Econ 312 Part A, Spring 2023

Final Exam

James J. Heckman

Due May 1st, 2023 at 11:59pm

This draft: April 24, 2023

1. [**5 pts**] Explain and contrast the roles of the probability of selection, $P(Z)$, sometimes called the "propensity score" in

   (a) Matching

   (b) Selection Models

   (c) Instrumental variable models

   In what sense is it a fundamental aspect of each model for identification? For estimation?

2. [**5 pts**] Using the Generalized Roy model, derive and interpret the weights on MTE required to form:

   (a) ATE

   (b) Treatment on the treated

   (c) The voting criterion

   (d) AMTE

   (e) PRTE

   (f) Quantile treatment effects

   (g) The IV estimator with Z being a vector of instrumental variables

(h) The matching estimator

3. [15 pts] **NOTE: the model, dataset, and parts of this question have been modified from the first issue of the exam. For example, the previous part (c) was dropped.** This question considers the time series earnings and consumption dynamics model presented in the handout, "Modelling the Income Process:"

$Y_{it} = \beta + U_{it},\ 1 \leq t \leq T$ (income);

$U_{it} = U_{i(t-1)} + \varepsilon_{it},\ \varepsilon_{it} = \delta\eta_{i(t-1)} + \eta_{i,t},\ \eta_{i,t}$ iid;

$C_{it} = \gamma E_t\left(\sum_{j=0}^{T-t} \frac{Y_{i(t+j)}}{(1+r)^j} \mid \mathcal{I}_t \right) + \omega_{it}$; (consumption, where $\mathcal{I}_t$ is the information at age t on which agents act)

$\omega_{it} = \psi_i + \lambda_{it}, \lambda_{it} \perp\!\!\!\perp \eta_{it'}$ all t, t'

$E(\lambda_{it}) = 0,\ E(\eta_{it}) = 0$ for all t

All innovations are mutually independent across i.

This question uses the dataset "q3_dt.csv" to answer the computational components, with $r = 0.1$.

(a) Show how to identify the contribution of the permanent income component, the permanent consumption component, and the transitory income component to income and consumption (this is % of variance explained). The permanent component of consumption is the part of consumption attributable to permanent component of income.

(b) Compute the contribution of the components listed in (a) and comment on the estimates.

(c) Write out and estimate all moment conditions for identifying the consumption and income models. State the information sets used by agents.

4. **[10 pts]** Using the handout, "Ability Bias, Errors in Variables, and Sibling Methods Panel Data" as a guide and factor-analyzing the given measurement system:

   (a) What is the rank of the factor model for measurements?

   (b) Consider a true model $Y_i = \alpha + \beta X_i + \varepsilon_i$. You observe (Y_i, X_i^*), where X_i^* is X_i measured with error so that $X_i^* = X_i + \eta_i$ with $E(\eta_i) = 0$, $\eta_i \perp\!\!\!\perp (X_i, \varepsilon_i)$. (ε_i, η_i) are iid and mutually independent and Var $\eta_i = \sigma_\eta^2$. You have $J > 1$ measurements on X.
   The measurement system is:

$$\begin{aligned} M_{ij} =& \lambda_j X_i + \omega_{ij} \\ j =& 1, \dots, J \text{ measurements} \\ E(\omega_{ji}) =& 0 \\ \omega_{ij} \perp\!\!\!\perp & \varepsilon_i, \text{ for all } j. \end{aligned}$$

   This is a factor model. Its rank is determined by the covariance matrix of M_{ij}.
   Show how to identify β using these measurements. Using the dataset "q4_dt.csv," which assumes that X_i^* and X_i are scalar variables, estimate β and provide standard errors using IV with the M_{ij} as instruments for X_i^*. All error terms are independent across i.

(c) Now suppose that $X_i = (X_{i1}, X_{i2})$ is a vector, with $\beta' = (\beta_1, \beta_2)'$. Suppose that only the first component is measured with error $X_{1i}^* = X_{1i} + \eta_{1i}, X_{2i}^* = X_{2i}$. You observe $(Y_i, X_{i1}^*, X_{i2}^*) = (Y_i, X_{i1}^*, X_{i2})$. Consider two approaches to estimating β_2:

    (i) Dropping X_{i1}^* and fitting the model using (Y_i, X_{i2}).

    (ii) Keeping X_{i1}^* and fitting the model using (Y_i, X_{i1}^*, X_{i2})

    Which approaches yields the smallest bias for β_2? The smallest mean squared error?

5. **[20 pts]** Consider the model

$$
\begin{aligned}
&Y_i = \alpha_i + \beta_i D_i \\
\text{where } &E(\alpha_i) = \bar{\alpha}, |\bar{\alpha}| < \infty \\
&E(\beta_i) = \bar{\beta}, |\bar{\beta}| < \infty \\
&Var\ \alpha_i < \infty, \quad Var\ \beta_i < \infty \\
&D_i = 1(\beta_i - C_i > 0)
\end{aligned}
$$

where $C_i = \psi(Z_i) + U_i$, with $U_i \perp\!\!\!\perp (\alpha_i, \beta_i)$ and $Z_i \perp\!\!\!\perp (\alpha_i, \beta_i, U_i)$. You observe (Y_i, D_i, Z_i).

This question uses the dataset "q56_dt.csv" to answer the computational components, where you will use the variables (indiv,Z1,Z2,D_q5,Y_q5) to answer them.

(i) Define sorting bias and selection bias for this model.

(ii) How does this model relate to the model of potential outcomes used by statisticians?

(iii) Define "monotonicity" in the sense of Imbens and Angrist and contrast with standard usage in mathematics.

