Web Appendix: Intergenerational Networks, Unemployment, and Persistent Inequality in South Africa

1 Intergenerational correlations in Industries

Table 1 reports coefficient estimates from systems of seemingly unrelated regressions, where each dependent variable is a dummy for working at baseline in a one-digit industry and the right hand side is a dummy for the father or mother being in that industry and a dummy for the parent working, in addition to race, age, and education fixed effects[1]. So as to not multiple count individuals, I use the fraction of working years spent in an industry or occupation for those who work in multiple industries or occupations. Sons' industries appear highly correlated with their fathers' while daughters' industries are highly correlated with their mothers'. While father-daughter and mother-son correlations are occasionally significant for an individual industry, they are always small in magnitude and are jointly insignificant.

A natural extension is to consider occupational correlations as well as industries. Table 2 reports coefficient estimates for occupations from seemingly unrelated regressions, following the methodology used for industries. The difference between the occupation and industry correlations are striking. Industries are correlated along gender lines, with most industries being associated with an 7-12% increase in likelihood of employment in that industry for children if their parents work in that industry and happen to be the same gender. Occupational correlations also are largest within mothers-daughter pairs, and look quite similar across the other three groups. Once again, they are often individually significant for a given occupation, but a joint test reveals that parent's occupations are not correlated with children's across the sample in any of the 4 parent-child pairs.

[1] Results are qualitatively unaffected by omitting these demographic controls.

2 Attrition

To test if attrition could be responsible for my results, I adopt two approaches. The first is a test similar to Becketti et al (1988). In this test, I use baseline data to test whether relationships between the explanatory variables of interest and the dependent variable are similar in the attrition and non-attrition sample. More specifically, we can regress

$$W_{i2002} = \beta X_{i2002} + \varepsilon_{i2002}$$

for two different samples: the full estimation sample and the sample that never attrits. If coefficients are statistically different between these two samples, then we need be concerned that the relationship between independent variables and working are different in the attriting sample, suggesting that bias is present[2]. This analysis is presented in the first two columns of table 3. Point estimates in the full sample are very similar to those in the full panel sample, as is precision. Moreover, we cannot reject that the coefficients on any of the employment trends are different from each other, nor can we jointly reject this difference, suggesting that, up to the power of this test, we have no evidence that the relationship between employment trends and working is different for attritors and non-attritors. However, concerns may remain that attrition bias is present. Therefore I make a variety of extreme assumptions about the attritors' behavior so as to include them in my estimation sample in a bounding exercise. In column 3 of Table 3, I assume that they are all working. In column 4, I assume that none are working, while in column 5, those whose fathers live in other provinces are assumed to work and those whose fathers living nearby are assumed to not work. Column 6 assumes the opposite. Even under these extreme assumptions, the pattern of coefficients remains the same throughout, and the network effects largely retain significance. Since attritors work less at baseline than non-attritors, we may think that column 2 represents the most likely case, where none of them are working. In this case, all coefficients remain similar to the baseline.

[2] One caveat is necessary: the baseline analysis used individual fixed effects to demean log employment, so that the deviations of employment represented the offer rate in that year in that industry. In this analysis, fixed effects are impossible; however, we can still demean employment in a given industry to give an estimate of whether the employment is hiring or firing at the time, that is, how large employment is relative to it's time average. This is algebraically equivalent to the levels used in the analysis with fixed effects (as the deviations are identical to the levels up to a person-specific constant), and as such could be used interchangably in the other specifications in this paper. Therefore, I use these industry employment deviations to perform this test.

3 Correlated Industries & Neighborhood Effects

Table 4 examines whether employment of young men is sensitive only to the employment trend in fathers' industries or rather to industries of a variety of network members. First, in column 2, I look at trends in the log employment in the industry of the male head of household. In households where the reported head was not an employed male, I define the male head as the spouse of the head if he is male and working, or the first employed male on the household roster at baseline who is too old to be considered in this study of young adults (older than 22 at baseline). The strong network effect remains on fathers who live in the province, and the point estimate remains similar and significant. In contrast, head of household effects are negative and hover around marginal significance, much as the father's employment trend on daughters and sons of absent fathers does. This provides further support to the wealth effect hypothesis – it appears that young men get wealth support from male heads of household who are not their fathers, but do not benefit from network connections.

