Economics 310
Fall 2004
University of Wisconsin-Madison

Menzie D. Chinn
Social Sciences 7418

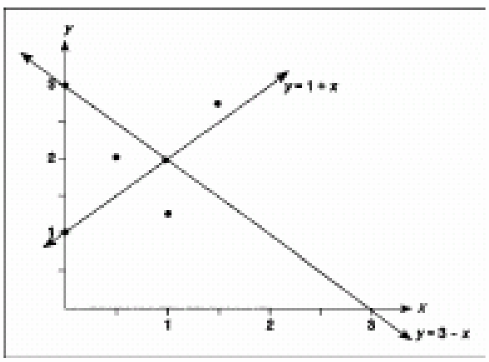
# Problem Set 5 Answers

This problem set is due in lecture on **Wednesday, December 15th**. No late problem sets will be accepted. **Be sure to show your work** (that is, do not use a spreadsheet or statistical program to generate your answers), and to write your name, ID number, as well as the name of your Teaching Assistant, on your problem set.

Answer all these problems. They are from the textbook, with the exception of Problem W which is written out.

- 12.12
- 12.30
- 12.52
- 13.28

12.12 a.




b. Choose $y = 1 + x$ since it best describes the relation of $x$ and $y$.

c.

| $y$ | $X$ | $\hat{y} = 1 + x$ | $y - \hat{y}$ |
|---|---|---|---|
| 2 | .5 | 1.5 | 2 – 1.5 = .5 |
| 1 | 1.0 | 2.0 | 1 – 2.0 = –1.0 |
| 3 | 1.5 | 2.5 | 3 – 2.5 = .5 |
| | | | Sum of errors = 0 |

| $y$ | $X$ | $\hat{y} = 3 - x$ | $y - \hat{y}$ |
|---|---|---|---|
| 2 | .5 | 3 – .5 = 2.5 | 2 – 2.5 = –.5 |
| 1 | 1.0 | 3 – 1.0 = 2.0 | 1 – 2.0 = –1.0 |
| 3 | 1.5 | 3 – 1.5 = 1.5 | 3 – 1.5 = 1.5 |
| | | | Sum of errors = 0 |

d. $\text{SSE} = \sum (y - \hat{y})^2$

SSE for 1st model: $y = 1 + x$, $\text{SSE} = (.5)^2 + (-1)^2 + (.5)^2 = 1.5$

SSE for 2nd model: $y = 3 - x$, $\text{SSE} = (-.5)^2 + (-1)^2 + (1.5)^2 = 3.5$

The best fitting straight line is the one that has the smallest least squares. The model $y = 1 + x$ has a smaller SSE, and therefore it verifies the visual check in part a.

e. Some preliminary calculations are:

$\sum x = 3 \quad \sum y = 6 \quad \sum xy = 6.5 \quad \sum x^2 = 3.5$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 6.5 - \frac{(3)(6)}{3} = .5$$

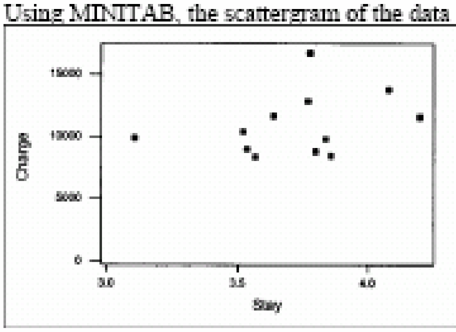
$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 3.5 - \frac{(3)^2}{3} = .5$$

$$\hat{\beta}_1 = \frac{.5}{.5} = 1;\ \bar{x} = \frac{\sum x}{3} = \frac{3}{3} = 1;\ \bar{y} = \frac{\sum y}{3} = \frac{6}{3} = 2$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 2 - 1(1) = 1 \Rightarrow \hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x = 1 + x$$

The least squares line is the same as the second line given.

