Internal Migration and Labour Market Outcomes in South Africa

Julia Garlick[*] Murray Leibbrandt[†] James Levinsohn[‡]

June 16, 2019

Abstract

It is well established, internationally and in South Africa that internal migration is a key livelihood strategy to get ahead for individuals and their households. This literature also highlights the importance of labour market opportunities in both the sending and receiving areas as one of the central planks of this broader point. We seek to make a contribution to this literature through an analysis of the South African case. We use five waves of panel data containing detailed individual and household information to establish facts on the extent, nature, and timing of migration by working age adults in South Africa and to describe the immediate and medium run changes in labour market status and earnings attendant on this migration. We then use synthetic control methods and difference-in-difference regressions to estimate the individual returns to migration. Our estimation work strongly supports the finding that migrants experience better labour market outcomes than comparable (on observables) people who do not migrate. We find that the income gains from migration are a one-time effect, at least in normal economic conditions. Migrants move onto a higher income path, but do not continue to climb, instead returning to the same (or lower) rate-of-change as non-migrants. For those who migrate during generally adverse economic conditions there is a very different pattern. In this case, it appears that there is no immediate income gain to migration, but in later intervals, they experience very large and significant returns. For them, migration is a longer-run investment. In general, migrants and non-migrants seem to be on the same income path prior to migration, but migrants resume this path, at a higher level, post migration. Unemployed adults who migrate

[*]University of Iowa
[†]University of Cape Town
[‡]Yale University

are significantly more likely to be employed than their peers who do not migrate in every time interval.

1 Introduction

This paper examines migration patterns and long-run returns to internal migration in South Africa. We use almost a decade of panel data from South Africa to analyse migration patterns and labour market returns from migration. With up to four waves of nationally representative post-migration panel data, this dataset gives us the unprecedented opportunity to study shorter-run and longer-run outcomes from migration in an African country.

The literature on measuring labour market returns to migration in Africa is sparse given the prominent role assigned to labour migration in seminal models of economic development (e.g. Lewis (1954), Harris and Todaro (1970), and more recently Young (2013).) The empirical side of internal labour migration research is better represented in South Asia. Recent studies, each focused on different aspects of temporary migration, include Morten (2013) (India) and Bryan, Chowdhury and Mobarak (2014) (Bangladesh.) While we know of no nationally representative studies examining the impact of internal migration on incomes in any African country, Beegle, Weerdt and Dercon (2011) examine the issue with a baseline sample of 912 households from the Kagera region of Tanzania.

In the South African context, research has focused on correlates of migration and endogenous household formation. For example, Hamoudi and Thomas (2014) examine the role the State Old Age Pension plays in household formation while Ardington, Case and Hosegood (2009) analyze the role that cash transfers play in enabling migration by essentially staking the migrant. Neither of these papers estimates the return to migration for any party involved.[1] These, and other papers, do establish the fluid nature of South African households, and raise the possibility of migration patterns outside those predicted by traditional models. A good example is the recent study by Hall and Posel (2018) which summarises this broad literature and its findings, and points out that one important consequence is that children have remained less urbanized than adults, and have grown up without co-resident parents.

This paper contributes to three literatures. First, we add to the sparse literature on labour market migration in African countries. Second, we provide evidence on the nature of migration: one of our descriptive results is that many adults migrate

[1] Prior empirical studies of migration in South Africa have, by necessity, relied on repeated cross-sectional data. See for example Posel and Casale (2006), Posel, Fairburn and Lund (2006), and Budlender and Lund (2011). Hall and Posel (2018) use the National Income Dynamics Study, mostly to profile migration. But they make good use of the data as a panel by tracking children and their mothers over time.

repeatedly over a nine year period, and, as far as we can determine, this is non-circular migration. This suggests that migration may reflect searching for new draws for employment in different markets, rather than a one-time re-allocation from rural to urban labour markets, or short-term, temporary migration. Third, we extend the literature on the returns to migration by looking at the short and longer-run labour market outcomes that result from on these complex migration patterns. With multiple post-migration waves, we can examine how income responds over time to migration, and assess whether migration causes a one-time level shift in income, or moves the migrant to a new, steeper income path. To do this, we use both difference-in-difference estimators and synthetic control methods.

2 Data

We use data collected in the first five waves of South Africa's National Income Dynamics Study (NIDS). These waves were collected in 2008, 2010, 2012, 2014-2015, and 2017. They comprise a nationally representative sample, when combined with supplied weights. The panel follows individuals as well as households. In the first wave, all household members (or their proxy) were surveyed, and are classified as continuing sample members (CSMs). In later waves, all CSMs are tracked and re-interviewed if possible, along with all co-residents in their current household. Co-residents who were not interviewed in 2008 are designated as temporary sample members (TSMs).[2] This makes NIDS an ideal dataset with which to study migration, since individuals are tracked even if they have left the original sampled household. If a CSM joins a new household, detailed information on their new co-residents and location is collected.

