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

Umakrishnan Kollamparambil

Abstract

Apart from being the first attempt at investigating the impact of education based positive assortative mating on income inequality in a developing country context with very high levels of inequality and low education levels, this study pioneers in the global literature in analyzing the nonlinear relationship between positive assortative mating and income inequality. Further, the study contributes by the use of unconditional quantile regression or the recentered influence function (RIF) method developed by Firpo et al. (2009) to determine the impact of sorting in marriage markets on income inequality using the fourth wave of the South African National income Dynamics Survey. The study finds convincing evidence of existence of positive assortative mating in South Africa. However, the strength of the relationship is seen to be weakening among younger cohorts as compared to older cohorts. The study further found a non-linear U-shaped relationship between income inequality and level of education based assortative mating.

1. Introduction

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 (Leibbrandt et al. 2012; Ismail and Kollamparambil 2015); education (Keswell and Poswell, 2004; Lam et al., 2011; Branson et al. 2012), race and gender dimensions (Mwabu and Schultz, 1996; Moll, 1998; Ntuli, 2007; Burger and van der Berg, 2011; Muller 2009 and; Kollamparambil and Razak 2011). All of the above factors intersect at household levels and the role of assortative mating in household formation becomes relevant. It is therefore surprising that no study has explicitly looked into the role of sorting in marriage markets on South Africa’s income inequality.

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). On similar lines, Sociology literature has articulated homogamy as leading to positive assortative mating (PAM) whereby man and woman with similar characteristics (physical and socio-economic characteristics like height, education, income etc.) tend to come together to form a household. Becker (1973) highlights the gain from such household formation as complementary resulting in the maximization of the sum of match outputs. This in turn determines not only intrahousehold allocation of resources, but also labour supply and household income (Chiappori 2015). 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. Therefore, an analysis of matching patterns and of the incentives thus generated can contribute to our understanding of long-term economic trends like income inequality.

There is near universal consensus on the existence of PAM within developed countries (Siow 2015, Rose 2001, Liu and Lu 2006), with Kuhn & Ravazzini (2017) being the only exception who did not find evidence of it in the Swiss context. However, when it comes to the trend in positive assortative mating and its contribution to increasing income inequality, there is little consensus. Gabrielli and Serio (2017) does not find a clear pattern of increasing assortative mating over the period 1980 and 2014 in Argentina. Gehlan and Lang (2016) point out that "Conclusions about changes in homogamy are sensitive to how educational groups are defined". The study finds that the trends observed in mating are not robust to reclassification of education levels.

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. Kuhn and Ravazzini (2017) 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 positive 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.

Apart from the fact that the existing studies have mostly focused on developed countries, none of them have considered the possibility of a non-linear effect of PAM on income inequality. In this paper we argue that the relationship between PAM and inequality need not be monotonic and is determined based on whether the initial increase in PAM is weighted more towards lower or higher end of the income distribution as explained further in the next section. Investigating this is particularly relevant in a developing country like South Africa, which is known for high income inequality and low average levels of education, which provides a very different context from that of developed countries. With 90% of the population having a per capita household income below the country mean, the non-monotonic relationship between PAM and income inequality hypothesized in this paper is a strong possibility in the South African context.

Further, the study contributes by the use of the recentered influence function (RIF) regression method developed by Firpo et al (2009) to determine the impact of sorting in marriage markets on income inequality. The advantage of this estimation technique over standard quantile conditional regression technique is that, while the results of the latter is interpreted conditional on the other covariates, the unconditional quantile regression approach affords a more direct interpretation of the effect on the entire distribution of income. This enables the use of unconditional quantile regressions to not just interpret the impact of PAM within each quantile but also derive the between effect across the entire income distribution. Further, RIF can be used to estimate the effect of changes in covariates on distributional statistics such as the variance, and the Gini coefficient using individual level cross-section data.

The study utilizes the 4th 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.

