Question 1:

On July 1st, the current BC sales tax will be replaced by a harmonized sales tax (HST) that will integrate the old sales tax into the federal GST. The total tax rate will now be 12 percent on all goods and services. This represents a tax increase for some services (e.g. restaurant meals) that did not used to be taxed under the old system, but a tax decrease for some other goods and services. The detailed information about taxes for each good and service before and after the introduction of the HST are available on the web site of the BC government.

You are hired by a marketing firm that has access to detailed data on the spending habits of a sample of 1000 British Columbians every month since January 2008. These data consist of monthly consumption spending for 100 different goods and services. They also contain some additional information about the age, education, gender and income of the people in the samples. Your first duty on the job is to write a short proposal explaining how you would plan to use these data for January 2008 to December 2010 to analyse the impact of sales tax changes on consumption. You need to include the following two elements in your proposal:

a) Write down two equations you plan to estimate using both the data at the individual level, and data at the aggregate level (average consumption over sample individuals for each good or service). Be specific in discussing which variables are included in the model, and what are the assumptions you are making about the error term.

b) Present the estimation method(s) you plan to use to estimate the model.

c) Your boss then gets your proposal reviewed by an external expert from UBC who asks you to provide a short answer to each of the following queries:

i) How do you plan to compute the standard errors on the estimated effect of the tax rate when using individual data?

ii) How do you plan to deal with the fact that some individuals have a zero consumption of some of the goods or services during a given month?

iii) The current BC sales tax is often set in response to consumption patterns the government wants to change. For instance, there is currently no BC sale tax on bicycles (the government wants to encourage more people to use bicycles), while there is a 21 percent tax on parking (the government wants to discourage people from using cars). Since this means the current sales tax is arguably endogenous, aren’t you worried your estimates will be biased?
Question 2:

Having successfully completed the task presented in Question 3, you get promoted to the “beer” department of your marketing firm. You now have to review an earlier study done by another economist working for the firm that looks at the choice of beer brand among people, as well as the total monthly consumption of beer. Another data set is available to look at this issue. This time you have a monthly sample of 2000 Canadians from all 10 provinces, with detailed information on monthly consumption of beer (number glasses or bottles consumed), as well as preferred brands (the one they drank most during the month). In this case, you have data from January 2005 to December 2009. The key right hand side variable of interest is the amount spent on advertising for each brand during a given month in a given province.

The other economist, who only has a BA, shows you results obtained by running simple OLS models. For instance, the OLS estimates for total monthly consumption (Total) is:

\[

total = 5.2 + 3.1 \text{ male} + 0.2 \text{ age} - 0.1 \text{ education} + 0.05 \log(\text{Advertisement})
\]

\[
\begin{array}{c}
(0.7) \\
(1.1) \\
(0.09) \\
(0.05) \\
(0.02)
\end{array}
\]

The standard errors (in parentheses) are simply obtained using the standard formula for OLS standard errors. Note that in this case, advertisement is total advertisement spending for all beer brands (in each month and province).

The other economist then shows you the models for the choice of brand. For instance, in the case of Sleeman, the estimated model (using OLS again) is:

\[
sleeman = 0.02 + 0.005 \text{ male} + 0.003 \text{ age} + 0.001 \text{ education} + 0.03 \log(\text{Ad}_{\text{Sleeman}})
\]

\[
\begin{array}{c}
(0.004) \\
(0.004) \\
(0.005) \\
(0.002) \\
(0.01)
\end{array}
\]

where Sleeman is a dummy variable indicating whether Sleeman is the preferred brand of the consumer, and Ad_{Sleeman} is the amount of advertising expenses on Sleeman only. Once again, the model is estimated by OLS, and the standard errors are the usual OLS standard errors. The other economist tells you that the results for Sleeman are typical in the sense that advertisement spending has a positive and significant effect for 12 of the 15 brands studied.

Your boss wants to know whether you think there is enough evidence there to tell beer company executives that advertisement does have a positive and significant effect on beer consumption. More specifically, you have to explain what is potentially wrong with the econometric analysis presented above, and how you would suggest to “fix it” using all the great tools you have learned in ECON 594 at UBC. [SUGGESTION: use a couple of bullet points for “what is wrong” and “what should be done instead”. Focus on specific problems and recommendations instead of generalities].
Question 3:

This question and the next one are based on an actual panel data set (the National Longitudinal Survey of Youth, NLSY) for years 1985 to 1994. In this first question we are interested in estimating the effect of education on log wages, controlling for a number of other covariates such as labour market experience (and its squared value) and a linear time trend. These variables are included in all the regressions discussed below, but their coefficients are not reported to save space.

