Inheriting the future:

The role of family background and neighbourhood characteristics in child schooling outcomes in South Africa

Justine Burns

DRAFT

30 October 2001

(All comments most welcome. The author can be contacted via e-mail: keswell@mweb.co.za)

Overview

The contribution of this paper will be to investigate the extent of intergenerational transmission of education status between parents and their children, as well as the relative importance of family background to neighbourhood effects on child schooling attainments in the province of KwaZulu-Natal in South Africa. There are few, if any, such empirical estimates for South Africa. What makes this an especially interesting question is that race and space have a particular relationship in South Africa (Dawood, 1994). This is a country in which neighbourhoods and communities were largely constructed and enforced through political and economic interventions by the Apartheid State. Influx control and the Group Areas Act, combined with racial disparities in state expenditures on social services, promoted and enforced racially homogenous and geographically distinct communities or neighbourhoods, with radically different socio-economic outcomes.

Intergenerational transmission of educational status

Much of the literature on intergenerational mobility has focused on the transmission of economic status, as measured by income. However, education is a major mechanism through which intergenerational social mobility is also affected, and may play an important complementary role in levelling an uneven playing field. Not only is education likely to have a positive impact on the chances of upward occupational mobility (Gazioglu, 1994), but it also raises the opportunity for upward income mobility. Thus, while much of the previous literature has focussed on the role of education in income mobility, there are good reasons to focus on educational mobility as an outcome on its own, especially in South Africa. First is the concern over the *processes* that may generate unequal outcomes over time, of which inherited status may be a feature. Furthermore, there is a strong correlation between the level of education and the standard of living. The poverty rate for individuals with no education is 69% compared with 54% for individuals with primary education, 24% for those with secondary education, and 3% for those with tertiary education. (PIR, 1995). Second, to the extent that one *is* concerned with income mobility, intergenerational schooling correlations may provide an upper bound to the true earnings correlation because parental tastes and wealth influence child's schooling much more than their adult earning capacities. (Behrman et al, 1980, (Haveman, 1995))[1] Moreover, key studies by Blau and Duncan (Blau, 1967) and Featherman and Hauser (Featherman, 1978) suggest that educational attainment is the main determinant of occupational status and that the educational and occupational status of fathers affects their son's occupational attainment primarily via the son's education.[2]

[1] A controversial study by Jencks et al **Jencks, C. et al.** *Inequality:A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books, 1972. argued that factors other than family characteristics, schooling and genetic inheritance, (they termed it "luck") explained more than half of the variation in socioeconomic status of an individual's occupation, and of this, three quarters was explained by income. Hence, they concluded that in USA, schooling had only a minor effect in reducing economic inequality.

[2] These studies focussed on intergenerational transmission between American men, notably father and son pairs.

Furthermore, schooling gap ratios (the measure which will be used in this paper) are important from an egalitarian point of view. Negative externalities affecting the acquisition of education may arise from being a member of a discriminated against group or community. (Datcher, 1982) To the extent that schooling gaps may be a function of factors other than natural endowments, one should be concerned with why it is that some children enjoy a normal schooling trajectory, while others start school late, repeat grades, or leave school early, particularly where such explanations correlate with race, class, income and location. More importantly, in the context of high unemployment and minimal job creation in South Africa, it is clear that education will become increasingly important as a signal to potential employers, and in generating livelihoods, thereby distinguishing the non-poor from the poor. Recent work by Hertz (2001) and Lam (2000) confirms that multiple equilibria exist in the returns to schooling, and that there are significant thresholds in terms of years of schooling that individuals need to reach in order to earn a decent wage. Hertz (2001) argues that the rate of return reaches its minimum at between 5 and 6 years of education, and increases thereafter until at least the 14^{th} year of education. Individuals with large schooling gaps are likely to find themselves below these thresholds, and thus prone to a poverty trap existence.

The impact of family background on child educational attainment

Numerous studies document the association between family background, parental schooling and the schooling of children. (Chase-Lansdale and Brooks-Gunn, 1994; Furstenberg, BrooksGunn, and Morgan, 1987; Garfinkel and McClanahan, 1986, Hauser and Featherman 1977, Huston, 1991, Sewell and Hauser 1975; Behrman (1997) and Behrman and Knowles (1999) are good survey articles). Mulligan, for example, (Mulligan, 1999) finds that across eight sets of estimates[3], the intergenerational correlation co-efficient on education attainment ranges from 0.14-0.45, averaging at 0.29. These studies almost always find a significant positive association between child's schooling and parental education, with mother's education being about 10% more important than father's education at the median of estimates that include both.[4](Haveman, 1995),(Hill, 1995),(Schultz, 1993, Case, 1991). Estimates for South Africa (Burns, 2000) largely accord with these trends.[5](See Appendix A) Key studies by Blau and Duncan (Blau, 1967) and Featherman and Hauser (Featherman, 1978) argue that the educational attainment of American men is the main known determinant of their occupational status, and that the educational and occupational status of father's affects their son's occupational attainment primarily via the son's education. Finally, parental completion of high school and one or two years of tertiary schooling typically has a larger effect on children's schooling than parental years of education beyond that, suggesting the presence of non-linearities.

---

[3] These estimates are based on data for the US, Germany, Malaysia, and Kalamazoo, Michigan.

[4] A notable exception is the study by Behrman and Taubman (1985) that estimated that in the USA, the impact of father's education on offspring education is larger than mother's education.

[5] These kinds of empirical results have been used to argue that a stronger case exists, *ceteris paribus*, for subsidies for female education as opposed to male education.

However, these kinds of simple OLS estimates are subject to upward bias, in that they do not control for unmeasured heritable traits and assortative mating, thus ignoring the possibility of intergenerational "ability" bias.[6] (Behrman, 2001) However, even in studies where intellectual ability has been included, the influence of father's education on son's education has maintained its relative position as the most important of parental-background influences. (Sewell, 1980) Furthermore, Bowles and Gintis (Bowles, 2001) argue that the genetic inheritance of traits contributing to the cognitive skills measured on IQ and related tests explain less than one twentieth of the intergenerational transmission of economic status. However, their study focuses on the transmission of income status, as opposed to educational attainment, and it is unclear whether the genetic inheritance of cognitive skills would play a larger role in the latter case. To the extent that these sources of bias are not corrected for, the OLS estimates should be treated with caution. Indeed, Behrman and Rosenzweig (Behrman, 2001), using data on twins, find that once heritability is controlled for, women's schooling has a marginally negative impact on child's schooling attainments.[7]

Empirical studies also document significant positive associations between household income and child's education status. Furthermore, the source of income may matter for child attainments. Earned income typically has a positive and significant effect on children's schooling attainments, while welfare income tends to have a small, sometimes negative impact on schooling. (Hill, 1987) However, income measures, in the absence of multiple observations, are subject to measurement error, as they provide a measure of permanent income with error. Moreover, such measures provide no information on the intra-household allocation of income, nor of the parental time devoted to children in the household.

There is also evidence that household structure affects outcomes. Blake (Blake, 1985) for example, argues that the influence of father's education on son's educational attainment is conditional on, and inversely related to the number of siblings. Furthermore, she finds

[6] There are two fundamental problems with these kinds of cross-sectional estimates. First of all, the co-efficient may just reflect ability bias, i.e. more able mothers obtain more schooling, and, if this ability is genetically transferred to their children, have more able children who also obtain more schooling. Furthermore, even among mothers with the same abilities, those with higher levels of schooling may have children with greater academic and labour market performance due to assortative mating. More educated women marry more educated men. **Behrman, J.R.; and Rosenzweig, M.R.** "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" *American Economic Review (forthcoming)*, 2001.

[7] They attribute this to the increased labour force participation by more educated women, thereby reducing time spent in the home with children. However, it is unclear why increased participation by women in the workforce should necessarily result in declining educational attainment by children, especially if the additional income generated is used to purchase substitute goods such as child care, and kindergarten. Furthermore, Hill and Stafford **Hill, C.R. and Stafford, F.** "Allocation of Time to Pre-School Children and Educational Opportunity." *Journal of Human Resources*, 1974, *9*(Summer), pp. 323-41. show that the quantity of time devoted by parents to children is positively related to parental education. Also, parental education suggests that parents value education, and would hold this to be so for their children. Behrman and Rosenzweig acknowledge that their result may be context-specific, and reference Behrman et al (1999) which reports strong evidence from rural India, where more educated women did not appreciably increase their labour force participation. Hence, they suggest that anticipating the consequences of investments in women's education requires attention to the role that schooling plays in marriage markets as well as in labour market opportunities for women.

non-linearities in educational mobility, with men with small and medium numbers of siblings experiencing high mobility, while men with large numbers of siblings experiencing very low educational mobility. Finally, the evidence concerning the impact of race is mixed. While a large number of sstudies (Haveman, 1995) find that race is not associated significantly with educational outcomes when family income and other background characteristics are controlled for, Altonji (Altonji, 1988) and Case et al (Case, 1991) find evidence that blacks get more schooling than whites with similar family background characteristics.

Neighbourhood effects and child schooling outcomes

There has also been much attention devoted to the impact of neighbourhoods on socio-economic outcomes, in addition to controlling for family background. The most obvious reason for including neighbourhood variables in such an investigation is that failure to do so may lead to a misleading picture in which the size of family background effects is overestimated. Cooper suggests that the typical intergenerational correlation regression equation has been fundamentally mis-specified by omitting the role of neighbourhood characteristics in the determination of children's attainments. (Cooper, 1995) Moreover, given South Africa's spatial legacy, it is reasonable to suspect that neighbourhoods play an important role in social outcomes, and policy debates need to be re-oriented accordingly. For example, if neighbourhood effects do have a relatively large impact on child schooling outcomes, then the argument in favor of investing in poor communities is strengthened by the evidence that not only would such investment improve standards of living in general (with better services and infrastructure) but would also have positive indirect impacts on educational outcomes.

Durlauf (Durlauf, 1997) characterises the recent work on neighbourhood effects as a memberships theory of inequality, in which individual outcomes are strongly influenced by one's membership in various groups, be they schools, firms, neighbourhoods, or even racial groups. Group members interact positively with one another with the result that group level interactions generate common group outcomes across group members. In effect, the characteristics of the groups to which the individual belongs condition the range of the individual's life economic prospects. (Durlauf, 1997) Furthermore, these interdependencies are not compensated or enforced by the market. Durlauf (Durlauf, 1997) argues that social stratification by income, race, education or any other factor, will enhance group differences, resulting in greater cross-sectional inequality, and declining social mobility. The combination of strong group effects and endogenous group formation can exacerbate an already unequal income distribution and lead to persistent inequality across generations.

Typically, the impact of neighbourhoods on individual outcomes has been ascribed to one of four general models. These include the contagion model, which emphasises the influence of peer effects; competitive models, in which neighbours compete with one another for scarce resources; models of relative deprivation, where individual attainments depend on the individual context relative their neighbours; and models of collective socialisation, which emphasise the importance of role models and monitoring as part of a child's socialisation process. (Jencks, 1990) Under the contagion and collective socialisation models, living in an affluent neighbourhood confers benefits on children, especially lower-income children. By way of contrast, models of relative deprivation and competition imply that living in an affluent neighbourhood holds negative outcomes for these same children.

