1 Estimation of PS9 Q2

For the model of PS9 Q2, use the posted panel data on income and consumption and estimate the identified parameters of earnings equation:

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

and for $U(C_t)$:

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

Observations across persons are assumed to be independent. For each dataset, use GMM to compute the identified parameters using the sample moments. Review the assumptions used in GMM and explain how you can use the available moments on consumption and income in a GMM procedure and how you weight them in securing the estimates.

Solution. We first revisit the equations from the previous problem set and then examine the assumptions behind the GMM procedure. We will be implementing GMM using a two-step variance-covariance estimator.

Equations for Estimation Recall that we had the following moments available from PS9 Q2:

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

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

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

$$[4] : \text{Cov}[\Delta C_t, \Delta Y_t] = \frac{r}{(1+r) - (1+r)^{-(T-t)}} \left[1 + \left(\frac{\theta}{1+r}\right)^2\right] \sigma_\omega^2 + \left(\frac{r}{(1+r) - (1+r)^{-(T-t)}}\right) \sigma_\lambda^2$$

where the unknowns are: $\rho, \theta, \sigma_\lambda^2, \sigma_\omega^2$. We will estimate this using two-step variance-covariance estimator.

- ▷ Notice that equations [3] and [4] have t in the equations, so we compute the cross-sectional variance across agents at each point in time. Since we have 45 time periods and lose one degree of freedom in computing ΔC_t, we have 44 moments for each of [3] and [4].

Assumptions for GMM Define $m(x)$ as the 4×1 vector of moments from the real data, where $x = N \times K$ matrix of data with K columns and N observations. Furthermore, define $m(x|\boldsymbol{\theta})$ as the vector of 4 moments from our model that correspond to the real-world moment data. In this case, we have

$$m\left(x|\boldsymbol{\theta}\right) = m\left(x|\rho, \theta, \sigma_\lambda^2, \sigma_\omega^2\right) = \begin{bmatrix} [1] \\ [2] \\ [3] \\ [4] \end{bmatrix}$$

where $[1]-[4]$ are the equations from above. Under this setup, the GMM estimator, in general, can be represented as

$$\hat{\boldsymbol{\theta}}_{GMM} = \arg\min_{\boldsymbol{\theta}} e\left(x|\boldsymbol{\theta}\right)^\top W e\left(x|\boldsymbol{\theta}\right)$$

where $e\left(x|\theta\right)$ is the moment error function defined as the percentage difference in the vector of moments from the data moments:

$$e\left(x|\boldsymbol{\theta}\right) \equiv \frac{m\left(x|\boldsymbol{\theta}\right) - m\left(x\right)}{m\left(x\right)}$$

and W is the 4×4 weighting matrix.

Two-step Estimation We implement GMM using a two-step variance-covariance estimator.

1. We first estimate the vector using an identity weighting matrix:

$$\hat{\boldsymbol{\theta}}_{1,GMM} : \arg\min_{\boldsymbol{\theta}} e\left(x|\boldsymbol{\theta}\right)^\top I e\left(x|\boldsymbol{\theta}\right)$$

2. Construct the covariance matrix from residuals:

$$\hat{\Omega}_2 = \frac{1}{N} e\left(x|\hat{\boldsymbol{\theta}}_{1,GMM}\right) e\left(x|\hat{\boldsymbol{\theta}}_{1,GMM}\right)^\top$$

3. Re-estimate using the weighting matrix W_2 defined as

$$W_2 \equiv \hat{\Omega}_2^{-1}$$

to obtain

$$\hat{\boldsymbol{\theta}}_{2,GMM} : \min_{\boldsymbol{\theta}} e\left(x|\boldsymbol{\theta}\right)^\top W_2 e\left(X|\boldsymbol{\theta}\right)$$

Estimation Results We estimate using Python and obtain the following results:

▷ Dataset 1: $\rho = 0.8123, \theta = 0.7599, \sigma_\lambda = 0.0000, \sigma_\omega = 0.0000$

▷ Dataset 2: $\rho = 0.8120, \theta = 0.7605, \sigma_\lambda = 0.0000, \sigma_\omega = 0.0000$

▷ Dataset 3: $\rho = 0.8345, \theta = 1.2637, \sigma_\lambda = 0, \sigma_\omega = 0.0883$

■

2 Estimation of PS9 Q3

For the model of Question 3 (from Problem Set 9), use the posted data set on income and consumption and estimate the parameters of the two models in Hyrshko's paper. (Hint: The data are not Hyrshko's.) Estimate and report the life cycle profiles of permanent and transitory components of income.

Solution. Recall from last problem set, the encompassing model was given by

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

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

The autocovariances, $\gamma_k = E[\Delta y_{it}\Delta y_{it-k}]$, in Hryshko (2012) where it assumes MA(1), $\tau_{it} = (1+\theta L)\epsilon_{it}$, were given by:

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

As provided in the Appendix of Hryshko (2012), the empirical moments of the autocovariances, m^d is given by:

$$m^d = vech(\sum_{i=1}^{N} \tilde{y}_i\tilde{y}_i')/N_{tt'},$$

where $\tilde{y}_i = (\Delta y_{i2}, \Delta y_{i3}, ..., \Delta y_{iT})$ (first differences in the idiosyncratic income of individual i) and $N_{tt'}$ is a vector of row dimentsion $T(T-1)/2$ with each element being the number of people contributing to estimating the autocovariances. Also, as done in Hryshko (2012), we consider three scenarios,

- ▷ Leave measurement errors unidentified

▷ Set $\sigma_{u,me} = 0$

▷ Set $\sigma_{u,me} = 0.25\sigma_\epsilon$

Let

$$m(\theta) = \arg\min \quad [m(\theta) - m^d]' I [m(\theta) - m^d]$$

be the miminum-distance estimator for autocovariances we are interested in, with which we can use the analytical equations above to identify and estimate σ_β^2 and σ_ξ^2 under both models. Although θ and σ_ϵ^2 are not identified as explained in the previous problem set (recall that we set $\sigma_{u,me}^2 = 0$ so we are not interested in it here), we can identify $(1+\theta^2)\sigma_\epsilon^2$ from γ_1, γ_2, and γ_k:

$$(1+\theta^2)\sigma_\epsilon^2 = 3\gamma_k - \gamma_1 - 2\gamma_2$$

The parameter estimates under each model are given by:

Table 1: Estimated Parameters

	HIP				
	σ_ξ^2	σ_β^2	θ	σ_ϵ^2	$\sigma_{u,me}^2$
$\sigma_{u,me}$ Unidentified	0	0.0004472301	0.2124076652	0.0277547314	0.0274249000
$\sigma_{u,me} = 0$	0	0.0004467711	0.1056152767	0.0558076215	0
$\sigma_{u,me} = 0.25\sigma_\epsilon$	0	0.0004467732	0.1122948164	0.0524877015	0.00328048134
	RIP				
	σ_ξ^2	σ_β^2	θ	σ_ϵ^2	$\sigma_{u,me}^2$
$\sigma_{u,me}$ Unidentified	0.006572908	0	0.155239260	0.028731539	0.022739318
$\sigma_{u,me} = 0$	0.006572903	0	0.086140329	0.051779061	0
$\sigma_{u,me} = 0.25\sigma_\epsilon$	0.006581531	0	0.091511136	0.048703528	0.0030439705

Notes: The table reports the identified parameter values from minimum-distance estimation under each model, with varying constraints on $\sigma_{u,me}$.

To decompose the variance of income into permanent and transitory shock components, note that

▷ Under HIP

$$\begin{aligned}
y_{iht} &= \alpha_i + \beta_i h + \tau_{iht} \\
\Rightarrow var[y_{iht}] &= var[\alpha_i] + var[\beta_i h] + var[\tau_{iht}] \\
&= var[\alpha_i] + \sigma_\beta^2 h^2 + var[\epsilon_{iht} + \theta\epsilon_{ih-1t-1}] \\
&= var[\alpha_i] + \sigma_\beta^2 h^2 + (1+\theta^2)\sigma_\epsilon^2 && \because \epsilon_{iht} \text{ stationary}
\end{aligned}$$

▷ Under RIP

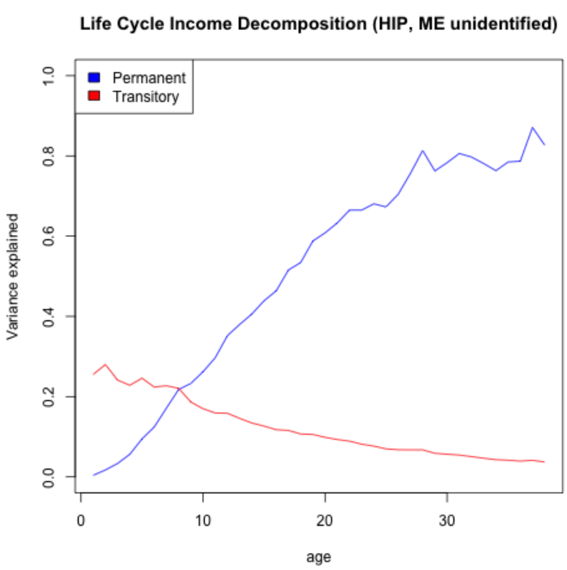
$$\begin{aligned} y_{iht} &= \alpha_i + p_{iht} + \tau_{iht} \\ \Rightarrow var[y_{iht}] &= var[p_{iht}] + var[\tau_{iht}] \\ &= var[\alpha_i] + var[p_{ihnt-1} + \xi_{iht}] + var[\epsilon_{iht} + \theta\epsilon_{ih-1t-1}] \\ &= var[\alpha_i] + var[\xi_{i1t-h+1} \cdots + \xi_{ih-1t-1} + \xi_{iht}] + var[\epsilon_{iht} + \theta\epsilon_{ih-1t-1}] \\ &= var[\alpha_i] + \sigma_\xi^2 h + (1+\theta^2)\sigma_\epsilon^2 \qquad \because \xi_{iht} \text{ i.i.d and } \epsilon_{iht} \text{ stationary} \end{aligned}$$

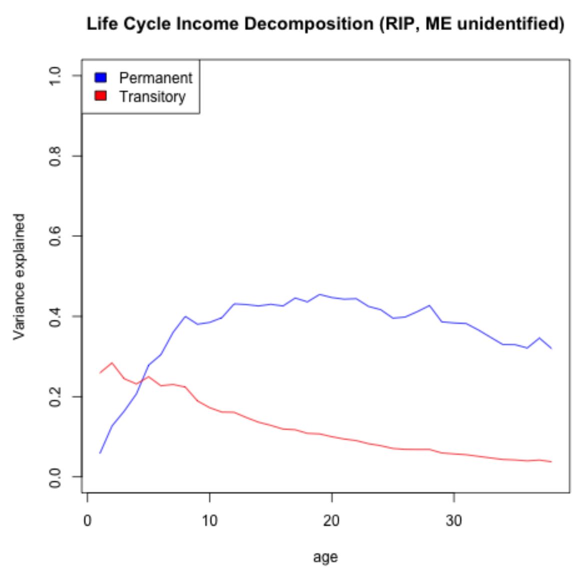
Recall that we showed earlier $(1+\theta^2)\sigma_\epsilon^2$ is identifiable when $\sigma_{u,me} = 0$.