Column 3 of table 4 considers neighborhood effects. In column 3, I condition on the fraction of black and coloured males in other households in the same neighborhood (measured as the enumeration area, which are approximately three block squares) who are working as well as the log employment in the modal industry among adult males in the neighborhood. This neighborhood youth employment appears very important for predicting young males' employment status, suggesting that neighborhood trends are quite important, and that this employment rate is capturing some signal about neighborhood effects. Of course, we can not firmly attribute this effect to neighborhood networks since this neighborhood employment rate is likely correlated with many other neighborhood trends. Including this neighborhood effect does not statistically impact the coefficient on fathers or its significance, as neighborhood employment rates turn are uncorrelated with trends in the fathers' industries. The modal industry variable appears insignificant, suggesting that the father is more than a random neighborhood man.

Columns 4 and 5 of table 4 utilize information on how a job was received. Looking at network jobs versus self-found jobs, we find that the trends in fathers employment remain strong for network jobs and absent for self-reported jobs in the presence of the correlated networks controls. Noteworthy to this paper, the effects of neighborhood employment trends seem to primarily be felt in self-sought jobs, suggesting that the self-sought job variable does carry a signal, and that the "catch-all" variable of neighborhood employment trends cannot be thought of as simply larger-scale networks.

4 Employment Data Construction

Provincial employment data are constructed by the author using two-digit industries from the September Labour Force Surveys in 2002-2004 using the sampling weights calculated by Statistics South Africa. These are nationally representative samples of a rotating panel survey, with each surveying 67836 to 73797 adults aged 15 and older. To alleviate sample size concerns, occupation-industry employment numbers are constructed by estimating the fraction of a two-digit industry in each occupation in each year using national data and multiplying the overall employment numbers in that industry by that fraction. The employment data are quite noisy at the two digit levels, especially for relatively small industries. Because the analysis will focus on log changes, the possibility of a small industry not being found in one period due to sampling error could create large outliers. As a result, I observe that my naive employment estimates for industry i at time t, $\widehat{Emp_{it}} = w_{it}\left(Emp_{it} + u_{it}\right)$, where u_{it} is sampling error, w_{it} represent sampling weights, and Emp_{it} represents the number of employed individuals observed in the national survey. Fortunately, we know the distribution of this sampling error: if share p_i of the population works in industry i and I draw n people randomly, then each sampled individual faces a multinomial distribution with probability p_i of being in industry i. Employment is a sum of binomials, which means that $u_{it} \sim N\left(0, p_i\left(1-p_i\right)n\right)$. However, employment estimates face a lower bound of zero, and hence in expectation I underestimate true employment for small industries as the sampling error is truncated. As a result, I correct my estimates so that u_{it} is mean zero, by estimating the mean sampling bias in small industries, using mean sampling weights for each year and average fraction of the population found employed in industry i from 2001-2004 as an estimate of p_i. The analysis presented is not sensitive to small changes in p_i estimates.

However, the presence of sampling error suggests another concern. Because there is some error in my estimates of employment, the coefficient may be biased downward due to attenuation bias, and the standard errors may be biased downwards as well (e.g. Murphy and Topel 1985). Indeed, this problem is present in any study which uses macro statistics constructed from household surveys. Fortunately, the variance of this sampling error is very small: using the delta method derives the distribution for the error in log employment to be $w_{it}^2 n p_{it}\left(1-p_{it}\right)/Emp_{it}^2$, which is small whenever employment in the industry is larger than the industry-specific mean sampling weight. As argued above, this error term is not normal in finite samples due to truncation for small industries, however, we can approximate it by assuming normality and estimating p_{it} from my employment estimates. This allows a Murphy-Topel (1985) correction term to the standard errors as a robustness check. In an unclustered equivalent of the baseline estimation, the largest standard error increases by 2.5 percent in magnitude, and no conventional significance thresholds are crossed, suggesting that this is not a major cause for concern. As a result, preference is given to reporting results with clustered standard errors.

