Solutions to Midterm Exam 1, Econ 551
Professor John Rust
Yale University
February, 1999

Part I Question 1

a. We are to calculate $\widehat{\beta} = (X'X)^{-1}X'y$, where

$$X'X = \begin{bmatrix} 24 & 0 \\ 0 & 2 \end{bmatrix} \quad X'y = \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix}.$$

From the result in a), we know the diagonal matrix $X'X$ has the simple inverse matrix given by

$$\begin{bmatrix} 1/24 & 0 \\ 0 & 1/2 \end{bmatrix}.$$

Therefore, we have

$$\widehat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} 1/24 & 0 \\ 0 & 1/2 \end{bmatrix} \times \begin{bmatrix} 1 & 0 \\ 0 & 3 \end{bmatrix} = \begin{bmatrix} 1/24 \\ 3/2 \end{bmatrix}.$$

b. By regressing y on X_1 and X_2 separately, we get $\widehat{\beta}_1 = (X_1'X_1)^{-1}X_1'y = (24)^{-1} \cdot 1 = 1/24$ $\widehat{\beta}_2 = (X_2'X_2)^{-1}X_2'y = (2)^{-1} \cdot 3 = 3/2$ So we have the same results from this naive estimators as those from the results in b).

c. This naïve approach works because the matrix $X'X$ is diagonal, i.e., the two vectors X_1 and X_2 are orthogonal to each other. We have 0's for the off-diagonal elements in $X'X$, which enables us to ignore X_2 vector when we run our regression of y on X_1, and vice versa. The intuition is simple: if two vectors are orthogonal they have nothing in common. So each regressor does its own job in explaining the vector y without taking the place of the other regressor.

d. Suppose we have a nonsingular $N \times K$ matrix of regressors $X = [X_1 X_2]$, where X_1 and X_2 are $N \times K_1$ and $N \times K_2$ matrices, respectively. Suppose further that we have the following blockwise orthogonality condition given by $X_1'X_2 = 0$, which implies each column in the submatrix

X_1 is orthogonal to that in the submatrix X_2. Then our $K \times 1$ OLS estimates vector $\widehat{\beta}$ will be

$$\widehat{\beta} = (X'X)^{-1}X'y = \begin{bmatrix} X_1'X_1 & 0 \\ 0 & X_2X_2 \end{bmatrix}^{-1} \cdot \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix}.$$

But using our results in a), we get

$$\widehat{\beta} = \begin{bmatrix} (X_1'X_1)^{-1} & 0 \\ 0 & (X_2X_2)^{-1} \end{bmatrix} \cdot \begin{bmatrix} X_1'y \\ X_2'y \end{bmatrix} = \begin{bmatrix} (X_1'X_1)^{-1}X_1y \\ (X_1'X_1)^{-1}X_1y \end{bmatrix}.$$

Therefore, the blockwise diagonality condition allows us to treat the two sets of regressors separately. As a special case, when we have K regressors all orthogonal to each other as in the pard f) below, we can estimate each of the K regression coefficients by $\widehat{\beta}_i = (X_i'X_i)^{-1}X_i'y, i = 1, \ldots, K$

e. Our regression equation is given by

$$y_i = \beta_1 D_{i1} + \beta_2 D_{i2} + ... + \beta_K D_{iK} + e_i, \quad i = 1, 2, \ldots, N$$

or in a matrix form,

$$y = \begin{bmatrix} D_1 & D_2 & ... & D_K \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_1 \\ \cdots \\ \beta_1 \end{bmatrix} + e$$

The mutual exclusiveness of the dummy variables implies $D_j'D_k = \sum_{i=1}^{N} D_{ij}'D_{ik} = 0$, for all $1 \leq$ j $\neq$k $\leq$ K. Based on this orthogonality condition, we can apply what we have obtained up to e). Therefore, we get $\widehat{\beta}_j = (D_j'D_j)^{-1}D_j'y$, $j = 1, \ldots, K$ Or in the summation form, $\widehat{\beta}_j = \sum_{i=1}^{N} D_{ij}y_i \;/\; \sum_{i=1}^{N} D_{ij}^2 = \sum_{i=1}^{N} D_{ij}y_i \;/\; \sum_{i=1}^{N} D_{ij}$ The second equality above holds because all the possible values of D_{ij} are just one or zero. Now let's interpret the meaning of $\widehat{\beta}_j$. Suppose we have S times of the occurrence of one. Then, the denominator is just S, and the numerator is the sum of y$_i$s over the set of is such that $D_{ij} = 1$. Hence, our $\widehat{\beta}_j$ can be interpreted as a mean of y over the subpopulation of individuals with $D_{ij} = 1$.

