Assortative mating and income inequality in South Africa: An unconditional quantile regression analysis

Umakrishnan Kollamparambil
School of Economic and Business Sciences
University of Witwatersrand
Johannesburg
South Africa
Email: uma.kollamparambil@wits.ac.za

1. Introduction

Stratification, defined as the tendency of agents with similar characteristics to interact with one another in isolation of others has been identified in economic theory as one of the factors that explains differences in the performance within a range of contexts like countries, firms and households (Durlauf 1996, Darity 2005). This, together with the realization that individuals are not randomly assigned to households but that assortative mating explains household formation has led literature on income inequality to acknowledge positive assortative mating in developed countries as one of the contributors to income inequality. The intuition behind the argument of positive assortative mating being that the gain from the marriage function is complementary resulting in the highest type man mating the highest type woman, and the second highest type man mates the second highest type woman, and so on based on various physical and socio-economic characteristics like height, education, income etc. Positive assortative matching therefore maximizes the sum of match outputs in the marriage market when male type and female type are complements in the match output function, referred to as the "supermodularity" condition of the match output function (Becker 1973).

The implications of positive assortative mating based on education or income on household income inequality is obvious for income inequality measured at household level based on per capita household income. Therefore, it is expected that household inequality would increase in the presence of education/income based positive assortative mating, other things remaining constant. This has been empirically analysed in various developed country contexts (Eika et al 2017). While there is near universal consensus among these studies on the existence positive assortative mating (Kuhn & Ravazzini 2017, is an exception in the context of Switzerland), evidence is mixed when it comes to trend in positive assortative mating and its impact on income inequality. Moreover, empirical analysis into assortative mating and its implication for inequality have been thus far restricted to developed country contexts. This study is important as it is undertaken in the context of South Africa, a developing country which is known for high income inequality and low average levels of education, which provides a very different context from that of developed countries.

Given that South Africa has the dubious distinction of being one of the most unequal societies in the world, it is little wonder that economists have studied various dimensions of its inequality. Inequality literature in the context of South Africa has focused on labour markets, education, race, region, gender dimensions (Mwabu and Schultz, 1996; Moll, 1998; Ntuli, 2007; Burger and van der Berg, 2011). While assortative mating is expected to intersect with labour markets, education and gender, no study has explicitly looked into the role of sorting in marriage markets on South Africa's income inequality. It is therefore important to ascertain the trends in sorting behavior in household formation in South Africa and analyse its impact on income inequality determination.

Apart from being the first attempt at investigating the impact of positive assortative matching on income inequality in a developing country context, this study pioneers in the global literature in analyzing the nonlinear relationship between positive assortative mating and income inequality over age cohorts to differentiate mating behavior and its impact over age groups. Further, the study contributes by the use of unconditional quantile regression or the recentered influence function (RIF) method developed by Firpo et al (2007, 2009) to determine the impact of sorting in marriage markets on income inequality using individual level estimation. The advantage of this estimation technique over quantile regression technique is that it can be used to estimate the effect of changes in covariates on any distributional statistics such as inter-quartile ranges, the variance, or the Gini coefficient.

The study utilizes the 4^{th} wave of the National income Dynamics Survey. Education rather than income is used for ascertaining assortative mating for dual reasons. Education is expected to suffer less from endogeneity caused through reverse causality as education is likely to have been achieved prior to the household formation and secondly, because individual income data within the household is ridden with problems of missing information and unreliability of reported data.

Recent trends in inequality in South Africa indicate that between racial group inequality persists, within racial group inequality contributes a larger share to overall inequality. Inequality within the African population group has increased substantially over the post-apartheid years with earnings inequality being the highest and increasing the most over time compared to other race groups (Leibbrandt et al. 2012).This study hence focusses on the African black population with the objective to explore the existence and trends of assortative matching and whether it impacts on income inequality.

The findings of the survey using robust OLS estimation to ascertain the extent of assortative mating and recentred influence function regression to understand its impact on inequality indicate that; a) assortative mating is positive across cohorts but is weaker among younger cohorts compared to older cohorts, b) assortative mating significantly contributes to income inequality and has not changed between age cohorts and, c) there is a non-linear "U" shaped relationship between positive assortative mating and income inequality in South Africa.

2. Literature review

Positive assortative Mating and inequality

While studies empirically investigating positive assortative mating in marriage markets within developed countries, and US in particular, are more common (Siow 2015, Rose 2001, Liu and Lu 2006), there are limited studies in the context of developing countries. Gabrielli and Serio (2017) is one such exception which finds strong evidence supporting positive education based assortative mating in Argentina using non-parametric and the parametric methods. Nevertheless, the study does not find a clear pattern of increasing assortative mating over the period 1980 and 2014. While the above mentioned studies find evidence of positive sorting, the results of Kuhn and Ravazzini (2017) are unique in that they found little evidence of homogamy in hourly wages with very weak correlation between between partners' realised earnings in the context of Switzerland.

