the review of income and wealth

Review of Income and Wealth
Series 59, Special Issue, October 2013
DOI: 10.1111/roiw.12066

EDUCATIONAL INHERITANCE AND THE DISTRIBUTION OF OCCUPATIONS: EVIDENCE FROM SOUTH AFRICA

BY MALCOLM KESWELL*

Southern Africa Labour and Development Research Unit, University of Cape Town

SARAH GIRDWOOD

Genesis Analytics

AND

MURRAY LEIBBRANDT

Southern Africa Labour and Development Research Unit, University of Cape Town

We analyze the role of educational opportunity in shaping inequality in the distribution of occupations in the long run. We use the timing of political events in the history of the struggle to end Apartheid to devise an identification strategy that permits a causal interpretation of the role of educational opportunity. We find evidence that educational opportunity has a strong conditioning effect on the distribution of occupations in steady state. In particular, African female children who inherit the same level of educational opportunity as their parents are 6 percentage points more likely to be in the bottom of the occupation distribution than if they were exposed to better educational opportunities. An alternative identification strategy based on matching on the probability of educational persistence suggests that this figure is approximately 10 percent for younger cohorts of African female children.

JEL Codes: I24, J24, J62

Keywords: dynastic inequality, intergenerational mobility, occupation

1. Introduction

Formal models of the emergence of poverty traps highlight the interplay of educational investments and occupational structure. A key feature of this literature is the idea that non-convexities in the production of human capital are induced by indivisibilities in its investment as well as imperfections in credit markets. In this class of models, the shape of the aggregate distribution of occupations (and therefore long-run inequality) is strongly dependent on the educational opportunities of the previous generation (Banerjee and Newman, 1993; Galor and Zeira, 1993; Atkinson and Bourguignon, 2000). In this paper we explore

Note: Murray Leibbrandt acknowledges the Research Chairs Initiative of the South African Department of Science and Technology and South African National Research Foundation for funding his work as the Research Chair in Poverty and Inequality. We especially thank Stephan Klasen, Patrizio Piraino, and two anonymous referees for helpful comments as well as participants of the IARIW Conference on "Measuring National Income, Wealth, Poverty, and Inequality in African Countries" held in Cape Town, September 28 to October 1, 2011.

*Correspondence to: Malcolm Keswell, Southern Africa Labour and Development Research Unit, School of Economics, University of Cape Town, Rondebosch 7700, Cape Town, South Africa (malcolm.keswell@uct.ac.za).

this relationship empirically for the case of South Africa. If the predictions of these models are correct, we would expect to see a gradient in the relationship between climbing the occupational ladder, and increasing educational opportunities.

Our paper is part of the broader literature on social mobility. In this literature a great deal of emphasis is often placed on the intergenerational linkage in occupations. For example, standard Markov models, as well as the less standard "mover–stayer" variety of such models, focus attention on the relationship in occupational status between parents and their children. In such studies, the occupational outcomes of the children are regressed against the occupational outcomes of their parents. However, viewed from the poverty traps literature cited above, these types of regressions are misspecified. The occupational outcomes of parents determine the educational opportunities of their children, since parents who have jobs that are higher up on the occupational ladder can afford better schooling for their children. To give a causal interpretation to the intergenerational association in occupations, one has to control for the child's opportunity set to acquire more schooling.

This is the problem we seek to address in this paper. In particular, we seek to understand the extent to which educational opportunity might be an important determinant of long-run occupational structure. More precisely, we ask: does having more or less educational *opportunity* affect the distribution of occupations in steady-state in a way that is consistent with the hypothesis that human capital matters for long-run occupational structure? Note that this question *transcends* the issue of identifying the causal effect of education on occupations. To put it differently, if in every generation, children face no impediments (market and non-market) to acquiring additional schooling then we should expect a relatively equal occupational distribution over time: children of relatively poor families could simply borrow against future earnings to finance educational investments and since ability is probably normally distributed in the population, unequal advantage in any generation is offset by a reversal in the next generation and so on. On the other hand, models of poverty traps are premised on the assumption that (unjust) advantage can persist. In the canonical model of a poverty trap, this assumption is usually encapsulated by the twin assumptions that educational investments are by nature lumpy, and that credit markets are imperfect. Our central interest in this paper is to look for evidence that would be consistent with this claim in the context of South Africa; a country well suited to the task at hand because it has historically witnessed very strong market and non-market distortions that could be expected to have affected at least three inter-related outcomes relevant to the question we seek to address: (1) the average attainment of schooling for the Black population; (2) the occupational return to acquired schooling; and (3) the rate at which previous (dis)advantage was allowed to be wiped out.

To investigate the empirical salience of persistence on long-run inequality, we have to allow for the possibility that the intergenerational rate of mobility in education does not follow a standard Markov process. There are many possible ways we might operationalize this idea, but a fitting candidate is the adapted Markov model known as the "mover–stayer" model. The intuition for this approach is not hard to see: with certainty, some fraction of the population does not move at all; i.e., they acquire the same level of schooling as their parents. This slight tweak in the framework is tantamount to a reformulation of the matrix of transition probabilities that govern the dynamics of educational attainment between generations and the main prediction is that *in general* this matrix will not contain a fixed-point vector so that the long-run distribution of occupations depends on the starting point. Thus the "stayers" will face very different trajectories in the long run. The fact that Black South Africans faced severe market and non-market impediments to acquiring more schooling seems to fit this framework quite well. It is of interest to know if the effects of these impediments might have long-lasting effects. It becomes all the more interesting to look at this issue now that average attainment among the Black population has risen sharply.

The paper is structured as follows. We begin in Section 2 by describing the data used in this study. There are two objectives of this discussion. The first objective is to demonstrate that there is variation in educational attainment by cohort, as this is crucial to identifying the effects we are interested in. The second objective is to describe our method of measuring occupational attainment.

Section 3 then outlines the core analytical framework of the paper. Our main objective here is to sketch a framework that is capable of backing out a causal effect of educational opportunity on the steady-state distribution of occupations. The precise way in which we do this proceeds in three steps. The first step is to estimate the extent to which a child's occupational status is conditioned by his or her parents' occupational status, controlling for educational opportunity. These estimates are then used to construct transition matrices of occupational mobility. In the second step, we use these transition matrices to compute steady-state distributions of occupations for each level of educational opportunity, and for each age cohort. This step in the analysis is key and we describe in detail the underlying models that give rise to the steady-state distribution that we later go on to estimate. In particular we consider the effect on the steady-state distribution under a standard first-order Markov process, and then compare this against a stochastic process of the mover–stayer variety, where some groups exhibit zero mobility with certainty. In the third and final step, we then compare these steady-state distributions between the young and older cohorts, holding the level of opportunity constant. The idea here is that if we are able to hold educational opportunity constant, then a comparison of the effect of any given level of opportunity between the old cohort and the young cohort captures the exogenous part of opportunity, and thus identifies the effect of interest.

Section 4 presents a test of the main hypothesis of the paper, which derives from the literature on poverty traps. The main empirical content of this body of theory is that we should expect to see a gradient between better educational opportunities and movement up the occupational ladder. We find that there is clear evidence supporting our hypothesis. Better educational opportunity decreases the mass at the bottom of the long-run distribution of occupations, whereas the reverse is true when children merely inherit the same level of educational attainment as their parents. In particular, we find that educational opportunity has a strong conditioning effect on the distribution of occupations in steady state. In particular, African female children who inherit the same level of educational opportunity as their parents are 6 percent more likely to be in the bottom of the occupation distribution than if they were exposed to better educational opportunities. An alternative identification strategy suggests that this figure is approximately 10 percent for younger cohorts of African female children.

The empirical identification of our results is far from clear cut, as omitted ability of the parent might also be correlated with the opportunity set (for acquiring education) of the child, through the occupational attainment of the parent. Our main approach to identification makes use of exogenous variation in schooling attainment derived from the "legacy effect" of Apartheid. The basic components of our empirical strategy are to control for essential features of educational opportunity, alongside the realized schooling attainment of children and occupational outcomes of their parents, and then to examine the effects of varying the level of opportunity for older and younger cohorts of children. The reasoning is that children schooled in the post-Apartheid era would have faced better opportunities for educational advancement than their older counterparts. In particular, we restrict attention to that subset of the population who are 50 years or older. The youngest members of this cohort of children will have completed schooling before the 1976 Soweto riots. In Section 4.2 we interrogate the identifying assumptions underlying our approach. We start by making the case for why prior to the 1976 Soweto uprising, Black educational attainment was exogenously driven. We then employ an alternative identification strategy based on a non-parametric matching approach. These alternative estimates corroborate our findings that educational persistence causes occupational persistence in the long run.

2. Data Description

Much is known about the levels and correlates of inequality in South Africa.[1] By contrast, strikingly little is known about the *dynamics* of inequality. This paper seeks to address this gap in our knowledge, by using data from the first wave of the South African National Income Dynamics Study (NIDS). This poses a challenge for the type of analysis we wish to conduct as it means that we have to synthetically extract the temporal dimensions of the data. It is possible to do this because parental and offspring education as well as occupational status is measured in this first wave.[2] We now turn our attention to describing the key variables in our analysis.

