Problem Set 8

Multivariate Statistics, 17.846, April 8, 2001

1.0 Griffiths, 9.6

See the STATA log file for instructions on how to do all of this quickly in STATA. We can talk about some advanced functions for STATA on Friday.

The signs appear to make sense, even though only one of them is significant. As the price of beef goes up, people eat less beef. As the price of lamb and pork increase, people shift away from pork and lamb to a substitute (beef), and consumption of beef increases. Also, people appear to eat more meat as wealth increases, though not much more.

The key thing about the variance/covariance matrix is that when you triple the number of observations, the standard deviations decrease by the square root of three. (When you quadruple the number of observations, the standard deviations are cut in half.) Obviously, this is cheating. But I’ve seen it done by mistake - for instance, what if you have data every fifth year, and you replicate it five times (once for each year). This would create serious problems. Time series (autocorrelation) is similar to this problem.

2.0 Griffiths 10.4

Part a)

The total variation is the total sum of squares. The unexplained variation is the error sum of squares. The explained variation is the model sum of squares. Let’s begin by backing out whatever we can:

\[
\hat{\sigma}^2 = \frac{e'e}{n-k} \quad RSS = e'e = (n-k)\hat{\sigma}^2 = (20-3)0.9466 = 16.0922
\]

\[
1 - R^2 = \frac{e'e}{TSS} \quad TSS = \frac{e'e}{1-R^2} = \frac{16.0922}{1-0.9466} = 301.35
\]

\[
MSS = TSS - RSS = 301.35 - 16.09 = 285.26
\]

Part b)

Looking at the T table for 17 degrees of freedom, we find a 95% critical value of 2.11. We take the square roots of the diagonals from the variance/covariance matrix to get the standard deviations of the variables, and multiple these standard deviations by the critical value to get the range:
Part c)
We can easily calculate:

\[
\frac{0.69914 - 1}{0.22029} = -1.366
\]

So 1 is 1.366 standard deviations away from the estimated value of 0.66914. This is below the 95% confidence threshold of 1.74 for a one sided test, so we would not reject the hypothesis that \( B_2 > 1 \). It of course appears more likely that \( B_2 < 1 \), but not so much more likely that we wouldn’t consider the competing hypothesis.

Part d)
We could do a standard (Rb-q) test, since we have the variance/covariance matrix and the beta vector. However, there is a faster way to test the null hypothesis that the model has no explanatory power at all (that is, that all the coefficients save for the constant are zero).

\[
\frac{\frac{R^2}{K - 1}}{\frac{1 - R^2}{N - K}} = 150.68
\]

Checking the F table for F[2,17], we find that this is well above the 3.59 threshold. We reject the hypothesis that the model has no explanatory power with considerable confidence.

Note that the R-squared test above testing the significance of the entire model is a reduced form of the R-squared goodness of fit test which can be used with any series of restrictions. If we have time, we can talk about how to impose restrictions on a model when running it in STATA, and then how to use results from running a restricted regression to test the restrictions.

Part e)
We can run the standard Wald test (Rb-q), where \( R = \) the identity matrix and \( q = \) a vector of zeroes, and we’d get:

\[
\frac{b'\hat{\sigma}^2(X'X)^{-1}b}{K - 1} = \frac{1}{2} \begin{bmatrix} 0.69914 & 1.7769 \end{bmatrix} \begin{bmatrix} 0.0485 & -0.0312 \\ -0.0312 & 0.0371 \end{bmatrix}^{-1} \begin{bmatrix} 0.69914 \\ 1.7769 \end{bmatrix} = 150.6
\]
So indeed, we do get the same answer. Note here that we only used the lower right 2X2 box from the full 3X3 variance/covariance matrix. We can get away with this because the VCV matrix is symmetric, and we are not at all concerned with the first coefficient (the constant).

3.0 Griffiths 10.4

Part a)

This is simply showing that the inverse is correct. I’m going to skip this part as it’s pretty straightforward. As a general note, I will point out that this is a special form of the standard partitioned inverse. Note something about the matrix D:

\[ D = I - x_1(x_1'x_1)^{-1}x_1 \]

The second piece is the projection matrix for a column vector of 1’s. Since we subtract this from the identity matrix I, D is equal to the residual projection matrix. Premultiplying D by any vector will produce the mean deviated vector. Something to keep in handy. In many textbooks, the vector of 1’s is written as \( i \), and thus \( D = I - i(i'i)^{-1}i' \). Sometimes, D is written as \( M_0 \). We will discuss these further when we talk about projection matrices.

