Economic inequality and HIV in South Africa

Nicklas Nordfors

Abstract

The aim of this thesis is to study the association between economic inequality and HIV in South Africa, since relatively recent research has suggested that economic inequality may be a structural driver in the HIV epidemic, especially in sub-Saharan Africa. Economic inequality is measured at the provincial and municipal level using an asset index to approximate household wealth. The asset index is constructed through multiple correspondence analysis using variables on household assets and amenities. Economic inequality is calculated using the Gini coefficient. The effects of inequality are estimated with a multilevel mixed-effects logistic regression model. When both men and women are included in the sample, municipal inequality proves to be insignificant while an increase in household wealth increases the odds of being HIV positive. Disaggregating the analysis by sex yields different results: for males, the effect is negligible. For females, higher levels of municipal inequality increases the odds of being HIV positive, and women living in the most unequal municipality is 4.383 times as likely to be HIV positive than women in the least unequal municipality are. Economic household wealth is also a significant predictor for HIV serostatus, where a higher level of wealth is associated with higher odds of being HIV positive. The data used in this thesis is from 2005 and is supplied by the Human Sciences Research Council.

Supervisor: Professor Dick Durevall
Local supervisor: Professor Margaret Chitiga-Mabugu
Bachelor Thesis in Economics
Department of Economics with Statistics
University of Gothenburg

Acknowledgements

The writing of this thesis was supported by a Minor Field Study stipend provided by the Swedish International Development Cooperation Agency.

I would like to thank Professor Margaret Chitiga for her assistance and her colleagues at the Human Sciences Research Council in Pretoria: Dr Khangelani Zuma for his help and explanations of the data set and Dr Njeri Wabiri for taking the time to see me and pointing out possible flaws. I would also like to thank Professor Ingrid Woolard at the University of Cape Town for her helpful suggestions and criticism. Finally, I would like to thank Professor Dick Durevall at the University of Gothenburg for his guidance and patience.

Contents

1 Introduction

A common perception amongst people is that the HIV/AIDS epidemic is closely related to poverty — e.g. (United Nations. Population Division, 2005) states that, "[p]overty and HIV are interrelated". While it is true that the rich or industrialized countries in general do have a low HIV prevalence, the correlation is not clear cut at a national level: many lower income countries have low prevalence rates, while middle income countries have high prevalence rates. Some of the countries with the highest HIV prevalence in Sub-Saharan Africa are relatively rich (United Nations. Population Division, 2005), one example being South Africa, an upper middle-income country.

In 2005, the HIV prevalence in South Africa was estimated to be 10.8 %, where the prevalence by sex was 8.2 % for males and 13.3 % for females. It should be noted that the prevalence rates are higher for certain subsets of the population, where e.g. Africans have a prevalence rate of 13.3 % and Whites only 0.9 %. South Africa is also a country of vast income and wealth disparities. In 2005, the Gini coefficient for South Africa was 0.72 (Bhorat et al., 2009). For comparison, in the late 2000's Sweden had a Gini coefficient of 0.26 and the United States 0.38. (OECD, 2013)

While research on HIV has often focused on poverty as an important driver in the HIV epidemic (United Nations. Population Division, 2005; Durevall and Lindskog, 2012), researchers have also started to study economic inequality and its association with the spread of HIV. This is also the purpose of this thesis: to study whether economic inequality has an impact on HIV in South Africa.

The impact of economic inequality, on both the provincial and municipal level, has on the probability of being HIV positive will be studied and estimated. The original hypothesis is that economic inequality does increase the probability of being HIV-positive, an hypothesis that is based on the results of several studies on economic inequality and HIV which have indicated that a higher economic inequality increases the probability of being HIV-positive, e.g. Durevall and Lindskog (2012); Fox (2012); Brodish (2014). Durevall and Lindskog (2012) used data from Malawi, while Fox (2012); Brodish (2014) are cross-country studies comparing economic inequality and HIV prevalence between different countries. All three studies make us of Demographic Health Survey data for Africa, and come to the conclusion that a higher degree of economic inequality is associated with a higher HIV prevalence.

To my knowledge, this thesis is the first to investigate the association

between economic inequality and HIV in South Africa. Hopefully, it can be an addition to the literature of economic inequality and HIV, as well as a study of possible structural factors contributing to the spread of HIV in South Africa.

The data set used for this thesis is supplied by the Human Science Research Council in South Africa. The data was collected through a survey on HIV and other related information; impact of different campaigns, attitudes but also personal and household information. The data covers all 9 provinces and a large share of the municipalities of South Africa and dates back to 2005, which is the most recent data of its kind in South Africa available.

To estimate the possible impact inequality has on the probability of being HIV-positive, a measure of wealth is needed. Since the data set lacks information on income and consumption expenditure, the wealth measure is created by constructing an asset index using multiple correspondence analysis, using information collected on household amenities and assets. The wealth or asset index is then used to calculate a measure of inequality, on both provincial and municipal levels. A multilevel mixed-effects logistic regression model is then used to estimate the effect of inequality on the probability of being HIV positive.

The results show that for the whole sample, economic inequality does not have an effect on the provincial nor on the municipal level. Household wealth or economic status proves to be a significant predictor, where the odds ratio of being HIV-positive increases when household economic wealth increases (economic status). When the regression analysis is disaggregated by sex, the results are different. For women, an increase in municipal inequality is associated with a higher odds ratio of being HIV positive: women living in the most unequal municipality is 4.383 times as likely to be HIV positive than women in the least unequal municipality are. For males, the effect of municipal inequality is the opposite: where the odds ratio decreases when municipal inequality increases. This effect is small however: the difference between the most and least unequal municipality is small, and the economic significance can be questioned.

2 Literature review

It is well known that HIV/AIDS has a tremendous effect on several levels of society, but the connection between economics and the HIV epidemic is

not self-evident. Apart from the physical effects of the HIV infection, there are also consequences to the economy, economic performance and personal finances. A survey on the household level in South Africa, showed that out of 771 households surveyed, the average age of the individual sick with AIDS was 35 years old, i.e. likely to be a vital income provider as well as a parent. 64 % of the AIDS-sick individuals surveyed were women (Steinberg, 2002). In a survey of households in the urban township Soweto in South Africa, Naidu and Harris (2006) finds that households are affected in various ways: some examples being burdensome funeral costs, taking care of a relative's orphaned children, devoting time caring for the sick, decline in productivity and increases in health care expenditure.

While most evidence has pointed to the fact that households and individuals are economically hit by higher morbidity and mortality (e.g. Naidu and Harris (2006)), the effect on the economy in a macroeconomic sense is more debated with inconclusive evidence and results. Although, as argued by Haacker in (Haacker et al., 2004, Chapter 2), an analysis of only aggregated economic variables has the consequence of missing microeconomic effects, it can still be worthwhile exercise in order to paint a broader picture.

In a cross-country analysis of 51 different countries, both developed and industrialized, Bloom and Mahal (1997) estimates the effect of the AIDS epidemic on economic growth and finds no evidence that the epidemic had an adverse effect on the economic growth of countries. Since Bloom and Mahal (1997), e.g. Bell et al. (2006) has argued that the AIDS epidemic will in fact peak before causing the most economic damage, because the prevalence rates for people aged 15-49 are high. Since AIDS kills mostly young adults, the human capital of these young adults will also be lost, which can lead to a chain reaction affecting future generations. Children finding themselves without parents will be disadvantaged in terms of human capital accumulation, thus affecting the economy negatively. Bell et al. (2006) apply their model on South Africa and find that had there not been an HIV-epidemic, there would have been a possibility of growth in per capita income.