Part I, Question 2

a-1. The value of the inner product of $\hat{y}$ and $\hat{\varepsilon}$ is zero.

a-2. The algebraic derivation of this is as follows. We know that:

$$\hat{\beta} = (X'X)^{-1}X'y$$

Hence,

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y = My$$

where

$$M = X(X'X)^{-1}X'$$

Furthermore,

$$\begin{aligned}\hat{\varepsilon} = y - \hat{y} = y - X\hat{\beta} &= y - X(X'X)^{-1}X'y \\ &= y - My = (I - M)y = Py\end{aligned}$$

Then,

$$\hat{\varepsilon} = Py \quad and \quad \hat{y} = My.$$

We also know that the matrices M and P are idempotent, [see exercise 3, part 1 from last year's midterm] thus,

$$P = P' and\ P = PP$$

and, analogously,

$$M = M' \quad and \quad M = MM$$

Given these properties, it's straightforward to compute the required inner product, in fact

$$\begin{aligned}\hat{y}'\hat{\varepsilon} &= (My)'(Py) = y'M'Py \\ &= y'M'(I - M)y = y'M'y - y'M'My \\ & \quad \textit{but, by idempotency,} \\ M' &= M \quad and \quad M'M = MM = M \\ & \quad \textit{hence} \\ \hat{y}'\hat{\varepsilon} &= y'My - y'My = 0.\end{aligned}$$

Some of you invoked the normal equations which is another perfectly legitimate way to proceed. In fact,

$$\begin{aligned}
\hat{y}'\hat{\varepsilon} &= (X\hat{\beta})'\hat{\varepsilon} = \hat{\beta}'X'\hat{\varepsilon} \\
&\quad \text{but by the least squares normal equations:} \\
X'\hat{\varepsilon} &= 0 \\
&\quad \text{so we have:} \\
\hat{y}'\hat{\varepsilon} &= \hat{\beta}'0 \\
&= 0.
\end{aligned}$$

Geometric argument: We are working with conventional assumptions. If N corresponds to the number of observations (i.e. the dimension of y and the number of rows of X), and K corresponds to the number of regressors (the number of columns of X), then $N > K$. This implies that the regressors span a linear subspace of R^n. Let's call this subspace L. When we do least-squares we seek to find the vector in $L \subset R^n$ which is closest (in the sense described in class) to the original vector of dependent observations ($\in R^n$). Given a matrix of regressors X, if the vector of dependent observations is y, the resulting best predictor of y is $\hat{y} = X\hat{\beta}$. It's apparent that $\hat{y}$belongs to L (it's a linear combination of the columns of X with weights given by the components of the vector $\hat{\beta}$). Furthermore, $\hat{y}$ is the orthogonal projection of y on the space spanned by X. $\hat{\varepsilon}$ is just the "distance" between y and $\hat{y}$. In order for this distance to be minimum, $\hat{\varepsilon}$ has to lay in a subspace of R^n which is orthogonal to the subspace L spanned by X. Hence, $\hat{y}$ and $\hat{\varepsilon}$ are just the two sides of a right angle triangle. Their orthogonality makes the inner product equal to zero.

a-3. The relationship between $\hat{y}$ and $\hat{\varepsilon}$ is just another way to assess the consequences of least-squares regressions. You are used to think in terms of orthogonality between X and $\hat{\varepsilon}$. You also know that the previous implies orthogonality between $\hat{\varepsilon}$ and all the vectors laying on the space spanned by the columns of X. Now, since $\hat{y}$ belongs to this space, the orthogonality property applies to it as well as to every other vector of the type Xa for a generic vector a conformable with X. Hence, we are just looking at the usual well-known problem from a different perspective. It follows that the implications of $\hat{y}'\hat{\varepsilon} = 0$ for regression analysis are the same as the implications of $X'\hat{\varepsilon} = 0$, namely the residuals don't contain information to improve on the estimates and so on.

b-1. We know that,

$$y = \hat{y} + \hat{\epsilon}$$

hence,

$$y'y = (\hat{y} + \hat{\epsilon})'(\hat{y} + \hat{\epsilon}) = \hat{y}'\hat{y} + \hat{\epsilon}'\hat{\epsilon} + 2\hat{y}'\hat{\epsilon}$$

