Implementing Wolfhard Kaus study on visible consumption in South Africa

May 19, 2016

Regressions on Income and Expenditure Survey

Kaus considers a black-dummy and coloured-dummy in regression of visible consumption against demographic, time, permanent income and household characterisitcs. Starting with no controls(I)[1] (with black-dummy and coloured-dummy), the coefficients for coloured and black are negative (i.e. black and coloured social groups spend less on visible consumption than the white population). However, these coefficients flip signs after adding income controls(II)[2] .

The coefficients for black and coloured groups are even higher when expenditure is added(III)[3]. The endogeneity of total expenditure requires instrumentation - Kaus uses instruments suggested by Charles, Roussanov and Hurst - including positive-income-dummy, cube-of-income-level, dummies for under-secondary-education, secondary-eduation and degree-education. Kaus then performs two 2SLS regressions - one with income controls and total-expenditure as control variables[4] and the other with education-dummies as control variables (while still using instruments)[5]. The instrumentation thus performed (Specification IV in Table 1) results in lower coefficients for the black and coloured dummies for both regressions.

[1] `lm(lnvis~black_dummy+coloured_dummy)`

[2] `lm(lnvis~black_dummy+coloured_dummy+ lninc)` - where lninc is log(income_household_head). Notice that only log of income level is found significant when used along side with income-level-cubic and income level.

[3] `lm(lnvis~black_dummy+coloured_dummy+ lninc+lnpinc)` - where lnpinc is log of total expenditure (which itself is a proxy for permanent income)

[4] `ivreg(lnvis~black_dummy+coloured_dummy+ lninc+ lnpinc | . - lnpinc + cbinc + lsecd + secd + degree)` - where lsecd,secd and degree are education dummies and cbinc is income-level-cubic.

[5] `ivreg(lnvis~black_dummy+coloured_dummy+ lnpinc +lsecd | . - lnpinc + cbinc+lninc +incpsv)`

	Controls	Black	Coloured
I	No Controls ($R^2 = .05$)	-1.03	-0.91
II	Income Controls ($R^2 = .26$)	0.37	0.30
III	Log Expenditure ($R^2 = 0.41$)	0.74	0.46
IV	2SLS Regression ($R^2 = \{0.41, 0.41\}$)	{0.62,0.71}	{0.41,0.43}
VI	2SLS Regression with Demographics ($R^2 = \{0.41, 0.41\}$)	{0.36,0.47}	{0.19,0.24}

Table 1: Results from regressing ln(visible_consumption) against respective set of control variables

	Controls	Black	Coloured
I	No Controls ($R^2 = .07$)	-0.22	-0.16
II	Income Controls ($R^2 = 0.09$)	-0.18	-0.14
III	Log Expenditure ($R^2 = 0.16$)	-0.10	-0.10
IV	2SLS Regression ($R^2 = \{0.14, 0.16\}$)	{-0.15,-0.10}	{-0.12,-0.10}
V	2SLS Regression with year-dummies ($R^2 = 0.14, 0.16$)	{-0.16,-0.10}	{-0.13,-0.10}
VI	2SLS Regression with Demographics ($R^2 = \{0.17, 0.17\}$)	{-0.16,-0.12}	{-0.11,-0.09}

Table 2: Results from regressing ln(visible_consumption) against respective set of control variables for years 2005 and 2010

When demographic params are added(VI), the coefficients are further lowered (without changing R^2 a lot) - for both types of regressions [6].

All of the above interpretations match with those reported in Kaus's paper. The regression V from the paper is ommitted for 1995 - as the data on visible consumption from 2005/2010 wasn't available from datafirst to be included in the regression with year-dummies. The data from 2000 - is of poor quality (as explained by Kaus) and does not have enough household characteristic fields for applying an analysis in the Kaus's paper - and was thus discarded altogether. For 2005/2010, the only field amongst all in the visible category available to us - is the number-of-cars-owned by the household (rather than cars purchased in the surveyed period). Hardly meant to conclusive on visible consumption, if owned cars were still to be considered the proxy for visible consumption, a reversal of signs after adding of income-controls is not observed. Other results equivalent to 1995 result-set are summarized in Table 2 (Notice that specification V is possible when more than an year is available for analysis). What the results suggest is only that unlike visible consumption, the ownership of cars is very much influenced by permanent income (and other wealth characteristics accounting for lower ownership of cars for the black and coloured groups).

Analysis

The first-change of sign when income controls are added implies that after accounting for income of household head, the coloured and black groups spend more on visible consumption. Total expenditure is not a perfect proxy of permanent income - which justifies instrumentation. After resolving these endogeneity concerns - with a regression of visible expenditure as dependent variable and total-expenditure as one of the control variables - the signs are revised - although R^2 does not change a lot. Adding demographic controls to the 2SLS regression - shows significance of age, urban/rural area and family size - but coefficients are still not lowered significantly. This confirms that coloured and black social groups spend a rather disprportionate amount on visible consumption.

The second claim that the paper is interested in is testing whether the rise in average group income reduces the visible consumption. The regression considers average-incomes of particular provinces for every group (black/white/coloured). This regression shows the effect more on black-social-group

[6] First: `ivreg(lnvis~black_dummy+coloured_dummy+ lninc+ lnpinc + age+ n_members + area_type| . - lnpinc + cbinc + lsecd + secd + degree)` Second: `ivreg(lnvis~black_dummy+coloured_dummy+ lnpinc +lsecd + age + n_members + area_type | . - lnpinc + cbinc+lninc +incpsv)` - where n_members is the number of members in the household and area_type is urban/rural factor

than the white-social-group - a difference that author argues points out alternate ways of signaling for white-social-group relative to black-social-group.

AER toolbox ivreg notation

Regressors and instruments in implementaiton of ivreg in R-toolbox AER are specified in a formula with two parts on the right-hand side. The specification - used in this note - can be either of the form - `y ~ ex + en | ex + in` or as `y ~ ex + en | . - en +in` (where `ex` is the set of exogenous variables, `en` is the set of endogenous variables and `in` the set of instruments) . The latter is the notation used throughout in this note.