McDonald and Roberts (2006) reaches the conclusion that the level of HIV/AIDS prevalence has reached the point where the effect on both human capital as well as infant mortality is large enough that it has started to affect incomes, i.e. noting that the HIV/AIDS epidemic has reached the point where it has an effect on economic interactions. According to Haacker and Crafts in (Haacker et al., 2004, Chapter 6), most of the direct effects on welfare caused by the epidemic is related to mortality. The increased mortality,

i.e. loss in life expectancy in the countries hardest hit by HIV/AIDS have erased the health gains that have happened over the last century (Haacker et al., 2004, Chapter 6). Haacker and Crafts also attempt to quantify the effect that HIV/AIDS has on welfare. For South Africa, they estimate that the increased mortality carries a welfare loss of around 80 % of national GDP.

While the behavioural factors behind the infection of HIV are well documented and frequently discussed, there is also evidence pointing to structural factors being an important factor in explaining the HIV/AIDS-epidemic as well as the importance of structural approaches to HIV prevention (Gupta et al., 2008). HIV/AIDS is a fact of life across the whole world, but the spread of the disease has been different in different parts of the world (Barnett et al., 2000). Barnett et al. (2000) argue that the profile of an epidemic in a society is largely defined by two variables: the degree of social cohesion and the overall level of wealth.

Seeley et al. (2012) also argue that there is a need to adress the social or structural drivers behind the HIV epidemic. The authors defines a social or structural driver as "core social processes and arrangements, reflective of social and cultural norms, values, networks, structures and institutions, that operate in concert with individuals behaviours and practices to influence HIV epidemics in particular settings". Examples of these social processes or arrangements is gender inequality, poverty, economic inequality and many more. These factors can influence the decision making behind using a condom, transactional sex and intergenerational sex, which in turn affect the probability of contracting HIV (Seeley et al., 2012).

Poverty is one of the structural factors that has been extensively studied in relation to the spread of HIV/AIDS. Despite the common perception of HIV as a disease of poverty, Gillespie et al. (2007) argues that this is an incorrect assumption. While acknowledging that poor households and individuals are harder hit once exposed to HIV, they are not more likely of becoming HIV-positive compared to wealthier households and individuals. Education does however seem to be a protective factor against HIV (Gillespie et al., 2007).

Other studies on the effect of poverty on the spread of HIV tell a similar story: a cross-country study on eight African countries showed that in all but two of the countries (Ghana and Lesotho), the HIV prevalence is in fact higher amongst the wealthiest 20% compared to the poorest 20% (Mishra et al., 2007). Nattrass (2009) draws similar conclusions, suggesting that there is no systematic relationship between poverty and HIV, while noting that in some cases malnutrition as well as economic vulnerability could increase the

likelihood of HIV infection (Nattrass, 2009).

Studying the urban poor using DHS data from 20 countries in Sub-Saharan Africa, Magadi (2013) examines the association between HIV and urban poverty. The author finds that the probability of being HIV-positive is higher for the urban poor compared to non-poor with similar characteristics. The study also reveals differences between the urban and rural population: in urban areas the poor have a higher probability of being HIV-positive compared to the non-poor, whereas in rural areas the poor have a lower probability of being HIV-positive compared to the non-poor (Magadi, 2013).

Contrary to the results and conclusions drawn in previously mentioned studies, Lopman et al. (2007), in a study on HIV prevalence in the province Manicaland in Zimbabwe, finds that poverty is a determinant in HIV infection. HIV is associated with poverty for males, where HIV prevalence is higher in the lowest wealth tercile compared to the highest wealth tercile. The same association was not prevalent for females. It should also be mentioned that most of the previously mentioned studies are cross-sectional studies on poverty and HIV/AIDS on the national level; Fox (2012) argues that this can contribute to the masking of intra-country differences, and to omitted variable bias (Brodish, 2014). Wabiri and Taffa (2013) finds that a low socio-economic index score is associated with a higher probability of being HIV-positive, while the inverse applies for those classified as middle or high on the index scores. Education proved to be a protective factor, as the probability of being HIV-positive decreased with increases in educational level.

While many studies cannot find evidence supporting the view of poverty as an important structural factor with regards to the HIV epidemic, economic inequality has been suggested to be an important structural factor. The association between inequality and HIV has also been confirmed by recent studies, e.g. Fox (2012). In a cross-sectional study of 16 African countries, using Demographic and Health Survey (DHS) data, Fox finds that the risk of HIV infection was higher in regions with a higher degree of wealth inequality. This association remained true even when taking individual wealth into account. A difference between rural and urban areas with regards to the effect of wealth on HIV infection was also found; in urban areas HIV prevalence was lower when wealth was higher. In rural areas however, HIV was higher with higher wealth (Fox, 2012). In a study on Malawi (using DHS data), Durevall and Lindskog (2012) find that among women aged 15-24 there is a strong relationship between regional economic inequality and a higher risk

of HIV infection. Meanwhile, they find differing effects regarding poverty; although a higher median wealth at both household and neighborhood level seems to lead to a higher risk of HIV infection, the results are not statistically significant (Durevall and Lindskog, 2012).

Brodish (2014) studied the association between economic inequality and HIV in six Sub-Saharan countries, each with a HIV prevalence rate of more than 5 %. The paper uses, similarly to Durevall and Lindskog (2012) and Fox (2012), a multilevel regression framework to evaluate the the probability of being HIV-positive dependent on cluster-level inequality. Brodish (2014) finds that an increase in economic inequality is associated with a higher probability of being HIV-positive. Likewise, an increase in wealth is also associated with a higher probability of being HIV-positive.

The hypothesis of this thesis is that an association between economic inequality and HIV does exist, based on the results of Fox (2012); Durevall and Lindskog (2012); Brodish (2014). However, the association between inequality and health is not self-evident — Leigh and Jencks (2007) argues that these mechanisms could be explained by three different hypotheses: the absolute income hypothesis, the relative income hypothesis and the society-wide effects hypothesis. The absolute income hypothesis states that the gains in health should diminish as individual income rises. Leigh and Jencks (2007) argue that while this is theoretically valid, empirical evidence is inconclusive. The relative income hypothesis posits that others' income can have an effect on a person's health, and not only one's own income. According to Wilkinson (1997), economic inequality can increase stress if an individual's well-being is partly determined by comparing themselves to others — Leigh and Jencks (2007) add that several studies show these types of comparisons to be more stressful when they involve people who have a lot in common with the subject, and differences between the two are relatively small.

Society-wide effects is the hypothesis that inequality has an effect on crime, public spending and social capital and trust. Empirically, a relationship between violent crime and economic inequality has been shown in both rich and poor countries, which could also affect stress due to an increased concern of falling victim. As for public spending, it is suggested that an increased heterogeneity in preferences could lead to lower government spending on health (Leigh and Jencks, 2007).