2.1. *Educational Attainment*

In order for our strategy to work, we have to be able to show that there is variation in educational attainment, between young and old cohorts as well as between generations, within and across cohorts. Table 1 contains the key elements necessary for making this case. By restricting the sample to those respondents aged between 20 and 35 and respondents aged 50 and older, we are able to look more

[1]See the review of this extensive literature in Leibbrandt *et al.* (2010).

[2]The NIDS Wave 1 dataset is described in Woolard *et al.* (2010) and the Wave 1 variables for use in the analysis of intergenerational mobility are described in detail in Girdwood and Leibbrandt (2009). Intergenerational mobility is one of the main themes in NIDS and special attention was given to this theme in the first wave of data collection. As additional waves are added to this panel dataset, other topics related to the themes addressed in this paper will become feasible. Examples are the dynamics of income and unemployment.

TABLE 1
Descriptive Statistics: Schooling of Children Aged 20–35

	Age 20–35				Age 50			
	African	Colored	White	Total	African	Colored	White	Total
Child's years of education	10.201 (3.040)	11.026 (3.310)	12.564 (2.485)	10.449 (3.087)	3.675 (4.010)	5.614 (4.172)	12.679 (2.846)	5.702 (5.275)
Mother's years of education	4.417 (4.443)	7.541 (4.081)	11.360 (3.238)	5.186 (4.774)	0.702 (2.116)	2.432 (3.331)	10.508 (3.271)	2.888 (4.696)
Father's years of education	4.348 (4.599)	8.032 (4.668)	11.616 (3.871)	5.178 (5.020)	0.845 (2.329)	2.621 (3.594)	10.543 (3.756)	3.011 (4.818)

Notes: Table shows mean years of schooling for the sample of children aged 20–35 and the corresponding schooling of the matched parental sample. Standard deviations in parentheses. Post-stratification weights are used.

closely at within-generation temporal patterns in schooling attainment. The parents of the younger cohort would be in their late 40s, 50s, and early 60s, whereas the parents from the older cohort would be from the generation born prior to 1945. Thus, in effect this table displays a picture of three generations of educational achievement.

The table shows clearly that mean education is increasing across generations: educational attainment appears to have doubled between the parental generation and the offspring generation (5.2 to 10.2 years). Interestingly, this pattern is closely mirrored for the older cohort as well (which shows an increase in average parental education from 2.95 years to 5.7 years). Not surprisingly, most of this effect is driven by changes in attainment for Africans and Coloreds. This pattern confirms similar findings using the 10 percent sample of the 1991 census (see, for example, Thomas, 1996).

White parents are, on average, more educated than their African and Colored counterparts in both generations, but the gap is shrinking because non-whites have experienced increases in average attainment whereas Whites have little scope for upward mobility. In the present generation, White parents have approximately 2.6 and 1.48 times more years of education than African and Colored parents, respectively; whereas a generation ago, White parents had approximately 13.6 times more and 4.17 times more years of education than African and Colored parents, respectively.

Further evidence of these patterns is reflected in Table 2, which shows the intergenerational transition probabilities for six educational categories by race over the average of parental years of schooling.[3]

[3]The columns of Table 2 refers to the six educational categories indexed by each row, lowest to highest, into which the average of parental schooling falls. The entries in each cell refer to the proportions of individuals occupying the relevant state, where columns sum to 1. These proportions are population weighted using the NIDS post-stratification weights. Further breakdowns of these results (not shown here but available from the authors upon request), by race, gender of the child, gender of the parent, and location reveal several other interesting patterns. For example, rural and Tribal Authority areas have exactly the same profiles in terms of education mobility, whereas urban dwellers are more mobile than their rural or tribal counterparts and experience greater upward mobility and downward mobility. In particular, an urban respondent has a 28 percent probability of attaining the same education level as his or her parent, a 62 percent probability of achieving an education level higher than that of his or her parent, and a 10 percent probability of obtaining an education level less than his or her parent.

TABLE 2
Education Transition Probabilities

	Parental Education Level						
Educational Categories	1 %	2 %	3 %	4 %	5 %	6 %	Total %
Full sample							
No education	32.0	5.6	1.5	1.4	0.5	0.2	15.9
Some primary	24.5	17.8	9.6	3.8	1.3	0.1	15.9
Lower secondary	12.7	16.8	14.7	7.7	6.4	1.6	12.0
Upper secondary	15.9	30.3	33.1	31.9	15.3	10.1	21.7
Completed secondary	12.3	22.9	32.9	36.2	48.3	27.1	23.4
Tertiary	2.7	6.6	8.2	18.9	28.2	60.9	11.1
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Africans							
No education	32.6	5.8	1.5	2.5	0.8	0.4	19.0
Some primary	24.1	17.3	9.9	5.5	3.2	0.4	18.1
Lower secondary	12.6	16.0	13.9	10.1	6.4	2.0	12.7
Upper secondary	15.7	30.6	34.2	33.7	16.5	17.5	22.3
Completed secondary	12.5	23.6	33.3	35.8	52.8	44.1	21.7
Tertiary	2.5	6.7	7.3	12.4	20.3	35.7	6.2
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Whites							
No education	5.1	0.0	0.0	0.0	0.3	0.0	0.3
Some primary	0.0	4.3	2.9	0.0	0.0	0.0	0.4
Lower secondary	28.3	44.5	14.4	2.6	6.4	1.2	7.3
Upper secondary	16.6	27.8	37.5	33.7	12.0	5.4	18.0
Completed secondary	31.2	22.5	34.4	36.8	47.0	18.4	35.5
Tertiary	18.7	1.0	10.8	26.9	34.4	75.0	38.5
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0
Coloreds							
No education	30.2	5.2	2.7	0.0	0.0	0.0	10.2
Some primary	35.3	26.9	15.1	5.6	0.7	0.0	19.0
Lower secondary	15.7	18.2	23.8	9.7	8.3	3.6	15.1
Upper secondary	10.8	28.0	28.1	26.1	24.8	8.0	21.4
Completed secondary	6.5	15.6	21.3	44.7	42.0	18.2	22.5
Tertiary	1.5	6.0	9.0	13.9	24.2	70.2	11.9
Total	100.0	100.0	100.0	100.0	100.0	100.0	100.0

Several features of these transition matrices accord with the story told by the descriptive statistics. First, there is greater mass below the diagonal than above it. Second, Africans and Coloreds are more mobile than Whites. Third, the rate of intergenerational mobility is clearly non-linear, since the very bottom and top quantiles show greater persistence than in any of the other quantiles, and this pattern is especially pronounced for Whites and to a lesser extent Coloreds.

To gain further purchase on these patterns, we look instead at the years of schooling transition. Figure 1 shows a 3-dimensional view of the educational transition matrix. The height of the surface is the unconditional probability that an individual, whose parents have i years of education, will have j years of education. The plot indicates that an individual with parents with the highest level of education is 915 times more likely to achieve that level over an individual whose parents have zero education.

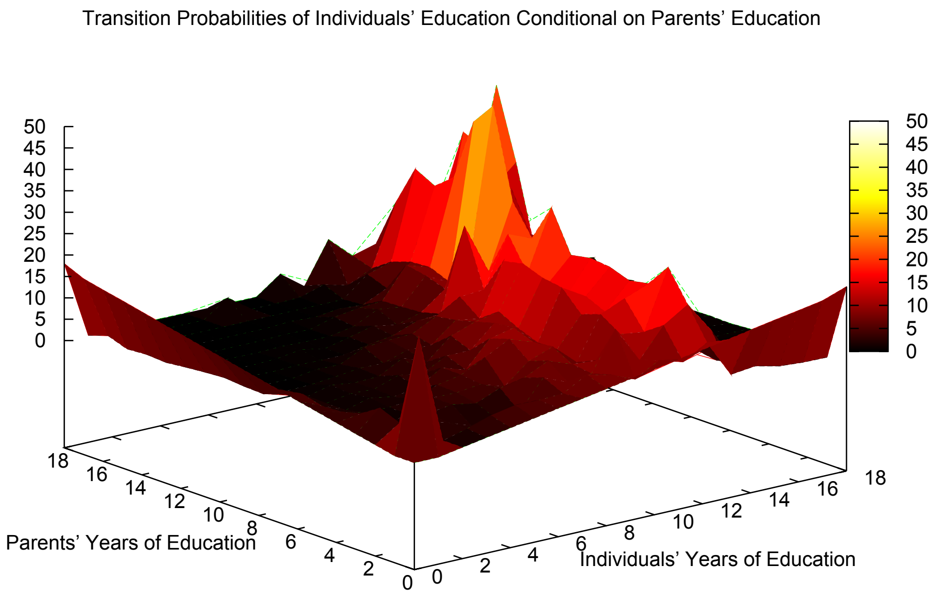

Figure 1. Unconditional Transition Probabilities: 3-D Surface Plots

Figure 2 shows where the highest concentrations of mass lies in Figure 1. The relatively even spread in mass at around 12 years of education for the child demonstrates the advances made in education as well as the weight of secondary school completion in this bivariate distribution. This plot also exhibits the non-linear attainment of education where larger gains in education years are made at the higher end of the parental education distribution than at the lower. A 45 degree line is also apparent, separating the relatively inactive side (dark areas) from the higher probabilities (ligher shaded areas). Figure 3 shows that a major source of this pattern is the increasing numbers of African females completing high school (or equivalent).

