Mira NPL algorithm is a swapping of the nested fixed point algorithm.

$$\boldsymbol{V} = \sum_{j \in \mathcal{D}} \boldsymbol{p}_j(\boldsymbol{V}) \circ [\boldsymbol{u}_j + \beta F_j \boldsymbol{V} + \boldsymbol{\varepsilon}^*_{jt}].$$

Rust's NFXP $\Rightarrow$ inner loop finds V as a fixed point, outer loop updates the parameters.

The CCPs are mappings of the value function:

$$\boldsymbol{p} = \Lambda(\boldsymbol{V}).$$

Rewriting $\boldsymbol{V}$ as:

$$\boldsymbol{V}(\boldsymbol{p}) = \left(I_X - \beta \sum_{j \in \mathcal{D}} \boldsymbol{p}_j \iota'_X \circ F_j \right)^{-1} \left(\sum_{j \in \mathcal{D}} \boldsymbol{p}_j \circ [\boldsymbol{u}_j + \boldsymbol{\varepsilon}^*_{jt}] \right).$$

Therefore, $\boldsymbol{p} = \Lambda(\boldsymbol{V}(\boldsymbol{p})) \equiv \Psi(\boldsymbol{p})$.**Aguirregabiria-Mira NPL**$\Rightarrow$ fixed point (in CCPs) in the outside loop and the update of the parameters in the inner loop.

Results in the Rust example

Table: Second Stage Estimation: Cost Function Parameters

Method	Parameter	Group 1, 2, 3	Group 4	Group 1, 2, 3, 4
NFXP	θ_R	11.87	10.12	9.75
		(1.95)	(1.36)	(0.89)
	θ_M	5.02	1.18	1.37
		(1.40)	(0.28)	(0.24)
CCP	θ_R	11.76	10.21	9.62
		(0.91)	(0.71)	(0.45)
	θ_M	-4.99	1.16	-3.95
		(22.49)	(13.48)	(9.35)
NPL	θ_R	11.72	10.19	9.59
		(0.91)	(0.71)	(0.45)
	θ_M	-2.20	1.43	-2.11
		(22.48)	(13.46)	(9.35)

Courtesy of José García-Louzao, Sergi Marin Arànega, Alex Tagliabracci, and Alessandro Ruggieri, who replicated Rust's paper for the replication exercise in the Microeconometrics IDEA PhD course in Fall 2014.

Performance (Monte-Carlo simulation)

N.obs.	Algorithm	θ_R		θ_M		Time (sec.)
	DGP	9.74		2.69		
$N = 100$	NFXP	10.07	(0.90)	2.83	(0.39)	157.56
	CCP	10.48	(0.40)	-2.17	(6.91)	0.56
$N = 1,000$	NFXP	9.78	(0.28)	2.67	(0.12)	160.57
	CCP	10.04	(0.14)	2.19	(2.55)	47.84
$N = 5,000$	NFXP	9.93	(0.11)	2.79	(0.05)	400.31
	CCP	10.09	(0.05)	2.92	(0.93)	59.11
$N = 10,000$	NFXP	9.85	(0.08)	2.73	(0.04)	1,030.70
	CCP	9.99	(0.04)	2.64	(0.66)	121.38
$N = 25,000$	NFXP	9.78	(0.05)	2.67	(0.02)	1,070.20
	CCP	9.92	(0.02)	2.74	(0.41)	133.38

Courtesy of José García-Louzao, Sergi Marin Arànega, Alex Tagliabracci, and Alessandro Ruggieri, who replicated Rust's paper for the replication exercise in the Microeconometrics IDEA PhD course in Fall 2014. Simulations executed with a MacBookPro, 2.5GHz Intel Core i5, 8GB RAM (50 replications, asymptotic standard errors from the first iteration are in parentheses).

Extensions: unobserved heterogeneity and competitive equilibrium models

Unobserved heterogeneity

Standard approach to introduce unobserved heterogeneity: **mixture distributions** (Heckman and Singer, 1984).