12.30 a. Using MINITAB, the scattergram of the data is:




b. $\sum x = 44.71 \qquad \sum y = 131{,}670 \qquad \sum xy = 493{,}117.7 \qquad \sum x^2 = 167.4615$
$\sum y^2 = 1{,}514{,}402{,}100$

$$\bar{x} = \frac{\sum x}{n} = \frac{44.71}{12} = 3.7258333 \qquad \bar{y} = \frac{\sum y}{n} = \frac{131{,}670}{12} = 10{,}972.5$$

$$SS_{xy} = \sum xy - \frac{(\sum x)(\sum y)}{n} = 493{,}117.7 - \frac{44.71(131{,}670)}{12}$$
$$= 493{,}117.7 - 490{,}580.475 = 2{,}537.225$$

$$SS_{xx} = \sum x^2 - \frac{(\sum x)^2}{n} = 167.4615 - \frac{44.71^2}{12}$$
$$= 167.4615 - 166.5820083 = .8794917$$

$$\hat{\beta}_1 = \frac{SS_{xy}}{SS_{xx}} = \frac{2{,}537.225}{.8794917} = 2884.876571 \approx 2884.877$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x} = 10{,}972.5 - 2884.876571(3.7258333) = 10{,}972.5 - 10{,}748.56929$$

$$= 233.93071 \approx 233.931$$

The fitted regression line is $= 233.931 + 2884.877x$

c. $SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 1{,}514{,}402{,}1000 - \frac{131{,}670^2}{12}$
$= 1{,}514{,}402{,}100 - 1{,}444{,}749{,}075 = 69{,}653{,}025$

$SSE = SS_{yy} - \hat{\beta}_1 SS_{xy} = 69{,}653{,}025 - 2{,}884.876571(2{,}537.225)$
$= 69{,}653{,}025 - 7{,}319{,}580.958 = 62{,}333{,}444.04$

$$s^2 = \frac{SSE}{n-2} = \frac{62{,}333{,}444.04}{12-2} = 6{,}233{,}344.404$$

$$s = \sqrt{s^2} = \sqrt{6{,}233{,}344.404} = 2{,}496.6667$$

We would expect to see most of the hospital charges to fall within $2s$ or 2($2,496.6667) = $4,993.3333 of the least squares line.

d. For $x = 4$, $\hat{y} = 223.931 + 2{,}884.877(4) = 11{,}763.439$

$\hat{y} \pm 2s \Rightarrow 11{,}763.439 \pm 4{,}993.3333 \Rightarrow (6{,}770.106,\ 16{,}756.772)$

e. Only one state (California) had an average hospital charge more than 2 standard errors from the least squares line. Thus, 11 out of 12 or 11/12 or .917 of the states had average hospital charges within 2 standard errors of the least squares line.

12.52 a. Some preliminary calculations are:

$\sum x = 0 \qquad \sum x^2 = 10 \qquad \sum xy = 20$
$\sum y = 12 \qquad \sum y^2 = 70$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 20 - \frac{0(12)}{5} = 20$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 10 - \frac{0^2}{5} = 10$$

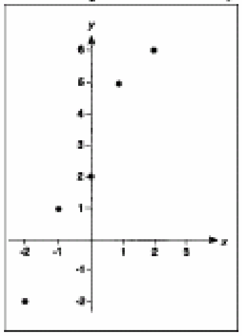
$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 70 - \frac{12^2}{5} = 41.2$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{20}{\sqrt{10(41.2)}} = .9853$$

$$r^2 = .9853^2 = .9709$$

Since $r = .9853$, there is a very strong positive linear relationship between $x$ and $y$.

Since $r^2 = .9709$, 97.09% of the total sample variability around  is explained by the l relationship between $x$ and $y$.



b. Some preliminary calculations are:

$$\sum x = 0 \qquad \sum x^2 = 10 \qquad \sum xy = -15$$
$$\sum y = 16 \qquad \sum y^2 = 74$$

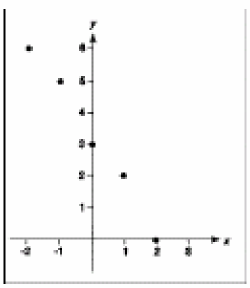
$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = -15 - \frac{0(16)}{5} = -15$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 10 - \frac{0^2}{5} = 10$$

$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 74 - \frac{16^2}{5} = 22.8$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{-15}{\sqrt{10(22.8)}} = -.9934$$

$$r^2 = (-.9934)^2 = .9868$$



Since $r = -.9934$, there is a very strong negative linear relationship between $x$ and $y$.

