APEC 8213: Econometric Analysis III
Spring 2026
Professor: Paul Glewwe
Teaching Assistant: Matt Braaksma

Problem Set 2

Handed out: February 2, 2026
Due: February 11, 2026

1. GMM and 2SLS

(30 points total)

Suppose we have the linear equation $Y = X'\beta + e$ with two sets of instruments Z_1 and Z_2. Then consider the following estimators of β:

$\hat{\beta}$: 2SLS using the instruments Z_1

$\tilde{\beta}$: 2SLS using the instruments Z_2

$\bar{\beta}$: GMM using the instruments $Z = (Z_1, Z_2)$
and the weight matrix $\boldsymbol{W} = \begin{pmatrix} (\boldsymbol{Z_1'Z_1})^{-1}\lambda & 0 \\ 0 & (\boldsymbol{Z_2'Z_2})^{-1}(1-\lambda) \end{pmatrix}$

for $\lambda \in (0, 1)$.

a) Find an expression for $\bar{\beta}$ which shows that it is a specific weighted average of $\hat{\beta}$ and $\tilde{\beta}$. **(28 points)**

b) Is this an efficient weight matrix to use for GMM? (Short answer, 1-2 sentences) **(2 points)**

2. Multinomial Logits and Probits

(50 points total)

Load the `heating` data, which contains a sample of 900 Californian households and their choice of heating system.

- `idcase`: id
- `depvar`: heating system
  - `gc` (gas central)
  - `gr` (gas room)
  - `ec` (electric central)
  - `er` (electric room)
  - `hp` (heat pump)
- `ic.z`: installation cost for heating system z (defined for the 5 heating systems)
- `oc.z`: annual operating cost for heating system z (defined for the 5 heating systems)
- `income`: annual income of the household
- `agehed`: age of the household head
- `rooms`: numbers of rooms in the house
- `region`

a) Estimate a multinomial logit model for heating system choice using only house characteristics. Report and interpret the coefficients. **(6 points)**

b) Estimate a conditional logit model using only the system-specific variables. Report and interpret the coefficients. **(6 points)**

c) Estimate a mixed logit model including both house and system-specific variables. Compare the results to (a) and (b). **(6 points)**

d) Estimate a mixed logit model that allows installation and operating costs to have alternative-specific effects. Test whether this more flexible specification is warranted compared to the mixed model in (c). **(7 points)**

e) Using the model from part (c), calculate the marginal effects of household income on the probability of choosing each heating system. Evaluate these at the mean values of all other covariates. Interpret your results. **(5 points)**

f) Using the same model, calculate how the probability of choosing a heat pump changes if its installation cost decreases by 20%. Compare this to the marginal effect of installation cost. **(5 points)**

g) Estimate two nested logit models for heating system choice. For each model: (i) report and interpret the inclusive value (dissimilarity) parameters for each nest, and (ii) test whether each inclusive value parameter is statistically different from 1 and explain how to determine if the nested structure is justified (show your code and results). **(7 points)**

- Model 1: Nests by energy source: (1) gas (gc, gr), (2) electric (ec, er, hp)
- Model 2: Nests by system type: (1) room systems (gr, er), (2) central systems (gc, ec, hp)

h) Combine the five heating systems into three categories: gas (`gc`, `gr`), electric (`ec`, `er`), and heat pump (`hp`). Estimate both a mixed multinomial probit and a mixed multinomial logit model for this grouped outcome. Compare the results to the nested logit from (g). **(8 points)**

3. Ordered Logits and Probits

(20 points total)

Load the `Persistence_preferences_rural_Guatemala` data from the replication files for the paper "Persistence of Individual and Social Preferences in Rural Settings." This dataset contains a panel of 1,262 agricultural households in rural Guatemala, surveyed annually from 2019 to 2022. It includes self-reported preference indicators, household socioeconomic characteristics, and contextual variables.

- `trust_people_0`: trust in people, scale 1–4 (1 = almost always can trust, 4 = almost always must be very careful)
- `head_compprim_above`: household head completed elementary or above
- `head_male`: household head is male
- `head_age`: household head age
- `head_spanish`: household head's main language spoken is Spanish
- `t_TV_radio`: household head owns TV or radio
- `HH_poor_PL2011`: household is poor (below $1.90/day, 2011 PPP)
- `Tot_AgriLand_ha`: household head agricultural land size (hectares)

a) Estimate an ordered logit model for trust in people (`trust_people_0`) as a function of household head education (`head_compprim_above`), sex (`head_male`), age (`head_age`), main language spoken (`head_spanish`), TV/radio ownership (`t_TV_radio`), poverty status (`HH_poor_PL2011`), and agricultural land size (`Tot_AgriLand_ha`). Report and interpret the coefficients. **(5 points)**

b) Estimate an ordered probit model for the same outcome and predictors. Compare the results to the ordered logit model. **(5 points)**

c) For a female household head, age 40, with completed elementary education, not poor, owns a TV or radio, main language spoken is Spanish, and owns 0.8 ha of agricultural land, calculate the predicted probability of each trust category using the ordered logit model. **(5 points)**

d) Discuss how the predicted probabilities change if the household head is male, holding other characteristics the same. **(5 points)**