Recent theoretical work suggests that with endogenous neighbourhood formation and local financing of education, stratification by economic status will occur.(Loury, 1976, Durlauf, 1993a, 1993b, Benabou, 1992a). If neighbourhoods are endogenously formed, and comprise economically homogenous groups, then children who grow up in a community with a high level of income and educational attainment may benefit more from their community environment than children from an identical family growing up in a less favorable community environment. This could be attributed to peer effects, where rich neighborhoods are characterized by high human capital investments while poor neighborhoods have low human capital investment (Benabou, 1992a), or network effects, under which multiple equilibria may exist in the distribution of occupations within neighbourhoods if one's success in the labour market depends on referrals by neighbourhood members. (Montgomery, 1991) Finally, Streufert (Streufert, 1991) argues that the absence of role models in poor communities prevents children from accurately assessing the marginal product of effort in school. Consequently, the intergenerational transmission of status from child to parent will differ across neighbourhoods.

The empirical literature tends to find that income and occupational status of the neighbourhood is positively associated with children's schooling outcomes. (Haveman, 1995, Hill, 1987, Jencks, 1990, Corcoran, 1992). After controlling for family background, Datcher (Datcher, 1982) finds that a 10% increase in zip-code area income raises educational attainment of men by one-tenth of a school year for both blacks and whites. Similarly, Brooks Gunn et al (Brooks-Gunn, 1993) find that even after controlling for family effects, neighborhood effects on school leaving remain significant. In particular, they find that the fraction of households in the neighbourhood with annual incomes of $30,000 or more has a significant negative effect on school dropout rates, and concur with Crane (Crane, 1991) that living in a neighbourhood with very few professional or managerial workers is associated with higher rates of early school leaving.[8]

Typically, studies of neighbourhood effects argue that the racial composition of a neighbourhood is insignificant, once family and other background variables have been accounted for, (Brooks-Gunn, 1993), although there are some exceptions. While Jencks et al (Jencks, 1990) found race had no significant effect on educational attainment after controlling for background variables, Corcoran et al (Corcoran, 1992) found a significant black advantage over white, while Massey et al (Massey, 1992) argue that residential

[8] While Crane (1991) interprets his results as evidence of contagion, Brooks-Gunn et al (1993) argue that given the non-significance of neighbourhood poverty and male joblessness, and the significance of affluent neighbours in the PSID data, their results are more consistent with the social control theories that stress the importance of role models.

segregation and poverty are key dimensions of social structure that affect personal behaviour. Sewell et al (Sewell, 1980) find that gender affects the pattern of schooling attainment. Finally, Case and Katz (Case, 1991) find a large degree of spatial correlation across neighbourhoods on child socio-economic outcomes, even after controlling for family background and individual characteristics.

Social capital may be important too. Putnam [Putnam, 1995 #8:296] finds that US states with high social capital scores[9] are the states where children do well, particularly in educational and health outcomes, even after controlling for factors such as racial composition, state affluence, economic inequality, adult illiteracy and schooling quality. Indeed, Putnam finds that social capital is the single most important predictor of standardized test scores for students, and is especially important in preventing children from dropping out of school. Furthermore, the education level of the adult population in the neighbourhood does not, in general, have a significant independent effect on child educational outcomes after poverty, social capital and demographics are controlled for, suggesting that "social capital may be most crucial for families who have fewer financial and educational resources" [Runyan, 1998 #35]. Possible explanations for this result could be that in communities with high civic engagement, parents are more actively involved in, and supportive of, their children's schooling activities. Coleman [Coleman, 1988 #4] argues that Catholic school success is not because of the individual attributes of the students, but rather the social structure in which the school is embedded, whereby the parents of students form relationships with each other both through PTA meetings and church activities. [10] Putnam [Putnam, 1995 #8:306] concludes that "…precisely what many high achieving suburban school districts have in abundance is social capital, which is educationally more important even than financial capital….where social connectedness is lacking, schools work less well, no matter how affluent the community."

However many of these studies assume a linear relationship between outcome variables, and family and neighborhood level regressors, and that parental and neighbourhood characteristics are additive. Yet, there is evidence that these relationships are nonlinear. Crane (Crane, 1991) finds evidence of non-linearities in the relationship between school dropout rates and the number of professional or managerial workers in the neighbourhood, and attributes this to a contagion effect. Using regression tree analysis, Cooper et al (Cooper, 1995) find that families which reside in relatively affluent or poor communities exhibit more persistence in economic status than families who reside in communities near the middle of the income distribution. Cooper at al (Cooper, 1995) conclude that there is substantial evidence of non-linearities across neighbourhoods, and that neighborhood income appears to affect the persistence of a family's economic status across generations

[9] Social capital provides a measure of whether state residents trust each other, join organisations, volunteer, vote and socialise with friends. Putnam finds that the correlation between state social capital and positive child development is 0.8.

[10] For contrary evidence, see Stephen L. Morgan and Aage, B. Sorenson: A Test of Coleman's Social Capital Explanation of School Effects." American Sociological Review 64, 1999, pg 661-681

Finally, while these studies argue that neighbourhoods are important, many fail to address the question of *how* important neighbourhoods are, relative to family background. Corcoran et al (Corcoran, 1992) argue that with the exception of the percentage of households receiving welfare assistance in the zip-code area, community level variables exhibit relatively little influence on men's economic outcomes. Similarly, Crane (Crane, 1991) argues that the available evidence suggests that neighbourhood effects on educational attainment are quite small. However, both sets of authors acknowledge that these small co-efficients could be caused by the crude level of aggregation using zip-code data, which may not proxy well for neighborhoods. Mis-defining the neighbourhood may bias estimates downward by a large amount because the error in measuring neighbourhoods introduces randomness. Datcher (Datcher, 1982) by way of contrast, argues that in general, neighbourhood characteristics are at least as important as family characteristics in explaining the gaps between black and white achievement in earnings. In specific relation to education outcomes, her methodology[11] suggests that changing the neighbourhood income and racial composition for the bottom quartile would account for 38% of the resulting increase in education, while family background variables would account for the remaining 62%.

Methodological problems in neighbourhood studies

Studies of neighbourhood effects are plagued by numerous problems, which suggest that estimates of neighbourhood effects should be treated circumspectly. The level of aggregation used to define a neighbourhood in most studies is at the census tract or zip code level[12], and it is unclear how accurately these reflect neighbourhoods in any sense. (Haveman, 1995, Datcher, 1982, Jencks, 1990, Brooks-Gunn, 1993, Crane, 1991) The difficulty of definition may obscure or magnify the impact of neighbourhood effects. For certain types of effects, such as the presence of role models for children, using a census district measure may be too large. Alternatively, census district measures may be more appropriate if the transmission mechanism depends on the characteristics of the local labour market. Similarly, to the extent that neighbourhood effects relate to differences in economic conditions between urban, suburban and rural communities, census district measures may be good proxies. Mis-defining the neighbourhood may bias estimates downwards by a large amount because the error in measuring neighbourhoods introduces

[11] Datcher {**Datcher, L.** "Effects of Community and Family Background on Achievement." *Review of Economics and Statistics*, 1982, *64*, pp. 32-41.} provides an interesting method for identifying the role of background factors in explaining education and earnings differences within and between race groups. She makes use of co-efficient estimates from intergenerational education regressions, that include estimates on neighbourhood and family level characteristics, as well as a measure of the differences in the average value of a particular variable between the bottom and top quartile. Then, if a_{ij} is the coefficient of Z_i in the regression for race group j, and ΔZ_{ij} is the difference in the average level of Z_i between the bottom and top quartile of race group j, then $a_{ij}\Delta Z_{ij}$ is the estimate of the direct effect o changing Z_i for the bottom quartile group.

[12] In the USA, census tracts contain, on average 4,200 people, and are formed by local committees appointed by the Census Bureau to approximate locally perceived "real" neighborhood boundaries.

randomness (Crane, 1991). Indeed, Brooks-Gunn et al (Brooks-Gunn, 1993) find that when they go from the census tract to zip code level, the co-efficient on the effect of affluent neighbours on school dropout rates increases significantly. One study that manages to deal with this definitional problem is that of Case and Katz (Case, 1991) who are able to define neighborhoods at a very disaggregated level by utilizing information on exact street addresses for the majority of their respondents, assigning individuals to neighbourhoods that are approximately one of two square blocks in size.

The formal empirical literature on neighbourhood effects has also been criticised for its failure to deal with identification issues, such as the self-selection problem. (Manski, 1993b) Given the large degree of homogeneity across households within a neighbourhood, it becomes difficult to separate out the independent effects of family and neighbourhood economic circumstances on child attainment ((Evans, 1992); (Manski, 1993b). This selection process may impart unknown bias into the empirical estimates, which is difficult to measure. Including both the socio-economic position of the family and the neighbourhood in the regression may over-control for family effects if the family's socio-economic position is itself caused by the neighbourhood factors.

However, Brooks Gunn et al (Brooks-Gunn, 1993) argue that, with the exception of income, it is unlikely that there is a strong causal link between most family level characteristics and neighbourhood conditions. Characteristics such as maternal education are determined much earlier. And while measures such as family income and family structure may be influenced by current neighbourhood conditions, they argue that they are more the product of choices made and constraints faced long before the point at which the survey data is collected. Datcher (Datcher, 1982) argues that in order to minimise the probability that neighbourhood variables are simply reflections of unobservable family characteristics, one should test the sensitivity of neighbourhood co-efficients by controlling for additional family characteristics, which are aimed at capturing family unobservables. [13] If neighbourhood coefficients are not affected when these additional variables are included, she argues that this indicates that the neighbourhood co-efficients do in fact represent the effect of neighbourhood quality on individual outcomes.

Despite these problems, the empirical evidence certainly suggests that group interactions have important effects. If anything, previous studies may have under-estimated these effects through their failure to examine the presence of non-linearities arising from group interactions, and because multicollinearity between individual and group characteristics may have resulted in statistical insignificance of group effects, even if they were present (Durlauf, 1997). Finally, Corcoran et al (Corcoran, 1992) argue that previous studies have underestimated the impact of family and neighbourhood variables, mainly because of measurement error[14] and reliance on unrepresentative homogenous samples.

[13] In her study, she includes additional family information on whether the family received welfare income or not, whether parents expected their children to go to college or not, and other measures of parental efficacy.

