Name: Nwabisa Nontenja

Student Number: 0711941F

Course: MCOM Economics Research Report

Declaration

I know that plagiarism is wrong. Plagiarism is to use another's work and pretend that it is one's own.

I have used the Harvard convention for citation and referencing. Each contribution to, and quotation in, this research report from the work(s) of other people has been attributed, and has been cited and referenced.

This report is my own work.

I have not allowed, and will not allow, anyone to copy my work with the intention of passing it off as his or her own work.

Signature _____

Date _____

ABSTRACT

The present study sought to investigate self-selection among internal and international migrants in Gauteng by making use of the Gauteng City Region Quality of Life Survey data. The present study also sought to disentangle the effects of observed and unobserved characteristics in the self-selection of migrants by conducting Oaxaca-Blinder decomposition on overall employment and self-employment outcome variables. Preliminary descriptive statistics indicated that international migrants experienced markedly higher levels of employment than both locals and internal migrants driven by higher rates of informal and self-employment. System GMM analysis of pseudo panel data confirmed these results and showed that international migrants had a higher probability of employment and self-employment. Oaxaca-Blinder decomposition indicated that unobserved characteristics explained the greatest share of the differences in the rates employment and self-employment of locals, internal migrants and international migrants. These results provide evidence for the positive selection of international migrants to Gauteng on unobservable characteristics relevant to the region's labour market.

Key Words

Self-Selection; Migration; Self-Employment; Employment

1. INTRODUCTION

Literature on migration has pointed to the existence of self-selection bias in the decision of individuals to migrate or not. Migrants are not randomly selected from the population of source countries (Borjas, 1987). Instead, individuals who choose to migrate may be systematically different from those who choose not to migrate. Migration represents an adjustment in an individual's portfolio human capital investments, along with activities such as education, training and healthcare, to maximise their overall lifetime returns (Gabriel and Schmitz, 1995). The economic model of migration is thus, one where utility maximising individuals will migrate if the expected utility of moving to an alternative location is greater than the expected utility of remaining in their current location. (Dostie and Leger, 2006). Gabriel and Schmitz (1995) state that individuals consider economic incentives, including wage differentials and differentials in employment opportunities between the country of origin and the destination and the cost of moving including direct outlays from relocation and psychic costs from leaving family and friends when making the decision to migrate. Furthermore, Gabriel and Schmitz (1995: p. 461) state that "life cycle factors such as age, marital status, presence of children, education and acquired labour market skills will affect how individuals evaluate differences between destination and the country of origin". It then follows that for a population that is heterogeneous in factors such as ability, motivation, education, and life cycle characteristics, the returns associated with migrating across different locations are likely to be heterogeneous (Dostie and Léger, 2009). The cost and benefits of migration and consequently, the incentive to migrate are likely to depend on both individual-specific and location-specific characteristics. Those who make the decision to migrate will therefore be a self-selected subset of the population.

Self-selection in migration has important implications for understanding the economic and sociological consequences of migration for the countries of origin and destination respectively

and for the migrants themselves (Chiswick, 2000). The ability to draw favourably selected migrants will have a positive impact on the destination country's economy and society. Migrants who possess high innate ability are also more likely to be successful in their adopted countries. However, the loss of high ability individuals to migration will adversely affect countries of origin (Chiswick, 2000). The impact on wages and employment of locals as well as the labour market performance of the migrants themselves is determined by the structure of the migrant population with respect to their abilities and human capital characteristics (Brucker and Trubswetter, 2004). Thus, the self-selection of migrants has important consequences for economic growth, labour markets and the fiscal balance of the welfare state (Brucker and Trubswetter, 2004).

Issues of self-selection in migration are of relevance to South Africa. The country has experienced a high level of both internal and international migration, particularly from the SADC region and from regions around the world. The country is a preferred destination for various categories of migrants and faces a host of migration-related challenges including the increased prevalence of undocumented migration, inadequate migration management policies and rising xenophobic sentiments and violence against migrants (International Organisation of Migration, 2014). There is therefore an imperative to study the characteristics of migrants to South Africa and how they are likely to affect socioeconomic outcomes in both South Africa and the various countries of origin (Kaestner and Malamud, 2011). The present study will seek to investigate the presence of self-selection of internal and international migrants to Gauteng, making use of the Gauteng City Region's Quality of Life Survey data. It will further seek to decompose the effects of observed and unobserved characteristics of migrants on self-selection and employment outcomes. The paper will proceed as follows: Section two presents a summary of the model of self-selection proposed by Borjas (1987). Section three will review relevant literature and end with the resultant research questions. Section four will discuss a description

of the data used in the study. Section five discusses the methodology of the study. Section six presents the descriptive statistics. Section seven will present the study's results followed by the conclusion in section eight.

2. A THEORETICAL MODEL OF SELF SELECTION

Borjas (1987) puts forward that the earnings of the immigrant population may be expected to differ from the earnings of the native population because of the endogeneity of the decision to migrate. The following summarises Borjas' (1987: pp. 533-534) approach. Suppose two countries: 0 denotes the country of origin while 1 denotes the country of destination. Residents of the home country have earnings distributed as $ln\ w_0 = \mu_0 + \varepsilon_0$, where $\varepsilon_0 \sim N\ (0,\ \sigma_0^2)$. The earnings facing this population were they to migrate to the destination country are given by $ln\ w_1 = \mu_1 + \varepsilon_1$, where $ln\ w_1 = \mu_1 + \varepsilon_1$, where $\varepsilon_1 \sim N\ (0,\ \sigma_1^2)$. The correlation coefficient of ε_0 and ε_1 is ρ.

