

Seminar 6

Endogeneity (Part II)

1 Measurement Error

1. We would like to run the following regression using "Life in Kyrgyzstan" household survey for 2012:

$$\text{Remittances} = \beta_0 + \beta_1 \text{income} + \beta_2 \text{hhsiz} + \beta_3 \text{males_over_15} + \beta_4 \text{child_under_5} + \beta_5 \text{female_hh_head} + \beta_6 \text{married_hh_head} + u$$

where

Remittances=amount of remittances sent by labor migrants in KGS:

- (a) H617="How much money did the migrant send during the last 12 months, for example, carried by migrant himself, by friends/relatives, or sent via bank/money transfer agencies?"
- (b) H502="Amount of the following income source per month, on average: Money transfers from persons living abroad";

income=income excluding remittances in KGS;

hhsiz=household size;

males_over_15=number of males over age 15;

child_under_5=number of children under age 5;

female_hh_head=dummy variable equal to "1" if the household head is "female" and "0" otherwise;

married_hh_head=dummy variable equal to "1" if the household head is "married" and "0" otherwise.

2. As the required data are contained in different Stata data files ("**hh1a**", "**hh4b**", "**hh5**", "**hh6b**" from the Household questionnaire, we need to merge them into one data file before proceeding with estimation.
 - (a) Import the Stata data file "**hh1a**" from the e-course platform.
 - i. Save it in your computer and name the file "**hh1a_new**".

- ii. Keep only the following variables: `hhid pid age h102 h104 h108`.
 - iii. Generate the new variable `hhsize`.
 - iv. Generate the new variable `males_over_15`.
 - v. Generate the new variable `child_under_5`.
 - vi. Generate the new variable `female_hh_head`.
 - vii. Generate the new variable `married_hh_head`.
 - viii. Drop the variables `pid, age, h102, h104, h108`.
 - ix. Delete duplicate observations.
 - x. Check if the variable `hhid` uniquely identifies observations in the data (i.e. if there are any duplicates), and save the changes.
- (b) Import the Stata data file "`hh4b`" from the e-course platform.
- i. Save it in your computer and name the file "`hh4b_new`".
 - ii. Generate the new variable `health_expenses_all`, which is the sum of expenditures on medicines and medical care, including dental care, in KGS per month.
 - iii. Drop the variables `n4b h403 h404 h405 health_expenses`
 - iv. Delete duplicate observations.
 - v. Check if the variable `hhid` uniquely identifies observations in the data (i.e. if there are any duplicates), and save the changes.
- (c) Import the Stata data file "`hh5`" from the e-course platform.
- i. Save it in your computer and name the file "`hh5_new`".
 - ii. Generate the new variable `all_income`, which is household income including remittances.
 - iii. Generate the new variable `remittance_2`, which is the amount of money transfers from persons living abroad.
 - iv. Generate the new variable `income`, which is household income excluding remittances.
 - v. Drop the variables `n5 h501 h502 all_income`.
 - vi. Delete duplicate observations.
 - vii. Check if the variable `hhid` uniquely identifies observations in the data (i.e. if there are any duplicates), and save the changes.

- (d) Import the Stata data file "hh6b" from the e-course platform.
- i. Save it in your computer and name the file "hh6b_new".
 - ii. Keep only the following variables: `hhid` `h617_s` `h617_c` `h617_t`.
 - iii. Generate the new variable `remittance_1`, which is the amount of money that the migrant sent during the last 12 months, for example, carried by migrant himself, by friends/relatives, or sent via bank-money transfer agencies.

Note: the official exchange rate (average for the whole year) for 2012:

USD-KGS: 47.01;
RUB-KGS: 1.51;
EUR-KGS: 60.44.
 - iv. Drop the variables `h617_s` `h617_c` `h617_t`.
 - v. Check if the variable `hhid` uniquely identifies observations in the data (i.e. if there are any duplicates), and save the changes.
- (e) Open the "hh1a_new" data file again and merge it with "hh4b_new".
- i. Now merge the newly merged data with `hh5_new`.
 - ii. Now merge the newly merged data with `hh6b_new`.
 - iii. As there are only 414 matched observations during the last merging (remittances were reported only for those households who reported having received them), we need to replace missing values for `remittance_1` with zero values.
 - iv. Now obtain summary statistics of `remittance_1` and `remittance_2` variables to see if there are any differences between them.

3. Run the linear regression given by

$$\text{Remittances} = \beta_0 + \beta_1 \text{income} + \beta_2 \text{hhsiz} + \beta_3 \text{males_over_15} + \beta_4 \text{child_under_5} + \beta_5 \text{female_hh_head} + \beta_6 \text{married_hh_head} + u$$

using `remittance_1` measure.

- (a) Now run the same regression but using `remittance_2` measure.
- (b) Compare the OLS coefficients for `income`, `hhsiz`, `males_over_15`, `child_under_5`, `female_hh_head`, `married_hh_head`. What can you say?

4. Now we would like to run the following regression

$$\text{Health_expenses_all} = \beta_0 + \beta_1 \text{remittances} + \beta_2 \text{income} + \beta_3 \text{hhsiz} + \beta_4 \text{males_over_15} + \beta_5 \text{child_under_5} + \beta_6 \text{female_hh_head} + \beta_7 \text{married_hh_head} + u$$

where

Health_expenses_all=expenditures on medicines and medical care, including dental care, in KGS per month.

- (a) Run this regression using **remittance_1** measure.
- (b) Now run the same regression but using **remittance_2** measure.
- (c) Compare the OLS coefficients for **remittance_1** and **remittance_2**.
What can you say?