Economic inequality has been though to affect the HIV epidemic through reduced social cohesion, where countries with a high social cohesion and high levels of wealth tend to have low infection rates, whereas countries

with both a low degree of social cohesion and low levels of wealth will be experience a more severe HIV epidemic. For countries with a high degree of social cohesion but low levels of wealth, the epidemic is thought to be slow-growing (Barnett et al., 2000). Barnett et al. (2000) does not explain exactly why social cohesion has this effect, but other studies have suggested that inequality may e.g. increase sexual concurrency. Sexual concurrency can be caused by differences in wealth, where women may be relatively poor and men rich, which can lead to transactional sex. Transactional sex can be attributed to women being either poor enough to need money for survival or wanting a better lifestyle, which can lead to taking on "sugar-daddies" (**?**Fox, 2012). Studies have also pointed to economic migration as a factor (**?**), since it can lead to individuals taking on informal partners both when moving or staying behind.

3 Data

The data used for this thesis is produced and distributed courtesy of Human Sciences Research Council of Pretoria, South Africa. The survey that is the foundation for this data is the South African National HIV Prevalence, HIV Incidence, Behaviour and Communication Survey, 2005. The data is cross-sectional, covering all nine provinces of South Africa. It also records four different geotypes: urban informal, urban formal, rural formal and rural informal areas (also called tribal areas). There are three data sets derived from the aforementioned survey: individual data on adults and youth, individual data on children and household data. For this thesis, only data on adults and youth as well as household data will be used.

A youth is defined as being between 15 and 24 years of age and an adult as 24+. Individuals classified as either youth or adult are 15 years of age or older. The adult and youth subset contains 16 398 observations, and 12 032 out of these were tested for HIV (Shisana, 2005, p. 33-34). All observations in the data set are identified by a municipal area code and an enumerator area code. The individual data contains information on HIV-status, age, educational level and other similar variables. The household data contains information on the size of the household, language spoken and information on household assets and amenities (Shisana, 2005). The household data and individual data was merged, but all individuals could not be matched to a household leading to a decrease in the number of observations, as can be seen

in Section 5.

Table 1: HIV prevalence for South Africa by sex and geotype and descriptive statistics of inequality and wealth variables.

Race	African	White	Coloured	Indian
Female	18.3 %	1.5 %	3.7 %	1.2 %
Male	13.7 %	2.2 %	2.7 %	2.3 %
Total	16.6 %	1.8 %	3.4 %	1.6 %
Geotype	Urban formal	U. informal	Rural formal	R. informal*
Female	8.3 %	23.7 %	12.8 %	16.5 %
Male	6.7 %	19.8 %	9.5 %	9.5 %
Total	7.7 %	22.3 %	11.2 %	14.3 %
	Mean	S.d.	Min	Max
Province inequality**	.6297	.0833	.5136	.8101
Municipal inequality**	.6299	.1268	.2159	.8636
Household wealth***	1.7475	.9968	0	4.7767

* Also referred to as tribal area ** Gini coefficients. *** Also referred to as economic status.

3.1 Sampling

The sampling frame used for the survey is based on a master sample constructed by Statistics South Africa. The master sample consisted of 1000 different enumerator areas (EA). An EA is a spatial area consisting of approximately 180 households in urban areas and 80-120 households in rural areas (Shisana, 2005, p. 9).

Testing for HIV serostatus was done by collecting DBS specimens and then sent to the Global Clinical Viral Laboratory in Durban, South Africa for processing. The HIV incidence testing was then performed at the National Institute for Communicable Diseases in Johannesburg, South Africa. Of the specimens collected, 100% of blots received in the laboratory were suitable for testing (Shisana, 2005, p. 17-18).

Out of 24 236 individuals that were considered eligible, 15 851 or 65.4 % agreed to be both interviewed and tested for HIV. It should be noted that some segments of the population refused at a higher rate, e.g. Whites and

Indians[1]. From the samples collected, 11.23% of the respondents that agreed to be tested were HIV positive, while 88.77% were negative (Shisana, 2005, p. 23-34). Further information on sampling methodology and response rates for different segments can be found in Shisana (2005).

4 Method

As previously stated, the aim of this thesis is to estimate a possible association between HIV and economic inequality in South Africa. A measure of inequality is created from an asset index, since there are none available in the data set or from other sources.

4.1 Variables

In the literature on economic inequality and HIV, a number of inequality measures are used. The Gini coefficient is common choice (Durevall and Lindskog, 2012; Brodish, 2014; Fox, 2012). Durevall and Lindskog (2012) also uses a measure of the distance between the 90th and 10th percentile of households in terms of scores of the created wealth index and Brodish (2014) uses an index based on the ratio of the mean wealth of the top 20% quintile and mean wealth of the bottom 20% quintile. McKenzie (2005) proposes an alternative measure of inequality, where the standard deviation of the first principal component in e.g. a community is divided by the standard deviation of the whole sample.

Measuring economic inequality, regardless on what level, requires a measure of economic status. Economic status (for a household) is commonly measured by three primary indicators (Rutstein and Johnson, 2004): household income, household consumption expenditure and/or household wealth.

While household income is the first choice for many economists, it is often difficult to measure accurately. Many people do not their income very well, especially in developing countries where self-employment and home production is important as well as a lack of tax infrastructure. Additional problems include respondents (both rich and poor) trying to hide income from

[1]The reasons behind this are not clear, but in a discussion with Prof Ingrid Woolard, she argued it could be due to e.g. Whites or Indians not feeling they are a part of society in South Africa, indicating a low degree of social cohesion.

interviewers, a temporal variability in income and a variation in income between members of the household which they might keep secret (Rutstein and Johnson, 2004). There are also difficulties in how to value home production (Rutstein and Johnson, 2004; Bollen et al., 2002). Only reporting the household income can also be problematic due to difficulties in valuing home production, capital gains and illicit income (Rutstein and Johnson, 2004).

Due to the problems and difficulties of obtaining correct information on household income, household expenditure can be chosen as a proxy for income (Rutstein and Johnson, 2004), where it can serve as a measure of household welfare in the short- and long-run (Filmer and Pritchett, 2001). Using consumption expenditures as a proxy can also be problematic, since there is usually only one respondent for each household who does not necessarily know the expenditure of other household members. Additionally, the time period and goods and services that should be measured are not evident. Both income and consumption expenditure are also problematic measures due to their volatility, since income in developing countries can be random and seasonal (Rutstein and Johnson, 2004).

Household wealth has been introduced as a viable alternative to both income and consumption expenditure as a measure of economic status; it is both easier to measure (e.g. since only one respondent is needed) and requires fewer questions to be accurately measured compared to collecting data on income or consumption expenditure (Bollen et al., 2002). It is also lauded as a better indicator of permanent economic status than the earlier mentioned measures (Rutstein and Johnson, 2004). Information on assets and amenities can then be used to construct indices, which are often called asset or wealth indices. These indices can then be used as an approximation for the long-run economic status of a household (Rutstein and Johnson, 2004; Filmer and Pritchett, 2001). The information on the types of variables that could be used for an asset index will vary, but common variables include floor type in the house, water source and type of housing (Rutstein and Johnson, 2004).