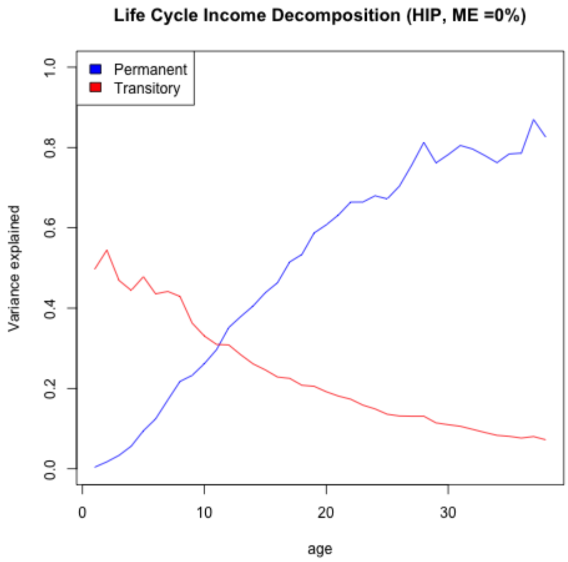
To decompose the life cycle profiles of income into permanent and transitory components,

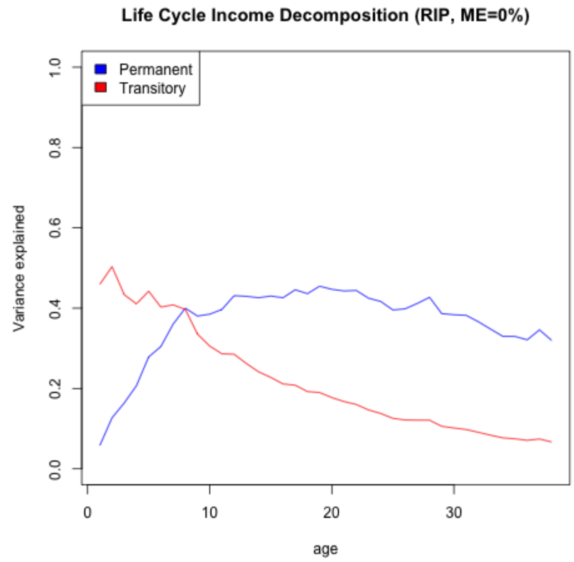
1. We estimate the cross-sectional variance within each age, $v\hat{a}r[y_{iht}]$.
2. Using the estimates we obtained previously,
   - ▷ We let the permanent component of income at age h be $\hat{\sigma}_\beta^2 h^2$ and $\hat{\sigma}_\xi^2 h$ under HIP and RIP, respectively.
   - ▷ We let the transitory component of income at age h be $(1+\hat{\theta})\hat{\sigma}_\epsilon^2$ under both model, since transitory shocks are stationary.
3. We divide the permanent and transitory component by $v\hat{a}r[y_{iht}]$ to calculate the variation of income explained by both components.

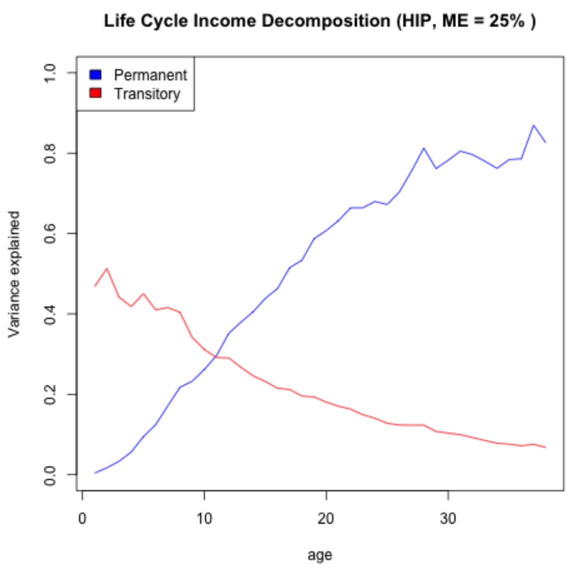
The following figures plot the life cycle profiles of permanent and transitory components of income under HIP and RIP, for the three cases of treating $u_{iht,me}$:

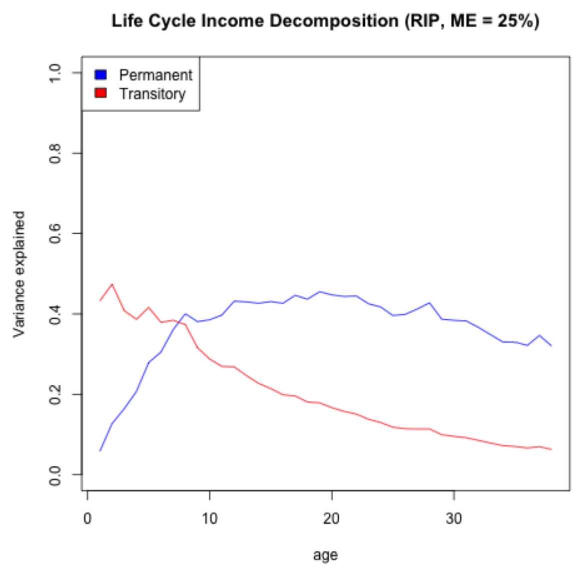




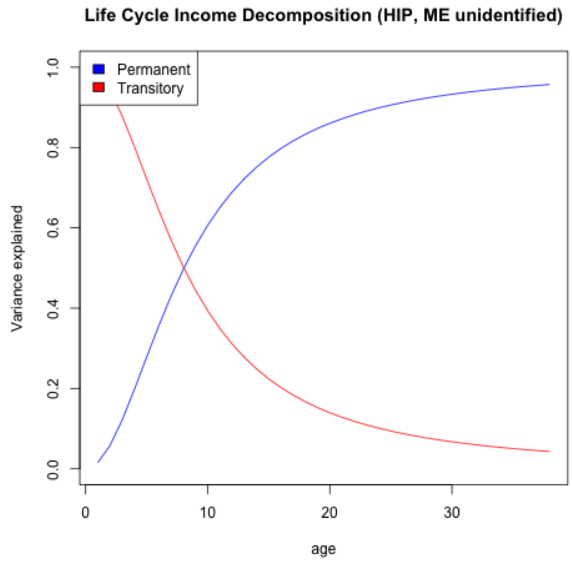


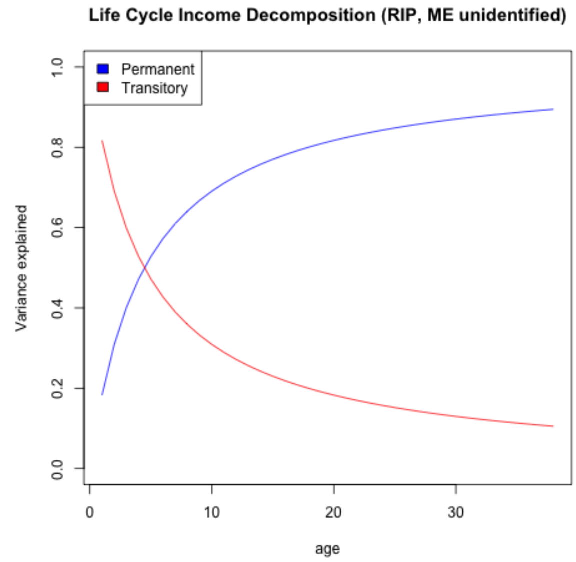


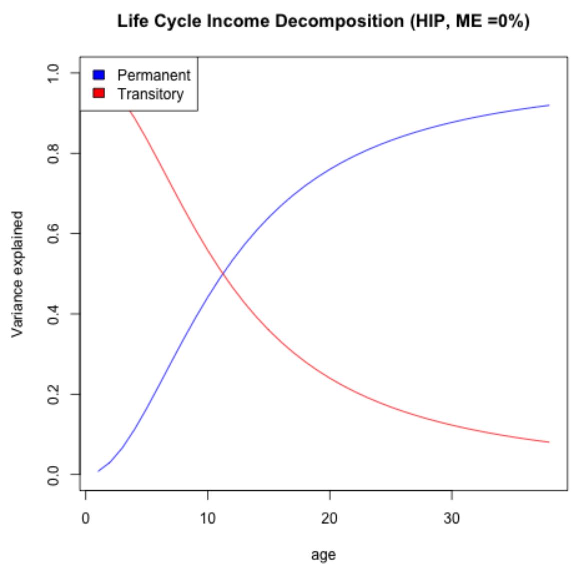


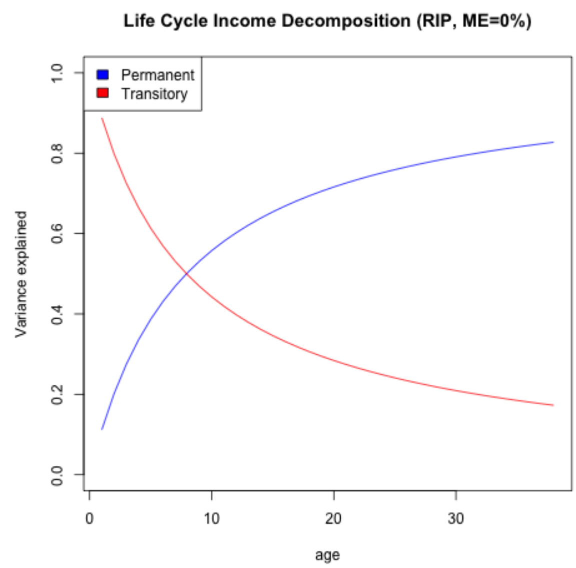


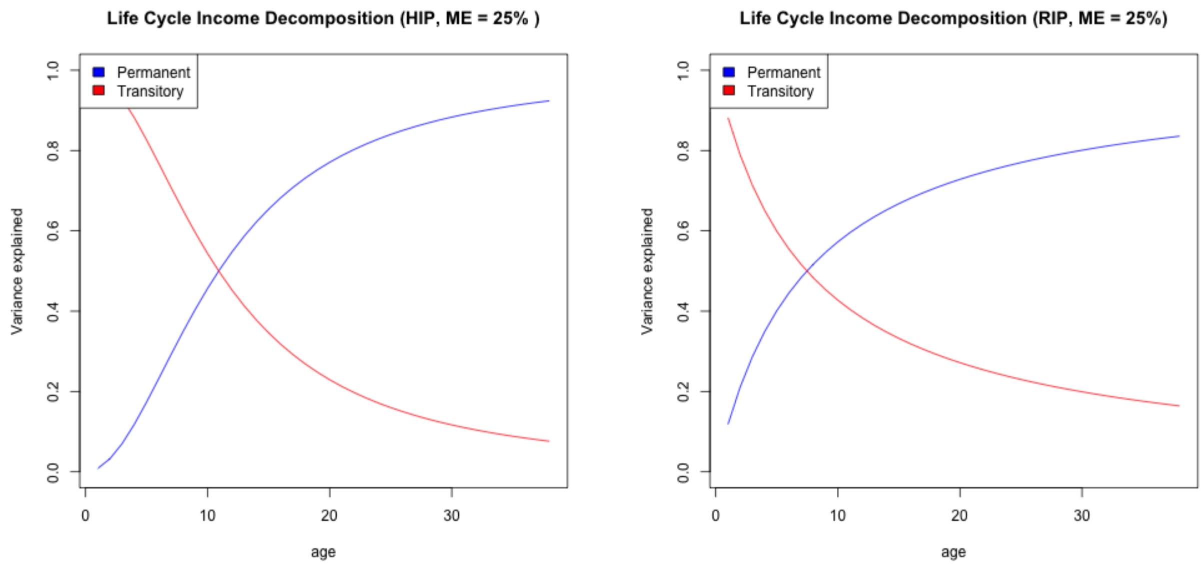
The following figures also plot the decomposition of the permanent and transitory components only as the fraction of their sum and bypassing $\hat{var}[y_{iht}]$ (hence, we obtain smoother plots),


Life Cycle Income Decomposition (HIP, ME unidentified)
Permanent
Transitory
Variance explained
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
age


Life Cycle Income Decomposition (RIP, ME unidentified)
Permanent
Transitory
Variance explained
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
age


Life Cycle Income Decomposition (HIP, ME =0%)
Permanent
Transitory
Variance explained
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
age


Life Cycle Income Decomposition (RIP, ME=0%)
Permanent
Transitory
Variance explained
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
age


Life Cycle Income Decomposition (HIP, ME = 25%)
Permanent
Transitory
Variance explained
age
0.0
0.2
0.4
0.6
0.8
1.0
0
10
20
30
Life Cycle Income Decomposition (RIP, ME = 25%)
Permanent
Transitory
Variance explained
age

■

3 Estimation of PS9 Q4

For the model of Question 4 in Problem Set 9, use the posted data sets and estimate $\alpha, E(\alpha), g, f$ and as many of the other parameters of the model as possible using the four methods discussed in part (c) of Problem Set 9 and the difference in differences estimator. Compare the performance of the difference-in-difference estimator with these estimators (bias, variance).

