

# 47-813 Econometrics III: Assignment 1 Tutorial

Jaepil Lee

Carnegie Mellon University

January 26, 2022

# Assignment 1

- Assumptions and model

- Derivation of  $v_0(x_{it}) - v_1(x_{it}) = \sigma \log \frac{p_0(x_{it})}{p_1(x_{it})}$ .

- Understanding the 'estimating equation':

$$\sigma \log \frac{p_0(x_{it})}{p_1(x_{it})} = \beta \mathbb{E}_x \left[ -\sigma \log p_1^{(1)}(x_{0it}) - c^{(1)}(x_{0it}) \right] + c(x_{it})$$

- Understanding the assumptions for identification of monetary costs and non-monetary costs of closing a bank by FDIC.
  - Understanding how the authors are estimating transitions in the paper (and why).

- Alternative Conditional Choice Probabilities

- Including a time trend.
  - Omitting the effect of the House of Representatives (denoted as House in the paper) and/or Senate.
  - Adding squared terms on time, House and/or Senate.

# What I will cover today . . .

- Hotz and Miller's inversion theorem and the implicit function theorem
- A review of dynamic discrete choice model and conditional choice probability estimation
  - Arcidiacono and Ellickson, "Practical methods for estimation of dynamic discrete choice models." (2011) is a good reference.
- A quick walk-through for Kang et al. (2015).
- Installing COMPECON Matlab toolbox
- My Replication vs K LW with time trend
- Some tips for assignment

## Hotz and Miller (1993): Inversion Theorem

- Wikipedia's page on [implicit function theorem](#) illustrates the idea of the Inversion Theorem in Hotz and Miller (1993).
- Suppose we have an  $m$ -dimensional space, parametrized by a set of coordinates  $(x_1, \dots, x_m)$ .
- We can introduce a new coordinate system  $(x'_1, \dots, x'_m)$  by supplying  $m$  functions  $h_1 \dots h_m$  each being continuously differentiable:

$$\begin{bmatrix} x'_1 \\ x'_2 \\ \vdots \\ x'_{m-1} \\ x'_m \end{bmatrix} = \begin{bmatrix} h_1(x_1, \dots, x_m) \\ h_2(x_1, \dots, x_m) \\ \vdots \\ h_{m-1}(x_1, \dots, x_m) \\ h_m(x_1, \dots, x_m) \end{bmatrix}$$

- The set of functions  $h_1, \dots, h_m$  allow us to find the new coordinates  $(x'_1, \dots, x'_m)$ , given the point's old coordinates  $(x_1, \dots, x_m)$ .
- We want to verify if the opposite is possible: given coordinates  $(x'_1, \dots, x'_m)$ , can we 'go back' and compute the original coordinates  $(x_1, \dots, x_m)$ ?

## Hotz and Miller (1993): Inversion Theorem

- The implicit function theorem provides an answer to this question.
- The new and old coordinates  $(x'_1, \dots, x'_m, x_1, \dots, x_m)$  are related by  $f = 0$ :

$$f(x'_1, \dots, x'_m, x_1, \dots, x_m) = (h_1(x_1, \dots, x_m) - x'_1, \dots, h_m(x_1, \dots, x_m) - x'_m)$$

- Now the Jacobian matrix of  $f$  at a certain point  $(a, b)$  where  $a = (x'_1, \dots, x'_m)$ ,  $b = (x_1, \dots, x_m)$  is given by

$$(Df)(a, b) = \left[ \begin{array}{ccc|ccc} -1 & \cdots & 0 & \frac{\partial h_1}{\partial x_1}(b) & \cdots & \frac{\partial h_1}{\partial x_m}(b) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -1 & \frac{\partial h_m}{\partial x_1}(b) & \cdots & \frac{\partial h_m}{\partial x_m}(b) \end{array} \right] = [-I_m \mid J]$$

where  $I_m$  is the  $m \times m$  identity matrix and  $J$  is the  $m \times m$  matrix of partial derivatives evaluated at  $(a, b)$  (we don't see  $a$  in the Jacobian in this particular application of the implicit function theorem).

- The implicit function theorem states that we can locally express  $(x_1, \dots, x_m)$  as a function of  $(x'_1, \dots, x'_m)$  if  $J$  is invertible,  $|J| \neq 0$ .