[14] **Solon, G.R.** "Intergenerational Income Mobility in the United States." *American Economic Review*, 1992, *82*, pp. 393-408. and **Zimmerman, D.J.** "Regression toward Mediocrity in Economic Stature." Ibid., (409-429). have shown that the value of any of the standard variables used to measure family economic

Neighbourhoods in South Africa

South Africa's spatial legacy and artificial creation and enforcement of neighbourhoods provide an interesting context in which to think about the impact of neighbourhood effects on individual outcomes. *De facto* residential segregation existed in South Africa prior to its formalization by the apartheid state.[15] South African segregation involved the overt use of state power to dictate social, political and economic relations between racial groups through the control and allocation of the spatial environment. It was explicitly and intentionally designed to prevent the emergence of a black bourgeoisie. (Stadler, 1987). Repressive legislation confined African individuals to the poorly-resourced homelands, while urban residential segregation was class based, with the top and middle end of the property market being accessible to the wealthy, most of whom were white, while lower income areas comprised mixed race groups (Dawood, 1994). With the passing of the Group Areas Act in 1950, however, the State intervened to re-engineer even these urban communities, and large-scale removals occurred.[16] Under Verwoed in the 1950s, a large urban African population was housed and granted rights to remain in white cities, in separate locations, with no political rights. However, during the 1960s and 1970s, policy reverted to being more restrictive, with African workers being sent back to the homelands in an attempt to stem the growth in urbanization. Influx control was abolished in 1986, and policy efforts focused on resolving the urban housing crisis. Minimal shelter was provided for the (predominantly African) urban poor and site-and-service schemes and squatter settlements became increasingly common. Spatial divisions between the rich and poor became increasingly stark with the increasing differentiation between wealthy inner city dwellers and the growing shack settlements of poor in-migrants. Concurrently, the composition of African townships began to change with the provision of better quality housing for the emerging middle class. From the mid 1980s onwards, the strict enforcement of the Group Areas Act was relaxed, and the Act was finally scrapped in 1991. However, the territorial exclusion of African townships from the white city was not addressed until the passing of the Local Government Transition Bill in 1993 (Robinson, 1996). Hence, it is only since 1993 that we really see the opportunity for what Durlauf (Durlauf, 1997) calls "associational redistribution". Thus, while the market mechanism has been a factor in residential segregation in South Africa, the law has also had a large hand in ensuring the emergence of racially and socio-economically homogenous groups.

---

status in any one year is noisy, leading to measurement error, causing the estimated correlation co-efficient and OLS estimates to be downward inconsistent. A partial remedy to this is to average the data over a number of years, in an attempt to improve the signal to noise ratio.

[15] In the early 1900s, pressure was exerted on individuals living on the land, predominantly African, to pay taxes and rents, largely in a bid to get them off the land. The Glen Grey Act of 1894 imposed severe limits on the size and number of holdings that Africans could own. The 1913 and 1936 Land Acts dispossessed thousands of African people of their land, and prohibited African ownership of land outside the areas designated for such purposes.

[16] While the Group Areas Act made provision for the creation of African areas or groups, none were ever proclaimed because the work and residential rights of African people were still governed by a different set of laws. Still, African people were affected by the Group Areas Act, in that they were moved out of areas that had been set aside for other groups.

Given that both group membership (based on race and space) and the socio-economic resources available to particular groups were dictated by the State, arguably, the degree of self-selection into and out of neighbourhoods was limited. High-income African households were constrained in their ability to self-select into a better neighbourhood, and could only ever self-select into a better African neighbourhood. Given the inadequate provision of state resources to African areas in general, it is unclear how large the benefits of such neighbourhood selection conditioned on race would have been. Similarly, poor white households were "protected" in the sense that they could only self-select into a poorer white neighborhood, whose average resources and infrastructure would have been better than even the mot affluent African neighbourhoods. With the scrapping of apartheid legislation post 1993, the opportunity for self-selection and "associational redistribution"(Durlauf, 1997) has increased. However, the legacy of racially motivated economic and political controls has become a binding constraint for many households in the new South Africa, with the result that self-selection is still severely limited. Even with the removal of the Group Areas Act, residential communities have remained largely segregated. The incorporation of advantaged and previously disadvantaged communities under single municipal authorities has taken time, as bureaucratic structures have had to be re-negotiated, with the result that previously disadvantaged communities have yet to see significant improvements in the provision of services or infrastructure to their neighbourhoods.

Schooling gaps in South Africa

In 1994, 90% of all 10-14 year olds in South Africa were enrolled in school, but only 55% were in the normal standard for their age. Over-age scholars are both a result of late entry into school and high repetition rates. Policy makers are concerned with both late entry and early exit from school. Early exit from school may mean lower future earnings for the individual, to the extent that they have less human capital. This may also constrain the individual's employment opportunities, causing them to spend time unemployed. Furthermore, to the extent that early exit from school is driven by the poverty of the household, with individuals being forced to find employment to support household members, this is cause for further concern. The outcome of early school exit may hold financial burdens for the state who will be forced to provide welfare assistance to these individuals. However, the overriding fact is that given South Africa's history of unequal spending on education across race groups, it is predominantly African students who experience schooling gaps. In the interests of egalitarian reform, this source of inequality of opportunity and outcome should be removed.

The schooling gap is defined as the potential years of education (i.e. number of years of education an individual would have if they started school at age 6 and advanced one grade every year) minus the number of years of education actually attained. However, given the fact that the raw schooling gap measure will increase with the age of the child, I use an alternative dependent variable in my regressions. The dependent variable, the schooling gap ratio, is defined as the ratio of the schooling gap to the potential years of schooling attainment of the child. This ratio provides an indication of the extent of the child's schooling career that is missing. For example, a schooling gap ratio of 0.5

indicates that the child has only half the years of schooling that he or she should have, while a ratio of 0.75 indicates that the child only has one quarter of the years of schooling that he or she should have.

The existence of schooling gaps is confirmed by the KIDS data, with the average schooling gap for the entire sample of children between the ages of 6-17 being just under 2 years[17]. This schooling gap increases with age as one would expect. This merely indicates that children leave school at a relatively early age. For example, the schooling gap for 9 to 12 year olds is not that much larger than for 6 to 8 year olds. However, the schooling gap is much larger for 13 to 15 year olds as compared to 9 to 12 year olds. This would suggest that individuals tend to leave school in their early teens.

However, to look only at the size of the schooling gap is misleading. The schooling gap ratio provides a better measure of shortfalls in education attainment. On average, the school gap is 42% of the number of years of education the child should have attained. This means that children attain approximately three fifths of their potential education. This schooling gap ratio is slightly higher for children in the 9-12 year age group than children in the 13-15 age group.

Table 1: Average schooling gaps, by gender and age group

All 6-17 year olds	*N*	*Mean*	*Std. Dev*	*Mean*	*Median*	*N*	*Mean*	*Median*	*N*
					Male			*Female*	
Average school gap	1380	1.92	1.31	2.05	2.00	707	1.79	1.50	673
Average schooling gap ratio	1380	42.37	27.10	44.51	41.67	707	40.12	39.29	673
6-8 year olds									
Average school gap	146	1.33	0.63	1.47	1.50	66	1.22	1.25	80
Average schooling gap ratio	146	58.73	30.10	61.74	75.00	66	56.25	62.50	80
9-12 year olds									
Average school gap	878	1.65	1.05	1.75	1.50	456	1.55	1.50	422
Average schooling gap ratio	878	39.22	27.08	41.67	39.29	456	36.56	32.14	422
13-15 year olds									
Average school gap	356	2.83	1.61	2.99	3.00	185	2.66	2.50	171
Average schooling gap ratio	356	43.43	23.17	45.36	45.00	185	41.34	41.67	171

Notes:
The average schooling gap is the difference between the number of years of education an individual would attain if they started school at age six, and progressed one grade each year minus their actual school attainment (in years). The average schooling gap ratio is the ratio of the schooling gap, to the potential number of years of education the child could have attained.

[17] I limit the sample to children between the age of 6 and 17 as this corresponds to children of school going age who attend primary and secondary school.

It is also the case that schooling gaps are higher for boys than for girls on average, while schooling gap ratios are smaller for girls than boys. This gender difference may indicate that more sons leave school early to find work as migrants, and to possibly support the household. However, considering average schooling gap ratios is somewhat unsatisfactory, especially if this phenomenon exhibits non-linearities.

Estimating family and neighbourhood effects in South Africa

At the core of this analysis is the notion that learning by children occurs partly as a result of a choice by the child to engage in educational activities, and partly as a result of environmental exposure (i.e. learning is affected by the school/neighbourhood/home environment). Changes in the home, school or neighbourhood environment could alter the learning experience of the child, and cause the child to re-allocate their time between learning and non-learning activities. Following Maynard (1977):

$$L_t = f(\overline{L}_{t-1}, g, A_t, E_t)$$

where L_t = learning in period t
L_{t-1} = the stock of knowledge at the start of period t
g = genetic endowment
A_t = learning activities chosen by the child in period t
E_t = environmental conditions in time t.

Educational success may be influenced both by an individual's traits as well as the characteristics of the individuals with whom the individual interacts. This could include peers at school, or the broader community. If these interactions are not random, but are more likely to take place with individuals who belong to a common group, then (following Bowles and Gintis (Bowles, 2001)):

$$s_o = b_{s_o y_p} y_p + b_{s_o s_p} s_p + b_{s_o x} x + e_{s_o}$$

where S_o represents the schooling attainment of the child, Y_p is parental income, S_p is parental schooling, and X is a vector of characteristics describing the average level of schooling, economic success, wealth etc of the group (or in this case, neighbourhood) to which the individual belongs.

In my preliminary analysis, a simple OLS regression model of the following form was estimated:

$$SG_i = a + b_1 X + b_2 Y + gN + e\ ,$$

where $Y = w + w_1 X + w_2 n + u$ [18]

where SG_i is the schooling gap ratio[19] of the *i*th individual, X is a vector of parental education, Y is household income, N is a vector of neighbourhood characteristics, and n is a subset of those neighbourhood characteristics.

The data used in this analysis comes from the KwaZulu Income Dynamics Study (KIDS), a panel data set covering approximately 1200 households, and 11400 individuals[20]. The first round of data was collected in 1993 under the auspices of the *Project for Statistics on Living Standards and Development* (PSLSD), which was the first ever nationally representative demographic and socio-economic survey to be conducted in South Africa. In 1998, a re-survey was conducted in the province of KwaZulu-Natal only. For this study, however, the sample will be limited to African households only, as preliminary investigation reveals that neighbourhood effects have minimal impact on child schooling gaps for children in Indian households. (see Appendix B, Tables 3 and 4) .

There are a number of possible data problems that arise when making use of this kind of cross-sectional data, including measurement error and sample attrition. Sample attrition bias is a potential source of bias that I have yet to correct for. The effect of measurement error is to exaggerate the dynamics, since not all of the observed intertemporal variation in the welfare indicator is due to mobility. In addition, OLS estimates will be downward biased. In estimation, this is the problem of errors in variables (Greene, 1996; Rendtel et al, 1998; Luttmer, 2000; Solon, 1989; Pritchett et al 2000.) One way to minimise this error is to average the welfare indicator across time periods, which I do. The variables used are the average of the reported values in 1993 and 1998. However, I do intend to take further steps to correct for measurement error, following the approach of Bowles[Bowles, 1972 #17], and Johnston [Johnston, 1963 #60].

This study, while prone to many of the same methodological problems outlined above, has some key features that will help to alleviate the extent of these problems. By incorporating an intergenerational element, the problem of selection is minimised in that children are not responsible for the quality of their neighbourhoods or schools, nor are

[18] If part of the impact of parents education on child schooling attainment operates via the influence of household income, and if this relationship is not explicitly modeled in the specification, then inferences regarding the impact of parental education and family income on child outcomes will be unreliable.