We also know that,

$$\hat{y}'\hat{\epsilon} = 0$$

thus,

$$y'y = \hat{y}'\hat{y} + \hat{\epsilon}'\hat{\epsilon}$$

But this implies that,

$$1 = (\hat{y}'\hat{y} + \hat{\epsilon}'\hat{\epsilon})/y'y if\, y'y \neq 0$$

and so,

$$1 \geq \hat{y}'\hat{y}/y'y \geq 0$$

given that

$$\hat{y}'\hat{y}, \quad \hat{\epsilon}'\hat{\epsilon} \quad and \quad y'y \geq 0.$$

Alternatively, you could have used the Pythagorean theorem by just invoking the geometric interpretation you eventually gave in point a. $\hat{y}$ and $\hat{\epsilon}$ are othogonal, hence

$$||y||^2 = ||\hat{y}||^2 + ||\hat{\varepsilon}||^2$$

By definition of norm,

$$||x||^2 = x'x$$

Then,

$$y'y = \hat{y}'\hat{y} + \hat{\epsilon}'\hat{\epsilon}$$

and, as before

$$1 = (\hat{y}'\hat{y} + \hat{\epsilon}'\hat{\epsilon})/y'y \quad if \quad y'y \neq 0$$

and,

$$1 \geq \hat{y}'\hat{y}/y'y \geq 0$$

given that

$$\hat{y}'\hat{y}, \quad \hat{\epsilon}'\hat{\epsilon} \quad and \quad y'y \geq 0.$$

b-2. c provides a measure of goodness of fit (and it doesn't coincide with the R^2!). $y'y = \sum_{i=1}^{n} y_i^2$ could be interpreted as a measure of variability of the observations, whereas $\hat{y}'\hat{y} = \sum_{i=1}^{n} \hat{y}_i^2$ might represent the dispersion of the fitted values. Note that traditional measures of variability are centered around the mean (i.e., $var(x) = \sum_{i=1}^{n}(x_i - \overline{x})^2$) The ratio $\hat{y}'\hat{y}/y'y$ can then be seen as the amount of dispersion in the data explained by the regression. Furthermore,

$$\hat{\epsilon} = 0 \Rightarrow \hat{\epsilon}'\hat{\epsilon} = 0 \Rightarrow \hat{y}'\hat{y} = y'y \Rightarrow c = 1$$

and the regression guarantees a perfect fit. If, on the other hand, y is perpendicular to the space spanned by the columns of X, then $\hat{y} = 0$ and

$$\hat{y} = 0 \Rightarrow \hat{y}'\hat{y} = 0 \Rightarrow \hat{\epsilon}'\hat{\epsilon} = y'y \Rightarrow c = 0$$

The matrix of covariates doesn't help explain the dependent variable and the fit is as bad as it can be.

Part I, Question 3

We are given the regression model $y = Z_1\beta_1 + Z_2\beta_2 + \epsilon$ where

$$Z_1' = \begin{pmatrix} 2 & 0 & 0 \end{pmatrix}, \quad Z_1' = \begin{pmatrix} 0 & 2 & 0 \end{pmatrix}, \quad \text{and} \quad y' = \begin{pmatrix} 3 & 2 & 3 \end{pmatrix}.$$

a. The model may be written in the form of

$$y = X\beta + \epsilon$$

where

$$\underset{3\times 1}{y} = \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix}, \quad \underset{3\times 2}{X} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} \quad \text{and} \quad \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \end{bmatrix}.$$

b.

$$X'X = \begin{bmatrix} 4 & 0 \\ 0 & 4 \end{bmatrix}.$$

Since $X'X$ is diagonal we have,

$$(X'X)^{-1} = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \quad \text{and} \quad X'X\,(X'X)^{-1} = I.$$

c. The OLS estimate of β

$$\widehat{\beta} = (X'X)^{-1} X'y = \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 6 \\ 4 \end{bmatrix} = \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix}$$

Hence the predicted value of y

$$\widehat{y} = X\widehat{\beta} = \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{3}{2} \\ 1 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix}$$

d. See diagram below.

e. The vector of residuals is:

$$\widehat{\epsilon} = y - X\widehat{\beta} = \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix}$$

1. So we have:

$$i'\widehat{\epsilon} = \begin{bmatrix} 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} = 3 \neq 0$$

   This nonzero value to the inner product $\langle i'\widehat{\epsilon} \rangle$ suggest that the 2 vectors are not orthogonal to each other. Given that the model does not contain a constant term, it suggest that i is not the subspace spanned by the matrix of regressors. $\langle i'\widehat{\epsilon} \rangle$ also happens to be the sum of the residuals of the regression, its non zeros value suggest that the sum of residuals is zero only in regressions containing a constant term.