The study uses robust OLS estimation to ascertain the extent of assortative mating and; recentred influence function regression to understand the association between positive assortative mating and income inequality. The results indicate that; a) assortative mating is positive across cohorts but is weaker among younger cohorts compared to older cohorts and, b) there is a non-linear "U" shaped relationship between positive assortative mating and income inequality in South Africa. The findings highlight that the association between PAM and income inequality is quite different in the context of a developing country with high levels of income inequality and low levels of education, from what has been observed in developed countries.

2. Non-linearity between PAM and Income inequality

Drawing upon the argument by Becker (1974; 1981) that positive matching is stronger on complementary characteristic but not on substitutable characteristics, Gehlan and Lang (2016) explain that it may be expected for men with high education levels to marry women not in the labor force and specialize instead in home production. This however may not be the preference in the lower or middle segment of the education distribution with increasing PAM as a result of growth in women's labor force participation. Therefore, the case made is that the assortative mating techniques and preferences could differ at different points of the matching criteria distribution resulting in varied impact of PAM on inequality. Hence an increase in the average PAM across the population, on the back of positive matching in the left-hand side of the distribution, is likely to result in decreased income inequality. Of course, further increases in average PAM would require that the matching behaviour pervades the entire distribution (including at higher education levels), resulting ultimately in increased income inequality. Thus, when we account for the differences in matching techniques across the underlying distribution, it is possible to highlight that non-linearity could emerge with respect to income inequality with initial increases in positive mating at lower levels resulting in decreased inequality before further increases in PAM increases income inequality.

The possibility of a non-linear relationship is illustrated using income distribution of six hypothetical households where total income determined is the sum of wages earned by head of household (i) and his/her spouse (s). Eq (1) has a PAM of 0 when all spouses have 0 education, while the heads have education ranging from one to six.

$$I0(1+0, 2+0, 3+0, 4+0, 5+0, 6+0) \qquad (1)$$

The PAM can be increased from its current level of 0 in eq(1) by either increasing the education levels of spouses of heads with educations levels to the left or right of the mid-point three. Eq (2) shows an increase in PAM to 0.16 through assortative mating on the left-hand side of the distribution while eq (3) achieves the same PAM through assortative mating on the right-hand side of the distribution.

$$I1(1+1, 2+0, 3+0, 4+0, 5+0, 6+0) \qquad (2)$$

$$I6(1+0, 2+0, 3+0, 4+0, 5+0, 6+6) \qquad (3)$$

Despite the same PAM levels, the Gini coefficients of (2) and (3) are not equal indicating that at similar level of average PAM, household incomes can depict different levels of inequality. When average PAM is 0 (depicted as I0), inequality measured as Gini coefficient is 0.27. Increased PAM of 0.16 due to positive matching in the lower segment of income distribution (I1 in eq 2), leads to reduced Gini coefficient of 0.22. On the other hand, at similar PAM level of 0.16, inequality increases to 0.4 when positive matching happens at higher end of the income distribution (I6 in eq 3). This is shown below (eq 4) where the level of matching (i+s) is shown to determine the resulting inequality level.

$$I6(1+0, 2+0, 3+0, 4+0, 5+0, 6+6) > I0(1+0, 2+0, 3+0, 4+0, 5+0, 6+0) < I1(1+1, 2+0, 3+0, 4+0, 5+0, 6+0) \qquad (4)$$

Therefore, the relationship between PAM and inequality is determined based on whether the initial increase in PAM is weighted more towards lower or higher end of the income distribution.

In the South African context of high levels of poverty and inequality, over 90% of households live below the mean income level. Therefore, as PAM increases, the weight is likely to be below the mean incomes resulting in reduced income inequality initially. However sustained increase in PAM is possible when positive assortative mating is observed also at higher income households, which leads to increased inequality. Hence, we hypothesize that a non-linear U shape relationship would be observed between PAM and income inequality in South Africa.

3. Methodology

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. Greenwood et al (2014) on the other hand uses ordinary least Squares (OLS) regression with husband's 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). The limitation to this approach is that it assumes that labour participation decisions are not impacted by the pairing factor. This assumption is however long drawn as labour participation of the spouse is very much a function of the wages of the household head. This study therefore chooses a different approach based on the revealed incomes rather than simulation techniques.