Given our assumptions, the log-likelihood can be written as:

$$\mathcal{L}_{\mathrm{N}}(\boldsymbol{\theta}, \boldsymbol{\pi}) = \sum_{\mathrm{i}=1}^{\mathrm{N}} \ln \left[\sum_{k=1}^{K} \pi_{k|\boldsymbol{x}_{i1}} \prod_{t=1}^{T} f_t(d_{it}, \boldsymbol{x}_{it+1} | \boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta}) \right].$$

As discussed in previous chapter, this log-likelihood function is **no longer additively separable**, and we cannot directly estimate it in two stages as before.

The EM Algorithm

The FOC that stems from maximizing the above likelihood is:

$$0 = \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\sum_{k=1}^{K} \pi_{k|\boldsymbol{x}_{i1}} \sum_{t=1}^{T} \left[\frac{\partial f_t(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \prod_{t' \neq t} f_{t'}(d_{it'}, \boldsymbol{x}_{it'+1}|\boldsymbol{x}_{it'}, \boldsymbol{\omega}^k; \boldsymbol{\theta}) \right]}{\sum_{k=1}^{K} \pi_{k|\boldsymbol{x}_{i1}} \prod_{t=1}^{T} f(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})}$$

$$= \sum_{\mathrm{i}=1}^{\mathrm{N}} \frac{\sum_{k=1}^{K} \pi_{k|\boldsymbol{x}_{i1}} \sum_{t=1}^{T} \left[\frac{\partial \ln f_t(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \prod_{t'=1}^{T} f_{t'}(d_{it'}, \boldsymbol{x}_{it'+1}|\boldsymbol{x}_{it'}, \boldsymbol{\omega}^k; \boldsymbol{\theta}) \right]}{\sum_{k=1}^{K} \pi_{k|\boldsymbol{x}_{i1}} \prod_{t=1}^{T} f(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})},$$

where the second expression is obtained by replacing $\partial f_t(\cdot)/\partial \boldsymbol{\theta} = f_t(\cdot)\partial \ln f_t(\cdot)/\partial \boldsymbol{\theta}$.

This FOC is easily reinterpreted using the Bayes' Rule because:

$$\frac{\pi_{k|\boldsymbol{x}_{i1}} \prod_{t'=1}^{T} f_{t'}(d_{it'}, \boldsymbol{x}_{it'+1}|\boldsymbol{x}_{it'}, \boldsymbol{\omega}^k; \boldsymbol{\theta})}{\sum_{k=1}^{K} \pi_{k|\boldsymbol{x}_{i1}} \prod_{t=1}^{T} f(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})} \equiv \varpi(k|\boldsymbol{d}_i, X_i; \boldsymbol{\theta}, \boldsymbol{\pi}),$$

where $\boldsymbol{\pi}$ is the vector including all the type probabilities $\pi_{k|\boldsymbol{x}_{i1}}$ for all k and i, $\boldsymbol{d}_i \equiv (d_{i1}, ..., d_{iT})'$ and $X_i \equiv (\boldsymbol{x}_{i1}, ..., \boldsymbol{x}_{iT})'$, and:

$$0 = \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{k=1}^{K} \sum_{t=1}^{T} \varpi(k|\boldsymbol{d}_i, X_i; \boldsymbol{\theta}, \boldsymbol{\pi}) \frac{\partial \ln f_t(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})}{\partial \boldsymbol{\theta}}.$$

The EM Algorithm (cont'd)

Dempster, Laird, and Rubin (1977) note that the following maximization problem delivers exactly **the same first order conditions**:

$$\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{t=1}^{T} \sum_{k=1}^{K} \varpi(k|\boldsymbol{d}_i, X_i; \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\pi}}) \ln f_t(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta}),$$

for $\hat{\boldsymbol{\pi}}$ that satisfies:

$$\hat{\pi}_{k|\boldsymbol{x_1}} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \varpi(k|\boldsymbol{d}_i, X_i; \hat{\boldsymbol{\theta}}, \hat{\boldsymbol{\pi}})\, \mathbb{1}\{\boldsymbol{x}_{i1} = \boldsymbol{x}_1\}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathbb{1}\{\boldsymbol{x}_{i1} = \boldsymbol{x}_1\}}.$$

