Measuring Economic Mobility in Namibia using Repeated Cross-sections

Thomas Ferreira
Department of Economics, Stellenbosch University

Preliminary Draft - Please do not sight

Abstract

Namibia has experienced sharp declines in poverty over the last two decades. Estimates from the World Bank suggest that the poverty rate at the national poverty line declined from around 70% in 1994 to 29% in 2009. Studying economic mobility in the country would provide great insights into understanding the sharp declines in poverty. Traditionally, econometric studies of economic mobility have hinged on the availability of panel data. For Namibia, however, no panel data exists. To overcome this limitation, this paper exploits a method developed by Dang et al. (2014) that uses repeated cross-sections to estimate bounds on the joint probability of staying in or out of poverty, or moving into or out of poverty over two periods. The method produces particularly narrow bounds for Namibia, when compared to other studies, and suggests that from 2003 to 2009, between 21.8% and 23.2% of the population moved out of poverty while between 10.5% and 18.9% of the population remained in poverty.

1 Introduction

Namibia is country with relatively high poverty rates and it is also one of the most unequal countries in the world with a gini coefficient of around 0.6. Using the national poverty line, the world bank estimates that poverty decreased from 69.3% in 1994 to 37.7% 2004 and then to 28.7% in 2010. Levine and Roberts (2013) suggests that the poverty rate was closer to 48.6% in 1994 but in both cases that decreases are large. With such large decreases in poverty over the period it would of great use to identify which groups were economically mobile and in what direction. In the fight against poverty alleviation identifying which groups are vulnerable to poverty and which groups make the transition out of poverty is of great use in designing effective policies.

Levine and Roberts (2013) argues that good data, and data that is more comparable over time will be a key guide for policy makers to achieve development goals in Namibia. Identifying economic mobility has relied mostly on panel data. Panel data allows you to follow the same individuals over time while collecting a wide range of information including income and consumption. Unfortunately for Namibia, no panel datasets exist. This limitation has been an issue in many countries and researchers have attempted at getting around the limitation by using pseudo panels which are created by combining multiple cross-sections over time.

Antman and McKenzie (2007) showed how cohorts in pseudo panels can be used to estimate mobility by estimating the effect of lagged income/consumption on current income/consumption in the presence of measurement error. Bourguignon et al. (2004) showed how it is possible to estimate the extent to which individuals are vulnerable to poverty. Both of these studies though require the use of at least three waves of cross-sectional data. With three waves estimates will be imprecise and ideally a large number of comparable cross-sections exist. For Namibia there are only 3 income and expenditure surveys. Dang et al. (2014) showed how to estimate bounds on the proportion of the population moving in and out of poverty using only two cross-sections of data. This is achieved by making assumptions on how the unobservable characteristics of individuals correlate over time. If for example it is assumed that errors are perfectly correlated over time, then a lower-bound to economic mobility will be estimated. Similarly, in the errors are independent over time, an upper bound to mobility will be measured.

This study estimates the extent of movements into and out of poverty in Namibia in the absence of panel data using the method proposed by Dang et al. (2014). The study starts with a brief overview of Namibia, and its history which has led to the levels of inequality and poverty witnessed in the country. This is followed by a description of the methodology developed by Dang et al. (2014) and the data used. The empirical results are then presented followed by concluding remarks.

2 Background on Namibia

2.1 Namibia pre and post independence.

After nearly 90 years of colonial rule Namibia gained independence in 1990. The period of colonial rule was violent. After the Herero and Nama wars between 1904 and 1908 Germany gained power in Namibia. Germany ended their rule in World War 1 and following occupation in 1915, South Africa (SA) gained administrative control in 1920 (Bank 2008). In 1966 SA rejected a United Nations (UN) mandate to place the country under a trusteeship arrangement and as a result the UN cancelled SA's mandate to rule the country.

In that same year the South West Africa People's Organization (SWAPO) declared war to free Namibia from colonial rule. SWAPO established bases in southern Angola after Angola gained independence in 1975. The organisation used guerilla tactics and northern Namibia became the centre of a war between South Africa and SWAPO. A UN-sponsored peace deal was finally reached in 1989 when Cuban troops who were supporting SWAPO left Angola and SA troops left Namibia. Elections were held in 1989, which SWAPO won easily. A constitution was adopted in February 1990 and in March independence was granted to the country (Bank 2008).

Under South African rule and just as in SA, apartheid was also enforced in Namibia. Under German administration the country was divided into two areas - the "Police Zone" and the rest. The Police zone was the area of land that was vacated by white farmers and stretched across the north-centre of the country from east to west. Basically areas just south of Namibia's northern border were excluded. In the Police zone the government created Native Reserves or Homelands where many natives were sent after being dispossessed of their land in favour of white farmers. The government did a lot to support white farmers by, for example, granting loans, drilling boreholes, and assisting with drought relief. On the other hand the government did little to support farmers in the homelands. In 1968 more segregation laws were instated with the government establishing 10 homelands and these areas were also granted self-rule. These included areas north of the police zone (Odendaal 2011). Werner (1993) points out that the reserves were essentially created to provide labour to the colonial economy. These policies led to a highly unequal society in Namibia.