Some of these patterns might be explained by the fact that Africans and Coloreds experienced a much higher growth rate in educational attainment in the post-World War II period than did Whites. For example, Louw *et al.* (2007) report that 40 percent of African individuals aged 21–25 in the 1970 census had never enrolled in school and only 1 percent had passed matric, whereas these attainment percentages had improved to 9 percent and 36 percent respectively, for the same age group in the 2001 census. Similar patterns are reported by Thomas (1996) for Africans born in the 1950s and 1960s relative to those born earlier.

2.2. *Occupational Status*

Our coding of occupations is based on the adaptation of the South African Standard Classification of Occupations (SASCO) suggested by Ziervogel and

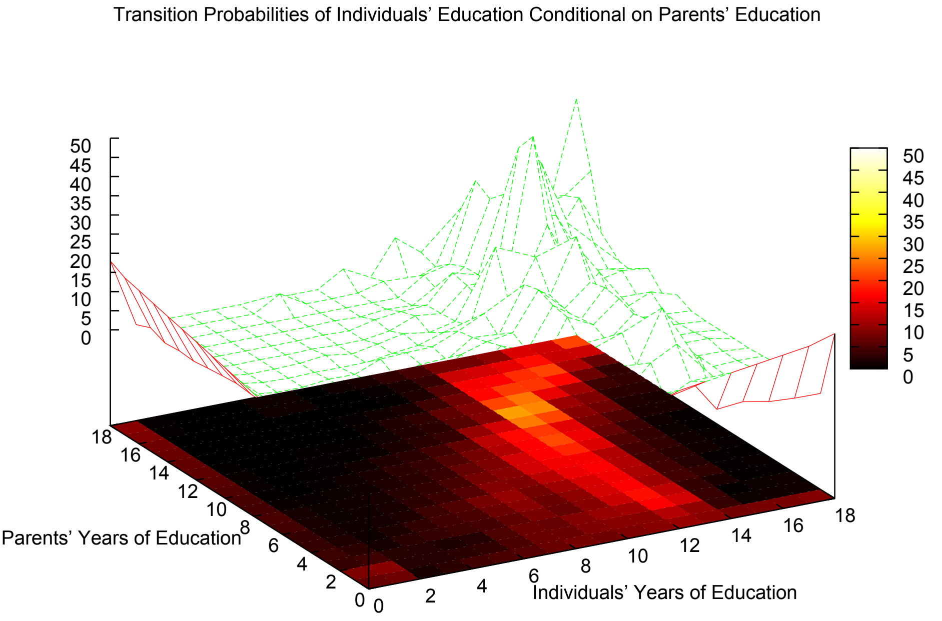

Figure 2. Unconditional Transition Probabilities: Heat plot

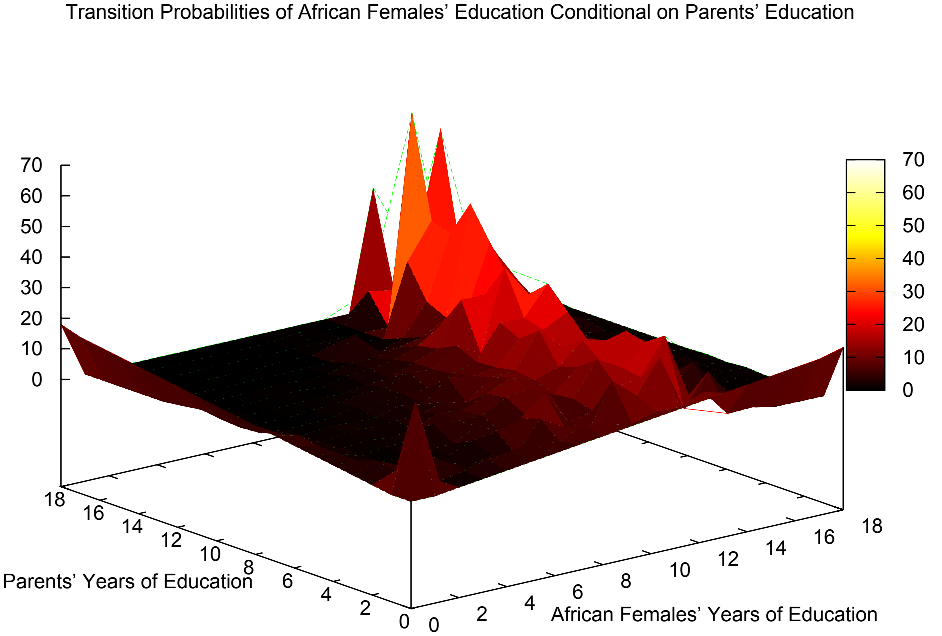

Figure 3. Unconditional Transition Probabilities: 3-D Surface Plots (African Females)

TABLE 3
Occupation Codes and Skill Level (SASCO)

Code	Major Group	Skill Level
1	Legislators, senior officials, and managers	n/a (4)
2	Professionals	4
3	Technicians and associate professionals	3
4	Clerks	2
5	Service workers and shop and market sales workers	2
6	Skilled agricultural and fishery workers	2
7	Craft and related trades workers	2
8	Plant and machinery operators and assemblers	2
9	Elementary occupations	1
0	Armed forces and unspecified occupations	n/a (1)

Crankshaw (2009). This coding convention, developed by Statistics South Africa, is based on the International Standard Classification of Occupation 1988 (ISCO-88) of the International Labour Office (ILO). Table 3 describes how the SASCO coding convention maps into skill levels. These skill levels are an ISCO-88 convention that aims to classify work, in the first instance, according to tasks and duties related to an occupation and, in the second instance, according to the relevant skills that are necessary for fulfilling the formal and practical requirements of a particular occupation (Bergman and Joye, 2001). The skill levels associated with each major group are based on education qualifications and thus serve to transform the SASCO occupational categories into a quasi-hierarchical variable, with ordered occupational levels.

We used the SASCO codes to construct three variables: child's occupation, mother's occupation, and father's occupation. The occupational status of the child was created by taking the 1-digit occupation codes from section E of the NIDS adult questionnaire for regular work 1, regular work 2, casual work, self-employed work, and the occupation code for when the individual once ever worked.[4]

A major problem with the skill levels embedded in the SASCO approach is that two of the four categories relate to tertiary education. Due to the sparse nature of tertiary education qualifications in South Africa, this approach is likely to lead to artifactual biases in the estimates of transition probabilities between the four states. To deal with this problem, we follow Ziervogel and Crankshaw (2009) by reassigning the ISCO-88 major groups to four skills groups that more accurately reflect the distribution of skills in South Africa. Table 4 shows the effects of our recoding exercise in this regard.

3. Modeling the Distribution of Occupations

Synthetic temporal comparisons of the distribution of occupations in repeated cross-sectional surveys can serve as a useful diagnostic about unfolding

[4]Precedence was given to the occupation from regular work 1 in the case of multiple jobs. Parental occupation variables include information from two sources: for non-resident or deceased parents, we used the respondent's recollection of their parents' current or last occupation captured in section D of the NIDS adult questionnaire; for resident parents, we used the respondent's self-reported occupational status reported in section E of the NIDS adult questionnaire.

TABLE 4
ISCO88 Occupation Skill Groups

Skill Level	Education Qualification
Level 1	Primary education (approx. 5 years)
Level 2	Secondary education (between 5 and 7 years)
Level 3	Tertiary education (between 3 and 4 years): not leading to a university degree
Level 4	Tertiary education (between 3 and 6 years): leading to a university degree or equivalent

inequality. But they are only descriptive and not predictive at the end of the day. To shed light on a *causal* question of the sort we have in mind requires a theory that: (a) adequately specifies the assumptions that underpin the dynamics of the processes governing transitions between relative positions within the bivariate occupational distribution; and (b) links these assumptions to the existence of a steady-state distribution. In order to analyze whether educational opportunity predicts occupational structure in steady-state, we have to know that some type of steady-state is actually possible. In this section, we outline two such frameworks. One permits the opportunity to play a role, and the other does not. Under the first framework, we make the strong assumption that the population is homogenous with respect to the rate of transition between occupational states; in this framework long-run outcomes are determined purely stochastically, so educational opportunities are not allowed to play any role. Therefore the first framework corresponds to a standard first-order Markov process. In the second framework, we explicitly relax these assumptions. In both cases, our objective is to derive the relevant steady-state distribution as these equations are the focal point of the empirical estimates presented in Section 4.

