Teacher Project

Chao Fu, Christopher Taber and Matthew Wiswall
University of Wisconsin

January 25, 2019

1 Model (Simplified)

Note that this is a model to be taken to the data-it is simplified to abstract from things we might want to include in constructing counterfactuals.

An individual enters the model as a college student endowed with characteristics X_i (college quality(x_{i1}), gender, and race) and type $\chi_i = \{\theta_i, \nu_i\}$ where θ_i represents unobsered ability and ν_{ij} represents the vector of tastes for occupations, $\nu_i = \{\nu_1, ..., \nu_J\}$. χ_i is known to individual i but unobservable by the researcher, affects productivity and preferences, and is correlated with $X : P(\chi|X)$. Teaching ability (ξ_i) is not observed ex ante but is learned in one period if the individual chooses to teach.

1. $t = 0$: One makes decision on college major $m_i \in \mathcal{M}$.

2. $t \geq 1$: Workers choose their occupations every period. There are J occupations on the labor market; let $j = 0$ be the home sector and $j = J$ be teaching. (we begin with $J = 2$, so, home, non-teaching, teaching). We can divide each period into 5 different segments that occur sequentially

(a) Advanced degree $MA = \left(MA^1, MA^2\right)$, where MA^1 is advanced degree not related to teaching (e.g., MBA, lawyer etc), MA^2 is advanced teaching related degree.

For those with $MA = (0,0)$, make a choice among $\{(1,0),(0,1),(0,0)\}$.

For those with $MA = (0,1)$, make a choice bw $\{(1,1),(0,1)\}$

For those with $MA = (1,0)$, make a choice bw $\{(1,1),(1,0)\}$

(b) If no teacher license, agents decide whether to obtain one (l).

(c) If a worker is not certified and not a teacher, it is determined whether they receive an offer to be a teacher

(d) Workers choose an occupation.

(e) They work and receive wages.

Note to Matt-Chao and I talked abou the timing. Its not ideal, but it is necessary to be able to easily solve the model.

1.1 Technology

Let $e_{it} = \{e_{itj}\}_{j=1}^{J}$ be a worker's labor experience at period t. ($\sum_{j=1}^{J} e_{itj}$ is general experience).

1.1.1 Wages

Worker $i's$ wage in occupation $j \in \{1, ...J\}$ is given by

$$\ln\left(w_{ijt}\right) = W_j\left(e_{it}, X_i, \theta_i, m_i, MA_{it}\right) + \eta_{itj},$$

$$\begin{aligned}\ln\left(w_{ijt}\right) =& W_j\left(e_{it}, X_i, \theta_i, m_i, MA_{it}\right) + \eta_{itj} \\ =& Z'_{ijt}\gamma_j + \delta_j\theta_i + \eta_{itj}.\end{aligned}$$

where $\eta_{itj} \sim N\left(0, \sigma_j^2\right)$ and Z_{ijt} is a vector of relevant observable variables for occupation j :

- For teaching Z_{iJt} includes up to quartic terms in teacher experience, MA_{it}, and an intercept. In this case ability does not influence wages because of contracts so, $\delta_J = 0$.
- For non-teaching Z_{ijt} includes polynomial terms in experience, X_i, major,, and an intercep. (I am not sure what to do about masters here)
- The η are not known when workers make occupation decisions.

1.1.2 Teaching Production

We use value added model to measure one's output. Output is stochastic:

$$o_{it} = Z'_{iot}\gamma_0 + \delta_o\theta_i + \xi_i + \varsigma_{it} \tag{1}$$

where Z_{iot} includes a polynomial in experience, X_i, major, certification, MA_{it}, and an intercept. ς_{it} is iid $N\left(0, \sigma_\varsigma^2\right)$ and $\xi_i \sim N\left(0, \sigma_\xi^2\right)$.

As mentioned ξ_i is like an experience good-it is revealed when someone teaches, but teachers have no private information about it before hand.

1.1.3 Teaching Offers

Let $\mu(l, j_{t-1}, \sum_{t=1}^{t} e_{itJ})$ represent the probability of receiving an offer, depending on license, last period choice, and total teaching experience.

1.2 Preferences

1.2.1 Post school flow utility

The flow in nonteaching occupations is

$$u_{ijt} = \alpha w_{ijt} + X_i\prime\gamma_j + \kappa_j 1\left(j = d_{t-1}\right) + \nu_{ij} + \epsilon_{ijt}$$

where ϵ_{ijt} is extreme value, ν_{ij} was defined above, and we might want to restrict the coefficient on college quality to zero.

Non-employment is normalized to 0 except for the extreme value term

$$u_{i0t} = \epsilon_{i0t}$$

and for teaching people also care about their ability to teach

$$u_{iJt} = \alpha w_{iJt} + X_i\prime\gamma_J + \kappa_J 1\left(J = d_{t-1}\right) + \lambda\left(Z'_{iot}\gamma_0 + \delta_o\theta_i + \xi_i\right) + \nu_{ij} + \epsilon_{ijt}$$

The ϵ_{ijt} is not revealed until the occupation decision is made in stage (c). In particular it is not known when the certification decision is made.