2.2 Current trends in socio-economic factors

Levine and Roberts (2013) studied changes in poverty and inequality in the 1990's in Namibia using the National Household Income and Expenditure surveys (NHIES) for 1993/1994 and 2003/2004. Due to changes in the two surveys direct comparisons were not possible. To account for this the authors applied techniques originally designed for small-area estimation by Elbers et al. (2003). The method allows researchers to impute income/consumption for datasets that do not have such information. This is done by estimating models of income/consumption in one dataset, and using the model parameters to then predict income in another dataset with the same explanatory variables, while adjusting the standard errors. Using this Levine and Roberts (2013) modelled income using the more reliable NHIES 2003/2004 and then imputed it for the 1993/1994 survey. They found that headcount poverty decreased from 48.6% to 37.7% over the period. Around two-thirds of the change could be attributable to growth in income over the period. Inequality though, remained virtually unchanged. The gini-coefficient for 1993/1994 was 0.628 and for 2003/2004 it was 0.6. The decrease over time was not statistically significant though. The proportion of inequality attributable to income also increased over the period.

Levine and Roberts (2013) also decomposed their results by geographic location and language and found that the effects of segregation policies during pre-independence still linger. Indeed, Levine (2007) estimated the Human Development Index for Namibia and found great inequalities in the HDI along lines of language and region. The three highest ranked regions - Karas, Erongo, and Khomas - were all part of the police zone before independence while the three lowest ranked - Ohangwena, Kavango, and Caprivi - were all north of the police zone.

3 Pseudo-panel techniques and economic mobility

Methods to study economic mobility have generally relied on the availability of panel data. Using panel data, individual earnings is followed over time to gain insights into the extent of mobility and the characteristics of individuals on different mobility paths. In many countries though, panel data is not available.

Fields (2005) notes that income mobility studies can be done on a macro or micro level and within macro mobility (which is the focus of his paper) there are six distinct concepts of mobility. Micro mobility studies are concerned with studying the correlates and determinants of mobility at the individual level. Macro mobility studies are concerned with aggregate levels of mobility in a country, and questions related to that. He further shows that even on the same dataset, depending on what measure was used, mobility could be increasing or decreasing [1].

There are three methods that have been developed to measure economic mobility or certain aspects of it in using repeated cross-sections of data: Bourguignon et al. (2004), Antman and McKenzie (2007), and Dang et al. (2014) (DLLM). Fields and Viollaz (2013) distinguish between the three approaches by the assumptions made about the structural parameters and the mobility question that the method attempts to answer. Along these characteristics they label Antman and McKenzie (2007) as a mean-based approach, and the other two as dispersion based approaches. Mean-based approaches follow cohorts of individuals over time from repeated cross-sectional data while dispersion-based approaches construct estimates of mobility from the variance of the error term.

The first two methods construct pseudo panels that follow cohorts over time - a idea first proposed by Deaton (1985)[2]. The last method uses prediction methods that follow on the small-area estimation method of Elbers et al. (2003).

Deaton (1985) suggested that while in repeated cross-sections, specific individuals cannot be followed over time, cohorts, defined as groups with fixed membership, can be followed. Deaton argued that repeated surveys should generate repeated random samples of cohorts if the samples or cohorts are large enough. From these repeated random samples it is then possible to create a time series from the summary statistics for the cohorts which can be used to estimate certain economic relationships for a cohort.

The DLLM method follows on from the small-area estimation methods developed by Elbers et al. (2003). The method was originally developed to estimate poverty at a very small geographic area. Census data contains representative samples at very small geographic levels, but usually lacks income data. Household surveys on the other hand usually contain accurate income data but samples are only representative at larger geographic regions. Elbers et al. (2003) used household surveys to create a prediction model for income based on variables in both surveys, and then predicted consumption for households in the census data. This paper uses the DLLM method to estimate mobility. To motivate it's use and as a general review of the literature the other two methods will be described next. The DLLM method is discussed under the methodology section.

[1] The six distinct concepts are: **time dependence** - how past well-being determines present well-being-, **positional movement** - how have people's economic position changed with regards to any number of measures such as ranks or deciles -, **share movement -** how people's share of total income has changed -, **income flux -** how large have people's changes in income been regardless of the direction-, **directional income movement -** how many people have moved (or not moved) in any given direction, and **mobility as an equalizer of longer-term incomes** - how has income inequality changed over time.

[2] For a detailed description of these two methods please refer to the appendix.

4 Data and Methodology

The data used for this exercise are come from the three available Namibia Household Income and Expenditure Surveys (NHIES) conducted in 1993/94, 2003/04, and, 2009/10. The purpose of the surveys were to gain accurate information on income, expenditure, and consumption in the country and the surveys serve as the only valid source from to calculate poverty estimates. All three surveys used a stratified two-stage sampling design. For the 2003/04 and 2009/10 surveys, the questionnaires were kept the same to make them as comparable as possible. The 1993/94 surveys differed from the later surveys in a number of important ways. It was half the sample size, later surveys used modern technology for data capturing, infrequent non-food expenditures were captured worse when compared to the later surveys, and finally, to be considered as a household member you had to be residing the house for at least 1 week in the past month, compared to two weeks in the later surveys (Levine and Roberts 2013)[3]. The sample sizes for the three surveys were 4 752 (1993/94), 10 920 (2003/04), and 9 801 (2009/2010) households.

Given that there are only three Income and Expenditure surveys for Namibia, the use of both methods discussed above becomes problematic as they will both give imprecise estimates. The DLLM method on the other hand requires only two survey periods. As noted earlier the method is influenced by the small area estimation method developed by Elbers et al. (2003) and relies on predicting what level of consumption individuals would have had in the period that they are not observed for. By making certain assumptions about the joint distribution of the error terms bounds can be obtained on transitions into and out of poverty.