The EM Algorithm (cont'd)

The algorithm is as follows:

- Guess $\boldsymbol{\theta}^{(m)}$ and $\boldsymbol{\pi}^{(m)}$.
- Expectation (E) step:

$$\varpi(k|\boldsymbol{d}_i, X_i; \boldsymbol{\theta}^{(m)}, \boldsymbol{\pi}^{(m)}) = \frac{\pi^{(m)}_{k|\boldsymbol{x}_{i1}} \prod_{t'=1}^{T} f_{t'}(d_{it'}, \boldsymbol{x}_{it'+1}|\boldsymbol{x}_{it'}, \boldsymbol{\omega}^k; \boldsymbol{\theta}^{(m)})}{\sum_{k=1}^{K} \pi^{(m)}_{k|\boldsymbol{x}_{i1}} \prod_{t=1}^{T} f(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta}^{(m)})}$$

$$\pi^{(m+1)}_{k|\boldsymbol{x}_1} = \frac{\sum_{\mathrm{i}=1}^{\mathrm{N}} \varpi(k|\boldsymbol{d}_i, X_i; \boldsymbol{\theta}^{(m)}, \boldsymbol{\pi}^{(m)})\, \mathbb{1}\{\boldsymbol{x}_{i1} = \boldsymbol{x}_1\}}{\sum_{\mathrm{i}=1}^{\mathrm{N}} \mathbb{1}\{\boldsymbol{x}_{i1} = \boldsymbol{x}_1\}}$$

- Maximization (M) step:

$$\boldsymbol{\theta}^{(m+1)} = \arg\max_{\boldsymbol{\theta}} \sum_{\mathrm{i}=1}^{\mathrm{N}} \sum_{t=1}^{T} \sum_{k=1}^{K} \varpi(k|\boldsymbol{d}_i, X_i; \boldsymbol{\theta}^{(m)}, \boldsymbol{\pi}^{(m)}) \ln f_t(d_{it}, \boldsymbol{x}_{it+1}|\boldsymbol{x}_{it}, \boldsymbol{\omega}^k; \boldsymbol{\theta})$$

The M-step of the algorithm reintroduces **additive separability**.

Arcidiacono and Jones (2003): updating of transition functions and utility parameters can be done separately.

Arcidiacono and Miller (2011)

CCP estimation with unobserved heterogeneity requires that CCPs and transitions are estimated **conditional on the unobserved type**.

Arcidiacono and Miller (2011) propose an implementation of the EM algorithm.

Arcidiacono and Miller's approach is also well suited for **time-varying** persistent unobserved heterogeneity.

Using again the Bayes' Rule (and the law of iterated expectations), the CCPs can be expressed as:

$$\begin{aligned}\Pr(j|\boldsymbol{x}_{it},k) = \frac{\Pr(j,k|\boldsymbol{x}_{it})}{\Pr(k|\boldsymbol{x}_{it})} &= \frac{\mathbb{E}[\mathbb{1}\{d_{it}=j\}\,\mathbb{1}\{k_{it}=k\}|\boldsymbol{x}_{it}]}{\mathbb{E}[\mathbb{1}\{k_{it}=k\}|\boldsymbol{x}_{it}]} \\ &= \frac{\mathbb{E}[\mathbb{1}\{d_{it}=j\}\varpi(k|\boldsymbol{d}_i,X_i;\boldsymbol{\theta},\boldsymbol{\pi})|\boldsymbol{x}_{it}]}{\mathbb{E}[\varpi(k|\boldsymbol{d}_i,X_i;\boldsymbol{\theta},\boldsymbol{\pi})|\boldsymbol{x}_{it}]}.\end{aligned}$$

Arcidiacono and Miller (2011) expands the E step of the EM algorithm to also **update the CCPs** as the sample analog of the above expression using the updated $\varpi(k|\boldsymbol{d}_i,X_i;\boldsymbol{\theta}^{(m)},\boldsymbol{\pi}^{(m)})$.