Since $r^2 = .9868$, 98.68% of the total sample variability around  is explained by the linear relationship between $x$ and $y$.

c. Some preliminary calculations are:

$$\sum x = 18 \qquad \sum x^2 = 52 \qquad \sum xy = 36$$
$$\sum y = 14 \qquad \sum y^2 = 32$$

$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 36 - \frac{18(14)}{7} = 0$$

$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 52 - \frac{18^2}{7} = 5.71428571$$

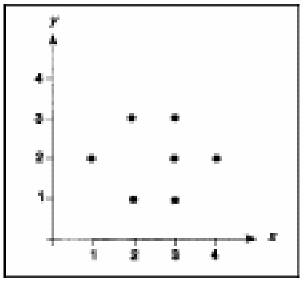
$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 32 - \frac{14^2}{7} = 4$$

$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{0}{\sqrt{5.71428571(4)}} = 0$$

$$r^2 = 0^2 = 0$$

Since $r = 0$, this implies that $x$ and $y$ are not related.

Since $r^2 = 0$, 0% of the total sample variability around  is explained by the linear relationship between $x$ and $y$.

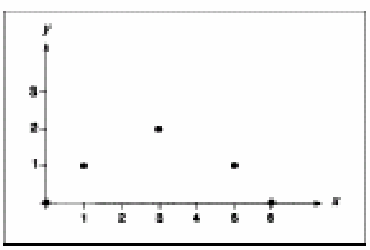


d. Some preliminary calculations are:

$$\sum x = 15 \qquad \sum x^2 = 71 \qquad \sum xy = 12$$
$$\sum y = 4 \qquad \sum y^2 = 6$$
$$SS_{xy} = \sum xy - \frac{\sum x \sum y}{n} = 12 - \frac{15(4)}{5} = 0$$
$$SS_{xx} = \sum x^2 - \frac{\left(\sum x\right)^2}{n} = 71 - \frac{15^2}{5} = 26$$
$$SS_{yy} = \sum y^2 - \frac{\left(\sum y\right)^2}{n} = 6 - \frac{4^2}{5} = 2.8$$
$$r = \frac{SS_{xy}}{\sqrt{SS_{xx}SS_{yy}}} = \frac{0}{\sqrt{26(2.8)}} = 0$$
$$r^2 = 0^2 = 0$$

Since $r = 0$, this implies that $x$ and $y$ are not related.

Since $r^2 = 0$, 0% of the total sample variability around  is explained by the linear relationship between $x$ and $y$.



13.28 a. The least squares prediction equation is:

$$\hat{y} = -4.30 - .002x_1 + .336x_2 + .384x_3 + .067x_4 - .143x_5 + .081x_6 + .134x_7$$

b. To determine if the model is adequate, we test:

$H_0$: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = \beta_5 = \beta_6 = \beta_7 = 0$
$H_a$: At least one $\beta_i \neq 0$, $i = 1, 2, 3, \ldots, 7$

The test statistic is $F = 111.1$ (from table).

Since no $\alpha$ was given, we will use $\alpha = .05$. The rejection region requires $\alpha = .05$ in the upper tail of the $F$-distribution with $\nu_1 = k = 7$ and $\nu_2 = n - (k + 1) = 268 - (7 + 1) = 260$. From Table IX, Appendix B, $F_{.05} \approx 2.01$. The rejection region is $F > 2.01$.

