Answer Keys to Problem Set 1

Microeconometrics I with Joan Llull
IDEA, Fall 2024

TA: Conghan Zheng

October 04, 2024

**This document is not intended to be a complete solution, but rather to provide some examples of interpretations and key points for the answer.*

How to make an interpretation[1]

The main principle: try to interpret and comment on them **as if it were your own research**.

- First, comment on the (1) sign (2) magnitude (3) significance of the effects.
- Then think about it further from an economic point of view:
  - Does this result make sense? Does it seem strange? If so, how could it be?
  - How does it compare to your previous results, if any? How could you explain it?
  - Based on all of your results, what conclusion can you draw about the model studied in this exercise?
- You don't have to go through all of the above steps of interpretation in every exercise. However, the more I can see that you have tried to make economic sense of what you find, the more tolerant I will be of your mistakes in code or interpretation.

[1] Credit for this page goes to Sophie Brochet.

Exercise 1.1

- Estimate β_1 in

$$w_{it}^H = \beta_0 + \beta_1 immigr_i^H + X_{it}\gamma + \eta_i + \varepsilon_{it}$$

using the RE model. Compare with the results from the FE model.

	(1) RE	(2) FE
h_immigr	-0.0687**	0
	(0.0233)	(.)
age	0.0804***	0.0785***
	(0.00215)	(0.00228)
age2	-0.000791***	-0.000753***
	(0.0000251)	(0.0000265)
female	-0.427***	0
	(0.0144)	(.)
black	-0.174***	0
	(0.0260)	(.)
asian	0.179***	0
	(0.0307)	(.)
_cons	6.533***	6.461***
	(0.0462)	(0.0488)
N	358688	358688
adj. R-sq		-0.030

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

Exercise 1.1

	(1) RE	(2) FE
h_immigr	-0.0687**	0
	(0.0233)	(.)
age	0.0804***	0.0785***
	(0.00215)	(0.00228)
age2	-0.000791***	-0.000753***
	(0.0000251)	(0.0000265)
female	-0.427***	0
	(0.0144)	(.)
black	-0.174***	0
	(0.0260)	(.)
asian	0.179***	0
	(0.0307)	(.)
_cons	6.533***	6.461***
	(0.0462)	(0.0488)
N	358688	358688
adj. R-sq		-0.030

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

- In an attempt to use the within-group estimator, the dummies for high-skilled immigrant, female, black, and Asian were dropped because the individual fixed effects already capture these **time-invariant** characteristics.

Exercise 1.2

- Compare FE, RE with OLS

	(1) RE	(2) FE	(3) OLS
h_immigr	−0.0687**	0	−0.0986***
	(0.0233)	(.)	(0.00436)
age	0.0804***	0.0785***	0.111***
	(0.00215)	(0.00228)	(0.00133)
age2	−0.000791***	−0.000753***	−0.00121***
	(0.0000251)	(0.0000265)	(0.0000157)
female	−0.427***	0	−0.429***
	(0.0144)	(.)	(0.00266)
black	−0.174***	0	−0.205***
	(0.0260)	(.)	(0.00506)
asian	0.179***	0	0.166***
	(0.0307)	(.)	(0.00572)
_cons	6.533***	6.461***	6.151***
	(0.0462)	(0.0488)	(0.0273)
N	358688	358688	358688
r2		0.00542	0.102
r2_a		−0.0304	0.102

Standard errors in parentheses
* p<0.05, ** p<0.01, *** p<0.001

- The estimated effects are similar in magnitude and sign across models. However, one may notice that OLS yields slightly larger effects; the **omitted variables** problem may have caused OLS to overestimate the wage gap.

Exercise 1.3

- Test RE vs. FE assumption

	Coefficients (b) FE	(B) RE	(b-B) Difference	sqrt(diag(V_b-V_B)) Std. err.
age	.0784598	.0803595	-.0018996	.0007729
age2	-.0007529	-.0007906	.0000376	8.43e-06

b = Consistent under H0 and Ha; obtained from **xtreg**.
B = Inconsistent under Ha, efficient under H0; obtained from **xtreg**.