[19] Schooling gap ratio is calculated as : (No. yrs education child should have- no. yrs education child actually has)/No. yrs education child should have

[20] In the original data set, there were 1212 households, of which 1040 (86%) are African, and 172 (14%) are Indian. These 1212 households translate into 11407 individuals, of whom 10423 (91%) are African, and 984 (9%) are Indian.

they the decision-makers deciding where to live, and thus are not responsible for the group effects they experience in these circumstances. Furthermore, given that both group membership (based on race and space) and the socio-economic resources available to particular groups were dictated by the State in South Africa, arguably, the degree of self-selection into and out of neighbourhoods was limited. High-income African households were constrained in their ability to self-select into a better neighbourhood, and could only ever self-select into a better African neighbourhood. Given the inadequate provision of state resources to African areas in general, it is unclear how large the benefits of such neighbourhood selection conditioned on race would have been. Similarly, poor white households were "protected" in the sense that they could only self-select into a poorer white neighborhood, whose average resources and infrastructure would have been better than even the mot affluent African neighbourhoods. With the scrapping of apartheid legislation post 1993, the opportunity for self-selection and "associational redistribution"(Durlauf, 1997) has increased. However, the legacy of racially motivated economic and political controls has become a binding constraint for many households in the new South Africa, with the result that self-selection is still severely limited. Even with the removal of the Group Areas Act, residential communities have remained largely segregated. The incorporation of advantaged and previously disadvantaged communities under single municipal authorities has taken time, as bureaucratic structures have had to be re-negotiated, with the result that previously disadvantaged communities have yet to see significant improvements in the provision of services or infrastructure to their neighbourhoods. Indeed, in the KIDS data set, between 1993 and 1998, only 76 households (6% of the sample) had moved to a different location, suggesting that geographic mobility was not large.

The arguments of Brooks-Gunn (Brooks-Gunn, 1993) suggest that it is unlikely that there is a strong causal link between most family level characteristics and neighbourhood conditions, perhaps other than income. However, it is unclear whether this is true in the case of South Africa, where enforced segregation has meant that multi-generational households have often lived in one community or neighbourhood for many decades. I will examine this relationship, using data in the KIDS sample on the number of years that a household has lived in a particular community. In addition, using the technique suggested by Datcher (Datcher, 1982), I will test the sensitivity of neighbourhood co-efficients by controlling for additional family characteristics, which are aimed at capturing family unobservables.

The neighbourhood variables employed in the study are all calculated for individuals aged twenty years and older, thus ensuring that the community variables reflect outcomes for a group distinct from those captured in the dependent variable. Furthermore, the nature of the dependent variable lessens the causality problem somewhat. While the usual problem is to understand the causality between, for example, individual employment status and community level unemployment rates, in this case it seems less plausible to argue that schooling gaps of currently enrolled children reflect cluster level unemployment among adults aged 20 and older, as opposed to the opposite causality, where cluster level unemployment causes schooling gaps. Furthermore, using the panel structure of the KIDS data set, and using data from the 1996 Census, allows one to obtain a measure of neighbourhood characteristics prior to the outcomes of interest in 1998, thereby avoiding the problem of concurrent measurement, in which neighborhood conditions may reflect the process underlying the developmental outcome but not be the cause of it (Brooks-Gunn, 1993).

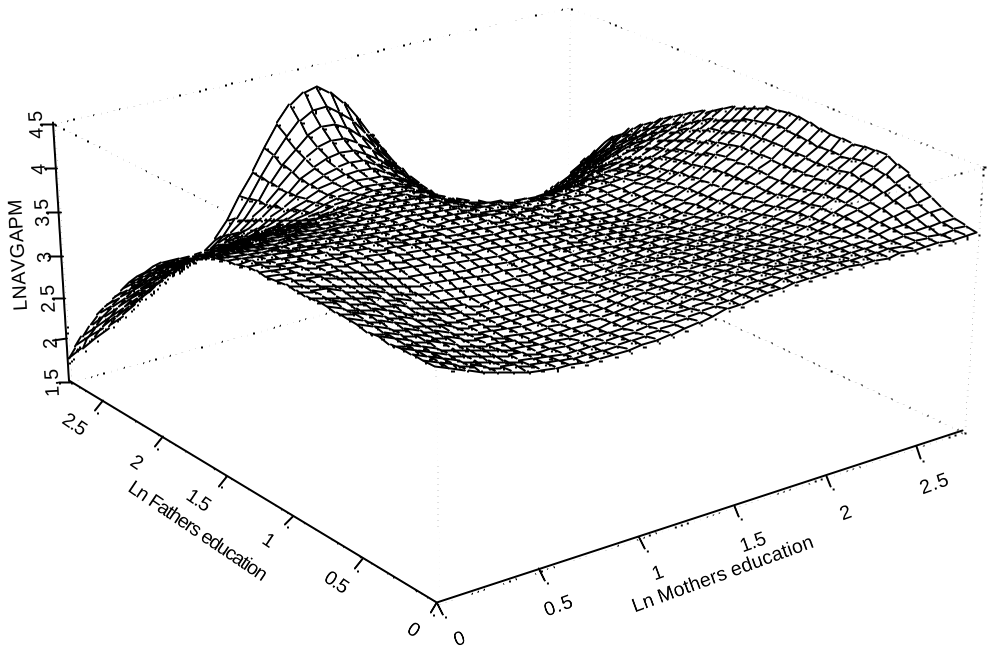
Preliminary findings

The simple OLS model described above was estimated using the KIDS data, and then re-estimated substituting neighbourhood level variables from the 1996 Census into the model (as a means of verifying the results). In most cases, the signs and significance of the co-efficients were the same. Simple OLS results (see Appendix B: Table 1) indicate that parents education has a significant impact on reducing schooling gap ratios. These results are significant, and accord with the estimates of Mulligan [Mulligan, 1999 #33], and other estimates of my own for South Africa.[21]

However, it appears that there are non-linearities in the relationship between child outcomes and parental education. Making use of the non-parametric smoothing technique of locally weighted (or nearest-neighbourhood) regression[22], a preliminary bivariate model (with quadratic polynomials used to calculate the smooth) of the form $lnY_t = lnX_{t-1} + lnZ_{t-1}$ was estimated, where X_{t-1} and Z_{t-1} refers to years of education of the mother and father respectively, while Y_t represents the schooling gap ratio of the child. The smoothed regression surface of the figure below shows the results of estimating this model.

[21] Note that the results presented in Appendix A relate to regression results for a sample of adults, all over the age of 25. These elasticities thus represent the intergenerational transmission of status for individuals who have completed their education.

[22] Cleveland and Devlin (1988) and Loader (1999) provide a discussion concerning the localisation weights and greater detail on the estimation algorithms used in a multivariate framework. Corak and Heitz (1999) applied this technique to estimating the intergenerational elasticity of earnings among Canadian men. The strength of this approach is that it does not assume any specific functional form between the dependant and independent variable. Moreover, it permits a graphical exploration of how β changes over portions of the data space, by using three-dimensional surface plots of a smoothed curve describing the data space. The basic approach is as follows: I first select a window of observations (defined as some fraction α of the total sample) in the neighbourhood of a specific data point of lnY_{t-1} (some outcome measure of the parents' generation). Call this data point x. Each observation in the window (other than x) is then weighted, with observations nearest to x given higher weights. Weighted least squares is then performed using some polynomial of lnY_{t-1} (linear or quadratic usually) and the estimated coefficients from this regression are then used to predict y, the lnY_t data point corresponding to x, where lnY_t is some outcome measure of the child's generation.



This plot suggests that the OLS specification of parental education (which suggests that parental education is complementary) may be incorrect, and that a different functional form is required. The non-parametric plot indicates that having a relatively uneducated mother need not be an obstacle to child educational success as long as paternal education is sufficiently high. Indeed, having both a highly educated mother and father does not seem to improve outcomes over this former case. Yet, having a mother with a moderate amount of education and a father with lots of education causes significant increases in schooling gap ratios. This indeed is a puzzle, and further modelling is required. At the very least, it suggests that the inclusion of both parents' education as explanatory variables should not necessarily be treated as a multiplicative function. One of the challenges of this paper will be to find the correct functional form in which to incorporate parental education into the model.

Household income also has a significant, although somewhat smaller, impact on reducing schooling gaps. This is to be expected, given that household income is likely to represent an upper bound on the total resources that parents have available to spend on their children. Furthermore, it appears that schooling gap ratios increase with age, suggesting that grade repetition and early school exit are more important than later start in this respect. Contrary to Coleman's suggestion[23] (Coleman, 1988), average family capital (adult:child ratio) is insignificant and has the wrong sign.

---

[23] Coleman **Coleman, J.S.** "Social Capital in the Creation of Human Capital." *American Journal of Sociology, Supplement*, 1988, *94*, pp. S95-120. suggests the use of this ratio as a proxy of social capital in

It is worth noting that the inclusion of neighbourhood variables into the regression equation does not, with the exception of the co-efficient on rural location (itself a proxy for neighbourhoods), alter any of the co-efficients on family background in any significant manner. This suggests that family background effects and neighbourhood variables are uncorrelated, a surprising result, which will need to be investigated further for its robustness (See Appendix B, Table 1). However, this does not rule out the existence of an underlying unobservable factor related to neighbourhoods that is correlated with schooling gap ratios, but not with family background variables, influencing the results. This possibility is merely noted, but cannot be corrected for with the current dataset.

Living in a community with relatively more affluent households (proxied by the fraction of households with monthly income greater than R3000) reduces schooling gaps significantly.[24] This accords with the findings of Brooks-Gunn et al (Brooks-Gunn, 1993) in their study using PSID data. Also, the average cluster unemployment rate has a negative impact on schooling gap ratios, even after controlling for urban and metro areas (not shown). Where unemployment is high, perhaps suggesting isolation from job opportunities and networks, grade repetition is lower and fewer children leave school early. Moreover, as the fraction of high status workers in the cluster increases, schooling gap ratios fall. The fraction of high status workers may be a significant indicator of neighbourhood quality for two reasons. On the one hand, these individuals may act as role models for children in the neighbourhood and give them an incentive to stay in school. In the presence of role models, children correctly evaluate the marginal product of effort in their schoolwork, and this reduces grade repetition. (Streufert, 1991). Alternatively, high status workers may be able to use their status or influence to bring better resources into the neighbourhood and make local institutions work better.[25] Arguably, in the same vein then, one could make the argument that where unemployment is high, this improves the evaluation of marginal product of effort in schoolwork, as children realize that to be successful in the labour market will require some minimum level of education. An alternative is that where jobs are scarce, there is no reason for children to leave school, and children realise that the only hope of being able to gain employment is to gain a good education. All of these results taken together provide tentative evidence that neighbourhood effects operate via a collective socialisation model. [26] However, as the data is insufficient to really be able to test these propositions, they will have to remain speculations.

---

the family, with the notion that as the number of adults increases relative to children in the household, supervision of children improves.

[24] Neighbourhood income measures may be a summary measure of the resources available within a community, and may be correlated with better schooling, information, job opportunities etc.

[25] Arguably, the presence of high status workers may reflect a selection process, with affluent individuals choosing to live in "good" neighbourhoods. **Crane, J.** "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping out and Teenage Childbearing." *American Journal of Sociology*, 1991, *96*(5), pp. 1226-59. However, given South Africa's spatial legacy and influx controls, it is unclear how applicable such a question is.

[26] Incorporating lagged exogenous variables into the OLS model such that

Neighbourhoods effects relative to family effects

Preliminary analysis to test for the relative importance of family background vis-à-vis neighbourhood effects suggests that changing family background variables would account for about two thirds of the change in child outcomes, while neighbourhoods account for the other third. (see Appendix D, Table 6) Using Datcher's method, the results suggest that changing neighbourhood characteristics for households at the 25th percentile to match those of households in the 75th percentile would account for one third of the resulting decline in child schooling gaps, while changing family background variables would account for the remaining two thirds decline. The same result is found simply by summing the standardized regression co-efficients.[27] Surprisingly, these results are very similar to Datcher's estimates for the US. However, in Table 7 (Appendix D) I re-estimate these effects, including in the regression model only those variables that were significant in the previous regressions. In this case, family effects account for 56% of the change in child schooling outcomes, while neighbourhood variables account for the remaining 44%.[28] This larger neighbourhood effect accords with my *a pri*ori expectations.