Competitive equilibrium and aggregate shocks

Extension of CCP estimation to competitive equilibrium is straightforward in a world with **no aggregate shocks**.

Individuals are price-takers $\Rightarrow$ treated as **nuisance** (non-fundamental) **parameters** of the model (estimation often enriched with aggregate data or individual pay-off data, like wages).

Aggregate shocks: Altuğ and Miller (1998) and Llull (2018b).

Key difficulty: integrate over future counterfactual paths for the aggregate conditions, for which CCPs cannot be (directly) recovered from the data.

$$v_{jt}(\boldsymbol{x}_t, r_t) = u_{jt}(\boldsymbol{x}_t, r_t) + \beta \int \int V_{t+1}(\boldsymbol{x}_{t+1}, r_{t+1}) dF_x(\boldsymbol{x}_{t+1}|\boldsymbol{x}_t, j) dF_r(r_{t+1}|r_t, \Omega_t).$$

where r_t is the aggregate variable and Ω_t is the relevant information at time t that is useful to predict r_{t+1}.

Solution: infer counterfactuals from choices observed under similar circumstances (see Altuğ and Miller, 1998, and Llull, 2020).

Application: Llull (2020)

Overview

Workers decide every year Blue Collar, White Collar, STEM, School, or Home.

Representative firm produces a single output combining labor skill units from each occupation with structures and equipment.

Capital (for today) and immigration processes are **specified outside of the model** —yet, allowed to be endogenous to aggregate conditions.

Knowledge spillovers: Intellectual Property Products (IPP) or patents, generated a by-product of using STEM labor and capital equipment in production $\Rightarrow$ externality.

Main trade-off for policy making: **competition** effects vs **spillovers**.

Firm (I): Knowledge Production

- **Innovation** (ΔI_t) is generated as a by-product of using equipment capital K_{Et} and STEM labor, measured in skill units S_{Tt}, in general production:

$$\Delta I_t = \xi_t K_{Et}^{\chi_1} S_{Tt}^{\chi_2}.$$

- **No restriction** on χ_1 and χ_2 (constant, decreasing, or increasing returns to scale).

- **Knowledge shock**:

$$\Delta \ln \xi_{t+1} = \pi_\xi + \sigma_\xi \upsilon_{\xi t} \text{ with } \upsilon_{\xi t} \sim \mathcal{N}(0,1).$$

Firm (II): Production Function

- The **production function** is:

$$Y_t = (\zeta_t I_t^{\varphi}) K_{St}^{\varsigma_t} \left\{ \alpha_t S_{Bt}^{\rho} + (1-\alpha_t) \left[\theta_t S_{Wt}^{\kappa} + (1-\theta_t) \left(\iota_t S_{Tt}^{\psi} + (1-\iota_t) K_{Et}^{\psi} \right)^{\frac{\kappa}{\psi}} \right]^{\frac{\rho}{\kappa}} \right\}^{\frac{1-\varsigma_t}{\rho}},$$

  where:

$$o_t \equiv \frac{\exp(\tilde{o}_0 + \tilde{o}_1 I_t)}{1 + \exp(\tilde{o}_0 + \tilde{o}_1 I_t)}, \qquad \text{for } o \in \{\varsigma, \alpha, \theta, \iota\}.$$

- **Spillovers**: both factor neutral/TFP-type (if $\varphi > 0$), and skill-biased (if $\tilde{o}_1 \neq 0$ for some $o \in \{\varsigma, \alpha, \theta, \iota\}$).

