1 Assumptions behind Identification

Compare the assumptions used to identify ex ante and ex post ATE, TOT, MTE, and LATE based on (a) matching, (b) IV, and (c) selection approaches. Use the cross-section model where observations are independently distributed:

$$
\begin{aligned}
Y &= DY_1 + (1-D)Y_0 \\
Y_1 &= \mu_1(X) + U_1, E[U_1] = 0 \\
Y_0 &= \mu_0(X) + U_0, E[U_0] = 0 \\
D &= 1(\mu_D(Z) \geq V), D \in \{0,1\} \\
(U_0, U_1, V) &\perp\!\!\!\perp X, Z
\end{aligned}
$$

Z and X may contain some elements in common. Specifically, compare the information available from (Y, Z, X) and that (Y, Z, D, X), i.e.,

- ▷ Information from IV assumptions using $E[Y|Z, X]$.
- ▷ Selection model information using $E[Y|Z, X, D]$.
- ▷ Matching information using Y, Z, X, D.

Use the relationship:

$$E[Y|Z,X] = E[Y|Z,X,D=1]P[D=1|Z,X] + E[Y|Z,X,D=0]P[D=0|Z,X]$$

to compare the estimators. Compute:

$$\frac{\partial E[Y|Z,X]}{\partial P(Z,X)}$$

for this model and relate to ATE and MTE. How do the derivatives of the control function compare to MTE? Compare the information sets of the agent and the econometrician under each approach.

Solution. The assumptions required to get identified estimates for each of these approaches are listed below

- ▷ **Matching Assumptions**:

$$
\begin{aligned}
(Y_1, Y_0) &\perp\!\!\!\perp D|X, \forall X, \\
0 < P[D=1|X] &< 1, \forall X.
\end{aligned}
$$

- ▷ **IV Assumptions**:

$$(Y_1, Y_0) \not\perp\!\!\!\perp D|X, Z$$
$$(Y_1, Y_0) \perp\!\!\!\perp Z|X$$
$$P(D = 1|Z, X) \textbf{ is a nondegenerate function of } Z \textbf{ given } X$$

▷ **Selection Assumptions**:

We now summarize 5 distinct information sets helpful in the analysis below.

1. $\sigma(I_{R^*})$, i.e. a *relevant information set*, is an information set with associated random variable I_R^* that satisfies conditional independence, i.e.

$$(Y_1, Y_0) \perp\!\!\!\perp D|I_{R^*}.$$

2. $\sigma(I_R)$, i.e. the *minimal relevant information set* is the smallest information set that satisfies conditional independence.

3. $\sigma(I_A)$, i.e. the *agent's information set* is the information set available to the agent at the time decisions to participate in treatment are made.

4. $\sigma(I_E)$, i.e. the *information set available to the economist*.

5. $\sigma(I_{E^*})$, i.e. the *information set used by the economist in conducting an empirical analysis*.

The structure of these information sets gives us

$$\sigma(I_R) \subseteq \sigma(I_{R^*})$$
$$\sigma(I_A) \subseteq \sigma(I_R)$$
$$\sigma(I_E) \subseteq \sigma(I_{E^*})$$

The information sets of the agent and the econometrician under each approach are

▷ **Matching**: $\sigma(I_R) \subseteq \sigma(I_E)$, i.e. the information set used in the analysis by the econometrician contains the relevant information set. We know from above that $\sigma(I_A) \subseteq \sigma(I_R)$, that the relevant set contains the agent's information set. Therefore, in matching, we are assuming that the econometrician knows what the agent knew at the time the agent chose to participate in treatment- a strong assumption.

Also note that we allow for the econometrician to have *more* information than the minimal relevant information and thus if the econometrician uses more or less than the relevant information set, conditional independence will not necessarily hold. Thus, the econometrician has to exactly control for I_A. In other words, the econometrician has to be able to partition its information set I_{E^*} to what the agent knew and the other irrelevant data it has access to.

- ▷ **IV**: Instrumental variables allows for $\sigma(I_A) \nsubseteq \sigma(I_E)$ or $\sigma(I_E) \nsubseteq \sigma(I_A)$, i.e. there are some things the agent knew that the econometrician does not and vice versa.

- ▷ **Selection**: In the selection approach, we want to express U_1 as

$$U_1 = E\left(U_1 \mid V\right) + \epsilon$$

where V is the same V that entered on the RHS of the expression for the choice equation. We can then write

$$E\left(Y_1 \mid X, Z, D = 1\right) = \mu_1\left(X\right) + \underbrace{E\left(U_1 \mid \mu_D\left(Z\right) > V\right)}_{\text{control function}}$$

where we can write $E\left(U_1 \mid V\right) = \frac{Cov(U_1,V)}{Var(V)}V$. So in this setting, similar to case of IV, we have that $\sigma\left(I_E\right) \subseteq \sigma\left(I_R\right) \subseteq \sigma\left(I_A\right)$, where the reverse relations do not in general hold. That is, the conditional independence assumption we relied on for matching is likely violated. However, because of the distributional assumptions we make using the control function, the econometrician is able to correct for the violation of the conditional independence assumption. Here, we have more structure than the IV case and so can recover more terms.

Below, we discuss what these methods are able to identify.