2. However we have

$$Z_1'\widehat{\epsilon} = \begin{bmatrix} 2 & 0 & 0 \end{bmatrix} \begin{bmatrix} 0 \\ 0 \\ 3 \end{bmatrix} = 0$$

   The value of this inner product is zero as expected given that the residuals by construction is orthogonal to the subspace spanned by the regressors, of which Z_1 is one of them.

f. To calculate and Py verify that $Py = \widehat{y}$

$$\begin{aligned} P &= X(X'X)^{-1}X' \\ &= \begin{bmatrix} 2 & 0 \\ 0 & 2 \\ 0 & 0 \end{bmatrix} \begin{bmatrix} \frac{1}{4} & 0 \\ 0 & \frac{1}{4} \end{bmatrix} \begin{bmatrix} 2 & 0 & 0 \\ 0 & 2 & 0 \end{bmatrix} \\ &= \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \end{aligned}$$

Hence

$$Py = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 2 \\ 3 \end{bmatrix} = \begin{bmatrix} 3 \\ 2 \\ 0 \end{bmatrix} = \widehat{y}$$

Notice that $P = X(X'X)^{-1}X'$ is a projection matrix into the subspace spanned by the matrix X. A projection matrix maps points into a subspace and every point in that subspace into itself.P projects every point in the n dimensional space into a subspace of dimension k, the number of columns in the regressor matrix X.

g. It turns out that the rank of P is 2. It does not have full column rank and is hence singular.

Part I, Question 4

Answers given in section 3 of Rust's lecture notes, Endogenous Regressors and Instrumental Variables.

Part II, Question 0

Answers given in section 9 of Rust's lecture notes, Endogenous Regressors and Instrumental Variables.

Part II, Question 1

a. $(X'RX)/49 = 22.262857$. The quantity $X'RX$ can't be negative since the matrix R is semidefinite positive. Let's assume R is a square matrix with dimension N, then

$$c'Rc \geq 0 \quad for\ every \quad c \in R^n$$

Let's now prove that $R$is postivie semidefinite. We know from question 1, part I, that R is idempotent, thus

$$\begin{aligned} R &= R' \\ &and \\ R &= RR \end{aligned}$$

Then, from 2)

$$c'Rc = c'RRc$$

and, from 1)

$$c'RRc = c'R'Rc = (Rc)'(Rc)$$

Hence,

$$c'Rc = (Rc)'(Rc) = z'z$$

with $z = Rc$ and

$$c'Rc = z'z = \sum_{i=1}^{n} z_i^2 \geq 0 \quad for\ every$$

In general, idempotent matrices are semidefinite positive as we just proved.

b. The standard deviation is $\sigma = 4.7183$, and the variance $\sigma^2 = 22.262857$. Basically everybody got this part right.

c. The answers to a. and b. are the same. In order to assess this result, let's look at the quadratic form $X'RX$ more closely. We already proved that, by idempotency of the matrix R, we can write

$$X'RX = (RX)'(RX)$$

Let's now examine the bloc RX. We know that $R = I - i(i'i)^{-1}i'$ Hence,

$$\begin{aligned} RX &= X - i(i'i)^{-1}i'X \\ &= X - i(i'i)^{-1}\overline{X}n \\ &and \\ (i'X)/n &= (\sum_{i=1}^{n} x_i)/n = \overline{X} \end{aligned}$$

and also

$$(i'i)^{-1} = 1/n$$

Then,

$$\begin{aligned} RX &= X - i(i'i)^{-1}\overline{X}n \\ &= X - i(1/n)\overline{X}n \\ &= X - i\overline{X} \end{aligned}$$

It is now clear that, given a generic vector X, the matrix R computes the difference between the original vector and a vector with components equal to the empirical mean of the vector itself. Thus,

$$\begin{aligned} X'RX &= (RX)'(RX) = (X - i\overline{X})'(X - i\overline{X}) \\ &= \sum_{i=1}^{n}(x_i - \overline{x})^2 \end{aligned}$$