3.1. *Equal Opportunities: Basic Markov Process*

We start by indexing generations as a discrete variable. An individual must occupy exactly one of a finite number of discrete states from the set of states $\mathcal{N} = \{1, \ldots, N\}$. Thus we define p_{ij} as the probability of an individual ending up in occupation level i after a single generation, given that this individual's parents started out in occupation level j in the previous time period. These p_{ij} define the following matrix of transition probabilities (which by definition have to be positive).

$$P = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix}. \tag{1}$$

This matrix describes a so called one-step transition process. It fully describes what it would take to reconcile differences in the distribution of occupations that can be expected to emerge within a generation. Since each individual starting in any given state must also end up in one of the N states, it must be the case that the columns of this matrix sum to one:

$$\sum_{j=1}^{N} p_{ij} = 1 \quad \forall \quad p_{ij} \geq 0.$$

In our formulation, these probabilities are subscripted p_{ij} implying that the first subscript indexes the offspring generation, whereas the second subscript indexes the parental generation. If we assume that p_{ij} are fixed and independent of generations (i.e., P is stationary), then the dynamics of this system follow a Markov process. To describe these underlying dynamics, we denote x_j^{n-1} as the fraction of a population of size N_1 that is in state j in generation $n - 1$, so that the total number of members of this population found in state j in generation n is given by $x_j^n N_1$. By stationarity, we have

$$x_i^{(n)} N_1 = \sum_{j=1}^{N} p_{ij} x_j^{(n-1)} N_1. \tag{2}$$

In words, the total number of members of this population that we can expect to find in state i in generation n is given by the sum over all of the members occupying state j in generation $n - 1$ that have moved into state i. Under the standard (first-order) Markov process, an individual must have been observed in one of the N states in generation $n - 1$, and then must move from state j to i within a generation (or more precisely, in the immediately preceding step of the Markov chain). Given that we can construct equations of this sort for all $N \cdot N$ possible transitions, we can put the resulting system of equations into matrix form (after dividing through by N_1), giving

$$\begin{pmatrix} x_1^{(n)} \\ x_2^{(n)} \\ \vdots \\ x_N^{(n)} \end{pmatrix} = \begin{pmatrix} p_{11} & p_{12} & \cdots & p_{1N} \\ p_{21} & p_{22} & \cdots & p_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ p_{N1} & p_{N2} & \cdots & p_{NN} \end{pmatrix} \begin{pmatrix} x_1^{(n-1)} \\ x_2^{(n-1)} \\ \vdots \\ x_N^{(n-1)} \end{pmatrix} \tag{3}$$

or more compactly, $\mathbf{x}^{(n)} = (x_1^{(n)}, x_2^{(n)}, \ldots, x_N^{(n)})' = P\mathbf{x}^{(n-1)}$ where we have denoted $\mathbf{x}^{(n-1)} = (x_1^{(n-1)}, x_2^{(n-1)}, \ldots, x_N^{(n-1)})'$. The right-hand side is a Markov chain over states 1, 2, . . . , N. The important thing to note about system (3) is that it is recursive, which implies that

$$\mathbf{x}^{(n)} = P^{(n)}\mathbf{x}^{(0)} \tag{4}$$

$$\mathbf{x}^{(n)} = P^{n}\mathbf{x}^{(0)}. \tag{5}$$

This is because the n-th power of the one-step transition matrix P is equal to the n-step transition matrix; $P^n = P^{(n)}$. If we restrict attention of P to Markov matrices, some power of which only ever has positive entries, then P is said to be a regular Markov matrix, and by a standard limit theorem of regular Markov chains (see, for example, Feller, 1950; Karlin and Taylor, 1975), we have the following result: (a) 1 is an eigenvalue of P of multiplicity 1; the absolute value of every other eigenvalue of P is always less than 1; eigenvalue 1 has eigenvector $\mathbf{w}_1$ with strictly positive components; if we normalize $\mathbf{w}_1$ by the sum of its components (which we can do by dint of the structure of the general solution of system 3), and call this new vector $\mathbf{v}_1$, then this new vector will represent a fixed-point *probability* vector.[5]

The empirical discussion presented in Section 4 starts by computing the matrix P, and then, by the standard result just stated, we know that the eigenvector corresponding to the eigenvalue 1 (which itself is always an eigenvalue of a matrix like P), will always be a fixed point vector. This is the starting point to characterizing the steady-state distribution. We will now use this fact about regular Markov matrices to develop the key equation we have to estimate to answer the main question posed in this paper.

3.2. *Unequal Opportunities: Mover–Stayer Markov Process*

In this section, we modify the standard Markov process to account for the possibility that better educational opportunities will affect the steady-state distribution in occupations. This can only happen if the conditional probability of a child's occupational choice, given parental occupation is affected by whether the child was exposed to better educational opportunities than their parents. For expositional purposes, let us denote two types of individuals: those who acquire more education than the highest level attained by either of their parents are denoted as "high opportunity children"; and those who acquire less education than the highest level attained by either of their parents are denoted as "low opportunity children." The key departure under this adapted Markov process is that we no longer assume that all types of children (high and low opportunity alike) are homogenous with respect to the transition matrix. In other words, the matrix P (which was key to the derivation of the steady-state under the standard model that we outlined above) no longer applies. We require separate matrices that will be unique to the specific type of individual, while at the same time, possessing the same properties as the original matrix P that guarantees that a steady-state of the distribution of occupations will exist for each type. Our objective below is to explain how these type-specific matrices are defined.

The main tweak to the model is that we now allow different groups (defined along opportunity type, race, and gender) to have different transition probabilities. This type of model is sometimes dubbed the "mover–stayer" Markov model. The application of this model to occupational mobility was first carried out by Blumen *et al.* (1955). Our formulation follows that of Goodman (1961). In particular, we denote as *movers* people who have some non-zero probability of

[5]More precisely, as $n \to \infty$, we must have $P^n \to W$, where the matrix W has identical rows all equal to $\mathbf{w}_1$, such that $\mathbf{w}_1 = \mathbf{w}_1 P$. The fact that $\mathbf{w}_1$ is left unchanged after post-multiplying by P means that $\mathbf{w}_1$ must be a fixed point vector of P.

changing occupational classes between generations and *stayers* as people who will persist in the same occupations as their parents with certainty. For the moment, we assume that the matrix of transition probabilities for the movers is constant across all movers and is given by the matrix M, whose elements m_{ij} are the probabilities of transition from the j-th occupational class in the parental generation to the i-th occupational class in the offspring generation. By contrast the matrix of transition probabilities of the stayers is given by the identity matrix. This process implies that someone observed in the present generation as occupying the i-th state, whose parents also occupied the i-th state, could persist in this category in two ways: the dynamics of the social structure determine that he is sure to remain stuck in the same occupational rung as his parents, or *by chance* he ends up in the same occupational rung with probability m_{ii}. As we will see momentarily, the distinguishing feature of this adapted Markov model is that the classic limit theorem for a standard Markov process will carry through for the movers in this set-up, but unlike the standard model, the limiting matrix of transition probabilities will not have a fixed-point (i.e., it will in general depend on the distribution of occupations in the parental generation).

Let the matrix S denote an $N \times N$ diagonal matrix with s_i representing the fraction of people in the i-th state who will stay there with certainty. Then, by the above description, we can write

$$p_{ij} = \begin{cases} (s_i(1) + (1 - s_i)m_{ii} & \text{if} \quad j = i \\ (1 - s_i)m_{ij} & \text{if} \quad j \neq i \end{cases} \tag{6}$$

or, more compactly

$$P = S + (I - S)M \tag{7}$$

and, for the n-th power of the matrix P, we have

$$P^{(n)} = S + (I - S)M^n \tag{8}$$

and for the k-th sub-population for the n-th generation, we have

$$P_k^{(n)} = S_k + (I - S_k)M_k^n. \tag{9}$$

Since M^n is regular, its limiting matrix, which we denote as V, has a fixed point vector (i.e., all the rows are the same). The same argument in representing the components of this fixed point vector as probabilities made above applies, so we can denote $\mathbf{v}_1$ as the fixed point probability vector of the movers. However, note that V is the limiting matrix of only the movers. The limiting matrix of the combined population of movers *and* stayers is given by the left-hand side of (9), which, as the right-hand side of this equation makes clear, will not have identical rows. Applying equation (4), we will have

$$\mathbf{x}^{(n)} = P^{(n)}\mathbf{x}^{(0)} \rightarrow [S + (I - S)V]\mathbf{x}^{(0)}. \tag{10}$$

We estimate this equation for several sub-populations (chiefly delineated by race and gender). Thus we can write the steady-state distribution for the k-th sub-population as

$$\mathbf{x}_k^{(n)} = P_k^{(n)}\mathbf{x}_k^{(0)} \rightarrow [S_k + (I - S_k)V_k]\mathbf{x}_k^{(0)}. \quad (11)$$

Our main goal in this paper is to estimate equation (11). Before we turn to exactly how we went about doing this, note that the k subscripting operationalizes the notion that steady-states can be group-specific. In the empirical implementation below, groups are defined along three dimensions: opportunity type, race, and gender. Notice also that V_k is the limiting matrix of M_k^n, where the latter is nothing but the n-th power of the occupational transition matrix for the k-th group. So to compute the steady-state vector $\mathbf{x}_k^{(n)}$, we first must estimate M_k^n, then compute the eigenvector that corresponds to the first eigenvalue of this matrix, and finally scale it by the sum of its components. We now turn to these details.

4. Empirical Implementation and Results

Under the standard Markov model, the matrix of transition probabilities P (equation 1) that is used to calculate the steady-state distribution of occupations (equation 5) is comprised of raw percentages of children in each occupation, conditional on the occupation of their parents. However, unequal opportunity operates to change this matrix, such that chance is allowed to affect the transition probabilities of the movers but not the stayers. In Section 3.2, we showed how this sort of heterogeneity in the transition process (as defined by the matrices S and M) leads to a different steady-state distribution (equation 11). Our goal in this paper is to estimate this key equation, in a way that permits us to ask a causal question about this heterogeneity: i.e., how does heterogeneity in one important dimension (educational opportunity) cause the steady-state distribution to change. In order to be able to answer this question, we need to compute equation (11). This in turn requires computation of the matrices M and S, as well as a specification of the measure of educational opportunity that we apply to this task. The empirical content behind these issues is taken up in this section.