- ▷ **What Matching Identifies**: With matching, we are able to identify

$$\begin{aligned} ATE(X) &= E[Y_1 - Y_0|X] \\ TOT(X) &= E[Y_1 - Y_0|X, D = 1] \\ MTE(X, U) &= E[Y_1 - Y_0|X, U] \end{aligned}$$

since in this case, they are all equal to each other since treatment is randomized for a given $X = x$. LATE is not well defined here. Also, this does not mean we can identify ATE, TOT, or MTE overall since treatment is only random conditional on X.

- ▷ **What IV Identifies**: IV identifies LATE:

$$LATE = \frac{E[Y|Z = z'] - E[Y|Z = z]}{E[D|Z = z'] - E[D|Z = z]} = E[Y_1 - Y_0|Compliers].$$

- ▷ **What Selection Identifies**: We can identify $\mu_1\left(X\right)$ and $\mu_0\left(X\right)$. Then, using our assumptions for the control function, we can identify $E\left(U_1 \mid \mu_D\left(Z\right) > V\right)$, $E\left(U_1 \mid \mu_D\left(Z\right) \leq V\right)$, etc. This allows us to recover the MTE, ATE, TOT, and LATE – so in principle we can recover more than when using the IV specification.

Finally, we compute $\frac{\partial E[Y|Z,X]}{\partial P(Z,X)}$:

$$\begin{aligned}
\frac{\partial E[Y|Z,X]}{\partial P(Z,X)} &= \frac{\partial}{\partial P(Z,X)}(E[Y_0] + E[Y_1 - Y_0|D=1,Z,X]P(Z,X)) \\
&= \frac{\partial}{\partial P(Z,X)}(E[Y_0] + E[Y_1 - Y_0|P(Z,X) > U_D]P(Z,X)) \\
&= \frac{\partial}{\partial P(Z,X)}(E[Y_0] + \int_0^{P(Z,X)} MTE(U_D)dU_D) \\
&= MTE(P(Z,X)).
\end{aligned}$$

■

2 Dynamic Model of Life Cycle Consumption

Consider a dynamic model of life cycle consumption $(C_t)_{t=1}^T$ over horizon T for agents facing an exogenous but uncertain income flow $\{Y_t\}_{t=1}^T$. Assets evolve according to

$$[1] : A_{t+1} = (1 + r) A_t + Y_t - C_t$$

and initial endowment A_0 is specified as exogenous. r is a known constant exogenous interest rate. Agents maximize

$$\sum_{t=0}^{T} \beta^t U(C_t)$$

subject to [1]. Also note that $\beta(1 + r) = 1$. Assume $U(C_t)$ is quadratic:

$$U(C_t) = \delta_0 + \gamma_1 C_t + \gamma_2 C_t^2, \quad \gamma_2 < 0, \gamma_1 > 0$$

and the Y_t process defined as:

$$\begin{aligned}
[A] : Y_t &= P_t + \tau_t \\
P_t &= \rho P_{t-1} + \lambda_t, \quad \mathbb{E}[\lambda_t] = 0 \\
\tau_t &= \omega_t + \theta \omega_{t-1} \\
\mathbb{E}[\omega_j] &= 0, \forall j
\end{aligned}$$

is stationary; ω_j are mutually independent, P_0 is fixed and exogenous. Agents at t know $\epsilon_{t-j}, j \leq t$ but not future ϵ_t. Suppose $T < \infty$.

Problem 2.1. Characterize the response of consumption C_t of an agent to a unit shock in ω_t and to a unit shock in λ_t. How do they change with age? With increases in ρ? With t? With increases in $\sigma^2_{\omega_t}$?

Solution. The agent's maximization problem is

$$\max_{\{C_t, A_{t+1}\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{T} \beta^t U(C_t)\right]$$
$$\text{s.t. } A_{t+1} = (1 + r) A_t + Y_t - C_t$$

and given the functional form for $U(\cdot)$, we have $U'(C_t) = \gamma_1 + 2\gamma_2 C_t$. Plugging in the budget constraint, we can rewrite:

$$\max_{\{A_{t+1}\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{T} \beta^t U((1 + r) A_t + Y_t - A_{t+1})\right]$$

Since $Y_t = P_t + \tau_t = \rho P_{t-1} + \theta\omega_{t-1} + \lambda_t + \omega_t$, we have:

$$\max_{\{A_{t+1}\}_{t=0}^{\infty}} \mathbb{E}\left[\sum_{t=0}^{T} \beta^t U\left(\left(1+r\right)A_t + P_t + \tau_t - A_{t+1}\right)\right]$$

▷ We can solve this problem using the Bellman approach.

* Writing the sequence problem as a Bellman equation:

$$v\left(A, P, \tau\right) = \max_{A'} \left\{U\left(\left(1+r\right)A + P + \tau - A'\right) + \beta\mathbb{E}_t\left[v\left(A', P', \tau'\right)|P, \tau\right]\right\}$$

* Since the shocks are independent:

$$v\left(A, P, \tau\right) = \max_{A'} \left\{U\left(\left(1+r\right)A + P + \tau - A'\right) + \beta\mathbb{E}_t\left[v\left(A', P', \tau'\right)\right]\right\}$$

* FOC and the EC:

$$[FOC] : U'\left(C\right) = \beta\mathbb{E}_t\left[v_1\left(A', P', \tau'\right)\right]$$
$$[EC] : v_1\left(A, P, \tau\right) = \left(1+r\right)U'\left(C\right)$$

* Plugging in the expression for $U\left(C\right)$, we have

$$C_t = \mathbb{E}_t\left[C_{t+1}\right]$$

▷ Given a finite time horizon of length $T - t$, we can set $A_{T+1} = 0$. Solving the budget constraint forward from this boundary condition, we obtain:

$$C_t = \frac{r}{\left(1+r\right) - \left(1+r\right)^{-(T-t)}}\left[A_t + \sum_{k=0}^{T-t}\left(\frac{1}{1+r}\right)^k \mathbb{E}_t\left[Y_{t+k}\right]\right]$$

which captures the essence of the Permanent Income Hypothesis – current income is determined by a combination of current non-human wealth A_t and human capital wealth Y_t.