Solution. Below I provide the parameters we can estimate through our four means of estimation: IV, matching, MLE, and sample selection. I also compare them to the differences-in-differences estimator.

▷ **IV**: As long as we have exogeneity and relevance of our instrument Z, IV/2SLS will identify α if treatments are homogeneous or $E[\alpha_i|Compliers]$ if treatments are heterogeneous. The results are below

Table 2: Instrumental Variables Estimation

	(1)	(2)	(3)
	Sample 1	Sample 2	Sample 3
Treated	0.203***	0.160***	0.536***
	(0.0337)	(0.0400)	(0.125)
X	0.0380***	0.0134	0.0535*
	(0.0112)	(0.0108)	(0.0251)
N	2000	2000	2000
R-sq	0.119	0.110	.

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

▷ **Matching**: If CIA holds, OLS will consistently estimate α or $E[\alpha_i]$. The estimates are below. We use the psmatch2 command in Stata to get the estimates.

Table 3: Matching Estimation

	(1)	(2)	(3)
	Sample 1	Sample 2	Sample 3
psmatch2: Treatment assignment	0.256***	0.246***	0.242***
	(0.0156)	(0.0146)	(0.0360)
N	2000	2000	2000
R-sq	0.119	0.124	0.022

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

▷ **MLE**: Assuming we know the joint distribution of (U_{it}, V_i) then we can use MLE to estimate α. We rewrite our model as

$$\begin{aligned}
Y_{it} &= \alpha_0 + \alpha_1 X_t + \alpha_2 D_k + U_{it} \\
U_{it} &= P_{it} + \tau_{it} \\
\tau_{it} &= \frac{1+\theta L}{1+\rho_\tau L}\epsilon_{it} \\
P_{it} &= \rho P_{i,t-1} + \lambda_{i,t} + \eta_{i,t} \\
0 &= E[\lambda_{it}] = E[\epsilon_{it}] = E[\eta_{it}].
\end{aligned}$$

This model specification gives us

$$\begin{aligned}
Y_{it} = &(\rho + \rho_\tau) y_{i,t-1} + \rho\rho_\tau y_{i,t-2} \\
&- \\
&+ \alpha_0(1 - \rho - \rho_\tau + \rho\rho_\tau) \\
&+ \alpha_1 X_t - \alpha_1(\rho + \rho_\tau) X_{t-1} + \rho\rho_\tau X_{t-2}
\end{aligned}$$

To estimate this, we will assume agents have perfect foresight and perfectly knows foregone earnings. With this assumption, MLE can estimate $\{\alpha_0, \alpha_1, \alpha_2, \sigma_V, \delta, \sigma_U\}$. The estimates are provided in the table below.

Table 4: Maximum Likelihood Estimation

	(1)	(2)	(3)
	Sample 1	Sample 2	Sample 3
α_0	2.229718	2.026524	1.930320
α_1	-0.040298	0.021210	0.025360
α_2	0.517121	0.502608	0.538602
σ_U	0.320858	0.268297	0.780008
σ_V	5.649151	1.469571	5.488767
δ	-1.167382	0.244819	-0.376760

▷ **Selection Model**: Here we use the two step selection model and the Heckman correction to estimate α. Specifically, we estimate the following probit regression for the first stage:

$$D_i = \gamma_0 + \gamma_1 X_{it} + \gamma_2 Z_i$$

and then use the fitted values to obtain the inverse mills ratio, and control for the inverse Mills ratio in our second stage and so our second stage is

$$Y_{it} = \alpha_0 + \alpha D_k + \alpha_1 X_t + \delta \textbf{Inverse Mill's Ratio} + U_{it}.$$

The results are in the table below for the coefficients on "Treated".

Table 5: Selection Estimation

	(1) Sample 1	(2) Sample 2	(3) Sample 3
Treated	0.273***	0.259***	0.217***
	(0.0176)	(0.0157)	(0.0376)
X	0.0380***	0.0135	0.0533*
	(0.0112)	(0.0107)	(0.0246)
Inverse Mills Ratio	0.137	0.187*	-0.573*
	(0.0722)	(0.0842)	(0.243)
N	2000	2000	2000
R-sq	0.126	0.127	0.027

Standard errors in parentheses

$^{*}\ p < 0.05$, $^{**}\ p < 0.01$, $^{***}\ p < 0.001$

▷ **Differences-in-Differences**: The DiD estimators are below and correspond to the "Treated" variable. We will have imperfect compliance and so our estimate is the "intend to treat" estimate, which in general will not be $E[\alpha_i]$.

We see that these results are quite close across samples. These results are quite different from IV and MLE, but the Heckman correction and matching results are consistent with the DiD.

■

Table 6: Differences-in-Differences Estimation

	(1) Sample 1	(2) Sample 2	(3) Sample 3
Treated	0.257***	0.247***	0.242***
	(0.0133)	(0.0126)	(0.0359)
treated_sample	0.00172	0.000747	-0.0460
	(0.0102)	(0.0102)	(0.0280)
X	0.0350***	0.0258***	0.0428**
	(0.00548)	(0.00519)	(0.0144)
N	6000	6000	6000
R-sq	0.097	0.103	0.011

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

4 Extension of PS9 Q2

Consider how to estimate the parameters of the utility function

$$U(C_t) = AC_t^{\gamma}, \quad \gamma < 1$$

for the earnings process governed by the earnings process in

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

Report estimates for the posted datasets.

Solution. Note that the provided utility function implies

$$\begin{aligned} U'(C_t) &= A\gamma C_t^{\gamma-1} \\ U''(C) &= A\gamma(\gamma-1) C_t^{\gamma-2} \end{aligned}$$

which implies

$$-C_t \frac{U''(C_t)}{U'(C_t)} = -C_t \frac{A\gamma(\gamma-1) C_t^{\gamma-2}}{A\gamma C_t^{\gamma-1}} = 1-\gamma$$

i.e. constant relative risk aversion. With this in mind, we can proceed similarly as before.

Agent's Problem The agent's maximization problem is

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

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:

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

- ▷ We can solve this problem using the Bellman approach. Writing the sequence problem as a Bellman equation:

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

▷ 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:

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

▷ Since $U'\left(C_t\right)=A\gamma C_t^{\gamma-1}$, we have:

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

which yields:

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

Propensity to Consume in Response to Risk Unlike the previous case with quadratic consumption, we **cannot** derive an analytical expression for the "consumption function" and hence the value of the propensity to consume in response to risk (income shocks) is not easily derivable. Therefore, we must estimate the Euler equation directly.

Moment Condition Recall that we had:

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

Dividing each side by $C_t^{\gamma-1}$, we have

$$[1]:\mathbb{E}_t\left[\left(\frac{C_{t+1}}{C_t}\right)^{\gamma-1}\right]=1$$

where γ is the parameter of our interest.

GMM Estimation Estimating the parameters using GMM yields the following results

▷ For dataset 1:

```
GMM estimation

Number of parameters =   1
Number of moments    =   1
Initial weight matrix: Identity                  Number of obs   =      8,800
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/gamma	.9957491	.0985072	10.11	0.000	.8026786	1.18882

Instruments for equation 1: _cons

▷ For dataset 2:

```
GMM estimation

Number of parameters =    1
Number of moments    =    1
Initial weight matrix: Identity                    Number of obs    =      8,800
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/gamma	.6546887	.9542119	0.69	0.493	-1.215532	2.52491

Instruments for equation 1: _cons

▷ For dataset 3:

```
GMM estimation

Number of parameters =    1
Number of moments    =    1
Initial weight matrix: Identity                    Number of obs    =      8,800
GMM weight matrix:     Robust
```

	Coef.	Robust Std. Err.	z	P>|z|	[95% Conf.	Interval]
/gamma	1	2.57e-16	3.9e+15	0.000	1	1

Instruments for equation 1: _cons

■

5 "Short Run vs. Long Run Response"

Answer the embedded question in the Koyck (1954) "Short Run vs. Long Run Response" lag handout

Problem 5.1. Let

- $\triangleright$ $y^* =$ optimal level (Long Run)
- $\triangleright$ $y_t =$ actual level
- $\triangleright$ $y_{t-1} =$ previous level

The objective is to pick current y_t to minimize the following loss function:

$$\underbrace{\frac{1}{2}\phi(y^* - y_t)^2}_{\text{cost of being away from optimum},\phi>0} + \underbrace{\frac{1}{2}\phi(y_t - y_{t-1})^2\eta}_{\text{cost of adjustment},\eta>0}$$

The FOC condition yields:

$$\begin{aligned}\text{FOC} : 0 &= -\phi(y^* - y_t) + \eta(y_t - y_{t-1}) \\ \Rightarrow y_t &= \frac{\phi}{\phi+\eta}y^* + \frac{\eta y_{t-1}}{\phi+\eta}\end{aligned}$$

Consider the partial adjustment model where $y^* = \tau(x)$

$$\frac{\partial y_t}{\partial x} = \underbrace{\frac{\phi}{\phi+\eta}}_{<1}\frac{\partial \tau(x)}{\partial x}.$$

Suppose $y^*_{i,t} = X_i\beta + U_{i,t}$ where $U_{i,t} = \rho U_{i,t-1} + \epsilon_{it}$ and ϵ_{it} are mutually correlated. You observe $Y_{i,t}$ and $Y_{i,t-1}$ and the specified X values.