4.1. *Estimating the Steady-State Distribution of Occupations*

4.1.1. Estimating M_k and S_k

A distinguishing feature of our approach is that the matrices S and M are not taken as fixed constants but rather estimated directly from the data. In this respect, our approach is much more in line with Goodman (1961). Estimation of these two unknowns is required in order to compute these steady-state distributions. Thus our empirical analysis begins by computing the probabilities of transitioning between occupation levels from one generation to the next; i.e., the elements of the matrix M in equation (7).

Keeping in mind that M is nothing more than a matrix of *predicted* probabilities rather than the raw percentages of people which constitutes the elements of P,

TABLE 5
Occupation Transition Probabilities

	Father's Occupation Level Full Sample				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	38.2	27.0	19.7	7.9	24.7
2	22.7	34.9	26.1	14.9	28.7
3	24.2	21.5	31.3	22.0	23.4
4	14.8	16.6	22.8	55.3	23.2
Total	100.0	100.0	100.0	100.0	100.0
	Mother's Occupation Level Full Sample				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	37.6	25.5	10.0	10.3	27.5
2	27.4	32.7	16.8	15.8	24.7
3	22.9	22.5	37.8	30.0	26.0
4	12.1	19.2	35.4	43.9	21.9
Total	100.0	100.0	100.0	100.0	100.0

the first question of empirical significance that we need to address is what to condition these probabilities on. Here we are guided by what can be learnt from a careful examination of the raw transition matrices.

Tables 5–6 are examples of matrices such as P, computed separately for each parent. The cells in these tables indicate the population weighted proportions of individuals falling into a given state, conditional on the state occupied by the previous generation. It is clear that there is substantial persistence in the very highest level of occupations (Managers and Professionals) in the case of fathers (55 percent), compared to mothers (44 percent). Part of this can be explained by the fact that the correlation in occupational status between parents is quite low (less than 50 percent): given the low correlation across parental occupation status, a high degree of persistence in the status of one parent would suggest a relatively lower degree of persistence in the status of the other parent. These differential patterns of persistence by the gender of the parent suggest that it will be important to control for the occupational statuses of both parents, as well as the gender of the child when modeling the conditional probabilities of the child's occupational status.

Table 6 gives a breakdown by race as well as the genders of the parents. It is clear from these initial diagnostics that Africans and Whites have greater persistence at the top of the father–child conditional distribution, whereas for Coloreds, the mother–child transition dominates persistence at the very top. Therefore, in modeling the conditional probabilities of the child's occupational status, we should also control for race.

The final variable we have to control for is our variable of interest: the educational opportunities of the child. We measure educational opportunity in terms of the educational levels of children relative to their parents, by constructing

TABLE 6

Occupation Transition Probabilities by Race

	Father's Occupation Level Africans				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	40.5	31.4	27.4	13.7	30.7
2	22.2	36.7	25.4	24.0	31.3
3	25.9	20.7	30.4	26.2	23.6
4	11.4	11.1	16.8	36.1	14.4
Total	100.0	100.0	100.0	100.0	100.0

	Mother's Occupation Level Africans				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	38.1	29.2	12.1	18.3	32.5
2	26.6	32.1	15.8	15.9	25.2
3	23.0	25.1	44.0	37.2	26.7
4	12.2	13.5	28.1	28.7	15.6
Total	100.0	100.0	100.0	100.0	100.0

	Father's Occupation Level Coloreds				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	33.6	26.5	6.3	27.6	25.8
2	28.6	34.0	45.2	15.8	32.6
3	19.3	22.0	19.9	19.3	21.1
4	18.5	17.5	28.5	37.3	20.5
Total	100.0	100.0	100.0	100.0	100.0

	Mother's Occupation Level Coloreds				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	35.5	20.4	16.4	14.8	26.4
2	34.9	35.4	28.3	18.4	31.6
3	18.6	17.0	45.1	24.2	21.7
4	11.0	27.2	10.2	42.6	20.3
Total	100.0	100.0	100.0	100.0	100.0

	Father's Occupation Level Whites				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	41.7	3.6	4.4	0.5	3.7
2	13.6	24.3	27.7	6.8	18.0
3	4.9	25.9	35.2	16.5	23.4
4	39.9	46.2	32.7	76.2	54.8
Total	100.0	100.0	100.0	100.0	100.0

	Mother's Occupation Level Whites				
ans_child_occ	1 %	2 %	3 %	4 %	Total %
1	8.4	6.1	5.3	2.2	3.5
2	30.3	11.1	15.8	16.8	16.8
3	52.3	6.9	29.0	25.9	26.8
4	9.0	75.9	49.8	55.1	52.9
Total	100.0	100.0	100.0	100.0	100.0

three dummy variables: children that acquire more schooling than their parents (*high opportunity children*), children that acquire less schooling than their parents (*low opportunity children*), and children that acquire the same amount of schooling than their parents (*same opportunity children*).

To compute the matrix of M_k^n, we estimate ordered logit regressions of the occupational level of the child using as explanatory variables age, age squared, completed schooling, dummies for race, gender, occupational level of the father, occupational level of the mother, age of the father, age of the mother, and dummies for the level of educational opportunity. These ordinal logit regressions underpin the computed transition matrices of the k-th group M_k and are shown in Table 7.[6,7]

Using the estimated coefficients from the regressions shown in Table 7, we then compute predicted probabilities for each possible contrast of the child–parent occupational transition matrix. In computing these predicted probabilities, we hold age and schooling at their sample means while conditioning on race, gender, educational opportunity, and parental occupational status. Since there are four occupation levels, this defines 16 possible contrasts. Since each contrast is for a given level of the child occupational variable against all possible levels of the parent occupation level, the conditional probability of the child being in a given occupation level is defined over four possible values, where the sum of these values must be equal to one. Collecting these probabilities into separate matrices (by each of the conditioning variables) results in M_k.[8] This matrix is a positive matrix, since we only make the computation for the k-th matrix if every possible contrast exists, with non-zero probability. Given this criterion, it also means that the columns of m_k must sum to one, so that M_k satisfies the properties as required by our set-up in Section 3.2.

A final detail concerns how we estimate the other aggregate that feeds into the computation of the steady-state distribution: i.e., the matrix of stayers S_k. In Goodman (1961), the estimator of the proportion of stayers in the i-th occupation collapses onto the proportion of people that stay in that category over n periods (for large n). For any given number of periods, the estimator is in fact even more

[6]Interestingly, notice that there appears to be some evidence that educational opportunity for the older cohort operates non-linearly in child and parental education. We thank an anonymous referee for pointing this out. This of course would suggest a non-linear specification of the two education variables. However, we chose to err on the side of parsimony since conditioning on too many variables when computing predicted probabilities increases the risk of the so called "empty cell problem" (see footnote 9 for more on this point). Moreover, since the non-linearity is inferred by looking across the three specifications reported in Table 7, these effects should be fully captured by the separate steady state distributions corresponding to each of these specifications. By contrast, differencing each opportunity specific steady-state vector from the benchmark of the basic Markov model and then examining the gradient of this difference (final column of Table 8) isolates the pure linear model versus the non-linear model which is represented by the undifferenced gradients labeled "Level 1, 2, etc." Thus, Table 8 fully reflects these non-linearities and so it is unclear what is to be gained by further complicating the underlying logit specifications reported in Table 7.

[7]Note also that we do not report the corresponding marginal effects as they are not of direct interest to us, but they can be made available upon request.

[8]The number of transition matrices obviously varies by the number of possible contrasts chosen for parental occupation status. Since we include both mothers' and fathers' occupation dummies as controls (see Table 7), for any given value of the gender, race, and educational opportunity dummies, there exist 16 possible combinations of parental occupation. To keep these computations manageable, we therefore only condition on the probability that *both* parents occupy the same occupational rung.