Table 1 gives the sample size for each wave by race.[3] In Wave 1, there were 28,226 individuals in the sample, 22,206 of whom were Black. These are the data to which sampling weights are later applied. The sample size increases in each wave due to the addition of new CSMs and the inclusion of new TSMs. CSMs leave the sample in three ways: death; international migration; or non-response, either because they could not be found, or because they refused to take part in the survey that year.[4] TSMs exit the sample when they are no longer co-resident with a CSM. By Wave 5, the sample includes 59,652 individuals, 47,746 of whom were Black.

Tables 2 and 3 provide more information on migrant demographics. Each column

[2]New biological or adopted children of CSMs become CSMs, but all other new household members are TSMs.

[3]The South African population is divided into several officially recognized racial groups, following the categories formalized by the Apartheid government. In official parlance, "Coloured" people are members of a long-standing and culturally distinct mixed-race population.

[4]The modal reason for missing CSMs is refusal, not sample exit or failure to find them.

represents migrants between two waves, and the final column defines migrants as those who have moved at any time between 2008 and 2017.

Looking at Table 2, we see that migration rates varied between 14% and 25% over the nine year period, and were similar for both genders. Individuals with a matric certificate are more likely to move than individuals who have not completed high school, which is in line with migration patterns in both developed and developing countries. Table 3 shows numbers of migrants (and non-migrants) for each interval, split into six age groups. Migration is most common for younger adults, with the highest rates for people between age 18 and 34. We will focus on migrants of working age, between ages 18 and 65, since these are the most likely labour market participants, and are more likely to be making their own decisions than people under 18.

3 Descriptive Results

Before moving into the main substance of the paper, two additional facts are important to frame our results. Table 4 shows the number of people affected by migration in each interval. This includes migrants themselves, individuals in households that sent a migrant the period before, and individuals in households that receive a migrant the period after. This table shows the strength of these panel data in allowing for such a view and giving a sense of the total impact of migration, in terms of those who may be affected by it. The first panel looks at the numbers of affected and unaffected individuals over the various time periods, while the second panel focuses only on individuals who were present in the first wave. The third panel includes only CSMs. Two things stand out: first, the majority of the sample in all waves is touched by migration; and second, a large share of households both send and receive a migrant. We focus on individual returns to migration, but it is important to be aware that, in a developing country with strong remittance networks, the returns to migration affect many more people than just the migrant.

Table 5 breaks migrants into categories based on the type of area they leave, and the type they enter. The first panel shows that the majority of moves are 'within type', ie. a migrant moves within rural areas, rather than migrating to an urban area. This is in contrast to traditional models of migration in developing countries, which suggest that migration is primarily a rural-to-urban move. The second panel includes only migrants who changed districts, as a proxy for a move of substantial distance. Most districts include both rural and urban areas, so people may change types of area without changing districts. Within this subset of migrants, the modal move is rural-to-urban; however, the majority of migration is still not rural-to-urban.

It is also useful to provide descriptive information on migration outcomes, to contextualise the causal results that will follow in Section 4. Table 6 shows income changes[5] and employment outcomes for adults of working age. The categories "Move" and "DNM" are, respectively, those who migrated between the two periods, and those who did not migrate. Each category is then split further by gender, or by education level. Table 6 provides no causal evidence on the effects of migration. These are simple cross-tabulations. The first panel shows income changes in a given interval. Three things stand out: first, in the first interval, migrants had lower income gains than non-migrants; second, for all other intervals, migrants' income gains were much larger than those for migrants; and third, neither gender nor education level has a systematic pattern of better returns. In the second panel, we examine employment or unemployment for the same categories. In every interval, migrants had a higher probability of employment than their non-migrant counterparts.

Finally, Table 7 presents evidence on repeat migration. This table is similar to a tree or flow diagram. Each node shows the number of people who have a particular migration history. The first panel counts those who moved between 2008 and 2010. The second row of each panel has two cells, splitting the sample again based on whether they moved between 2010 and 2012. Analogously, the third row splits the sample yet again based on whether they moved between 2012 and 2015. The fourth row splits based on whether someone moved between 2015 and 2017. Thus, the cells in the fourth row of each panel represent the number of people who followed a particular migration pattern. The majority of the sample never moves (this is the far-left cell in the fourth row of the second panel). However, 24% of the sample moves more than once, which translates to 50% of migrants moving at least twice. This suggests that migration is not a one-time event for many. This is a more nuanced reallocation of labour than simple rural-to-urban migration, in which migrants may search repeatedly in different markets.