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 individuals 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. The level of educational attainment has changed in South Africa over the recent decades and as such requires that education variable is normalized to enable comparison across cohorts. The analysis of the PAM trends and its inequality is based on normalized education variable.

Our empirical investigation is divided into two parts: the first is to analyse whether the phenomenon of assortative mating is applicable in the South African context using robust OLS regression estimation and, the second is to tease out the association between positive assortative mating and income inequality using the RIF regression estimations with quantiles, variance and the Gini coefficient.

Assortative Mating

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 5 below is used to test for assortative mating along the lines of Greenwood et al (2014):

$$E_i^s = \alpha + \theta E_i^h + \sum_{a\in\tau}\gamma_a Cohort^h + \sum_{a\in\tau}\delta_a Cohort^h * E_i^h + X_i'\beta + \varepsilon_i \text{ , with } \varepsilon_i \sim N(0,\sigma) \quad (5)$$

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 $a \in \tau \equiv \{\ 2\ (51\ to\ 60\ years), 3\ (age\ 41\ to\ 50\ years),$ $4\ (age\ 31\ to\ 40\ years, and\ 5\ (age\ 21\ to\ 30\ years)\}$. Cohort 1 (age over years) is excluded as the benchmark. Further, interaction terms between age cohorts and head of household are included to understand the differences in the mating patterns between 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 γ's control for the secular change in educational attainments of spouses across cohorts. Again, based on the descriptive statistics already undertaken, the expectation is that younger cohorts have higher education compared to older cohorts. The coefficients of the interaction (δ) between age cohorts and head of household allows us to identify whether the assortative matching has changed with age cohorts. This allows a time dimension to the issue at hand even while using cross-section data.

Further control variables like geography type (dummy variable =1 for rural and 0 otherwise), gender of the spouse (dummy variable =1 if spouse 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.

A review of literature did not yield studies that 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 Recentered Influence Function (RIF) regression to gain insights into assortative mating and income inequality using individual level data.

Recentered Influence Function

The second part of the analysis involves teasing out the association between positive assortative mating and 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).

This study adopts a relatively new parametric technique, the recentered influence function (RIF) regression, which estimates the impact of changes in the distribution of the explanatory variables on quantiles (or other functionals like Gini or variance) of the unconditional (marginal) distribution of an outcome variable (Firpo et al 2009). The advantage of an unconditional regression is in enabling interpretation of the estimated relationship in a broader context on the distribution of the outcome variable. Firpo et al (2009) explains this brilliantly comparing the conditional and unconditional quantiles in the context of the impact of unionization on wage dispersion. The difference between the conditional coefficients at the 90th and 10th quantile cannot be interpreted as indicative of the impact of unionization on the overall wage dispersion. Such interpretation is possible however with unconditional quantile regressions. The further advantage of the RIF regression is that it can be generalized to other distributional statistics such as the Gini and the variance of the outcome variable. Studies such as Kollamparambil (2019) and (Alejo et al. 2014) have used RIF to analyse inequality in various other contexts.

Hence, the RIF regression or the unconditional quantile regression is suitable to identify inequality determinants at the individual level and has been recently introduced in the analysis of income inequality (Alejo et al 2014, Sakellario 2012). RIF regression is similar to the standard OLS regression except that it replaces the dependent variable, per capita household income, with the recentred influence function , RIF (y:G) , where G is the distributional Gini / Variance of per capita household income in our analysis. The dependent variable hence can be interpreted as income inequality (estimated separately using Gini coefficient and Variance).

$$E[RIF(Y;G)|X]=X'\beta \qquad (6)$$

Then, by the law of iterative expectations:

$$G=E(RIF(y;G))=E_x[E(RIF(y;G)|X]=E(X)'\beta \quad (7)$$

Where X indicates the matrix of explanatory variables including the positive assortative mating index and other control variables like the combined years of education of spouses/education of the household head, cohort dummies, interaction between cohort dummies and the education variable, race dummy (African =1, 0 otherwise), geographical region dummy (Rural =1, 0 otherwise), sex of the household head (Male household head=1, 0 otherwise), employment status of spouse (Spouse employed =1, 0 otherwise) and household size. Separate estimations are undertaken with squared term of PAM index included to the baseline regression to comprehend the existence of non-linearity in the relationship between positive assortative mating and income inequality.

