

### Problem Set #1

#### I. Okun's Law

In *The Political Economy of Prosperity* Arthur Okun argues that the unemployment rate is a good proxy variable for all cyclical variables and that the relationship between unemployment changes and actual output change is stable. Okun's Law states that for every one percentage point that unemployment is above  $u^*$ , the natural rate of unemployment, there is a  $\beta\%$  gap between potential and actual GDP.

$$P_t = X_t [1 + \beta(u_t - u^*)]$$

where  $P$  is potential GDP and  $X$  is actual GDP. We will assume that potential GDP grows smoothly over time,

$$P_t = P_o e^{g_1 t + g_2 t^2}$$

where  $t$  represents time. By taking logs where  $x = \log X$  these equations become

$$p_t = x_t + \beta(u_t - u^*) + \text{second order terms}$$

and

$$p_t = p_o + g_1 t + g_2 t^2$$

Substituting and setting  $u^* = 0.05$ , we have

$$x_t = p_o + g_1 t + g_2 t^2 - \beta(u_t - 0.05)$$

or

$$x_t = (p_o + 0.05\beta) + g_1 t + g_2 t^2 + \beta u_t$$

giving us a regression equation

$$x_t = \alpha_t + g_1 t + g_2 t^2 + \beta u_t + \varepsilon_t$$

Quarterly data on real GDP and the unemployment rate from 1948:1 to 2009:2 are available in the file "**okun.dta**" in our course folder<sup>1</sup>.

You will need to generate a new variable for the log of real GDP. Type the following command in the command line:

```
gen lgpd = log(gdp)
```

You have a time trend variable called `tr`. You should create a squared time trend using the following command:

```
gen tr2 = tr*tr
```

You can run the above regression by using the command

```
regress lgpd tr tr2 unemp
```

---

<sup>1</sup> This file contains quarterly data from 1948:1 – 2009:2 on five variables: YEAR (year of the observation), QTR (quarter of the observation), GDP (real GDP), UNEMP (unemployment rate), and TR (time trend).

What is your estimate of Okun's coefficient? Test the null hypothesis that  $\beta = -2$  at the 5% level. Some economists have observed that the Okun's law coefficient has changed over time. Estimate the Okun's law coefficient over the following two periods: 1948:1 – 1977:4 and 1978:1 – 2009:2.

**regress lgdp tr tr2 unemp if tin(1948q1, 1977q4)**

What is your estimate of Okun's coefficient in each period? Test the null hypothesis that  $\beta = -2$  in each period.

## II. Housing Prices

You have been hired to manage a housing construction project in Claremont, CA. In order to establish home selling prices, you need to know something about the relationship between selling price and various relevant variables. Data on selling price (in thousands of dollars.), living area (in square feet), lot size (in acres), number of bedrooms, number of baths, and the year the house was built for 83 random homes sold in Claremont, CA is provided in the file “**houses.dta**” in the course folder<sup>2</sup>. The point of this exercise is to understand how different variables affect the selling price of a home.

1. Estimate the univariate regression equation:

$$PRICE_i = \alpha + \beta_1 LIV_i + \varepsilon_i$$

where price is the dependent variable and living area is the independent variable.

2. Report and interpret your estimates of  $\alpha$  and  $\beta$ .

3. What is the  $R^2$  and the standard error of the regression (root MSE)? In what units is it measured?

4. Estimate the following equations by adding one variable at a time. First add lot size, then number of rooms, and then number of baths, and then year to the above equation.

$$PRICE_i = \alpha + \beta_1 LIV_i + \beta_2 LOT_i + \varepsilon_i$$

$$PRICE_i = \alpha + \beta_1 LIV_i + \beta_2 LOT_i + \beta_3 ROOMS_i + \varepsilon_i$$

$$PRICE_i = \alpha + \beta_1 LIV_i + \beta_2 LOT_i + \beta_3 ROOMS_i + \beta_4 BATHS_i + \varepsilon_i$$

$$PRICE_i = \alpha + \beta_1 LIV_i + \beta_2 LOT_i + \beta_3 ROOMS_i + \beta_4 BATHS_i + \beta_5 YEAR_i + \varepsilon_i$$

What is the effect on  $R^2$ , adjusted- $R^2$  and the standard error of the regression of adding each variable one at a time? Does this imply that the predictive reliability of the regression model improved or worsened with each additional variable?