▷ Since

$$Y_t = \rho P_{t-1} + \lambda_t + \omega_t + \theta\omega_{t-1}$$

and

$$C_t = \frac{r}{\left(1+r\right) - \left(1+r\right)^{-(T-t)}}\left[A_t + \sum_{k=0}^{T-t}\left(\frac{1}{1+r}\right)^k \mathbb{E}_t\left[Y_{t+k}\right]\right]$$

▷ Write

$$\begin{aligned}
\Delta C_t &= C_t - C_{t-1} \\
&= C_t - \mathbb{E}_{t-1}\left[C_t\right] \\
&= \frac{r}{\left(1+r\right) - \left(1+r\right)^{-(T-t)}}\left[A_t - \mathbb{E}_{t-1}\left[A_t\right] + \sum_{k=0}^{T-t}\left(\frac{1}{1+r}\right)^k \mathbb{E}_t\left[Y_{t+k}\right] - \mathbb{E}_{t-1}\left[\sum_{k=0}^{T-t}\left(\frac{1}{1+r}\right)^k \mathbb{E}_t\left[Y_{t+k}\right]\right]\right] \\
&= \frac{r}{\left(1+r\right) - \left(1+r\right)^{-(T-t)}}\left[A_t - \mathbb{E}_{t-1}\left[A_t\right] + \sum_{k=0}^{T-t}\left(\frac{1}{1+r}\right)^k \left\{\mathbb{E}_t\left[Y_{t+k}\right] - \mathbb{E}_{t-1}\left[\mathbb{E}_t\left[Y_{t+k}\right]\right]\right\}\right]
\end{aligned}$$

From the Law of Iterated Expectations and the fact that $A_t = \mathbb{E}_{t-1}[A_t]$:

$$\Delta C_t = \frac{r}{(1+r)-(1+r)^{-(T-t)}}\left[\sum_{k=0}^{T-t}\left(\frac{1}{1+r}\right)^k \{\mathbb{E}_t[Y_{t+k}] - \mathbb{E}_{t-1}[Y_{t+k}]\}\right]$$

Note that

$$Y_t = P_t + \tau_t = \rho P_{t-1} + \theta\omega_{t-1} + \lambda_t + \omega_t$$

and thus $\mathbb{E}_t[Y_{t+k}] - \mathbb{E}_{t-1}[Y_{t+k}]$ can be explicitly written out for each value of k:

$$\begin{aligned}
k = 0 : &\ \mathbb{E}_t[Y_t] - \mathbb{E}_{t-1}[Y_t] \\
&= (\rho P_{t-1} + \theta\omega_{t-1} + \lambda_t + \omega_t) - (\rho P_{t-1} + \theta\omega_{t-1} + \mathbb{E}_{t-1}[\lambda_t] + \mathbb{E}_{t-1}[\omega_t]) \\
&= \lambda_t + \omega_t \\
k = 1 : &\ \mathbb{E}_t[Y_{t+1}] - \mathbb{E}_{t-1}[Y_{t+1}] \\
&= (\rho P_t + \theta\omega_t + \mathbb{E}_t[\lambda_{t+1} + \omega_{t+1}]) - (\rho\mathbb{E}_{t-1}[P_t] + \theta\mathbb{E}_{t-1}[\omega_t] + \mathbb{E}_{t-1}[\lambda_{t+1} + \omega_{t+1}]) \\
&= \rho(P_t - \mathbb{E}_{t-1}[P_t]) + \theta\omega_t \\
(\because P_t = \rho P_{t-1} + \lambda_t) &= \rho\lambda_t + \theta\omega_t \\
k = 2 : &\mathbb{E}_t[Y_{t+2}] - \mathbb{E}_{t-1}[Y_{t+2}] \\
&= (\rho\mathbb{E}_t[P_{t+1}] + \theta\mathbb{E}_t[\omega_{t+1}] + \mathbb{E}_t[\lambda_{t+2} + \omega_{t+2}]) \\
&- (\rho\mathbb{E}_{t-1}[P_{t+1}] + \theta\mathbb{E}_{t-1}[\omega_{t+1}] + \mathbb{E}_{t-1}[\lambda_{t+2} + \omega_{t+2}]) \\
&= \rho(\mathbb{E}_t[P_{t+1}] - \mathbb{E}_{t-1}[P_{t+1}]) \\
&= \rho(\mathbb{E}_t[\rho P_t + \lambda_{t+1}] - \mathbb{E}_{t-1}[\rho P_t + \lambda_{t+1}]) \\
&= \rho\lambda_t
\end{aligned}$$

and for $k \geq 3$, the argument is identical to $k = 2$. Therefore, we have:

$$\Delta C_t = \frac{r}{(1+r)-(1+r)^{-(T-t)}}\left[\{\lambda_t + \omega_t\} + \frac{\theta}{1+r}\omega_t + \rho\lambda_t \sum_{k=1}^{T-t}\frac{1}{(1+r)^k}\right]$$

Since

$$\sum_{k=1}^{T-t}\frac{1}{(1+r)^k} = \frac{\frac{1}{1+r}\left[1 - \frac{1}{(1+r)^{T-t}}\right]}{1 - \frac{1}{1+r}} = \frac{(1+r)-(1+r)^{-(T-t)}}{r}$$

the sum can be simplified to

$$\Delta C_t = \frac{r}{(1+r)-(1+r)^{-(T-t)}}\left[\{\lambda_t + \omega_t\} + \frac{\theta}{1+r}\omega_t\right] + \rho\lambda_t$$

Now we examine how consumption responds to these shocks.