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 income inequality.

Measuring Positive Assortative Mating

The key challenge in this analysis is to construct a variable to capture the degree of assortative mating. Gehlan and Lang(2016) has highlighted the sensitiveness of findings based on the measure. We therefore undertake two classifications, first, with total years of education that results in 23 categories and second, with eight education level categories (no education, primary education, middle school education, high school education, matriculation, technical education, undergraduate university degree, post graduate and higher degrees), as a sensitivity test. As is noted in the descriptive statistics, education levels have significantly increased among the younger cohorts. Therefore, in order to measure the level of assortative mating, we first normalise the education variable for both the household head and spouse, which results in the education variables being range bound between 0-1. The difference in the normalised education measure can be written as:

$$EduGap_i = |E_i^s - E_i^h| \quad (8)$$

This is further refined in the form of a positive assortative index below:

$$PAMIndex_i = 1 - |E_{ni}^s - E_{ni}^h| \quad (9)$$

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. PAM index is included in estimations in its squared form as well to check for non-linear relationship with income inequality.

4. Data

Education variable, when considered in years of education, is seen to vary from 0 to 23 years; and when classified in levels, ranges from 0 to 8 (Table 1). 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 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.

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
Education gap (years, normalised)	0.16	0.15		0.87
Head Education (level)	4.0	2.24	0	8
Spouse Education (level)	3.9	2.21	0	8
Education gap (level)	1.28	1.21	0	7
Education gap (level, normalised)	0.17	0.18	0	1
Household size	5.03	2.48	1	19
Binary Variables				
Rural	0.38		0	1
Male head	0.71		0	1
African	0.72		0	1
Spouse employed	0.61		0	1
Head employed	0.71		0	1
Spouse economically active	0.70		0	1
Head economically active	0.79		0	1

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. Income inequality is also seen to follow an inverted “U” shape with Gini being lowest in the youngest and oldest cohorts. 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 as indicated by the percentile ratios, nevertheless the average PAM index is substantially higher for younger cohorts as compared to the older ones.

Table 2:A cohort-wise summary of key variables of interest					
Age cohort (years), observations	Head Education years, Normalised	Spouse Education years, Normalised	Head Education level, Normalised	Spouse Education level, Normalised	Household Income (Rands)
cohort 1 (60 >), 559	6.61, 0.289 (5.94, 0.26)	6.90, 0.299 (5.91, 0.26)	2.86, .35 (2.28, 0.28)	2.92, 0.36 (2.27, 0.28)	5961.05 (7569.95)
cohort 2 (51-60), 590	8.26, 0.359 (6.00, 0.26)	8.24, 0.358 (5.98, 0.25)	3.50, 0.44 (2.42, 0.28)	3.48, 0.43 (2.24, 0.28)	10272.56 (38760.46)
cohort 3 (41-50), 527	10.88, 0.47 (5.51,0.24)	10.80, 0.47 (5.64, 0.25)	4.42, 0.55 (2.02, 0.25)	4.43, 0.55 (2.03, 0.25)	9214.93 (16512.46)
cohort 4 (31-40), 397	12.64, 0.55 (4.72, 0.21)	12.02, 0.52 (4.74, 0.21)	5.04, 0.63 (1.74, 0.22)	4.82, 0.60 (1.74, 0.22)	7388.6 (10066.62)
cohort 5 (21-30), 127	12.71, 0.55 (3.86, 0.17)	12.05, 0.52 (3.89, 0.17)	5.06, 0.63 (1.51, 0.19)	4.80, 0.60 (1.45, 0.18)	6300.21 (6848.16)
Total 2200	9.72, 0.423 (6.00, 0.26)	9.47, 0.412 (5.97, 0.23)	4.00, 0.50 (2.24, 0.28)	3.9, 0.49 (2.2, 0.28)	7924.45 (20255.78)