- **TFP shock**:

$$\Delta \ln \zeta_{t+1} = \pi_\zeta + \sigma_\zeta \upsilon_{\zeta t} \text{ with } \upsilon_{\zeta t} \sim \mathcal{N}(0,1).$$

Firm (III): Profit Maximization

- Firms **maximize profits**:

$$\max_{S_{Tt}, S_{Wt}, S_{B_t}, K_{Et}, K_{St}} \left\{ \begin{array}{r} Y(\zeta_t, I_t, S_{Tt}, S_{Wt}, S_{B_t}, K_{Et}, K_{St}) - (r_{Kt} + \delta_E) K_{Et} \\ -(r_{Kt} + \delta_S) K_{St} - r_{Tt} S_{Tt} - r_{Wt} S_{Wt} - r_{Bt} S_{Bt} \end{array} \right\},$$

- Input **demands** are given by the first order conditions on this maximization problem.

- **Externality**: firms take I_t as given, and do not internalize it when setting wages or interest rates.

Workers (I): Life Cycle

Life cycle:

- Let $a \in \{16, ..., 65\}$ denote age.
- **Natives**: born at age $a = 16$, die at age $a = 65$ with certainty, and make yearly decisions.
- **Immigrants**: enter at age $a = \tilde{a}$, die at age $a = 65$, make yearly decisions, and there is no return migration.

Workers (II): Choice Set

Choice sets:

- Three to five mutually exclusive choices: **blue collar** ($d_a = B$), **white collar** ($d_a = W$), **STEM** ($d_a = T$), **school** ($d_a = S$), and **home** ($d_a = H$).

- **Dropping out from school** is an absorbing state, so S not available if $d_{a-1} \neq S$.

- **STEM** occupation has a **college degree** requirement, so T is not available if $E_a < 16$.

- Thus, define the state-dependent choice set as $\mathcal{D}(h_a)$, and define the indicator variables $d_{ja} = \mathbb{1}\{d_a = j\}$.

Workers (III): State Variables I

Observable Idiosyncratic State Variables:

- State vector $h_a = (a, \ell, E_a, d_{a-1}, n_a, \tilde{a})'$.
- **Age** a: evolves deterministically.
- **Type** $\ell \in \mathcal{L}$: observable. Gender (2) $\times$ race for natives (3) and country of origin for immigrants (7) $\Rightarrow$ 20 types.
- **Education** E_a: endogenous, number of years of education. Initial education E_{16} or $E_{\tilde{a}}$ is exogenously given.
- **Preschool children** $n_a \in \mathcal{C} \equiv \{0, 1, 2+\}$: evolves stochastically with an exogenous process that depends on type, education level, and age.
- **Age at entry** $\tilde{a}$: exogenously given.
- **Previous period choice** d_{a-1}: transition costs.

Workers (IV): State Variables II

Aggregate State Variables:

- **Skill prices** $r_t \equiv (r_{Bt}, r_{Wt}, r_{Tt})'$: determined in equilibrium.
- **Information to predict** their future values ϖ_t: see below.

Idiosyncratic shocks:

- **Productivity shock** η_a: it affects the productivity in all occupations. I.i.d over time with $\eta_a | h_a \sim \mathcal{N}(0,1)$.
- **Taste shock** ϵ_a: defined so that:

$$\varepsilon_a \equiv (\sigma_{B\ell}\eta_a + \epsilon_{Ba}, \sigma_{W\ell}\eta_a + \epsilon_{Wa}, \sigma_{T\ell}\eta_a + \epsilon_{Ta}, \epsilon_{Sa}, \epsilon_{Ha})',$$

where $\sigma_{j\ell}$ is defined below, and:

$$F_\varepsilon(\varepsilon_a) \equiv \exp\left\{-\left[\left(e^{-\varepsilon_{Ba}/\varrho} + e^{-\varepsilon_{Wa}/\varrho} + e^{-\varepsilon_{Ta}/\varrho}\right)^\varrho + e^{-\varepsilon_{Sa}} + e^{-\varepsilon_{Ha}}\right]\right\},$$

with $\varrho \equiv \sqrt{1 - \text{Corr}(\varepsilon_{ja}, \varepsilon_{ka})}$ for any $j, k \in \{B, W, T\}$.