TABLE 7
OCCUPATIONAL MOBILITY: ORDINAL LOGIT

	B_Young b/se	B_Older b/se	C_Young b/se	C_Older b/se	D_Young b/se	D_Older b/se	E_Young b/se	E_Older b/se
Age	0.313***	0.303***	0.312***	0.332***	0.320***	0.278***	0.313***	0.229***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.01)
Age squared	–0.004***	–0.003***	–0.004***	–0.003***	–0.004***	–0.002***	–0.004***	–0.000***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Male	–0.058***	0.698***	–0.058***	0.730***	–0.056***	0.738***	–0.103***	0.427***
	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)	(0.00)	(0.01)
Black	–0.422***	0.069***	–0.418***	0.397***	–0.416***	0.173***	–0.403***	0.296***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
Colored	0.271***	0.944***	0.263***	1.229***	0.273***	1.116***	0.179***	0.536***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
ans_father_occ==2	0.376***	0.682***	0.375***	0.708***	0.374***	0.639***	0.378***	0.527***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
ans_father_occ==3	–0.170***	–0.113***	–0.174***	0.337***	–0.174***	–0.082***	–0.199***	0.165***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
ans_father_occ==4	0.437***	–0.180***	0.428***	0.130***	0.425***	–0.267***	0.384***	–0.123***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
ans_mother_occ==2	0.439***	1.376***	0.444***	1.025***	0.444***	1.316***	0.461***	1.416***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.02)
ans_mother_occ==3	0.947***	–0.312***	0.949***	–1.002***	0.949***	–0.476***	0.923***	–0.970***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)
ans_mother_occ==4	0.656***	0.941***	0.660***	0.343***	0.656***	0.686***	0.639***	0.508***
	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.02)	(0.01)	(0.03)
Child's years of education	0.303***	0.471***	0.297***	0.220***	0.302***	0.454***	0.244***	0.168***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Parental education	0.074***	–0.004	0.079***	0.242***	0.075***	0.043***	0.071***	0.135***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
High educational opportunity	–0.025***	–1.815***						
	(0.01)	(0.01)						
Low educational opportunity			–0.065***	–1.715***				
			(0.01)	(0.03)				
Same educational opportunity					0.050***	1.815***		
					(0.00)	(0.01)		
Predicted educational persistence							–5.943***	–12.315***
							(0.06)	(0.08)
cut1	7.871***	11.177***	7.826***	12.524***	7.995***	12.243***	6.852***	8.778***
	(0.10)	(0.50)	(0.09)	(0.49)	(0.10)	(0.48)	(0.09)	(0.36)
cut2	9.680***	13.120***	9.635***	14.396***	9.805***	14.196***	8.667***	10.717***
	(0.10)	(0.50)	(0.09)	(0.49)	(0.10)	(0.48)	(0.09)	(0.36)
cut3	11.602***	14.296***	11.559***	15.525***	11.727***	15.395***	10.615***	11.996***
	(0.10)	(0.50)	(0.09)	(0.49)	(0.10)	(0.48)	(0.10)	(0.36)
R-squared	0.164	0.326	0.164	0.309	0.164	0.332	0.168	0.347
BIC	2,756,662.5	547,302.9	2,756,578.5	561,477.5	2,756,556.8	542,528.2	2,744,506.9	530,429.1
lrx2	541,373.1	264,978.4	541,457.1	250,803.8	541,478.8	269,753.2	553,528.7	281,852.2

Notes: Marginal effects. Standard errors in parentheses. (d) for discrete change of dummy variable from 0 to 1. *$p < 0.05$, **$p < 0.01$, ***$p < 0.001$.

complicated and involves the fixed point probability vector for the movers. Since a period in our model represents a generation, it would not be feasible to apply this estimator: even if one were to observe more than two generations, the resulting sample sizes would be vanishingly small and then the curse of selection bias will bite as dynasties comprised of three or more generations are likely to be unobservably different than two-generation dynasties. We therefore employ a rather more parsimonious estimator of s_i: the unconditional proportion of parent–child pairs who do not change occupation categories between generations. We could in principle also use the conditional proportion of parent–child pairs who do not change occupation categories between generations (i.e., like in Goodman (1961), both s_i and m_{ii} could be identified off the same matrix of occupation transition probabilities). However, since the process governing the stayers (with probability one) is by definition different to that governing the movers, the more parsimonious approach makes more sense to us.

4.1.2. Estimates of the Long-Run Distribution $\mathbf{x}_k^{(n)}$

As mentioned above, the Markovian framework imposes the restriction that the mobility matrix has to be a positive matrix. This restriction implies that if some transitions occur with zero probability, the long-run distribution cannot be computed. In what follows, we only present results for African females as all other race–gender groups have at least one zero-probability transition. This restriction also makes sense given our identification strategies, which are discussed at length in Section 4.2.[9]

Table 8 reports the main results of the paper. Panel A shows the estimated distribution of occupations in steady state for a younger cohort of 20–35 year olds, and Panel B reports the same for an older cohort of workers aged 50 and older.

We interpret the results of the Basic Markov Model (row 1) as the predicted long-run distribution under equal educational opportunities. This steady state is the empirical counterpart to equation (5). The key assumption here is that no member of the population faces the prospect that their parents will have occupied the same rung as they are expected to with certainty. In a sense then, this line of the table (for both panels) tell us what we would expect to be the situation where only chance is allowed to determine future outcomes (i.e., under a a pure first-order Markov process where we assume a homogenous transmission process). As equation (5) makes clear, in this model there are no stayers so educational mobility cannot amplify any pre-existing differences in occupational structure. The resulting steady state therefore is unconditional so it is not all that surprising that the distribution of occupations for the younger cohort appears relatively flat. There is

[9]We cannot compute predicted probabilities for all 16 transitions for all four race groups because of a methodological limitation: when an individual occupies some transition with zero probability, there is no way to assign a predicted probability to that transition. When using specifications B, C, and D of Table 7, we encounter this problem for Colored males and Whites because these population groups never make certain conditional transitions (e.g., downwardly mobile occupation but upwardly mobile schooling). With the alternative identification strategy employed in Section 4.2.2, this problem is exacerbated as we also encounter the same difficulty for African males. Therefore to make consistent comparisons in Table 8, we can only compare African females under the two assumptions: one where educational mobility is exogenous, as in rows 2, 3, and 4 in both Panels A and B; the other where it is endogenous (row 5 of both panels).

TABLE 8

Steady-State Distribution of Occupations for African Females

	Level 1	Level 2	Level 3	Level 4	Difference
Panel A: 20–35 Year Olds					
Basic Markov Model	0.25	0.27	0.30	0.18	0.00
Mover–Stayer Markov Model (high opportunity)	0.20	0.51	0.23	0.07	0.19
Mover–Stayer Markov Model (low opportunity)	0.29	0.49	0.18	0.05	0.26
Mover–Stayer Markov Model (same opportunity)	0.49	0.36	0.11	0.03	0.33
Mover–Stayer Markov Model (endogenous persistence)	0.22	0.50	0.22	0.06	0.20
Panel B: 50 Year Olds and Older					
Basic Markov Model	0.30	0.24	0.11	0.35	0.00
Mover–Stayer Markov Model (high opportunity)	0.33	0.46	0.08	0.13	0.25
Mover–Stayer Markov Model (low opportunity)	0.48	0.37	0.05	0.10	0.31
Mover–Stayer Markov Model (same opportunity)	0.35	0.50	0.09	0.06	0.31
Mover–Stayer Markov Model (endogenous persistence)	0.37	0.47	0.08	0.08	0.30

Notes: Mover–stayer steady state distributions are conditional on educational opportunity, measured in terms of the educational categories of children relative to their parents, where "B" is more opportunity, "C" is less opportunity, and "D" is the same opportunity. The educational categories used here are detailed in Table 2. More opportunity is when children obtain more schooling than their fathers (i.e., at least one category higher). Less opportunity is when children obtain less schooling than their fathers. The same opportunity is when children obtain the same schooling as their fathers. The k-th row in the table is $\mathbf{x}_k^{(n)} = P_k^{(n)}\mathbf{x}_k^{(0)} \rightarrow [S_k + (I - S_k)V_k]\mathbf{x}_k^{(0)}$, where V_k is the limiting matrix of M_k^n, the matrix of predicted transition probabilities from an ordinal logit regression of the occupational level of the child against age, age squared, completed schooling, dummies for race, gender, occupational level of the father, occupational level of the mother, age of the father, age of the mother, average parental years of schooling, and educational opportunity. S_k is a diagonal matrix of the same dimension as M_k and $\mathbf{x}_k^{(0)}$ is parental distribution of occupations. Since each contrast is for a given level of the child occupational variable against all possible levels of the parent occupation level, these probabilities must sum to one.

a bit more dispersion in the distribution of the older cohort. This too is not that surprising, given that this model would predict far more accurately for the younger generation which would have been exposed to a more meritocratic labor market structure where chance would play a stronger role.

Rows 2–5 of both panels in Table 8 report the steady-state distributions when educational background is varied. The final column of Table 8 labeled "Difference" is a measure of the difference between the predicted probability of being in the bottom half of the occupational distribution in steady state, and what is predicted by the Basic Markov Model. Setting aside the question of causation for the moment, the first important point to note is that our hypothesis of a gradient between educational opportunity and occupational advancement is borne out by these results. Focusing on Panel A, we see that, relative to the Basic Markov Model, children with higher educational opportunities have an 19 percent higher chance of being in the bottom half of the distribution, whereas this difference increases to 33 percent and 26 percent for children with the same or lower educational opportunities, respectively. Since we are effectively holding educational opportunity constant when looking at any one row among rows 2–4 of either panel, the comparison reflected in the final column can also be interpreted as the direct effect of labor market structure on the steady-state distribution of occupations.

We also see a similar pattern for the older cohort of children (Panel B): higher opportunity children have a 25 percent higher probability of being in the bottom half of the distribution in steady-state compared to 31 percent for children with the same or lower educational opportunities as their parents. Again, if we restrict our attention to any given row in Panel B, these results cannot be due to differences in educational background since within a row educational opportunity is always held constant. The interesting question that now arises of course, is whether we can give a causal interpretation: i.e., does increasing educational opportunity cause these predicted shifts in the distribution, or do they merely reflect the types of unobservable influences mentioned earlier? We now turn to making the case for a causal interpretation.