Source: Author calculation based on weighted NIDS data. Standard Error in parentheses

Table 3: A cohort-wise income inequality, PAM, Head-spouse education correlation					
Age cohort (years)	Gini	Percentile ratios (90/10, 90/50, 10/50)	PAM index: years, level	Pearson*** (year, level)	Spearman *** (year, level)
cohort 1 (60 >)	0.587	12.15, 4.35, 0.36	0.627, 0.676 (0.361, 0.374)	0.655, 0.633	0.665, 0.645
cohort 2 (51-60)	0.737	15.52, 4.66, 0.3	0.667, 0.712 (0.339, 0.347)	0.631, 0.585	0.645, 0.602
cohort 3 (41-50)	0.591	16.86, 4.6, 0.27	0.766, 0.800 (0.275, 0.267)	0.576, 0.543	0.576, 0.539
cohort 4 (31-40)	0.584	16.99, 4.52, 0.27	0.832, 0.851 (0.194, 0.183)	0.468, 0.443	0.488, 0.449
cohort 5 (21-30)	0.567	18.93, 4.71, 0.25	0.880, 0.885 (0.137, 0.137)	0.434, 0.395	0.477, 0.403
Total	0.634	16.34, 4.69, 0.29	0.732, 0.767 (0.306, 0.305)	0.652, 0.623	0.674, 0.6437

Source: Author calculation based on weighted NIDS data. Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

A non-parametric preliminary analysis finds strong support for positive assortative mating in South Africa across age cohorts using Pearson correlation and Spearman's rank correlation (Table 3). The correlation between the years of education of spouses is seen to be highest in older cohorts and although still high and significant at 99 percent confidence level, the strength of positive association is seen to reduce consistently among younger cohorts using both measures of correlation.

5. Econometric Results

Assortative mating in SA

Education of spouses are seen to be positively associated at very high confidence level indicative of positive assortative mating using both level (models 3 & 4) and year (models 1 & 2) measures of education (Table 4). 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 as compared to the bench of older cohorts (over 60 years of age). 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 spouses of male heads of household have higher education as compared to spouses of female heads. All results are consistent across normalized measurement of education using years as well as levels of completed education.

Table 4: Robust Ordinary Least Squares regression: Dependent variable, Spouse's Education

VARIABLES	(1) Years#	(2) Years#	(3) Level#	(4) Level#
Head education	0.590***	0.648***	0.553***	0.616***
	(0.0162)	(0.0239)	(0.0166)	(0.0242)
Head Education *cohort2		-0.0530		-0.0687*
		(0.0383)		(0.0391)
Head Education *cohort3		-0.0814*		-0.0951**
		(0.0421)		(0.0435)
Head Education *cohort4		-0.206***		-0.201***
		(0.0523)		(0.0543)
Head Education *cohort5		-0.285***		-0.298***
		(0.103)		(0.103)
cohort2	0.00591	0.0253	0.0100	0.0402**
	(0.0102)	(0.0172)	(0.0112)	(0.0205)
cohort3	0.0454***	0.0777***	0.0598***	0.105***
	(0.0108)	(0.0215)	(0.0118)	(0.0256)
cohort4	0.0541***	0.157***	0.0678***	0.182***
	(0.0120)	(0.0296)	(0.0132)	(0.0350)
cohort5	0.0651***	0.211***	0.0723***	0.249***
	(0.0188)	(0.0588)	(0.0207)	(0.0676)
African	-0.0107	-0.00868	-0.0165*	-0.0142
	(0.00903)	(0.00902)	(0.00993)	(0.00991)
Rural	-0.0379***	-0.0381***	-0.0437***	-0.0438***
	(0.00862)	(0.00860)	(0.00947)	(0.00944)
Head male	0.0189**	0.0190**	0.0177*	0.0187*
	(0.00870)	(0.00869)	(0.00956)	(0.00955)
Constant	0.152***	0.130***	0.202***	0.173***
	(0.0136)	(0.0153)	(0.0154)	(0.0176)
Observations	2,555	2,555	2,555	2,555
R-squared	0.442	0.447	0.408	0.413

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1
Results 1) and 2) use the normalised education years, while 3) & 4) use normalised education levels

Having confirmed significant evidence of positive assortative mating in South Africa, we turn to looking at its association with income inequality in the next section.