Workers (V): Wages

Wages:

- **Skill units** times **market price**.
- Let $s_j(h_a, \eta_a)$ denote **skill units**.
- Then: $w_j(h_a, \eta_a, r_t) \equiv r_{jt} s_j(h_a, \eta_a)$.
- Skill units are of the form: $s_j(h_a, \eta_a) \equiv \exp(\tilde{s}_j(h_a) + \sigma_{j\ell}\eta_a)$.
- **Exclusion restriction**: number of children not in $\tilde{s}_j(\cdot)$.

Workers (VI): Flow Payoffs

Alternative-specific period utility functions:

- **Log utility** function plus additive taste shocks (consumption in non-working alternatives embedded in parameters).

- **Working in occupation** $j \in \{B, W, T\}$:

$$u_j(h_a, \varepsilon_a, r_t) \equiv \ln w_j(h_a, \eta_a, r_t) + \epsilon_{ja} = \ln r_{jt} + \tilde{s}_j(h_a) + \varepsilon_{ja}.$$

- **School**:

$$u_S(h_a, \varepsilon_a, r_t) \equiv \tau_0(\ell)\,\mathbb{1}\{E_a < 12\} + \tau_1(\ell)\,\mathbb{1}\{12 \leq E_a < 16)\} \\ +\tau_2(\ell)\,\mathbb{1}\{E_a \geq 16\} + \varepsilon_{Sa}.$$

- **Home**:

$$u_H(h_a, \varepsilon_a, r_t) \equiv \vartheta(\ell, n_a) + \varepsilon_{Ha}.$$

- **Stationarity**, except for skill prices.

Workers (VII): Intertemporal Decisions

Value functions:

- **Lifetime discounted utility**:

$$\mathbb{E}_t\left[\sum_{l=0}^{65-a}\beta^l\left(\sum_{j\in\mathcal{D}(h_{a+l})} d_{ja+l}[\tilde{u}_j(h_{a+l},r_{t+l})+\varepsilon_{ja+l}]\right)\right].$$

- **Bellman equation**:

$$V(h_a,\varepsilon_a,r_t,\varpi_t)=$$
$$\max_{\{d_{ja}\}_{j\in\mathcal{D}(h_a)}}\textstyle\sum_{j\in\mathcal{D}(h_a)} d_{ja}\left\{\tilde{u}_j(h_a,r_t)+\varepsilon_{ja}+\beta\,\mathbb{E}_t[V(h_{a+1},\varepsilon_{a+1},r_{t+1},\varpi_{t+1})]\right\}.$$

- **Ex-ante value function**:

$$\bar{V}(h_a,r_t,\varpi_t)\equiv\int V(h_a,\varepsilon,r_t,\varpi_t)dF_\varepsilon(\varepsilon).$$

- **Alternative-specific value function**:

$$v_j(h_a,r_t,\varpi_t)\equiv\tilde{u}_j(h_a,r_t)+\beta\int\sum_{h\in\mathcal{H}_{a+1|h_a}}\bar{V}(h,r,\varpi)P_h(h|h_a,j)dF_r(r,\varpi|\varpi_t,r_t).$$

Equilibrium

- **Labor supply**: aggregation of skill units of individuals working in j:

$$S_{jt}^{(S)}(r_t) = \iint \sum_{h \in \mathcal{H}} d_{jt}^*(h, \epsilon + \eta, r_t, \varpi_t) s_j(h, \eta) P_h(h) dF_\epsilon(\epsilon) d\Phi(\eta).$$

- **Labor demand**: first order conditions on firm's problem.

- **Equilibrium**: market clearing conditions:

$$S_{jt}^{(S)}(r_t^*) = S_{jt}^{(D)}(r_t^*) \equiv S_{jt}^*.$$

Expectations

- Individuals use $F_\varepsilon(\varepsilon_a)$, $P_h(h_{a+1}|h_a, d_a)$, and $F_r(r_{t+1}|r_t, \varpi_t)$ to **form expectations** about future state variables.

- All others have been defined above, $F_r(r_{t+1}|r_t, \varpi_t)$ is defined here.

- **Rational expectations**: ϖ_t should include the whole distribution of state variables (intractable).