But this is the numerator of the empirical variance, $\hat{\sigma}^2$. Hence,

$$X'RX/49 = (\sum_{i=1}^{n}(x_i - \overline{x})^2)/49 = \hat{\sigma}^2.$$

and this value coincides with the value delivered by Gauss when we square the standard deviation. Let's look at the same problem somewhat differently. Let's assume we are seeking to find a statistics capable of summarizing the information contained in the vector $X = (x_1, x_2, ..., x_n)$. The model we have in mind can be set-up as follows

$$X = \vec{i}\beta + \varepsilon$$

where $\vec{i} = (1, 1, \ldots, 1)$, $\varepsilon = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n)$ and β is a scalar coefficient. The model says that the information in the vector X can be approximated by using a vector $\vec{i}\beta$ with constant components equal to β. We know what the formula for the least squares estimator of β, $\hat{\beta}$, is, namely

$$\hat{\beta} = (i'i)^{-1}i'X$$

But, as we saw before,

$$\hat{\beta} = (i'i)^{-1}i'X = \frac{1}{n}\sum_{i=1}^{n}x_i = \overline{X}$$

Hence, the vector which best approximates X is a vector with components equal to the empirical mean of X. Let's now go back to the original problem in order to examine RX again.

$$\begin{aligned} RX &= X - i(i'i)^{-1}i'X \\ &= X - i\hat{\beta} \\ &= X - \hat{X} \\ &= X - i\overline{X} \\ &\textit{and also} \\ &= \hat{\varepsilon} \end{aligned}$$

where $\hat{\varepsilon}$ is the error term of the model we just outlined. Then,

$$X'RX = (RX)'(RX) = \hat{\varepsilon}'\hat{\varepsilon} = SSE \quad \textit{of the model}$$

Part II, Question 2

The linear regression model is as following;

$$Y = \beta_1 \times \vec{i} + \beta_2 X_2 + \beta_3 X_3 + \epsilon$$

$$where \vec{i} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \quad X_2 = \begin{bmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \\ 6 \\ 7 \\ 8 \\ 9 \\ 10 \\ 11 \end{bmatrix} \quad X_3 = \begin{bmatrix} 1 \\ 3 \\ 5 \\ 7 \\ 9 \\ 11 \\ 13 \\ 15 \\ 17 \\ 19 \\ 21 \end{bmatrix}$$

This model can be written in a simpler way;

$$Y = X\beta + \epsilon$$

where $X = (\vec{i}, X_2, X_3)$ and $\beta = (\beta_1, \beta_2, \beta_3)'$.

a. No, we can not compute OLS estimates of the three unknowns $(\beta_1, \beta_2, \beta_3)$ because $X'X$ is singular and is not invertible. The columns of matrix X is linearly dependent because we can perfectly predict X_3 from i and X_2(i.e. $2X_2 - i = X_3$.) Note that Gauss shows the inverse of $X'X$, even though it is not invertible. This is because Gauss computes it by some approximation.

b. Yes, there is a multicollinearity problem in this regression such as $2X_2 - i = X_3$.

c. We have two possible cases for this part. First, if you eliminate X_3, the linear regression model is

$$\begin{aligned} Y &= \beta_1 \times i + \beta_2 X_2 + \epsilon \Rightarrow Y = X\beta + \epsilon \\ where\ X &= \begin{pmatrix} i & X_2 \end{pmatrix} \text{ and } \beta = \begin{pmatrix} \beta_1 & \beta_2 \end{pmatrix}' \end{aligned}$$

then by OLS formula, $\widehat{\beta} = (X'X)^{-1}X'Y$, we can get $\widehat{\beta_1} = -12$, $\widehat{\beta_2} = 2$. Next, we have to find out R^2. By the definition, $R^2 = 1 - \frac{SSE}{SST}$ where $SSE = \sum_i (y_i - \widehat{y}_i)^2$ and here $y_i = \widehat{y_i}$ for $i = 1, ..11$. Therefore, $SSE = 0$ and $R^2 = 1 - \frac{0}{SST} = 1$. It means that the model fits to the data perfectly. Second, if you eliminate X_2, the linear regression model is

$$\begin{aligned} Y &= \beta_1 \times i + \beta_3 X_3 + \epsilon \Rightarrow Y = X\beta + \epsilon \\ where\ X &= \begin{pmatrix} i & X_3 \end{pmatrix} \text{ and } \beta = \begin{pmatrix} \beta_1 & \beta_3 \end{pmatrix}' \end{aligned}$$

then by $\widehat{\beta} = (X'X)^{-1}XY$, we have$\widehat{\beta_1} = -11$, $\widehat{\beta_3} = 1$. Like the first case, we have that $SSE = 0$ and thus $R^2 = 1$(perfect fit). Note that we should not eliminate i because if we eliminate i, we can not depend on R^2 when you compare the model. This is because R^2 of the regression included a constant term may be higher or lower than that excluded a constant term regardless of the truth of the model.