Firstly, model individuals' consumption in both periods using the same set of variables. For individual i in period $t = 1$ and $t = 2$:

$$y_{it} = \beta x_{it} + e_{it}, \tag{1}$$

where y_{it} is the income for individual i in period t. Assume then that there is some poverty line z_t for period t. We are interested in the joint probability of being either poor or non-poor in the first period and also the in the second. In the case of measuring the extent to which individuals transition out of poverty between two periods you need to estimate the probability of individuals being in poverty in the first period but not in the second:

$$P(y_{i1} < z_1 \, and \, y_{i2} > z_2). \tag{2}$$

In repeated cross-sections though we do not observe the same individuals in both periods which makes estimating the probability impossible. The probability though, can be rewritten as:

$$P(e_{i1} < z_1 - \beta_{i1} \, and \, e_{i2} > z_2 - \beta_{i2}). \tag{3}$$

This shows that the probability depends on the joint distribution of e_{i1} and e_{i2} which captures the autocorrelation of the parts of consumption that are not explained by the observable characteristics in those periods. Intuitively, the less correlated e_{i1} and e_{i2} was, the more mobility there was. By making assumptions on the joint distribution of the error, bounds can be obtained on mobility. At the extremes,

3 There was an explicit attempt at improving this in the later surveys. There were only seven categories for fees and loans and seven commodities for this section in the 1993/94 questionnaire. This was expanded greatly in the later surveys (Levine and Roberts 2013).

one can assume that the errors are either perfectly correlated or not correlated at all. This will give the non-parametric upper- and lower-bound estimates of mobility.

Dang et al. (2014) state two assumptions that must be satisfied for their method to hold. The first is the that underlying population sampled is the same in both surveys. This can be verified by comparing time invariant characteristics of a cohort over time. The second assumption is that the errors are positive quadrant dependent (PQD) which means that the correlation between the error terms cannot be negative. They provide three reasons why the assumption is reasonable. Firstly, income or consumption shocks are expected to be fairly persistent over time. Secondly, poverty is fairly persistent over time, which implies that the joint probability of being in poverty in both periods is expected to be higher than the product of the individual probabilities of being poor in both periods. Finally, the authors note that while for some individuals the correlation between errors would be negative, for the majority it is expected to be positive. To satisfy these assumptions, they note that it is best to limit the sample to an age cohort that is not too young or too old, such as ages 25 - 55 which is a standard practise in pseudo-panel analysis. From these assumptions they propose the following theories to obtain the bounds on mobility.They also show that the estimates calculated in equations 4-11 are robust to classical measurement error and also to non-classical measurement error if the PQD assumption is not violated[4].

Independent errors

Firstly, the upper-bounds on mobility, and also the lower bounds on immobility, are obtained when corr(e_{i1},e_{i2}) = 0.

The upper-bound estimate of the probability of being poor in period 1 and not poor in period 2 is

$$P(y_{i1}^{2u} < z_1 \, and \, y_{i2} > z_2) = P(e_{i1} < z_1 - \beta_1' x_{i2}) P(e_{i2} > z_2 - \beta_2' x_{i2}), \tag{4}$$

where y_{i1}^{2u}refers to the estimated upper-bound consumption of period 2 households in period 1 which is $y_{i1}^{2u} = \beta_1' x_{i2} + e_{i1}$.

The upper-bound estimate of the probability of not being poor in period 1 but being poor in period 2 is given by

$$P(y_{i1}^{2u} > z_1 \, and \, y_{i2} < z_2) = P(e_{i1} > z_1 - \beta_1' x_{i2}) P(e_{i2} < z_2 - \beta_2' x_{i2}). \tag{5}$$

The lower-bound estimate of the probability of not being poor in both periods is

$$P(y_{i1}^{2u} > z_1 \, and \, y_{i2} > z_2) = P(y_{i2} > z_2) - P(y_{i1}^{2u} < z_1 \, and \, y_{i2} > z_2). \tag{6}$$

and, the lower-bound estimate of the probability of being poor in both periods is

$$P(y_{i1}^{2u} < z_1 \, and \, y_{i2} < z_2) = P(y_{i2} < z_2) - P(y_{i1}^{2u} > z_1 \, and \, y_{i2} < z_2). \tag{7}$$

Completely dependant errors

Secondly, when corr(e_{i1},e_{i2}) = 1 the lower-bound estimates of mobility can be obtained and similarly the upper-bound estimates of immobility.

[4] Refer to appendix 1 of Dang et al. (2014) for the proof.

The lower-bound estimate of the probability of being poor in period 1 and not poor in period 2 is

$$P(y_{i1}^{2l} < z_1\, and\, y_{i2} > z_2) = P(\eta e_{i2} < z_1 - \beta_1' x_{i2}) - P(e_{i2} \leq z_2 - \beta_2' x_{i2}), \tag{8}$$

where $y_{i2}^{2l}$is the estimated lower-bound consumption of period 2 households for period 1 which is $y_{i1}^{2l} = \beta_1' x_{i2} + \eta e_{i2}$ and $\eta = \sqrt{\frac{Var(e_{i1})}{Var(e_{i2})}}$. In this instance it is not necessary to repeat all the calculations 500 times as no random sampling takes place.