Test of H0: Difference in coefficients not systematic

chi2(2) = (b-B)'[(V_b-V_B)^(-1)](b-B)
= **53.91**
Prob > chi2 = **0.0000**

- Why do I get negative χ^2 that cannot be?[2]
  - Negative χ^2: The assumptions on which the Hausman test is based did not hold.
  - Reason: You used the `constant` option on the `hausman` command in Stata. The `constant` option should be included when you believe that the constant has a common interpretation across the two models.
  - Theoretically, the `sigmamore` or `sigmaless` option in Stata could help you solve this problem, but that is not the case here.

[2] Credit: Sophie Brochet

Exercise 1.4

$$y_{jst} = \theta p^H_{jst} + Z_{jst}\delta + \mu_j + \nu_s + \phi_t + (\mu_j \times \nu_s) + (\mu_j \times \phi_t) + (\nu_s \times \phi_t) + \xi_{jst}$$

Regression Results on Yearly Wage

	Dependent Variable:			
	Log(Wage)		Hours Worked	
	(1)	(2)	(3)	(4)
Share: H-skill migrant	-0.005	0.272	-6.432	2.080
	(0.230)	(0.215)	(14.863)	(13.726)
Share: H-skill	0.267***		7.199	
	(0.062)		(4.008)	
Share: Female	-0.144**		-0.617	
	(0.054)		(3.517)	
Share: Black	-0.286*		-27.119***	
	(0.121)		(7.801)	
Share: Asian	0.433		13.799	
	(0.222)		(14.321)	

- For an unbalanced panel, we use sample weights when aggregating the data to a lower frequency to keep the data representative.
- High-skilled Workers vs. High-skilled Migrants

Exercise 2.1

- Estimate

$$h_{it} = \alpha(a_{it}) + \beta(a_{it})married_{it} + x'_{it}\gamma + \eta_i + \varepsilon_{it}$$

- $\alpha(a)$ and $\beta(a)$ are functions of age a. That is, you should estimate the equation for each age group a.
- For simplicity, alpha and beta are constants for each age. This is similar to LSDV, you add a set of age and age×married dummies to your command.
- Then $\alpha(a)$ is the health curve for singles, $\alpha(a) + \beta(a)$ is the one for married people, and $\beta(a)$ is the marriage health gap.
- Note: You can't include all the age categories and a constant term at the same time, there will be perfect collinearity.

Exercise 2.2

- Estimate

$$h_{it} = \phi h_{it-1} + \alpha(a_{it}) + \beta(a_{it}) married_{it} + x'_{it}\gamma + \eta_i + \varepsilon_{it},$$

- Why dynamic panel? With ϕh_{it-1} and dynamic panel methods, we can deal with:
  - feedback from health shocks on marriage probabilities, and
  - self-selection due to innate unobserved heterogeneity

