**ECO375H1F: Applied Econometrics I**
**Summer 2019**

Prof. Martin Burda
Department of Economics, Unviersity of Toronto

**Graded Problem Set #2**

**Submission details:** All work should be performed in one R Notebook file. Your answers should appear as comments, R commands and the corresponding output. Submit the "Preview" version of the file in nb.html or pdf format through Quercus by the due date.

**Part 1**
Use the data in *hprice3.txt* and consider the following model of U.S. house price determinants:

$$\begin{aligned} lprice &= \beta_0 + \beta_1 year + \beta_2 age + \beta_3 agesq + \beta_4 nbh + \beta_5 cbd + \beta_6 inst \\ &\quad + \beta_7 rooms + \beta_8 area + \beta_9 land + \beta_{10} baths + \beta_{11} dist + u \end{aligned}$$

where *lprice* is log of a house price, *year* denotes old or new type, *age* is the house age, *agesq* is age$^2$, *nbh* is neighborhood rank, *cbd* is the distance to the central business district, *inst* is the distance to an interstate highway, *rooms* is the number of rooms in the house, *area* is the square footage of the house, *land* is the square footage of the lot, *baths* is the number of baths, and *dist* is the distance of the house to the nearest industrial site.

1. Use OLS to estimate the model and output your estimates and the usual standard errors.

2. Report the outcome of the Breusch-Pagan test for heteroskedasticity and comment on heteroskedasticity being present or absent.

3. Obtain the heteroskedasticity-robust standard errors. Report any important differences with the usual standard errors.

4. Estimate the model using Generalized Least Squares (GLS), using the usual exponential function for $h(x)$, and report the output.

**Part 2**
Use the data in *hsng2.txt* and consider the following model of rent determinants:

$$rent = \beta_0 + \beta_1 hsngval + \beta_2 pcturban + u$$

where *rent* is the median rent is a U.S. city, *hsngval* is median house value in the city, and *pcturban* is the percentage of urban population in the state where the city is located.

1. Use OLS to estimate this equation and output your estimates.

2. We suspect that *hsngval* is endogenos and hence the OLS estimates are biased. We maintain that *pcturban* is exogenous.

3. Estimate the model with 2SLS, instrumenting *hsngval* with $faminc$ (median family income in the city) and *popden* (population density in the city), assuming these instruments are exogenous in the rent equation.

4. Test for weak instruments, regressor exogeneity, and overidentifying restrictions. State your conclusions.