Is the model identified? Consider three cases

- $\triangleright$ $X_{it} = \bar{X}$: a constant for all i, t
- $\triangleright$ X_{it} varies over people and time
- $\triangleright$ $X_{it} = X_t$ is the same variable over persons and time

---

Solution. The parameters we are interested in identifying are β, ϕ, and η. From the FOC condition, we can write

$$\begin{aligned}Y_{i,t} &= \frac{\phi}{\phi+\eta}Y^*_{i,t} + \frac{\eta}{\phi+\eta}y_{i,t-1} \\ &= \frac{\phi}{\phi+\eta}\beta X_{i,t} + \frac{\eta}{\phi+\eta}Y_{i,t-1} + \frac{\phi}{\phi+\eta}U_{i,t},\end{aligned}$$

where $E[Y_{i,t-1}U_{i,t}] \neq 0$ due to serial correlation. Let

$$\gamma = \frac{\phi}{\phi+\eta},$$
$$1-\gamma = \frac{\eta}{\phi+\eta},$$

then we have:

$$Y_{i,t} = \gamma\beta X_{i,t} + (1-\gamma)Y_{i,t-1} + \gamma U_{i,t}.$$

We can then express $U_{i,t}$ as:

$$\begin{aligned} U_{i,t} &= \frac{1}{\gamma}Y_{i,t} - \beta X_{i,t} - \frac{1-\gamma}{\gamma}Y_{i,t-1} \\ \Rightarrow U_{i,t-1} &= \frac{1}{\gamma}Y_{i,t-1} - \beta X_{i,t-1} - \frac{1-\gamma}{\gamma}Y_{i,t-2}. \end{aligned}$$

Because $U_{i,t} = \rho U_{i,t-1} + \epsilon_{i,t}$, we have:

$$\begin{aligned} U_{i,t} &= \rho U_{i,t-1} + \epsilon_{i,t} \\ &= \frac{\rho}{\gamma}Y_{i,t-1} - \rho\beta X_{i,t-1} - \rho\frac{1-\gamma}{\gamma}Y_{i,t-2} + \epsilon_{i,t}. \end{aligned}$$

Substituting the above expression for $U_{i,t}$ back into the above equation for $Y_{i,t}$ yields:

$$\begin{aligned} Y_{i,t} &= \gamma\beta X_{i,t} + (1-\gamma)Y_{i,t-1} + \gamma\left(\frac{\rho}{\gamma}Y_{i,t-1} - \rho\beta X_{i,t-1} - \rho\frac{1-\gamma}{\gamma}Y_{i,t-2} + \epsilon_{i,t}\right) \\ &= \gamma\beta X_{i,t} + (1-\gamma)Y_{i,t-1} + \rho Y_{i,t-1} - \gamma\rho\beta X_{i,t-1} - \rho(1-\gamma)Y_{i,t-2} + \gamma\epsilon_{i,t} \\ &= \gamma\beta X_{i,t} + (1-\gamma+\rho)Y_{i,t-1} - \gamma\rho\beta X_{i,t-1} - \rho(1-\gamma)Y_{i,t-2} + \gamma\epsilon_{i,t} \end{aligned}$$

Note that now all of the regressors on the right are uncorrelated with $\epsilon_{i,t}$, so we can identify the parameters above given considerable variations in the variables. Recall that $Y_{i,t}$, $Y_{i,t-1}$, $X_{i,t}$, and $X_{i,t-1}$ are observed. Now, consider the three cases:

▷ $X_{it} = \bar{X}$: a constant for all i, t

  * Here, we would have:

$$\begin{aligned} Y_{i,t} &= \gamma\beta\bar{X} + (1-\gamma+\rho)Y_{i,t-1} - \gamma\rho\beta\bar{X} - \rho(1-\gamma)Y_{i,t-2} + \gamma\epsilon_{i,t} \\ &= \gamma\beta(1-\rho)\bar{X} + (1-\gamma+\rho)Y_{i,t-1} - \rho(1-\gamma)Y_{i,t-2} + \gamma\epsilon_{i,t} \end{aligned}$$

  * Since $\bar{X}$ is not a random variable, we cannot identify the regressed coefficient, $\gamma\beta(1-\rho)$. So, we can only identify $1-\gamma+\rho$ and $\rho(1-\gamma)$ but not γ nor ρ separately.
  * But, having identified $1-\gamma+\rho$ and $\rho(1-\gamma)$ we can identify $\gamma(1-\rho)$ since $(1-\gamma+\rho)-\rho(1-\gamma) = 1-\gamma(1-\rho)$, and having identified $\gamma\beta(1-\rho)\bar{X}$, we can identify β by dividing $\gamma(1-\rho)\bar{X}$.
  * Hence, β is identified, but ρ and $\gamma = \frac{\phi}{\phi+\eta}$ are not identified.

▷ X_{it} varies over people and time

* Here, we have as before:

$$Y_{i,t} = \gamma\beta X_{i,t} + (1-\gamma+\rho)Y_{i,t-1} - \gamma\rho\beta X_{i,t-1} - \rho(1-\gamma)Y_{i,t-2} + \gamma\epsilon_{i,t}$$

* Since $X_{i,t}$ and $X_{i,t-1}$ are separately observed and vary across individuals, we can identify $\gamma\beta$, $1-\gamma+\rho$, and $\gamma\rho\beta$. From the three equations with three unknowns, so we can identify all of β, ρ, and $\gamma = \frac{\phi}{\phi+\eta}$ separately. To see this:

  - From the coefficient on $X_{i,t}$ ($= \gamma\beta$) and the coefficient on $X_{i,t-1}$ ($= \gamma\rho\beta$), we can identify ρ.
  - Having identified ρ, we can identify γ from the coefficient on $Y_{i,t-1}$ ($= 1-\gamma+\rho$).
  - Having identified γ, we can identify β from the coefficient on $X_{i,t}$ ($= \gamma\beta$).

▷ $X_{it} = X_t$ is the same variable over persons and time

* Here, we have

$$Y_{i,t} = \gamma\beta X_t + (1-\gamma+\rho)Y_{i,t-1} - \gamma\rho\beta X_{t-1} - \rho(1-\gamma)Y_{i,t-2} + \gamma\epsilon_{i,t}$$

* Since we still have variations over X_t and the lagged term X_{t-1} per each $Y_{i,t}$, the parameters are identified just as in the previous case. Hence, we can identify all of β, ρ, and $\gamma = \frac{\phi}{\phi+\eta}$ separately.

■

6 "Alternative Methods for Evaluating the Impact of Interventions"

Answer the questions in the "Alternative Methods for Evaluating the Impact of Interventions" handout. For the data set in Question 3. of this problem set, compare the performance (bias, variance) of synthetic cohort estimators with panel data estimators.

Problem 6.1. (Slide 20) If the costs of program participation are independent of U_{it} for all t (so both W_i and τ_i are independent of U_{it}) then $E[U_{it}d_i] = 0$ only if the unobservables in period t are (mean) independent of the unobservables in period k or

$$E[U_{it}|U_{ik}] = 0, \forall t > k.$$

Solution. We have

$$\begin{aligned}
E[U_{it}d_i] &= E[U_{it}|d_i = 1]Pr[d_i = 1] \\
&= E[U_{it}|W_i\phi + \frac{\alpha}{r} - X_{ik}\beta + \tau_i - U_{ik} > 0]Pr[W_i\phi + \frac{\alpha}{r} - X_{ik}\beta + \tau_i - U_{ik}] \\
&= E[U_{it}|g(W_i, X_{ik}, U_{ik})]Pr[g(W_i, X_{ik}, U_{ik})] \\
&= E[U_{it}|h(U_{ik})]Pr[g(W_i, X_{ik}, U_{ik})] && (U_{it} \perp\!\!\!\perp W_i, \tau_i)
\end{aligned}$$

Thus, if $E[U_{it}|U_{ik}] = 0$ and U_{it} is mean indepedent of U_{ik}, then $E[U_{it}|h(U_{ik})] = 0$, and thus $E[U_{it}d_i] = 0$.

■

Problem 6.2. (Slide 79) Assume

- ▷ $E_{k-1}(U_{ik}) = E[U_{ik}|U_{i,k-1}, ..., U_{i,k-N}]$.
- ▷ S_i and X_{ik} are known as of period $k - 1$ when the enrollment decision is being made.
- ▷ X_{it} is distributed independently of U_{ij} for all t and j.

Then,

$$E[d_i|U_{it}, U_{i,t-1}, ..., U_{i,k-1}, ..., U_{i,k-N}] = E[d_i|U_{i,k-1}, ..., U_{i,k-N}].$$

Solution. By assumption 2, we have

$$d_i = 1\{S_i - X_{ik}\beta - U_{ik} + \frac{\alpha}{r}\}.$$

Then,

$$\begin{aligned} E[d_i|U_{it}, U_{i,t-1}, ..., U_{i,k-1}, ..., U_{i,k-N}] &= E[1\{S_i - X_{ik}\beta - U_{ik} + \frac{\alpha}{r}\}|U_{it}, U_{i,t-1}, ..., U_{i,k-1}, ..., U_{i,k-N}] \\ &= E[g(S_i, X_{ik}, U_{ik})|U_{it}, U_{i,t-1}, ..., U_{i,k-1}, ..., U_{i,k-N}]. \end{aligned}$$

By the assumptions, we have that S_i and X_{ik} do not depend on any U_{ij}'s and that $E_{k-1}[U_{ik}]$ only depends on the previous N periods of residuals. Thus,

$$E[d_i|U_{it}, U_{i,t-1}, ..., U_{i,k-1}, ..., U_{i,k-N}] = E[d_i|U_{i,k-1}, ..., U_{i,k-N}]$$

■

Problem 6.3. (Slide 105) Assume

▷ U_{it} is covariance stationary.

▷ U_{it} has a linear regression on U_{ik} for all t (i.e., $E[U_{it}|U_{ik}] = \beta_{tk}U_{ik}$).

▷ U_{it} is mutually independent of (X_{ik}, S_i) for all t.

▷ α is common to all individuals (so the model is of the fixed coefficient form).

▷ The environment is one of perfect foresight where decision rule $d_i = 1\{S_i - Y_{ik} + \frac{\alpha}{r}\}$ determines participation.

Then, the following characterizes the data

$$\begin{aligned} E[(U_{it} - U_{it'})(X_{it} - X_{it'})] &= 0 \qquad \text{for some } t > k > t' \\ E[(U_{it} - U_{it'})d_i] &= 0 \qquad \text{for some } t > k > t' \end{aligned}$$

Solution. Let $t = k + j$ and $t' = k - j$. Note that by the first and second assumptions, there exists a δ such that

$$\begin{aligned} U_{it} &= \delta U_{ik} + \omega_{it} \\ U_{it'} &= \delta U_{ik} + \omega_{it'} \end{aligned}$$

and $E[\omega_{it}|U_{ik}] = E[\omega_{it'}|U_{ik}] = 0$.

Thus, $U_{it} - U_{it'} = \omega_{it} - \omega_{it'}$. Observe we also have

$$E[U_{it}|d_i = 1] = \delta E[U_{ik}|d_i = 1] + E[\omega_{it}|d_i = 1].$$

But from the third assumption we have $E[\omega_{it}|d_i = 1] = 0$. By construction, $\omega_{it} \perp\!\!\!\perp U_{ik}$. The third assumption tells us ω_{it}, i.e. a component of U_{it}, is independent of (X_{ik}, S_i). Since $d_i = S_i - X_{ik}\beta - U_{ik} + \frac{\alpha}{r}$, a function of variables that are independent of ω_{it}, $\omega_{it} \perp\!\!\!\perp d_i$. Thus, we also have mean independence and so

$$E[\omega_{it}|d_i = 1] = E[\omega_{it}] = 0.$$

By a similar argument, we have

$$E[\omega_{it'}|d_i = 1] = E[\omega_{it'}] = 0$$

and thus

$$E[U_{it} - U_{it'}|d_i = 1] = E[\omega_{it} - \omega_{it'}|d_i = 1] = 0.$$

Since mean independence implies uncorrelatedness, we have

$$E[(U_{it} - U_{it'})d_i] = 0.$$

■

Problem 6.4. For the data set in Question 3 of this problem set, compare the performance (bias, variance) of synthetic cohort estimators with panel data estimators.

Solution. Below are the results when running a pooled regression (synthetic cohort) on the samples vs including fixed effects (panel), we find that the coefficient estimates do not change but the standard errors are smaller in the panel data estimate. Thus, we have a more efficient estimator when using panel data.