Since the observed value of the test statistic falls in the rejection region ($F = 111.1 > 2.01$), $H_0$ is rejected. There is sufficient evidence to indicate that the model is adequate for predicting the logarithm of the audit fees at $\alpha = .05$.

c. $\hat{\beta}_3 = .384$. For each additional subsidiary of the auditee, the mean of the logarithm of audit fee is estimated to increase by .384 units.

d. To determine if the $\beta_4 > 0$, we test:

$H_0$: $\beta_4 = 0$
$H_a$: $\beta_4 > 0$

The test statistic is $t = 1.76$ (from table).

The $p$-value for the test is .079. Since the $p$-value is not less than $\alpha$ ($p = .079 \not< \alpha = .05$), $H_0$ is not rejected. There is insufficient evidence to indicate that $\beta_4 > 0$, holding all the other variables constant, at $\alpha = .05$.

e. To determine if the $\beta_1 < 0$, we test:

$H_0$: $\beta_1 = 0$
$H_a$: $\beta_1 < 0$

The test statistic is $t = -0.049$ (from table).

The $p$-value for the test is .961. Since the $p$-value is not less than $\alpha$ ($p = .961 \not< \alpha = .05$), $H_0$ is not rejected. There is insufficient evidence to indicate that $\beta_1 < 0$, holding all the other variables constant, at $\alpha = .05$. There is insufficient evidence to indicate that the new auditors charge less than incumbent auditors.

Problem W. Below are plotted data for the trade balance as a share of GDP for the US and US GDP. The idea is that when the US economy booms, imports rise, and the trade balance deteriorates.

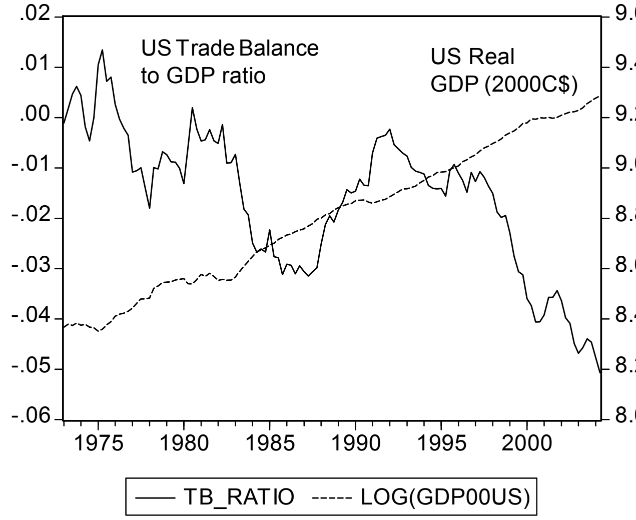



**Figure 1:** Time series for Trade Balance to GDP Ratio and US Real GDP. Sources: BEA.

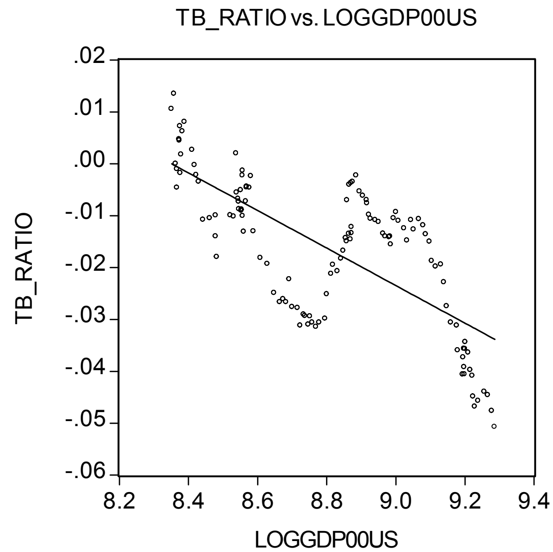



**Figure 2:** Scatter plot of Trade Balance Ratio and US Real GDP.

Dependent Variable: TB_RATIO
Method: Least Squares
Date: 12/07/04 Time: 20:08
Sample: 1973:1 2004:2
Included observations: 126

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 0.302216 | 0.027757 | 10.88792 | 0.0000 |
| LOGGDP00US | -0.036190 | 0.003152 | -11.48326 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.515370 | Mean dependent var | -0.016365 |
| Adjusted R-squared | 0.511462 | S.D. dependent var | 0.014116 |
| S.E. of regression | 0.009867 | Akaike info criterion | -6.383591 |
| Sum squared resid | 0.012071 | Schwarz criterion | -6.338571 |
| Log likelihood | 404.1663 | F-statistic | 131.8653 |
| Durbin-Watson stat | 0.099030 | Prob(F-statistic) | 0.000000 |

a. In words, interpret the coefficient on *LOGGDP00US* (where this is the log of *GDP00US*).