One question which this paper will address is how these results, once verified as being robust, should be interpreted. This is not a trivial matter as different interpretations provide for different policy prescriptions. One possible interpretation is that impediments to residential mobility should be removed, (and perhaps even assistance provided) to assist households in poorer neighbourhoods to relocate to more affluent neighbourhoods. In part, this was one of the intended consequences of the scrapping of the Group Areas Act in South Africa. An alternative interpretation of these results is that if the existing conditions, opportunities and services in poorer neighbourhoods were improved to match the levels of those in more affluent neighbourhoods, this would generate significant positive effects on child schooling outcomes, along with improved standards of living for all. I favor the latter interpretation.

$SG_{it} = a + b_1 X_{t-1} + b_2 Y_{t-1} + gN_{t-1} + e$,
where $Y_{t-1} = w + w_1 X_{t-1} + w_2 n_{t-1} + u$[26]

Produces very similar results. (See Appendix B:Table 5)

[27] These results hold regardless of whether I include all the regression co-efficients in the calculations, or if I restrict the included co-efficients to those that are significant in the regression.

[28] One issue to be resolved in these calculations is the treatment of the unemployment result. If the current results are to be believed, it suggests that high unemployment has an indirect positive impact on reducing schooling gaps, as children recognise the importance of staying in school to get an education. This has obviously strange policy implications and clearly, one would not want to advocate increasing unemployment as a means of lowering schooling gaps. The challenge for this work is to investigate the robustness of these preliminary findings, before any final conclusions can be drawn about the relative importance of family and neighbourhoods in these matters.

Peer effects and school quality

A recent extension of this work has been to incorporate a role for peer effects and school quality. While the analysis thus far has conceptualised the "common group" to which the individual belongs as the neighbourhood, it is also possible to think of peer groups in the same way. Arguably, children are influenced significantly by the behaviour of their peers, and the average educational attainments of children in the community may have an influence on individual schooling gaps. From Table 8 in Appendix D, it is clear that the average schooling gap of one's peers in the community has a very large and significant impact on individual schooling gaps, and could arguable negate the positive effects of having educated parents. Furthermore, with the inclusion of this "peer effect", some of the neighbourhood variables lose their significance.

This is an interesting result, and certainly deserves more attention. However, it is difficult to know whether this result is truly indicative of peer effects, or whether it is merely a proxy for poor schooling quality. Are high average schooling gaps merely a reflection of inadequate schooling resources, overcrowded classrooms[29] and insufficient teachers, or do they reflect some type of peer effect, where attitudes and behaviours are influenced and conditioned by the group to which one belongs?

Indeed, a weakness of this work to date is its failure to include any dimension of schooling quality in the analysis. In part, this is owing to a lack of adequate data on these aspects, as well as a reflection of the difficulty of measuring schooling quality, as opposed to quantity. Data that can be used to proxy for schooling quality was only collected in the 1998 round of interviews in the KIDs study. Not only was coverage incomplete, but the reliability of this data is questionable. However, I have incorporated these variables into the regression analysis (limited to African households only), in order to try and assess the robustness of my findings to date. I find some evidence that schooling quality, proxied by the student to classroom ratio, does matter. Overcrowded classrooms are associated with rising schooling gaps. Once I add in these school quality controls, only one neighbourhood measure, namely the fraction of households earning more than R3000 per month, remains significantly associated with schooling gaps. (Table 9, Appendix D).

As a preliminary attempt to control for the possible endogeneity between schooling quality and household/neighbourhood characteristics, I run the same regression for African households that have been resident in the same community since 1990. Thus, by the time this data was collected in 1998, the household would have been resident in the community for at least 8 years, and prior to the scrapping of the Group Areas Act. As argued earlier, self-selection among African households was fairly limited to begin with, and this is a simple, albeit incomplete, way to try to control for self-selection biases. These results are shown in Table 10 of Appendix D, and again, I find that the student to classroom ratio, as well as the number of secondary schools present in the community are significantly associated with schooling gaps. For this subset of households, cluster

[29] The average schooling gaps of children in the community and the student to classroom ratio (a proxy for school quality) are correlated 0.4.

unemployment rates and the fraction of households earning more than r3000 per month also remain significant. Thus, both schooling quality and neighbourhood characteristics matter.[30]

However, in a final regression in Table 11 (Appendix D), I include peer effects, neighbourhood effects and school quality variables. Once again, I find that peer effects are large and significant, and with their inclusion, the schooling quality variables become insignificant, as do most of the neighbourhood variables. Only family variables remain significant as before.

Estimating robust results

Clearly, simple OLS regressions are insufficient to adequately explain the role of family background and neighbourhood effects on schooling outcomes. There are a number of ways I plan to test for the robustness of these estimates:

1. ***Aggregation**:* Currently, the neighbourhood variables from the KIDS data set provide information across all the households interviewed in each census cluster. In some cases, these households were in close enough proximity to each other to constitute a neighborhood, but in others, this may not have been the case Using the original KIDS survey maps and the recorded information on street addresses, I will follow the approach of Case and Katz (Case, 1991) in defining neighbourhoods in the KIDS data set in the true sense of the word. Similarly, the census-matched data currently provides information at the level of a census district, which is likely to be too aggregated. Linking the KIDS-mapped communities to census sub-districts will enable me to extract more locally disagregated census neighbourhood information from the Census, which can be used as a means of verification.

2. ***Age effects***: Brooks-Gunn et al (Brooks-Gunn, 1993) argue that neighbourhood characteristics will have a greater direct impact on child outcomes during late adolescence than during early childhood, since adolescents have more immediate direct and intense interactions with neighbours and neighborhood institutions. There is already evidence that schooling gaps vary by age[31], with grade repetition becoming more serious as the child gets older. I will estimate my equations for two subsets of children, those aged 6-12, and those aged 13-17. A priori, one would expect neighbourhood level effects to be larger for the latter group.

3. ***Functional form**:_* A key contribution of this paper will be to investigate alternative functional forms for the way in which parental education should be included in the analysis. Preliminary evidence indicates that the current double-log specification in the OLS regressions is inadequate. Once functional form has been corrected for, the relative importance of neighbourhood and family effects will be re-estimated, in order to see whether a change in the functional form affects my simple OLS results

[30] Indeed, one might argue that schooling quality is itself a neighbourhood characteristic.
[31] See Appendix D.

significantly. In addition, this paper will take up the challenge laid down in the literature of investigating non-linearities between neighborhood variables and individual outcomes.

4. ***Disentangling neighbourhood effects:***

Preliminary evidence suggests that endogeneity problems are not severe, as the co-efficients on family background variables do not change significantly when neighbourhood variables are added to the model. However, in the simplest case, I will follow Datcher (Datcher, 1982) and include additional family controls to test the sensitivity of the neighbourhood co-efficients.

<u>*A. Peer effects*</u>: Arguably, peer effects, in conjunction with neighbourhood effects, may be an important factor in schooling gap ratios. Case and Katz (Case, 1991), in their study of peer effects, propose the use of probit equations in alleviating endogeneity problems.[32] Following their methodology, I will run two sets of probit equations. Model 1 estimates individual schooling outcomes as a function of individual and family background characteristics, and records the log-likelihood for these models. The second set of probits is identical to the first, but now includes a measure of the schooling gap ratios of peers in the community. I will then employ log-likelihood tests to test for the joint significance of the explanatory variables of peers.

Model 1: $Y_i^* = X_i b + u_{i,} \quad Y_i = 1$ if $Y_i^* \rangle 0$ and $Y_i^* = 0$ otherwise.

Model 2: $Y_i^* = X_i b + lX_j + u_i, \; Y_i = 1$ if $Y_i^* \rangle 0$ and $Y_i^* = 0$ otherwise

where X_i is a vector of family background characteristics, and X_j is the schooling gap ratio of other children in the neighbourhood. $Y_i = 1$ if the child has a schooling gap ratio above the mean schooling gap ratio for the entire sample.

<u>*B. Treatment effects:*</u> In addition, I will control for so-called treatment effects. Following Greene (Greene, 1997) a sample selection model can be applied to treatment effects.

$$SG_c = b'x_i + dN_i + e_i$$

where SG_i is the schooling gap ratio of the *ith* individual, and N_i is a dummy variable indicating whether the individual lives in a good quality or poor quality neighbourhood. However, using OLS, d doesn't measure the value of living in a good neighbourhood if the typical child who lives in a good neighbourhood would have a low schooling gap ratio irrespective of whether they lived in a good or poor quality neighbourhood. In this case, d will over-estimate the treatment effect. Thus, neighbourhood quality will be modelled as a probit equation:

---

[32] Following Chamberlain (1985) they argue that in the case of time series data, one can add lagged exogenous variables to the model and test their significance. If current behaviour fails to respond to lagged exogenous variables, then this is evidence that no interdependence is taking place.

$$N_i^* = gw_i + u_i$$

$N_i = 1$ *if* $N_i^* > 0$, *0 otherwise*

The model is then estimated using the two-step Heckman procedure. An alternative is to employ latent variable econometric techniques. (For a fuller discussion of these methods, see Appendix E)

Conclusion

Preliminary evidence indicates that both family and neighbourhood characteristics play an important, and relatively equal, role in child schooling outcomes. The challenge for this paper will be to test these simple OLS results for their robustness, particularly by investigating superior model specifications and controlling for peer effects. Should the general findings of this preliminary analysis hold, however, it will be an important step towards highlighting the continued role of inherited inequalities, both in terms of family background and neighbourhood status, in shaping South Africa's future.

Appendix A

Table 1.: Persistence in education status between parents and children, controlling for the education of both parents.