The lower-bound estimate of the probability of not being poor in period 1 but being poor in period 2 is given by

$$P(y_{i1}^{2l} > z_1\, and\, y_{i2} < z_2) = P(e_{i2} < z_2 - \beta_2' x_{i2}) - P(\eta e_{i2} \leq z_1 - \beta_1' x_{i2}). \tag{9}$$

The upper-bound estimate of the probability of not being poor in both periods is

$$P(y_{i1}^{2l} > z_1\, and\, y_{i2} > z_2) = P(y_{i2} > z_2) - P(y_{i1}^{2l} < z_1\, and\, y_{i2} > z_2). \tag{10}$$

and, the upper-bound estimate of the probability of being poor in both periods is

$$P(y_{i1}^{2l} < z_1\, and\, y_{i2} < z_2) = P(y_{i2} < z_2) - P(y_{i1}^{2l} > z_1\, and\, y_{i2} < z_2). \tag{11}$$

Parametric Approach

Dang et al. (2014) further show that it is possible to obtain thinner bounds, parametrically by assuming that the errors follow a statistical distribution such as a bivariate normal distribution with a correlation coefficient ρ. Using true estimates of ρ from from countries comparable to the country being studied, one can obtain sharper estimates. Assuming that ρ is equal to one and zero will give the parametric equivalents of the non-parametric estimates above.

Assuming that $e_1\, and\, e_2$ have a bivariate normal distribution with standard deviations $\sigma_{e_1}\, and\, \sigma_{e_2}$and a correlation coefficient $\rho \geq 0$, then the probability that households are poor in period 1 and not poor in period 2 is:

$$\begin{aligned} P(y_{i1} < z_1\, and\, y_{i2} > z_2) = &\quad P(e_{i1} < z_1 - \beta_1' x_{i2}) P(e_{i2} > z_2 - \beta_2' x_{i2}) \\ = &\quad \Phi(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho), \end{aligned} \tag{12}$$

where $\Phi(.)$ refers to the bivariate normal cumulative distribution function.

Using the bounds approach it is assumed that the true ρ lies somewhere between an upper and lower bound value $\rho_u\, and\, \rho_l$where $0 < \rho_l < \rho_u < 1$. A lower value of ρimplies a higher level of mobility. Thus the lower and upper bound estimates of the joint probabilities the different poverty statuses are:

$$P(y_{i1} < z_1\, and\, y_{i2} < z_2) = \Phi(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \tag{13}$$

$$P(y_{i1} < z_1\, and\, y_{i2} > z_2) = \Phi(\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, -\frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \tag{14}$$

$$P(y_{i1} > z_1 \, and \, y_{i2} < z_2) = \Phi(-\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, \frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \quad (15)$$

$$P(y_{i1} > z_1 \, and \, y_{i2} > z_2) = \Phi(-\frac{z_1 - \beta_1' x_{i2}}{\sigma_{e_1}}, -\frac{z_2 - \beta_2' x_{i2}}{\sigma_{e_2}}, -\rho) \quad (16)$$

where$\rho = \rho_u$for the upper bound estimates and $\rho = \rho_l$for the lower bound estimates.

Models and Estimation Procedure

To obtain the non-parametric bounds of mobility consumption is modelled for both periods 1 and 2 using the same set of variables and the model specified in equation 1. From the period 1 model the vectors β_1 and $e_1$are obtained. y^{2u} is then calculated using the observable characteristics of households in period 2 with the variable coefficients of period 1 (β_1) and a bootstrapped sample (with replacement) of the errors (e_{i1}). To obtain consistent estimates of equations 4-7, this process is repeated 500 times and the averages of the estimates are taken. y^{2l} is calculated in the same manner but using the true observed errors for individuals in period 2, e_{i2}, which scaled by $\eta = \sqrt{\frac{Var(e_{i1})}{Var(e_{i2})}}$.

Income is measured at the household level and per adult equivalent income is used in the analysis . This is used because Namibia's poverty line is set at per adult equivalent levels. The log of per adult equivalent household income is used as the dependant variable. The sample is limited to household heads. The choice of independent variables is limited because time invariant characteristics need to be used [5]. In the NHIES surveys there are not many variables that satisfy this. To obtain the estimates different variable combinations are used, starting with the most plausible time invariant variables and then adding variables that do not necessarily satisfy the constraint.