## Hotz and Miller (1993): Inversion Theorem

- Hotz and Miller (1993) applies this idea to the relationship between the CCPs and the conditional value functions for each  $x$  and  $t$ .
- They consider a  $(J - 1)$ -dimensional vector of CCPs,  $p := (p_1, p_2, \dots, p_{J-1})$ . This is because there is linear dependence in the vector of all CCPs:  
$$\sum_{j=1}^J p_j = 1.$$
- The corresponding vector of differences of conditional value functions is  $\Delta v := (\Delta v_1, \dots, \Delta v_{J-1})$  where  $\Delta v_j = v_j - v_J$  ( $J$  is the normalizing choice here).
- Us econometricians observe the  $(J - 1)$ -dimensional CCPs as a coordinate in a  $(J - 1)$ -simplex for each  $x$  and  $t$  from the data. We also know that the following relation holds for  $j = 1, \dots, J - 1$ :

$$p_{jt}(x) = \int_{-\infty}^{\infty} G_{jt} \left( \epsilon_j + \Delta v_{jt}(x) - \Delta v_{1t}(x), \dots, \epsilon_j, \dots, \epsilon_j + \Delta v_{jt}(x) \middle| x \right) d\epsilon_j$$

where  $G$  is the cdf of  $(\epsilon_1, \dots, \epsilon_J)$  and  $G_j$  is its partial derivative with respect to the  $j$ -th argument.

## Hotz and Miller (1993): Inversion Theorem

- Hotz and Miller were wondering, given the CCPs  $p \in \Delta^{J-1}$ , can we 'go back' and find the original coordinates  $\Delta v = (\Delta v_1, \Delta v_2, \dots, \Delta v_{J-1}) \in \mathbb{R}^{J-1}$ ?
- Appealing to the implicit function theorem, we compute:

$$(Df)(\Delta v, p) = \left[ \begin{array}{ccc|ccc} -1 & \cdots & 0 & \frac{\partial Q_1}{\partial v_1}(p) & \cdots & \frac{\partial Q_1}{\partial v_{J-1}}(p) \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -1 & \frac{\partial Q_{J-1}}{\partial v_1}(p) & \cdots & \frac{\partial Q_m}{\partial v_{J-1}}(p) \end{array} \right] = [-I_{J-1} \mid \mathcal{J}]$$

- If the determinant of the right half of the Jacobian,  $\mathcal{J}$ , is invertible, yes, we can invert  $(p_1, p_2, \dots, p_{J-1})$  to  $(v_1 - v_J, v_2 - v_J, \dots, v_{J-1} - v_J)$ .
- By the virtue of  $G$  being continuously differentiable (or, "well-defined", they say), Hotz and Miller prove that  $\mathcal{J}$ , is invertible.
- ...and this is how they found the inversion theorem!

## CCPs and CVFs when $\varepsilon$ follows i.i.d T1EV

- In many cases, you will see  $\varepsilon$  follows i.i.d. T1EV so let's see some expressions in a binary case. Let's say we observe  $p_1$  and  $\Delta v = \Delta v_1 = v_1 - v_2$ .

$$\begin{aligned} p_1(x) &= Q_1(\Delta v_1(x), x) = \int_{-\infty}^{\infty} \int_{-\infty}^{\Delta v_1(x) + \varepsilon_1} g(\varepsilon_1, \varepsilon_2 | x) d\varepsilon_2 d\varepsilon_1 \\ &= \frac{\exp(\Delta v_1(x))}{1 + \exp(\Delta v_1(x))} \end{aligned}$$

$$\log p_1(x) = \Delta v_1(x) - \log(1 + \exp(\Delta v_1(x)))$$

- The ex-ante value function is  $V(x) = \log(\sum_{j=1,2} \exp(v_j(x))) + \gamma$ .<sup>1</sup> Using this, let's compute the correction term,  $\psi_1(x) := V(x) - v_1(x)$ :

$$\begin{aligned} \psi_1(x) &= \log\left(\sum_{j=1,2} \exp(v_j(x))\right) + \gamma - v_1(x) \\ &= \gamma - \log\left(\frac{\exp(v_1(x))}{\sum_{j=1,2} \exp(v_j(x))}\right) = \gamma - \log p_1(x). \end{aligned}$$

<sup>1</sup>If you are interested in derivation, see [here](#).  $\gamma \approx 0.577215$  is the Euler-Macheroni constant.