The above discussed studies (with the exception of Kuhn and Ravazzini 2017) do not take the analysis forward to investigate the impact of assortative mating on household income inequality. Given that Kuhn and Ravazzini (2017) did not find evidence of positive assortative mating, it is not surprising that they conclude that the observed Gini coefficient of realised earnings is not different from the Gini in a scenario where partners match independently of their earnings. Other studies that consider the inequality impact of positive assortative mating are Cancian and Reed (1998) and, Schwartz (2010), Eika et al (2017), Hryshka (2015) and Greenwood et al (2014). All of these studies have been undertaken in developed country contexts and find sorting in marriage to be a determinant of income inequality. There is however contention on whether changing trends in assortative mating has contributed to increasing income inequality. While Cancian and Reed (1998) and, Schwartz (2010) conclude that an increase in assortative mating has led to a rise in income inequality; Eika et al (2017), Hryshka (2015) and Greenwood et al. (2014) do not find evidence of assortative mating contributing substantially to increasing inequality. This is indicative that while sorting in household formation plays an important role in the cross section it is not an important factor in explaining the increase in inequality in recent years.

Literature thus far has used various techniques to ascertain assortative mating ranging from conventional correlation coefficient or standardized correlation used by Lui & Lu (2006) that controls for the variation in skill distribution while comparing the degree of assortative mating over time or across countries, to regression based analysis. Mare (1991) and Siow (2015) use log linear models for contingency tables and log odds ratio respectively, to provide estimates of the changing association between couples' educational characteristics while controlling for shifts in their marginal distributions. These models represent the association between husband's and wife's education in terms of a single parameter that represents the odds that husbands and wives share the same rather than different education levels. Greenwood et al (2014) on the other hand uses ordinary least Squares (OLS) regression with husbands education interacted with dummies for years to estimate change in positive assortative mating over time. The current study derives inspiration from Greenwood et al (2014) but uses age cohorts in lieu of year dummies to estimating changes in mating patterns between age groups.

In analysing the impact of positive assortative mating on income inequality, literature has tended to make use of counterfactual experiments that help quantify the change in Gini or other measures of income inequality by undertaking random matching of partners in the data and comparing it vis-à-vis revealed matches. Eika (et al 2017) constructs income distributions under alternative counterfactual scenarios, including keeping the marital sorting parameter used to match couples, the education distribution of men and women, or the economic returns to education, fixed at base year t0. The study then employs semiparametric decomposition method proposed by DiNardo et al. (1996) to quantify the relative importance of changes in educational composition, returns to education, and educational assortative mating for the rise in household income inequality. Similar approach of comparing actual couples to randomly paired simulated couples, are adopted by Hryshko et al. (2015) and Greenwood et al (2014).

A review of literature did not yield studies that link positive assortative mating with income inequality in the context of a developing country. Further, none of the existing studies take into account age-cohort based differences in mating patterns or possible non-linear relationship between positive assortative mating and income inequality. Lastly, this study is also the first to use unconditional quantile regression (RIF) to gain insights into assortative mating and income inequality using individual level data.

Income inequality in South Africa

The overall level of total inequality is determined to a large extend by the wage gaps between races and different groups of workers like men-women, non-migrant -migrants, and workers in the formal-informal economy etc. It is therefore not surprising that literature focuses on these stratification in analyzing inequality. While income inequality between racial groups persists in South Africa, within racial group inequality contributes a larger share to overall inequality and has increased substantially over the post-apartheid years. Leibbrandt et al. (2012) indicate that earnings inequality within the African population group is highest and has increased the most over time. Therefore although most studies on income inequality are aligned across race lines, this study prefers to focus on income inequality within the African population in South Africa that account for over 70 percent of the country's population.

Apart from race, gender wage-gap has also been highlighted as contributing to inequality. Studies in South Africa confirm the existence of gender wage gap, but they lack consensus on the level and trend of it in the country. Muller (2009) and Ntuli (2007) found contradictory evidence with the former finding a narrowing of gender wage gap between 1995 and 2006; and the latter finding that the gap had increased between 1995 and 2004. Kollamparambil and Razak (2011) found evidence on lines of Muller (2009) that the gender based discrimination had declined in the period 2001-2007. Further, evidence is also emerging that increasing rural-urban migration has also had a positive impact on inequality in South Africa (Kollamparambil 2017). This shows the changing dynamics of inequality in South Africa. All of the above factors intersect at household levels and become relevant in our analysis while assessing the role of assortative mating in determining household income inequality.

Education has a bearing in determining the labour market effect on inequality through both employment effect and wage effect (ILO 2015). The importance of education in determining income inequality in South Africa has been highlighted by many studies (Keswell and Poswell, 2004; Lam et al., 2011). Branson et al (2012) reveals the nexus between education and the labour in driving household inequality through the role of education in determining the probability of employment as well as the earnings returns for those that have employment. The study finds that South Africa has experienced a skills twist with the returns to matric and postsecondary education rising and the returns to levels of education below this remaining constant.

The high levels of unemployment rate in South Africa is well documented with narrow unemployment rate being higher than 25% and the level of unemployment among youth being over 50% (Zaakirah and Kollamparambil 2015). Unemployment aside, wage inequality also contributes substantially to inequality. Thus, being employed is not sufficient to eliminate inequality as a household is likely to continue in the lower deciles of the household income distribution with an earner with unskilled employment. Leibbrandt et al. (2012) decomposes income inequality and show that while labour market income accounted for 83 per cent of income inequality in 1993, this increased to 85 per cent in 2008. While these decompositions confirm the importance of rising unemployment as a key driver of inequality, they also emphasise the importance of the rising inequality of earnings for those households with access to labour market earnings. Education being a key determinant of both employment status as well as the income decile of those employed becomes pertinent. This study hence focuses on education based assortative mating among Africans as a determinant of inequality controlling for other aspects relating to gender, region and age.