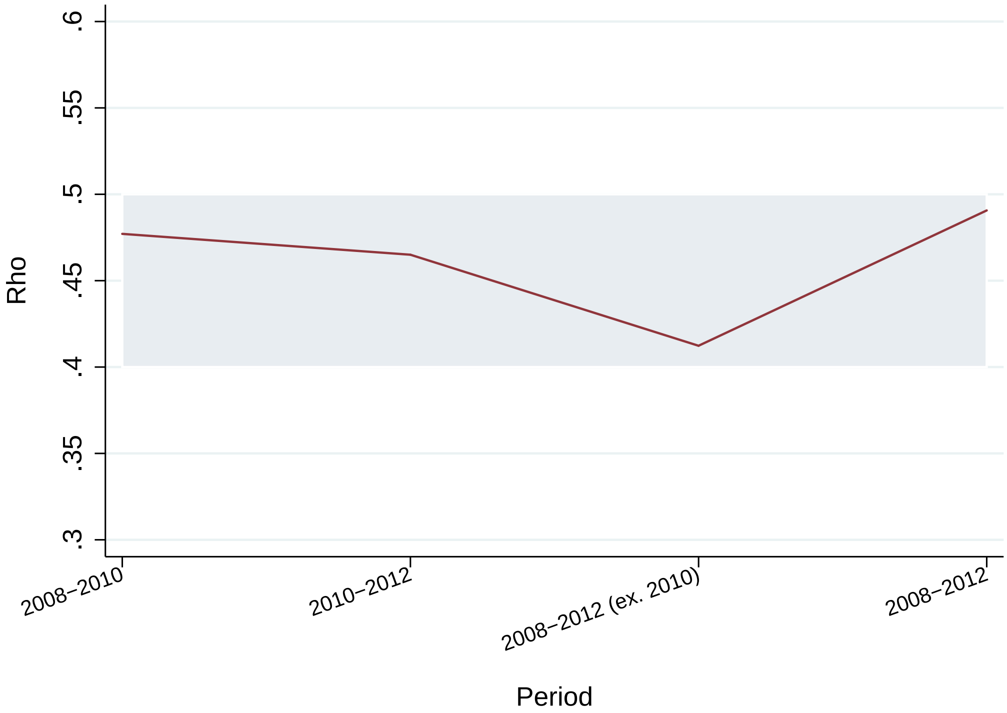
For the parametric estimates, the estimation procedure also involves prediction income in period 1 for households in period 2. Once this done it is easy to calculate the joint probability of poverty status using the CDF for a bivariate normal distribution with correlation ρ. Consumption is modelled using OLS the controls for race, gender, age, education, ownership of a car, and, region and rural.

To get an approximation forρ panel data from South Africa is used. South Africa is similar to Namibia in a number of ways. Both were under apartheid rule that created highly unequal societies. In the early 1990's both also abolished apartheid and held democratic elections. The National Income Dynamics Study (NIDS) data is used. The first round of NIDS was conducted in 2008 with subsequent waves in 2010 and 2012. All three waves are used to estimate the correlation of the error term over time. Different time permutations are also used to get consistent estimates of ρ. It is calculated between 2008 and 2010, 2010 and 2012, 2008 and 2012, and, 2008, 2010 and 2012. Regression models for NIDS that were used to estimate ρ can be found in the appendix. Figure shows that estimates of Rho all the estimates lie between 0.41 and 0.49. As a conservation measure the upper and lower bounds of Rho are taken to be 0.35 and 0.5.

For the first model gender, home language, years of education and age at the time of the first survey are used. In the second specification locational variables are added. This is problematic because there could have been a lot of migration, specifically to Windhoek in the Khomas region. All models and results are estimated on the household level using the where applicable, the characteristics of the household

[5] This is needed because $y_{i1}^{2u} = \beta_1' x_{i2} + e_{i1}$.

Figure 1: Estimates of ρ for South Africa from NIDS



Source: Own calculations.

head. The samples are also limited households where the head was between the ages of 25 and 55at the time of the first survey.

5 Results

To start, the OLS models that are used to predict income are shown in tables 1 and 2. Given that these are prediction models is it rather important to focus on the R^2. The prediction values of the model are high given that such few controls are added. For 1993/94 the most basic model (Model 1) has a $R^2$of 0.38 while for 2003/04 it is 0.42 or 0.43. The difference for 2003/2004 between tables 1 and 2 stem from the fact that different estimation samples are used given that the samples are limited to individuals who were between the ages of 25 and 55 at the time of the first survey. For 2009/10 Model has a $R^2$of 0.48. When regional indicators are added to the models the R^2 increases in all cases and the increases are large 2003/2004. For 1993/94 the R^2 increases to 0.44 while for the other years it increases to 0.51.

Tables 3 give the estimation results for the estimates to mobility using the non-parametric assumptions around the errors. Starting with mobility between 1993/94 and 2003/04 the bounds on the joint probability are fairly wide which makes discussion around what happened less fruitful. The bounds narrow between model 1 and 2 which is to be expected but they remain wide. What is suggested by the results is that between 20 and 35% of the population remained poor of the period. Between 2 and 21% moved out of poverty over the period and similarly, between 2 and 18% of the population moved into

Table 1: Model of consumption - 1993/93 - 2003/04

	1993/1994		2003/2004	
	Model 1	Model 2	Model 1	Model 2
Primary	0.15^{**}	0.10	0.43^{***}	0.28^{***}
	(0.06)	(0.06)	(0.10)	(0.09)
Secondary	0.56^{***}	0.40^{***}	1.44^{***}	1.15^{***}
	(0.07)	(0.07)	(0.11)	(0.10)
Tertiary	1.25^{***}	1.07^{***}	2.86^{***}	2.58^{***}
	(0.10)	(0.09)	(0.13)	(0.13)
Female	-0.22^{***}	-0.11^{**}	-0.50^{***}	-0.42^{***}
	(0.05)	(0.04)	(0.09)	(0.08)
Age	-0.02	-0.03	0.03	0.03
	(0.04)	(0.04)	(0.07)	(0.06)
Age^2	0.00	0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Controls				
Language	X	X	X	X
Region		X		X
Adj. R^2	0.38	0.44	0.43	0.51
Num. obs.	2125	2125	2131	2131