Each one percent change in real US GDP causes a 0.036 percentage point decrease in the trade balance to GDP ratio. Technically, it is $\Delta tb / \Delta y$ where *tb* is the trade balance to GDP ratio and *y* is log real GDP.

b. Conduct a one-sided t-test using a 1% significance level, for the following hypothesis test:

$H_0$: $\beta_1 = 0$

$H_A$: $\beta_1 < 0$

$$t = \frac{-0.0362 - 0}{0.0032} \approx -11.31$$

The critical value at 1% level for a one tailed test, using 120 degrees of freedom (actually there are 124 d.f.), is -2.358. Since -11.31 < -2.358, reject the null hypothesis.

c. Calculate the standard error of the regression, using the statistics reported in the output (show your work!).

$s = \sqrt{\frac{SSR}{n-2}} = \sqrt{\frac{0.012071}{124}} \approx 0.00987$ which matches the entry under "S.E. of regression".

d. Calculate the value of the "S.E. of regression" using the "Sum of squared resid" (also termed the Sum of Squared Errors in the textbook).

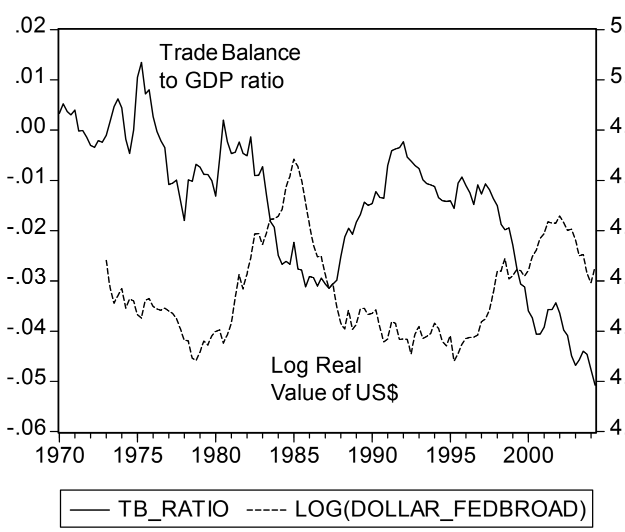
$s = \sqrt{\frac{SSR}{n-2}} = \sqrt{\frac{0.012071}{124}} \approx 0.00987$ which matches the entry under "S.E. of regression".

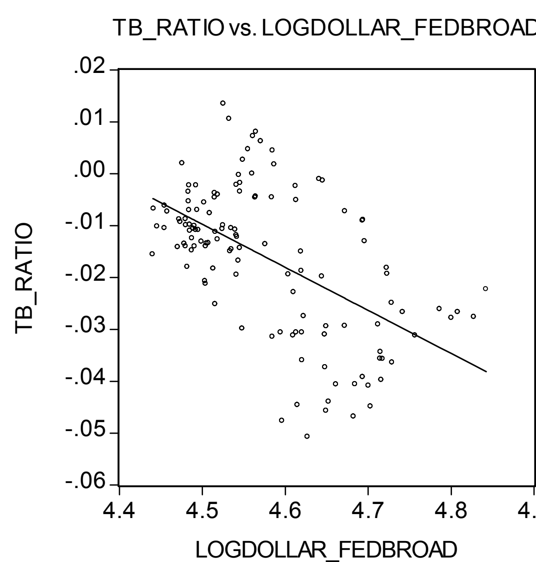
e. Calculate the R-squared using the values for SSE and the "S.D. dependent var" (which is the standard deviation of the dependent variable).

