**MA Econometrics Examination**
**August, 2014**

**Total Time: 8 hours**
**MA students are required to answer from A and B.**
**PhD students are required to answer from A, B, and C.**

## PART A
**(Answer any TWO from Part A)**

1. ***Fill in all of the ?????? in Table 2 (see next page). There are 8 numbers to calculate.***

**REFERRING TO TABLE 2**

a. Define and interpret R-squared
b. Define the relationship between F and R-squared
c. Define and interpret the 95% CI for "Family Size"
d. Interpret the coefficient on "Married"

**USING BOTH TABLES 1 & 2**

e. Why does adding "Married" create a relatively large change on the coefficients on Family Size and Family Size^2 but cause relatively small change the coefficients on Age and Age^2?
f. Calculate the marginal effect of being one year older than the mean age (40.66272) of the sample.
g. If you were going to add a variable to this regression, what variable would you add? Explain the variable choice (why?), what the expected sign would be, and how you interpret the coefficient on that variable.

*Did you calculate the missing numbers (??????) in Table 2? You should have 8 numbers in addition to answering the above questions.*

**Table 1**

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| | | | | Number of obs | |
| | | | | F( 4, 12397) | |
| Model | 1.31E+12 | 4 | 3.28E+11 | Prob > F | 0.000 |
| Residual | 1.48E+13 | 12397 | 1.19E+09 | R-squared | |
| Total | 1.61E+13 | 12401 | 1.30E+09 | Root MSE | 34523 |

| Wage Income | Coef. | Std.Err. | t | P>t | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Age | 3358.85 | 140.11 | 23.97 | 0.00 | 3084.22 | 3633.49 |
| Age^2 | -35.59 | 1.74 | -20.41 | 0.00 | -39.01 | -32.17 |
| Family Size | 3208.56 | 504.99 | | | | |
| Family Size^2 | -442.89 | 58.92 | -7.52 | 0.00 | -558.38 | -327.39 |
| Constant | -50889.86 | 2691.38 | -18.91 | 0.00 | -56165.38 | -45614.34 |

**Table 2**

| Source | SS | df | MS | | |
|---|---|---|---|---|---|
| | | | | Number of obs | ?????? |
| | | | | F( 5, 12396) | ?????? |
| Model | 1.59E+12 | 5 | 3.17E+11 | Prob > F | 0.000 |
| Residual | 1.45E+13 | 12396 | 1.17E+09 | R-squared | ?????? |
| Total | 1.61E+13 | 12401 | 1.30E+09 | Root MSE | 34203 |

| Wage Income | Coef. | Std.Err. | t | P>t | [95% Conf. | Interval] |
|---|---|---|---|---|---|---|
| Age | 2825.81 | 143.12 | 19.74 | 0.000 | 2545.27 | 3106.34 |
| Age^2 | -31.09 | 1.75 | -17.74 | 0.000 | -34.53 | -27.65 |
| Family Size | -263.60 | 549.42 | ?????? | ?????? | ?????? | ?????? |
| Family Size^2 | -138.37 | 61.68 | -2.24 | 0.025 | -259.27 | -17.47 |
| Married | ?????? | 751.98 | 15.29 | 0.000 | 10025.89 | 12973.88 |
| Constant | -36431.60 | 2829.10 | -12.88 | 0.000 | -41977.07 | -30886.13 |

2. ***Probability Theory, Distributions, and More***
   a. State Bayes' Theorem
   b. In words, describe what Bayes' Theorem means
   c. What distribution is below:

$$f(x;\lambda) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0, \\ 0 & x < 0. \end{cases}$$

   d. Find the first and second moments of this distribution directly from the distribution itself.
   e. Find the first and second moments using the moment generating function. Also, discuss if these moments are the same or different from the moments you found in **part d** and why or why not.
   f. This distribution has a property called "memoryless." Prove that this distribution is memoryless.
   g. Name a practical application of this distribution, i.e., where does it naturally occur?
   h. Name the third and fourth central moments – *name them do not calculate them.*
   i. Define and draw the PDF and CDF of the distribution.