4.2. *Do Better Educational Opportunities Cause Changes in the Distribution?*

4.2.1. Basic Identification Strategy

It is not unreasonable to assume that the racial variation in educational attainment under Apartheid was exogenously driven. After all, for most of the period in question, it was one of the explicit intentions of Apartheid policy to socially engineer a racially organized class structure in South Africa, and educational policy was a major component of this strategy (Wilson and Ramphele, 1989; Fiske and Ladd, 2004; Louw *et al*., 2007). Our basic identification strategy is this: prior to the Soweto riots of 1976, shifts in educational policy was probably not strongly determined by individual action, whereas after this watershed moment in history, the innovations undertaken by the Apartheid state were probably in response to collective action. Indeed, the trigger for the 1976 Soweto riots was the rejection of the language of Afrikaans as the medium of instruction in Black African schools. Prior to 1976, changes in *average* attainment of Black Africans are arguably better explained by the slow but incremental opening up of greater schooling opportunities by the State alongside other marginal changes that might have improved the quality of schooling; for example, the abolition of the policy of "Bantu education." The assumption we make is that these pre-1976 changes are driven more by supply-side factors. Certainly the key demand side factor (mass mobilization in the domain of education) did not materialize more or less until the mid-1970s, culminating in the Soweto riots of 1976. The period of Apartheid up until 1976 can therefore be thought of as a natural experiment that affected the Black population. The variation in educational attainment of adult children who are likely to have completed their schooling before the Soweto riots (i.e., children aged 50 years or older at the date of survey) can be treated as exogenous because differences in preferences for schooling or unobserved ability are unlikely to have been binding for this cohort since the very purpose of policies like "Bantu Schooling" was to artificially limit opportunities. Therefore, one source of exogenous variation in educational opportunity comes from restricting attention to this group of adult children. Thus, the gradients in educational opportunity reported in Panel B of Table 8 can be given a causal interpretation. In particular, we notice a monotonic relationship between increasing the level of opportunity and the shape of the occupational distribution in steady state (the long-run probability of being in the bottom half of the occupation distribution is 79 percent for children with high educational opportunities compared to 85 percent for children with low educational opportunities).

The long-run treatment effect of better educational opportunities on climbing the occupational ladder is 6 percent. Of course, pure chance events could shift the distribution over time. One way of partialling out this effect is look at the difference-in-difference treatment effect. Interestingly, this does not change the result (31% – 25% = 6%). This finding is of course completely consistent: chance *cannot* be expected to play a strong role when ceilings are imposed on both educational and occupational attainment, as is generally understood to be the case for this old cohort of Black Africans. For example, the reservation of certain categories of employment for low-skilled Whites was still very much in force throughout the 1970s.

The dismantling of Apartheid is another potential source of exogenous variation. As the descriptive statistics indicated (Table 1), there are sharp differences in average attainment between the younger and older cohorts in our sample. Younger African children have almost triple the amount of schooling than older African children, and younger African parents have more than quadruple the amount of schooling than their older counterparts. The increases in attainment for Africans were sharp enough so that restricting attention only to the population of young people schooled after Apartheid (1994–2008) would be a plausible way of extracting exogenous variation in attainment. However, there are two potential problems with this strategy. First, with increased opportunities to acquire more and better quality schooling, choice arguably plays a greater role and so it is likely that persistence in educational status could be correlated with preferences, more so for this very young cohort than for the older cohort schooled before 1976. Second, even if preferences played no role, many of the individuals falling into this cohort would too young to have ever worked before.

Our solution to this problem is to include children that would have benefited from the shifts in educational spending that took place in the 1980s. The increased public spending on Black schools that began in the early 1980s in part has its roots in the establishment of the tri-cameral parliament which saw the establishment of partial suffrage to Coloreds and Indians and the devolution of powers to the so-called self-governing territories under Black control. By reaching back into these final years of the Apartheid period, we are able to exploit the variation in educational outcomes that came about during this period as a consequence of these incremental political changes. Therefore, in addition to presenting results for the older pre-1976 cohort, we also present results for a younger cohort of children aged 20–35. This restriction makes it possible for a member of this cohort to obtain at least half their schooling (6 years) after 1983.

Panel A of Table 8 shows that like in the case of the older cohort, the probability of being in the bottom of the occupation distribution in the long run decreases with better educational opportunities. However we interpret these effects more cautiously: for this younger cohort the probability of persistence is likely to be endogenous since the greater educational opportunities that started to become available during the 1980s probably would have allowed for preferences to be correlated with educational outcomes in a much more binding way than it would have for the older cohort. Since it is possible that persistence might be endogenous for this younger cohort, we employ an alternative identification strategy (detailed

TABLE 9

Logit Estimates of the Determinants of Intergenerational Persistence in Educational Attainment

Variable	Coefficient	(Std. Err.)
Age	0.025**	(0.007)
Age squared	0.000**	(0.000)
Male	–0.153**	(0.055)
Colored	–0.284**	(0.082)
Rural	0.644**	(0.078)
Number of children younger than 6 years old	–0.089	(0.066)
Studied mathematics in grade 12	–0.118	(0.096)
Completed some form of tertiary studies	–1.029**	(0.118)
Number of household residents	0.011	(0.008)
Intercept	–2.954**	(0.164)
N	11,919	
Log-likelihood	–5,073.338	
$\chi^2_{(9)}$	2,003.623	

Significance levels: †10%, *5%, **1%.

in the next subsection) as a final robustness check to corroborate these causal inferences.

4.2.2. Alternative Identification Strategy

As an alternative identification strategy, we employ the use of a non-parametric kernel matching methodology. We begin by constructing a binary indicator of whether the child obtained the same level of schooling as the mother. We then run a logit regression (Table 9) using this indicator as the dependent variable and compute the predicted probability of persisting in the same educational level as the parents. We check that this predicted probability adequately balances the data.[10]

The fact that conditioning on the propensity score balances the data lends support to the assumption that treatment status is conditionally mean independent.[11]

[10]The results of these balance tests are not reported for reasons of space but are available upon request. In brief however, over 11 strata of the propensity score, we fail to reject the null hypothesis that the difference in means for each covariate between those that persist with the same education as their parents is the same as those that do not persist. The only covariate to not balance is the household size variable and this is only true for the first stratum. Looking at the standard error for this variable in Table 9, note that it is not individually significant. We are therefore satisfied that the reported specification adequately balances the data.

[11]To be clear, the results of the balance test are *suggestive* that treatment status will be mean independent once we condition on the propensity score. There is no guarantee that this is the case. Indeed, this would be true of any study that uses a matching methodology. Hence, we are ultimately forced to assume conditional mean independence. However we have a reasonable level of confidence in this assumption: the dependent variable in Table 9 is whether or not the child persists in the same educational level as the parent so that both advantage and disadvantage is preserved. Under *Apartheid*, much of this had to do with skin color. However as labor markets became more meritocratic, we would expect other factors to have started to play a role: gender, age, location, ability, and household demographics. The regressors contained in Table 7 represent proxies for each of these factors we hypothesize to be important and indeed many of the variables turn out to be significant. Therefore, while it might be true that the model suffers from omitted variable bias, this only presents a problem if indeed the missing variable is *uncorrelated* with all of the included regressors. This possibility seems quite remote given the dependent variable in question.

TABLE 10

Average Treatment Effect of Educational Persistence on Probability of Occupational Persistence

Kernel	Kernel Function $K(s)$	Optimal Bandwidth	ATT Matched	ATT Unmatched	t-ratio Analytical	t-ratio Bstrap
Gaussian	$(2\pi)^{-1/2}\exp(-s^2/2)$	0.300	0.171	0.168	5.29	3.79
Gaussian	$(2\pi)^{-1/2}\exp(-s^2/2)$	0.022	0.121	0.168	3.54	3.75
Epanechnikov	$\frac{3}{4}(1-s^2)\cdot\mathbf{1}(\lvert s\rvert<1)$	0.049	0.120	0.168	3.52	3.43
Quartic	$\frac{15}{16}(1-s^2)^2\cdot\mathbf{1}(\lvert s\rvert<1)$	0.058	0.120	0.168	3.53	3.69
Rectangular	$\frac{1}{2}(\lvert s\rvert<\mathbf{1})$	0.019	0.119	0.168	3.41	3.17
Tricube	$\frac{70}{80}(1-s^3)^3\cdot\mathbf{1}(\lvert s\rvert<1)$	0.500	0.157	0.168	4.83	3.24

Notes: The first ATT estimate using the Gaussian kernel uses a fixed global bandwidth of 0.30. The tricube kernel is not defined for an optimal bandwidth so we set the bandwidth level in that case to be 0.5. The remaining kernel estimators use an optimal bandwidth calculated according to Silverman's (1986) plug-in formula. Bootstrapped standard errors are over 250 replications. Matching occurs over the common support of the estimated propensity score.