		First generation to second generation				***Second generation to third generation***					
Dep. Variable: Ln of scaled education	***Parent***	***Maternal grandparents to mothers***		***Paternal grandparents to fathers***		***Parents to child***		***Parents to daughter***		***Parents to sons***	
<u>Controls in regression</u>		Beta[a]	Adj. R-sq (df)	Beta[b]	Adj. R-sq (df)	Beta[c]	Adj. R-sq (df)	Beta[d]	Adj. R-sq (df)	Beta[e]	Adj. R-sq (df)
Age	*Mother*	0.23	0.39	0.13	0.31	0.34	0.32	0.37	0.29	0.31	0.33
	Father	0.26	1051	0.27	672	0.21	697	0.21	353	0.20	335
Age and child is African	*Mother*	0.24	0.40	0.19	0.38	0.34	0.32	0.36	0.29	0.31	0.35
	Father	0.21	1050	0.12	671	0.18	696	0.21	352	0.14	334
Age and child lives in Rural area	*Mother*	0.22	0.42	0.14	0.37	0.33	0.33	0.35	0.31	0.30	0.35
	Father	0.21	1050	0.16	671	0.17	696	0.18	352	0.14	334
Age and Child is female	*Mother*					0.34	0.32				
	Father					0.21	696				
Age, child is African and in rural area	*Mother*	0.22	0.42	0.17	0.39	0.33	0.33	0.36	0.31	0.31	0.36
	Father	0.20	1049	0.10	670	0.17	695	0.19	351	0.13	333
Age, child is female, African and lives in rural area	*Mother*					0.33	0.33				
	Father					0.17	694				

Appendix B:

Table 2. OLS regression using average values of regressors across 1993-1998

Dependent Variable: Ln of average schooling gap ratio of child						
	Reg 1: family variables only			*Reg 2: Family and neighbourhood variables*		
	B	Beta		B	Beta	
(Constant)	9.32		*	9.75		*
	(1.56)			(1.56)		
Ln of mother's education (scaled up by 1)	-0.22	-0.16	*	-0.22	-0.16	*
	(0.08)			(0.08)		
Ln of father's average education (scaled up by 1)	-0.17	-0.14	*	-0.15	-0.12	**
	(0.07)			(0.07)		
Child lives in female headed household	0.24	0.08	**	0.20	0.07	**
	(0.10)			(0.10)		
RURAL	0.28	0.10	*	0.05	0.02	
	(0.11)			(0.11)		
Child is female	-0.10	-0.05		-0.11	-0.05	
	(0.08)			(0.08)		
Average age of child, 1993-1998	-1.03	-1.78	*	-1.00	-1.74	*
	(0.27)			(0.27)		
Average age of child squared, 1993-98	0.05	1.85	*	0.04	1.81	*
	(0.01)			(0.01)		
Average family capital-adult:child ratio 1993-1998	0.01	0.00		0.03	0.02	
	(0.05)			(0.05)		
Estimated avge. real HH income, 1993-1998	-0.09	-0.08	**	-0.05	-0.05	
	(0.04)			(0.04)		
Avge cluster unemployment rate 1993-98				-0.95	-0.07	***
				(0.58)		
Fraction of HH with average real HH income less than or equal to R1000 per month				0.13	0.03	
				(0.33)		
Fraction of HH with average real HH income more than R3000 per month				-0.83	-0.13	***
				(0.49)		
Average fraction of high status workers in cluster 1993-98				-2.39	-0.30	*
				(0.90)		
Average fraction of high status workers in cluster squared				4.24	0.27	**
				(1.83)		
Adj R-squared	0.14			0.16		
df	691			686		

*Sig at 1% level; ** Sig at 5% level; *** Sig at 10% level; Beta represents the standardised co-efficient.*

Table 3: OLS regression, by race group (including a squared term for high status works in the community)

Dependent Variable: Ln of average schooling gap ratio of child	**Full sample**				**African households only**				**Indian households only**			
	B	*Std error*	***Beta***		***B***	*Std error*	***Beta***		***B***	*Std error*	***Beta***	
FAMILY VARIABLES												
Constant	**9.75**	1.56		*	**10.26**	1.71		*	**7.10**	4.54		
Child lives in female headed Household	**0.20**	0.10	**0.07**	**	**0.18**	0.10	**0.07**	***	**0.95**	0.67	**0.16**	
Child is female	**-0.11**	0.08	**-0.05**		**-0.09**	0.08	**-0.04**		**-0.22**	0.20	**-0.12**	
Average age of child	**-1.00**	0.27	**-1.74**	*	**-1.08**	0.30	**-1.87**	*	**-0.90**	0.71	**-1.92**	
Average age of child, squared	**0.04**	0.01	**1.81**	*	**0.05**	0.01	**1.93**	*	**0.04**	0.03	**2.00**	
Ln of mother's education (scaled up by 1)	**-0.22**	0.08	**-0.16**	*	**-0.25**	0.08	**-0.17**	*	**0.34**	0.47	**0.10**	
Ln of father's average education (scaled up by 1)	**-0.15**	0.07	**-0.12**	**	**-0.13**	0.07	**-0.10**	***	**-0.44**	0.33	**-0.19**	
Estimated avge. real HH income	**-0.05**	0.04	**-0.05**		**-0.06**	0.04	**-0.05**		**-0.11**	0.18	**-0.10**	
Average family capital: Adult to child ratio in HH	**0.03**	0.05	**0.02**		**0.03**	0.06	**0.02**		**-0.08**	0.15	**-0.07**	
NEIGHBORHOOD VARIABLES												
Child lives in rural HH	**0.05**	0.15	**0.02**		**0.00**	0.17	**0.00**		**2.01**	0.89	**0.34**	**
Avge cluster unemployment rate 1993-98	**-0.95**	0.58	**-0.07**	***	**-1.09**	0.70	**-0.06**		**-1.33**	5.05	**-0.05**	
Average fraction of high status workers 1993-98	**-2.39**	0.90	**-0.30**	*	**-2.50**	1.29	**-0.22**	**	**6.51**	6.59	**1.25**	
Fraction of high status workers squared	**4.24**	1.83	**0.27**	**	**4.51**	4.16	**0.12**		**-7.25**	9.45	**-1.04**	
Fraction of HH with average real HH income less than or equal to R1000 per month	**0.13**	0.33	**0.03**		**0.19**	0.34	**0.04**		**4.36**	3.71	**0.24**	
Fraction of HH with average real HH income more than R3000 per month	**-0.83**	0.49	**-0.13**	***	**-0.62**	0.65	**-0.06**		**-0.89**	0.87	**-0.22**	
Adjusted R-squared	**0.16**				**0.14**				**0.04**			
df	**686**				**603.00**				**68.00**			

=Sig at1%, **=Sig at 5%; *=Sig at 10%; ****=Sig at 15%*

Table 4: By race, excluding a squared term for fraction of high status workers in cluster

Dependent Variable: Ln of average schooling gap ratio of child	***African households***				***Indian households***				***Full sample***			
FAMILY VARIABLES	***B***	***Std. Error***	***Beta***		***B***	***Std. Error***	***Beta***		***B***	***Std. Error***	***Beta***	
Constant	10.13	1.70		*	8.20	4.29		***	9.47	1.56		*
Female headed Household	0.18	0.10	0.07	***	0.97	0.67	0.16	****	0.19	0.10	0.07	***
Child is female	-0.09	0.08	-0.04		-0.24	0.20	-0.13		-0.10	0.08	-0.05	
Average age of child	-1.07	0.30	-1.84	*	-0.94	0.71	-2.00		-0.98	0.27	-1.71	*
Average age of child, squared	0.05	0.01	1.90	*	0.04	0.03	2.07		0.04	0.01	1.76	*
Ln of mother's education (scaled up by 1)	-0.25	0.08	-0.18	*	0.33	0.47	0.10		-0.22	0.08	-0.16	*
Ln of father's average education (scaled up by 1)	-0.13	0.07	-0.10	***	-0.49	0.33	-0.21	****	-0.15	0.07	-0.11	**
Estimated avge. real HH income	-0.05	0.04	-0.05		-0.10	0.18	-0.09		-0.05	0.04	-0.04	
Average family capital	0.03	0.06	0.02		-0.06	0.15	-0.05		0.03	0.05	0.02	
NEIGHBORHOOD VARIABLES												
Rural	-0.01	0.17	0.00		1.73	0.81	0.29	**	0.05	0.15	0.02	
Avge cluster unemployment rate 1993-98	-1.27	0.68	-0.07	***	0.33	4.55	0.01		-1.46	0.54	-0.11	*
Average fraction of high status workers 1993-98	-1.24	0.57	-0.11	**	1.53	1.19	0.30		-0.61	0.46	-0.08	
Fraction of HH with average real HH income less than or equal to R1000 per month	0.27	0.34	0.05		2.55	2.85	0.14		0.45	0.30	0.10	****
Fraction of HH with average real HH income more than R3000 per month	-0.40	0.62	-0.04		-1.23	0.75	-0.30	***	-0.33	0.44	-0.05	
Adjusted R-squared	0.14				0.04				0.16			
df	604				69				687			

=Sig at1%, **=Sig at 5%; *=Sig at 10%; ****=Sig at 15%*

Table 5: OLS regression, using lagged dependent variables

Dependent Variable: *Dependent Variable: Ln of schooling gap ratio of child* 1998						
	Regression 1: Family variables			*Regression 2: family and neighbourhood vars*		
	B	Beta		B	Beta	
(Constant)	5.08		*	5.51		*
	(0.97)			(0.98)		
RURAL	0.16	0.06		-0.04	-0.01	
	(0.11)			(0.14)		
Female headed HH	0.28	0.09	*	0.26	0.09	*
	(0.10)			(0.10)		
Estimated HH income, 1993	0.00	0.00		0.02	0.02	
	(0.04)			(0.04)		
Child is female	-0.13	-0.06	***	-0.14	-0.06	***
	(0.08)			(0.08)		
Age of child, 1993	-0.30	-0.50		-0.29	-0.48	
	(0.22)			(0.21)		
Age of child squared, 1993	0.01	0.32		0.01	0.32	
	(0.01)			(0.01)		
Ln of mother's education (scaled up by 1)	-0.15	-0.10	**	-0.16	-0.11	**
	(0.08)			(0.08)		
Ln of father's average education (scaled up by 1)	-0.21	-0.15	*	-0.18	-0.13	*
	(0.07)			(0.07)		
Family capital: adult to child ratio in HH	-0.05	-0.03		-0.04	-0.03	
	(0.05)			(0.05)		
Child's schooling gap ratio, 1993 (scaled up by 1)	0.17	0.29	*	0.16	0.27	*
	(0.02)			(0.02)		
Cluster Unemployment rate 1993				-0.99	-0.11	**
				(0.41)		
Fraction of HH with average real HH income less than or equal to R1000 per month in 1993				0.40	0.08	
				(0.32)		
Fraction of HH with average real HH income more than R3000 per month in 1993				-0.54	-0.08	
				(0.45)		
Fraction of high status workers in cluster, 1993				-1.18	-0.19	***
				(0.64)		
Fraction of high status workers in cluster squared , 1993				1.52	0.14	
				(1.20)		
Adjusted R-squared	0.19			0.21		
df	734.			729		

*Sig at 1% level; ** Sig at 5% level; ***Sig at 10% level;

Figures in brackets are standard errors of the co-efficients

Appendix D: Calculating the relative importance of family and neighbourhood effects

Table 6: Size of family effects relative to neighbourhood effects

Datcher's method	***Unstandardised co-efficient A***	***Difference (75th percentile - 25th percentile) B***	***A*B***	***Summing the standardized co-efficients***
Constant	10.13			
Female headed households	0.18		-0.18	-0.07
Child is female	-0.09		0.09	0.04
Average age of child	-1.07			
Average age of child squared	0.05			
Ln of mother's education (scaled up by 1)	-0.25	1.16	-0.29	-0.18
Ln of father's average education (scaled up by 1)	-0.13	1.32	-0.17	-0.10
Estimated avge. real HH income	-0.05	1.10	-0.06	-0.05
Average family capital	0.03	0.77	0.02	0.02
TOTAL FAMILY EFFECTS			***-0.59***	***-0.34***
Avge cluster unemployment rate 1993-98	-1.27	0.11	-0.14	-0.07
Average fraction of high status workers 1993-98	-1.24	0.16	-0.20	-0.11
Fraction of HH with average real HH income less than or equal to R1000 per month	0.27	0.30	0.08	0.05
Fraction of HH with average real HH income more than R3000 per month	-0.40	0.11	-0.04	-0.04
Rural location	-0.01	0.01	0.00	0.00
TOTAL NEIGHBOURHOOD EFFECTS			**-0.3**	**-0.17**
Total			***-0.89***	***-0.51***
Family effects as proportion of total (%)			***0.66***	***0.67***
Neighbourhood effects as proportion of total (%)			***0.34***	***0.33***

Note: These calculations are based on the co-efficients from Table 4 in Appendix B.