3. Data and Methodology

Most studies covered under literature review have chosen to study the impact of assortative matching on income inequality over a period of time to answer the question of whether sorting in marriage has changed over time and its contribution to income inequality. One of the challenges in replicating similar studies in a developing country context like South Africa is the non-availability of reliable individual income data over sufficiently long enough time period to ascertain shifts in trends. While household income inequality data in the National Income Dynamics Survey (NIDS) is considered to be reliable, we do not see much meaning in comparing changes over time based on the various waves (four waves in approximately 2 year intervals between 2008 to 2015) of NIDS because of its panel nature given the limited change one would expect in the household formation due to either separation/divorce or death of these individuals over the period. We therefore restrict to the use of the 4th wave of NIDS as a cross-section using weights to account for systematic non-response and attrition to ensure national representability. The study makes use of age cohort-based analysis of Africans over the age of 20 years. This approach provides us with additional insights that are not accorded by existing studies. It allows an indication of transition happening across the age groups in assortative matching as well as its implication for income inequality. Further given the restrictions of data that analysis in developing countries are constrained by, it allows an early indication of changes in mating behavior through the younger cohorts.

Education variable is considered in years of education as well as levels of education. The years of education is seen to vary from 0 to 23 years (Table 1). For purpose of latter variable, we define 8 education classifications based on the highest attained level of education for the head and spouse respectively(no education, primary school, middle school, high school, matriculation, some college, bachelor's degree, advanced degree). The term married is used in a broad sense to indicate any cohabiting relationship within the household and is inclusive of unmarried partners living together. For the purpose of the study we use the term 'spouse' to refer to the partner of the household head. The study sample consists of only households with cohabiting partners. In other words, single partner households are excluded from the study.

Table 1 provides us with the descriptive summary of variables included in the analysis. While the averages are provided, the high levels of standard deviation of variables like income and education is revealing of the South African context. While 72 percent of household heads are seen to be male, a non-negligible proportion is headed by female. It is for this reason that this study prefers a gender neutral approach in the empirical analysis comparing the education of the "spouse" (female/male) with that of the household head (male/female). From the averages, the differences in the education levels of head and the spouse is not seen to be substantial. Nevertheless it may be pointed out that the variation in the education of the spouse is marginally higher than that of the household head. In terms of the levels of education, it is evident that there is high level of variation in educational levels of individuals among both household head and spouse categories. The median as well as modal level of education for both the household head and the spouse is middle school that involves 5 to 8 years of formal education.

Table 1: Descriptive statistics				
Variable	Mean	Std. Dev.	Min	Max
Household income per capita	7169.95	7799.83	11.91	67471.8
Combined education (years)	19.30	10.36	0	45
Head education (years)	9.65	5.73	0	23
Spouse education (years)	9.63	5.84	0	23
Education gap (years)	3.70	3.77	0	20
Household size	5.03	2.48	1	19
Binary Variables				
Rural	0.38		0	1
Male head	0.72		0	1
Spouse economically active	0.30		0	1
Head economically active	0.26		0	1
Education Level	Household Head(%)	Spouse (%)		
1 (no education)	16.6	11.21		
2 (primary school)	14.84	19.22		
3 (Middle school)	35.16	31.12		
4 (high school)	15.52	18.08		
5 (Matriculation)	4.11	6.63		
6 (some college)	6.6	4.81		
7 (bachelor's degree)	4.79	5.95		
8 (advanced degree)	2.28	2.97		

Source: Calculated from weighted NIDS sample

A cohort-wise summary of some of our key variables is provided in Tables 2 and 3. Household income is seen to follow the expectations under the life cycle hypothesis with an inverted U-shaped curve where income peaks just before retirement age. On the other hand education of both the head of the household as well as the spouse is seen to be increasing consistently from under 7 years for the oldest cohort to over 12 years amongst the youngest cohort. It is therefore not surprising that the combined education in years of both spouses have increased consistently among younger cohorts. Education gap has however also increased marginally among younger cohorts but this cannot be taken as being indicative of weaker positive assortative mating among younger cohorts as similar trend is not visible with PAM index. Measures of positive assortative mating, both calculated in terms of year and education level, indicate an increase among younger cohorts. The decline in Gini of pre-retirement cohorts (younger than 60 years) raises questions against this backdrop.

Table 2:A cohort-wise summary of key variables of interest					
Age cohort (years)	observations	Per capita household income (Rands)	Household head Education_years	Spouse Education years	Income gini
cohort 1 (60 >)	271	5961.05	6.61	6.90	0.52196
		(7569.95)	(5.94)	(5.91)	
cohort 2 (51-60)	291	10272.56	8.26	8.24	0.74868
		(38760.46)	(6.00)	(5.98)	
cohort 3 (41-50)	248	9214.93	10.88	10.80	0.6187
		(16512.46)	(5.51)	(5.64)	
cohort 4 (31-40)	142	7388.6	12.64	12.02	0.58821
		(10066.62)	(4.72)	(4.74)	
cohort 5 (21-30)	33	6300.21	12.71	12.05	0.51433
		(6848.16)	(3.86)	(3.89)	
Total	1065	7924.45	9.72	9.47	0.6338
		(20255.78)	(6.00)	(5.97)	

Source: Author calculation based on weighted NIDS data. Standard Errors in parenthesis

