Problem Set #5
14.74
Due Monday, April 19, 3004 by 6:00pm

1. **Savings: Evidence from Thailand**
Paxson (1992), in her article entitled "Using Weather Variability to Estimate the Response of Savings to Transitory Income in Thailand" (which is in the course reader), attempts to estimate the marginal propensity to save out of transitory income. She states that "a finding that these marginal propensities are high would indicate that farmers do use savings to smooth consumption" (p. 15).

   Read the article, since you will need to refer to specific tables in the article to complete the following empirical exercise. This exercise uses the data set paxson.dta. At the beginning of your Stata session, type "**`set memory 20m`**" and also "**`set matsize 200`**" because you will be using more memory and variables than the default setting.

   a. What is the problem with measuring the marginal propensity to save out of total income by regressing savings on total income?

   b. Paxson only has data on total income (**`inc`**). She wants to decompose it into a permanent income component (call it **`incperm`**) and a transitory income component (call it **`inctrans`**). In this part, you will replicate the results in Table 3 and obtain estimates of **`incperm`** and **`inctemp`**.
      i. Run the regression that Paxson ran in Table 3, Column 1 (the income equation). Include all the variables you see in the table, as well as the region dummies mentioned in the footnote. (When I do this, I get numerically identical estimates as Paxson.) This part is not hard, but it gets tedious because of all the variables that must be included. Note you will have to form dummy variables for year and region.)
      ii. Using the coefficients estimated in (i), Paxson constructs a predicted value for permanent income as follows. She multiplies the permanent characteristics by their respective coefficients from (i), and adds it all up to form **`incperm`** (see equation 2). [Stata hint: Suppose in (i), I just ran "**`reg inc x z`**" where **`x`** is a permanent characteristic and **`z`** is a transitory characteristic. Then I would form **`incperm`** by typing: "**`gen incperm = _b[x]*x`**". So note the syntax: the "**`_b[x]`**" refers to the coefficient estimated for the independent variable, and you ask Stata to multiply it by **`x`** in order to form the predicted value of permanent income.] Generate the variable **`incperm`**. What is the mean of **`incperm`**?
      iii. Paxson obtains a predicted value for transitory income in a similar way. She multiplies the transitory characteristics by their respective coefficients from (i), and adds it all up to form **`inctrans`** (see equation 3). Which variables from (i) are considered to be transitory characteristics? Generate the variable **`inctrans`**. What is the mean of **`inctrans`**?
      iv. Paxson also has a category called unexplained income, defined as: **`inc`** – **`incperm`** – **`inctrans`**. Form this variable, and call it **`incunexp`**. What is the mean of **`incunexp`**?

v. Now you are ready to estimate the effect of income on savings. Run the regression that Paxson ran in Table 4, Column 2 (using **save2** as the savings measure). Include all the variables you see in the table, as well as the year dummies mentioned in the footnote. (When I do this, I get very close to Paxson's results, but not exactly the same. So don't worry if you don't exactly replicate her results.) Paxson interprets the coefficient for **inctrans** as the marginal propensity to save out of transitory income. Based on your regression results, what is the marginal propensity to save out of each dollar of transitory income? Test the null hypothesis that this marginal propensity equals one. [Stata hint: Suppose you just ran "**reg save2 inctemp x1 x2**" and you wanted to test whether the coefficient for **inctemp** equals zero. You can just look at the Stata regression table and see whether the t-statistic is above the critical value, or see whether zero is in the confidence interval, or see whether the p-value is <= the significance level. Or, you can type "**test inctemp=0**" and the p-value will be given. Now, for a test of the coefficient for **inctemp** equal to one, you can either look at the confidence interval or type "**test inctemp=1**".]

c. You are worried that Paxson's estimates are biased because she omitted lots of variables from the Table 4 regression (savings) that she included in the Table 3 regression (income). You believe that rain deviations have an effect on savings only through its effect on transitory income. What you don't believe is that the permanent characteristics mentioned in (ii) have no effect on savings other than through permanent income. For example, some regions may have a greater tradition of saving than others, so the region dummies should affect both permanent income directly and savings directly. As another example, families with more working age males may be less inclined to save since they feel more confident of the income stream, so the household demographic characteristics should affect both permanent income directly and savings directly.