[5] Income changes are calculated as follows. Only labour market income is considered, from all jobs a respondent reports. Many people have zero labour market income, including both those who are unemployed and would like to work, and those who are out of the labour force. We wish to be able to comment on income changes for those who are in the labour force, and who gain or lose employment. This group of people may have zero labour market income in one wave, and positive labour market income in the prior or following wave. Thus, we assign labour market income in a three step process. First, we sum labour market income for all survey respondents. Second, we log-transform this income, which automatically assigns missing values to all zero earners. Third, we identify all respondents who have zero income in a given year, but positive income before or after that, and we manually replace their missing income to zero. Finally, we calculate change in income by taking the difference in income between two survey waves. This means that, for example, someone who is unemployed in 2010 and employed in 2012 shows a positive income change between 2010 and 2012, instead of being missing from the analysis. Someone who is employed in 2008 and unemployed in 2010 will show a negative income change. We believe that this is a more complete reflection of the income changes between two waves, since it does not ignore the income changes among those who change employment status.

4 Results

4.1 Methods

To calculate the returns to migration, we need to infer how a migrant would have fared absent the migration event. The central difficulty we face is the role of selection, and the existence of confounding factors that affect both the probability of migration and the returns to migration. Our first methodology uses the synthetic control approach, to identify or construct comparable non-migrants and migrants, along with differenced data to remove time-invariant individual effects. Our second methodology is standard difference-in-difference regressions, in which time-invariant differences between individuals are removed, and which requires only that the parallel trends assumption hold.[6]

Neither of these approaches identify the same 'return to migration' that an experimental paper might pursue, since only selection on observable factors can be addressed using the synthetic control, or matching, method. We argue that this is not a weakness, however. In some sense, there are two ways to define the return to migration. The first is to think of the effect of migrating on a random population member, or in experimental terminology, the average treatment effect of migration. The second is to consider the effect of migrating on a migrant, or the treatment on the treated. Both of these definitions are of interest. We do not identify the average treatment effect, but matching on observables does identify the treatment on the treated, as shown below.

In formal terms, matching assumes that the distribution of potential incomes of migrants and non-migrants are independent of migration conditional on the set of covariates, X. Let D denote migration status, with $D = 1$ for migrants (migrant-households) and $D = 0$ for non-migrants. Similarly, Y_1 is income after migration and Y_0 is income for non-migrants in the corresponding period. Then the assumption underlying matching is that

$$(Y_1, Y_0) \perp\!\!\!\perp D|X \tag{1}$$

If this is true, then conditional on covariates X, non-migrants have the same income distribution that migrants would have experienced without migration, and migrants have the same income distribution that non-migrants would have experienced had they migrated. Matching estimators can then calculate the return to migration by creating a weighted sample of non-migrants such that the distribution of observable characteristics in each group is the same. However, assuming that the returns to

[6]In the case of migration, the parallel trends assumption is that the labour market outcomes of migrants and non-migrants would have evolved similarly, had the migrant not migrated. We provide some evidence in favour of this assumption in Tables 10 and 11.

migration do not affect the migration decision, even with a large selection of control variables, is probably wrong.

Heckman, Ichimura, Smith and Todd (1998) and Rosenbaum and Ruben (1983) demonstrate that a weaker condition is sufficient for a valid matching estimator, namely

$$E\left(Y_0|P(X), D=1\right) = E\left(Y_0|P(X), D=0\right) \tag{2}$$

where $P(X) = Pr\left(D=1|X\right)$. The use of the index $P(X)$ avoids the dimensionality problem that arises with using a large number of covariates, and only mean-independence of the non-migration income is assumed. This amounts to allowing the returns to migration to differ across migrants and non-migrants, while requiring that the non-migration incomes of each group have the same mean. Individuals can self-select based on their expected post-migration income, provided their incomes without migration do not differ. This is the result that Ham, Li and Reagan (2011) use to justify their matching estimator. Because it does not claim mean equality for Y_1, this estimator cannot be used to measure the average return to migration for the population, or even for a sub-sample of likely migrants. It can only measure the returns for those who migrated, because only Y_0 is assumed equivalent for migrants and non-migrants. It does not speak to the income that non-migrants would experience if they migrated, but only to the income that migrants would experience had they not migrated.

Even in this less restrictive case, matching estimators may still be biased compared to experimental estimators. The extent and sources of this bias were studied in detail by Heckman, Ichimura and Todd (1997) in their evaluation of non-experimental relative to experimental methods using a US job-training program. They identify three contributors: nonoverlapping support between treatment and control populations; different distributions of covariates X within the two populations; and genuine selection bias due to selection on unobservables. In the cases they examine, the larger share of measured bias was due to the first two contributors, not to true selection bias. If matching methods are correctly applied, these first two sources of bias can be eliminated and the remaining bias in measurements, due to selection on non-observables, will be small.[7]

The two additional sources of bias that commonly arise in nonexperimental evaluations are due to geographic mismatch between treatment and control groups, and the use of different survey instruments (Heckman et al. (1997)). For our purposes, the latter is not of concern. Information on both migrants and non-migrants was collected in the same nationally representative survey. We additionally have access to sufficiently detailed geographic information to place migrants and non-migrants into

[7] Performing bounds tests, to address this concern, is on our agenda.

similar (pre-migration) labor markets, which increases the plausibility that Y_0 is truly equivalent for both groups.