(iv) Define the MTE for this model.

(v) Suppose $Z_i = (Z_{i1}, \ldots, Z_{iK})$, $K > 1$ but you use Z_{i1} alone to instrument for D_i. Derive the expression for IV using Z_{i1} applied to D_i. Does this IV satisfy the Imbens-Angrist "monotonicity" condition? Why or why not? Explain when it does. Derive the IV (Z_{i1}) weight for the MTE.

(vi) Is what IV identifies in the case considered in part (v) ever economically interesting? Suppose that you estimate the model using the first coordinate of Z as an instrument, but condition all the other instrument values at realized values. Formally characterize the estimand and compare with the estimands for (v). Does this instrument necessarily satisfy monotonicity? Compare to the case where you fix $(Z_2, Z_3, \ldots, Z_K) = (z_2, \ldots, z_K)$ at the *same* values across people.

(vii) Compare and interpret your estimates for (v) and (vi).

Illustrate **all** of your answers empirically using the posted dataset.

6. [**5 pts**] Consider the questions in Question (5), but now suppose that

$$D_i = \mathbf{1}(E(\beta_i - C_i | Z_i) > 0)$$

. Answer all the questions in Question (5) for this model. This question uses the dataset "q56_dt.csv" to answer the computational components, where you will use the variables (indiv,Z1,Z2,D_q6,Y_q6) to answer them.

7. [**5 pts**] What does matching on $P(Z)$ (nearest neighbor) identify for the data and model of Questions (5) and (6)? Using the relevant datasets, estimate and plot the MTE as a function of P for the general model and for the matching model. Compare to the estimator of $E(Y_1 - Y_0)$ based on matching on P $(E(Y_1 - Y_0|P = p))$.

8. [**35 pts**] Consider the model of earnings and program participation developed in Heckman (1977), Heckman and Robb (1985), and applied in Ashenfelter and Card (1984).

   Training only occurs in period k and there are no earnings. Opportunity costs (foregone earnings) in period k are Y_{0ik}, and tuition costs or subsidy, C_i.

   Individual i's potential earnings are determined by the following equations:

$$Y_{0it} = \beta + \theta_i + U_{it} \quad \text{all } t \tag{1}$$

and

$$Y_{1it} = Y_{0it} + \alpha_i \quad t > k \tag{2}$$

so that the individual causal effect of training is:

$$Y_{1it} - Y_{0it} = \alpha_i \quad t > k. \tag{3}$$

The error term U_{it} is an AR(1) process:

$$U_{it} = \rho U_{i(t-1)} + \varepsilon_{it} \tag{4}$$

where ε_{it} are iid shocks with $E(\varepsilon_{it}) = 0$ and $\theta_i, \varepsilon_{i,t}$, and α_i are mutually independent and independent across people.

The decision to participate made in period k is:

$$D_i = \begin{cases} 1 \quad \text{iff } \alpha_i/r - Y_{0ik} - C_i > 0 & \text{and } t > k \\ 0 & \text{otherwise} \end{cases} \tag{5}$$

where $C_i = Z_i'\phi + V_i$ with $E(V_i) = 0$ and where Z_i is an observed characteristic that affects cost of training, $(U_{it}, \theta_i, V_i) \perp\!\!\!\perp Z_i$, $V_i \perp\!\!\!\perp (\alpha_i, \theta_i$. We observe Y_{it} where $Y_{it} = Y_{0it}$ for $t < k$, $Y_{it} = (1 - D_i)Y_{0it}$ for $t = k$, and

$$Y_{it} = D_i Y_{1it} + (1 - D_i)Y_{0it}. \tag{6}$$

The mean effect of training in the population: $E(\alpha_i)$.

Mean effect of training on those who receive training is $E(\alpha_i | D_i = 1)$.

The three samples (datasets) you will use in this question are called "q8s1_dt.csv", "q8s2_dt.csv", and "q8s3_dt.csv." The three samples were constructed using one of the three data generating process: (i) α_i is heterogeneous but individuals choose treatment using $E[\alpha_i]$ instead of α_i, (ii) α_i is homogenous (constant), and (iii) no restrictions (heterogeneous α_i which is also known in making treatment choice). As part of answering the questions, you should determine which sample corresponds to which data generating process.

(a) For this model, plot the panel earnings dynamics for all three posted samples. Use the estimated propensity score to test for essential heterogeneity ($\alpha_i \not\perp\!\!\!\perp D_i$) "sorting."

(b) For each sample, estimate $E(\alpha), E(\alpha|D=1)$ using

  (i) IV estimator (with $P(Z)$ as instrument)

  (ii) The selection bias (i.e., control function) estimator

  (iii) The difference in differences estimator (-1,3) where -1 refers to the period just before k (i.e. $k-1$) and 3 refers to 3 periods after: k, (i.e. $k+3$).

  (iv) The difference in differences estimator (-3,3)

  (v) The cross section estimator comparing participant and nonparticipant earnings in period $k+4$

  (vi) The nearest neighbor matching estimator based on preprogram earnings.

  Explain the differences among the estimators and estimates within each sample and across samples.

(c) Under what conditions does the model satisfy the parallel trends assumption required in difference in differences estimators? Which versions of the model guarantee that this condition is met? Consider three cases:

  i. Agents know $Y_{i,k}$ and C_i when deciding if $D_i = 1$.

ii. Agents do not know $Y_{i,k}$ and C_i and use expected values where the expected values are true (population) expected values.

iii. A version of (ii) where agents anticipated values of Y_{ik} are biased down by $\omega_{i,k} = \$1,000$ and C_i up by \$200.