Source: Own calculations . Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

Table 2: Model of consumption - 2003/04 - 2009/10

	2003/2004		2009/2010	
	Model 1	Model 2	Model 1	Model 2
Primary	0.52^{***}	0.32^{***}	0.30^{***}	0.24^{***}
	(0.08)	(0.07)	(0.04)	(0.04)
Secondary	1.40^{***}	1.01^{***}	0.77^{***}	0.66^{***}
	(0.08)	(0.08)	(0.04)	(0.04)
Tertiary	3.06^{***}	2.61^{***}	1.66^{***}	1.51^{***}
	(0.10)	(0.10)	(0.05)	(0.05)
Female	-0.53^{***}	-0.44^{***}	-0.28^{***}	-0.24^{***}
	(0.06)	(0.05)	(0.03)	(0.03)
Age	0.01	0.02	0.01	0.01
	(0.04)	(0.03)	(0.02)	(0.02)
Age^2	-0.00	-0.00	-0.00	-0.00
	(0.00)	(0.00)	(0.00)	(0.00)
Language	X	X	X	X
Region		X		X
Adj. R^2	0.42	0.51	0.48	0.51
Num. obs.	3627	3627	5474	5474

Source: Own calculations . Significance levels: * p<0.1, ** p<0.05, *** p<0.01.

poverty. between 41 and 58% of the population remain above the poverty line. The bounds on the conditional probabilities are very wide. What can be said though is that there was a more than 50% chance that those who were poor in 1993/94 were also poor in 2003/04. Similarly there was a high chance of staying out of poverty.

For the period 2003/04-2009/10 the bounds are much tighter, especially on the joint probability of moving out of poverty and staying out of poverty. Using model 1 estimates suggest that between 19.65 and 22.85% of the population moved out of poverty over the period. Using model 2 this estimate is higher between 23.4 and 24.49% moving out of poverty. Around 60% of the population were not poor in both periods while between 10 and 20% of population were poor in both periods. When comparing the results from the two different periods, it seems as if upward mobility increased in the last decade while downwards mobility decreased.

Table 3: Non-parametric bounds (Joint Probability)

		1993/94 - 2003/04		**2003/04 - 2009/10**	
		Lower	**Upper**	**Lower**	**Upper**
Period1	**Period2**		**Model 1**		
Poor	Poor	36.72	19.68	18.90	9.00
Poor	Not Poor	2.50	21.13	19.65	22.85
Not Poor	Poor	2.10	17.70	0.02	10.57
Not Poor	Not Poor	58.67	41.49	61.43	57.58
Period1	**Period2**		**Model 2**		
Poor	Poor	34.19	21.33	18.88	11.14
Poor	Not Poor	3.01	18.55	23.40	24.49
Not Poor	Poor	4.63	16.05	0.04	8.42
Not Poor	Not Poor	58.16	44.07	57.68	55.95

Source: Own calculations.

Table 4 shows the results from the parametric estimation. The first set of results are the parametric equivalents of the non-parametric results while the second set are whit the estimates of ρ obtained from South Africa. Comparing the two sets shows how much the bounds can be tightened if reliable estimates of ρcan be obtained. The estimates assuming ρ=1 and ρ=0 are close to the estimates from the non-parametric estimates. The bounds on the proportion who moved out of poverty and also stayed out of poverty have widened though.

Using estimates of ρ from South Africa produces very tight bounds. The results from Model 1 suggest that between 12 and 14% of the population were chronically poor in 2009/10. between 58 and 60% of households were not poor in both periods. Between 21 and 23% of the population moved out of poverty while between 4 and 6% of household in Namibia moved in to poverty. The results are again encouraging and suggest that there was substantial upward mobility over the period while there was little vulnerability to poverty.

Table 4: Parametric bounds (Joint Probability)

		$\rho = 1$ **and** $\rho = 0$		$\rho = 0.5$ **and** $\rho = 0.35$	
		Lower	**Upper**	**Lower**	**Upper**
Period1	**Period2**			**Model 1**	
Poor	Poor	18.87	9.57	14.04	12.61
Poor	Not Poor	16.54	25.84	21.38	22.80
Not Poor	Poor	0.00	9.30	4.84	6.26
Not Poor	Not Poor	64.59	55.29	59.75	58.33
Period1	**Period2**			**Model 2**	
Poor	Poor	18.92	11.75	15.38	14.24
Poor	Not Poor	19.41	26.58	22.95	24.09
Not Poor	Poor	0.00	7.17	3.54	4.68
Not Poor	Not Poor	61.67	54.50	58.13	56.99

Source: Own calculations .

6 Conclusion

This paper has mad a first attempt at estimating the extent of economic mobility in Namibia. A lack of panel data and also only a small number of cross-sectional surveys have made it difficult in the past to estimate aspects of economic mobility. A new method developed by that only needs two cross-sectional surveys has given opportunity to explore mobility in Namibia for the first time. It is question that deserves to be studied as Namibia, like South Africa, gained independence and held democratic elections for the first time in the early 1990's. Identifying what has happened since the transitions can give vital insights into the success of government since independence.

The method proposed by Dang et al. (2014) estimates bounds to joint probability of poverty statuses over two periods by making assumptions on how the unobservable components of consumption are correlated over time. At the one extreme the errors can be completely dependant which intuitively gives a lower bound to economic mobility. Assuming that errors are independent over time gives an upper bound to mobility. With these distributional assumptions about the errors income for households in one period can be predicted for the other period.

The method produced surprisingly tight bounds for Namibia even when using the most conservative prediction models. The results suggest that there has been substantial upward mobility in Namibia and that upward mobility has increased substantially since the 1990's. Vulnerability to poverty has also decreased.