# Static Discrete Choice and CCP

- Let's consider a static model of a utility maximizing agent:
  - Choice:  $j \in \mathcal{J} := \{1, \dots, J\}$  and state  $x \in \mathcal{X} := \{1, \dots, X\}$ .
  - $u(x, j)$  is the utility the agent enjoys given choice  $j$  and state  $x$ .
  - $\varepsilon := (\varepsilon_1, \dots, \varepsilon_J)$  is the vector of idiosyncratic shocks.
  - $d^o(x, \varepsilon) := (d_1^o(x, \varepsilon), \dots, d_J^o(x, \varepsilon))$  is the vector of indicator functions that characterize the decision rule.
- The agent solves:  $\max_{j \in \mathcal{J}} u(x, j) + \varepsilon_j$ .
- The decision rule is obvious:

$$d_j^o(x, \varepsilon) = \mathbb{1}[u(x, j) + \varepsilon_j \geq u(x, k) + \varepsilon_k \quad \forall k \in \mathcal{J}].$$

# Static Discrete Choice and CCP

- We only observe  $\{(d_n, x_n)\}_{n=1}^N$  where  $d_n := (d_{1n}, \dots, d_{Jn})$  and  $d_{jn} := \mathbb{1}(\text{agent } n \text{ chose } j)$ .
- We are interested in how much can we say about  $u(x, j)$  from the 'conditional choice probabilities (CCP)' we compute from the data:

$$\begin{aligned} p(d_j = 1|x) &= \mathbb{E}_\varepsilon (\mathbb{1}[u(x, j) + \varepsilon_j \geq u(x, k) + \varepsilon_k \quad \forall k \in \mathcal{J}]) \\ &= \mathbb{E}_\varepsilon (\mathbb{1}[u(x, j) - u(x, k) \geq \varepsilon_k - \varepsilon_j \quad \forall k \in \mathcal{J}]) \end{aligned}$$

- Assuming that we know the distribution of  $\varepsilon \sim G$ , we can compute the RHS. Let's assume  $\varepsilon \sim$  i.i.d T1EV. Then,

$$p(d_j = 1|x) = \frac{\exp(u(x, j) - u(x, k))}{\sum_{l \in \mathcal{J}} \exp(u(x, l) - u(x, k))}$$

- Adding or multiplying all utilities by a constant does not change the CCPs. Thus, we say that the utilities are identified up to normalization.

## Static Discrete Choice and CCP

- So, how do we estimate  $u(x, j) - u(x, k) \forall j \in \mathcal{J}$  for each  $x \in \mathcal{X}$ ?
- We first choose which choice-specific utility to normalize for each  $x \in \mathcal{X}$ . For convenience, a lot of applications normalize choice  $j = 1$  for all  $x \in \mathcal{X}$  by setting  $u(x, 1) = 0$  for all  $x \in \mathcal{X}$ .<sup>2</sup>
- We have  $(J - 1) \times X$  (linearly independent) equations and  $(J - 1) \times X$  “parameters” to estimate (is it under/exactly/over-identified?):
  - The equations are  $p(d_j = 1|x) - \frac{\exp(u(x, j) - u(x, 1))}{1 + \sum_{k \neq 1} \exp(u(x, k) - u(x, 1))} = 0$  for  $j \in \{2, \dots, J\}$  for each  $x$ .
  - The parameters are  $u(x, j)$  for  $j = \{2, \dots, J\}$  for each  $x$ .
- By differencing the log of CCPs, we can instead use the equation below and avoid computing the denominator:

$$\log \frac{p(d_j = 1|x)}{p(d_1 = 1|x)} = u(x, j) - u(x, 1)$$

---

<sup>2</sup>Arcidiacono and Miller (2019) shows why normalizing like this without ‘actually knowing’  $u(x, j)$  for one of the choices for each  $x \in \mathcal{X}$  is misleading.

# Static Discrete Choice and CCP

- Let's follow 'plug-in GMM' approach:
  - Estimate the CCPs  $p(d_j = 1|x)$  from the data:  $\hat{p}(d_j = 1|x)$ .
  - Plug the estimates of the CCPs in the set of equations above.
  - Construct a criterion function with the equations above and minimize the criterion function using the parameters  $u(x, j)$ .
- In other words, we are using  $u(x, j)$ 's to match the set of equations whose element looks like below:

$$\log \frac{\hat{p}(d_j = 1|x)}{\hat{p}(d_1 = 1|x)} = u(x, j) - u(x, 1) = u(x, j) - 0 = u(x, j)$$

- If  $u(x, j)$  is allowed to be as flexible as possible, then the estimate is straightforward:  $\log \frac{\hat{p}(d_j=1|x)}{\hat{p}(d_1=1|x)}$ . If  $u(x, j)$  is parameterized, then it resembles over-identified GMM (more equations than the number of parameters).