We then go on to estimate the effect of persistence in education on persistence in occupation (defined analogously). These results are reported in Table 10. As the table shows, children that persist in the same educational level have a 12.1 percent higher probability of persisting in the same occupations. Moreover this treatment effect, which is obtained by matching on the predicted probability of persistence, does not change substantially when we estimate a treatment effect that is not based on matching by the predicted probability, but rather by taking the difference in average occupational persistence between the group with educational persistence and the group without education persistence. In particular, the unmatched treatment effect has the same sign as the matched treatment effect and has a magnitude of 16.8 percent. Table 10 also shows that the treatment effects are invariant to the choice of kernel function used and in all cases we use an optimal bandwidth specification based on Silverman's plug in rule. Notice also that the optimal bandwidths are very localized (meaning only individuals with very similar predicted probabilities of educational persistence are considered valid controls). Using such small bandwidth choices of course could involve a bias-efficiency trade-off. However note that the treatment effects are very significant across the different kernel and bandwidth choices. Moreover, when we fix the bandwidth choice to 0.3 and use a straightforward Gaussian kernel, the matched and unmatched treatment effects are virtually identical (17.1 versus 16.8 percent, respectively). So while we do not dismiss the possibility of bias, these experiments suggest that the level of bias is probably quite small and does not affect the qualitative finding that persistence in educational attainment causes persistence in occupational attainment.

The similarity between the matched and unmatched treatment effects tends to lend support to the exogeneity assumption implicit in the other specifications reported in Table 7 (i.e., those that use the observed indicator of educational opportunity). Nevertheless, we also report a final specification of the ordinal logit model of occupations using as our measure of education persistence the predicted probability of persistence. As the last two columns of Table 7 indicate, the coefficient on predicted educational persistence is negative, as we would expect. Using these last two specifications, we then recompute the steady-state distribution of occupations for individuals that persist with the same level of education as their parents. These new results appear as two new rows of Table 8 (one in each panel) under the label "Mover–Stayer Markov Model (endogenous persistence)."

The final step in using this alternative identification strategy is to compute the treatment effect on the long-run occupation distribution. Again we take the difference in the predicted probability of being in the bottom half of the distribution under the Markov model as our benchmark and compare it against the same quantity for the mover–stayer model under the case of "endogenous persistence." Notice that the probability of being in the bottom half of the distribution for the older cohort for this case is 0.84, whereas the probability under the Basic Markov Model is 0.54. Thus, the difference in the difference is 0.30 (which is the number reported in the last row and last column of Table 8). Therefore for the older cohort (for which we think the case for causal inference is strongest) the Basic Markov Model substantially under-predicts the extent of inequality in the long run.

Notice also that the results of our matching methodology provide an indirect test of the basic identification strategy employed for this older cohort (i.e., restricting the sample to adult children who completed their schooling prior to the Soweto riots of 1976). If indeed educational opportunities were exogenously driven before 1976, we should expect the difference in difference estimate on the probability of being in the bottom of the distribution to be quite similar between the two identification strategies. This indeed is the case: 0.30 compared to 0.31 (i.e., comparing the last two rows of Table 8).

Finally, we can compute one other treatment effect of interest: the difference in the difference reported in the final column between the older and younger cohorts, for the case of endogenous persistence is estimated to be 0.30 – 0.2 = 0.1. At first glance it is not immediately obvious that this comparison identifies a treatment effect of educational opportunity, since by definition we are holding this constant (i.e., the comparison is between a younger and an older cohort who persist with the same education as their parents). However, from Table 1 it is clear that persistence for the younger generation means something qualitatively different to the case for the older cohort as the average attainment for parents quadruples from the older to the younger cohorts. This happens while the education distribution becomes more equal for the parents, while becoming much more unequal for the children (tripling in the mean at the same time). Then looking at rows 4 and 5 of Table 8, we also notice that the big shift in the occupation distributions when we control for endogeneity of education persistence is a 27 percent reduction in the long-run probability of being in job category 1, which is redistributed into the three higher occupational levels (an increase of 14 percent going into level 2, 11 percent into level 3, and 3 percent into level 4). Levels 3 and 4 are defined as Technicians/Associate Professionals and Professionals respectively so it is likely that the individuals predicted to get into these occupations in the long run will more likely be better qualified. So while it is true that some of the 10 percent treatment effect relates to changes to pure-labor market structure brought about as Apartheid ended (e.g., the abolition of the type of job-reservation policies talked about earlier), it is also true that probably many of the individuals more likely to benefit from these new occupations in the long run will have been those whose parents were better educated.

5. Conclusion

The main goal of the paper was to estimate the long-run distribution of occupations in South Africa. The standard approach in this type of research program is to model the distribution in steady-state, by assuming that the matrix of transition probabilities is homogenous across members of the population. However, since the probabilities that make up the elements of such a matrix also reflect other things about the transition process, one cannot usually ask causal questions about the mechanisms through which a correlation in the occupational outcomes of parents and their children might come to exist. One such causal connection, common in models of poverty traps where occupational structure is the main driver of inequality, is that of educational opportunity. The key contribution of this paper is on this issue.

Empirical evidence of the connection between educational opportunity and occupational structure is quite rare, because the educational outcomes of an individual are likely to be *determined* by the parents' choices of occupations. We make several contributions in this regard. First, we explicitly model the probability that a child occupies the same or a different rung on the occupational ladder as her parents. Second, in modeling this probability, we control for both the educational attainment of the child, as well as the level of educational opportunity of the child, measured by whether the child attained more, less, or the same amount of schooling than her parents. Third, these conditional probabilities then constitute separate transition matrices by level of educational opportunity, race, and gender. These matrices are then used to the compute the steady-state distribution under the more plausible assumption that different race–gender pairings facing different levels of educational opportunity result in different long-run occupation distributions. Our final contribution is to argue that political events in the history of the struggle to end Apartheid provide the raw material for an identification strategy that would permit a causal interpretation of the role of educational opportunity.

We find evidence that educational opportunity has a strong conditioning effect on the distribution of occupations in steady state. While the paper presented several types of treatment effects that could be of interest, our main finding is that African female children who inherit the same level of educational opportunity as their parents are 6 percent more likely to be in the bottom of the occupation distribution than if they were exposed to better educational opportunities. An alternative identification strategy suggests that this figure is approximately 10 percent for younger cohorts of African female children.

A final point worth highlighting concerns identification. Since we restricted our analysis to one wave of the NIDS data, it was imperative to achieve identification through the methods discussed (i.e., the timing of political events and propensity score matching). Of course one could take the alternative route by assuming fixed preferences for schooling and then exploiting the panel structure of the data to achieve identification. While this approach would be a fruitful avenue for future research, we do not pursue it in the present paper as it introduces a non-trivial complication of our analytical approach because of the problem of non-random attrition.

References

Atkinson, A. and F. Bourguignon, "Introduction: Income Distribution and Economics," in A. Atkinson and F. Bourguignon (eds), *Handbook of Income Distribution*, vol. 1 of Handbooks in Economics, Elsevier Science, Amsterdam, 1–58, 2000.

Banerjee, A. V. and A. F. Newman, "Occupational Choice and the Process of Development," *Journal of Political Economy*, 101(2), 274–98, 1993.

Bergman, M. and D. Joye, *Comparing Social Stratification Schemas: CAMSIS, CSP-CH, Goldthorpe, ISCO-88, Treiman and Wright*, SIDOS, Neuchatel, 2001.

Blumen, L., M. Kogan, and P. J. McCarthy, *The Industrial Mobility of Labor as a Probability Process*, Cornell University Press, Ithaca, NY, 1955.

Feller, W., *An Introduction to Probability Theory and Its Applications*, vol. 1, Wiley, New York, 1950.

Fiske, E. and H. Ladd, *Elusive Equity: Education Reform in Post-Apartheid South Africa*, Brookings Institution, Washington, DC, 2004.

Galor, O. and J. Zeira, "Income Distribution and Macroeconomics," *Review of Economic Studies*, 60(1), 35–52, 1993.

Girdwood, S. and M. Leibbrandt, "Intergenerational Mobility," National Income Dynamics Discussion Paper, Number 10, SALDRU, 2009.

Goodman, L. A., "Statistical Methods for the Mover-Stayer Model," *Journal of the American Statistical Association*, 56(296), 841–68, 1961.

Karlin, S. and H. M. Taylor, *A First Course in Stochastic Processes*, Academic Press, Boston, MA, 1975.

Leibbrandt, M., I. Woolard, A. Finn, and J. Argent, "Trends in South African Income Distribution and Poverty Since the Fall of Apartheid," OECD Social, Employment and Migration Working Papers No. 101, 2010.

Louw, M., S. Van der berg, and D. Yu, "Convergence of a Kind: Educational Attainment and Intergenerational Social Mobility in South Africa," *South African Journal of Economics*, 75(3), 548–71, 2007.

Silverman, B., *Density Estimation for Statistics and Data Analysis*, Chapman and Hall, London, 1986.

Thomas, D., "Education Across Generations in South Africa," *American Economic Review*, 6(2), 330–4, 1996.

Wilson, F. and M. Ramphele, *Uprooting Poverty: The South African Challenge*, David Philip, Cape Town, 1989.

Woolard, I., M. Leibbrandt, and L. de Villiers, "The South African National Income Dynamics Study: Design and Methodological Issues," *Studies in Economics and Econometrics*, 34(3), 7–24, 2010.

Ziervogel, C. and O. Crankshaw, "Inter-Generational Occupational Mobility Amongst Blacks in the Mitchells Plain Magisterial District, Cape Town: Evidence from the Khayelitsha/Mitchells Plain Survey," *Urban Forum*, 20, 235–51, 2009.