5. Should you use the  $R^2$  or the adjusted  $R^2$  as a measure of "goodness of fit"? Explain.

6. Consider the very last regression estimates for

$$PRICE_i = \alpha + \beta_1 LIV_i + \beta_2 LOT_i + \beta_3 ROOMS_i + \beta_4 BATHS_i + \beta_5 YEAR_i + \varepsilon_i$$

What would be the predicted selling price of a home given:

living area = 1900 square feet

lot size = .25 acres

number of bedrooms = 3

number of baths = 2.5

<sup>2</sup> This file contains 83 observations on the following seven variables: OBS (observation number), PRICE (price of the home), LIV (square footage of the house), LOT (lot size), ROOM (number of bedrooms), BATHS (number of bathrooms), and YEAR (year built).

year built = 1968

7. Test the 5 hypotheses that each individual explanatory variable is statistically significantly different from zero. *i.e.*  $H_0: \beta_i = 0$  where  $i = 1, 2, 3, 4$  and 5. Explain your results.

8. What would you recommend to be the "best" model of home selling price? Define what you mean by "best". Which variables would you keep, and which might you want to discard? Explain your rationale.

### III. Consumption

According to Robert Hall<sup>3</sup>, no variable apart from current consumption should be of any value in predicting future consumption. In support of this hypothesis he estimated the following equations:

$$\begin{aligned} C_t &= \alpha + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \beta_3 C_{t-3} + \beta_4 C_{t-4} \\ C_t &= \alpha + \beta_1 C_{t-1} + \gamma_1 Y_{t-1} \\ C_t &= \alpha + \beta_1 C_{t-1} + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \gamma_3 Y_{t-3} + \gamma_4 Y_{t-4} \end{aligned}$$

where  $C_t$  is real per capita consumption of nondurables and services and  $Y_t$  is real per capita personal disposable income. He found the coefficients on lagged terms apart from lagged consumption to be insignificant. The data for this exercise is available in the file "hall.dta" in the course folder<sup>4</sup>. You will have to create the appropriate variables from the variables provided in the dataset.

Estimate the following Hall equations over the updated sample period 1990:1 to 2009:2 and test the following hypotheses. To specify this regression in Stata, type something like

```
regress con L.con L.inc
regress con L.con L2.con L3.con L4.con
regress con L.con L.inc L2.inc L3.inc L4.inc
```

as your equation. **L** is called a lag operator. Make sure that the sample periods for the restricted and unrestricted regressions are exactly the same.

$$\begin{aligned} C_t &= \alpha + \beta_1 C_{t-1} + \gamma_1 Y_{t-1} \\ * \quad H_0: \gamma_1 &= 0 \end{aligned}$$

$$\begin{aligned} C_t &= \alpha + \beta_1 C_{t-1} + \beta_2 C_{t-2} + \beta_3 C_{t-3} + \beta_4 C_{t-4} \\ * \quad H_0: \beta_2 = \beta_3 = \beta_4 &= 0 \end{aligned}$$

$$\begin{aligned} C_t &= \alpha + \beta_1 C_{t-1} + \gamma_1 Y_{t-1} + \gamma_2 Y_{t-2} + \gamma_3 Y_{t-3} + \gamma_4 Y_{t-4} \\ * \quad H_0: \gamma_1 = \gamma_2 = \gamma_3 = \gamma_4 &= 0 \end{aligned}$$

Do the Hall results hold up over the updated sample period? Explain.

<sup>3</sup> Robert Hall, "Stochastic Implications of the Life Cycle - Permanent Income Hypothesis: Theory and Evidence," *Journal of Political Economy*, December 1978, pp. 971-987.

<sup>4</sup> This file contains quarterly data from 1990:1 – 2009:2 on six variables: YEAR (year of the observation), QTR (quarter of the observation), POP (population), CND (real consumption of nondurables), CSR (real consumption of services), and DPI (real disposable personal income).