▷ **First, consider a unit shock to ω_t.**

This is the transitory component. It is scaled by

$$\frac{r}{(1+r)-(1+r)^{-(T-t)}}$$

which depends on the time horizon (= age). It is unaffected by ρ and σ_ω^2 and is increasing in t.

▷ **Next, consider a unit shock to λ_t.**

This is the permanent component, so the shock is scaled by ρ. Therefore it is increasing in ρ and unaffected by age, t or σ_ω^2. ■

Problem 2.2. Show how to identify the model using panel data on individuals $(C_t, Y_t)_{t=1}^T$. Assume shocks are independent across agents.

Solution. First, use the income data to place the following covariance restrictions:

▷ For variance of ΔY_t:

$$\begin{aligned}\Delta Y_t &= Y_t - Y_{t-1} \\ &= (P_t - P_{t-1}) + (\tau_t - \tau_{t-1}) \\ &= (\rho P_{t-1} + \lambda_t - P_{t-1}) + (\omega_t + \theta\omega_{t-1} - (\omega_{t-1} + \theta\omega_{t-2}))\end{aligned}$$

and since

$$\mathrm{Var}[P_t] = \frac{\sigma_\lambda^2}{1-\rho^2}$$

it must be that

$$[1] : \mathrm{Var}\left[\Delta Y_t\right] = \sigma_\lambda^2 \left[\frac{(1-\rho)^2}{1-\rho^2}\right] + \sigma_\omega^2 \left[1 + \theta^2 + (1-\theta)^2\right]$$

▷ For covariance:

$$\begin{aligned}&\mathrm{Cov}\left(\Delta Y_t, \Delta Y_{t-1}\right) \\ =&\,\mathrm{Cov}((\rho P_{t-1} + \lambda_t - P_{t-1}) + (\omega_t + \theta\omega_{t-1} - (\omega_{t-1} + \theta\omega_{t-2}))\,, \\ &(\rho P_{t-2} + \lambda_{t-1} - P_{t-2}) + (\omega_{t-1} + \theta\omega_{t-2} - (\omega_{t-2} + \theta\omega_{t-3}))) \\ =&\, \sigma_\lambda^2 \left[\frac{\rho(1-\rho)^2}{1-\rho^2} + (\rho - 1)\right] + \sigma_\omega^2 \left[(\theta-1) + \theta(\theta-1)\right]\end{aligned}$$

so

$$[2] : \mathrm{Cov}\left[\Delta Y_t, \Delta Y_{t-1}\right] = \sigma_\lambda^2 \left[\frac{\rho(1-\rho)^2}{1-\rho^2} + (\rho-1)\right] + \sigma_\omega^2 \left[(\theta-1) + \theta(\theta-1)\right]$$

▷ For variance of ΔC_t:

Since we had

$$\Delta C_t = \frac{r}{(1+r) - (1+r)^{-(T-t)}} \left[\{\lambda_t + \omega_t\} + \frac{\theta}{1+r}\omega_t\right] + \rho\lambda_t$$

it follows that

$$\begin{aligned}[3] : \mathrm{Var}\left[\Delta C_t\right] = &\left(\frac{r}{(1+r) - (1+r)^{-(T-t)}}\right)^2 \left[1 + \left(\frac{\theta}{1+r}\right)^2\right] \sigma_\omega^2 + \rho^2\sigma_\lambda^2 \\ &+ \left(\frac{r}{(1+r) - (1+r)^{-(T-t)}}\right)^2 \sigma_\lambda^2\end{aligned}$$

Therefore, from the income data, we can use $[1]$, $[2]$, $[3]$ to estimate the model parameters of interest.

Second, we can also use both income and consumption data and get the covariance between consumption and income:

$$
\begin{aligned}
&\operatorname{Cov}\left[\Delta C_t, \Delta Y_t\right] \\
=&\operatorname{Cov}\left[\begin{array}{c} \frac{r}{(1+r)-(1+r)^{-(T-t)}}\left[\{\lambda_t+\omega_t\}+\frac{\theta}{1+r}\omega_t\right]+\rho\lambda_t, \\ (\rho P_{t-1}+\lambda_t-P_{t-1})+(\omega_t+\theta\omega_{t-1}-(\omega_{t-1}+\theta\omega_{t-2})) \end{array}\right] \\
=&\frac{r}{(1+r)-(1+r)^{-(T-t)}}\left[1+\left(\frac{\theta}{1+r}\right)^2\right]\sigma_\omega^2+\left[\frac{r}{(1+r)-(1+r)^{-(T-t)}}+\rho\right]\sigma_\lambda^2
\end{aligned}
$$

can even provide us with an over-identifying restriction. ■

Problem 2.3. Suppose we add to equation $[A]$ measurement error M_t where M_t is i.i.d and $\mathbb{E}\left[M_t\right]=0, \operatorname{Var}\left[M_t=\sigma_M^2\right]$. What parameters can be identified using only income? Using income and consumption?