3. ***Estimator:*** Estimating the Variance of the LS Estimator
   *a.* We know $var(b) = \sigma^2(X'X)^{-1}$, but $\sigma^2$ is an unknown parameter. Therefore in order to find $var(b)$, we need to find a good estimator for $\sigma^2$. *Derive an estimator of* $\sigma^2$.
   b. Is it unbiased? Prove this is an unbiased estimator?
   *c.* If you take the square root of this estimator what is that called?
   *d.* How and why do we use the estimator you just derived?

## PART B
**(Answer any TWO from Part B)**

### Question 4

Given the following linear model in matrix notation

$$Y = X\beta + U$$

a. Derive the OLS estimator of the beta-coefficient vector $\beta$ under the assumption of the white-noise errors, and derive the variance-covariance matrix of $\beta$.
b. Assuming a simple heteroscedastic form: $\sigma_t^2 = \gamma\, Z_t^2$ (and no autocorrelation), present the variance-covariance matrix of U, and the decomposition matrix P.
c. How would you perform a BP test to see if the variance structure is of the type presented in b? Show each step.
d. Why is the BP test better than the Goldfeld-Quandt test?
e. Simplifying the model as follows: $y_t = \beta_0 + \beta_1 x_t + u_t$, and using the heteroscedastic structure as presented in Q 4 b, derive the weighted least squares model.
f. What are the consequences of applying the OLS method to a heteroscedastic model. (No proof needed.)

### Question 5

Briefly explain the following

a. Compare and contrast the three estimation methods – OLS, GLS, and MLE?
b. Explain the meaning of the following: "efficiency implies consistency," but "consistency does not imply efficiency."
c. Heteroscedasticity is exclusively a cross-sectional problem. True or false? Explain. Autocorrelation is exclusively a time series problem. True or false? Explain.
d. The following distributions --t, F, and Chi-squared—are considered the members of the normal family. Explain.
e. The OLS estimator $(\widehat{\beta})$ under the assumption of autocorrelation or heteroscedasticity is still unbiased, but the variances are biased. True or false. Explain.
f. Why do we use marginal effect instead of the regular coefficient to explain the result of logit/probit models? Explain.
g. The OLS is not appropriate for the binary logit model. It has to be estimated using the MLE method. Explain.

## Question 6

Imagine we know with certainty that the following model fully describes the true state of the supply and demand for wheat.

First, the demand for wheat in any year, ***Q(t)***, is a function of the price of wheat, ***pw(t)***, the income of the average individual, ***I(t)***, and the price of corn, ***pc(t)***.

Second, in any year the price of wheat is a function of the amount of wheat brought to market, ***Q(t)***, and a weather index, ***W(t)***, that is positively related to the amount of wheat that is harvested. Third, the error terms in the supply and demand functions are due purely to measurement errors—that is, there are no omitted variables in the model. Thus, we have the following two equation model:

***Demand:*** $Q(t) = \alpha 0 + \alpha 1\, pw(t) + \alpha 2\, I(t) + \alpha 3\, pc(t) + \varepsilon(t)$ ***(1) and***

***Supply:*** $pw(t) = \beta 0 + \beta 1\, Q(t) + \beta 2 W(t) + \eta(t)$ ***(2)***

We assume that the error terms each are normally distributed with a mean of zero and a constant variance. Moreover, we assume that the two error terms are independent of each other—that is, we are assuming that:

$E(t) \sim N(0,\sigma 2\_\varepsilon)$, $\eta(t) \sim N(0,\sigma 2\_\eta)$, ***and*** $E(\varepsilon(t).\eta(t)) = 0$.

Finally, we assume that income, the price of corn, and the weather index are non-stochastic variables—i.e., these variables are independent of the two error terms.

a. Present the endogenous variables. List the predetermined variables.
b. Identify the demand and the supply equations using the order condition.
c. Identify the supply equation using the rank condition.
d. Present the 2 SLS method of estimation.
e. Using the supply equation as an example, discuss the generated regressor bias problem and show how you would correct for it.
f. Using the Hausman method, how would you test to see if the ***Q(t)*** variable in the supply equation is endogenous?