Comparisons between 1993 and 2003 are conducted using the September 2003 Labour Force Survey and 1993 October Household Survey, another nationally representative survey which had 91,494 adults fifteen years and older.

Table 1: Industry Correlations

Child's Industry	Father's Industry Sons	Father's Industry Daughters	Mother's Industry Sons	Mother's Industry Daughters
Informal Services	0.035	0.011	0.002	0.075***
	(0.050)	(0.028)	(0.011)	(0.017)
Agriculture	0.126***	0.045**	0.022	0.140***
	(0.021)	(0.020)	(0.034)	(0.036)
Manufacturing	0.093***	0.01	0.029*	0.079***
	(0.019)	(0.011)	(0.016)	(0.017)
Construction	0.086***	0.022**		
	(0.018)	(0.009)		
Retail	0.092***	0.02	0.032*	0.077***
	(0.020)	(0.012)	(0.017)	(0.018)
Transportation	0.053***	0.014	0.015	0.049
	(0.020)	(0.012)	(0.058)	(0.031)
Business	0.079***	0.025*	0.053*	0.086***
	(0.025)	(0.015)	(0.028)	(0.024)
Services	0.091***	0.02	0.028*	0.071***
	(0.019)	(0.011)	(0.015)	(0.017)
F-test: Parent Effect	54.13	11.2	8.22	30.28
(p-value)	0.00	0.19	0.31	0.00
F-test: Excl. Agriculture	37.1	8.4	8.18	24.13
(p-value)	0.00	0.30	0.23	0.00
N	687	649	687	649

Notes

1 Specifications are seemingly unrelated regressions where the dependent variable is a dummy for the child working in a one-digit industry and the independent is a dummy for the parent working in that industry

2 Includes controls parent working, and fixed effects for each age, education level and racial group

3 Mining and Utility provision are excluded due to sample size

Table 2: Occupational Correlations

	Father's Occupation		Mother's Occupation	
Child's Occupation	Sons	Daughters	Sons	Daughters
Senior Officers & Managers	0.027**	0.019	0.021	0.034**
	(0.013)	(0.012)	(0.015)	(0.015)
Professionals	0.026*	0.022*	0.025*	0.045**
	(0.015)	(0.013)	(0.014)	(0.014)
Technicians	0.026*	0.022	0.023	0.036**
	(0.016)	(0.014)	(0.015)	(0.015)
Clerks	0.037**	0.022	0.024*	0.040***
	(0.018)	(0.015)	(0.015)	(0.015)
Service and Sales Workers	0.029*	0.021*	0.025*	0.040***
	(0.015)	(0.013)	(0.014)	(0.015)
Agr. and Fisheries Workers	0.030*	0.013	0.021	0.02
	(0.017)	(0.019)	(0.026)	(0.038)
Craft and Trades Workers	0.029**	0.021*	0.025*	0.039***
	(0.013)	(0.012)	(0.015)	(0.015)
Plant and Machine Operators	0.026*	0.02	0.026*	0.039***
	(0.014)	(0.012)	(0.014)	(0.014)
Elementary Occupations	0.029**	0.022*	0.026*	0.042***
	(0.014)	(0.012)	(0.014)	(0.014)
F-test: Parent Effect	5.85	3.56	3.77	11.02
(p-value)	0.75	0.94	0.93	0.28
N	829	819	829	819

Notes

1 Specifications are seemingly unrelated regressions where the dependent variable is a dummy for the child working in a one digit occupation and the independent is a dummy for the parent working in that occupation

