METADATA: Linked spatial-attitudinal dataset for South Africa

Data preparation methodology

In order to test whether people's individual attitudes towards inequality and redress are influenced by their lived experience of inequality it was necessary to perform three main data preparation steps. The first step involved attaching area-level measures of lived experience of inequality (and area-level poverty rate) to the individual survey respondent-level South African Social Attitudes Survey (SASAS) dataset. The second step involved reviewing the SASAS questionnaire and retaining only those survey questions (plus the merged area-level inequality and poverty measures) that were regarded as being potentially important control variables. The third step involved splitting the retained SASAS dataset into the poor/non-poor subset to enable appropriately specified models to be run for both groups.

Methodology for merging area-level data into SASAS

Attaching area-level variables to individual-level survey records requires details of the geographical home location of each survey respondent. The sampling methodology underpinning the SASAS data collection utilised the Enumeration Area (EA) geography in the stratification process and so each survey respondent has an EA code. In order to preserve respondent confidentiality, the EA codes are excluded from the standard SASAS datasets that are made available to researchers. However, for the purpose of this ESRC-funded research project, special dispensation was granted by the Human Sciences Research Council (HSRC) to enable the respondents' home EA codes to be included in the base survey dataset. The EA code represents the key data linkage variable through which the area-level variables were attached to the individual-level survey records.

The datazone geography was generated by combining EAs in such a way as to maximise adherence to a number of rules, including population size thresholds and population homogeneity measures. As such, EAs nest perfectly within datazones and so it was possible to generate an EA-to-datazone lookup table. This EA-to-datazone lookup table was matched into the individual-level survey response records using the EA code as the common link variable.

Once the datazone code was successfully attached to the survey records, it was then possible to merge in the area-level inequality and poverty measures using the datazone

code as the common link variable. Upon completion of this matching process the EA code was deleted from the matched dataset to preserve respondents' confidentiality.

The matched dataset therefore consisted of all the individual-level survey variables plus the home datazone code and datazone-level inequality and poverty measures.

Selecting potentially important control variables

The 2009 SASAS questionnaire contained a large number of survey questions. Only a minority of these questions related to issues around inequality (as the survey also contained various other modules on topics such as experience of crime and fear of crime). In order to make the modelling process manageable, a review was undertaken of the entire SASAS questionnaire and a subset of questions identified that were deemed to be particularly important for this component of the research project. The objective here was to ensure that the models controlled for as many measureable explanatory factors as possible in order to maximise confidence in any observed effects on the response variables by 'experience of inequality'. Table 4.2 lists the SASAS variables that were selected for inclusion in the base modelling dataset.

Table 4.2: Variables selected for base modelling file

	Variable name	Variable type	Variable description
	uniqueid	-----	Individual survey respondent unique identifier code
	benchwgt	-----	Composite survey weight
	ineqavr	Ordinal	Q186. To what extent do you agree or disagree that differences in income in South Africa are too large?
	govredr	Ordinal	Q187. To what extent do you agree or disagree that it is the responsibility of the government to reduce the differences in income between people with high incomes and those with low incomes?
	age	Numerical	Q238. Age of respondent in completed years.
	agesq	Numerical	Derived 'Aged squared' indicator, based on Q238
	race	Categorical	Respondent's population group (taken from 'respondent selection procedure' questions).
	marstat	Categorical	Q239. What is your current marital status?
	hhper	Numerical	Number of persons in this household (taken from 'respondent selection procedure' questions)
	assetindex	Numerical	Derived 'Asset Index' indicator, based on 25 separate items (Q267-Q300): e.g. Q281. Does your household have a washing machine (in working order)?
	edu	Categorical	Q242. What is the highest level of education that you have