Table 3: A cohort-wise summary of measures of assortative mating				
Age cohort (years)	PAM index (level)	PAM index (year)	Combined education	Education gap
cohort 1 (60 >)	0.604	0.602	13.56	-0.24
	(0.240)	(0.383)	(-10.83)	(-4.95)
cohort 2 (51-60)	0.701	0.648	16.53	0.02
	(0.239)	(0.359)	(-10.85)	(-5.15)
cohort 3 (41-50)	0.774	0.734	21.78	0.09
	(0.223)	(0.308)	(-9.94)	(-5.16)
cohort 4 (31-40)	0.835	0.821	24.79	0.65
	(0.179)	(0.235)	(-8.1)	(-4.91)
cohort 5 (21-30)	0.859	0.860	24.94	0.7
	(0.172)	(0.210)	(-6.47)	(-4.38)
Total	0.715	0.718	19.33	0.17
	(0.241)	(0.328)	(-10.9)	(-5.04)

Source: Author calculation based on weighted NIDS data. Standard Errors in parenthesis

Table 4: Correlation between Education of Spouses across age cohorts			
		Education (level)	
Age cohort (years)	Education (in years) Pearson correlation	Pearson correlation	Spearman rank correlation
cohort 1 (60 >)	0.7058**	0.1799**	0.6623**
cohort 2 (51-60)	0.6539**	0. 4286 **	0.6078**
cohort 3 (41-50)	0.6262**	0. 5422 **	0.6118**
cohort 4 (31-40)	0.5537**	0.4516 **	0.5889**
cohort 5 (21-30)	0.6077**	0.3745 **	0.5252**
Total	0.6581**	0.5247 **	0.6743**

Source: Author calculation based on weighted NIDS data.

A non-parametric preliminary analysis finds strong support for positive assortative mating in South Africa across age cohorts using Pearson correlation as well as Spearman rank correlation. The correlation between the years of education of spouses is seen to be highest in older cohorts and although still high, the strength of positive association is seen to reduce consistently among younger cohorts.

Methodology

Our empirical investigation is divided into two parts: the first is to analyse whether the phenomenon of assortative mating is applicable in South African context, and the second is to tease out the impact of this on income inequality. Most literature focus heavily on non-parametric methodologies, but given our constraints with data we prefer regression approach in this section. A robust OLS estimation of the equation below is used to test for assortative mating:

$$E_i^s = \alpha + \theta E_i^h + \sum_{a\in\tau} \gamma_a Cohort^h + \sum_{a\in\tau} \delta_a Cohort^h * E_i^h + X_i'\beta + \varepsilon_i \text{ , with } \varepsilon_i \sim N(0,\sigma)$$

Where, E_i^h and E_i^s are the education (years/Level) of the head of household and his/her spouse respectively for the i[th] household. The age based cohorts are dummy variables taking value 1 for head of household belonging to a cohort and 0 otherwise, where $a \in \tau \equiv \{\ 2\ (51\ to\ 60\ years), 3\ (age\ 41\ to\ 50\ years),$

$4\ (age\ 31\ to\ 40\ years, and\ 5\ (age\ 21\ to\ 30\ years)\}$. It is set up so that interaction terms between age cohorts and head of household are included to understand the differences in mating patterns between cohorts. The coefficient θ measures the impact of household heads's education on his/her spouse for the baseline cohort 1 (age above 60 years). A positive and significant θ would indicate the existence of positive assortative mating. The coefficient δ

gives the additional cohort specific impact of a household head's education on his spouses relative to the baseline cohort 1. Insignificant δ for a cohort would indicate that the degree of assortative mating does not differ from that of cohort 1, while a positive and significant δ for a cohort would indicate a higher degree of positive assortative mating while a negative and significant δ for a cohort would indicate that the degree of homogamy is lower for that cohort relative to cohort1. The differences in δ over cohorts indicates the changes in assortative mating across cohorts and can be broadly interpreted as indicative of changes over time. The γ's control for the secular change in educational attainments of spouses across cohorts.

The key estimation coefficient of interest is "θ", where a positive and significant coefficient would indicate positive assortative mating. This expectation is fueled by theory as well as the significant correlation coefficients observed in the section above. The expectation with regard to the age cohort coefficients (γ) based on descriptive statistics already undertaken, is that younger cohorts have higher education compared to older cohorts.

Further control variables like geography type (dummy variable =1 for rural and 0 otherwise), gender of the head of household (dummy variable =1 if head of household is male and 0 otherwise) are included in the X vector. Robust estimation is undertaken to control for heteroscedasticity. VIF is reported and considered to be at acceptable levels. Normality of error terms is tested and statistics included in results.

The second part of the analysis involves teasing out the impact of positive assortative mating on income inequality. This study adopts parametric techniques through the estimation of unconditional quantile regression or RIF with Gini of household income as the distributional parameter as the dependent variable. Proposed by Firpo et al (2007, 2009), RIF is a regression method to estimate the impact of explanatory variables on quantiles of the unconditional marginal distribution an outcome variable. The method consists of running a regression of the recentred influence function of the unconditional quantile on the explanatory variables. Besides quantiles, this can be generalized to other distributional statistics like variance and Gini. This study uses Gini as the preferred measure of income inequality.