The simplest asset indices are simply constructed by using the sum of the chosen variables possessed by households, which means giving all variables equal weight in the index. It is however unlikely that all of the variables included in the construction of the index is of equal importance, which could prove to be problematic (Wai-Poi et al., 2008). One possible remedy to this problem is weighting the index using prices for each variable — although this also creates the problem of obtaining accurate price information (Wai-

Poi et al., 2008; Filmer and Pritchett, 2001). In order to avoid the aforementioned problems, data reduction methods are used, e.g. principal components analysis (PCA), to calculate the weights of the index using variables on asset ownership of households (Filmer and Pritchett, 2001). An alternative to PCA is multiple correspondence analysis (MCA) or factor analysis (Traissac and Martin-Prevel, 2012; Wabiri and Taffa, 2013). It is argued that MCA is the better choice over PCA when analyzing categorical variables (Booysen et al., 2008), since MCA was designed for the analysis of categorical variables while PCA was designed for the analysis of continuous variables (Booysen et al., 2008; Howe et al., 2012; Burger et al., 2006).

Both PCA and MCA derived index scores will contain both negative and positive values (Fox, 2012), which means that some inequality measures, e.g. the Gini coefficient, cannot be calculated (McKenzie, 2005; Sahn and Stifel, 2003). To circumvent this problem, the index scores can be transformed, most commonly through additive transformation (i.e. adding value so all values are greater than zero) (Fox, 2012). However, this transformation can have redistributive effects on the distribution (Sahn and Stifel, 2003). Instead, Fox (2012) uses an exponential transformation — by taking the exponential of a value instead of using the additive method, the distribution is less affected (Wai-Poi et al., 2008).

Wealth or asset indices have been criticized for being too urban-focused in its choice of variables as well as its inability to distinguish between different levels of poverty. Possible solutions include collecting information that would better reflect rural wealth and creating separate indices for urban and rural areas (Rutstein, 2008). The first option is moot for this thesis since the information is already collected. Using separate indices for rural and urban areas canals complicate the analysis, since they are not necessarily comparable (Rutstein, 2008).

Because there are a limited number of variables to use in the construction, and for comparability, a single asset index is calculated using all of the available variables on household assets and amenities are used. These variables include: source of drinking water, access to electricity, type of toilet facility, source of energy for cooking and if there is a working refrigerator, radio, TV and land line telephone in the household. Since the variables used in constructing my wealth indices either were discrete from the beginning, or recoded to be discrete, I use an MCA approach in calculating the scores for the reasons mentioned by Wai-Poi et al. (2008); Traissac and Martin-Prevel (2012); Booysen et al. (2008); Burger et al. (2006). As described earlier, the

index is then transformed using an exponential transformation of the index scores, in order to calculate Gini coefficients. The distribution of the index scores, before and after the exponential transformation, can be seen in Figure A.1, where is it also possible to see that the distribution has not changed significantly.

The scores are then used to calculate a Gini coefficient as a measure of economic inequality, on both a provincial level and municipal level. The Gini coefficients are based on the asset or wealth indices calculated for each province and municipality and therefore, the Gini coefficient are measures of the inequality within provinces and municipalities.

4.2 Empirical analysis

The dependent variable of interest is HIV, a categorical binary variable measuring HIV serostatus, where 1 indicates an HIV positive individual and 0 indicates HIV negative. It follows that since the dependent variable is categorical, it is of interest to estimate the probability of $HIV = 1$.

The data used for this thesis is nested or multileveled. Since sampling was done according to geographical location; respondents belong to a municipality, each municipality is a part of a larger province which in turn is a part of South Africa. The observations, which are on an individual level, are therefore nested within larger "unit", i.e. municipalities, provinces and state. Following the empirical framework of Durevall and Lindskog (2012); Fox (2012); Brodish (2014) a multilevel logistic regression model is used to estimate the impact of inequality on the odds of an individual being HIV positive. Multilevel modeling allows you to estimate how much each level of the data contributes to the model, and can account for the heterogeneity between groups (Primo et al., 2007).

The three-level mixed-effects logistic regression model can be formulated as follows, where individuals i are nested in a municipality j which is in turn nested in a province k (Rabe-Hesketh and Skrondal, 2008):

$$logit\{P(HIV_{ijk} = 1|x_{ijk}, \varsigma_{jk}^{(2)}, \varsigma_{k}^{(3)})\} = \beta_1 + \beta_2 x_{2ijk} + ... + \beta_{11} x_{11,k} + \varsigma_{jk}^{(2)} + \varsigma_{k}^{(3)} \tag{1}$$

where $\varsigma_{jk}^{(2)}$ is a random intercept which varies over the municipalities, i.e. level 2, and $\varsigma_{k}^{(3)}$ is a random intercept varying over provinces. x_{ijk} is vector that contains the independent variables. Independent variables with the

indexing ijk are individual level variables, jk municipal level and k province level.

The province and municipal level variables will include inequality calculated for both levels. The variable of most interest is the municipal inequality variable since it is the lowest level of inequality variable. The control variables on the individual level will include race, individual/household wealth, civil status, geotype, age and the education level which is a recoding of information on the highest educational qualification achieved, where low education represents no schooling up to grade 6-7, mid education from grade 8-9 to grade 11. Finally, high education represents grade 12 to a postgraduate degree and each category represents around a third of the sample.

A correlation matrix of the independent variables is presented in Table A.2. Looking at Table A.2, there are two cases of possible collinearity: for provincial and municipal wealth. These variables measure the average wealth in each province and municipality respectively, but since province wealth / provincial inequality and municipal wealth / municipal inequality are highly correlated, the wealth variables are omitted. Testing for heteroskedasticity issues in logistic regression models can be complicated, which is the rationale behind using robust standard errors when estimating the regressions.

Previous research has included regression analysis on sub-samples of the data set used, to find out whether there are differences between groups, e.g. Durevall and Lindskog (2012) on 15-24 year-old women, Fox (2012) on rural and urban sub-samples and Brodish (2014) on males and females. The same is done in this thesis and the results are presented in the appendix, although the regression analysis on rural and urban and 15-24 year-olds probably suffer from lack of power due to small sample size.

5 Results

In this section, the results of the aforementioned regression analysis is presented. Table 2 contains five different specifications of the estimates of the three-level mixed-effects logistic regression model for the whole sample, i.e. both males and females. The first model contains only municipal and provincial inequality, and the last model is the full specification with an interaction term between $female\ municipalinequality$. The inequality measures are highly significant, as can be seen in the first model. However, when household wealth (economic status) is introduced, they turn insignificant and remains

so even after adding all independent variables. It is only when introducing the interaction term that municipal inequality, i.e. the variable of interest, is statistically significant, suggesting that there is a difference between men and women with respect to inequality.

Therefore, a disaggregate analysis by sex is performed and presented in Table 3. It should be noted that the number of observations for females is 5 683 for women and only 3 583, which could affect the results. In this table, we can more clearly see the effects that municipal inequality has. As suggested by specification five of Table 2, there is evidence of a difference in effects. Municipal inequality for women is a significant predictor of HIV serostatus (OR = 3.263) and when inequality increases, the probability of a woman being HIV positive increases. For men, inequality is also a significant predictor (OR = 0.0899) and has a negative effect. The economical significance for men can be questioned: if municipal inequality increased by 0.65 (which is the difference between the least and most unequal municipalities), the odds would only decrease by 0.058. For women, the difference would mean that a woman in the most unequal municipality is 2.1 times as likely to be HIV positive.