I am putting in switching costs with the κ but we probably want something more sophisticated. The cost from going Teach-unemployed-Teach seems smaller than going Not teaching-Teaching.

1.2.2 Certification and Masters Cost

These are only relevant if the person has not already acheived them.

The utility cost of receiving a masters is

$$\begin{aligned} C_{sit} =& Z_{is}'\gamma_s + \delta_s \theta_i - \omega_{1it} \\ \equiv& c_{si} - \omega_{1it} \end{aligned}$$

where ω_{sit} is iid extreme value and Z_{is} includes an intercept and X_i. The cost of not receiving a masters is $-\omega_{0it}$

The utiliy cost of certification is

$$\begin{aligned} C_{cit} =& Z_{il}'\gamma_l + \delta_l \theta_i - \varpi_{1it} \\ \equiv& c_{ci} - \varpi_{1it} \end{aligned}$$

where ω_{it} is iid extreme and Z_{il} includes an intercept, X_i, and masters. Similarly the cost of not getting certivied is ϖ_{0it}.

There will also be an extreme value draw in the case where you don't get a masters or certivied.

These error terms are revealed sequentially at the time decisions are made, so when workers make a decision about Masters, they have not yet observed the cost of getting certified.

1.3 Individual Decision

State variables

- Pre-determined and permanent: X, χ
- Endogenous and change only once: m, l, MA (major, license, MA),EJ where EJ is an indiator for whether the person has ever taught before in which case they know

ξ_i.

- Dynamic and changing each period: e (experience), j_{t-1};
- Transitory shocks post college (independent over time) choice-specific taste shocks (ϵ), wage shocks (η), masters cost (ω), and certification cost (ϖ)
- Draw on whether one gets an offer as a teacher. We will denote the realizion of this draw as $\mathcal{J}_t$ where $\mathcal{J}_t = \{0, ..., J\}$ if one has the option of teaching and $\mathcal{J}_t = \{0, ..., J-1\}$ otherwise
- Shocks in college (iid): taste (ε).

1.3.1 Post College

At time $1 \leq t < T$,tdecisions are made at three different points in time so we need three different value functions:

- Beginning of period (masters decision): $V_t^a\left(X, \chi, l_{t-1}, MA_{t-1}, e_t, \xi, EJ_{t-1}, d_{t-1}, \omega_t\right)$
- Second part of period (certification decision): $V_t^b\left(X, \chi, l_{t-1}, MA_t, e_t, \xi, EJ_{t-1}, d_{t-1}, \varpi_t\right)$
- Fourth part (occupation decision): $V_t^d\left(X, \chi, l_t, MA_t, e_t, \xi, EJ_{t-1}, d_{t-1}, \mathcal{J}_t, \epsilon_t\right)$

Solving backwards

$$V_t^d\left(X, \chi, l_t, MA_t, e_t, \xi, EJ_{t-1}, d_{t-1}, \mathcal{J}_t, \epsilon_t\right) = \max_{j \in \mathcal{J}_t} E_{\eta_t, \omega_{t+1}}\left(u_{ijt} + \beta V_{t+1}^a\left(X, \chi, l_t, MA_t, e_{t+1}, \xi, EJ_t, j, \omega_{t+1}\right)\right)$$

where we use subscripts on expectations to indicate what the expectation is taken over.

For people who are not yet certified ($l_{t-1} = 0$)

$$V_t^b\left(X, \chi, 0, MA_t, e_t, \xi, EJ_{t-1}, d_{t-1}, \varpi_t\right) = \max_{l \in \{0,1\}}\left(lc_{ci} - \varpi_{lit} + E_{\mathcal{J},\varepsilon}\left[V_t^d\left(X, \chi, l, MA_t, e_t, \xi, EJ_{t-1}, d_{t-1}, \mathcal{J}_t, \epsilon_t\right)\right]\right)$$

and for people who are already certified

$$V_t^b\left(X,\chi,1,MA_t,e_t,\xi,EJ_{t-1},d_{t-1},\varpi_t\right) = E_{\mathcal{J},\varepsilon}\left[V_t^d\left(X,\chi,1,MA_t,e_t,\xi,EJ_{t-1},d_{t-1},\mathcal{J}_t,\epsilon_t\right)\right]$$

Similarly for people wihout a masters ($MA_{ti-1}=0$)

$$V_t^a\left(X,\chi,l_{t-1},0,e_t,\xi,EJ_{t-1},d_{t-1},\omega_t\right) = \max_{S\in\{0,1\}}\left(Sc_{si}-\omega_{Sit}+E_{\varpi}\left[V_t^b\left(X,\chi,l_{t-1},S,e_t,\xi,EJ_{t-1},d_{t-1},\varpi_t\right)\right]\right)$$

and for people with a masters

$$V_t^a\left(X,\chi,l_{t-1},1,e_t,\xi,EJ_{t-1},d_{t-1},\omega_t\right) = E_{\varpi}\left[V_t^b\left(X,\chi,l_{t-1},1,e_t,\xi,EJ_{t-1},d_{t-1},\varpi_t\right)\right]$$