# Dynamic Discrete Choice and CCP

- Most of the things we just saw in the static case are applicable in a dynamic framework with just a few more factors: transition probabilities  $f_{jt}(x_{t+1}|x_t)$ , a discount factor  $\beta \in [0, 1]$ , and time horizon  $T \leq \infty$ .
- The agent now solves: 
$$\max_{d_t^*} \mathbb{E}_{d_t^*} \left[ \sum_{\tau=t}^T \beta^{\tau-t} (u(x_\tau, j_\tau) + \varepsilon_{j,\tau}) | x_t, \varepsilon_t \right]$$
  - $d_t^*$  is a set of decision rules  $d_t^o(x_\tau, \varepsilon_\tau)$  for the state at  $t$ ,  $(x_t, \varepsilon_t)$ , and for all possible realizations of  $(x_\tau, \varepsilon_\tau)_{\tau=t+1}^T$ .
  - The expectation is taken over all possible realization of  $(x_\tau, \varepsilon_\tau)$  induced by the decision rule  $d_t^*$ .
  - In other words, the agent decides what to do today by reflecting all optimal choices the agent will choose in the future given  $\beta, G, f_{jt}(x_{\tau+1}|x_\tau), T$ .

# Dynamic Discrete Choice and CCP

- We saw in the lecture that:

$$\begin{aligned}V_t(x_t, \varepsilon_t) &= \max_{d_t^*} u(x_t, j_t) + \varepsilon_{j_t} + \mathbb{E}_{d_t^*} \left[ \sum_{\tau=t+1}^T \beta^{\tau-t} (u(x_\tau, j_\tau) + \varepsilon_{j_\tau}) \mid x_t, \varepsilon_t \right] \\&= \max_j u(x_t, j) + \varepsilon_{j_t} + \beta \int_{\varepsilon} \sum_{x_{t+1}} V_{t+1}(x_{t+1}, \varepsilon) f_{j_t}(x_{t+1} \mid x_t) g_{t+1}(\varepsilon \mid x_{t+1}) d\varepsilon \\&= \max_j u(x_t, j) + \varepsilon_{j_t} + \beta \sum_{x_{t+1}} \bar{V}_{t+1}(x_{t+1}) f_{j_t}(x_{t+1} \mid x_t).\end{aligned}$$

- Let's make this look similar to the static case! We already have  $\varepsilon_{j_t}$ . How can we make the rest look similar to  $u(x, j)$ ? This is where conditional value function comes in.
- Recall that  $v_{j_t}(x_t)$  is the life-time utility that you commit to choose  $j$  at  $t$  and choose optimally from  $t+1$  and onward:

$$v_{j_t}(x_t) := u(x_t, j_t) + \beta \sum_{x_{t+1}} \bar{V}_{t+1}(x_{t+1}) f_{j_t}(x_{t+1} \mid x_t)$$

## Dynamic Discrete Choice and CCP

- The optimization problem at  $t$  now looks like:  $\max_j v_{jt}(x_t) + \varepsilon_{jt}$
- The decision rule now looks also similar:

$$d_{jt}^o(x_t, \varepsilon_t) = \mathbb{1}[v_{jt}(x_t) + \varepsilon_j \geq v_{kt}(x_t) + \varepsilon_k \quad \forall k \in \mathcal{J}].$$

- Assuming  $\varepsilon_t$  is i.i.d. T1EV for every  $t$ , we have:

$$p(d_{jt} = 1 | x_t) = \frac{\exp(v_{jt}(x_t))}{\sum_{k \in \mathcal{J}} \exp(v_{kt}(x_t))} = \frac{\exp(v_{jt}(x_t) - v_{kt}(x_t))}{\sum_{l \in \mathcal{J}} \exp(v_{lt}(x_t) - v_{kt}(x_t))} \quad (1)$$

for some choice  $k \in \mathcal{J}$  for each  $x_t$ .