Household wealth (economic status) also increases the odds for females (OR = 1.292), i.e. should household wealth increase by 1 point, the odds of being HIV positive is 1.292 times higher. For men, household wealth is statistically insignificant at a 5 % significance level, although significant at a 10 % significance level. It should be noted that should we accept a significance level of 10 %, it would mean that household wealth increases the odds ratio for men (OR = 1.132) of being HIV positive. The effect for women with regards to household wealth is consistent with some studies, e.g. Mishra et al. (2007); Nattrass (2009); Gillespie et al. (2007) (although these analyses are not disaggregated), and the effect for men is consistent with e.g. Fox (2012) who did not find household wealth to have a significant effect (again not a disaggregated analysis).

For the other independent variables, the two analyses have much in common, although some variables are statistically significant for the one but no the other. A high education is protective compared to both low and mid education, and the odds of being HIV positive, compared to ages 15-19, is higher for all age categories except for 60 and above. As an individual grows older, they will probably have more sexual encounters, which could explain this effect. The lower odds of individuals aged 60+ could be explained by mortality among HIV positive individuals or perhaps less sexual activity.

Compared to Africans (Black), the odds for being HIV positive is lower for Coloured, White and Indian individuals. Being married is associated with lower odds compared to those who have never been married (singles).

There differences between different types of areas (geotypes) is noticeable. Compared to urban formal residency, the odds of being HIV positive are higher for urban informal residency, while the odds are lower for both rural formal and rural informal. This effect is not statistically significant for the separate analyses of Table 3.

Additional specifications using both OLS and logistic regression, performed as an extra robustness check with and without clustered standard errors, can be found in Table 4. In the same table, the reader will also find regressions on subsets of the sample. Per Durevall and Lindskog (2012), an analysis on only 15-24 year-olds, who are likely to be recently infected, is presented. Per Fox (2012), separate analyses for urban and rural areas are presented as well. The reason for not including these in the main text is the small sample size, which could bias the estimates. For all logistic regressions, the odds ratios are presented.

Figure 1 shows the predicted probability of HIV infection with regards to economic status, divided into males and females. As the graph shows, both males and females have an increased risk of being HIV positive as household wealth (economic status) increases.

Figure 1: Predicted probability of household wealth, fixed and random effects.

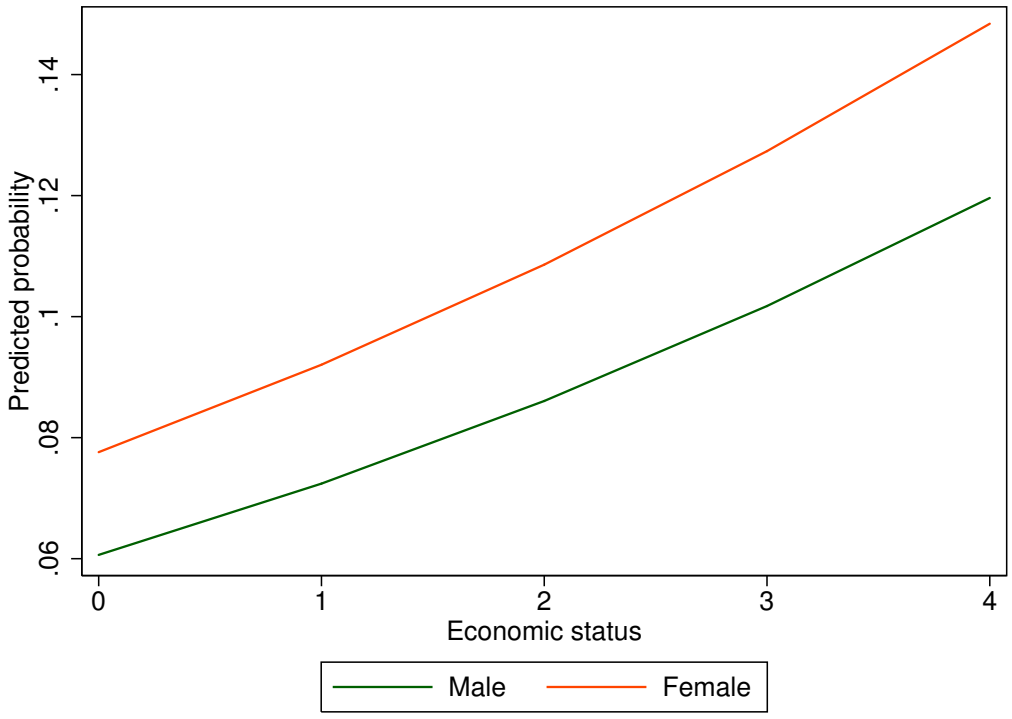


Table 2: Three-level mixed-effects logistic regression model.

	(1)	(2)	(3)	(4)	(5)
Independent variables	HIV	HIV	HIV	HIV	HIV
Province level					
Provincial inequality	0.00504*	0.00952*	0.0107*	0.0330*	0.0330*
	(-2.53)	(-2.28)	(-2.18)	(-2.09)	(-2.11)
Municipal level					
Municipal inequality	0.318*	1.427	1.190	0.899	0.262*
	(-2.24)	(0.62)	(0.29)	(-0.26)	(-2.32)
Female × Municipal inequality					7.431***
					(3.72)
Individual level					
Economic status		1.595***	1.607***	1.231***	1.235***
		(13.06)	(12.39)	(5.03)	(5.16)
Female			1.425***	1.337***	0.405**
			(4.86)	(5.42)	(-3.20)
Low education			0.868	1.404*	1.407*
			(-1.42)	(2.13)	(2.20)
Mid education			1.122	1.431***	1.425***
			(1.30)	(4.09)	(4.02)
Age					
20 to 24				3.308***	3.311***
				(7.22)	(7.27)
25 to 29				6.338***	6.406***
				(10.83)	(11.04)
30 to 34				6.164***	6.205***
				(9.75)	(9.76)
35 to 39				5.749***	5.763***
				(6.61)	(6.63)
40 to 44				3.488***	3.503***
				(5.80)	(5.81)
45 to 49				2.286***	2.267***
				(3.70)	(3.69)
50 to 54				1.707*	1.712**
				(2.58)	(2.58)
55 to 59				1.035	1.034
				(0.09)	(0.09)
60 and above				0.593*	0.593*
				(-2.33)	(-2.32)

Table 2 – *Continued from previous page*

	(1)	(2)	(3)	(4)	(5)
Independent variables	HIV	HIV	HIV	HIV	HIV
Race					
White				0.196***	0.197***
				(-7.09)	(-6.97)
Coloured				0.312***	0.315***
				(-7.62)	(-7.54)
Indian				0.0964***	0.0962***
				(-6.49)	(-6.46)
Other				0.232	0.232
				(-1.22)	(-1.21)
Geotype					
Urban informal				1.187*	1.187*
				(2.57)	(2.49)
Rural informal				0.766*	0.775*
				(-2.17)	(-2.07)
Rural formal				0.654**	0.652**
				(-2.78)	(-2.86)
Marital status					
Married				0.689***	0.688***
				(-3.98)	(-4.02)
Widowed				1.317*	1.325**
				(2.54)	(2.62)
Divorced				1.002	1.005
				(0.01)	(0.02)
Other				1.013	1.009
				(0.06)	(0.04)
Constant	5.730	0.578	0.468	0.245	0.503
	(1.32)	(-0.42)	(-0.57)	(-1.13)	(-0.56)
Province					
Constant	0.458**	0.432**	0.440**	1.115	1.118
	(-2.87)	(-2.98)	(-2.91)	(1.56)	(1.58)
Province > Municipality					
Constant	0.438***	0.465***	0.483***	1.107	1.098
	(-5.90)	(-5.15)	(-5.05)	(1.86)	(1.66)
Observations	11784	9441	9319	9279	9279

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3: Three-level mixed-effects logistic regression model.