First consider the coefficients estimated using OLS and first differences. Standard errors are reported using three different methods: conventional OLS standard errors, standard errors robust to arbitrary forms of heteroskedasticity, and clustered (on the individual) standard errors:

<table>
<thead>
<tr>
<th>Estimation method:</th>
<th>OLS</th>
<th>First-differences</th>
</tr>
</thead>
<tbody>
<tr>
<td>Estimated coefficients:</td>
<td>0.0861</td>
<td>0.0207</td>
</tr>
<tr>
<td>Standard errors: OLS:</td>
<td>0.0009</td>
<td>0.0059</td>
</tr>
<tr>
<td>Robust:</td>
<td>0.0010</td>
<td>0.0080</td>
</tr>
<tr>
<td>Clustered:</td>
<td>0.0021</td>
<td>0.0079</td>
</tr>
</tbody>
</table>

a) The reported standard errors are quite sensitive to the procedure used to compute them. First, standard errors are always larger when the first-differenced estimator is used instead of OLS. Second, OLS standard errors are relatively unaffected by heteroskedasticity, but more than double when we cluster. Third, unlike the case of OLS, clustering actually reduces standard errors (relative to robust) in the first difference specification. Explain as best as you can why we see these three patterns in the standard errors.

b) Which method of computing standard errors would you recommend in each of these two cases (OLS and first-differences)? Explain.

You now estimate the model using the random effect and fixed-effect/within methods. The random effect estimator is 0.0760 (clustered standard error of 0.0021) while the fixed effect estimator is 0.0274 (clustered standard errors of 0.0067). When you perform a Hausman test for these two models you get a chi-square statistics of 285 with a p-value of 0.0000.

c) Discuss two possible scenarios under which, as observed in the above results, the first-difference and within estimators are substantially lower than the OLS and random effect estimators. What is your preferred estimator under each of the two scenarios? Should you use the Hausman test to select your preferred estimator under these two scenarios?
Question 4:

This time we only use a single cross-section from the NLSY for the year 1990. In addition to the explanatory variables used in question 1 (education, years of experience and its square), we also include a dummy variable for union status on the right hand side of the log wage equation.

The results below show a number of estimates of the coefficient on the union status in regressions that always control for education, experience and experience square. The first specification is the standard OLS model. But since we are worried about a possible correlation between the union status and the error term, we control for “ability” in the second specification by including an “IQ-type” variable (the Armed Force Qualifying Test, or AFQT) in the regression. The third specification uses AFQT as an instrumental variable for union status instead. Since we are not sure about the validity of AFQT as an instrument, the last specification adds a second instrument, so that we can perform an overidentification test. This additional instrumental variable is the local unemployment rate. The rationale for using this instrument is that union jobs are better protected, so that we are relatively more likely to observe a union instead of a non-union job when the unemployment rate is high.

Here are the results from the four specifications, with robust standard errors in parentheses

<table>
<thead>
<tr>
<th>Specification:</th>
<th>OLS(1)</th>
<th>OLS(2)</th>
<th>TSLS(3)</th>
<th>TSLS(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Effect of Unions</td>
<td>0.140</td>
<td>0.163</td>
<td>-2.575</td>
<td>-2.474</td>
</tr>
<tr>
<td>(0.017)</td>
<td>(0.016)</td>
<td>(0.500)</td>
<td>(0.447)</td>
<td></td>
</tr>
<tr>
<td>AFQT in regression</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Instruments:</td>
<td>---</td>
<td>---</td>
<td>AFQT</td>
<td>AFQT + un. rate</td>
</tr>
<tr>
<td>Overid. Test:</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>0.276</td>
</tr>
<tr>
<td>[p-value]</td>
<td></td>
<td></td>
<td>[0.599]</td>
<td></td>
</tr>
</tbody>
</table>

a) The AFQT variable always has a strong and positive effect on log wages. Based on the results above, do you think AFQT and union status are positively or negatively correlated? Explain.

b) The critical value of a Hausman test of OLS (column 1) against TSLS (column 3 or 4) is 3.84. Based on the information above, would you reject the null hypothesis that the OLS estimate of the effect of union on wages is consistent?

c) Do you find the estimated effect of unions in column 4 economically and/or intuitively plausible? If not, you are then confronted with a dilemma since the Hausman test and the overidentification test appear to suggest that the model in column 4 should be preferred to the simple OLS model in column 1. Can you think of a scenario under which you should ignore the results of these two tests and conclude that the TSLS are not valid? Discuss.