2 Controls for the parent working and fixed effects for each age and education level and racial group.

Table 3: Attrition Robustness

	(1)	(2)	(3)	(4)	(5)	(6)
Log Employment,	0.052	0.083	0.075	-0.017	0.107	-0.049
Father's Industry	(0.196)	(0.259)	(0.108)	(0.044)	(0.112)	(0.048)
Male*	-0.405	-0.66	-0.148	-0.13	-0.125	-0.152
Father's Industry Employment	(0.318)	(0.402)	(0.183)	(0.105)	(0.187)	(0.118)
Father in Province*	-0.158	-0.151	-0.207*	-0.044	-0.222*	-0.029
Father's Industry Employment	(0.204)	(0.266)	(0.113)	(0.053)	(0.116)	(0.058)
Male*Father in Province*	0.611*	0.802*	0.361*	0.290**	0.293	0.357***
Father's Industry Employment	(0.328)	(0.411)	(0.190)	(0.114)	(0.193)	(0.128)
F-stat: Joint Male Test	4.16	2.76	9.95	7.47	7.35	9.96
(p-value)	(0.02)	(0.06)	(0.00)	(0.00)	(0.00)	(0.00)
F-stat: Joint Female Test	1.75	0.68	8.17	2.27	7.97	3.35
(p-value)	(0.17)	(0.51)	(0.00)	-(0.10)	(0.00)	-(0.04)
Observations	4097	3224	12309	12309	12309	12309
Sample	Year 2002	Non-Attrit	Full	Full	Full	Full
Number of individuals	4097	3224	4103	4103	4103	4103
R-squared	0.2	0.22	0.16	0.03	0.08	0.08

Notes

1 Presents OLS estimates. The dependent variable in all regressions is an indicator for working.

2 Columns (1) and (2) include only 2002 and condition on dummies for paternal presence, gender, the interaction, and racial group as well as age and year fixed effects.

3 Employment in columns (1) and (2) is measured as the deviation of employment in that industry in 2002 from it's 4 year average.

4 In Column 3, all attritors are assumed to be working, while in column 4, none are. Column (5) presumes that attritors whose fathers are absent are working while those with present fathers are not, while column 6 assumes the opposite.

5 Standard errors are clustered at the household level.

6 Columns 3-6 are conditional on age, year, and individual fixed effects, and the father's industry is fixed to be his baseline industry.

Table 4: Correlated Networks

	(1)	(2)	(3)	(4)	(5)
Job Search Method	Any	Any	Any	Network	No Help
Log Employment,	-0.184*	-0.175	-0.117	-0.197	0.066
Father's Industry	(0.111)	(0.111)	(0.104)	(0.135)	(0.086)
Father in Prov *	0.280**	0.312***	0.251**	0.331**	-0.069
Father's Industry Employment	(0.117)	(0.120)	(0.113)	(0.143)	(0.088)
Log Employment,		-0.068	-0.066	-0.119**	0.042
Male Head Industry		(0.049)	(0.048)	(0.051)	(0.031)
Local Young Male			0.138***	0.027	0.103***
Employment Rate			(0.042)	(0.041)	(0.030)
Log Employment,			0.016	0.046	-0.017
Neighborhood Modal Industry			(0.028)	(0.029)	(0.019)
F-test: Fathers	4.84	5.42	4.64	5.21	0.31
(p-value)	(0.008)	(0.005)	(0.010)	(0.006)	(0.734)
F-test: Others			4.48	2.79	4.55
(p-value)			(0.004)	(0.040)	(0.004)
Observations	4977	4977	4908	4850	4850
Number of individuals	1830	1830	1810	1810	1810
R-squared	0.06	0.06	0.07	0.03	0.06

Notes

1 Presents OLS estimates. In Columns (1)-(3) and (6), an Indicator for working is the dependent variable. In column (4), the dependent variable is an indicator of both working in a job found with network help, and in column 5 it is an indicator for working and having gotten the job by yourself.

2 The Male Head is the male head of household or the first working male on the household roster. The neighborhood youth employment rate is the fraction of males in other households in the same sampling cluster as the respondent who are working, while the modal industry is the modal industry among adult males in that sampling cluster.

3 Standard Errors are clustered at the household level.

4 All regressions are conditional on age, year, and individual fixed effects, and the father's industry is fixed to be his baseline industry.