For the matching estimators, a comprehensive set of demographic, household, and geographic variables are used to identify similar migrants and non-migrants, using variable values in the period immediately prior to the move. The same set of variables is used as the set of control variables for the difference-in-difference estimators.

4.2 Returns to Migration

4.2.1 Immediate Returns to Migration

The immediate returns to migration are shown in Tables 8 and 9. For Table 8, the sample is working age adults who are active in the labour market, and the outcome variable is the change in log individual labour market income for these adults between the two years in the column headers. Table 9 shows the probability of employment for working age adults in the period after migration, for those who were unemployed in the period before migration. Results from the propensity score matching are shown in the row titled "PSM", and results from the difference-in-difference estimators are shown in the row titled "DD". For each estimator, the coefficient on migration is shown in the "beta" row, the standard error on this coefficient is in the "se" row, and the number of observations is shown in "N".

Table 8 highlights some interesting contrasts. Migrants experienced larger income gains in all intervals except the first. Those who migrated between 2008 and 2010 experienced smaller, or not significantly different, returns than those who did not migrate. This interval spanned the height of the Great Recession, and we believe these results were the result of that particularly difficult economy. In Table 9, unemployed adults who migrated were significantly more likely to be employed than their peers who did not migrate, in every time interval. Both these tables thus suggest that migrants experience better labour market outcomes than comparable (on observables) people who do not migrate.

4.2.2 Longer-run Returns

The situation becomes even more interesting when we consider returns over time. One powerful feature of this dataset is that we can assess outcomes for migrants over multiple post-migration periods. The converse is that we can also look at pre-migration outcomes for some groups of migrants. Tables 10 and 11 show the results from such analyses. The samples and outcomes for these tables are analogous to those in Tables 8 and 9, respectively. In Table 10, each panel measures the change in log individual labour market income between the two years in the column headers. The first panel looks at these changes for migrants who moved between 2008 and 2010. Thus, the first column is the same result as in Table 8; the immediate income change for migrants relative to non-migrants. The second column shows the income change between 2010 and 2012, still comparing the returns of those who migrated before 2010, and those who did not. In other words, the second column is 't+2', if pre-migration is 't', and immediate post-migration is 't+1'. Similarly, column three is 't+3', and column four is 't+4'.

With the added information in these latter three columns, we see that although migrants had poor or indifferent income growth immediately following their migration, they had large and significant growth in the next three intervals. Again,we suspect that this reflects the Great Recession: migrants may have fared worse during the recession, but immediately thereafter they experience very strong growth, suggesting a catch-up effect. In the second panel, we compare people who moved between 2010 and 2012, to those who did not. Their immediate returns to migration are shown in column two. Column one, for this group, is the lag interval, or 't-1'. For this group of migrants, their strong initial income growth tapers off to insignificance in latter periods, for the matching results, while the difference-in-difference results show negative income changes. The third panel compares people who moved between 2012 and 2014/2015 and those who did not. Now, the third column shows their immediate returns, and the first and second columns are both lag intervals. Again, we see large and substantial income growth for migrants relative to non-migrants in the immediate case, but these taper off in the second post-migration interval. Finally, the fourth panel shows income changes over time for those who moved between 2015 and 2017, compared to those who did not. We observe only the immediate outcomes for this group, but they too experienced large, positive, and significant returns.

There are three take-aways from this table. First, it appears that the income gains from migration are a one-time effect, at least in normal economic conditions. Migrants move onto a higher income path, but do not continue to climb, instead returning to the same (or lower) rate-of-change as non-migrants. Second, migration during abnormal economic conditions has a very different pattern. For these migrants, it appears that there was no immediate income gain to migration, but in later intervals, they experienced very large and significant returns. For them, migration was a longer-run investment. Third, there is no clear pattern when looking at income changes in the lag period. Based on this, we cannot say that people who experience a negative income shock are more likely to move, or that a positive income shock allows people to fund their move. One interesting interpretation of this is that it suggests that migrants and non-migrants genuinely were on the same income path prior to migration, and thus provides support to the parallel trends assumption underlying both estimators.

Table 11 presents employment results in the same structure. Each panel shows the probability of employment for those who moved in a particular interval. The first panel is for those who moved between 2008 and 2010, the second for those who moved between 2010 and 2012, etc. Each column shows the probability of employment in a particular year. This sample is restricted to those who were unemployed in the period prior to migration. In all immediate post-migration periods, previously unemployed migrants experienced increased probability of finding a job, relative to unemployed non-migrants. There is no clear pattern for either lag and lead periods.