The results are encouraging, especially in a country with such high inequality. The next step of the analysis involves identying specifically who the groups are that are chronically poor, upwardly mobile, and vulnerable, and then understanding why. In the absence of panel data this is a challenge but using similar methods on geographically smaller areas is a start.

References

Antman, F. and McKenzie, D. J.: 2007, Earnings Mobility and Measurement Error: A Pseudo-Panel Approach, *Economic Development and Cultural Change* **56**, 125–161.

Bank, W.: 2008, Republic of Namibia - Addressing Binding Constraints to Stimulate Broad Based Growth : A Country Economic Report, *World Bank Other Operational Studies 12601*, The World Bank.

Bourguignon, F., Goh, C.-c. and Kim, D. I.: 2004, Estimating individual vulnerability to poverty with pseudo-panel data, *Policy Research Working Paper Series 3375*, The World Bank.
URL: *http://ideas.repec.org/p/wbk/wbrwps/3375.html*

Dang, H.-A., Lanjouw, P., Luoto, J. and McKenzie, D.: 2014, Using Repeated Cross-Sections to Explore Movements into and Out of Poverty., *Journal of Development Economics* **107**, 112 – 128.

Deaton, A.: 1985, Panel data from time series of cross-sections, *Journal of Econometrics* **30**(1-2), 109–126.
URL: *http://ideas.repec.org/a/eee/econom/v30y1985i1-2p109-126.html*

Elbers, C., Lanjouw, J. O. and Lanjouw, P.: 2003, Micro-level Estimation of Poverty and Inequality., *Econometrica* **71**(1), 355 – 364.

Fields, G. S.: 2005, The Many Facets of Economic Mobility, Retrieved 21-07-2015, from Cornell University, ILR School site: http://digitalcommons.ilr.cornell.edu/articles/230/.

Fields, G. and Viollaz, M.: 2013, Can the Limitations of Panel Datasets be Overcome by using PPseudo-Panel to Estimate Income Mobility.

Levine, S.: 2007, Trends in Human Development and Human Poverty in Namibia, *Background paper to the namibia human development report.*, United Nations Development Program.

Levine, S. and Roberts, B.: 2013, Robust Estimates of Changes in Poverty and Inequality in Post-Independence Namibia., *South African Journal of Economics* **81**(2), 167 – 191.

Odendaal, W.: 2011, Land Grabbing in Namibia: A case study from the Omusati region, northern Namibia, *Paper presented at the international conference on global land grabbing. 6-8 april 2011*, Land Deals Politics Initiative, University of Sussex.

Werner, W.: 1993, A brief history of land disposession in Namibia., *Journal of Southern African Studies* **19**(1), 135.

A Alternative Pseudo-panel methods to Study Mobility

Antman and McKenzie (2007)

Antman and McKenzie (2007) show how pseudo-panels can be used to measure income mobility in the absence of panel data and also deal with situations where income is measured with error. They argue that aggregating individual level data to a certain level averages out individual measurement error, while the fact that households are only observed in one time period implies the measurement error observed in different periods will be for different households.

Their mobility measure of interest is the coefficient of lagged income on income in a regression model:

$$Y_{it}^{*} \quad = \quad \alpha + \beta Y_{it-1}^{*} + \mu_{it}, \tag{17}$$

where Y_{it}^{*} is income for individual i in period t, and β is coefficient of interest. A value of β greater than 1 indicates income divergence while a value less than 1 indicates income convergence in the economy. In their model income is measured with error though. Instead of observing true income, Y^{*}, we observe Y_{it} which is equal to:

$$Y_{it} = Y_{it}^{*} + \varepsilon_{it} \tag{18}$$

Substituting 18 into 17 leads to the following equation:

$$Y_{it} = \alpha + \beta Y_{it-1} + \mu_{it} + \varepsilon_{it} - \beta \varepsilon_{it-1}. \tag{19}$$

As is the case in the absence of panel data we do not observe individual i in periods t and $t-1$. To deal with this problem the authors start by rewriting equation 19 in terms of cohort averages:

$$\bar{Y}_{c(t)t} = \alpha + \beta \bar{Y}_{c(t)t-1} + \bar{\mu}_{c(t)t} + \bar{\varepsilon}_{c(t)t} - \beta \bar{\varepsilon}_{c(t)t-1}, \tag{20}$$

where, for example, $\bar{Y}_{c(t)t}$ is the average income for all individuals belonging to cohort c observed in time t for time t [6]. $\bar{Y}_{c(t)t-1}$is the average income for cohort c, observed in time t, for period $t-1$. In the absence of panel data this is also not observed. To deal with this the authors replace $\bar{Y}_{c(t)t-1}$with $\bar{Y}_{c(t-1)t-1}$which leads to:

$$\bar{Y}_{c(t)t} = \alpha + \beta \bar{Y}_{c(t-1)t-1} + \bar{\mu}_{c(t)t} + \bar{\varepsilon}_{c(t)t} - \beta \bar{\varepsilon}_{c(t)t-1} + \lambda_{c(t)t}, \tag{21}$$

where, $\lambda_{c(t)t} = \beta(\bar{Y}_{c(t)t-1} - \bar{Y}_{c(t-1)t-1})$.