Solution. Recall that we had the following equations from income data:

$$
\begin{aligned}
&[1]:\operatorname{Var}\left[\Delta Y_t\right]=\sigma_\lambda^2\left[\frac{(1-\rho)^2}{1-\rho^2}\right]+\sigma_\omega^2\left[1+\theta^2+(1-\theta)^2\right] \\
&[2]:\operatorname{Cov}\left[\Delta Y_t, \Delta Y_{t-1}\right]=\sigma_\lambda^2\left[\frac{\rho(1-\rho)^2}{1-\rho^2}+(\rho-1)\right]+\sigma_\omega^2\left[(\theta-1)+\theta(\theta-1)\right] \\
&[3]:\operatorname{Var}\left[\Delta C_t\right]=\left(\frac{r}{(1+r)-(1+r)^{-(T-t)}}\right)^2\left[1+\left(\frac{\theta}{1+r}\right)^2\right]\sigma_\omega^2+\rho^2\sigma_\lambda^2 \\
&\quad+\left(\frac{r}{(1+r)-(1+r)^{-(T-t)}}\right)^2\sigma_\lambda^2
\end{aligned}
$$

Now we add measurement error. Then the equations change to:

$$
\begin{aligned}
&[1]:\operatorname{Var}\left[\Delta Y_t\right]=\sigma_\lambda^2\left[\frac{(1-\rho)^2}{1-\rho^2}\right]+\left[\sigma_\omega^2+\sigma_M^2\right]\left[1+\theta^2+(1-\theta)^2\right] \\
&[2]\operatorname{Cov}\left[\Delta Y_t, \Delta Y_{t-1}\right]=\sigma_\lambda^2\left[\frac{\rho(1-\rho)^2}{1-\rho^2}+(\rho-1)\right]+\left[\sigma_\omega^2+\sigma_M^2\right]\left[(\theta-1)+\theta(\theta-1)\right] \\
&[3]:\operatorname{Var}\left[\Delta C_t\right]=\left(\frac{r}{(1+r)-(1+r)^{-(T-t)}}\right)^2\left[1+\left(\frac{\theta}{1+r}\right)^2\right]\sigma_\omega^2+\rho^2\sigma_\lambda^2 \\
&\quad+\left(\frac{r}{(1+r)-(1+r)^{-(T-t)}}\right)^2\sigma_\lambda^2
\end{aligned}
$$

As it is clear from the first two moments, σ_ω^2 and σ_M^2 cannot be told apart from income data alone. Note that σ_λ^2 can still be identified.

With income and consumption, we can now use $\operatorname{Cov}\left[\Delta C_t, \Delta Y_t\right]$ as we have done in the previous part to identify the variance of the transitory shocks. So having a panel data allows us to estimate the same parameters as before when we did not have measurement errors. ■.

Problem 2.4. How does access to data on C_t aid in identifying agent information sets?

Solution. Here we stick to Professor Heckman's suggested interpretation that measurement error represents "superior information" that is observed by the individual but not by an econometrician. As we have shown previously, the panel data on (Y_t, C_t) allows us to compute $\text{Cov}\left[\Delta Y_t, \Delta C_t\right]$, thereby enabling us to estimate σ_M^2. In other words, the access to data on C_t helped us identify σ_M^2 in the agent's information set, which was not possible with just income data alone. ■

Problem 2.5. Using the posted panel data sets on income and consumption, estimate the parameters of $[A]$ and $U\left(C_t\right)$.

Solution. Skipped. ■

3 RIP to HIP

Read the posted handout, "RIP to HIP," based on Hyrshko.

Problem 3.1. Define HIP and RIP. What information is assumed to be known to the agent in each model? How would you test this information assumption from income data alone?

Solution. Consider the following model for individual labor income:

$$\log(Y_{iht}) = \alpha_t + \gamma_t' X_{iht} + y_{iht}, \tag{A}$$

where Y_{iht} is labor income of individual i with h years of labor experience in year t, α_t is some time-fixed effect, X_{iht} are observable covariates, and y_{iht} is the idiosyncratic income not explained by the observables.

Suppose we can characterize the stochastic process of idiosyncratic income under the following model:

$$y_{iht} = \underbrace{\alpha_i + \beta_i h}_{\text{heterogeneity}} + \underbrace{p_{iht} + \tau_{iht}}_{\text{risk}} + \underbrace{u_{iht,me}}_{\text{measurement error}}, \tag{B}$$

where

- α_i: individual i's initial level of income
- β_i: individual i's growth rate of income
- $p_{iht} = p_{iht-1} + \xi_{iht}$: the permanent stochastic component of income, where ξ_{iht} is a mean-zero shock to the permanent component
- $\tau_{iht} = \theta(L)\epsilon_{iht}$: the (transitory) stochastic component of income, where $\theta(L)$ is a moving average polynomial in L and ϵ_{iht} is a mean-zero shock to the transitory component
- $u_{iht,me}$: a mean-zero measurement error + purely transitory shock