■

Table 7: Estimating alpha: Synthetic Cohort vs Panel

	(1)	(2)	(3)	(4)	(5)	(6)
	Sample 1	Sample 1	Sample 2	Sample 2	Sample 3	Sample 3
Post-Treatment=1 × Treated=1	0.483***	0.483***	0.558***	0.558***	0.621***	0.621***
	(0.0268)	(0.0238)	(0.0241)	(0.0230)	(0.0699)	(0.0360)
N	2000	2000	2000	2000	2000	2000
R-sq	0.353	0.512	0.407	0.482	0.081	0.768
Type of Data	SC	Panel	SC	Panel	SC	Panel

Standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

7 Tradeoffs in Assumptions required to identify α

For the model used in the reference cited in Question 6., consider the tradeoff in assumptions required to identify α. Specifically, consider the following sources of data:

Problem 7.1. Panel Data

Solution. We will go over the three methods given in the slides to estimate α using panel/longitudinal data.

- ▷ **Fixed Effects Method:** Here, if we assume

$$E[U_{it} - U_{it'}|D_i, X_{it} - X_{it'}] = 0 \qquad\qquad \forall t, t', t > k > t'.$$

Then we can run a difference regression equation that will consistently estimate α:

$$E[Y_{it} - Y_{it'}|d_i, X_{it} - X_{it'}] = (X_{it} - X_{it'})\alpha_1 + \alpha D_i.$$

The assumption will hold if we have $U_{it} = \phi_i + \varepsilon_{it}$ where ϕ_i is a permanent component of the residual and ε_{it} is a transitory residual. Both are mean zero random variables that are independent of each other and ε_{it} is independent of all other transitory residuals at different dates. Will also hold if Z is independent of the transitory residuals except perhaps ε_{ik}. With access to two periods of data, we can just identify α. We in fact have 20 periods of data and thus α will be over-identified, and thus we can test these assumptions.

- ▷ U_{it} **follows an AR(1) process:** Suppose U_{it} follows a first-order autoregression:

$$U_{it} = \rho U_{i,t-1} + \nu_{it}$$

where $E[\nu_{it}]$ and they are mutually independent and $\rho \neq 1$.

Substituting this into our main specification we get

$$Y_{it} = [X_{it} - \rho^{t-t'} X_{it'}]\alpha_1 + (1 - \rho^{t-t'})D_i\alpha + \rho^{t-t'}Y_{it'} + \sum_{j=0}^{t-(t'+1)} \rho^j \nu_{i,t-j}, t > k > t'$$

and thus the regression below will consistently estimate α as the set of observations per person gets large:

$$E[Y_{it}|X_{it}, X_{it'}, d_i, Y_{it'}] = [X_{it} - \rho^{t-t'} X_{it'}]\alpha_1 + (1 - \rho^{t-t'})D_i\alpha + \rho^{t-t'}Y_{it'}$$

- ▷ U_{it} **is covariance stationary:** We have to assume here that

* $E[U_{it}U_{i,t-j}] = E[U_{it'}U_{i,t'-j}] = \sigma_j, \forall j \geq 0$.
* Have access to pre and post training data.
* $pE[U_{it}|D_i = 1] \neq 0$.

Using the following moments of the data:

$$m_1 = (\sum (Y_{it'} - \bar{Y}_{t'})(Y_{i,t'-j} - \bar{Y}_{t'-j}))/I$$

$$m_2 = (\sum (Y_{it} - \bar{Y}_t)(Y_{it'} - \bar{Y}_{t'}))/I$$

$$m_3 = (\sum (Y_{it'} - \bar{Y}_{t'})D_i)/I$$

and noting plim$m_1 = \sigma_j$, plim$m_2 = \sigma_j + \alpha pE[U_{it'}|D_i = 1]$, and plim$m_3 = pE[U_{it'}|D_i = 1]$, we have

$$\textbf{plim}\frac{m_2 - m_1}{m_3} = \alpha$$

Overall, these assumptions are not very strong and an advantage of panel data is that we typically only need two data points for the same individuals to consistently estimate α. Even better, the more points we have these assumptions start to become testable. However, there are some advantages of repeated-cross sections that we will see below.

■

Problem 7.2. Repeated cross-section on individuals

Solution. With access to repeated cross-sections on individuals, as long as the samples are random we can use the methods we used above with panel data, but in slightly different ways. This results in obvious trade-offs between the two methods.

▷ **Fixed Effects Method:** We require the same assumption as in using fixed effects with panel data of common trends. Then the differences of the sample average will be a consistent estimator of α, i.e.

$$\hat{\alpha} = (\bar{Y}_t^{(1)} - \bar{Y}_t^{(0)}) - (\bar{Y}_{t'}^{(1)} - \bar{Y}_{t'}^{(0)})$$

Thus, this is not really different from the panel data method. As long as the sampling is random with the repeated samples, α can be consistently estimated.

▷ U_{it} **follows an AR(1) process:** In this case, we must assume we have three post-period data points. Assume for simplicity $X_{it}\alpha_1 = \alpha_t$ Thus,

$$\begin{aligned}
\textbf{plim}\bar{Y}_j^{(1)} &= \alpha_t + \alpha + E[U_{ij}|D_i = 1] \\
\textbf{plim}\bar{Y}_j^{(0)} &= \alpha_t + E[U_{ij}|D_i = 0].
\end{aligned}$$

The AR(1) structure gives us

$$\begin{aligned}
E[U_{i,t+1}|D_i = 1] &= \rho E[U_{i,t}|D_i = 1] \\
E[U_{i,t+1}|D_i = 0] &= \rho^2 E[U_{i,t}|D_i = 0] \\
E[U_{i,t+2}|D_i = 1] &= \rho E[U_{i,t}|D_i = 1] \\
E[U_{i,t+2}|D_i = 0] &= \rho^2 E[U_{i,t}|D_i = 0].
\end{aligned}$$

Thus we can estimate α since

$$\hat{\rho} = \frac{(\bar{Y}_{t+2}^{(1)} - \bar{Y}_{t+2}^{(0)}) - (\bar{Y}_{t+1}^{(1)} - \bar{Y}_{t+1}^{(0)})}{(\bar{Y}_{t+1}^{(1)} - \bar{Y}_{t+1}^{(0)}) - (\bar{Y}_{t}^{(1)} - \bar{Y}_{t}^{(0)})}$$

is a consistent estimator for ρ and so

$$\hat{\alpha} = \frac{(\bar{Y}_{t+2}^{(1)} - \bar{Y}_{t+2}^{(0)}) - \hat{\rho}(\bar{Y}_{t+1}^{(1)} - \bar{Y}_{t+1}^{(0)})}{1 - \hat{\rho}}$$

is consistent for α. The main drawback here compared to panel data is that we need at least 3 post training data points, so more data is needed compared to longitudinal data. However, if there is measurement error in the data, this method could still be advantageous since classical measurement error will be averaged out in the repeated cross sections method.

▷ U_{it} **is covariance stationary:** The key difference here is that we do not need the strong stationarity assumptions we had with panel data. Here, we only need stationary for variances; covariances can differ. That is, with panel data we had $E[U_{it}U_{i,t-j}] = E[U_{it'}U_{i,t'-j}] = \sigma_j, \forall j \geq 0$; here, we only need this to hold for $j = 0$.

■

Problem 7.3. Repeated cross-section on aggregates (say cohorts) but not individuals.

Solution. Similar to aggregate data below and having repeated cross-sections on individuals. Anything we did for the individuals, we could do for the cohorts, since the subjects are not linked. Just like in the aggregate data case, we need to know the fraction treated in each point in time. Unlike with aggregate data, we can measure heterogeneous effects across cohort groups.

■

Problem 7.4. Cross-section data

Solution. As discussed in the last problem set, with a cross section we can identify α using instrumental variables (IV), matching, selection models, or maximum likelihood estimation (MLE). The assumptions require for each step follow

- ▷ IV: The "only" assumptions required here are that we have a variable Z_i that is exogenous with respect to U_{it} and it is correlated with treatment d_i. Then, we can use Z_i as an instrument for d_i. Exogeneity is a strong assumption, but the advantage of the approach of instrumental variables is that we make no assumptions on the distributions of V_i or U_{it}.

- ▷ Matching: Here we need conditional independence ($(Y_{1i}, Y_{0i}) \perp\!\!\!\perp d_i | X_{it}, Z_i$) and $0 < P(d_i = 1|Z_i, X_{it}) < 1$. Thus, we are making quite strong assumptions about the joint conditional distributions of potential outcomes and treatment.

- ▷ Selection Models: If we assume that $Z_i \perp\!\!\!\perp V_i$ and that we either know or can estimate the functional form of the distribution V_i, we can use the Roy sample selection model to estimate α. Specifically, we use a probit model to estimate the propensity score $P[d_i = 1|Z_i] = F(-Z_i\gamma)$. If we further assume $Z_i \perp\!\!\!\perp U_{it}$, then we can use $F(-Z_i\hat{\gamma})$, fitted values of a probit, as an instrument for d_i. That is, we have at the population level

$$E[Y_{it}|X_{it}, Z_i] = X_{it}\beta + \alpha(1 - F(-Z_i\gamma)).$$

  Replacing $F(-Z_i\gamma)$ with $F(-Z_i\hat{\gamma})$ we can consistently estimate α by estimating the model above using OLS. Regardless, we have to make assumptions about the joint distributions of U_{it}, Z_i, V_i for this to work.

- ▷ MLE: If the joint density of (U_{it}, V_i) is assumed known up to a finite set of parameters, we can estimate α using MLE. We do so by estimating the equation

$$E[Y_{it}|X_{it}, Z_i] = X_{it}\beta + d_i\alpha + d_iE[U_{it}|d_i = 1, Z_i] + (1 - d_i)E[U_{it}|d_i = 0, Z_i].$$

■

Problem 7.5. Aggregate data where the identity of training status is not known but in the aggregate the proportion of people in training is known

---

Solution. First begin with only having access to two dates of data. If we assume a time homogenous environment ($X_{it}\beta_t = X_{it'}\beta_{t'}$) and that the samples that generated this data is random we can get a consistent estimate of α. Specifically, note

$$\begin{aligned} \mathbf{plim}\bar{Y}_t &= E[\beta_t + \alpha d_i + U_{it}] = \beta_t + \alpha p && t > k \\ \mathbf{plim}\bar{Y}_{t'} &= E[\beta_{t'} + \alpha d_i + U_{it}] = \beta_{t'} && t' < k \end{aligned}$$

where $p = E[d_i]$. Then, a consistent estimate of α would be

$$\mathbf{plim}\frac{\bar{Y}_t - \bar{Y}_{t'}}{\hat{p}} = \alpha$$

where $\hat{p}$ is a consistent estimator of p (the proportion estimate we have access to). If we have access to more than two dates we can even relax the time homogeneity assumption, although we have to make some assumption of the functional form. For example, if

$$\beta_t = \sum_{j=0}^{L-2} \pi_j t^j$$

then having access to L repeated cross sections allows us to consistently estimate the $L-1$ coefficients above and α.

■

8 Ashenfelter's Dip

For the models used in Questions 6 and 7, suppose α is a random coefficient model, i.e., α_i varies in the population. Consider two cases:

- ▷ $\alpha_i \perp\!\!\!\perp d_i$
- ▷ $\alpha_i \not\perp\!\!\!\perp d_i$

Derive "Ashenfelter's dip". Does stationary of the environment help in identifying the model?