			ever completed?
	empl	Categorical	Q246. What is your current employment status?
	spoor	Ordinal	Q151. Would you say that you and your family are… e.g. 'wealthy/very comfortable'
	topbott00	Numerical	Derived indictor (transformed to 0-100 scale) based on Q198. In our society there are groups which tend to be towards the top and groups which tend to be towards the bottom. Where would you put yourself on a scale of 1 to 10, where 10 is the top and 1 the bottom?
	ssocmobc	Ordinal	Q2. In the last 5 years, has life improved, stayed the same or gotten worse for people like you?
	futmob	Ordinal	Q3. Do you think that life will improve, stay the same or get worse in the next 5 years for people like you?
	jobprest	Ordinal	Q200. Please think about your present job (or your last one if you don't have one now). If you compare this job to the job your father had when you were 15, would you say that the level of status of your job is (or was)… e.g. 'Much higher than your father's'
	classconind	Numerical	Derived 'Class conflict index' indicator, based on multiple separate items, e.g. Q195. In your opinion, in South Africa how much conflict is there between the working class and the middle class?
	groupdis	Ordinal	Q59. Would you describe yourself as being a member of a group that is discriminated against in this country?
	meritind	Numerical	Derived 'Merit factor index', based on multiple separate items, e.g. Q164. How important is hard work for getting ahead in life?
	exogind	Numerical	Derived 'Exogenous factors index', based on multiple separate items, e.g. 160. How important is coming from a wealthy family for getting ahead in life?
	q68r	Ordinal	Q168. How important is a person's race for getting ahead in life?
	polideol	Categorical	Q235. In political matters, people talk of 'the left' and 'the right' or 'liberal' and 'conservative'. Where would you place your views on this scale?
	anc	Ordinal	Derived 'ANC voter' indicator based on Q231. If there were a national election tomorrow, for which party would you vote?
	geotype	Categorical	Statistics South Africa Census 2001 enumeration area Geotype classification
	dz_code	-----	Datazone unique identifier code
	mun_name	-----	Municipality name
	prov	-----	Province name
	exposure_of_poor	Ratio	aLDPxyi* exposure to inequality measure developed above in Chapter 2

	exposure_of_rich	Ratio	aLDPyxi* exposure to inequality measure developed above in Chapter 2
	inc	Ratio	Proportion of datazone population that is classified as being deprived on the 'Income and Material Deprivation Domain' of the SAIMD 2001 at datazone level

Splitting the file into poor and non-poor

As part of this project, two separate but complementary measures of the 'lived experience of inequality' were developed and analysed: one relating to the experience of the poor, and the other relating to the experience of the non-poor. In order to test whether a person's experience of inequality influences their attitudes towards inequality and options for redress, it is important to include the inequality measure that corresponds to the person's own poverty status. In short, the measure of inequality experienced by the poor is only relevant for people who would be classified as poor, whilst the measure of inequality experienced by the non-poor is only relevant for the people who would be classified as non-poor. In order to achieve this data and model configuration it was therefore necessary to split the SASAS dataset into two distinct subsets: one consisting of respondents defined as poor and one consisting of respondents defined as non-poor.

There is currently no single definitive poverty threshold for South Africa. For the purpose of developing a measure of lived experience of inequality in this research project, the distinction between poor and non-poor is provided by the South African Index of Multiple Deprivation 2001 at Datazone level (SAIMD 2001). According to the Income and Material Deprivation Domain of the SAIMD 2001, 73% of the total population of South Africa is defined as poor, with the remaining 27% defined as non-poor. This 73%-26% split is therefore adopted for the purposes of the modelling in this chapter.

Although SASAS does contain a question on 'income', a considerable proportion of respondents either declined to provide an answer or stated that they did not know their income. This resulted in a substantial amount of missing data in the income variable of the SASAS dataset and therefore it was not possible to use the income variable as the basis for classifying people as poor/non-poor. However, SASAS also contains a series of questions relating to people's material asset ownership which, for the purpose of this research project, were combined together to form a composite asset index. The overall SASAS dataset was therefore sorted according to the respondents' score on the asset index variable, resulting in a ranking from the person with the lowest material asset ownership to the respondent with the highest material asset ownership. The SASAS dataset was then split into a poor subset and a non-poor subset, with the poor subset consisting of the 73% of

respondents with the lowest scores on the asset index, and the non-poor subset consisting of the 27% of respondents with the highest scores on the asset index.

The poor subset contained a total of 2120 cases, while the non-poor subset contained a total of 1036 cases. In the model development stage discussed below, the base dataset for the poor subset contained the measure of inequality experienced by the poor (namely, the 'exposure_of_poor' variable listed in Table 4.2 above), whilst the base dataset for the non-poor subset contained the measure of inequality experienced by the non-poor (namely, the 'exposure_of_rich' variable listed in Table 4.2 above). Apart from the differential specification of 'experience of inequality', the composition of variables in the poor and non-poor base datasets was identical.