5 Conclusion

Migration in South Africa, as in many African countries, is very much a story about endogenous household formation. Some members of an original household migrate and create a new household or join another already existing household. In this context, examining the economic returns to migration is a multi-layered exercise. Our descriptive analysis shows that, in the South African context at least, this is not only an intellectual curiosity. The majority of the sample in all waves is touched by migration and a large share of households both send and receive a migrant. Therefore, in addressing the drivers and consequences of this migration, we are seeking to understand an important component of the livelihood strategies of South Africans, especially Black South Africans.

Our analysis starts by describing some important general features of this migration milieu. The majority of moves are 'within type'; for example, a migrant moves within rural areas, rather than migrating to an urban area. This is in contrast to traditional models of migration in developing countries, which suggest that migration is primarily a rural-to-urban move. The modal move is rural-to-urban; however, the majority of migration is still not rural-to-urban. As many of these many of these moves are not long-range moves, clearly migration should not be considered as a once off, discrete move but rather as a process that could involve a number of moves.

Within this important broader context this paper focusses hard on understanding labour market returns to migration for individuals aged 18-65. For this group, 52% of the sample never moves over the ten-year period. However, 50% of the migrants move at least twice. This suggests that, for many of this group of potential labour force participants, migration is not a one-time event, suggesting a more nuanced reallocation of labour than simple rural-to-urban migration with migrants often searching repeatedly in different markets. Our descriptive analysis also highlights some important earnings and employment trends. In the first interval, migrants had lower income gains than non-migrants but, for all other intervals, migrants' income gains were much larger than those for migrants. Gender and education level sub-groups suggest that these findings do not differ by gender and education level. When we examine employment or unemployment in general and by the same categories we find that migrants had have a probability of employment than their non-migrant counterparts.

We use synthetic control methods and difference-in-difference regressions to estimate the individual returns to migration relative to appropriate counterfactual non-migrants. These estimations suggest that the income gains from migration are a one-time effect, at least in normal economic conditions. Migrants move onto a higher income path, but do not continue to climb, instead returning to the same (or lower) rate-of-change as non-migrants. Second, migration during abnormal economic conditions has a very different pattern. For these migrants, it appears that there was no immediate income gain to migration, but in later intervals, they experienced very large and significant returns. For them, migration was a longer-run investment. Third, there is no clear pattern when looking at income changes in the lag period. Based on this, we cannot say that people who experience a negative income shock are more likely to move, or that a positive income shock allows people to fund their move. One interesting interpretation of this is that it suggests that migrants and non-migrants genuinely are on the same income path prior to migration, but migrants resume this path at a higher level. Unemployed adults who migrate are significantly more likely to be employed than their peers who do not migrate, in every time interval. Our estimation work strongly supports the finding that migrants experience better labour market outcomes than comparable (on observables) people who do not migrate.

References

Ardington, Cally, Anne Case, and Victoria Hosegood, "Labor Supply Responses to Large Social Transfers: Longitudinal Evidence from South Africa," *American Economic Journal: Applied Economics*, 2009, *1*, 22–48.

Beegle, Kathleen, Joachim De Weerdt, and Stefan Dercon, "Migration and Economic Mobility in Tanzania: Evidence from a Tracking Survey," *The Review of Economics and Statistics*, 2011, *93* (3), 1010–1033.

Bryan, Gharad, Shyamal Chowdhury, and Ahmed Mushfiq Mobarak, "Underinvestment in a Profitable Technology: The Case of Seasonal Migration in Bangladesh," *Econometrica*, 2014, *September.*

Budlender, D. and F. Lund, "South Africa: A legacy of family disruption," *Development and Change*, 2011, *42* (4), 925–46.

Hall, Katharine and Dorrit Posel, "Fragmenting the family? The complexity of household migration strategies in postapartheid South Africa.," *World Institute for Development Economic Research (UNU-WIDER)*, 2018.

Ham, John, Xianghong Li, and Patricia Reagan, "Matching and Semi-Parametric IV Estimation, A Distance-Based Measure of Migration, and the Wages of Young Men," *Journal of Econometrics*, April 2011, *161* (2), 208–227.

Hamoudi, Amar and Duncan Thomas, "Endogenous Coresidence and Program Incidence: Reexamining South Africa's old age pension," *Journal of Development Economics*, 2014, *109*, 30–37.

Harris, John and Michael Todaro, "Migration, unemployment and Development: A Two-Sector Analysis," *American Economic Review*, 1970, *60*, 126–142.

Heckman, J., H. Ichimura, and P. Todd, "Matching as an Econometric Evaluation Estimator: Evidence from Evaluating a Job Training Program," *Review of Economic Studies*, 1997, *64*, 605–654.

—— , —— , J. Smith, and P. Todd, "Characterizing Selection Bias Using Experimental Data," *Econometrica*, 1998, *66*, 1017–1098.

