1. (35 points) In a recent study, Clark Nardinelli and Curtis Simon\(^1\) test whether there is "customer discrimination" against blacks and Hispanics in the market for baseball cards. The model they use to test for discrimination is a tobit model. The dependent variable is the log(Pi/Pc) where Pi is the price of player i’s card and Pc is the price of a common card (the lower bound on card prices). The independent variables are listed in the table below along the tobit coefficients and t-statistics.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Coefficient</th>
<th>T-statistic</th>
<th>Mean of variable</th>
<th>Mean of X * coefficient</th>
</tr>
</thead>
<tbody>
<tr>
<td>Constant</td>
<td>-1.28</td>
<td>1.31</td>
<td>1</td>
<td>-1.28</td>
</tr>
<tr>
<td>Black</td>
<td>-0.20</td>
<td>1.40</td>
<td>0.23</td>
<td>-0.05</td>
</tr>
<tr>
<td>Hispanic</td>
<td>-0.55</td>
<td>2.90</td>
<td>0.14</td>
<td>-0.08</td>
</tr>
<tr>
<td>Hits</td>
<td>0.0030</td>
<td>4.50</td>
<td>944</td>
<td>2.83</td>
</tr>
<tr>
<td>Doubles</td>
<td>0.0008</td>
<td>0.50</td>
<td>149</td>
<td>0.12</td>
</tr>
<tr>
<td>Triples</td>
<td>0.0020</td>
<td>0.60</td>
<td>26</td>
<td>0.05</td>
</tr>
<tr>
<td>Home runs</td>
<td>0.0032</td>
<td>4.40</td>
<td>95</td>
<td>0.30</td>
</tr>
<tr>
<td>Sigma</td>
<td>0.6400</td>
<td>14.80</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Sigma represents the standard deviation of the residual in the tobit regression. The last column provides the product of the coefficient and the mean of the relevant explanatory variables.

a. Based on the information provided, for the person with the characteristics equal to the sample mean but who is white (not black or Hispanic), compute the probability that the player’s card price exceeds the common price (i.e. that log(Pi/Pc)>0). Show your work. Tables are attached for the standard normal p.d.f. and c.d.f.

c. If Pc=$1, what is the predicted card price (in $) for the person described in (a)? Note that log(.) is the natural log. Show your work.

d. Based on the information provided, compute the expected value of log(Pi/Pc) for the person described in (a) conditioned upon knowledge that the player’s card price exceeds the common card price. Show your work.

e. For the person in (a), what is the marginal effect of an additional home run on the expected value of log(Pi/Pc)? Explain.

2. (35 points) Some utility companies charge according to "time of use" (TOU). The idea is that peak loads are more costly to utility companies so they charge higher prices during the peak periods. A recent study criticized earlier work that had not controlled for sample selection bias in earlier studies of how TOU pricing would affect utility consumption. In the "experimental" studies, consumers were given a choice of TOU pricing where prices were higher in the peak than the off-peak, as opposed to uniform pricing where prices were identical at all times.

Suppose for example, that the pricing schedules are:

TOU: \[ \text{utility cost} = 0.02(\text{peak KW}) + 0.01(\text{non-peak KW}) \]
Uniform: \[ \text{utility cost} = 0.015(\text{peak KW} + \text{non-peak KW}) \]

where peak KW and non-peak KW refers to the kilowatts consumed in the peak versus off-peak time periods.

a. What type of consumer is likely to choose the TOU scheme? Explain and be as precise as possible in describing the conditions under which the consumer would choose TOU.

Suppose that you estimate a regression of peak usage (i.e. peak KW) on family characteristics (e.g. income, family size, size of house) and a dummy for TOU pricing.

b. Would you expect a positive or negative coefficient on the TOU dummy? why?

c. Would you expect the coefficient on the TOU dummy to be upward or downward biased given the sample selection problem? Justify your answer.

d. Suppose that you re-estimated the model described above and included the appropriate inverse Mills ratios in the model. Would you expect a positive or negative coefficient estimate on the inverse Mills ratios? Why? What does the coefficient represent?

3. William Evans and Ioannis Kessides\(^2\) studied the effect of local market power on airline pricing using the following model.

\[ \ln(P_{ij}) = X_{ij}B + v_j + e_{ij} \]

where \(P_{ij}\) is the price that airline \(i\) charges on route \(j\), \(X_{ij}\) is a vector of coefficients describing airline \(i\)'s service in route \(j\), \(v_j\) is an intercept specific to route \(j\) that the authors treat as either a fixed or random effect, and \(e_{ij}\) is an i.i.d. error term. Each city pair is a separate route (e.g. Cincinnati/Phoenix is a route).

The authors estimate several variations of the model: OLS, (route specific) random effects, and (route specific) fixed effects. The results are given in the table below. For the fixed effects model, consider only the first column of estimates.

---

a. In the OLS and random effects model, ln(Milesj) [the log of the number of miles for route j] and its square are included. Why can't this variable be included in the fixed effects model? Explain why the regression would "break down" and justify your answer.

b. The authors test whether the assumptions necessary for the random effects model are satisfied by testing for equality of the random and fixed effects estimates. They reject the hypothesis that the coefficients are equal. What assumptions are necessary for the random effects model to be an appropriate representation of the fixed effects model? Describe the assumptions in the context of the regression represented.

c. Comparing the ols and fixed effects estimates, the impact of airport market share increases from .201 to .562 when fixed effects are accounted for. The measure of market share is a firm's sales in the airport as a percent of all sales in the airport. What could account for the large increase? What does it say about the relationship between the fixed effects and other variables in the regression? Explain and justify your answer.