$R^2 = 1 - (SSE/SS_{yy})$. While $SS_{yy}$ is not reported in the regression output, notice that the standard deviation of the dependent variable y *is* reported. In fact $SS_{yy} = (SD^2)x(n-1) = (0.014116)^2x(125) = 0.024907682$

Hence, $R^2 = 1 - (SSE/SS_{yy}) = 1 - (0.012071/0.024907682) = 0.515370$ (which matches the regression output entry for this measure).

The above story omits some other relevant factors. When the rest-of-the-world booms, then their imports (our exports) rise, so the rest-of-world GDP (*GDP96_ROW*) should enter. But what should also matter is the real value of the US dollar (*DOLLAR_FEDBROAD*), which determines how expensive American goods are relative to foreign. Figures 3 and 4 depict the trade balance and the dollar value (in logs).



Then the following multiple regression might be estimated:

Dependent Variable: TB_RATIO
Method: Least Squares
Date: 12/07/04 Time: 20:16
Sample: 1973:1 2004:2
Included observations: 126

| Variable | Coefficient | Std. Error | t-Statistic | Prob. |
|---|---|---|---|---|
| C | 1.222852 | 0.121408 | 10.07225 | 0.0000 |
| LOGGDP00US | -0.170249 | 0.025061 | -6.793372 | 0.0000 |
| LOGGDP96_ROW | 0.127615 | 0.023124 | 5.518824 | 0.0000 |
| LOGDOLLAR_FEDBROAD | -0.064503 | 0.006561 | -9.830690 | 0.0000 |

| | | | |
|---|---|---|---|
| R-squared | 0.771246 | Mean dependent var | -0.016365 |
| Adjusted R-squared | 0.765621 | S.D. dependent var | 0.014116 |
| S.E. of regression | 0.006834 | Akaike info criterion | -7.102585 |
| Sum squared resid | 0.005698 | Schwarz criterion | -7.012544 |
| Log likelihood | 451.4628 | F-statistic | 137.1082 |
| Durbin-Watson stat | 0.265394 | Prob(F-statistic) | 0.000000 |

f. Interpret the coefficient on *LOGGDP00US* in this context.

Each one percent change in real US GDP causes a 0.170 percentage point decrease in the trade balance to GDP ratio, *holding everything else constant*. Technically, it is $\partial tb / \partial y$ where *tb* is the trade balance to GDP ratio and *y* is log real GDP.

g. Is the coefficient on *LOGGDP00US* still statistically significantly different from zero? Is the coefficient economically different?

$$t = \frac{-0.170 - 0}{0.025} \approx -6.80$$

The critical values at 1% level for a two tailed test, using 120 degrees of freedom (actually there are 122 d.f.), is -2.617 and 2.617. Since -6.80 < -2.617 so reject the null hypothesis. Hence, the coefficient is statistically different from zero. Whether the coefficient is economically different from zero depends upon one's perspective.

h. Calculate the standard error of the regression, using the statistics reported in the output (again, show your work!).

$s = \sqrt{\frac{SSR}{n-(k+1)}} = \sqrt{\frac{0.005698}{122}} \approx 0.00683$ which matches the entry under "S.E. of regression".

i. Form a 95% confidence interval around the coefficient on *LOGDOLLAR_FEDBROAD*.

Note a 95% confidence interval implies α = 0.05, α /2 = 0.025; also there are 122 degrees of freedom.

The confidence interval is given by $\hat{\beta}_3 \pm t_{\alpha/2} \times s_{\beta_3}$ = -0.064503 ± 1.98 x 0.006561
= (-0.077521, -0.051512`)

j. If the dollar were to depreciate by 20%, what would happen to the trade balance, and by how much?

$\Delta tb = \left(\frac{\partial b}{\partial r}\right) \times \Delta r$ where r is the log real value of the dollar.

$\Delta tb =$ $-0.064503 \times -0.20 = 0.0129$ or in other words, the trade balance to GDP ratio improves by 1.3 percentage points.

9.12.2004 e310ps5a_f04