In this regression, I include only those regressors that appear as significant in earlier regressions. On this basis, changing family background variables would account for 56% of the decrease in child schooling gap ratios, while neighbourhood effects would account for the remaining 44%.

Table 7: Size of family effects relative to neighbourhood effects, when only significant variables are included in regression

Dependent Variable: Dependent Variable: Ln of average schooling gap ratio of child	*Unstandardized Coefficients*	*Std. error*	*Standardized Coefficients*	*Sig.*
FAMILY VARIABLES				
Constant	4.42	0.17		*
Ln of mother's education (scaled up by 1)	-0.25	0.07	-0.18	*
Ln of father's average education (scaled up by 1)	-0.16	0.07	-0.13	**
Female headed HH	0.15	0.10	0.05	****
Total family effects			***-0.36***	
NEIGHBOURHOOD VARIABLES				
Avge cluster unemployment rate 1993-98	-1.17	0.49	-0.09	**
Average fraction of high status workers 1993-98	-1.44	0.30	-0.19	*
Total neighbourhood effects			***-0.28***	
df	759			
Adjusted R-squared	0.13			
Total effects			***-0.64***	
Family effects as proportion of total (%)			***0.56***	
Neighbourhood effects as proportion of total (%)			***0.44***	

Note: Female headed household is added as a negative (dummy effect)

=Sig at 1%; **=Sig at 5%; *=Sig at 10%; ****=Sig at 15%*

Table 8: Incorporating peer effects into the regression

Dependent Variable: Ln of average schooling gap ratio of child	*Full sample*		*African households only*	
	B	***Beta***	***B***	***Beta***
(Constant)	6.170	*	6.581	*
	(1.576)		(1.718)	
Ln of school gap of peers in cluster	0.803	0.263 *	0.891	0.251 *
	(0.150)		(0.165)	
Ln of mother's education	-0.126	-0.090 ***	-0.136	-0.096 ***
	(0.073)		(0.075)	
Ln of father's average education	-0.145	-0.113 **	-0.124	-0.096 ***
	(0.067)		(0.069)	
Child lives in female headed HH	0.162	0.058 ***	0.144	0.054 ****
	(0.093)		(0.096)	
Child is female	-0.089	-0.041	-0.077	-0.035
	(0.072)		(0.078)	
Age of child	-0.912	-1.598 *	-1.037	-1.805 *
	(0.254)		(0.276)	
Age of child, squared	0.040	1.654 *	0.046	1.875 *
	(0.011)		(0.012)	
Average fraction of high status workers 1993-98	-0.269	-0.035	-0.900	-0.079 **
	(0.433)		(0.540)	
Fraction of HH with average real HH income less than or equal to R1000 per month	0.206	0.046	0.095	0.017
	(0.282)		(0.317)	
Fraction of HH with average real HH income more than R3000 per month	-0.123	-0.020	0.088	0.008
	(0.386)		(0.567)	
Avge cluster unemployment rate 1993-98	-1.100	-0.088 **	-0.818	-0.049 *****
	(0.500)		(0.630)	
HH is in rural area	-0.096	-0.037	-0.140	-0.041
	(0.130)		(0.150)	
adj R-squared	0.18		0.17	
df	752		649	

=Sig at 1% level; ** = Sig at 5% level; *= Sig at 10% level; ****=Sig at 15% level; *****Sig at 20% level*
Figures in brackets are standard errors of the co-efficients

Table 9: Incorporating schooling quality (using 1998 data only)

Dependent Variable: Ln of average schooling gap ratio of child, 1998						
	B	***Beta***		***B***	***Beta***	
Constant	-7.058		*	-7.062		*
	(0.925)			(0.842)		
Child lives in female headed HH	0.264	0.071	*	0.313	0.088	*
	(0.102)			(0.095)		
Child is female	-0.238	-0.077	*	-0.247	-0.080	*
	(0.085)			(0.082)		
Age of child, 1998	1.756	3.295	*	1.719	3.236	*
	(0.144)			(0.139)		
Age of child (1998) squared	-0.065	-3.067	*	-0.064	-3.016	*
	(0.006)			(0.006)		
Ln mother's education, 1998	-0.242	-0.130	*	-0.249	-0.133	*
	(0.067)			(0.065)		
Ln father's education, 1998	-0.110	-0.062	***	-0.124	-0.070	**
	(0.064)			(0.061)		
Ln real household income, 1998	-0.046	-0.032				
	(0.045)					
Fraction of HHs in cluster earning more than R3000 (real) per month, 1998	-0.804	-0.076	***	-1.211	-0.114	*
	(0.463)			(0.314)		
Fraction of high status workers 1998	-0.799	-0.039				
	(0.849)					
Number of secondary schools in cluster, 1998	-0.132	-0.091	*	-0.150	-0.102	*
	(0.041)			(0.040)		
Student to classroom ratio in secondary schools in cluster, 1998	0.002	0.081	**	0.002	0.084	*
	(0.001)			(0.001)		
Teacher to student ratio in secondary schools in cluster, 1998	-0.408	-0.008				
	(1.517)					
Adjusted R-squared	0.25			0.260		
df	976			1045		

=Sig at 1% level; ** = Sig at 5% level; *= Sig at 10% level; ****=Sig at 15% level; *****Sig at 20% level*
Figures in brackets are standard errors of the co-efficients

Table 10: OLS results for African households resident in the community since at least 1990. (Preliminary control for endogeneity)

Dependent Variable: Ln of average schooling gap ratio of child, 1998	***B***	***Beta***
Constant	-7.351	*
	(0.899)	
Child lives in female headed HH	0.288	0.081 *
	(0.101)	
Household is rural	-0.086	-0.019
	(0.147)	
Child is female	-0.297	-0.096 *
	(0.087)	
Age of child, 1998	1.801	3.405 *
	(0.147)	
Age of child (1998) squared	-0.067	-3.157 *
	(0.006)	
Ln Mother's education, 1998	-0.259	-0.138 *
	(0.070)	
Ln father's education, 1998	-0.134	-0.076 **
	(0.066)	
Cluster Unemployment rate, 1998	-1.351	-0.053 ***
	(0.829)	
Fraction of HHs in cluster earning more than R3000 (real) per month, 1998	-1.030	-0.096 *
	(0.390)	
Number of secondary schools in cluster, 1998	-0.164	-0.111 *
	(0.045)	
Student to classroom ratio in secondary schools in cluster, 1998	0.003	0.094 *
	(0.001)	
Adj R-squared	0.29	
df	890	

=Sig at 1% level; ** = Sig at 5% level; *= Sig at 10% level; ****=Sig at 15% level; *****Sig at 20% level*
Figures in brackets are standard errors of the co-efficients

Table 11: Peer effects, neighbourhood variables and schooling quality

Dependent Variable: Ln of average schooling gap ratio of child, 1998		
	B	***Beta***
Constant	-13.571	*
	(1.260)	
Ln of schooling gap ratio of peers in community, 1998	1.637	0.332 *
	(0.210)	
Ln of mother's education, 1998	-0.135	-0.072 **
	(0.067)	
Ln of father's education, 1998	-0.088	-0.049 *****
	(0.064)	
Child lives in female headed HH	0.228	0.062 **
	(0.100)	
Child is female	-0.226	-0.073 *
	(0.083)	
Age of child, 1998	1.792	3.362 *
	(0.140)	
Age of child (1998) squared	-0.066	-3.125 *
	(0.006)	
Ln of real HH income, 1998	-0.033	-0.023
	(0.044)	
Fraction of HHs in cluster earning less than R1000 (REAL) per month, 1998	-0.412	-0.057 ****
	(0.274)	
Cluster Unemployment rate 1998	-0.230	-0.010
	(0.792)	
Fraction of high status workers in cluster, 1998	0.394	0.019
	(0.919)	
Fraction of HHs in cluster earning more than R3000 (real) per month, 1998	-0.003	0.000
	(0.508)	
Number of secondary schools in cluster, 1998	-0.048	-0.033
	(0.044)	
Student to classroom ratio in secondary schools in cluster, 1998	0.000	0.002
	(0.001)	
Teacher to student ratio in secondary schools in cluster, 1998	0.662	0.014
	(1.487)	
Adj R-squared	0.3	
df	973	

=Sig at 1% level; ** = Sig at 5% level; *= Sig at 10% level; ****=Sig at 15% level; *****Sig at 20% level*
Figures in brackets are standard errors of the co-efficients

Appendix E: Disentangling neighbourhood effects

A. Peer effects: Arguably, peer effects, in conjunction with neighbourhood effects, may be an important factor in schooling gap ratios. Case and Katz (Case, 1991), in their study of peer effects, propose the use of probit equations in alleviating endogeneity problems.[33] Following their methodology, I will run two sets of probit equations. Model 1 estimates individual schooling outcomes as a function of individual and family background characteristics, and records the log-likelihood for these models. The second set of probits is identical to the first, but now includes a measure of the schooling gap ratios of peers in the community. I will then employ log-likelihood tests to test for the joint significance of the explanatory variables of peers.

Model 1: $Y_i^* = X_i b + u_i,\quad Y_i = 1$ if $Y_i^* \rangle 0$ and $Y_i^* = 0$ otherwise.

Model 2: $Y_i^* = X_i b + lX_j + u_i,\ Y_i = 1$ if $Y_i^* \rangle 0$ and $Y_i^* = 0$ otherwise

where X_i is a vector of family background characteristics, and X_j is the schooling gap ratio of other children in the neighbourhood. $Y_i=1$ if the child has a schooling gap ratio above the mean schooling gap ratio for the entire sample.

This model can easily be adapted to test explicitly for neighbourhood effects, as opposed to peer effects, via:

Model 1: $Y_i^* = X_i b + u_i,\quad Y_i = 1$ if $Y_i^* \rangle 0$ and $Y_i^* = 0$ otherwise.

Model 2: $Y_i^* = X_i b + lZ + u_i,\ Y_i = 1$ if $Y_i^* \rangle 0$ and $Y_i^* = 0$ otherwise

where Z is now a vector of neighbourhood level characteristics and all other variables are as above.

[33] Following Chamberlain (1985) they argue that in the case of time series data, one can add lagged exogenous variables to the model and test their significance. If current behaviour fails to respond to lagged exogenous variables, then this is evidence that no interdependence is taking place.

B. Treatment effects: Following Greene (Greene, 1997) a sample selection model can be applied to treatment effects.

$$SG_c = b'x_i + dN_i + e_i$$

where SG_i is the schooling gap ratio of the *ith* individual, and N_i is a dummy variable indicating whether the individual lives in a good quality or poor quality neighbourhood. However, using OLS, d doesn't measure the value of living in a good neighbourhood if the typical child who lives in a good neighbourhood would have a low schooling gap ratio irrespective of whether they lived in a good or poor quality neighbourhood. In this case, d will over-estimate the treatment effect.

Thus, we should model neighbourhood quality as a probit equation:

$$N_i^* = gw_i + u_i$$

$N_i = 1$ *if* $N_i^* > 0$, *0 otherwise*

Note: One option for measuring neighbourhood quality is to use the available data to construct an HDR-type index for each community. An alternative is to rank communities on the basis of their average deviations from the sample mean for a range of neighbourhood quality measures, such as unemployment rates, income levels etc.