Assortative mating and Inequality

The unconditional quantile regression results (Table 5) provide insights into the non-monotonic relationship between PAM and income distribution. At lower income levels (10th quantile) PAM has a positive and significant impact on incomes. The relationship stays significant and positive but lessens in magnitude at the 50th quantile and at the 90th quantile is seen to be negative and significant. Since these are unconditional quantile estimations, we are able to interpret the difference in the impact of PAM between the quantiles as the equalizing effect of PAM across the entire income distribution. Further we are able to quantify the contribution of PAM to overall income dispersion as -2.6 based on the difference between the coefficients of the 90th and 10th quantile.

Table 5: Unconditional quantile regressions (UQR)-Per capita household income

VARIABLES	(1) OLS	(2) UQR _10	(3) UQR_50	(4) UQR_90
PAM (years)	0.511*	1.979**	0.611**	-0.612*
	(0.270)	(0.915)	(0.265)	(0.361)
Education years	0.0968***	0.0424***	0.103***	0.136***
	(0.00518)	(0.0110)	(0.00636)	(0.0108)
PAM index* cohort1	-0.458	-0.754	-0.759**	0.283
	(0.326)	(1.108)	(0.356)	(0.431)
PAM index* cohort2	-0.648**	-2.048**	-0.445	0.250
	(0.306)	(1.014)	(0.315)	(0.407)
PAM index* cohort3	-0.423	-1.797*	-0.560	0.441
	(0.320)	(1.027)	(0.352)	(0.417)
PAM index* cohort4	-0.0279	-0.113	-0.459	1.128**
	(0.407)	(1.319)	(0.474)	(0.565)
PAM index* cohort5	1.117	-1.651	1.467	1.331
	(0.816)	(1.505)	(1.016)	(1.418)
cohort1	0.384	0.420	0.522*	0.316
	(0.262)	(0.936)	(0.285)	(0.312)
cohort2	0.718***	1.723**	0.588**	0.314
	(0.252)	(0.871)	(0.256)	(0.306)
cohort3	0.463*	1.824**	0.464	-0.156
	(0.268)	(0.892)	(0.297)	(0.311)
cohort4	-0.0673	0.237	0.243	-0.996**
	(0.346)	(1.157)	(0.407)	(0.450)
cohort5	-1.083	1.676	-1.341	-1.250
	(0.725)	(1.377)	(0.917)	(1.231)
African	-0.307***	-0.405***	-0.151**	-0.478***
	(0.0542)	(0.0993)	(0.0704)	(0.104)
Rural	-0.286***	-0.601***	-0.256***	-0.0527
	(0.0551)	(0.138)	(0.0738)	(0.0823)
Head male	0.646	-0.471*	0.736	1.882***
	(0.515)	(0.263)	(0.610)	(0.699)
Spouse employment	-0.423***	-0.570**	-0.417***	-0.489***
	(0.0929)	(0.261)	(0.129)	(0.121)
Spouse male	0.521	-0.430	0.545	1.664**
	(0.515)	(0.264)	(0.611)	(0.700)
Household size	-0.160***	-0.177***	-0.156***	-0.134***
	(0.0109)	(0.0324)	(0.0133)	(0.0166)
Constant	6.054***	5.084***	5.744***	6.686***
	(0.568)	(0.869)	(0.657)	(0.763)
Observations	1,794	1,794	1,794	1,794
R-squared	0.397	0.127	0.286	0.204

Standard errors in parentheses, *** p<0.01, ** p<0.05, * p<0.1,
UCR: Unconditional Quantile Regression, Education variable is normalised.