Lewis, W. Arthur, "Economic Development with Unlimited Supplies of Labor," *Manchester School*, 1954, *22*, 139–191.

Morten, Melanie, "Temporary Migration and Endogenous Risk Sharing in Village India," 2013. Stanford University and Federal Reserve Bank of Minneapolis.

Posel, D. and D. Casale, "Internal labour migration and household poverty in post-apartheid South Africa," in H. Bhorat and R. Kanbur, eds., *Poverty and Policy in Post-Apartheid South Africa*, Vol. Poverty and Policy in Post-Apartheid South Africa, HSRC Press- Cape Town, 2006, pp. 351–365.

Posel, Dorit, James Fairburn, and Frances Lund, "Labour migration and households: A reconsideration of the effects of the social pension on labour supply in South Africa," *Economic Modelling*, 2006, *23* (5), 836–853.

Young, Alwyn, "Inequality, the Urban-Rural Gap, and Migration," *Quarterly Journal of Economics*, November 2013, *128*, 1727–1785.

6 Tables and figures

Table 1: Sample size, by race and by year/wave

Race	2008	2010	2012	2014/2015	2017	Overall
Black	22,206	28,358	30,974	35,301	37,390	47,746
	78.67	*80.64*	*80.87*	*81.62*	*78.19*	*80.04*
Coloured	4,156	4,840	5,298	5,900	6,393	7,668
	14.72	*13.76*	*13.83*	*13.64*	*13.37*	*12.85*
Asian/Indian	429	469	478	512	951	994
	1.52	*1.33*	*1.25*	*1.18*	*1.99*	*1.67*
White	1,435	1,498	1,553	1,535	3,083	3,244
	5.08	*4.26*	*4.05*	*3.55*	*6.45*	*5.44*
Total	28,226	35,165	38,303	43,248	47,817	59,652

Table 2: Characteristics of migrants and non-migrants

	2008-2010	2010-2012	2012-2015	2015-2017	2008-2017
Moved	4,039	5379	6886	7020	14904
	15.32	*16.20*	*18.98*	*17.04*	*34.04*
Did not move	23,231	25945	26199	30225	44773
	88.12	*78.15*	*72.23*	*73.36*	*102.26*
Female, moved	2075	2814	3636	3739	7860
Female, dnm	12631	14109	14306	16455	24297
Male, moved	1964	2564	3250	3280	7042
Male, dnm	10600	11836	11892	13768	7860
% Movers female	*7.24*	*7.90*	*9.38*	*8.54*	*16.12*
% female, movers	*14.11*	*16.63*	*20.27*	*18.52*	*24.44*
% male, movers	*15.63*	*17.81*	*21.46*	*19.24*	*47.26*
Matric, moved	849	881	1330	1543	2200
Matric, dnm	3395	3496	3800	4606	2149
No matric, moved	3158	3251	4899	4668	9876
No matric, dnm	19690	20389	20520	21917	13816
% Movers, matric	*21.19*	*21.32*	*21.35*	*24.84*	*18.22*
% matric, movers	*20.00*	*20.13*	*25.93*	*25.09*	*50.59*
% no matric, movers	*13.82*	*13.75*	*19.27*	*17.56*	*41.68*

Table 3: Age distribution of migrants

	2008-2010		2010-2012		2012-2015		2015-2017		2008-2017	
	Moved	DNM	Moved	DNM	Moved	DNM	Moved	DNM	Moved	DNM
Under 18	1574	9846	1971	10681	2601	10625	2559	12104	5562	5957
% moved	13.78		15.58		19.67		17.45		48.29	
18-24	831	2869	1140	3424	1526	3234	1511	3684	2331	1451
% moved	22.46		24.98		32.06		29.09		61.63	
25-34	785	2734	1085	3260	1383	3434	1544	4250	1940	1722
% moved	22.31		24.97		28.71		26.65		52.98	
35-49	534	3744	719	4067	849	4173	865	4599	1445	3013
% moved	12.48		15.02		16.91		15.83		32.41	
50-64	211	2570	295	2890	362	3109	379	3646	575	2386
% moved	7.59		9.26		10.43		9.42		19.42	
Over 65	85	1448	152	1586	160	1585	159	1902	273	1504
% moved	5.54		8.75		9.17		7.71		15.36	