If U_i and e_i are correlated, then for individuals living in good quality neighbourhoods:

$$\begin{aligned} E\left[SG_i \middle| N_i = 1\right] &= b''x_i + d + E\left[e_i \middle| N_i = 1\right] \\ &= b'x_i + d + r s_e l\left(-g'w_i\right) \qquad (1) \end{aligned}$$

For individuals living in poor quality neighbourhoods, the corresponding equation is:

$$E\left[SG_i \middle| N_i = 0\right] = b'x_i + r s_e \left[\frac{-f(g'w_i)}{1 - j(g'w_i)}\right] \qquad (2)$$

The difference in the expected schooling gap between individuals living in good quality neighbourhoods versus poor quality neighbourhoods is:

$$E(\left[SG_i \middle| N_i = 1\right] - E\left[SG_i \middle| N_i = 0\right] = d + r s_e \left[\frac{f_i}{j_i(1 - j_i)}\right] \qquad (3)$$

This can be estimated using the two-step Heckman procedure. First, the probit equation is estimated by maximum likelihood to obtain estimates of g. For each observation in the selected sample, we then calculate $\hat{l}_i = f(\hat{g}w_i)/j(\hat{g}w_i)$ and $\hat{d}_i = \hat{l}_i(\hat{l}_i + \hat{g}w_i)$. This is used to estimate b and $b_l = r s_e$ by least squares regression of *SG* on *x* and $\hat{l}$.

An extension to this model is to follow the latent variable model specified in Manski et al. (Manski, 1992) I assume that the children aged 6-17 on the KIDS data set are drawn from a population of school children, each of whom is characterised by values for the variables (SG_h, SG_l, z, x). Here, x is the vector of observed covariates describing family background, while z is a binary variable indicating whether the individual lives in a good quality (N=1) or poor quality neighbourhood (N=0). Each child is characterised by two hypothetical outcomes: they could have a low or high schooling gap ratio, depending on the quality of neighbourhood they live in. Of these two outcomes, one is realised and the other is latent.

Possible outcomes:

A: N=1: SG_h=1; SG_h= 0 otherwise

In a good quality neighbourhood, a child can have a below average schooling gap ratio (SG_h=1, meaning they fare better in school) or an above average schooling gap ratio (SG_h= 0, indicating high degree of grade repetition)

B. N=0: SG_l=1; SG_l = 0 otherwise

In a poor quality neighbourhood, a child can have a below average schooling gap ratio (SG_l=1) or an above average schooling gap ratio. (SG_l)

SG_h is only realised when N=1, while SG_l is realised only when N=0.Thus, of these two possible outcomes, one is latent.

Thus, $SG = SG_h(N) + (1-N)SG_l$

The effect of neighbourhood quality on schooling gap performance will be

$P(SG_h=1|N) - P(SG_l=1|N)$

This measures how the probability of having a below average schooling gap ratio would vary if neighbourhoods were not self-selected but exogenously assigned.

The parametric latent variable model (Maddala, 1983) can be set up as a system of three equations:

$N = 1$, *if* $Bx+u > 0$
$N=0$ *otherwise*

$SG_h = 1$ *if* $Cx + e_h > 0$
$SG_h = 0$ *otherwise*

$SG_l = 1$ *if* $Cx+A+e_l>0$
$SG_l = 0$ *otherwise*

Here, B and C are parameter vectors, while x is a vector of covariates including the intercept term. The scalar parameter A allows the intercept term in the equation explaining SG_h to differ from that in the equation explaining SG_l. The contribution of unobserved covariates to the determination of child outcomes and neighbourhood quality is represented by the disturbances (u, eh, el), which are assumed to be statistically independent of x and distributed trivariate normal, with zero mean and variances equal to one.
It follows that:

$P(SG_h=1| x)=F(Cx)$
$P(SG_l=1| x)=F(Cx+A)$

If one assumes that u is statistically independent of (e_h,e_l), i.e. that the unobserved factors that affect neighbourhood quality and schooling gap ratios are unrelated, then,

$P(SG_h=1| x)=P(SG_h=1| x, z)$
and
$P(SG_l=1|x)=P(SG_l=1|x, z)$.

Hence, a child's latent schooling gap ratio outcomes are statistically independent of neighbourhood quality, conditional on the covariates x. When this holds, neighbourhood quality is said to be exogenous. This implies that $P(SG_h=1| x, z=0) = (Cx)$ *and* $P(SG_l=1| x, z=1) = F(Cx+A)$.

The parameters C and A can be estimated by maximising this binary probit likelihood.
Alternatively, one could estimate the original system of three equations using maximum likelihood methods.

Bibliography

Dawood, Z. "Race and Space: Dispossession through the Group Areas Act," Surplus People's Project, 1994.

Haveman, R.; and Wolfe, B. "The Determinants of Children's Attainments: A Review of Methods and Findings." *Journal of Economic Literature*, 1995, *XXXIII*(December), pp. 1829-78.

Jencks, C. et al. *Inequality:A Reassessment of the Effect of Family and Schooling in America*. New York: Basic Books, 1972.

Blau, P and Duncan, D.D. *The American Occupational Structure*. New York: Wiley, 1967.

Featherman, D and Hauser, R. *Opportunity and Change*. New York: Academic Press, 1978.

Datcher, L. "Effects of Community and Family Background on Achievement." *Review of Economics and Statistics*, 1982, *64*, pp. 32-41.

Mulligan, C.B. "Galton Versus the Human Capital Approach to Inheritance." *Journal of Political Economy*, 1999, *107*(6), pp. S184-S223.

Hill, A.M. and King, E.M. "Women's Education and Economic Well-Being." *Feminist Economics*, 1995, *1*(2), pp. 21-46.

Schultz, P.T. "Returns to Women's Education," E. M. a. H. King, M.A., *Women's Education in Developing Countries: Barriers, Benefits, and Policies*. Baltimore and London: John Hopkins University Press, 1993, 51-99.

Case, A.; and Katz, L. "The Company You Keep: The Effects of Family and Neighbourhood on Disadvantaged Youths," National Bureau of Economic Research, 1991.

Burns, J. "Inheriting the Future from the Past: Intergenerational Persistence in Educational Status in South Africa," University of Massachusetts, Amherst, 2000.

Behrman, J.R.; and Rosenzweig, M.R. "Does Increasing Women's Schooling Raise the Schooling of the Next Generation?" *American Economic Review (forthcoming)*, 2001.

Sewell, W.H., Hauser, R.M. and Wolf, W.C. "Sex, Schooling and Occupational Status." *American Journal of Sociology*, 1980, *86*, pp. 551-83.

Bowles, S. and Gintis, H. "The Inheritance of Economic Status: Education, Class and Genetics," M. Feldman, *Genetics, Behaviour and Society*. Oxford: Elsevier, 2001,

Hill, C.R. and Stafford, F. "Allocation of Time to Pre-School Children and Educational Opportunity." *Journal of Human Resources*, 1974, *9*(Summer), pp. 323-41.

Hill, M. and Duncan, G. "Parental Family Income and the Socioeconomic Attainment of Children." *Soc. Sci Res*, 1987, *16*(1), pp. 39-73.

Blake, J. "Number of Siblings and Educational Mobility." *American Sociological Review*, 1985, *50*(February), pp. 84-94.

Altonji, J.G. "The Effects of Family Background and School Characteristics on Education and Labour Market Outcomes," Northwestern University, 1988.

Cooper, S.J.; Durlauf, S.N. and Johnson, P.A. "On the Evolution of Economic Status across Generations." *Social Systems Research Institute Reprint Series*, 1995, *422*, pp. 50-58.

Durlauf, S.N. "The Memberships Theory of Inequality: Ideas and Implications," University of Wisconsin, Madison, 1997.

Jencks, C. ; Meyer, S.E. "The Social Consequences of Growing up in a Poor Neighborhood," a M. G. H. M. Laurence E. Lynn Jr., *Inner-City Poverty in the United States*. Washington, D.C.: National Academy, 1990,

Loury, G. "Essays on the Theory of the Distribution of Income," MIT, 1976.

Durlauf, S.N. "A Theory of Persistent Income Inequality," University of Wisconsin, Madison, 1993a.

____. "Neighborhood Feedbacks, Endogenous Stratification and Income Inequality," University of Wisconsin, Maddison, 1993b.

Benabou, R. "Workings of a City: Location, Education and Production." 1992a.

Maynard, R.A. "The Effects of the Rural Income Maintenance Experiment on the School Performance of Children". *The American Economic Review*, 1977, 67 (1), pp. 370-375.

Montgomery, J. "Social Networks and Labour-Market Outcomes: Towards an Economic Analysis." *American Economic Review*, 1991, *81*, pp. 1408-18.

Streufert, P.A. "The Effect of Underclass Isolation on Schooling Choice," University of Wisconsin, Madison, 1991.

Corcoran, M.; Gordon, R.; Laren, D.; and Solon, G. "The Association between Men's Economic Status and Their Family and Community Origins." *Journal of Human Resources*, 1992, *27*(4), pp. 575-601.

Brooks-Gunn, J.; Duncan, G.J.; Klebanov, P.K.; and Sealand, N. "Do Neighbourhoods Influence Child and Adolescent Development?" *American Journal of Sociology*, 1993, *99*(2), pp. 353-95.

Crane, J. "The Epidemic Theory of Ghettos and Neighborhood Effects on Dropping out and Teenage Childbearing." *American Journal of Sociology*, 1991, *96*(5), pp. 1226-59.

Massey, D.S.; Gross, A.B.; and Eggers, M.L. "Segregation, the Concentration of Poverty, and the Life Chances of Individuals." *Social Science Research*, 1992, *20*(4), pp. 397-420.

Manski, C.F. "Identification of Endogenous Social Effects: The Reflection Problem." *Review of Economic Studies*, 1993b, *60*(3), pp. 531-42.

Evans, W; Oates, , W. and Schwab,R. "Measuring Peer Group Effects: A Study of Teenage Behaviour." *Journal of Political Economy*, 1992, *100*(5), pp. 966-91.

Solon, G.R. "Intergenerational Income Mobility in the United States." *American Economic Review*, 1992, *82*, pp. 393-408.

Zimmerman, D.J. "Regression toward Mediocrity in Economic Stature." *American Economic Review*, 1992, *82*(409-429).

Stadler, A. *The Political Economy of Modern South Africa*. Cape Town, South Africa: David Phillip, 1987.

Robinson, J. *Power of Apartheid: State, Power and Space in South African Cities*. Butterworth-Heinemann Ltd, 1996.

Coleman, J.S. "Social Capital in the Creation of Human Capital." *American Journal of Sociology, Supplement*, 1988, *94*, pp. S95-120.

Anderson, K.G.; Lam, D; and Kaplan, H. "Grade Repetition, Schooling Attainment, and Family Background in South Africa," University of Ann Arbor, Michigan, 2000.

Greene, W.H. *Econometric Analysis*. New Jersey: Prentice Hall, 1997.

Manski, C.F.; Sandfeur, .D., McLanahan, S. and Powers, D. "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation." *Journal of the American Statistical Association*, 1992, *87*(417), pp. 25-37.

Maddala, .S. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge: Cambridge University Press, 1983.