- **Alternative**: similar to Altuğ & Miller (REStud 1998), also related to Krusell & Smith (JPE 1998), Lee & Wolpin (Ecma 2006, JE 2010), and Llull (REStud 2017):

$$\ln r_{jt+1} = \ln r_{jt} + \Xi_{0j} + \Xi_{1j}\sigma_\zeta \upsilon_{\zeta t+1} + \Xi_{2j}\sigma_\xi \upsilon_{\xi t+1} + \sigma_{\Upsilon j}\Upsilon_{jt+1}, \quad j \in \{B, W, T\},$$

  where $\Upsilon_{jt+1} \sim \mathcal{N}(0,1)$ is a **prediction error** uncorrelated across occupations, and independent of all variables.

- Thus, ϖ_t is **redundant** given r_t, and omitted hereinafter.

Preliminaries I

- **Micro data**: linked March CPS for 1993–2015 and SIPP 1988–2007 (one year panels).

- **Aggregate data**:
    - For estimation and simulation: IPP, structures, equipment, cohort sizes, output, and fertility process.
    - For simulation only: distributions of initial characteristics of natives and immigrants

- **Parameters assumed to be known**:

$$\beta = 0.95, \quad \delta_E = 11.93\%, \quad \delta_S = 2.88\%$$

Preliminaries II

Parameters to be estimated:

- **Wages**: skill unit production function $\{\tilde{s}_j(h_a)\}_{j\in\{B,W,T\}}$, and conditional wage variances $\{\sigma_{j\ell}\}_{j\in\{B,W,T\}}^{\ell\in\mathcal{L}}$.

- **Utility**: GEV correlation parameter ϱ, schooling utility $\{\tau_k(\ell)\}$ for $k \in \{0,1,2\}$, home utility $\vartheta(\ell, n_a)$.

- **Production function**: IPP capital generation function χ_1 and χ_2, production function φ, $\tilde{\varsigma}_0$, $\tilde{\varsigma}_1$, $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, $\tilde{\theta}_0$, $\tilde{\theta}_1$, $\tilde{\iota}_0$, $\tilde{\iota}_1$, ρ, κ, and ψ, aggregate shock processes π_ξ, π_ζ, σ_ξ, and σ_ζ.

- **Expectations**: reduced form parameters $\{\Xi_{0j}, \Xi_{1j}, \Xi_{2j}, \sigma_{\Upsilon j}\}$ for $j \in \{B, W, T\}$ (which are not fundamentals of the model but part of the solution).

Identification and estimation

CCP estimation:
break the estimation in pieces and
avoid solving for value functions or equilibrium
(no extra assumptions)

Identification and estimation

CCP estimation

$\Downarrow$

Estimate the CCPs $\tilde{p}(h_a, t)$ nonparametrically $\Rightarrow$ Estimate the wage regression

$\Downarrow$ $\quad \{\tilde{s}_j(h_a)\}_{j\in\{B,W,T\}}$, $\{\sigma_{j\ell}\}_{j\in\{B,W,T\}}^{\ell\in\mathcal{L}}$

Estimate the production function parameters $\Leftarrow$ $r_t,\ S_{jt}$

$\chi_1,\ \chi_2,$
$\varphi,\ \{\tilde{\varsigma}_k, \tilde{\alpha}_k, \tilde{\theta}_k, \tilde{\iota}_k\}_{k\in\{0,1\}},$
$\rho,\ \kappa,\ \psi,$
$\pi_\xi,\ \pi_\zeta,\ \sigma_\xi,\ \sigma_\zeta$

$\xi_t,\ \zeta_t$ $\Rightarrow$ "Estimate" the expectation rule for the baseline economy $\Rightarrow$ $\left\{\begin{array}{c}\Xi_{0j}\\ \Xi_{1j}\\ \Xi_{2j}\\ \sigma_{\Upsilon j}\end{array}\right\}_{j\in\{B,W,T\}}$