Solution. We define Ashenfelter's Dip to be

$$\begin{aligned} E_t[y_{it}] - E[y_{it}|D_k = 1] &= E_t[\alpha_0 + \alpha_1 X_t + U_{it}] - E_t[\alpha_0 + \alpha_1 X_t + U_{it}|E_k[\frac{\alpha}{r} - C - Y_{ik}] > 0] \\ &= E_t[U_{it}] - E_t[U_{it}|E_k[\frac{\alpha}{r} - C - Y_{ik}] > 0] \\ &= E_t[U_{it}] - E_t[U_{it}|E_k[U_{ik}] < \frac{\alpha}{r} - C - \alpha_0 - \alpha_1 X_k] \end{aligned}$$

Since

$$\begin{aligned} & E_k[\frac{\alpha}{r} - C - Y_{ik}] > 0 \\ \iff & E_k[\frac{\alpha}{r} - C - \alpha_0 - \alpha_1 X_k - U_{ik}] > 0 \\ \iff & \frac{\alpha}{r} - C - \alpha_0 - \alpha_1 X_k > E_k[U_{ik}] \end{aligned}$$

Stationarity will help in this model. If it is stationary, we can properly estimate this Ashenfelter's dip term and correct for it when estimating α_i. We cannot generally do this if we do not have stationarity.

If $\alpha_i \perp\!\!\!\perp d_i$, then we do not have to worry about Ashenfelter's dip since $E_t[y_{it}] = E[y_{it}|D_k = 1]$. If $\alpha_i \not\perp\!\!\!\perp d_i$, then selection is an issue and we have to worry about Ashenfelter's dip.

■

9 Heterogeneity in Binomial

Consider a Bernouli model:

$$
\begin{aligned}
&D_{i,t} = \mathbf{1}(\mathbf{Z}_{it}\delta + \beta D_{i,t-1} + U_{it} > 0) \\
&U_{it} = \phi_i + \epsilon_{it} \quad E(\phi_i) = 0 \text{ and } E(\epsilon_{it}) = 0 \\
&\epsilon_{it} \text{ iid, } \phi \perp\!\!\!\perp (\epsilon_{it})_{t=1}^{T} \text{ all } i \\
&[\phi_i, (\epsilon_{it})_{t=1}^{T}] \perp\!\!\!\perp \{Z_{it}\}_{t=1}^{T}
\end{aligned}
$$

Problem 9.1. For the case $Z_{it} = \bar{Z}$, constant for all i, discuss the runs patterns (e.g.1,0,1;1,1,0;0,0,0; etc.) associated with $\beta > 0$ and $\beta = 0$. How can you use data on runs patterns to test for state dependence? What values of the parameters produce a Markov model but not a pure Bernoulli model?

Solution. Consider $\beta = 0$ first. Then, we have

$$
\begin{aligned}
D_{i,t} &= \mathbf{1}(\bar{Z}\delta + U_{it} > 0) \\
&= \mathbf{1}(\bar{Z}\delta + \phi_i + \epsilon_{it} > 0) \\
\Rightarrow Pr[D_{i,t} = 1 | D_{i,t-1} = 0] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z}\delta)].
\end{aligned}
$$

Since ϵ_{it} are iid and independent of ϕ_i for all i, the associated probability with runs patterns would be given as

$$
Pr[d_{i1}, d_{i2}, ..., d_{it}] = \prod_{j=1}^{t} p_{ij}^{d_{ij}} (1 - p_{ij})^{1-d_{ij}}.
$$

Hence, this is a pure Bernoulli model with state independence for the runs patterns:

$$
Pr[D_{i,t} = 1 | D_{i,t-1} = 1] = Pr[D_{i,t} = 1 | D_{i,t-1} = 0] = Pr[D_{i,t} = 1].
$$

Now, consider $\beta > 0$. As before, we have $\epsilon_{it} \perp\!\!\!\perp \phi_i$ for $\forall i$, but now probability of $D_{i,t} = 1$ depends on $D_{i,t-1}$, i.e. we have state dependence. To see this:

$$
\begin{aligned}
D_{i,t} &= \mathbf{1}(\bar{Z}\delta + \beta D_{i,t-1} + \beta D_{i,t-1} + U_{it} > 0) \\
&= \mathbf{1}(\bar{Z}\delta + \beta D_{i,t-1} + \phi_i + \epsilon_{it} > 0) \\
\Rightarrow Pr[D_{i,t} = 1 | D_{i,t-1} = 1] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z} + \beta)] \\
Pr[D_{i,t} = 1 | D_{i,t-1} = 0] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z})]
\end{aligned}
$$

Note that by Bayes rule,

$$\begin{aligned}Pr[D_{i,t}=1|D_{i,t-1}=1] &= \frac{Pr[D_{i,t}=1, D_{i,t-1}=1]}{Pr[D_{i,t-1}=1]} \\ &\approx \frac{\#[D_{i,t}=1, D_{i,t-1}=1]}{\#[D_{i,t-1}=1]} \\ Pr[D_{i,t}=1|D_{i,t-1}=0] &= \frac{Pr[D_{i,t}=1, D_{i,t-1}=0]}{Pr[D_{i,t-1}=0]} \\ &\approx \frac{\#[D_{i,t}=1, D_{i,t-1}=0]}{\#[D_{i,t-1}=0]},\end{aligned}$$

where $\#[A]$ is the number of times we observe the event A empirically. Hence, using the empirical counterparts of the two conditional probabilities, we can test state dependence, i.e. $\beta > 0$, with $Pr[D_{i,t}=1|D_{i,t-1}=1] > Pr[D_{i,t}=1|D_{i,t-1}=0]$.

Note that with $\beta = 0$, the model is simply a pure Bernoulli model since the runs observations are drawn i.i.d from a Bernoulli distribution with $p_{it} = Pr[\epsilon_{it} > -(\phi_i + \bar{Z}\delta]$. However, with $\beta > 0$, recall that $D_{i,t}$ depended on $D_{i,t-1}$:

$$\begin{aligned}Pr[D_{i,t}=1|D_{i,t-1}=1] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z} + \beta)] \\ Pr[D_{i,t}=1|D_{i,t-1}=0] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z})].\end{aligned}$$

Hence, with $\beta > 0$ the model is a Markov model involving serial correlation with the preceding period. And as explained before, if $\beta = 0$, then the model is simply Bernoulli.

■

Problem 9.2. Suppose $\beta = 0$, but $U_{i,t} = \rho U_{i,t-1} + \epsilon_{it}$, $\rho > 0$, ϵ_{it}, iid. What are the runs patterns for this model?

Solution. With $\beta = 0$ and $U_{i,t} = \rho U_{i,t-1} + \epsilon_{it}$, we have:

$$\begin{aligned}D_{i,t} &= \mathbf{1}(Z_{it}\delta + U_{it} > 0) \\ &= \mathbf{1}(Z_{it}\delta + \rho U_{i,t-1} + \epsilon_{it} > 0) \\ \Rightarrow p_{it} := Pr[D_{i,t}=1] &= Pr[\epsilon_{it} > -(Z_{it}\delta + \rho U_{i,t-1})].\end{aligned}$$

Now, note that

$$\begin{aligned}Pr[D_{i,t}=1|D_{i,t-1}=1] &= Pr[\epsilon_{it} > -(Z_{it}\delta + \rho U_{i,t-1})|U_{i,t-1} > -Z_{it}\delta] \\ &> Pr[\epsilon_{it} > -(Z_{it}\delta + \rho U_{i,t-1})|U_{i,t-1} \leq -Z_{it}\delta] \qquad \because \epsilon_{i,t} \perp\!\!\!\perp Z_{it}, \rho > 0 \\ &= Pr[D_{i,t}=1|D_{i,t-1}=0],\end{aligned}$$

where the second inequality follows from the fact that $\epsilon_{i,t}$ and Z_{it} are independent and that $\rho > 0$ implies that with high U_{it-1} we will likely have high U_{it} as well, so $D_{it} = 1$ is more likely.

Hence, the runs for this model exhibits a similar pattern with the $\beta > 0$ case in part (a), having state dependence with $D_{i,t} = 1$ being more probable after observing $D_{i,t-1} = 1$ than $D_{i,t-1} = 0$.

■

Problem 9.3. For the model with $U_{it} = \phi_i + \epsilon_{it}$, suppose $Z_t = \bar{Z}$. Compute $Pr(D_t = 1|D_{t-1} = 1)$ and show that $Pr(D_t = 1|D_{t-1} = 1) > Pr(D_t)$.

Solution. Recall from (a), the conditional probabilities can be computed as:

$$\begin{aligned}
Pr[D_{i,t} = 1|D_{i,t-1} = 1] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z} + \beta)] \\
&= E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta - \beta)] && \epsilon_i \perp\!\!\!\perp \phi_i \\
Pr[D_{i,t} = 1|D_{i,t-1} = 0] &= Pr[\epsilon_{it} > -(\phi_i + \bar{Z})] \\
&= E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta)] && \epsilon_i \perp\!\!\!\perp \phi_i.
\end{aligned}$$

and, if $\beta > 0$, we have $Pr[D_{i,t} = 1|D_{i,t-1} = 1] > Pr[D_{i,t} = 1|D_{i,t-1} = 0]$. Note that by LIE:

$$\begin{aligned}
Pr[D_t = 1] &= Pr[D_{i,t} = 1|D_{i,t-1} = 1]Pr[D_{i,t-1} = 1] + Pr[D_{i,t} = 1|D_{i,t-1} = 0]Pr[D_{i,t-1} = 0] \\
&< Pr[D_{i,t} = 1|D_{i,t-1} = 1]Pr[D_{i,t-1} = 1] + Pr[D_{i,t} = 1|D_{i,t-1} = 1]Pr[D_{i,t-1} = 0] \\
&= Pr[D_{i,t} = 1|D_{i,t-1} = 1]\underbrace{(Pr[D_{i,t-1} = 1] + Pr[D_{i,t-1} = 0])}_{=1} \\
&= Pr[D_{i,t} = 1|D_{i,t-1} = 1],
\end{aligned}$$

where the inequality in the second line follows from the fact that $Pr[D_{i,t} = 1|D_{i,t-1} = 1] > Pr[D_{i,t} = 1|D_{i,t-1} = 0]$. Hence, we have obtained the desired relationship: $Pr[D_{i,t} = 1|D_{i,t-1} = 1] > Pr[D_t = 1]$.

■

Problem 9.4. Write down the likelihood function for the model.

Solution. For notational brevity, we let $q_{it} := Pr[D_{i,t} = 1|D_{i,t-1} = 1]$ and $r_{it} := Pr[D_{i,t} = 1|D_{i,t-1} = 0]$. Recall from (c),

$$\begin{aligned}
q_{it} &= Pr[D_{i,t} = 1|D_{i,t-1} = 1] = E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta - \beta)] \\
r_{it} &= Pr[D_{i,t} = 1|D_{i,t-1} = 0] = E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta)].
\end{aligned}$$

Also, recall from (a), we had $p_{it} = Pr[D_{i,t} = 1]$. Note that because there is no D_{i0} to condition on in $t = 1$, we simply have $p_{i1} = E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta)]$.