If we wanted to invert any partitioned matrix, the general form for the matrix is:

\[ M = \begin{bmatrix} A_{11} & A_{12} \\ A_{21} & A_{22} \end{bmatrix} \text{ Yields } M^{-1} = \begin{bmatrix} A_{11}^{-1}(I + A_{12}FA_{21}A_{11}^{-1}) - A_{11}^{-1}A_{12}F \\ -FA_{21}A_{11}^{-1}F \end{bmatrix} \text{ where } \]

\[ F = (A_{22} - A_{21}A_{11}^{-1}A_{21}) \]

You can of course rearrange this so that a slightly different F is in the upper left corner. Also, if M is symmetric you can simplify it even more.

Part b)

Note that \( X'y = (x_1, X_s')y = \begin{bmatrix} x_1'y \\ X_s'y \end{bmatrix} \)

When calculating \( b_s \) we only look at the second row of \((X'X)^{-1}\) times \( X'y \). We are not concerned with \( b_1 \) here, so this simplifies matters a bit.
Problem Set 8 April 18, 2001

Part c)

We know the residuals of the regression are orthogonal to the x’s by construction., which we write:

\[-(X_s'DX_s)^{-1}X_s'x_1^\prime y + (X_s'DX_s)^{-1}(X_s' y) = \]

\[-(X_s'DX_s)^{-1}X_s'\left(\frac{x_1'x_1'}{T}\right)(y) + (X_s'DX_s)^{-1}(X_s' y) = \]

\[(X_s'DX_s)^{-1}X_s'(I - \frac{x_1'x_1'}{T}))(y)\]

Part d)

We know the residuals of the regression are orthogonal to the x’s by construction., which we write: \(x'\hat{\varepsilon} = 0\)

Next we show: \(D\hat{\varepsilon} = \left(I - \frac{x_1'x_1'}{T}\right)\hat{\varepsilon} = \hat{\varepsilon} - \frac{x_1'x_1'}{T}\hat{\varepsilon} = \hat{\varepsilon} - \frac{x_1'0}{T} = \hat{\varepsilon}\)

Part e)

\[Dy - DX_s b_s = D(x_1b_1 + X_s b_s + \hat{\varepsilon}) - DX_s b_s = Dx_1 b_1 + D\hat{\varepsilon} = 0b_1 + \hat{\varepsilon} = \hat{\varepsilon}\]

Part f)

The result here is true of all projection matrices (and thus also residual projection matrices, since these are also projection matrices too). The symmetry is easy to see. The idempotency is shown below:

\(\left(I - \frac{x_1'x_1'}{T}\right)\left(I - \frac{x_1'x_1'}{T}\right) = II - \frac{x_1'x_1'}{T} - \frac{x_1'x_1'}{T} + \frac{x_1'x_1'}{T}\frac{x_1'x_1'}{T} = \)

\(I - 2I\frac{x_1'x_1'}{T} + \frac{x_1'x_1'x_1'x_1'}{TT} = I - 2I\frac{x_1'x_1'}{T} + \frac{x_1'Tx_1'}{TT} = I - 2I\frac{x_1'x_1'}{T} + \frac{x_1'Tx_1'}{TT} = I - I\frac{x_1'x_1'}{T}\)

By the way, all projection matrices are symmetric and idempotent. This is an often used fact, and helps us with a lot of things.

Part g)
\[ \hat{\epsilon}'\hat{\epsilon} = (Dy - DX_s b_s)'(Dy - DX_s b_s) = \]

\[ y'D'Dy - y'D'DX_s b_s + (-b_s'X_s'D'Dy) + b_s'X_s'D'DX_s b_s = \]

(Noting that \( D'D = DD = D \))

\[ y'Dy - y'D'X_s b_s - b_s'X_s'Dy - b_s'X_s'DX_s b_s \]

Note next that \((Dy)' = y'D'\) and that

\[ Dy = DX_1 b_1 + DX_s b_s + D\hat{\epsilon} = 0 + DX_s b_s + \hat{\epsilon} \]

and thus:

\[ b_s'X_s'Dy = b_s'X_s'(DX_s b_s + \hat{\epsilon}) = b_s'X_s'DX_s b_s \]

since \(X_s\hat{\epsilon} = 0 \)

and we also note that

\[ y'D'X_s b_s = (DX_s b_s + \hat{\epsilon})'X_s b_s = (DX_s b_s)'X_s b_s = b_s'X_s'DX_s b_s = b_s'X_s'DX_s b_s \]

Going back to the original equation, we can now show that:

\[ \hat{\epsilon}'\hat{\epsilon} = y'Dy - b_s'X_s'DX_s b_s - b_s'X_s'DX_s b_s + b_s'X_s'DX_s b_s = y'Dy - b_s'X_s'DX_s b_s \]

Part h)

As noted above, the residual projection matrix yields the mean deviated vector when pre-multiplied by any vector. It is also idempotent and symmetric, so we can easily get the sum of the mean-deviated squared terms out:

\[ y'Dy = y'D'Dy = (Dy)'Dy \]

\[ Dy = \left( I - \left( \frac{x_1 x_1'}{T} \right) \right)y = y - \left( \frac{x_1 x_1'}{T} \right)y = y - x_1 \frac{1}{T} \sum_{t=1}^{T} y_t = y - x_1 \bar{y} \]

Note that \( x_1 \bar{y} \) is just a vector containing the mean of \( y \) in each place. So the vector \( Dy \) contains \( y \) - the mean in each spot. The dot product of \( Dy \) and \( Dy \) is the sum of the squared deviations of \( y \).