Estimate $p(h_a, r_t)$ nonparametrically $\Rightarrow$ Estimate school utility

Estimate $p(h_a, r_t)$ nonparametrically $\Rightarrow$ Estimate home utility

Estimate school utility $\Downarrow$ $\color{green}{\tau_0(\ell), \tau_1(\ell), \tau_2(\ell)}$

Estimate home utility $\Downarrow$ $\vartheta(\ell, n_a)$ and ϱ

Summary of Estimated Worker Parameters

- Similarity of results across estimations:
  - **with and without IV correction** for measurement error (except in the selection correction term — the instrumented variable).
  - **with CPS and SIPP**.
- Wages:
  - **Returns to education**: 9.6% for STEM (plus college requirement), 10.7% for white collar, 6% for blue collar.
  - **Lower returns for immigrants**: reduction of 3.1%, 2.7%, and 3.1% in STEM, white collar, and blue collar respectively.
  - **Experience**: hump shape, similar across occupations.
  - **Transition costs**: significant, increasing in education, slightly convex in experience.
  - **Potential experience abroad**: less productive than domestic experience.

Summary of Estimated Worker Parameters

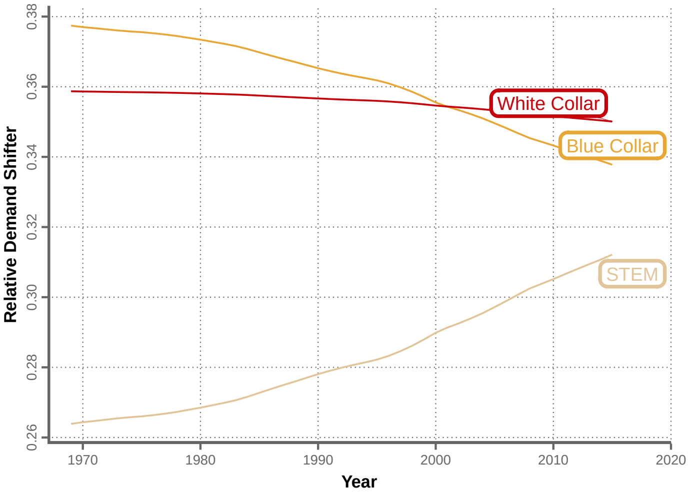
- **Types**: gender and origin/race gaps in line with expected.
- **Variances**: similar across groups, and somewhat larger for white collar.
- **GEV parameter**: correspond to a 91% correlation between the shocks across alternatives.

- Utilities:
  - **School**: College and especially graduate education is more costly than high school.
  - **Home**: Female have extra positive utility of being home from children, male have it negative.
  - **Types**: gender and origin/race gaps in line with expectations
- Precise estimates.

Summary of Production Function Parameters

- Production function:
  - **Elasticities of substitution**: equipment capital and STEM are relative complements.
  - **Aggregate shocks**: small negative TFP drift, smaller and positive IPP drift, variances of about 1.3-1.4%.
  - **Expectations**: biggest role on TFP shock, no drift, variance of the prediction error of about 2%.
- Externality:
  - **Precision**: precision should be increased (refinements).
  - **IPP production function**: constant or slightly increasing returns to scale, with a role for STEM (smaller than the role for equipment, though).
  - **TFP expernality**: estimated to be around 0.4.
  - **Skill-biased externality**: see next slide.

Figure: Relative Demand Shifters for Each Labor Input



Note: The figure represents three combinations of the estimated values of ς_t, α_t, θ_t, and ι_t that are associated to the relative demand for each of the indicated labor inputs. The statistic associated to blue collar labor is $\frac{\alpha_t}{1-(1-\alpha_t)(1-\theta_t)(1-\iota_t)}$, the one associated to white collar is $\frac{(1-\alpha_t)\theta_t}{1-(1-\alpha_t)(1-\theta_t)(1-\iota_t)}$, and the one associated to STEM is $\frac{(1-\alpha_t)(1-\theta_t)\iota_t}{1-(1-\alpha_t)(1-\theta_t)(1-\iota_t)}$.