Problem Set #3

I. Trade
The trade deficit is a major economic issue in the U.S. Consider the following model of U.S. imports:

\[
\ln \text{IMPORTS}_t = \beta_1 + \beta_2 \ln \text{GDP}_t + \beta_3 \ln \text{CPI}_t + \beta_4 \ln \text{EXCH}_t + u_t
\]

where IMPORTS are nominal imports of goods and services, GDP is nominal gross domestic product, CPI is the consumer price index, and EXCH is the nominal exchange rate. Quarterly data from 1947:1 to 2017:2 are found in “import.dta” in the course folder. Note that we had fixed exchange rates before 1973, and we have no measure of the foreign price level in this regression.

a. Estimate the parameters of this model using data for the fourteen years from 1995:1 to 2008:4. Make sure you take logs, and note the specific time period here.

```
regress limp lgdp lcpi lexch if tin(1995q1,2008q4)
```

Are \( \beta_2, \beta_3, \) and \( \beta_4 \) statistically significant?

b. Issue the command `estat vif` after running the regressions above. What do you learn?

c. Calculate the correlation matrix for IMPORTS, GDP, CPI, and EXCH over the same time period using the command

```
correlate limp lgdp lcpi lexch if tin(1995q1,2008q4)
```

What do you notice?

d. Run the following regression over the same time period:

i) \( \ln \text{GDP}_t = A_1 + A_2 \ln \text{CPI}_t + A_3 \ln \text{EXCH}_t + u_t \)

ii) \( \ln \text{CPI}_t = B_1 + B_2 \ln \text{GDP}_t + B_3 \ln \text{EXCH}_t + u_t \)

iii) \( \ln \text{EXCH}_t = C_1 + C_2 \ln \text{GDP}_t + C_3 \ln \text{CPI}_t + u_t \)

What can you say about the nature of multicollinearity in these data?

e. Estimate the parameters of this model from part a using the data from 1995:1 to 1996:4. What do you notice? Can you explain what happened?

f. Estimate the parameters of this model from part a using the data from 1995:1 to 2015:2. What do you notice? Can you explain what happened? Are \( \beta_2, \beta_3, \) and \( \beta_4 \) statistically significant now?

g. A common procedure for reducing collinearity with time series data is to first-difference the data and then to work with the first-differenced data. Logarithmic first-differencing is particularly attractive, since \( \ln y_t - \ln y_{t-1} \) equals \( \ln (y_t / y_{t-1}) \), which for small changes can be interpreted as the percentage change in \( y \) from period \( t-1 \) to period \( t \). Estimate the above model using the logarithmic first difference for the period 1995:1 to 2008:4. \(^1\)

\[
\ln \text{IMPORTS}_t - \ln \text{IMPORTS}_{t-1} = a_1 + \beta_2 (\ln \text{GDP}_t - \ln \text{GDP}_{t-1}) + \beta_3 (\ln \text{CPI}_t - \ln \text{CPI}_{t-1}) + \beta_3 (\ln \text{EXCH}_t - \ln \text{EXCH}_{t-1}) + u_t
\]

Compare the estimates of \( \beta_2, \beta_3, \) and \( \beta_4 \) to those above. Are \( \beta_2, \beta_3, \) and \( \beta_4 \) statistically significant now?

NOTE: Multicollinearity does not violate any of the assumptions of the Gauss Markov Theorem.

\(^1\) `gen var1 = D.var` creates a new variable “var1” which is the first difference in the variable “var.”
II. More Money
Using the famous square-root formula for money demand along with a partial adjustment model, we have the following money demand regression equation from our previous problem set:
\[ \ln M_t = \alpha + \beta \ln M_{t-1} + \gamma \ln Y_t + \delta \ln i_t + \varepsilon_t \]
a. Estimate a money demand equation for M1 using U.S. data in “money.dta.” After running the regression, issue the commands
\texttt{predict yhat}
\texttt{predict res, residuals}
to create the variables yhat (fitted Y) and res (fitted errors) from the regression. Use the command \texttt{twoway scatter res yhat} to generate a scatter plot of the two variables, res and yhat. Do you see any heteroskedasticity?
The command \texttt{rvfplot} (residual-versus-fitted plot) will do the same thing.
b. Look at your residuals over time. You can use the command \texttt{twoway scatter res time} to generate a scatter plot of the two variables, res and time. Do you see any heteroskedasticity?
c. Look at the relationship between your residuals and ln(GDP). You can use the command \texttt{twoway scatter res lgdp} or \texttt{rvpplot lgdp}. Do you see any heteroskedasticity?
d. Formally test for heteroskedasticity using White’s general test. After running the original regression, issue the following Stata command:
\texttt{estat imtest, white}
e. Test for heteroskedasticity using the Breusch-Pagan/Cook-Weisberg test. After running the regression, issue the following Stata command:
\texttt{estat hettest}
f. Regress the square of the fitted errors (res*res) on the square of ln(GDP). Do they move together? If they do, you might use weighted least squares.
g. Re-estimate the money demand equation using weighted least squares (WLS). Generate a new variable called “wts” where \( wts = 1/\sqrt{\ln GDP^2} \). Use the following general command to estimate a weighted least squares regression:
\texttt{regress y x1 x2 [weight=wts]}
Compare these regression estimates to your regression estimates from above.
h. An alternative remedy for heteroskedasticity is to estimate the regression with White heteroskedasticity consistent standard errors. You can do so in Stata by using the regress option \texttt{vce(robust)}
\texttt{regress y x1 x2, vce(robust)}
Compare this regression with the regression in part a. How does it differ? What are the benefits of this option?

III. Chickens
Lets examine the demand for chickens in the United States from 1960 to 1982. Data on per capita consumption of chickens (in pounds) (PCCON), real disposable per capita income (DISP), real retail price of chickens (PRICE), and the composite real price of chicken substitutes (SUBPR) are available on the course web page under “chicken.dta.”

Consider the following three regression specifications:
\begin{align*}
\text{PCCON}_t &= \gamma_1 + \gamma_2 \text{DISP}_t + \gamma_3 \text{PRICE}_t + \varepsilon_t \\
\ln \text{PCCON}_t &= \alpha_1 + \alpha_2 \ln \text{DISP}_t + \alpha_3 \ln \text{PRICE}_t + \nu_t \\
\ln \text{PCCON}_t &= \beta_1 + \beta_2 \ln \text{DISP}_t + \beta_3 \ln \text{PRICE}_t + \beta_4 \ln \text{SUBPR}_t + u_t
\end{align*}
(A) (B) (C)
a. Estimate equations (A) and (B), and use the Ramsey RESET test to determine the better regression specification. `estat ovtest`
b. Suppose (C) is the true model, but you have estimated (B). What are the theoretical consequences? What kind of specification tests would you do?
c. Suppose you estimate (C), the true model, and $\beta_4$ turns out to be statistically insignificant. Should we drop the price of substitute products as an explanatory variable in the demand function? Does this mean there is no specification error if we fit (B) to the data?
d. Now assume that (B) is the true demand function. What type of specification error is committed in this instance from estimating (C)? What are the theoretical consequences of this specification error? Illustrate your point with the regression results.

_Due Wednesday 1 November_