d. First, for (-6,-10,6) the regression model becomes $Y = -6 \times i + -10X_2 + 6X_3 + v$. Then, $v = 0$, $SSE = 0$, and thus $R^2 = 1$. Second, for(-10,-2,2) the regression model becomes $Y = -10 \times i + -2X_2 + 2X_3 + v$. Then, $v = 0$, $SSE = 0$, and thus $R^2 = 1$. Therefore, both answers fit to the data perfectly. We can not judge which estimate fits better to the data in terms of the R^2. Since we have $2X_2 - i = X_3$, we can rewrite the original model as the following;

$$\begin{aligned} Y &= \beta_1 \times i + \beta_2 X_2 + \beta_3 X_3 + \epsilon \ = \beta_1 \times i + \beta_2 X_2 + \beta_3 (2X_2 - i) + \epsilon \\ &= (\beta_1 + \beta_3) \times i + (\beta_2 + 2\beta_3) X_2 + \epsilon \\ &= \gamma_1 \times i + \gamma_2 X_2 + \epsilon \end{aligned}$$

where we let $\gamma_1 = \beta_1 + \beta_3$ and $\gamma_2 = \beta_2 + 2\beta_3$. From the first part of (c), we have $\widehat{\gamma_1} = -12$ and $\widehat{\gamma_2} = 2$ (i.e. $\widehat{\beta_1 + \beta_3} = -12$ (1) and $\widehat{\beta_2 + 2\beta_3} = 2$ (2)) Here, we have two restrictions, (1) and (2), but three unknowns $(\beta_1, \beta_2, \beta_3)$ because of the multicollinearity. We are unable to find $\widehat{\beta_1}$, $\widehat{\beta_2}$ and $\widehat{\beta_3}$ separately, but we can find only $\widehat{\beta_1 + \beta_3}$ and $\widehat{\beta_2 + 2\beta_3}$. Therefore we can't exactly identify the three unknown parameters, and thus there are infinite number of possible parameter vectors. Jim's and Tom's estimators are examples of many possible values. As a result, three unknown parameters are unidentified. When there is a problem of the multicollinearity, we have the identification problem.

Part II, Question 3.

a. I do not expect students to quote any particular theories here: a good common sense answer received full credit. Here is a slightly more "theoretical" explanation. There are theories of money holdings that derive the optimal holdings of *real money balances* for transactions purposes, accounting for the "transactions costs" of going to the bank to switch funds from a checking account or interest bearing time deposit account into cash. See for example, J. Tobin (1956) "The Interest-Elasticity of Transactions Demand for Cash" *Quarterly Journal of Economics* and J.A. Frenkel and B. Jovanovic (1980) "On Transactions and Precautionary Demand for Money" *Quarterly Journal of Economics* (warning: these articles make use of advanced probability theory and may not be easy reading, but by the end of Econ 161 you should be able to at least get the general point of these articles after reading them.) The optimal policy in these theories takes the form of an (s, S) rule: when cash balances decumulate to the lower limit s the individual goes to the bank and withdraws cash. When cash balances accumulate to the upper limit S the individual goes to the bank to deposit the cash in the checking or time deposit account. The higher the rate of spending and income, the more often one needs to go to the bank, so to economize on transactions costs the theory predicts that higher money balances are held. In terms of the regression equation, the theoretical prediction for β_1 is positive. When the interest rate is higher the opportunity cost of holding cash balances is higher, so all other things equal the individual chooses to hold less cash. Thus the theoretical prediction for β_2 is negative. There is no theoretical prediction for β_0 at this level of generality, but note that β_0 does determine the average level of cash balances and is a complicated function of the (s, S) rule and other aspects of the problem.

b. Do the regression with the following Gauss commands:

```
load gnp,cpi,r_3mo,m2;
m=log(m2./cpi);
y=log(gnp./cpi);
x=ones(rows(m),1)~y~r_3mo;
beta=inv(x'x)*x'm;
beta';
 -0.76819  1.20619 -0.01538
```

Thus, the coefficient estimates of β_2 and β_3 are consistent with the theoretical predictions. Note: some students used the natural log, `ln` instead of the base 10 log above. The coefficient estimates are $\hat{\beta} = (-.3336, 1.20619, -.00668)$. Either answer is equally correct. Note: for some reason the ascii version of `CPI` is different than the Gauss `fmt` binary version and lead to different regression results: $\hat{\beta} = (-.66429, 1.1428, -0.013)$ with natural log and $\hat{\beta} = (-.2885, 1.1428, -0.005)$ with base 10 logs. Again I gave full credit for correct answers based on the ascii data or the binary Gauss data.