RIF regression hence is suitable to identify inequality determinants at the individual level and has been recently introduced in the analysis of income inequality (Alejo et al 2012, Sakellario 2012). RIF regression is similar to the standard OLS regression except that it replaces the dependent variable, individual happiness level, with the recentred influence function , RIF (y:G) , the distributional Gini (G) of household income in our analysis. The dependent variable hence can be interpreted as income inequality.

$E[RIF(Y;G)|X]=X'\beta$

Then, by the law of iterative expectations:

$G=E(RIF(y;G))=E_X[E(RIF(y;G)|X]=E(X)'\beta$

Where X indicates the matrix of explanatory variables including the combined years of education of spouses, and the positive assortative mating index described further down this section.

The influence function is computed using the fact that the expected value of the influence function is equal to zero and the law of iterated expectations, we can express the distributional statistic of interest as the average of the conditional expectation of the RIF given the covariates. The average derivatives computed using the RIF-regressions yield the partial effect of a small location shift in the distribution of covariates on the distributional statistic of interest. Firpo et al (2009) call this parameter Unconditional Partial Effect (UPE). By approximating the conditional expectations by linear functions, the coefficients of these RIF-regressions indicate by how much the functional of the marginal outcome distribution is affected by an infinitesimal shift to the right in the distribution of the regressors. The β coefficients hence can be interpreted as the marginal impact of a small change in E(X) on the Gini index.

The key challenge in this analysis is to construct a variable to capture the degree of assortative mating. The simplest measures used in this regard are the sum of the educational attainment of both partners for i^{th} household($CombEdu_i$). The limitations of this approach are obvious because a measure of 22 can be the result of one partner accounting for 20 and the other for 2 or both partners having 11 years of education. But in the context of high degree of positive assortative mating established through the first stage of analysis, this measure may still be retained as meaningful. Another simple measure would be to construct the difference between the educational attainment of partner ($EduGap_i$). A negative effect of the variable is expected under the theory of positive assortative mating. Next we construct an index based on ratio of sum of educational attainment divided by the gap. This can be considered to be an improved measure of assortative mating when education is considered in years and level of education where gap is normalised with the combined education of partners to make allowances for the increases levels of education among younger age cohorts.

$$CombEdu_i = E_i^s + E_i^h$$

$$EduGap_i = |E_i^s - E_i^h|$$

These measures are refined further in the form of a positive assortative index below:

$$PAMIndex_i = 1 - \frac{|E_i^s - E_i^h|}{E_i^s + E_i^h}$$

PAM index by definition would range from 0-1. Perfect positive assortative matching index is given by a value of 1 and complete absence of positive assortative matching is given by 0. The key determinant of the impact of positive assortative mating on income inequality would be the sign and significance of the coefficient of the PAM index. Based on literature, PAM index is expected to have a positive effect on income inequality. PAM index is included in estimations in its squared form as well to check for non-linear relationship. Studies reviewed in literature thus far have not investigated possible non-linear relationship between PAM and income inequality, therefore this can be cited as a further contribution of this study.

4. Econometric analysis

Assortative mating in SA

Education of spouses, whether measured in years or levels, are seen to be positively associated at very high confidence level indicative of positive assortative mating. The years of education of spouses is seen to be increasing consistently among youngers cohorts. This is consistent with the increasing averages of education among younger cohorts observed in Table 2. However, an interesting finding coming out through the interaction of education of household head and cohorts is that positive assortative mating is weakening significantly among younger cohorts aged between 20 years and 40 years. Another expected finding is that education level of spouses in rural areas are significantly lower than in urban areas. It is also interesting to note that male spouses have lower education than female spouses. All results are consistent across measurement of education using years of completed education and highest level of education achieved.

Table 5: Positive Assortative Mating: Results robust OLS regression

VARIABLES	Spouse Education years	Spouse Education years	Spouse Education level	Spouse Education level
Head education	0.626***	0.627***	0.692***	0.692***
	(0.0516)	(0.0485)	(0.0467)	(0.0448)
cohort2	1.728**	1.668**	0.584**	0.559**
	(0.827)	(0.826)	(0.238)	(0.236)
cohort3	3.720**	3.653**	1.383***	1.358***
	(1.562)	(1.560)	(0.266)	(0.263)
cohort4	4.765***	4.694***	1.911***	1.883***
	(1.212)	(1.210)	(0.470)	(0.466)
cohort5	-	8.821***	-	2.817**
		(2.539)		(1.414)
Head Education *cohort2	-0.0822	-0.0827	-0.0251	-0.0247
	(0.105)	(0.104)	(0.0235)	(0.0230)
Head Education *cohort3	-0.195	-0.195	-0.0726***	-0.0721***
	(0.126)	(0.125)	(0.0232)	(0.0227)
Head Education *cohort4	-0.242**	-0.242**	-0.102***	-0.101***
	(0.101)	(0.0995)	(0.0369)	(0.0364)
Head Education *cohort5	-	-0.538***	-	-0.171***
		(0.201)		(0.015)
Rural	-1.167***	-1.164***	-0.464***	-0.460***
	(0.403)	(0.392)	(0.112)	(0.109)
Male Head	-0.987**	-0.931*	-0.415***	-0.399***
	(0.496)	(0.476)	(0.120)	(0.116)
Constant	3.313***	3.355***	1.073***	1.090***
	(0.481)	(0.471)	(0.204)	(0.199)
Observations	1,065	1,119	1,064	1,117
S-K test (Prob>chi2)	0.1145	0.1039	0.4819	0.3073
VIF	4.96	4.77	4.78	4.65
R-squared	0.390	0.388	0.384	0.382

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The declining trend of positive assortative mating is noted as educational levels are starting to rise among the younger cohorts but this is not widespread enough resulting in higher education gaps among partners in the household. Having found significant evidence of positive assortative mating in South Africa, we turn to looking at its impact on income inequality in the next section.