Table 6: Recentred influence Function regression-Per capita household income

VARIABLES	(1) Variance	(2) Gini	(3) Variance	(4) Gini	(5) Variance	(6) Gini
PAM #	-1.847**	-0.0631***	-3.726***	-0.172***	-4.427***	-0.133***
	(0.726)	(0.0220)	(0.851)	(0.0349)	(1.017)	(0.0309)
PAM squared			2.976***	0.0968***	3.529***	0.107***
			(0.769)	(0.0243)	(0.798)	(0.0242)
Education years	0.100***	0.00172***	0.0566***	0.00186***	0.104***	0.00182***
	(0.0140)	(0.000423)	(0.0206)	(0.000423)	(0.0139)	(0.000422)
PAM index* cohort1	0.764	0.0239	-0.00353	0.0408	-0.0755	-0.0102
	(0.878)	(0.0266)	(0.0225)	(0.0268)	(0.758)	(0.0230)
PAM index* cohort2	2.007**	0.0641**	0.0531***	0.0765***	1.038	0.0260
	(0.825)	(0.0250)	(0.0175)	(0.0251)	(0.688)	(0.0209)
PAM index* cohort3	1.314	0.0447*	0.0378**	0.0491*	0.0350	-0.00258
	(0.863)	(0.0262)	(0.0173)	(0.0261)	(0.714)	(0.0216)
PAM index* cohort4	2.041*	0.0558*	0.0635***	0.0433	0.144	-0.0104
	(1.096)	(0.0332)	(0.0222)	(0.0332)	(0.976)	(0.0296)
PAM index* cohort5	2.406	0.0554	0.0486	0.0309	0.185	-0.0208
	(2.199)	(0.0667)	(0.0433)	(0.0667)	(2.152)	(0.0653)
cohort1	0.0812	0.00269	0.651	-0.00862	0.810	0.0317*
	(0.707)	(0.0214)	(0.491)	(0.0215)	(0.605)	(0.0184)
cohort2	-1.145*	-0.0422**	-0.736*	-0.0509**	-0.330	-0.0106
	(0.679)	(0.0206)	(0.432)	(0.0206)	(0.567)	(0.0172)
cohort3	-1.172	-0.0413*	-1.002**	-0.0447**	-0.126	-0.00267
	(0.721)	(0.0219)	(0.455)	(0.0218)	(0.604)	(0.0183)
cohort4	-1.914**	-0.0530*	-1.810***	-0.0420	-0.331	0.00208
	(0.933)	(0.0283)	(0.603)	(0.0283)	(0.842)	(0.0255)
cohort5	-2.228	-0.0517	-1.416	-0.0305	-0.361	0.0122
	(1.954)	(0.0592)	(1.128)	(0.0592)	(1.922)	(0.0583)
African	-0.0407	0.00304	-0.0496	0.00264	-0.0376	0.00321
	(0.146)	(0.00443)	(0.145)	(0.00441)	(0.146)	(0.00442)
Rural	0.400***	0.0176***	0.432***	0.0175***	0.404***	0.0178***
	(0.149)	(0.00450)	(0.148)	(0.00448)	(0.148)	(0.00448)
Male Head	1.215	0.0294	0.974	0.0271	1.145	0.0273
	(1.388)	(0.0421)	(1.400)	(0.0419)	(1.383)	(0.0419)
Spouse employed	-0.311	-0.00334	-0.239	-0.00150	-0.262	-0.00196
	(0.250)	(0.00759)	(0.249)	(0.00758)	(0.250)	(0.00758)
Spouse male	1.066	0.0261	0.827	0.0235	0.967	0.0231
	(1.389)	(0.0421)	(1.401)	(0.0419)	(1.384)	(0.0420)
Household size	0.0774***	0.00452***	0.0814***	0.00461***	0.0803***	0.00461***
	(0.0294)	(0.000892)	(0.0295)	(0.000888)	(0.0293)	(0.000889)
Constant	0.357	0.0746	0.513	0.0940**	-0.0782	0.0543
	(1.531)	(0.0464)	(1.464)	(0.0465)	(1.493)	(0.0453)
Observations	1,794	1,794	1,794	1,794	1,794	1,794
R-squared	0.050	0.059	0.063	0.068	0.057	0.066

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

[#] PAM variable in models (1) - (4) is based on normalised years of education and for (4) - (6) is based on normalised levels of education.