c. The second regression is basically the same as the first regression but in nominal rather than real amounts. To the extent that there is no "money illusion", theory predicts that the results from the second regression should be basically the same as the first regression. Since $Y = \texttt{GNP}/\texttt{CPI}$, we have $\log(Y) = \log(\texttt{GNP}/\texttt{CPI}) = \log(\texttt{GNP}) - \log(\texttt{CPI})$, and similarly for $\log(M)$. Working through the algebra, we see that if there is no money illusion we should have the parameter restriction $\beta_3 = (1 - \beta_1)$ holding in the second money demand equation since this restriction guarantees that the second regression equation is equivalent to the first:

$$
\begin{aligned}
\log(M_t) &= \beta_0 + \beta_1 \log(Y_t) + \beta_2 r_t + \beta_3 \log(P_t) + \epsilon_t \\
\log(M_t) &= \beta_0 + \beta_1 \log(Y_t) + \beta_2 r_t + (1-\beta_1)\log(P_t) + \epsilon_t \\
\log(M_t) - \log(P_t) &= \beta_0 + \beta_1 (\log(Y_t) - \log(P_t)) + \beta_2 r_t + (1-\beta_1)\log(P_t) + \epsilon_t \\
\log(m_t) &= \beta_0 + \beta_1 \log(y_t) + \beta_2 r_t + \epsilon_t
\end{aligned}
$$

d. Do the second regression with the following Gauss commands:

```
x=ones(rows(m),1)~log(gnp)~r_3mo~log(cpi);
beta=inv(x'x)*x'log(m2);
beta';
 -0.33583 1.2001 -0.00675 -0.19640
 -0.77329 1.2001 -0.15555 -0.19640 /* if ln is used instead of log */
 -0.74465 1.1119 -0.01547 -0.07839 /* if ln is used with ascii data */
 -0.32340 1.1119 -0.00672 -0.07839 /* if log is used with ascii data */
```

We see that although the signs of $\hat{\beta}_1$ and $\hat{\beta}_2$ are consistent with the theoretically predicted signs, the restriction that $\hat{\beta}_3 = (1-\beta_1)$ does not hold exactly, but seems to be very close to holding: $\hat{\beta}_3 = -.1964$ in the second regression versus $(1-\hat{\beta}_1) = -.20619$ in the second regression. The difference might just be due to random estimation "noise". (However note that different CPI indices in the binary vs. ascii version of the data set seem to have had a big effect on the results and restriction doesn't hold as closely when the ascii version of the CPI is used). Later in Econ 161 we will show how to test the hypothesis that the "true coefficients" satisfy the restriction $\beta_3 = (1-\beta_1)$. However without doing a formal hypothesis test (which I didn't expect anyone to do), it seems reasonable to conclude that the empirical evidence is consistent with the hypothesis that there is no money illusion.

e. The plausibility of the estimated coefficients has already been discussed above. The R^2 for the first regression is computed using the following Gauss commands:

```
e=m-x*inv(x'x)*x'm;
rsqr=1-e'e/(sumc((m-meanc(m))^2);
rsqr;
     0.99392
```

The R^2 for the second regression is computed similarly and it has an $R^2 = .99945$, slightly higher than the first regression. This is to be expected since the second regression has an extra coefficient β_3 and this takes away one degreee of freedom improving the regression fit.

f. The Gauss commands for computing the marginal propensity to consume are:

```
load gnp,c;
x=ones(rows(gnp),1)~gnp;
beta=inv(x'x)*x'c;
beta';
     -19.842   0.6435
```

so the estimated aggregate marginal propensity to consume is 64.35 cents out of each dollar of GNP.

g. The Gauss commands to compute the serial correlation coefficient of the residuals from the consumption function regression are as follows:

```
ec=c-x*inv(x'x)*x'c;
ecl=ec[1:23];  /* Lagged errors at t-1 */
ecf=ec[2:24];  /* Current errors at t  */
stdl=sqrt(meanc((ecl-meanc(ecl))^2));
stdf=sqrt(meanc((ecf-meanc(ecf))^2));
rho=meanc((ecf-meanc(ecf)).*(ecl-meanc(ecl)))/(stdl*stdf);
rho;
   0.5654
```

Note: the Gauss `stdc` command divides by $T-1$ instead of T. If you used `stdc` instead of the direct commands above (which divide by T to compute variance), you would get $\rho = .5408$. I did not take off any points for this calculation, although it is technically incorrect. Thus, the residuals of the consumption function equation are positively serially correlated. This is a common occurrence in time series regression problems and could indicate the presence of other omitted serially correlated variables affecting consumption. However the presence of serial correlation does not contradict the normal equations for least squares, which only require the error vector to be orthogonal (i.e. uncorrelated with) the constant term and GNP.