We define HIP (Heterogeneous Income Profiles) to be the idiosyncratic income process when we assume $p_{iht} = 0$ for all t:

$$y_{iht} = \underbrace{\alpha_i + \beta_i h}_{\text{heterogeneity}} + \underbrace{\tau_{iht}}_{\text{risk}} + \underbrace{u_{iht,me}}_{\text{measurement error}},$$

and we define RIP (Restricted Income Profiles) to be the idiosyncratic income process when we assume $\beta_i = 0$:

$$y_{iht} = \underbrace{\alpha_i}_{\text{heterogeneity}} + \underbrace{p_{iht} + \tau_{iht}}_{\text{risk}} + \underbrace{u_{iht,me}}_{\text{measurement error}}.$$

Hence, under HIP model, we assume that there is heterogeneity in income growth rates across individuals which grow deterministically ($\beta_i \neq 0$), while the stochastic component (risk) is persistent but only transitory ($p_{iht} =$

$0, \tau_{iht} \neq 0$). Also, individual i has a perfect knowledge of her own β_i and makes her consumption decisions accordingly. On the other hand, under RIP model, we assume that income growth rates do not grow determinisitcally (so $\beta_i = 0$) but follow a random walk along with a stationary transitory shock ($p_{iht} \neq 0$, $\tau_{iht} \neq 0$). Here, individual i can distinguish the transitory (τ_{iht}) and permanent (p_{iht}) shock.

We can test the information assumption under HIP and RIP from income data alone using the approach in Hryshko (2012), which:

1. Conducts a Monte Carlo simulation study on small unbalanced panels replicated from the income growth data from the Panel Study of Income Dynamics (PSID) to identify and estimate the relevant parameters under the true model (B). The identified parameters relevant to each model are:
   - ▷ HIP: σ^2_β
   - ▷ RIP: σ^2_ξ
   - ▷ Unidentified under both model: σ^2_ϵ, θ, and $\sigma^2_{u,me}$
   - ▷ To see this, in Hryshko (2012) where it assumes $\tau_{hit} = (1 + \theta L)\epsilon_{iht}$,
     - * One can obtain y_{iht} and $\Delta y_{iht} = y_{iht} - y_{iht-1}$ by regressing (A) and taking first differences
     - * Since Δy_{iht} is expressed as

       $$\Delta y_{iht} = \beta_i + \xi_{iht} + \theta(L)\Delta\epsilon_{iht} + \Delta u_{iht,me},$$

       the autocovariances $\gamma_k = E[\Delta y_{iht} \Delta y_{iht-k}]$ are expressed as:

       $$\begin{aligned} \gamma_0 &= \sigma^2_\xi + \sigma^2_\beta + \left(1 + (1-\theta)^2 + \theta^2\right)\sigma^2_\epsilon + 2\sigma^2_{u,me} \\ \gamma_1 &= \sigma^2_\beta - (\theta - 1)^2\sigma^2_\epsilon - \sigma^2_{u,me} \\ \gamma_2 &= \sigma^2_\beta - \theta\sigma^2_\epsilon \\ \gamma_k &= \sigma^2_\beta, \quad k \geq 3 \end{aligned}$$

     - * As shown in Hryshko, we can identify σ^2_β with γ_k for $k \geq 3$ and σ^2_ξ with

       $$\sigma^2_\xi = E\left[\Delta y_{iht} \sum_{j=-2}^{2} \Delta y_{iht+j}\right] - 5\sigma^2_\beta = \sum_{j=-2}^{2} \gamma_j - 5\sigma^2_\beta,$$

       but cannot identify σ^2_ϵ, θ, nor $\sigma^2_{u,me}$ individually.
2. Tests the assumptions under each model on the estimates of the identified parameters from (1) using the actual PSID data.

Problem 3.2. How can you distinguish these two earnings processes using only income data?

Solution. Yes. As shown in 3a, we can estimate σ^2_ξ, σ^2_β, and γ_k's. To distinguish the two earnings processes, we can test $\sigma^2_\beta = 0$ to reject HIP or $\sigma^2_\xi = 0$ to reject RIP.

Problem 3.3. How does access to consumption data help?

Solution. Access to consumption data helps us to also identify and estimate σ_ϵ^2, θ, and $\sigma_{u,me}^2$, which we were not able to when using the income data alone. To see this,

- ▷ Under RIP with MA(1), change in consumption, Δc_{iht} follows:

$$\Delta c_{iht} = \xi_{iht} + \alpha_T \epsilon_{iht},$$

so that

$$var(c_{iht}) = var(c_{iht-1}) + \sigma_\xi^2 + \alpha_T^2 \sigma_\epsilon^2$$

- ▷ Whereas under HIP with AR(1), it follows

$$\Delta c_{iht} = \alpha_T \epsilon_{iht},$$

so that

$$var(c_{iht}) = var(c_{iht-1}) + \alpha_T^2 \sigma_\epsilon^2$$

- ▷ Since σ_ξ^2 and α_T^2 were identified from the income data and $var(c_{iht})$ can be directly identified from the consumption data, we can identify σ_ϵ^2 individually from the above relationships.
- ▷ Having identified σ_ϵ^2, we can then identify θ and $\sigma_{u,me}^2$ from the equations for γ_0 and γ_1 obtained in 3(a).

■

Problem 3.4. Why does any of this matter for determining agent welfare?

Solution. To be able to pinpoint which model (HIP or RIP) is appropriate to explain idiosyncratic income helps policymakers to effectively address consumption inequality. If the HIP model holds, then an appropriate policy may be to subsidize human capital investments by the disadvantaged who have low income growth rates. However, if the RIP model holds, then it may be of best interest to educate the public about the importance of risk-sharing and encourage self-insurance through products offered by the financial system such as insurance policies, stocks, bonds, and etc.

