

Implementing Wolfhard Kaus study on visible consumption in South Africa

May 18, 2016

Regressions on Income and Expenditure Survey

With no controls(I)¹, the coefficients for coloured and black are negative, but these coefficients flip signs after adding income controls(II)².

These coefficients for black and coloured groups are even higher when expenditure is added(III)³. The endogeneity of total expenditure is a valid concern and 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-education and degree-education. Kaus performs two 2SLS regression - one with income controls and total-expenditure as control variables⁴ and the other with education-dummies as control variables (while still using instruments)⁵. The instrumentation thus performed (IV) results in lower coefficients for the black and coloured dummies for both regressions.

When demographic params are added(VI), the coefficients are further lowered (without changing R^2 a lot) - for both types of regressions⁶.

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 +in cpsv)`

⁶First: `ivreg(lnvis~black_dummy+coloured_dummy+ lninc+ lnpinc + age+ n_members + area_type| . - lnpinc`

	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 $\ln(\text{visible_consumption})$ against respective set of control variables for years 2005 and 2010

The regression V from the paper is omitted because the data on visible consumption for 2005/2010 wasn't available from datafirst. The data from 2000 - is of poor quality (as explained by Kaus) and does not have enough household characteristics for a detailed analysis. For 2005/2010, the field we have information for amongst all fields in the visible categories- is the number-of-cars-owned by the household (rather than cars purchased in the surveyed period). 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 are summarized in Table 2.

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. This is partly because of permanent income - which results in lowered coefficients as total-expenditure is introduced. After resolving the endogeneity concerns (i.e. in a regression of visible expenditure as dependent variable and total-expenditure as one of the control variables), the signs are revised - although R^2 doesn't 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.

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 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.

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