Assortative mating and Inequality

The PAM index based on both years and level of education is seen to have a significant impact on income inequality (Tables 6 and 7). The impact of PAM index based on level of education and income inequality is positive with increase in assortative mating leading to increased income inequality. These results corroborate the findings of Eika et al (2017), Hryshka (2015) and Greenwood (2014) in US and other developed countries, indicating that results of the relationship hold in countries with high income inequality and low education levels. The finding however is contrary to Kuhn & Ravazzini (2017) in the context of Switzerland.

However, the relationship between positive assortative mating and income inequality is more nuanced as indicated by the non-linear relationship revealed in Tables 6 and 7. As expected, at lower levels of PAM index, an increase in positive assortative mating reduces inequality but at higher levels of PAM index this effect turns positive indicating an increase in income inequality. This is indicative of a 'U'-shaped relationship between positive assortative mating and income inequality. The impact of PAM index, based on both education years and level, is not seen to change significantly across age cohorts indicating a stability in the relationship. This provides strength to the conclusion that increase in inequality cannot be attributed to increased positive assortative mating inequality among younger cohorts in South Africa. On the contrary, reduced degree of positive assortative mating among younger cohorts (Table 5) has contributed to moderating the increase in income inequality among younger cohorts in South Africa.

Apart from positive assortative mating, the combined education of spouses, measured by years and level, has a positive impact on income inequality. In terms of the separate impact of education of partners, the education of the head of the household, but not that of the spouse, is seen to contribute positively to household inequality. Inequality is not seen to differ substantially across age-cohorts in South Africa.

Other findings emanating from the result are that income inequality is higher among rural areas compared urban areas and that households headed by male have higher inequality as compared to those headed by females. That rural areas as well as female headed households have higher levels of poverty in South Africa is well documented (World Bank 2018). Further, the same study found that the rich benefited more from consumption expenditure growth between 2006 and 2015 than the poor and those in the middle resulting in further polarisation with rural regions. Further results indicate that bigger household size contributes to reducing income inequality. Our findings validate the conclusion of Keller (2008) and Ebrahim et al (2013) who highlight the role of household formation strategy to counter unemployment and poverty in South Africa.

Table 6: RIF regression results: Impact of PAM Index (years of education) and Income inequality (Gini)

VARIABLES	Gini	Gini	Gini	Gini
Household head Education	0.00455**			
	(0.00215)			
Spouse Education	0.000515			
	(0.00206)			
Combined Education		0.00247**		
		(0.00120)		
PAM index	-0.356***	-0.357***	-0.304**	-0.00950
	(0.121)	(0.121)	(0.118)	(0.0522)
PAM index Squared	0.316***	0.317***	0.300***	
	(0.108)	(0.108)	(0.108)	
PAM index* cohort2	-0.0500	-0.0465	-0.0383	-0.0515
	(0.0778)	(0.0778)	(0.0778)	(0.0779)
PAM index* cohort3	-0.0816	-0.0918	-0.0898	-0.119
	(0.0843)	(0.0839)	(0.0840)	(0.0836)
PAM index* cohort4	0.0924	0.0961	0.0967	0.0187
	(0.151)	(0.151)	(0.151)	(0.149)
PAM index* cohort5	-0.389	-0.397	-0.354	-0.468
	(0.414)	(0.414)	(0.415)	(0.414)
Cohort 2	0.0382	0.0352	0.0342	0.0369
	(0.0341)	(0.0340)	(0.0340)	(0.0341)
Cohort 3	0.0435	0.0446	0.0500	0.0598*
	(0.0329)	(0.0329)	(0.0328)	(0.0327)
Cohort 4	-0.00948	-0.0104	-0.00153	0.0151
	(0.0410)	(0.0410)	(0.0409)	(0.0405)
Cohort 5	0.0633	0.0625	0.0647	0.0868
	(0.0743)	(0.0743)	(0.0744)	(0.0743)
Rural	0.0617***	0.0607***	0.0523***	0.0566***
	(0.0202)	(0.0201)	(0.0198)	(0.0198)
Male Household head	0.0562***	0.0569***	0.039**	0.0386**
	(0.0206)	(0.0202)	(0.019)	(0.0195)
Spouse employed	-0.0129	-0.0106	-0.0123	-0.0110
	(0.0315)	(0.0315)	(0.0315)	(0.0316)
Head employed	0.00684	0.00552	0.00594	0.00903
	(0.0337)	(0.0337)	(0.0337)	(0.0338)
Household size	-0.0203***	-0.0202***	-0.0216***	-0.0211***
	(0.00408)	(0.00408)	(0.00403)	(0.00404)
Constant	0.598	0.589	0.635	0.629

	(0.614)	(0.614)	(0.614)	(0.616)
Observations	1,017	1,017	1,017	1,017
R-squared	0.061	0.059	0.055	0.048

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: RIF regression results: Impact of PAM Index (Education level) and Income inequality (Gini)