- There is one more complication, though; there's recursive component:

$$\begin{aligned} & v_{jt}(x_t) - v_{kt}(x_t) \\ &= u(x_t, j) - u(x_t, k) + \beta \left( \sum_{x_{t+1}} \bar{V}_{t+1}(x_{t+1}) [f_{jt}(x_{t+1} | x_t) - f_{kt}(x_{t+1} | x_t)] \right) \end{aligned}$$

and  $\bar{V}_\tau(x_\tau) = \gamma + \log(\sum_{j \in \mathcal{J}} \exp(v_{j\tau}(x_\tau)))$ .

# Dynamic Discrete Choice and CCP

- How can we use equation (1) and bypass computing  $\bar{V}_{t+1}(x_{t+1})$ ?
- This is where the inversion theorem from Hotz and Miller (1993) and finite dependence from Arcidiacono and Miller (2011, 2020) come in.
- The inversion theorem implies that there exists a function  $\psi : \mathcal{X} \rightarrow \mathbb{R}^J$  such that  $\psi_j(p(x)) = \bar{V}(x) - v_j(x)$  where  $p(x)$  is the vector of CCPs. If  $\varepsilon \sim \text{T1EV}$ , then  $\psi_j(x) = \gamma - \log(p_j(x))$ .
- Let's cover a DDC model with a terminal choice:
  - $j \in \mathcal{J}$  is a terminal choice if i) the agent moves to the 'absorbing state' deterministically next period, or ii) the agent chooses it once, she must choose it for the rest of her life (Hotz and Miller, 1993).
  - If choice 1 is the terminal choice with property (i), then

$$\begin{aligned}\bar{V}_{t+1}(x_{t+1}) &= \psi_1(p_{t+1}(x_{t+1})) + v_{1,t+1}(x_{t+1}) \\ &= \psi_1(p_{t+1}(x_{t+1})) + u_{1,t+1}(x_{t+1}) + \beta \sum_{x_{t+2}} \bar{V}_{t+2}(x_{t+2}) f_{1,t+1}(x_{t+2}|x_{t+1}) \quad (2) \\ &= \psi_1(p_{t+1}(x_{t+1})) + u_{1,t+1}(x_{t+1}) + \beta \bar{V}_{t+2}(a), \quad a \text{ is the absorbing state.}\end{aligned}$$



# Dynamic Discrete Choice and CCP

- Remember that we started with some  $v_{jt}(x_t)$  and  $v_{kt}(x_t)$  at  $t$ . We want to avoid computing ex-ante value function  $\bar{V}(x)$ , which again contains  $v_{j\tau}(x_\tau)$ 's.
- We see that regardless of what choice the agent made at  $t$ , she reaches at state  $a$  deterministically at  $t + 2$  if she chooses the terminal choice at  $t + 1$ .

$$f_{1,t+1}(x_{t+2}|x_{t+1})f_{jt}(x_{t+1}|x_t) = f_{1,t+1}(x_{t+2}|x_{t+1})f_{kt}(x_{t+1}|x_t), \forall x_{t+2} \in \mathcal{X}$$
$$f_{1,t+1}(a|x_{t+1})f_{jt}(x_{t+1}|x_t) = f_{1,t+1}(a|x_{t+1})f_{kt}(x_{t+1}|x_t) = 1$$

- Using this feature, we can cancel the expectation of the ex-ante value function at  $t + 1$ .

$$v_{jt}(x_t) = u_{jt}(x_t) + \beta \sum_{x_{t+1}} \left[ \frac{\psi_1(p_{t+1}(x_{t+1})) + u_{1,t+1}(x_{t+1}) + \bar{V}_{t+2}(x_{t+2}|x_{t+1})f_{1,t+1}(x_{t+2}|x_{t+1})}{\beta \sum_{x_{t+2}} \bar{V}_{t+2}(x_{t+2}|x_{t+1})f_{1,t+1}(x_{t+2}|x_{t+1})} \right] f_{jt}(x_{t+1}|x_t)$$

$$v_{kt}(x_t) = u_{kt}(x_t) + \beta \sum_{x_{t+1}} \left[ \frac{\psi_1(p_{t+1}(x_{t+1})) + u_{1,t+1}(x_{t+1}) + \bar{V}_{t+2}(x_{t+2}|x_{t+1})f_{1,t+1}(x_{t+2}|x_{t+1})}{\beta \sum_{x_{t+2}} \bar{V}_{t+2}(x_{t+2}|x_{t+1})f_{1,t+1}(x_{t+2}|x_{t+1})} \right] f_{kt}(x_{t+1}|x_t)$$

$$v_{jt}(x_t) - v_{kt}(x_t) = u_{jt}(x_t) - u_{kt}(x_t) + \beta \sum_{x_{t+1}} \left[ \frac{\psi_1(p_{t+1}(x_{t+1})) + u_{1,t+1}(x_{t+1})}{\beta \sum_{x_{t+2}} \bar{V}_{t+2}(x_{t+2}|x_{t+1})f_{1,t+1}(x_{t+2}|x_{t+1})} \right] [f_{jt}(x_{t+1}|x_t) - f_{kt}(x_{t+1}|x_t)]$$