	Females	Males
Independent variables	HIV	HIV
Province level		
Provincial inequality	0.0309**	0.0263
	(-2.77)	(-1.77)
Municipal level		
Municipal inequality	3.263**	0.0899***
	(2.87)	(-4.64)
Individual level		
Economic status	1.282**	1.132
	(3.24)	(1.79)
Low education	1.393	1.482*
	(1.72)	(2.17)
Mid education	1.462***	1.431*
	(3.96)	(2.29)
Age		
20 to 24	3.295***	3.103*
	(8.23)	(2.48)
25 to 29	6.189***	7.555***
	(12.74)	(4.66)
30 to 34	4.911***	11.65***
	(9.47)	(5.73)
35 to 39	3.865***	14.32***
	(4.87)	(7.18)
40 to 44	2.178***	9.402***
	(3.83)	(5.38)
45 to 49	1.431	6.131***
	(1.44)	(5.61)
50 to 54	0.913	5.949***
	(-0.27)	(6.27)
55 to 59	0.717	2.402
	(-1.19)	(1.08)
60 and above	0.337***	1.868
	(-3.92)	(1.45)
Race		
White	0.193***	0.204***
	(-5.61)	(-7.52)
Coloured	0.296***	0.296***
	(-6.83)	(-7.23)

Table 3 – *Continued from previous page*

	Females	Males
Independent variables	HIV	HIV
Indian	0.0561***	0.192***
	(-4.12)	(-13.10)
Other	0.670	0.670
	(-0.30)	(-0.30)
Geotype		
Urban informal	1.114	1.366*
	(0.89)	(2.16)
Rural informal	0.908	0.611
	(-0.75)	(-1.64)
Rural formal	0.735*	0.547*
	(-2.44)	(-2.03)
Marital status		
Married	0.631**	0.663*
	(-3.05)	(-2.26)
Widowed	1.370*	1.974*
	(1.99)	(1.96)
Divorced	0.923	1.123
	(-0.49)	(0.29)
Other	1.129	0.604
	(0.40)	(-1.29)
Constant	0.186	0.797
	(-1.44)	(-0.15)
Province		
Constant	1.086*	1.132
	(2.03)	(1.38)
Province > Municipality		
Constant	1.017	1.096
	(0.21)	(0.52)
Observations	5683	3583

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

6 Conclusion and discussion

The main finding of this thesis is that municipal inequality is a statistically significant predictor of the probability of being HIV positive for adults and youth in South Africa after controlling for absolute levels of household wealth, but only when the analysis is disaggregated by sex. This result confirms the initial hypothesis, that there is an association between economic inequality and HIV in South Africa, and at least partially confirms the results of previous research. As such, this thesis also provides an indication that men and women are affected in different ways by economic inequality. While men were little affected by higher levels of inequality, for women the odds of being HIV positive increased substantially with higher levels of inequality.

The difference in impact between males and females has not been observed in previous studies on economic inequality and HIV: Durevall and Lindskog (2012) chose a sample of women aged 15-24 and did not include males because of sample size issues, making a comparison between sexes impossible. Fox (2012) does not analyze the samples separately by gender and while Brodish (2014) does perform separate analyses of male and females, there were no large differences between the two groups, especially not concerning the effect of economic inequality.

The reason why females are more negatively affected by economic inequality than men are not self-explanatory. Possible explanations could be an association between economic inequality and riskier sexual behaviour, as found in Durevall and Lindskog (2012). As mentioned in Section 2, economic inequality could lead to more transactional sex, which can both increase the number of sexual partners and lead to an earlier sexual debut for women, which may contribute to increasing the risk of HIV infection in sexual networks (Durevall and Lindskog, 2012; Brodish, 2014). Other reasons, which were not explored in this thesis, could include gender inequality and domestic or sexual violence towards women.

Like municipal inequality, household wealth (economic status) proved to be a significant predictor, but only for women. As economic status increases, the probability of being HIV positive also increases, thus discrediting the perception mentioned in the introduction, that "HIV is a disease of poverty". These results are in line with some of the previous studies on poverty and HIV, while discrediting others. It should however be noted that the sample size of men is significantly smaller than the sample size of women, which makes drawing too many conclusions regarding household wealth a little

tricky.

As Brodish (2014) concludes, since this study analyzes cross-sectional data, it is only possible to draw conclusions regarding one point in time. It is possible that there are issues of reverse causality, where being HIV positive can affect an individual's economic situation negatively. Other possible weaknesses in the analysis include the construction of the asset index where a possible difference of indicators of wealth between rural and urban areas were not taken into account due to limited availability of variables, which is one of the weaknesses brought up with regards to asset indices (Rutstein, 2008; Filmer and Pritchett, 2001). In order to account for these possible differences was made by including a control variable for the genotype, i.e. whether the area in question is rural/urban and formal/informal. Another potential weakness of the index is that some variables may reflect services that are available to a whole community rather than specific households, e.g. the use of electricity often necessitates an infrastructure that is privately provided and the same goes for other variables, such as piped water (Filmer and Pritchett, 2001).

Another weakness is the use of administrative boundaries for regional or neighborhood inequality, where municipalities were used instead of defining neighborhoods through e.g. analyzing geographic data to limit the scope of the thesis. Nonetheless, since the analysis concerned inequality only within the administrative boundaries of provinces and municipalities, some information could have been lost — it is not necessarily the case that the effects of inequality would remain within the official boundaries. Further research could study the differences between rural and urban areas as well as the underlying factors which could explain why females in South Africa are more affected by economic inequality than men seem to be.

References

Barnett, T., A. Whiteside, and J. Decosas (2000). The Jaipur paradigm–a conceptual framework for understanding social susceptibility and vulnerability to HIV. *South African Medical Journal = Suid-Afrikaanse tydskrif vir geneeskunde 90*(11), 1098.

Bell, C., S. Devarajan, and H. Gersbach (2006). The long-run economic costs of AIDS: a model with an application to South Africa. *The World Bank Economic Review 20*(1), 55–89.

Bhorat, H., C. Van der Westhuizen, and T. Jacobs (2009). Income and non-income inequality in post-apartheid south africa: What are the drivers and possible policy interventions? *Available at SSRN 1474271*.

Bloom, D. E. and A. S. Mahal (1997). Does the AIDS epidemic threaten economic growth? *Journal of Econometrics 77*(1), 105–124.

Bollen, K. A., J. L. Glanville, and G. Stecklov (2002). Economic status proxies in studies of fertility in developing countries: Does the measure matter? *Population Studies 56*(1), 81–96.

Booysen, F., S. Van Der Berg, R. Burger, M. Von Maltitz, and G. Du Rand (2008). Using an asset index to assess trends in poverty in seven sub-saharan african countries. *World Development 36*(6), 1113–1130.

Brodish, P. H. (2014). An association between neighbourhood wealth inequality and HIV prevalence in Sub-Saharan Africa. *Journal of Biosocial Science*, 1–18.