VARIABLES	Gini	Gini	Gini	Gini
Household head Education	0.0125*			
	(0.00643)			
Spouse Education	0.00329			
	(0.00635)			
Combined Education		0.00786**		
		(0.00348)		
PAM level index	-0.307**	-0.349**	-0.276**	0.00199
	(0.144)	(0.136)	(0.133)	(0.0571)
PAM level index Squared	0.287**	0.315***	0.270**	
	(0.122)	(0.118)	(0.116)	
PAM level index* cohort2	-0.00780	-0.0101	-0.0229	-0.0408
	(0.0747)	(0.0747)	(0.0746)	(0.0744)
PAM level index* cohort3	-0.0348	-0.0453	-0.0676	-0.108
	(0.0818)	(0.0809)	(0.0804)	(0.0787)
PAM level index* cohort4	0.119	0.119	0.0922	0.0298
	(0.150)	(0.150)	(0.150)	(0.148)
PAM level index* cohort5	-0.359	-0.363	-0.348	-0.468
	(0.415)	(0.415)	(0.416)	(0.413)
Cohort 2	0.0227	0.0217	0.0274	0.0338
	(0.0332)	(0.0332)	(0.0331)	(0.0331)
Cohort 3	0.0289	0.0303	0.0433	0.0568*
	(0.0326)	(0.0325)	(0.0321)	(0.0316)
Cohort 4	-0.0182	-0.0182	-0.00145	0.0120
	(0.0405)	(0.0405)	(0.0399)	(0.0396)
Cohort 5	0.0582	0.0584	0.0685	0.0897
	(0.0746)	(0.0746)	(0.0746)	(0.0742)
Rural	0.0639***	0.0630***	0.0533***	0.0566***
	(0.0201)	(0.0201)	(0.0197)	(0.0196)
Male Household head	-0.0589***	-0.0581***	-0.0627***	-0.0612***
	(0.0217)	(0.0216)	(0.0216)	(0.0216)
Spouse employed	-0.0114	-0.00934	-0.0111	-0.0101
	(0.0317)	(0.0316)	(0.0316)	(0.0317)
Head employed	0.00855	0.00754	0.00762	0.00800
	(0.0337)	(0.0337)	(0.0338)	(0.0339)
Household size	-0.0200***	-0.0198***	-0.0214***	-0.0210***
	(0.00409)	(0.00409)	(0.00404)	(0.00405)
Constant	0.617***	0.631***	0.683***	0.642***
	(0.0671)	(0.0653)	(0.0613)	(0.0588)
Observations	1,015	1,015	1,015	1,015
R-squared	0.059	0.058	0.053	0.048

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

5. Conclusion.

Positive assortative mating has been highlighted in literature as one of the contributing factors to income inequality in developed countries. While there is consensus on the existence of positive assortative mating in developed countries and its contribution to income inequality, its trend and contribution to increasing income inequality is still up for debate. The current study is the first to analyse the impact of positive assortative mating in the context of a developing country with very high levels of inequality and low education levels.

Although South Africa has rich literature on income inequality, the role of assortative mating in determining its inequality has not been analysed before. This phenomenon intersects with gender, education and labour markets, and although these have been accepted as contributing to South Africa's high income inequality, it has not been assessed within the framework of positive assortative mating. Apart from this, the current study undertook a cohort based analysis of assortative mating and hence makes a contribution to literature outside of South Africa as well. Further, even amongst global literature this is the first to posit a non-linear relationship between positive assortative mating and income inequality.

Initial investigations using simple correlation statistics found strong evidence for positive assortative mating in South Africa but finds that the strength of this association to be weakening among younger cohorts. It is therefore not surprising that income inequality among younger cohorts were also found to be lower compared to older pre-retirement cohorts. More robust multivariate regression analysis found similar evidence of positive assortative to be stronger among older cohorts as compared to younger cohorts. The evidence from South Africa is in line with those from developed countries (Siow 2015, Rose 2001) in validating the existence of positive assortative mating. While there are no South Africa based studies to compare our findings against, the trend derived from age-cohort analysis, indicates that positive assortative mating is weakening and is aligned with the findings of Liu and Lu (2006) for the period 1960-2000 in the US. It however goes against Siow (2015) who observes an increase in supermodularity between 1970 and 2000 in US.

Using the recentred influence function regression, the study further looked at the relationship between assortative mating and income inequality and found a non-linear relationship between income inequality and level of education based assortative mating. The results posit a U-shaped relationship between positive assortative mating and income inequality. This indicates that an increase in positive assortative mating at lower and higher levels of assortative mating leads to a reduction and increase in inequality respectively.

In the absence of studies in South African context to compare with, we baseline our results against the studies in the developed country contexts. While Eika et al (2017), Hryshka (2015) and Greenwood (2014) , Cancian and Reed (1998) and, Schwartz (2010) found evidence of positive assortative mating contributing to inequality in developed countries, these studies only investigated linear relationships. The nonlinear relationship between positive assortative mating and income inequality has thus far not been analysed in literature and can be said to be a major finding of this study. Our study however did not find any major shift across the age-cohorts in terms of the impact of positive assortative mating on income inequality. The increase in inequality hence cannot be pinned on changing pattern of mating across age cohorts in the South African context. This finding is in line with Eika et al (2017), Hryshka (2015) and Greenwood (2014) who attribute the role of positive assortative mating in explaining cross-sectional inequality but not to changing trends in inequality. This goes against the findings of Cancian and Reed (1998) and, Schwartz (2010) who attribute increase in inequality to increase in the degree of positive assortative mating.