Furthermore, knowing which model is appropriate allows us to infer about the welfare costs of business cycles. Under RIP, the welfare loss of business cycles is huge as consumption reacts significantly to the permanent shock to income, whereas under HIP, the welfare loss is not so big as shocks to income are not permanent so agents do not have to drastically alter their consumption decisions.

■

4 Earnings Dynamics

Consider the following model of earnings dynamics based on Heckman and Robb (*Journal of Econometrics*, 1985): Earnings evolve as

$$\begin{aligned}
y_{it} &= \alpha_0 + \alpha_1 X_t + \alpha_2 D_k + U_{it} \\
U_{it} &= P_{it} + \tau_{it}, & t = 1, ..., T \\
P_{it} &= \rho P_{i,t-1} + \lambda_{it} + \eta_i \\
B(L)\tau_{it} &= A(L)\varepsilon_{it} \\
E[\lambda_{it}] &= 0 & \lambda_{it} \perp\!\!\!\perp \eta_i, (\varepsilon_{it})_{t=1}^T \\
E[\varepsilon_{it}] &= 0 \\
E[\eta_i] &= 0 \\
B(L) &= 1 - \rho_\tau L \\
A(L) &= 1 + \theta L
\end{aligned}$$

Income maximizing agents are able to choose to participate in a program at times $1 < k < \infty$.
If they participate, they get (*ex post*) a boost of $\alpha > 0$ per period in their earnings after period k. Assume $T \to \infty$ (for simplicity) so $\frac{\alpha}{r}$ is the *ex post* gain in terms of present value of income maximization. r is the nonstochastic interest rate.
Participation in the program costs Y_{ik} (forgone earnings) and some psychic costs $C = Z\delta + V$.

Problem 4.1. Justify the decision role for participation in the program:

$$D_k = 1(E_k[\frac{\alpha}{r} - C - Y_{ik}] > 0)$$

where E_k means information at k. Characterize the possible agent information sets under two conditions:

- ▷ α is constant.
- ▷ α_i varies among individuals, assuming either Y_{ik} is known or is not known at k but C is known.

Solution. Let $PV_k(0)$ be the expected net present value of not being trained in the perspective of date k and let $PV_k(1)$ be the expected net present value of being trained at date k. Then,

$$PV_k(0) = E_{k-1}[\sum_{j=0}^{\infty}\left(\frac{1}{1+r}\right)^j Y_{i,k+j}]$$
$$PV_k(1) = E_{k-1}[\sum_{j=1}^{\infty}\left(\frac{1}{1+r}\right)^j (Y_{i,k+j}+\alpha) - C]$$
$$PV_k(1) - PV_k(0) = E_{k-1}[\sum_{j=1}^{\infty}\left(\frac{1}{1+r}\right)^j \alpha - Y_{ik} - C] = E_{k-1}[\frac{\alpha}{r} - Y_{ik} - C].$$

Thus, if training at date k boosts lifetime earnings, it is natural to think an agent will choose to train at that date. Therefore, we are justified in saying

$$D_k = 1(E_k[\frac{\alpha}{r} - C - Y_{ik}] > 0).$$

The agent information sets are discussed below for the different conditions:

- ▷ α **is constant:** $\sigma(I_A) = \{\alpha, r, C, \{Y_{it}, U_{it}, P_{i,t}, \tau_{i,t}, \lambda_{i,t}, \varepsilon_{i,t}, X_{it}\}_{t=0}^{k-1}, \eta_i, D_k\}$.
- ▷ α_i **varies and** Y_{ik} **is known**: $\sigma(I_A) = \{\alpha_i, r, C, Y_{ik}, \{Y_{it}, U_{it}, P_{i,t}, \tau_{i,t}, \lambda_{i,t}, \varepsilon_{i,t}, X_{it}\}_{t=0}^{k-1}, \eta_i, D_k\}$.
- ▷ α_i **varies and** Y_{ik} **is unknown**: $\sigma(I_A) = \{\alpha_i, r, C, \{Y_{it}, U_{it}, P_{i,t}, \tau_{i,t}, \lambda_{i,t}, \varepsilon_{i,t}, X_{it}\}_{t=0}^{k-1}, \eta_i, D_k\}$. However, instead of α_i, they may only know $E_{k-1}[\alpha_i]$.

■

Problem 4.2. For each information set, characterize $Cov(D_k, Y_t)$, $1 < t < \infty$.

Solution. Note that

$$Cov(D_k, Y_{it}) = \alpha_1 Cov(D_k, X_{it}) + \alpha_2 Var(D_k) + Cov(D_k, U_{it}).$$

Let $p = P(D_k = 1)$. Since D_k is a bernoulli random variable we have

$$Var(D_k) = p(1-p).$$

Also,

$$Cov(D_k, U_{it}) = E[D_k U_{it}] - E[D_k]E[U_{it}] = E[D_k U_{it}]$$

since

$$\begin{aligned} E[U_{it}] &= \rho_t E[P_t] + E[\lambda_t] \\ &= \rho_t E[P_t] && (E[\lambda_t] = 0) \\ &= 0 && (P_t \textbf{ is stationary}). \end{aligned}$$

We can further simplify using the total law of probability:

$$Cov(D_k, U_{it}) = E[D_k U_{it}] = E[U_{it}|D_k = 1]P[D_k = 1] = pE[U_{it}|D_k = 1].$$

Thus we so far have

$$Cov(D_k, Y_{it}) = \alpha_1 Cov(D_k, X_{it}) + \alpha_2 p(1-p) + pE[U_{it}|D_k = 1]$$

If we further assume X_{it} is exogeneous we have $Cov(D_k, X_{it}) = 0$ and then

$$Cov(D_k, Y_{it}) = \alpha_2 p(1-p) + pE[U_{it}|D_k = 1]$$

■

Problem 4.3. For each information set, discuss identification of α (or $E[\alpha]$) and the MTE both the fixed coefficient and for random coefficient version using

▷ IV

▷ Matching

▷ Two-step selection models

▷ MLE

Assume (only for convenience) normally distributed errors.