# Dynamic Discrete Choice and CCP

- Taking logs in equation (1) and using appropriate differencing, we get:

$$\begin{aligned} \log \frac{p_t(d_j = 1|x_t)}{p_t(d_k = 1|x_t)} &= v_{jt}(x_t) - v_{kt}(x_t) \\ &= u_{jt}(x_t) - u_{kt}(x_t) + \beta \left( \sum_{x_{t+1}} \left( [\psi_1(p_{t+1}(x_{t+1})) + u_{1,t+1}(x_{t+1})] \right) \right. \\ &\quad \left. \times [f_{jt}(x_{t+1}|x_t) - f_{kt}(x_{t+1}|x_t)] \right) \end{aligned}$$

- So, the identifying equation has just (a little) more complicated form on the RHS in the dynamic case with future utilities, transitions, discount factor, and the correction term ( $\psi$ , which is again just a specific function of CCPs)
- Finite dependence (Arcidiacono and Miller, 2011, 2020) generalizes this concept when we do not necessarily have such special choices in a discrete space.

# Dynamic Discrete Choice and CCP

- Convince yourself that we have the same number of equations as the number of parameters in a dynamic case, too. However, as we start parameterizing the utilities, then we have overidentification.
- Estimation is a plug-in GMM (again):
  - Estimate the CCPs and transition matrices from the data.
  - Stack all the identifying equations you need to identify the parameters.
  - Plug the estimates of the CCPs and transition probabilities in the identifying equations.
  - Construct a criterion function.
  - Minimize the criterion function with the parameters (i.e., the components that determine  $u_{jt}(x_t)$ ).

## Kang, Lowery, Wardlaw (2015)

- This paper estimates the FDIC's decision to close a bank using a dynamic discrete choice model.
- 'Closing a bank' is the terminal choice.
- Once a bank is closed, it moves to a set of absorbing states (" $\Delta$ ").
- The FDIC gets zero utility once the bank is in an absorbing state ( $u(x) = 0$ ; however, it still gets idiosyncratic shocks,  $\varepsilon_1$ , from the closed bank throughout the FDIC's life).

## KLW Estimation Procedure

- The cost function is a sum of monetary costs and non-monetary costs:

$$c(x_{it}) = MC(x_{it}|\theta_{mc}) + \tilde{x}'_{it}\theta_{nmc} \quad (3)$$

- $MC(x_{it}|\theta_{mc})$  is estimated separately with Tobit regression.
- CCPs are estimated using a flexible logit. Each bank-specific variable has 3 lags, each approximated using a cubic spline method.

$$p_1(x_{it}) = \Lambda(\{B(b_{is})\}_{s=t,\dots,t-3}, \{m_{is}\}_{s=t,\dots,t-3}) \quad (4)$$

- After estimating the CCPs and relevant parameters for the monetary cost function, KLW estimate non-monetary costs using the following equation:

$$\sigma \log \frac{p_0(x_{it})}{p_1(x_{it})} = \beta \mathbb{E}_x[-\sigma \log p_1^{(1)}(x_{0it}) - c^{(1)}(x_{0it})] + c(x_{it}) \quad (5)$$

- The moment condition that yields the main estimating equation is the orthogonality between the subset of bank-specific variables  $\tilde{z}_{it}$  and  $v(z_{it}, \hat{\pi}, \theta_{nmc}, \hat{\theta}_{mc})$ , which is (RHS-LHS) of equation (5).

$$\mathbb{E}[\tilde{z}_{it} v(z_{it}, \hat{\pi}, \theta_{nmc}, \hat{\theta}_{mc})] = 0 \quad (6)$$

# KLW Estimation Procedure

- 1 Estimate  $\theta_{mc}$  from the actual monetary payouts by the FDIC using a Tobit model.
- 2 Estimate the transition process for each variable  $x_{it}$ . Table 4 shows the estimation result.
- 3 Estimate the probability of closure with a flexible logit as a function of all  $x_{it}$  and determinants of the transition process.