Then, the likelihood function is given by:

$$\begin{aligned}
\mathcal{L}(\beta, \delta, f_\phi, f_\epsilon, \bar{Z}|d_{i1}, ..., d_{it}) &= Pr[D_{i1} = d_{i1}, ..., D_{it} = d_{it}] \\
&= \left(\prod_{j=2}^{t} Pr[D_{ij} = d_{ij}|D_{ij-1} = d_{ij-1}]\right) Pr(D_{i1} = d_{i1}) \\
&= \left(\prod_{j=2}^{t} Pr[D_{ij} = d_{ij}|D_{ij-1} = d_{ij-1}]\right) (d_{i1}p_{i1} + (1 - d_{i1})(1 - p_{i1})) \\
&= \left(\prod_{j=2}^{t} d_{ij-1}Pr[D_{ij} = d_{ij}|D_{ij-1} = 1] + (1 - d_{ij-1})Pr[D_{ij} = d_{ij}|D_{ij-1} = 0]\right) \\
&\times (d_{i1}p_{i1} + (1 - d_{i1})(1 - p_{i1})) \\
&= \left(\prod_{j=2}^{t} d_{ij-1}(d_{ij}Pr[D_{ij} = 1|D_{ij-1} = 1] + (1 - d_{ij})Pr[D_{ij} = 0|D_{ij-1} = 1])\right. \\
&\left. + (1 - d_{ij-1})(d_{ij}Pr[D_{ij} = 1|D_{ij-1} = 0] + (1 - d_{ij})Pr[D_{ij} = 0|D_{ij-1} = 0])\right) \\
&\times (d_{i1}p_{i1} + (1 - d_{i1})(1 - p_{i1})) \\
&= \left(\prod_{j=2}^{t} d_{ij-1}(d_{ij}q_{ij} + (1 - d_{ij})(1 - q_{ij})) + (1 - d_{ij-1})(d_{ij}r_{ij} + (1 - d_{ij})(1 - r_{ij}))\right) \\
&\times (d_{i1}p_{i1} + (1 - d_{i1})(1 - p_{i1})),
\end{aligned}$$

where

$$\begin{aligned}
q_{it} &= E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta - \beta)] \\
r_{it} &= E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta)] \\
p_{i1} &= E_\phi[1 - F_\epsilon(-\phi_i - \bar{Z}\delta)].
\end{aligned}$$

■

Problem 9.5. Read the "Heterogeneity in the Binomial Model" handout. Relate the model of $8a$ to the beta-binomial model in that handout.

Solution. Note that, beta-binomial model with $B(a, b)$ exhibits:

$$Pr[D_{it} = 1|D_{it-1} = 1, ...D_{i0} = 1] = \frac{a + t + 1}{a + b + t + 1},$$

which is monotonically increasing in t. However, the model of 9a is Markov depending only on the previous state:

$$Pr[D_{it} = 1|D_{it-1} = 1, ...D_{i0} = 1] = Pr[D_{it} = 1|D_{it-1} = 1] = Pr[\epsilon_{it} > -(\phi_i + \bar{Z} + \beta)].$$

Hence, the difference is that the respective probability increases in t under the beta-binomial model, but remains constant in the model of 9a.

■

10 "Modeling the Income Process"

Answer the questions in the "Modeling the Income Process" handout.

Problem 10.1. If the order of the MA process is one in the levels, then to implement this we will need at least six individual-level observations to construct this moment. **Question: Show this.**

Solution. We define

$$g_{i,a,t} := \Delta y_{i,a,t} = y_{i,a,t} - y_{i,a-1,t-1}.$$

Recall the key moment condition for identifying the variance of the permanent shock, $\zeta_{i,a,t}$, is given as:

$$E(\zeta_{i,a,t}^2) = E\left[g_{i,a,t}\left(\sum_{j=-(1+q)}^{(1+q)} g_{i,a+j,t+j}\right)\right],$$

where q is the order of the moving average process in the original levels equation.

Here, since we have $q = 1$, we have:

$$E(\zeta_{i,a,t}^2) = E\left[g_{i,a,t}\left(\sum_{j=-2}^{2} g_{i,a+j,t+j}\right)\right],$$

so identification requires $g_{i,a-2,t-2}$, ..., $g_{i,a+2,t+2}$. Using the definition of $g_{i,a,t}$ above, it follows that we need $y_{i,a-3,t-3}$, ..., $y_{i,a+2,t+2}$. Hence, we will need at least six individual-level observations to construct the above moment.

■

Problem 10.2. An alternative specification with very different implications is one where

$$\ln Y_{i,a,t} = \rho \ln Y_{i,a-1,t-1} + d_t(\Delta X'_{i,a,t}\beta + \Delta v_{i,a,t}) + m_{i,a,t},$$

where h_i is a fixed effect while $v_{i,a,t}$ follows some MA process and $m_{i,a,t}$ is measurement error. This process can be estimated by method of moments following a suitable transformation of the model.

Define $\theta_t = d_t/d_{t-1}$ and quasi-difference to obtain:

$$\ln Y_{i,a,t} = (\rho + \theta_t) \ln Y_{i,a-1,t-1} - \theta_t \rho \ln Y_{i,a-2,t-2} + d_t(\Delta X'_{i,a,t}\beta + \Delta v_{i,a,t}) + m_{i,a,t} - \theta_t m_{i,a-1,t-1}) \quad (8)$$

In this model the persistence of the shocks is captured by the autoregressive component of $\ln Y$ which means that the effects of time varying characteristics are persistent to an extent. Given estimates of the levels equations in (8) the autocovariance structure of the residuals can be used to identify the properties of the error term $d_t \Delta v_{i,a,t} + m_{i,a,t} - \theta_t m_{i,a-1,t-1}$. **Question: Prove this.**

Solution. Let

$$U_{i,a,t} := d_t \Delta v_{i,a,t} + m_{i,a,t} - \theta_t m_{i,a-1,t-1},$$

so that

$$\ln Y_{i,a,t} = (\rho + \theta_t) \ln Y_{i,a-1,t-1} - \theta_t \rho \ln Y_{i,a-2,t-2} + d_t \Delta X'_{i,a,t} \beta + U_{i,a,t},$$

where all of the regressors on the right are uncorrelated with the residual, $U_{i,a,t}$. We can identify $d_t \beta$ here and $d_{t-1}\beta$ from the regression of $\ln Y_{i,a-1,t-1}$, so we can identify $\theta_t = \frac{d_t}{d_{t-1}} = \frac{d_t \beta}{d_{t-1}\beta}$. Since we can identify $\rho + \theta_t$, we can thus identify ρ as well.

Recall $v_{i,a,t} = \epsilon_{i,a,t} - \theta_t \epsilon_{i,a-1,t-1}$ (from slide 13). Then, we have

$$\begin{aligned}
U_{i,a,t} &= d_t \Delta v_{i,a,t} + m_{i,a,t} - \theta_t m_{i,a-1,t-1} \\
&= d_t (v_{i,a,t} - v_{i,a-1,t-1}) + m_{i,a,t} - \theta_t m_{i,a-1,t-1} \\
&= d_t (\epsilon_{i,a,t} - \theta_t \epsilon_{i,a-1,t-1} - \epsilon_{i,a-1,t-1} + \theta_{t-1} \epsilon_{i,a-2,t-2}) + m_{i,a,t} - \theta_t m_{i,a-1,t-1} \\
&= d_t (\epsilon_{i,a,t} - (1+\theta_t) \epsilon_{i,a-1,t-1} + \theta_{t-1} \epsilon_{i,a-2,t-2}) + m_{i,a,t} - \theta_t m_{i,a-1,t-1} \\
\Rightarrow var[U_{i,a,t}] &= d_t^2 (\sigma^2_{\epsilon,t} + (1+\theta_t)^2 \sigma^2_{\epsilon,t-1} + \theta^2_{t-1} \sigma^2_{\epsilon,t-2}) + \sigma^2_{m,t} + \theta_t^2 \sigma^2_{m,t-1}
\end{aligned}$$

Recall from slide 23,

$$\begin{aligned}
\sigma^2_{\epsilon,t} &= \frac{E[g_{i,a,t} g_{i,a-2,t-2}]}{\theta_t} \\
\sigma^2_{m,t} &= -E[g_{i,a,t} g_{i,a-1,t-1}] - \frac{(1+\theta_t)^2}{\theta_t} E[g_{i,a,t} g_{i,a-2,t-2}].
\end{aligned}$$

Since we can identify all the autocovariance moments involved in $\sigma^2_{\epsilon,t}$,$\sigma^2_{\epsilon,t-1}$,$\sigma^2_{\epsilon,t-2}$, $\sigma^2_{m,t}$, and $\sigma^2_{m,t-1}$, and, as explained before, we can identify θ_t and θ_{t-1}, so we can identify all the variables in $var[U_{i,a,t}]$ except for d_t. In other words, as asked in the question, we can use the parameter estimates from the regression model (8) and use the autocovariance structure of the residuals to partially identify the variance of $U_{i,a,t}$ (everything except for d_t), so a reasonable assumption on the support of d_t would help us identify the bounds on the variance of $U_{i,a,t} = d_t \Delta v_{i,a,t} + m_{i,a,t} - \theta_t m_{i,a-1,t-1}$.

■

Problem 10.3. Alternatively, the fixed effect with the autoregressive component can be replaced by a random walk in a similar type of model. This could take the form

$$\ln Y_{i,a,t} = d_t (X'_{i,a,t} \beta + p_{i,a,t} + v_{i,a,t-1}) + m_{i,a,t}$$

In this model $p_{i,a,t} = p_{i,a,t-1} + \zeta_{i,a,t}$ as before, but the shocks have a different effect depending on aggregate conditions. **Question: Prove this.**

Solution. With $p_{i,a,t} = p_{i,a,t-1} + \zeta_{i,a,t}$, we have

$$\begin{aligned}
\ln Y_{i,a,t} &= d_t(X'_{i,a,t}\beta + p_{i,a,t} + v_{i,a,t-1}) + m_{i,a,t} \\
&= d_t(X'_{i,a,t}\beta + p_{i,a,t-1} + \zeta_{i,a,t} + v_{i,a,t-1}) + m_{i,a,t} \\
&= d_t(X'_{i,a,t}\beta + p_{i,a,t-1} + v_{i,a,t-1}) + d_t\zeta_{i,a,t} + m_{i,a,t}.
\end{aligned}$$

Note that the effect of $\zeta_{i,a,t}$ on $\ln Y_{i,a,t}$ depends on d_t. Since d_t varies with aggregate conditions, a unit permanent shock exerts a differential effect on individual income depending on aggregate conditions.

■

- Nevertheless the following two key moment conditions identify the parameters of the ARCH process, conditional on the unobserved heterogeneity (ν and ξ):

$$E_{t-2}\left(g_{i,a+q+1,t+q+1}g_{i,a,t} - \theta\gamma_t - \gamma g_{it+q}g_{i,a-1,t-1} - \theta\nu_i\right) = 0 \qquad \text{Transitory}$$

$$\begin{aligned}
&E_{t-q-3}\left[g_{i,a,t}\left(\sum_{j=-(1+q)}^{(1+q)} g_{i,a+j,t+j}\right)\right. \\
&\left. -\varphi_t - \varphi g_{i,a-1,t-1}\left(\sum_{j=-(1+q)}^{(1+q)} g_{ia+j-1t+j-1}\right) - \xi_i\right] = 0 \qquad \text{Permanent}
\end{aligned} \tag{13}$$

- The important point here is that it is sufficient to know the order of the MA process q.
- We do not need to know the parameters themselves.
- **Question: Show why this is true.**

Problem 10.4. Question: Show why this is true.

Solution. The proof of the results follow from Meghir and Pistaferri (2004). Recall from the slides:

$$\begin{aligned}
E_{t-1}(\epsilon^2_{i,a,t}) &= \gamma_t + \gamma\epsilon^2_{i,a-1,t-1} + v_i && \text{transitory} \\
E_{t-1}(\zeta i, a, t^2) &= \varphi_t + \varphi\zeta^2_{i,a-1,t-1} + \xi_i && \text{permanent}
\end{aligned}$$

Taking the first lag of the transitory variance function in above, pre-multiplying by the MA coefficient θ, and applying the law of iterated expectations on the qth-order autocovariance equation ($E_{t-1}[g_{i,t+q+1}g_{i,t}] = E_{t-1}[\epsilon^2_{it}]$), yield the desired relationship for the transitory component equation.