In short, you question the exclusion restrictions that Paxson has imposed to estimate the savings equation. She has assumed that the permanent characteristics can be excluded from the savings equation. You believe this will cause omitted variable bias.

You now fix this omitted variable bias, but recognize that in so doing you will not be able to estimate the marginal propensity to save out of permanent income. This is because **incperm** is a linear combination of the permanent characteristics, and now you will be including all the permanent characteristics in the regression [recall that the right-hand side variables in a regression must be linearly independent].

i. Run the regression that you ran in part (a)(v), but this time drop **incperm** and **incunexp**, and add all the permanent characteristics from the regression in part (a)(i). Based on your regression results, what is the marginal propensity to save out of each dollar of transitory income? Test the null hypothesis that this marginal propensity equals one.
ii. Compare the result in (i) to the result in part (a)(v). Are they similar? (No Stata needed here – just eyeball your coefficients.)
iii. What exclusion restriction have we assumed here in order to estimate the marginal propensity to save out of transitory income? [Hint: The answer is in the preamble to this question.] Specify exactly which variables you have used to estimate transitory income but not to estimate savings. Do you think omitting these variables will cause

omitted variable bias in the savings equation, considering all the other control variables we have included? Why or why not?

d. In fact, using a two-step process (as we've been doing – we form predicted values for transitory income, then plug it into the savings regression) is inefficient. We can use two-stage least squares (2SLS) to estimate the marginal propensity to save out of transitory income in a single step. 2SLS is an *instrumental variables* estimation procedure.

Let us assume that the rainfall deviations variables only affect transitory income, not permanent income. This is to say that any variation is **inc** that is caused by the rainfall deviations can be interpreted as variations in **inctrans**. If we are willing to make this assumption – and it seems reasonable – then we can use 2SLS to estimate the following equation:

$$\mathbf{save2} = \alpha + \beta * \mathbf{inc} + \mathbf{X}p + \varepsilon$$

where **X** contains all the independent variables used in part (b)(i). Obviously, **inc** is endogenous (income and savings are determined simultaneously, so we can expect **inc** to be correlated with the error term). We will use the rainfall deviations variables to instrument for **inc**. Then β will give the estimate the marginal propensity to save out of transitory income.

i. In order to be a valid instrumental variable, a variable must satisfy two conditions. The first is that it be correlated with the variable it is instrumenting for (in this case, income). The second is that it be uncorrelated with the error term. Do you think these conditions hold here? Explain.

ii. Run 2SLS. [Stata hint: Suppose you want to estimate the above **save2** equation using variables **z1** and **z2** as instruments for **inc**. Then type "**ivreg save2 X (inc = z1 z2 X)**". So note the syntax. What's in the parentheses is the first stage, where you tell Stata what your endogenous regressor is, what your instrumental variables are, and what your other independent variables are. What's outside the parentheses is the second stage, which is the name for the above **save2** equation.] Based on your regression results, what is the marginal propensity to save out of each dollar of transitory income? Test the null hypothesis that this marginal propensity equals one.

iii. Compare the result in (ii) to those in (a)(v) and (b)(i). Are they similar? (No Stata needed here – just eyeball your coefficients.)

iv. Does the result here suggest that households use savings to smooth consumption?