Burger, R., F. Booysen, S. v. d. Berg, and M. v. Maltitz (2006). *Marketable wealth in a poor African country: Using an index of consumer durables to investigate wealth accumulation by households in Ghana*. Number 2006/138. Research Paper, UNU-WIDER, United Nations University (UNU).

Durevall, D. and A. Lindskog (2012). Economic inequality and HIV in Malawi. *World Development 40*(7), 1435–1451.

Filmer, D. and L. H. Pritchett (2001). Estimating wealth effects without expenditure data or tears: An application to educational enrollments in states of India. *Demography 38*(1), 115–132.

Fox, A. M. (2012). The HIV–poverty thesis re-examined: Poverty, wealth or inequality as a social determinant of HIV infection in sub-Saharan Africa? *Journal of Biosocial Science 44*(04), 459–480.

Gillespie, S., S. Kadiyala, and R. Greener (2007). Is poverty or wealth driving HIV transmission? *AIDS 21*, S5–S16.

Gupta, G. R., J. O. Parkhurst, J. A. Ogden, P. Aggleton, and A. Mahal (2008). Structural approaches to HIV prevention. *The Lancet 372*(9640), 764–775.

Haacker, M. et al. (2004). *The macroeconomics of HIV/AIDS*. International Monetary Fund.

Howe, L. D., B. Galobardes, A. Matijasevich, D. Gordon, D. Johnston, O. Onwujekwe, R. Patel, E. A. Webb, D. A. Lawlor, and J. R. Hargreaves (2012). Measuring socio-economic position for epidemiological studies in low-and middle-income countries: a methods of measurement in epidemiology paper. *International Journal of Epidemiology*, dys037.

Leigh, A. and C. Jencks (2007). Inequality and mortality: long-run evidence from a panel of countries. *Journal of Health Economics 26*(1), 1–24.

Lopman, B., J. Lewis, C. Nyamukapa, P. Mushati, S. Chandiwana, and S. Gregson (2007). HIV incidence and poverty in Manicaland, Zimbabwe: is HIV becoming a disease of the poor? *AIDS (London, England) 21*(Suppl 7), S57.

Magadi, M. A. (2013). The disproportionate high risk of HIV infection among the urban poor in Sub-Saharan Africa. *AIDS and Behavior 17*(5), 1645–1654.

McDonald, S. and J. Roberts (2006). Aids and economic growth: A human capital approach. *Journal of Development Economics 80*(1), 228–250.

McKenzie, D. J. (2005). Measuring inequality with asset indicators. *Journal of Population Economics 18*(2), 229–260.

Mishra, V., S. Bignami-Van Assche, R. Greener, M. Vaessen, R. Hong, P. D. Ghys, J. T. Boerma, A. Van Assche, S. Khan, and S. Rutstein (2007). HIV infection does not disproportionately affect the poorer in sub-Saharan Africa. *AIDS 21*, S17–S28.

Moulton, B. R. (1990). An illustration of a pitfall in estimating the effects of aggregate variables on micro units. *The Review of Economics and Statistics*, 334–338.

Naidu, V. and G. Harris (2006). The cost of hiv/aids-related morbidity and mortality to households: preliminary estimates for soweto: economics. *South African Journal of Economic and Management Sciences = Suid-Afrikaanse Tydskrif vir Ekonomiese en Bestuurswetenskappe 9*(3), p–384.

Nattrass, N. (2009). Poverty, sex and HIV. *AIDS and Behavior 13*(5), 833–840.

OECD (2013). *Income inequality*. OECD Factbook 2013: Economic, Environmental and Social Statistics. OECD Publishing.

Primo, D. M., M. L. Jacobsmeier, and J. Milyo (2007). Estimating the impact of state policies and institutions with mixed-level data. *State Politics & Policy Quarterly 7*(4), 446–459.

Rabe-Hesketh, S. and A. Skrondal (2008). *Multilevel and longitudinal modeling using Stata*. STATA press.

Rutstein, S. O. (2008). The DHS Wealth Index: Approaches for rural and urban areas. *Calverton Maryland Macro International MEASURE DHS 2008 Oct.*.

Rutstein, S. O. and K. Johnson (2004). The DHS Wealth Index. DHS Comparative Reports no. 6. *Calverton: ORC Macro*.

Sahn, D. E. and D. C. Stifel (2003). Urban–rural inequality in living standards in Africa. *Journal of African Economies 12*(4), 564–597.

Seeley, J., C. H. Watts, S. Kippax, S. Russell, L. Heise, and A. Whiteside (2012). Addressing the structural drivers of hiv: a luxury or necessity for programmes? *Journal of the International AIDS Society 15*(Suppl 1).

Shisana, O. (2005). *South African national HIV prevalence, HIV incidence, behaviour and communication survey, 2005*. HSRC Press.

Steinberg, M. (2002). *Hitting Home: How households cope with the impact of the HIV/AIDS epidemic: A survey of households affected by HIV/AIDS in South Africa*. Henry J. Kaiser Family Foundation.

Traissac, P. and Y. Martin-Prevel (2012). Alternatives to principal components analysis to derive asset-based indices to measure socio-economic position in low-and middle-income countries: the case for multiple correspondence analysis. *International Journal of Epidemiology 41*(4), 1207–1208.

United Nations. Population Division (2005). *Population, development and HIV/AIDS with particular emphasis on poverty: the concise report*, Volume 247. United Nations Publications.

Wabiri, N. and N. Taffa (2013). Socio-economic inequality and HIV in South Africa. *BMC Public Health 13*(1), 1037.

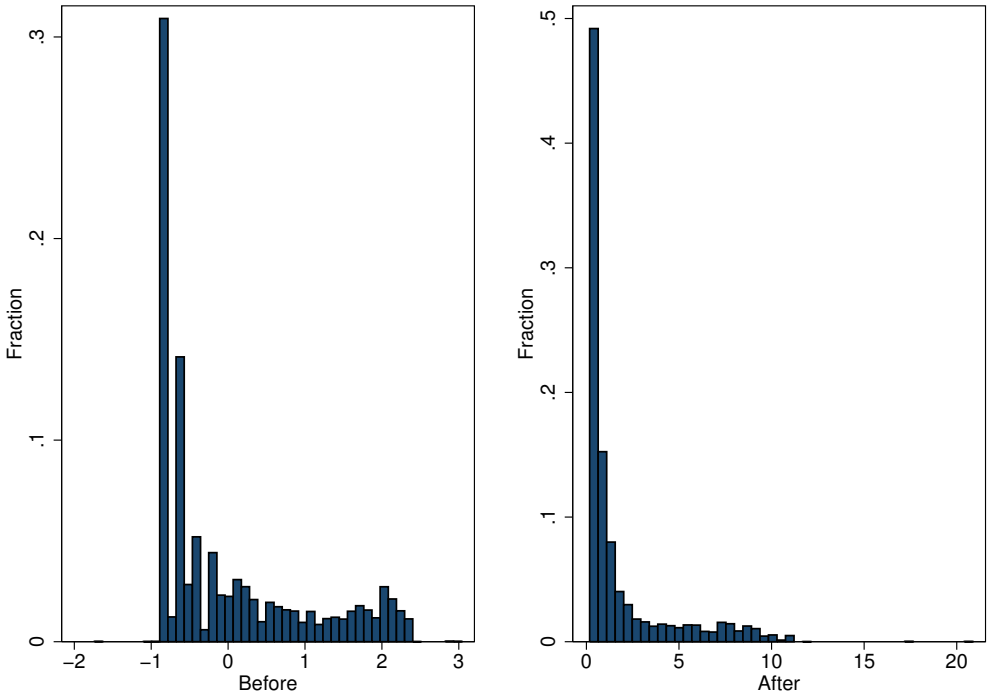
Wai-Poi, M., S. Spilerman, and F. Torche (2008). Economic well-being: concepts and measurement using asset indices. *Working Paper No. 27*.