Solution. The identification steps are below.

▷ α

* **IV**: As long as Z, X_{it} are exogeneous, we can use these as instruments and estimate α by 2SLS. When α is constant, we also recover MTE since $MTE = LATE$.
* **Matching**: Here, we would need X_{it} to be enough to ensure treatment is independent of potential outcomes. If that is the case, then running an regression of our specification should give us α. However, if foregone earnings Y_{ik} (an unobservable) is in the information set of the agent, conditional independence will not hold.

  * **Two-step selection models**: As long as the model is well specified, we can estimate α. Further, we can recover MTE since MTE$= \alpha$.
  * **MLE**: If we know the joint density of (U_{it}, V_i), then MLE can be used to estimate α. Just like in IV and the selection model, we have MTE$= \alpha$ and thus it is recovered.

- $E[\alpha_i]$.

  * **IV**: As long as Z, X_{it} are exogeneous, we can use these as instruments and run 2SLS. However, since the treatments are heterogeneous, we will not estimate $E[\alpha_i]$ but

$$E[\alpha_i|Compliers].$$

  LATE is a weighted average of MTE's and so α_i or the MTE is not recovered pointwise.
  * **Matching**: Here, we would need X_{it} to be enough to ensure treatment is independent of potential outcomes. If that is the case, then running an regression of our specification will give us $E[\alpha_i|X]$. If we have full common support we can retrieve $E[\alpha_i] = E[E[\alpha_i|X]]$. MTE is not recovered here since we do not recover the distribution of α_i.
  * **Two-step selection models**: As long as the model is well specified, we can estimate $E[\alpha_i]$ and recover MTE.
  * **MLE**: If we know the joint density of (U_{it}, V_i), then MLE can be used to estimate $E[\alpha_i]$ and the MTE.

■

Problem 4.4. Suppose that you have two independent samples taken at the same time. They have identical distributions of regressors and errors. One sample has no access to the program ever. One sample has access to the program only at time period k. How does access to this data improve your ability to identify α or $E[\alpha]$? Consider two cases:

- Both samples are exposed to a common time trend $\mu(t) = fe^{-gt}$.
- $g_{treatment} > g_{control}$.
- Compare the identification different achieved from difference in difference estimators with that of the estimators discussed in (c).

Solution. When we have access to panel data we can use the difference-in-difference estimator to estimate α or $\mathbb{E}[\alpha]$. Notice that we can recover both α and the unconditional expectation $\mathbb{E}[\alpha]$, which was not necessarily the case in part (c). If both samples are exposed to a common time trend the the second difference will eliminate this common trend. If there are different trends for both samples, we in general cannot identify α or $\mathbb{E}[\alpha]$ from the second difference – although we could if we know both g_{treated} and g_{control}, which in general we do not.

■

5 Marschak's Maxim

What is Marschak's Maxim? Discuss in the context of the two equation model in the simultaneous causality handout restriction. (Hint: ISI, 2008, Econometric Causality.)

Solution. Marschak (1953) noted that the answers to many policy questions did not require such detailed knowledge – a combination of parameters is all that is necessary and, moreover, it is often possible to identify the desired combination without identifying the individual components. The immediate implication is that forecasting or evaluating policies may only require partial knowledge of the full simultaneous equations system.

To see this in the simulatneous equation setting, consider:

$$[2a] : Y_1 = \alpha_1 + \gamma_{12} Y_2 + \beta_{11} X_1 + \beta_{12} X_2 + U_1$$
$$[2b] : Y_2 = \alpha_2 + \gamma_{21} Y_1 + \beta_{21} X_1 + \beta_{22} X_2 + U_2$$

and trasform them into the reduced-form counterparts:

$$Y_1 = \pi_{10} + \pi_{11} X_1 + \pi_{12} X_2 + \epsilon_1$$
$$Y_2 = \pi_{20} + \pi_{21} X_1 + \pi_{22} X_2 + \epsilon_2$$

Mechanically, we have:

$$\pi_{11} = \frac{\beta_{11} + \gamma_{12}\beta_{21}}{1 - \gamma_{12}\gamma_{21}}$$
$$\pi_{12} = \frac{\beta_{12} + \gamma_{12}\beta_{22}}{1 - \gamma_{12}\gamma_{21}}$$
$$\pi_{21} = \frac{\beta_{21} + \gamma_{21}\beta_{11}}{1 - \gamma_{12}\gamma_{21}}$$
$$\pi_{22} = \frac{\beta_{22} + \gamma_{12}\beta_{12}}{1 - \gamma_{12}\gamma_{21}}$$

The maxim tells us that if we are interested in the effect of X_1, X_2 on Y_1, Y_2, then there is really no need to fully estimate the simulatenous causality models, which often requires strong assumptions.

■