The overall conclusion seems to be that positive assortative mating is a significant factor that contributes to the income inequality in South Africa. Despite a reduction in positive assortative mating among younger generation, we do not see a difference in the inequality impact of positive assortative mating. The current study is limited by a cross-sectional analysis and can be taken forward in future to incorporate panel data that allows comparison over time. This will also allow a decomposition analysis to quantify the contribution of positive assortative mating and changing trends on income inequality in South Africa. Further multi-dimensional assortative mating can be considered in future research. Nevertheless, this study is a significant contribution in providing a benchmark study for further research on the subject.

References

Branson, N., Garlick, J., Lam, D., Leibbrandt, M. (2012). *Education and Inequality: The South African Case.* A Southern Africa Labour and Development Research Unit Working Paper Number 75. Cape Town: SALDRU, University of Cape Town

Burger, C. and Van Der Berg, S. (2011). *Modelling cognitive skills, ability and school quality to explain labour market earnings differentials*, Stellenbosch Economic Working Papers: 08/11.

Cancian, M and Reed, D. 1998. "Assessing the Effects of Wives' Earnings on Family Income Inequality." *The Review of Economics and Statistics*, 80 (1): 73-79.

Darity, W. (2005). Stratification economics: The role of intergroup *inequality Journal of Economics and Finance*, Vol 29 (2), 144-153. https://doi.org/10.1007/BF02761550

Durlauf, S.N. (1996) A Theory of Persistent Income Inequality, *Journal of Economic Growth*, 1: 75-93.

Ebrahim, A., Woolard, I. and Leibbrandt, M. (2013*). Unemployment and Household formation.* Cape Town: SALDRU, University of Cape Town. SALDRU Working Paper Number 126/ NIDS Discussion Paper 2013/8.

Eika, Lasse, Magne Mogstad, and Basit Zafar. 2014. "*Educational Assortative Mating and Household Income Inequality*." National Bureau of Economic Research Working Paper 20271.

Gabrielli M. F. & Serio, M. (2017) Testing Assortative Mating: Evidence from Argentina, *Economic Analysis Review*, Vol 32, No 2.

Greenwood, Jeremy, Nezih Guner, Georgi Kocharkov and Cezar Santos. 2014. "Marry Your Like: Assortative Mating and Income Inequality." *American Economic Review:* Papers & Proceedings, 104 (5): 348-353.

Hryshko, Dmytro, Chinhui Juhn and Kristin McCue. 2015. "Trends in Earnings Inequality and Earnings Instability among U.S. Couples: How Important is Assortative Matching?" Working Papers 2015-01, University of Alberta, Department of Economics.

Ismail Z & Kollamparambil U (2015) Youth Unemployment Duration and Competing Exit States: What Hides Behind Long Spells of Black Youth Unemployment in South Africa?, *African Development Review*, Volume27, Issue3.

Keller, Sonja. "Household formation, poverty and unemployment: The case of rural households in South Africa." *South African Journal of Economics* 72, no. 3 (2004): 437-483.

Keswell, M. and Poswell, L. (2004). Returns to education in South Africa: A retrospective sensitivity analysis of the available evidence, *South African Journal of Economics*, 72(4): 834-860.

Kollamparambil U & Razak A (2016) Trends in Gender wage gap and Discrimination in South Africa, *Indian Journal of Human Development*, Vol 10, Issue 1.

Kollamparambil U (2015) Impact of Internal In-Migration on Income Inequality in Receiving Areas: A District Level Study of South Africa, *Journal of Development Studies*, Volume 53, 2017 - Issue 12.

Leibbrandt, Finn and Woolard (2010), Describing and decomposing post-apartheid income inequality in South Africa, *Development Southern Africa*, Volume 29, 2012 - Issue 1

Liu H & Lu J (2006), Measuring the degree of assortative mating, *Economics Letters* 92 (2006) 317–322

Mare, Robert. D. 1991. .Five decades of educational assortative mating..*American Sociological Review*. 56(1): 15-32.

Moll, P. (1998). Primary Schooling, Cognitive Skills and Wages in South Africa, Economica, 65(258): 263 – 284.

Muller, C. (2009). *Trends in the gender wage gap and gender discrimination among part-time and full-time workers in post-apartheid South Africa* (Working Paper No. 124). Development Policy Research Unit, School of Economics, University of Cape Town.

Mwabu, G. and Schultz, P. (1996). Education Returns Across Quantiles of the Wage Function: Alternative Explanations for Returns to Education by Race in South Africa, *The American Economic Review*, 86(2):335 – 339.

Ntuli, M. (2007). Exploring Gender Wage Discrimination in South Africa, 1995- 2004: A Quantile Regression Approach. IPC Working Paper Series: 56.

Rose, Elaina. 2001. *Marriage and assortative mating: How have the patterns changed?* Working Paper 353330, University of Washington.

Schwartz, C. R. 2010. “Earnings Inequality and the Changing Association between Spouses’ Earnings.” *American Journal of Sociology*, 115 (5): 1524-1557.

Siow, A. 2015. "Testing Becker's Theory of Positive Assortative Matching," *Journal of Labor Economics*, University of Chicago Press, vol. 33(2), pages 409-441.