Table 4: Migrants, and those affected by migration

	2008-2010		2010-2012		2012-2015		2015-2017		2008-2017	
Any mover										
Unaffected	13,658	*39.86*	15,942	*39.16*	13,969	*29.92*	17,950	*38.45*	3,471	*5.82*
Affected	20,604	*60.14*	24,773	*60.84*	32,714	*70.08*	28,738	*61.55*	56,206	*94.18*
Sending	8,986	*43.61*	11,642	*46.99*	15,880	*48.54*	15,688	*54.59*	21,404	*38.08*
Receiving	17,507	*84.97*	18,375	*74.17*	24,385	*74.54*	22,222	*77.33*	42,524	*75.66*
Original Wave 1 only										
Unaffected	13,658	*48.76*	13,568	*48.27*	11,203	*39.89*	12,699	*48.75*	3,260	*11.55*
Affected	14,355	*51.24*	14,538	*51.73*	16,881	*60.11*	13,349	*51.25*	24,966	*88.45*
Sending	8,986	*62.60*	9,624	*66.20*	11,367	*67.34*	9,586	*71.81*	21,404	*85.73*
Receiving	11,258	*78.43*	11,645	*80.10*	13,668	*80.97*	10,392	*77.85*	23,077	*92.43*
CSM migration only										
Unaffected	19,007	*63.29*	19,953	*55.61*	18,616	*45.46*	22,653	*51.89*	19,332	*32.39*
Affected	11,023	*36.71*	15,930	*44.39*	22,334	*54.54*	20,999	*48.11*	40,345	*67.61*
Sending	8,986	*81.52*	9,624	*60.41*	11,367	*50.90*	9,586	*45.65*	21,404	*53.05*
Receiving	6,119	*55.51*	10,051	*63.09*	14,116	*63.20*	13,076	*62.27*	25,394	*62.94*

Table 5: Direction of migration, for all movers and for those who changed districts

	2008-2010		2010-2012		2012-2015		2015-2017		2008-2017	
Panel A: All working-age migrants										
UU	1184	*48.31*	2417	*48.40*	4692	*49.81*	6349	*52.78*	3915	*38.70*
UR	208	*8.49*	209	*4.19*	504	*5.35*	686	*5.70*	565	*5.59*
RR	560	*22.85*	1539	*30.82*	3043	*32.30*	4109	*34.16*	3351	*33.13*
RU	499	*20.36*	829	*16.60*	1181	*12.54*	886	*7.36*	2285	*22.59*
Panel B: Changed district, only										
UU	253	*29.15*	268	*21.49*	632	*29.85*	496	*25.91*	504	*28.31*
UR	148	*17.05*	136	*10.91*	313	*14.79*	441	*23.04*	115	*6.46*
RR	134	*15.44*	257	*20.61*	330	*15.59*	316	*16.51*	546	*30.67*
RU	333	*38.36*	586	*46.99*	842	*39.77*	661	*34.54*	615	*34.55*

Table 6: Labour market outcomes, for working age respondents

		2008-2010	2010-2012	2012-2015	2015-2017	2008-2017
Panel A: Changes in individual labour market income						
Move	Male	-0.4849	2.4783	2.5630	2.0164	3.0031
	Female	-0.0142	2.3387	2.0177	1.9839	3.1388
DNM	Male	0.3077	1.1714	1.2810	1.1999	2.0989
	Female	0.4302	0.9810	1.3695	1.2099	2.7239
Move	Matric	-0.3979	2.1961	2.4995	1.9888	3.0740
	No matric	-0.1890	2.3063	1.9660	1.7781	2.8883
DNM	Matric	0.2803	1.0713	1.2386	1.1890	2.1609
	No matric	0.3926	0.9415	1.3232	1.1778	2.1147
Panel B: Employment shares						
Move	Male	0.5700	0.5163	0.5988	0.6276	0.6088
	Female	0.3388	0.3646	0.4092	0.4211	0.4250
DNM	Male	0.3829	0.4102	0.4677	0.4967	0.4944
	Female	0.2736	0.2995	0.3461	0.3591	0.3524
Move	Matric	0.6200	0.6273	0.7067	0.6539	0.7285
	No matric	0.3558	0.3735	0.4112	0.4420	0.4493
DNM	Matric	0.5545	0.5710	0.6295	0.6068	0.6765
	No matric	0.2621	0.2886	0.3314	0.3540	0.3365

Table 7: Migration patterns

Panel A:

mover_2=1 **3,136**							
mover_3=1 **1,495**				mover_3=0 **1,576**			
mover_4=1 **547**		mover_4=0 **922**		mover_4=1 **546**		mover_4=0 **993**	
mover_5=1 **256** ***1.35***	mover_5=0 **279** ***1.47***	mover_5=1 **161** ***0.85***	mover_5=0 **746** ***3.94***	mover_5=1 **242** ***1.28***	mover_5=0 **303** ***1.60***	mover_5=1 **176** ***0.93***	mover_5=0 **805** ***4.25***

Panel B:

mover_2=0 **17,114**							
mover_3=1 **2,261**				mover_3=0 **14,519**			
mover_4=1 **1,072**		mover_4=0 **1,145**		mover_4=1 **2,430**		mover_4=0 **11,726**	
mover_5=1 **466** ***2.46***	mover_5=0 **580** ***3.06***	mover_5=1 **325** ***1.71***	mover_5=0 **800** ***4.22***	mover_5=1 **1,044** ***5.51***	mover_5=0 **1,343** ***7.09***	mover_5=1 **1,531** ***8.08***	mover_5=0 **9,895** ***52.21***

Working age sample, unweighted. Each node records whether an individual moved in a given interval. mover_2 records a move between Wave 1 and Wave 2, mover_3 between Wave 2 and Wave 3, and so on. The number below the node is a count of how many people reach that node, with that specific history of moves.