2. **Insurance and Incentives**

Consider a village of a large number of identical farmers. All farmers are owner-cultivators; i.e., they own and work their own land. They can choose how much effort p to apply to their land, where p is either p_L or p_H ($p_L < p_H$). With probability p their crop is a success, and their output is H; with probability $(1-p)$ their crop is a failure, and their output is $L<H$. For consumption equal to y, utility over consumption and effort is given by $u(y) - C(p)$. $C(p)$ is the cost of effort; assume that $C(p_L) = 0$ and $C(p_H) = K$, a positive constant. Farmers are risk-averse.

a. Assume there is no insurance. What condition must be satisfied for farmers to choose effort level p_H? For the remainder of this problem, assume that this condition is satisfied.

What is each farmer's expected consumption under this choice of effort? What is their expected utility?

b. What does risk-aversion imply about the shape of u? Explain why we interpret this condition on u as risk-aversion.
c. Suppose now that the farmers form a group insurance mechanism. Assume that each farmer's effort level is observable to all other farmers. Consider a full insurance mechanism which requires that each farmer choose effort p_H, all farmers pool their output, and then each farmer gets an equal, nonrandom share.
   i. Is such a mechanism feasible? Explain why? (Hint: There are two conditions that must be satisfied; all farmers must choose p_H and their consumption must be nonrandom).
   ii. What is each farmer's expected consumption under this mechanism? How does this compare with expected consumption in (a)?
   iii. What is each farmer's expected utility under this mechanism? How does this compare with the expected utility in (a)?
   iv. Explain the reason for the difference in your answers to parts (ii) and (iii).
d. Now suppose that each farmer's choice of effort level is *unobservable*. Is the mechanism in part (c) still feasible? Why or why not?
e. We continue to assume that each farmer's choice of effort level is unobservable. We will now solve for the optimal, feasible insurance mechanism. Consider the following mechanism: After output is produced, all farmers pool their output. If a farmer's output was H, he gets to consume h; if his output was L, he gets to consume l. In the case of full insurance, $h=l$.
   i. What "budget constraint" must be satisfied to ensure that the total insurance payments made equals the total amount taken in?
   ii. What condition must be satisfied to ensure that all farmers choose p_H? Show that this condition is violated in the case of full insurance.
   iii. What is each farmer's expected consumption from this insurance mechanism?
   iv. Argue from (i)-(iii) that $L<l<h<H$ and that expected utility under this mechanism is greater than expected utility under no insurance (i.e., part (a)) but lower than expected utility under full insurance (i.e., part (c)).
f. Suppose $L=1$ (i.e., the number one, not lower-case "L") and $u(y) = ln(y)$. [Hint: Recall that $ln(ab) = ln(a) + ln(b)$ and $ln(a/b) = ln(a) – ln(b)$.
   i. Calculate the optimal h and l based on part (e).
   ii. Show that expected consumption is the same as in the case of no insurance.
   iii. Show that farmers face less variation in consumption than in the no insurance case, but more variation than in the full insurance case.

3. **Credit as Insurance**

Read the following article in your course reader: Christopher Udry (1990), "Credit Markets in Northern Nigeria: Credit as Insurance in a Rural Economy," *World Bank Economic Review* 4(3), 251-269.

For each part below, please limit your response to about one paragraph (the entire question should take about a page). Do not summarize the article; please answer specifically and precisely the questions asked.

a. In the Nigerian villages that Udry studies, there is active borrowing and lending within villages and within kinship groups. How does Udry demonstrate that there are risk-sharing arrangements between borrowers and lenders? Please refer to *specific results in tables* to support your answer.
b. Someone might argue that Udry's empirical results are due to limited liability (which leads to sharecropping arrangements between borrowers and lenders). In what way would this person be right? Why can't this person's explanation fully account for Udry's empirical results? Please refer to *specific results in the tables* to support your answer.
c. Relate Udry's findings to lecture regarding the difficulty of enforcement: how can Udry's findings explain why it is often not economically feasible for formal institutions to lend to, or provide insurance for, poor, rural households? Even though informal insurance mechanisms may exist and operate well, what is a limitation that they have which formal institutions could improve upon?