Part II, Question 4.

a. Demand curves should ordinarily be downward sloping (unless we have a Giffen good) so β_1 should be negative. Also, if soybeans are a normal good we should have $\beta_2 > 0$. There is litte one can say about the constant term except that it should be sufficient large positive number so that quantity demanded is not negative.

b. Supply curves should be upward sloping so α_1 should be positive. More rainful should lead to higher production (unless there is excessive rainfall) so α_2 should also be positive. There is little we can say about the constant term except that it should be sufficient large positive number so that quantity demanded is not negative.

c. The program `sdchk.gpr` was used to load in the `soy.asc` data and do OLS regressions for the supply and demand curves. The section of the code that does the regression is as follows:

```
load data[200,4]=soy.asc;
p=data[.,1];
q=data[.,2];
d_shocks=data[.,3];
s_shocks=data[.,4];
n=rows(data);
xd=ones(n,1)~p~d_shocks;
beta=inv(xd'xd)*xd'q;
beta';
     1.733 .267 .359
```

The OLS estimates lead to the counterintuitive finding that the demand for soybeans is positively sloped. Is this because Soybeans are a Giffen good? No! Read on.

The results for the OLS regression of the supply curve parameters are:

```
xs=ones(n,1)~p~s_shocks;
alpha=inv(xs'xs)*xs'q;
alpha';
     8.039  .353  .625
```

The OLS estimates lead to the intuitively plausible finding that the supply curve for soybeans is positively sloped.

e. If there is only a single positively sloped supply curve, then obviously all of the equilibrium (p, q) must lie on this curve as demand shifts up and down due to variations in y. If one tries to estimate the demand curve by OLS, it is intuitively obvious that the estimated demand curve will be upward sloping: essentially you are not estimating the demand curve but the supply curve in this case. This is the essence of the problem of *simultaneous equations bias* that is an aspect studied in more advanced econometrics courses. Simultaneous equation bias is a special case of *endogeneity* in which one or more of the X variables in a regression is correlated with the error term. In this case the price variable p in the regression is correlated with the demand error ϵ_d since it is easy to verify that the equilibrium value of p that sets supply equal to demand is given by:

$$p = \frac{\beta_0 - \alpha_0 + \beta_2 y - \alpha_2 r + \epsilon_d - \epsilon_s}{\alpha_1 - \beta_1}$$

and since p contains ϵ_d, it will be positively correlated with the error term ϵ_d if $\alpha_1 > \beta_1$, and negatively correlated otherwise. Presumably the true α_1 is positive and the true β_1 is negative, so that p is positively correlated with ϵ_d. One can show that this leads to a problem where the OLS estimate $\hat{\beta}_1$ is *upward biased.* In some cases this upward bias can be so strong that the estimated value of $\hat{\beta}_1$ can be positive even if the true β_1 is negative. Is this what is happening in this case? Is there any way to get around the problem of simultaneous equations bias? Yes! Read on.

The method of 2-stage least squares is one way to get around the problem of simultaneous equations bias. It is an example of an *instrumental* *variables estimator* (IV). The idea is to *project* the endogenous variable p in the demand regression on *exogenous variables* (also known as *instrumental variables* which are known not to be correlated with the error term ϵ_d.

f. In this case it seems plausible that the variables y and r could serve as suitable instrumental variables. Indeed, from the way I constructed them they are in fact exogenous variables since they are independent of (and therefore uncorrelated with) ϵ_d and ϵ_s. In addition $\vec{e}$, the $N \times 1$ vector of 1's is also a suitable exogenous variable. Below is the Gauss code that I used to compute the 2SLS estimates

```
z=ones(n,1)~d_shocks~s_shocks;
xhat=z*inv(z'z)*z'x;
yhat=z*inv(z'z)*z'y;
beta2sls=inv(xhat'xhat)*xhat'yhat;
beta2sls';
   8.554  -1.635  .354
```

So the 2SLS estimate of β_1 is now negative, as we would expect. You can use similar methods to compute 2SLS estimate of α. In general you can see that the 2SLS estimates of β are much closer to the true values that I used to generate the artificial data in this problem than the OLS estimates.