Table 8: Returns to migration, based on changes in individual labour market income

		2008-2010	2010-2012	2012-2015	2015-2017
PSM	beta	-0.2523	0.5011	0.2522	0.5134
	se	*0.1270*	*0.1610*	*0.1015*	*0.1185*
	N	10,859	8,461	12,652	8,233
DD	beta	0.0696	0.3928	0.4468	0.5746
	se	*0.1163*	*0.1316*	*0.0886*	*0.1051*
	N	9,796	7,177	10,642	7,378

Table 9: Returns to migration, based on the probability of employment for the unemployed

		2010	2012	2014/2015	2017
PSM	beta	0.1039	0.1493	0.0785	0.1414
	se	*0.0206*	*0.0162*	*0.0135*	*0.0147*
	N	6,100	8,874	9,023	7,598
DD	beta	0.0469	0.1228	0.0701	0.1359
	se	*0.0163*	*0.0140*	*0.0128*	*0.0131*
	N	6,100	7,765	7,688	6,766

Table 10: Longer-run returns to migration, based on changes in individual labour market income

			2008-2010	2010-2012	2012-2015	2015-2017
Mover_2	Psmatch	beta	-0.2523	1.4932	0.2213	-0.0800
		se	*0.1270*	*0.1360*	*0.1548*	*0.1477*
		N	10,859	9,912	7,477	6,074
	DD	beta	0.0696	1.6715	0.2572	0.0899
		se	*0.1163*	*0.1264*	*0.1365*	*0.1361*
		N	9,796	9,491	7,330	6,072
Mover_3	Psmatch	beta	-0.0216	0.5011	-0.0024	-0.0886
		se	*0.1496*	*0.1610*	*0.1521*	*0.1447*
		N	7,899	8,461	8,264	6,815
	DD	beta	0.3883	0.3928	-0.4195	-0.3823
		se	*0.1464*	*0.1316*	*0.1487*	*0.1512*
		N	6,440	7,177	7,063	5,914
Mover_4	Psmatch	beta	-0.4237	0.1049	0.2522	-0.0791
		se	*0.1523*	*0.1076*	*0.1015*	*0.0887*
		N	6,843	12,270	12,652	10,831
	DD	beta	0.0445	0.1439	0.4468	-0.2760
		se	*0.1584*	*0.1099*	*0.0886*	*0.0830*
		N	5,965	9,882	10,642	9,393
Mover_5	Psmatch	beta	-0.0981	0.1844	-0.1557	0.5134
		se	*0.1514*	*0.1254*	*0.1116*	*0.1185*
		N	6,927	8,860	9,652	8,233
	DD	beta	0.4074	0.1057	-0.1601	0.5746
		se	*0.1553*	*0.1223*	*0.0997*	*0.1051*
		N	5,983	7,109	7,806	7,378

Table 11: Longer-run returns to migration, based on the probability of employment for the unemployed

			2010	2012	2014/2015	2017
Mover_2	Psmatch	beta	0.1039	0.0159	0.0414	0.0209
		se	*0.0206*	*0.0240*	*0.0216*	*0.0226*
		N	6100	5813	5634	4648
	DD	beta	0.0469	0.0157	0.0070	0.0237
		se	*0.0163*	*0.0202*	*0.0188*	*0.0189*
		N	6100	5813	5634	4648
Mover_3	Psmatch	beta	0.0287	0.1493	0.0444	0.0475
		se	*0.0147*	*0.0162*	*0.0196*	*0.0209*
		N	5961	8874	6997	5608
	DD	beta	0.0312	0.1228	0.0120	-0.0238
		se	*0.0113*	*0.0140*	*0.0184*	*0.0202*
		N	5423	7765	6402	5010
Mover_4	Psmatch	beta	0.0032	0.0101	0.0785	0.0260
		se	*0.0149*	*0.0131*	*0.0135*	*0.0159*
		N	4,713	7,310	9,023	6,840
	DD	beta	-0.0061	0.0056	0.0701	0.0241
		se	*0.0152*	*0.0099*	*0.0128*	*0.0146*
		N	4,437	6,329	7,688	6,076
Mover_5	Psmatch	beta	-0.0096	0.0206	0.0181	0.1414
		se	*0.0142*	*0.0124*	*0.0120*	*0.0147*
		N	4,818	7,385	9,425	7,598
	DD	beta	0.0363	-0.0020	0.0069	0.1359
		se	*0.0147*	*0.0099*	*0.0124*	*0.0131*
		N	4,453	6,402	7,730	6,766