To get to an estimable equation, Antman and McKenzie (2007) assume that the law of large numbers holds. This implies that as $n_c \to \infty$: $\frac{1}{n_c}\sum_{i=1}^{n_c} \varepsilon_{it} \to 0$ and $\lambda_{c(t)t} = \beta(\bar{Y}_{c(t)t-1} - \bar{Y}_{c(t-1)t-1}) \to 0$. Thus 21 becomes the estimable equation:

$$\bar{Y}_{c(t)t} = \alpha + \beta \bar{Y}_{c(t-1)t-1} + \bar{\mu}_{c(t)t} \tag{22}$$

Bourguignon et al. (2004)

The authors follow on work by Deaton and Paxson (1994) who studied the second-order moments of cohorts in a pseudo panel context. Bourguignon et al. (2004) simplify their idea by stating:

> *" if it may be assumed that all individuals within a cohort face a stochastic earning process that has common characteristics, these characteristics may be recovered at the aggregate level, without observing actual earning paths."*

[6] $\bar{Y}_{c(t)t} = \frac{1}{n_c}\sum_{i=1}^{n_c} Y_{i(t),t}$

Their model assumes earnings of individual i from cohort j at time t follows the following equation:

$$y_{it}^{j} = \beta_t^j x_{it}^j + e_{it}^j, \tag{23}$$

where y is the log of earnings, x is a group of individual characteristics, and e is the residual term contains two components of earnings: the transitory component, and, the unobserved component of permanent earnings.

Assume that e_{it}^j follows an autoregressive process *AR(1)*:

$$e_{it}^j = \rho^j e_{it-1}^j + \varepsilon_{it}^j \tag{24}$$

where,ρ is a parameter measuring the persistence of earnings shocks, and,ε_{it}^j is, what the authors call, the innovation in earnings which has a variance $\sigma_{\varepsilon jt}^2$.

In the absence of panel data it is not possible to observe e_{it}^j and e_{it-1}^j which implies that the above model cannot be estimated. They argue though, that under the assumption that that individuals enter and exit the labour market at random between 2 periods it is still possible to estimate $\rho^j$and $\sigma_{\varepsilon jt}^2$ by estimating the following equation:

$$\sigma_{ejt}^2 = \rho^{j2}\sigma_{ejt-1}^2 + \sigma_{\varepsilon jt}^2. \tag{25}$$

The above equation can easily be estimated if a minimum of three time periods are available. Three periods though will not lead to very precise estimates. If it is further assumed that the innovation term, , has a mean of zero and variance $\sigma_{\varepsilon jt}^2$ and that future predictions of variables, coefficients (apart from the constant), and, the variance in innovation are available, it is possible to estimate an individual's vulnerability to poverty[7]. The probability of an individual in period t falling below a poverty line, Z_{t+1}, in period $t+1$ is:

$$P(y_{it+1}^j < Z_{t+1} | x_{it}^j, x_{it+1}^j, \hat{\beta}_{t+1}^j, \hat{\sigma}_{\varepsilon jt+1}^2) = \Phi(\frac{Z_{t+1} - \hat{X}_{it+1}^j \hat{\beta}_{t+1}^j - \hat{\rho}^j e_{it}^j}{\hat{\sigma}_{\varepsilon jt}^2}). \tag{26}$$

[7] Variables such as age and gender are easy to predict in the future. If time variant characteristics want to be used, stationarity needs to be assumed .(Bourguignon et al. 2004, : 5)

B Regression models

Table 5: Models for the estimation of ρ from NIDS

	2008-2010		2010-2012	
	2008	2010	2010	2012
Race(Ref: Black)				
Coloured	−0.016	0.057	0.081	0.073
Asian/Indian	0.467***	0.178	0.437**	0.360*
White	0.656***	0.660***	0.736***	0.750***
Female	−0.623***	−0.615***	−0.687***	−0.676***
Age	0.023	−0.003	−0.024	0.030
Age^2	−0.000	0.000	0.000	−0.000
Education	−0.057***	−0.044***	−0.067***	−0.078***
Education^2	0.011***	0.010***	0.011***	0.012***
Car	0.652***	0.696***	0.759***	0.492***
Rural	−0.269***	−0.330***	−0.246***	−0.209***
Province	*Yes*	*Yes*	*Yes*	*Yes*
Observations	2525	2395	2171	2315
R^2	0.512	0.494	0.483	0.497
ρ	0.477		0.465	

	2008-2012		2008-2010-2012		
	2008	2012	2008	2010	2012
Race(Ref: Black)					
Coloured	0.229**	0.103	0.209*	0.154	0.016
Asian/Indian	0.395**	0.390***	0.257	0.257	0.069
White	0.857***	0.671***	0.827***	0.745***	0.705***
Female	−0.608***	−0.533***	−0.709***	−0.690***	−0.643***
Age	0.039	0.025	0.052*	0.002	0.047
Age^2	−0.000	−0.000	−0.001	0.000	−0.000
Education	−0.072***	−0.055***	−0.080***	−0.081***	−0.080***
Education^2	0.011***	0.011***	0.012***	0.012***	0.013***
Car	0.623***	0.539***	0.612***	0.636***	0.481***
Rural	−0.325***	−0.212***	−0.312***	−0.362***	−0.210***
Province	*Yes*	*Yes*	*Yes*	*Yes*	*Yes*
Observations	2293	2311	1725	1629	1737
R^2	0.494	0.498	0.486	0.484	0.502
ρ	0.412			0.491	

Source: Own calculations . Significance levels: * p<0.1, ** p<0.05, *** p<0.01.