- 4 Run simulation to:

- draw projected period-ahead state variables using the estimates from step 2 ("transitions")
- generate the corresponding period-ahead monetary costs from step 1
- generate the corresponding CCPs from a flexible logit function from step 3

Repeat 5000 times per observation, that is, per bank-quarter. Then, calculate the average across the 5,000 simulations as a sample analogue to the expectation of the CCPs. Similarly, calculate the expectation of the period-ahead state variables as well.

- 5 Using the expected period-ahead CCPs and state variables, estimate  $\theta_{nmc}$  from estimating equation (5) using continuously updated GMM.

# Installing COMPECON Matlab toolbox

- <https://pfackler.wordpress.ncsu.edu/compecon-download/>
- Download, unzip, and put it in a folder you put your toolboxes
- I use C:\Program Files\MATLAB\R2021b
- Add COMPECON folder in path in Home → Set Path → Add Folder (or Add with Subfolders)

## Code: Overview

- The baseline estimation is done with `logit_KLW.do`, `ccpnext_param_KLW.m`, `param_data_prep_KWL.m`, and `Avar_param_obj_KLW.m`.
- There are eight files to run, as instructed in `estimation_procedure.txt`
  - `logit_KLW.do`
  - `ccpnext_param_KLW.m` % Takes about 7 minutes
  - `param_data_prep_KLW.m`
  - `GMM1st_KLW.m`
  - `afterGMM1st_KLW.m`
  - `findmin_cont_KLW.m`
  - `theta_save_KLW.m` % Takes a long time
  - `Avar_param_obj_KLW.m`
- Once this is done ( $\sim 15$  minutes), run either:
  - `load('output_files\theta_KLW.mat', 'theta_KLW')` % shows the estimates (sign is reversed)
  - `load('output_files\T_KLW.mat', 'T_adjust_A')` % shows the estimates (sign is reversed) and the standard errors



## Code: Including time trend

- 1 In `logit_KLW.do`, extend logit regression to include `v79`, a simple index for each year-quarter. Then, estimate the conditional choice probability with flexible logit.
- 2 In `ccpnext_param_KLW.m`, include time trend. The modifications only load the coefficients from the logit regression with time trend and utilize them in simulating one period-ahead CCPs from simulated period-ahead bank characteristics.
- 3 In `param_data_prep_KWL.m`, include time trend in the code similar to the second file.
- 4 In `Avar_param_obj_KLW.m`, change the dimension of the Variance-covariance matrix to include time trend and added time trend variable in `Xlogit_next` in second to the last column.
- 5 In all `m`-files in `function_files` folder, all codes refer to expected CCP, the expected CCP in the next period, and the log of expected CCP in the next period. They are estimated with flexible logit without time trend. These need to be adjusted to include time trend. The replication code has where you should change.

## Replication vs KLM with time trend

	Time trend	KLM	Change
$\beta$	0.95*** (0.01)	0.96*** (0.01)	1% 0%
$\sigma$	713 (492)	646 (430)	-9% -13%
Intercept	388885*** (82830)	507025*** (83595)	30% 1%
log(Assets)	-80842*** (12324)	-93753 (15998)	16% 30%
log(Assets <sup>2</sup> )	3729*** (565)	4334*** (782)	16% 38%
NP Loans/Assets	52769 (34654)	56343 (37543)	7% 8%
Net Income/Assets	48488** (22998)	48087** (22202)	-1% -3%
RE Owned/Assets	175473** (82638)	188034** (94043)	7% 14%
House	795 (402)	849** (458)	7% 14%
Senate	-166 (131)	-207 (156)	25% 19%

## Some tips in changing the code

- In this assignment, you will be changing the state space in estimating the conditional choice probabilities. You also may want to change the transition, utilities, or orthogonality conditions in the criterion function in Matlab accordingly.
- When changing the logit regression for the CCPs, think carefully how that affects the utilities (in this paper, the cost function) and transitions. Justify your changes to the utilities and transitions, if any.
- If you can, parallelize simulation using `parfor` from Matlab's Parallel Computing Toolbox as simulations take a lot of time.
- Takeaway: Ideally, you want to estimate the CCPs nonparametrically to most flexibly capture the CCPs. If you have reasons to think that is impossible, you can instead put some parametric assumptions and proceed, like this paper.