Exercise 2.1 and 2.2

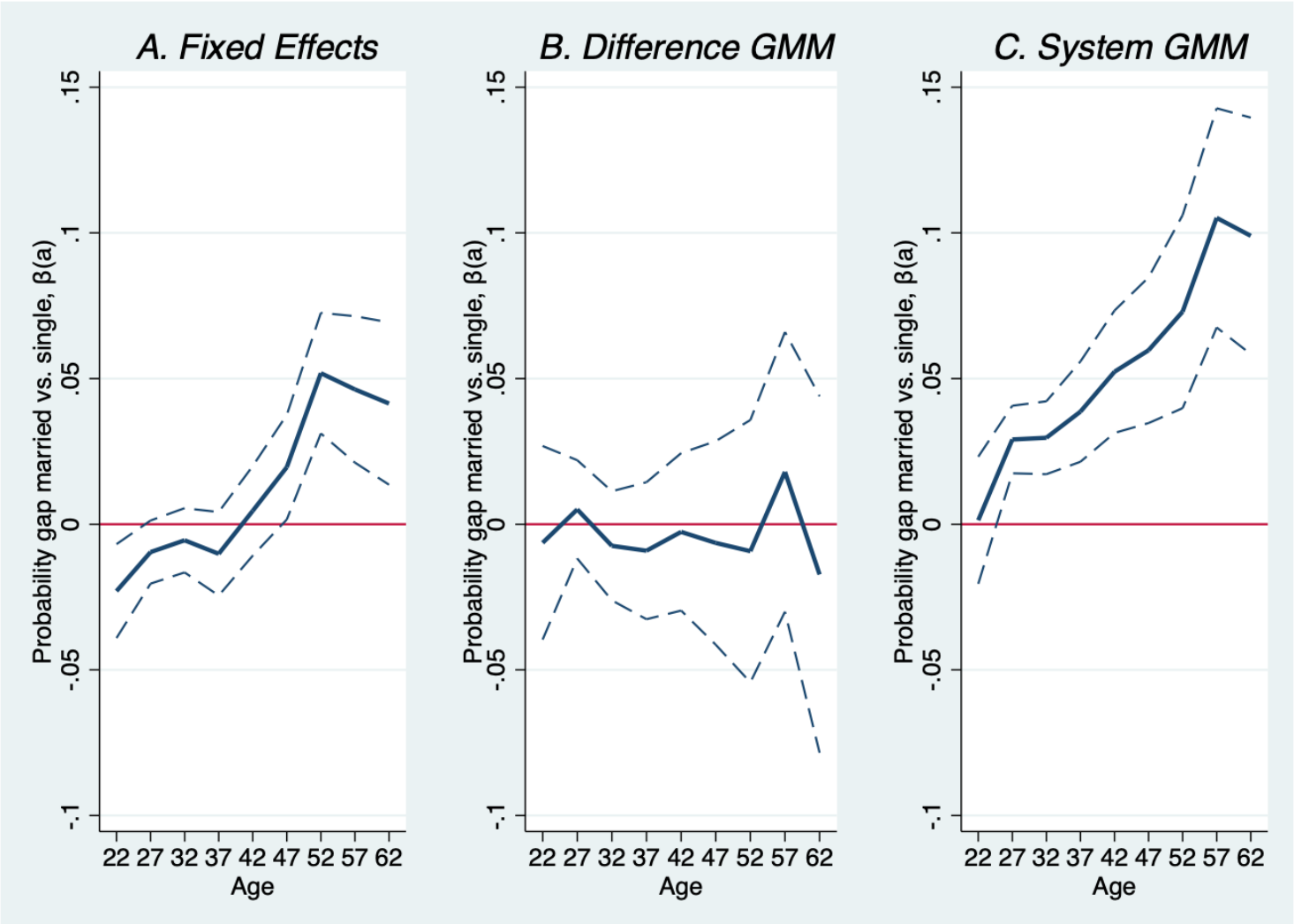
Interpretation of the Results: Example[3]

- In the following, I will discuss the behaviour of the coefficients (the signs, their magnitude and how they change by ages) for the three (Static FE, A-Bond, A-Bover) at the same time, given that they exhibit similar behaviours.
  - I've omitted the age group of 62 years, which means that the coefficients of the rest of the age groups should be interpreted as differences with respect to this group.
  - A positive and significant coefficient for the age groups ranging from 22 up to 47 imply that, in comparison to the oldest age group, the former ones enjoy a better health.
  - Another interesting feature is the effect that being married has on the reported health of the individual ...
  - ...
  - In A-Bond and A-Bover esimations with up to 4 lags (we put restriction on the number of lags to fix the overfitting problem), I observe that the lags of the dependent variable have a positive effect on the current reported health of the individuals, so here we could presume that we did well by introducing a dynamic panel data approach.
  - ... When we include the dynamics, the fact that childbearing is a decision based on past characteristics and results in a service flow is taken into account.
  - We can also see how the income and education affect positively to the reported health, which can be understood as ...

[3] Credit goes to Naia Ormaza Zulueta. With her permission, the text here is adapted from her solution in Fall 2021.

Exercise 2.3

- Plot marriage health gap (don't forget the CI bands)



Marriage Health Gap: Indentification

Guner et al., 2018

$$h_{it} = \phi h_{it-1} + \alpha(a_{it}) + \beta(a_{it}) married_{it} + x'_{it}\gamma + \eta_i + \varepsilon_{it},$$

- Assumption: additive separability of η
- If the effects α and β are not time-varying for a given individual (one is not observed over different ages), then identification relies exclusively on individuals who change their marital status.
- **Problem**: usually you don't observe enough people changing marital status within the time horizon of your sample.
- **Solution**: time-varying $\alpha(a)$ and $\beta(a)$ for an given individual.
- Those who never married in your data contribute to the married health curve, those who always married contribute to the single health curve, those who changed marital status contribute to the marriage health gap.