Part i)

First note that: \( \text{cov}(b) = \hat{\sigma}^2 (X_s'DX_s) = \frac{\hat{\epsilon}'\hat{\epsilon}}{T-K} (X_s'DX_s) \)
In the step above, we just moved around some scalar values, and then cancelled out the inverse. Note that, the sum of squared errors, and it is therefore a scalar value.

All we need to worry about now is the top piece in the numerator. To tackle this piece, we need to show that:

\[
b_s \left( \frac{\hat{\epsilon}'\hat{\epsilon}}{T-K} (X_s'DX_s)^{-1} \right)^{-1} b_s = \frac{b_s'X_s'DX_s b_s}{K-1} = \frac{\hat{\epsilon}'\hat{\epsilon}}{T-K}
\]

which we know from Part g) and Part f). This just says that the numerator is the Model Sum of Squares - the part of the variance of Y that is explained by the model (and hence the part that is not explained by the error). The D matrix really just takes out of everything the part that can be explained by a constant vector - in other words, it just mean deviations.

4.0 Griffiths, 10.10 a, b, c

See the log file for the regressions. In Part b) we fail to reject the hypothesis of a unitary price elasticity. We do reject the joint hypothesis of a unitary price elasticity and income elasticity. To run this by hand, you could adopt one of two techniques. First, you could write the R matrix:

\[
\begin{bmatrix}
0 & -1 & 0 \\
0 & 0 & 1
\end{bmatrix}
\]

and the q matrix:

\[
\begin{bmatrix}
1 \\
1
\end{bmatrix}
\]

And then calculate the test statistic for a standard F-test. You need to calculate the variance/covariance matrix for the estimated coefficients, of course. You can get this from STATA using the command shown.

Another way would be to implement the restrictions into the model, and then to use a goodness of fit test (which requires less matrix manipulation). Thus, we generate a new dependent variable:

\[
\ln(\tilde{q}) = \ln(q) - \tilde{B}_2 \ln(p) - \tilde{B}_3 \ln(y) = \ln(q) - (-1) \ln(p) - (1) \ln(y) = \ln(q) + \ln(p) - \ln(y)
\]
We would then run the regression against the remaining unrestricted variables - in this case, against the constant. See the log file for the result. The coefficient on the constant when we run a regression without other independent variables is just the mean of the dependent variable. The residual sum of squares and the total sum of squares are identical. We can easily calculate how much explanatory power our restricted model has by taking the total variance of \( \ln(q) \) and subtracting the variance of \( \ln(\tilde{q}) \). We note that the restricted model has less explanatory power than the unrestricted model (stands to reason, Least Squares gets its name from somewhere). But how much less? Significantly less? You can easily run this test using the equation from 11.6.5. (We can talk about this test in class.) This is equivalent to the Rb-q test, and will get you precisely the same answer.

\[
F = \frac{SSE_R - SSE_U}{J} = \frac{5.449 - 4.1935}{2} = \frac{4.1935}{30 - 3} = 4.032
\]

Which is really close to the 4.04 that STATA gives us (discrepancy from rounding error).

Constructing a confidence interval is somewhat harder. First, we select the F critical value for a 95% confidence region for 2 restrictions and 27 degrees of freedom: 3.35.

Next, we write R as: \( R = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \) but we leave the vector q undefined. This R vector will essentially give pull out the two coefficients: \( B_2 \) and \( B_3 \), and also the chunk of the variance/covariance matrix that only deals with these two coefficients (leaving out the coefficient on the constant). If we write this out and place in the estimated values where they are required (using, for instance, equation 10B.12 from GHJ), we get:

\[
\begin{bmatrix} -0.566 - B_2 & 1.434 - B_3 \\ -0.3068 & 2.0025 \end{bmatrix}^{-1} \begin{bmatrix} 0.0523 & -0.3068 \\ -0.3068 & 2.0025 \end{bmatrix}^{-1} \begin{bmatrix} -0.566 - B_2 \\ 1.434 - B_3 \end{bmatrix} < 3.35
\]

If we multiply everything, we get a quadratic equation out. This equation will define an ellipse. All values of the coefficients for which the left hand side is less than 3.35 are within the 95% confidence interval.