Wilkinson, R. G. (1997). Socioeconomic determinants of health. health inequalities: relative or absolute material standards? *BMJ: British Medical Journal 314*(7080), 591.

A Additional Graphs and Tables

A.1 Graphs

Figure 2: The distributions of the asset index before and after exponential transformation.



A.2 Tables

Table 4: Regression table showing the results of the robustness checks (OLS, logistic) and further disaggregation of the multilevel analysis.

	OLS	Logistic	Urban	Rural	15-24	Not robust
Independent variables	HIV	HIV	HIV	HIV	HIV	HIV
Province level						
Provincial inequality	-0.208***	0.0231**	0.0254*	0.0726	0.00223	0.0330*
	(-5.82)	(-2.81)	(-2.00)	(-0.51)	(-1.57)	(-2.07)
Municipal level						
Municipal inequality	-0.0341	0.320	0.101**	1.523	0.0605	0.262*
	(-0.81)	(-1.53)	(-2.89)	(0.82)	(-1.17)	(-2.09)
Female × Municipal inequality	0.0927	7.395***	10.22***	5.866	19.73	7.431**
	(1.94)	(3.62)	(5.48)	(1.83)	(1.08)	(3.05)
Individual level						
Economic status	0.0198***	1.202***	1.262	1.271**	1.678***	1.235***
	(3.68)	(4.10)	(1.96)	(3.13)	(3.82)	(4.37)
Female	-0.0358	0.406**	0.283***	0.570	0.509	0.405*
	(-1.12)	(-3.08)	(-4.86)	(-1.06)	(-0.36)	(-2.27)
Low education	0.0239*	1.349	1.577	1.184	0.897	1.407**
	(2.28)	(1.85)	(1.71)	(0.71)	(-0.20)	(3.00)
Mid education	0.0338***	1.395***	1.614***	1.186	0.542*	1.425***
	(4.06)	(3.53)	(7.32)	(0.67)	(-2.22)	(3.72)
Age						
20 to 24	0.0928***	3.220***	3.975***	3.347**		3.311***
	(8.64)	(7.28)	(5.63)	(3.01)		(8.44)
25 to 29	0.185***	6.183***	8.806***	6.722***		6.406***
	(11.46)	(11.24)	(8.98)	(4.52)		(12.22)
30 to 34	0.177***	6.036***	8.331***	8.651***		6.205***
	(10.85)	(9.53)	(9.54)	(5.55)		(11.63)
35 to 39	0.169***	5.688***	9.762***	6.753***		5.763***
	(10.35)	(6.80)	(9.75)	(3.66)		(11.05)
40 to 44	0.111***	3.512***	5.655***	3.700**		3.503***
	(7.52)	(6.29)	(9.32)	(3.07)		(7.22)
45 to 49	0.0779***	2.291***	3.531***	2.167**		2.267***
	(5.38)	(4.13)	(4.26)	(2.69)		(4.20)
50 to 54	0.0586***	1.713**	1.532	4.006**		1.712*
	(4.05)	(2.70)	(1.09)	(3.24)		(2.48)
55 to 59	0.0313*	1.084	1.864	0.916		1.034
	(2.06)	(0.22)	(1.30)	(-0.09)		(0.12)
60 and above	0.00781	0.602*	0.820	0.918		0.593*

Table 4 – *Continued from previous page*

	OLS	Logistic	Urban	Rural	15-24	No robust
Independent variables	HIV	HIV	HIV	HIV	HIV	HIV
	(0.59)	(-2.44)	(-0.56)	(-0.24)		(-2.16)
Race						
White	-0.0836***	0.181***	0.190***	3.369***		0.197***
	(-8.34)	(-7.03)	(-6.11)	(3.93)		(-5.90)
Coloured	-0.0949***	0.249***	0.301***		0.302	0.315***
	(-10.79)	(-22.00)	(-9.66)		(-1.87)	(-7.38)
Indian	-0.107***	0.134***	0.0923***		0.0715***	0.0962***
	(-10.71)	(-5.69)	(-6.65)		(-10.30)	(-8.48)
Other	-0.116**	0.197	0.370			0.232
	(-2.86)	(-1.36)	(-0.77)			(-1.40)
Geotype						
Urban informal	0.0451**	1.302***			1.138	1.187
	(2.90)	(3.87)			(0.30)	(1.39)
Rural informal	-0.0215	0.872			0.524	0.775*
	(-1.76)	(-0.98)			(-1.10)	(-2.06)
Rural formal	-0.0315*	0.711			1.107	0.652**
	(-2.43)	(-1.92)			(0.14)	(-2.73)
Marital status						
Married	-0.0483***	0.672***	0.604*	0.572***	0.663	0.688***
	(-4.68)	(-4.00)	(-2.44)	(-3.52)	(-0.44)	(-3.92)
Widowed	0.00225	1.353*	1.304	1.121	2.565	1.325
	(0.15)	(2.39)	(1.18)	(1.50)	(0.97)	(1.79)
Divorced	-0.0231	0.968	0.884	1.397		1.005
	(-1.22)	(-0.14)	(-0.32)	(0.88)		(0.02)
Other	-0.0176	0.949	1.272	0.863	0.872	1.009
	(-0.74)	(-0.23)	(0.70)	(-0.29)	(-0.21)	(0.04)
Constant	0.169***	0.670	0.809	0.0933	3.569	0.503
	(5.16)	(-0.43)	(-0.17)	(-0.66)	(0.37)	(-0.64)
Province						
Constant			1.118	1.179	1.269	1.118
			(1.28)	(1.65)	(1.47)	(1.72)
Province > Municipality						
Constant			1.074	1.033	1	1.098*
			(0.78)	(0.69)	(0.57)	(2.03)
Observations	9279	9279	5153	2160	1530	9279

t statistics in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Please note that some variables are missing due to a lack of observations.

Table 5: A correlation matrix table of independent variables.

	Correlation table								
	P. ineq.	M. ineq.	E.S.	P. wealth	M. wealth	Female	Low ed.	Mid ed.	High ed.
Province inequality	1								
Municipal inequality	0.263***	1							
Economic status	-0.0871***	-0.426***	1						
Province wealth	0.774***	0.432***	-0.258***	1					
Municipal wealth	0.453***	0.884***	-0.333***	0.611***	1				
Female	-0.00603	-0.0113	0.0225*	-0.0351***	-0.0176*	1			
Low education	-0.0414***	-0.166***	0.320***	-0.0887***	-0.140***	0.0575***	1		
Mid education	0.0652***	-0.0319***	0.0726***	0.0643***	-0.00198	-0.00838	-0.385***	1	
High education	-0.0215*	0.0768***	-0.197***	0.00252	0.0491***	-0.0288**	-0.153***	-0.266***	1
N	12584								

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$