1.3.2 Major Choice $t=0$

$V_0\left(X,\chi,\varepsilon\right)=$

$$\max_m\left\{\begin{array}{c} Z_{im}'\gamma_k+\delta_m\theta+\rho_m'\nu+\varepsilon_m+ \\ \pi E_\omega V_1^a\left(X,\chi,0,0,0,\xi,0,0,\omega_t\right)\end{array}\right\}.$$

$v_m\left(X,\chi\right)$ is non-pecuniary value of choosing major m, which depends on own quality and school-major-specific quality, e.g., it is painful to go to non-edu major if one's SAT is low (help to explain that on average edu students are worse than other majors).

1.4 Unobserved Heterogeneity

This is nonstandard because we have two different types of heterogenity boiled down to one.

- Unobserved ability θ_i is like a factor structure-we have observable measures of these to infer it

- The permanant occupation tastes ν_i are more like traditional unobserved heterogenity

For the first type we assume that we have K_{BB} measures of unobserved heterogeneity in the B&B and K_{NC} measures in the North Carolina data, some of which overlap.

We assume that for each measure we have a known parametric model (could be a regression with normal error or could be something like a logit) so that for each $D \in \{BB, NC\}$ and each $k = 1, ..., K_D$

$$Y_{ikD} = f_{kD}(\theta_i, e_{ikD}; \zeta_{kD})$$

where

- Each f_{kD} is known up to finite parameter ζ_{kD}
- The distribution of e_{ikD} is known and independent across people and measures
- The f_{kBB} and ζ_{kBB} is the same as f_{kNC} and ζ_{kNC} when the measures coincide

Life will be simpler if the support of θ_i is finite and it is naturally ordered. So I would propose assuming that it has L_θ different values $(\theta(1), ..., \theta(L_\theta))$ and use an ordered logit so

$$Pr(\theta_i = \theta(\ell)) = \Lambda(X_i'\gamma_\theta + d_\ell) - \Lambda(X_i'\gamma_\theta + d_{\ell-1})$$

with $d_0 = -\infty$ and $d_{L_\theta} = \infty$.

Empircially identifying the distribution of ν_i is going to be difficult because we can't use MLE as we have two different data sets. I would assume that each has the discrete distribution dictated by Gauss quadrature where each marginal distribution takes on L_ν points. We then have just three parameters to identify-2 variances and a correlation.

This gives a total of $L_\theta \times L_\nu \times L_\nu$ different types.

1.5 Estimation

My preference would be to estimate the factor structure first for two reasons. Computatoinally it will be better. Intuitively to me it is better-it is forcing identification of types to come from the measures.

Esimtation is done in three steps:

1. Estimate distribution of θ_i and measurement system ζ_{kBB} from B&B data
2. Estimate measurement system for North Carolina ζ_{KNC}
3. Estimate the rest of the parameters of the model using indirect inference

Some thoughts:

- Ideally we could do 1 and 2 together, but in practice we need to keep the two different data sets on two different computers so we can't
- We might think about dividing 3 into two steps-first estimating the post school model, and then estimate the major choice using $B\&B$ only. The issue here is whether that works with ν_i
- How we deal with θ_i is up for grabs. We could simulate the measurement system as part of the simulation and then have auxiliary parameters that depend on the measured variables. Alternatively we could define the auxiliary parameters as condition only on θ_i and use Bayes theorem in creative ways to estimate these from the data (given the measurement model)

1.6 Identification

Thoughts:

- For most of the standard selection issues, for better and for worse the factor structure does a lot of the heavy lifting.
- The elasticity of major choice with respect to compensation
  - We talked about this last week
  - Actual model is more complicated than the simple and probably identified from functional form-but even then we still need to pick auxiliary moments we think will be helpful in identifying it
  - Complications are-we have unemployment and logit error terms and only one α so in short run elasticity into occupation is pinned down to elasticity into work, in the longer run we also have ν_i to worry about and identify, the fact that we have expected wage levels but separability is in terms of logs would help. None of this is atractive to me.
  - We want an exclusion restriction. Some ideas:
    * Chao mentioned using local labor market conditions. My concern (as we discussed) is labor demand. We would be assuming one could always find a teaching job in a recession, but this seems problematic. Another way to do it would be to allow the probability of getting an offer to depend on a recession-in this case we could identify the elasticity from the exit rate from teaching. This would be assuming that the elasticity of exit is closely related to the elasticity of entrance (they won't be identical because of the learning about teaching ability).
    * Matt and I discussed using ability as the exclusion. This would essentially mean assuming that θ_i and ν_i are independent. Its a bit odd in that we have people caring about their value added as a teacher-so its weird to assume

they don't care about their ability outside of teaching.

* Another possibility is experience. By assuming that preferences are stable, the hazard rate out of teaching would help identify this. Again not ideal as there are other reasons why the hazard might be falling.