For the permanent component equation, note that the autocovariance equations can be used to obtain:

$$\begin{aligned}
E_{t-q-3}\left[g_{i,a,t}\left(\sum_{j=-q-1}^{q+1} g_{i,a+j,t+j}\right)\right] &= E_{t-q-3}[\zeta^2_{i,a,t}] \\
[\text{LIE}] \Rightarrow &= E_{t-q-3}[E_{t-1}[\zeta^2_{i,a,t}]] \\
&= E_{t-q-3}[\varphi_t + \varphi\zeta^2_{i,a-1,t-1} + \xi_i].
\end{aligned}$$

Also, we can obtain the following using analogous steps:

$$E_{t-q-3}\left[g_{i,a-1,t-1}\left(\sum_{j=-2}^{2} g_{i,a+j-1,t+j-1}\right)\right] = E_{t-q-3}[\zeta^2_{i,a-1,t-1}].$$

Therefore, we simply have:

$$\begin{aligned}
&E_{t-q-3}\left[g_{i,a,t}\left(\sum_{j=-2}^{2} g_{i,a+j,t+j}\right) - \varphi_t - \varphi g_{i,a-1,t-1}\left(\sum_{j=-2}^{2} g_{i,a+j-1,t+j-1}\right) - \xi_i\right] \\
&= E_{t-q-3}\left[\varphi_t + \varphi\zeta^2_{i,a-1,t-1} + \xi_i - \varphi_t - \varphi g_{i,a-1,t-1}\left(\sum_{j=-2}^{2} g_{i,a+j-1,t+j-1}\right) - \xi_i\right] \\
&= E_{t-q-3}\left[\varphi_t + \varphi\zeta^2_{i,a-1,t-1} + \xi_i - \varphi_t - \varphi\zeta^2_{i,a-1,t-1} - \xi_i\right] = 0,
\end{aligned}$$

so we are done.

■

- In this case the moment conditions become

$$\begin{aligned}
&E_{t-3}\left(\Delta g_{i,a+q+1,t+q+1} g_{i,a,t} - d_t^T - \gamma \Delta g_{it+q} g_{i,a-1,t-1}\right) = 0 \qquad \text{Transitory} \\
&E_{t-q-4}\left[\Delta g_{i,a,t}\left(\sum_{j=-(1+q)}^{(1+q)} g_{i,a+j,t+j}\right)\right. \\
&\left. -d_t^P - \varphi \Delta g_{i,a-1,t-1}\left(\sum_{j=-(1+q)}^{(1+q)} g_{ia+j-1t+j-1}\right)\right] = 0 \qquad \text{Permanent}
\end{aligned} \tag{14}$$

  where $\Delta x_t = x_t - x_{t-1}$. **Question: Show this.**
- In practice, however, as Meghir and Pistaferri (2004) found out, lagged instruments suggested above may be only very weakly correlated with the entities in the expectations above.

Problem 10.5. Question: Show why this is true.

Solution. We will use the results in 10.4 to prove this claim. Let $q = 1$. First, consider the transitory part.

Suppose we let $d_t^T = \theta(\gamma_t - \gamma_{t-1})$. Then,

$$\begin{aligned}
E_{t-3}[\Delta g_{i,a+2,t+2} g_{i,a,t} - d_t^T - \Delta\gamma g_{i,a+1,t+1} g_{i,a-1,t-1}] &= E_{t-3}[\Delta g_{i,a+2,t+2} g_{i,a,t} - \theta(\gamma_t - \gamma_{t-1}) - \Delta\gamma g_{i,a+1,t+1} g_{i,a-1,t-1}] \\
\text{[add and subtract } \theta v_i] \Rightarrow &= E_{t-3}[g_{i,a+2,t+2} g_{i,a,t} - \theta\gamma_t - \theta v_i - \gamma g_{i,a+1,t+1} g_{i,a-1,t-1}] \\
&\quad - \underbrace{E_{t-3}[g_{i,a+1,t+1} g_{i,a-1,t-1} - \theta\gamma_{t-1} - \theta v_i - \gamma g_{i,a,t} g_{i,a-2,t-2}]}_{=0 \text{ from 10.4}} \\
&= E_{t-3}[g_{i,a+2,t+2} g_{i,a,t} - \theta\gamma_t - \theta v_i - \gamma g_{i,a+1,t+1} g_{i,a-1,t-1}] \\
\text{[L.I.E]} \Rightarrow &= E_{t-3}[E_{t-2}[g_{i,a+2,t+2} g_{i,a,t} - \theta\gamma_t - \theta v_i - \gamma g_{i,a+1,t+1} g_{i,a-1,t-1}]] \\
[10.4] \Rightarrow &= E_{t-3}[0] = 0,
\end{aligned}$$

so we are done.

Now, consider the permanent part. Suppose we let $d_t^P = \varphi_t - \varphi_{t-1}$, then the permanent moment becomes:

$$\begin{aligned}
&E_{t-5}\left[\Delta g_{i,a,t} \sum_{j=-2}^{2} g_{i,a+j,t+j} - d_t^P - \varphi\Delta g_{i,a-1,t-1} \sum_{j=-2}^{2} g_{i,a+j-1,t+j-1} - \xi_i\right] \\
&= E_{t-5}\left[\Delta g_{i,a,t} \sum_{j=-2}^{2} g_{i,a+j,t+j} - (\varphi_t - \varphi_{t-1}) - \varphi\Delta g_{i,a-1,t-1} \sum_{j=-2}^{2} g_{i,a+j-1,t+j-1} - \xi_i\right] \\
&= E_{t-5}\left[g_{i,a,t} \sum_{j=-2}^{2} g_{i,a+j,t+j} - \varphi_t - \varphi g_{i,a-1,t-1} \sum_{j=-2}^{2} g_{i,a+j-1,t+j-1} - \xi_i\right] \\
&\quad \underbrace{- E_{t-4}\left[g_{i,a-1,t-1} \sum_{j=-2}^{2} g_{i,a+j-1,t+j-1} - \varphi_{t-1} - \varphi g_{i,a-2,t-2} \sum_{j=-2}^{2} g_{i,a+j-2,t+j-2} - \xi_i\right]}_{=0 \text{ from 10.4}} \\
&= E_{t-5}\left[g_{i,a,t} \sum_{j=-2}^{2} g_{i,a+j,t+j} - \varphi_t - \varphi g_{i,a-1,t-1} \sum_{j=-2}^{2} g_{i,a+j-1,t+j-1} - \xi_i\right] \\
\text{[L.I.E]} \Rightarrow &= E_{t-5}\left[E_{t-4}\left[g_{i,a,t} \sum_{j=-2}^{2} g_{i,a+j,t+j} - \varphi_t - \varphi g_{i,a-1,t-1} \sum_{j=-2}^{2} g_{i,a+j-1,t+j-1} - \xi_i\right]\right] \\
[10.4] \Rightarrow &= E_{t-5}[0] = 0,
\end{aligned}$$

so we are done.

■

11 Cross-Section Bias: Age, Period, and Cohort Effectss

Answer the questions in the “Cross-Section Bias: Age, Period, and Cohort Effect” handout.

Problem 11.1. Generalize this analysis for the case of polychotomous variables for age, period, and cohort effect.

Solution. Recall that we have the following definitions for the age effect, the period effect, and the cohort effect.

▷ Age effects are variations linked to biological and social processes of aging specific to individuals.

▷ Period effects result from external factors that equally affect all age groups at a particular calendar time.

▷ Cohort effects are variations resulting from the unique experience/exposure of a group of subjects (cohort) as they move across time.

The generalization that we consider is replacing the continuous variables with polychotomous variables. Recall that the original equation was given as:

$$\ln W = \beta_0 + \beta_1 a + \beta_2 y + \beta_3 e + \beta_4 s + \beta_5 c + u$$

where a, y, c are all continuous variables. Instead, consider replacing each variable on the RHS with a corresponding polychotomous variable. To do this, define the vectors of dummy variables

$$\mathbf{W}, \mathbf{A}, \mathbf{Y}, \mathbf{E}, \mathbf{S}, \mathbf{C}$$

For example, $\mathbf{A}$ is a vector such that for person age a, the ath element is equal to 1 and all the remaining elements are zero. Analogously, we can define the coefficient vectors $\mathbf{B}_j$ appropriately where we normalize the first element in each vector to zero to avoid trivial linear dependencies associated with using dummy variables (see Goldberger, 1968), yielding:

$$\ln \mathbf{W} = \mathbf{B}_0 + \mathbf{B}_1\mathbf{A} + \mathbf{B}_2\mathbf{Y} + \mathbf{B}_3\mathbf{E} + \mathbf{B}_4\mathbf{S} + \mathbf{B}_5\mathbf{C}$$

The corresponding equation for the identities are:

▷ $e_i = a_i - s_i$ is equivalent to requiring that for an observation with the eth element of $\mathbf{E}$ non-zero, the jth element of $\mathbf{A}$ is non-zero if and only if the $(e + j)$th element of $\mathbf{A}$ is non-zero.

▷ $y_i = a_i + c_i$ is equivalent to requiring that for an observation with the yth element of $\mathbf{Y}$ non-zero, the jth element of $\mathbf{A}$ is non-zero if and only if the $(y - c)$the element of $\mathbf{C}$ is non-zero.

So we see that the linear dependencies generated in the polychotomous case is essentially analogous to the continuous case, and the $\mathbf{B}$s can be identified using a similar restriction.

■

12 Keane-Wolpin and Roy

Show the relationship between the Keane-Wolpin model and the Roy model. Relate the Keane-Wolpin "types" to factors in the Roy model.

Solution. Keane and Wolpin (1994) develop a model in which an agent decides among K possible alternatives in each of T (finite) periods of time. With each choice is an immediate reward $(R_k (S (t)))$ that is known to the agent at time t but partly unknown from the perspective of periods prior to t. In this model, unlike the Roy model, the agents are forward-looking and their objective at any time τ is to maximize the expected rewards over the remaining time horizon:

$$\max_{\{d_k(t)\}_{k\in K}} \mathbb{E}\left[\sum_{\tau=t}^{T} \delta^{\tau-t} \sum_{k\in K} R_k(\tau)\, d_k(\tau) \,|S(t)\right]$$

where $S(t)$ is the state space and $d_k(\tau)$ equals one if alternative k is chosen. Therefore, this setup implies that the model is a finite-horizon dynamic programming problem and looks very different from the setup in the Roy model.

While the setup looks quite different, the intuition is rather similar here compared to that of the Roy model. Specifically, they are considering self-selection in three dimensions: schooling, work, and occupational choice. Moreover, they are extending the static deterministic setting of those models to one in which decision making is sequential and the environment is uncertain. Therefore, for example, current school attendance decisions depend on the probabilities attached to future occupational choices.

In the model, the authors define a type k for individuals to introduce heterogeneity. Individuals gain comparative advantage in this manner, and thus each type solves the above dynamic program with different initial conditions. In the original Roy model of self selection, Roy argued that there are three factors affecting the self-selection choice. In this setup, the "types" in Keane-Wolpin corresponds to the fundamental distribution of skills and abilities, which Roy cites as one of the important factors driving occupational choices.

$\blacksquare$