The first and second RIF estimations uses variance and the Gini coefficient measures of income inequality respectively. The PAM index is calculated based on the normalized years of education and is seen to have a significant negative impact on income inequality in both estimations (Table 6). Therefore, the indication is that an increase in positive assortative mating is associated with a decline in income inequality. These results differ from the findings of Eika et al (2017), Hryshka (2015) and Greenwood et al (2014) in US and other developed countries, indicating that results of the relationship differ in countries with high income inequality and low education levels.

However, the relationship between PAM index and income inequality is more nuanced as indicated by the non-linear relationship revealed in our estimations three to six (Table 6). As expected in a context of high levels of poverty and inequality, 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 as hypothesized in section two.

The education status of the household, measured as the combined education of spouses as well as the education of the household head, is seen to have a positive and significant impact on income inequality across age cohorts. This provides strength to the conclusion that disparity in education contributes to inequality in South Africa.

Other findings emanating from the result are that income inequality is higher among rural areas compared urban areas and that bigger household size contributes to reducing income inequality. The sex of the head of the household and the employment status of the spouse, are not seen to contribute significantly to household inequality when controlled for other variables. Leibbrandt et al. (2012) indicate that earnings inequality within the African population group is highest and has increased the most over time however our results do not find African households to have higher inequality compared to other races when controlled for positive assortative mating and other variables.

6. Conclusion.

Literature has highlighted the phenomenon of positive assortative mating in the context of 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 posit a non-linear relationship between positive assortative mating and income inequality in global literature.

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.

Initial investigations using simple correlation statistics found strong evidence for positive assortative mating in South Africa but finds the strength of this association to be weakening among younger 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. Siow (2015) also did not find more positive assortative matching in 2000 compared to 1970.

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 household per capita income inequality and years of education based assortative mating using multiple measures of inequality as well as education based positive assortative mating. An increase in positive assortative mating at lower levels of assortative mating is found to be associated with a reduction in income inequality; while an increase in positive assortative mating at higher levels of assortative mating is found to be associated with an increase in income inequality.

In the absence of studies in the 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. Our study also found a positive and significant impact of combined education of the spouses on income inequality. However, the increase in inequality in the South African context cannot be pinned on positive assortative mating among the younger age cohorts. 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. Our findings hence differ from those 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. However, unlike in developed country contexts, we observe in South Africa a negative association between positive assortative mating and income inequality at lower levels of PAM. The relationship however becomes positive at higher levels of PAM. The resulting U-shaped relationship is explained by the highly unequal South African society where initial increase in PAM is the result of positive assortative mating at lower levels of the distribution contributing to a decline in inequality. Further increase in PAM requires assortative mating at higher ends of the distribution resulting in increased inequality explaining the U-shaped relationship in the South African context.

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 for future research. Nevertheless, this study is a significant contribution in providing a benchmark study for further research on the subject.

References

Alejo, J., Gabrielli, M., & Sosa-Escudero, W. (2014). The Distributive Effects of Education: an Unconditional Quantile Regression Approach. *Revista de Análisis Económico – Economic Analysis Review, 29*(1), 53-76.

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.

Chiappori P. A., Dias, M. C., & Meghir, C. 2018. "The Marriage Market, Labor Supply, and Education Choice," *Journal of Political Economy*, vol. 126(S1), pages 26-72.

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.

Firpo, S., Fortin, N. and Lemieux, T. (2009), Unconditional Quantile regressions, *Econometrica*, Vol 77, no.3.

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.
Kollamparambil, U. (2019) Happiness, Happiness Inequality and Income Dynamics in South Africa, *Journal of Happiness Studies*, Forthcoming. DOI: https://doi.org/10.1007/s10902-019-00075-0
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