These equations describe the earnings distributions facing a given individual that is considering emigration. This framework decomposes individual earnings into a part due to observable socioeconomic variables (μ_0 and μ_1) and a part due to unobserved characteristics (ε_0 and ε_1). The parameter μ_1 is the mean income that residents from the home country would earn in the destination country if all citizens of the country of origin were to migrate. This level need not be the same as that of the destination country's native population since the average skills for the two population may differ. Borjas (1987) assumes that these intercountry differences in skill are standardised, and hence μ_1 also gives the earnings of the average native worker in the destination country. The migration decision can be modelled into

$$ln\ w_1 - ln\ w_0 - c > 0$$

Where c is the cost of migration (Cristina, 2007; Chiquar and Hanson, 2002). Let Q_1 be the income differential between the average migrant and the average person in country 0, Q_1 be the

income differential between the average migrant and the average native person in the country of destination. Let $k = \sigma_1 / \sigma_0$.

Borjas (1987) therefore identifies the following cases of interest that summarise the quality differentials between migrants and the native born:

Positive selection, $Q_1 > 0$ and $Q_0 > 0$. In this situation, the "best" persons leave the country of origin and when they get to the country of destination, they outperform the native population. Borjas (1987) goes on to explain that the necessary and sufficient conditions for positive selection to occur are that $\rho > \min (1/ k, k)$. If ρ is sufficiently high and if income is more dispersed in the country of destination than in the country of origin, the immigrants arriving in the country of destination are indeed selected from the upper tail of the country of origin's income distribution and will outperform the native born.

Negative selection, $Q_0 < 0$ and $Q_1 < 0$. In this type of selection, the country of destination draws persons from the lower tail of the country of origin's income distribution and these immigrants do not perform well in the country of destination's labour market. The condition for negative selection is that $\rho > \min (1/ k, k)$. Negative selection again requires that ρ be sufficiently positive but that the income distribution be more unequal in the country of origin than in the country of destination.

Refugee sorting, $Q_0 < 0$ and $Q_1 > 0$. Here, the country of destination draws below average migrants from the country of origin but they outperform the native born upon arrival. The condition for negative selection is that $\rho < \min (1/ k, k)$.

The Borjas (1987) model shows that contrary to prior studies such as Chiswick (1978), there is no general law stating that migrants must be positively selected. It can even be expected that, under a reasonable set of conditions, individuals who choose to migrate may be those with below average earnings and productivity (Borjas, 1991).

3. LITERATURE REVIEW

Self-selection in the decision to migrate has received wide-ranging attention in the literature and has gained importance as researchers seek to understand immigration and its impact on migrating individuals, the countries and communities they leave behind and the countries and communities that receive them.

3.1 Theoretical Approaches to Self-Selection

One of the earliest contributions was Chiswick's (1978) seminal study that examined the effect of migrant characteristics and years since migration on the earnings of foreign-born adult white men in the United States. Chiswick (1987: p. 899) hypothesised that whether migrant earnings eventually surpassed, equalled or remained below the earnings of locals depended not only on the time that has passed since migration and location-specific skills but also on whether immigrants are positively or negatively selected on "innate labour market ability and work motivation". Chiswick (1978: p. 899) argued that because migration is more profitable for individuals who are more able and more motivated, the self-selection of migrants implies that for the same demographic characteristics, "immigrants to the US have more innate ability or motivation relevant to the labour market that native-born persons". A multivariate linear regression on cross-sectional data confirmed the author's positive selection hypothesis. He finds that "although immigrants initially earn less than the native born, their earnings rise more rapidly with U.S. labour market experience, and after 10 to 15 years, their earnings equal and then exceed that of the native born".

Borjas (1985) points out the limitations of cross-sectional data in the study of self-selection in migration. He states that the analyses of cross-sectional data in "first-generation" studies including Chiswick (1978) led to the conclusion that immigrant earnings and years since migration are positively correlated and that immigrant earnings tend to overtake the earnings of comparable native workers. However, Borjas (1985) argues that these findings can be

explained by immigrant assimilation over time or by differences in the skills mixes and abilities of successive migrant cohorts. Therefore, cross-sectional analysis may tend to overstate the positive selection of migrants. Borjas (1985) therefore made use of cohort analysis to track synthetic cohorts over time and in contradiction to Chiswick (1978), found evidence of the negative selection of migrants. He found that cross-sectional analysis did overstate the rate of migrant assimilation and that subsequent migrant cohorts were unlikely to eventually to see their earnings equal or exceed those of native-born individuals. To control for age and cohort effects, the literature has shifted to the analysis of cohort and longitudinal data when investigating self-selection as in Borjas (1987) and Borjas, Bronars and Trejo (1992b) while more recent examples include Chiquiar and Hanson (2002), Dostie and Léger (2006) and Nakosteen, Westerlund and Zimmer (2008).

Results of empirical studies on migrant selection vary. Borjas *et al*., (1992b) made use of the Roy (1951) model of the effect of selection on the distribution of output and earnings as a conceptual framework to analyse internal migration flows in the US. They use longitudinal survey data giving insight into the labour market activity of young people. Their results suggest that "interstate differences in the returns to skills are a major determinant of both the size and skill composition of internal migration flows" with highly skilled individuals relocating to states that offer higher skills premiums (Borjas *et al*., 1992b: p. 4). In their seminal paper, Chiquiar and Hanson (2002) made use of data from the Mexico and U.S. population censuses to test Borjas' (1987) negative-selection hypothesis. They find that Mexican immigrants, while much less educated than U.S. natives, are on average more educated than residents of Mexico and were Mexican immigrants in the U.S. to be paid according to current skills prices in Mexico, they would tend to occupy the middle and upper portions of Mexico's wage distribution. This result contradicts the negative selection hypothesis and suggests that

immigrants from Mexico are positively selected in terms of observable skills (Chiquiar and Hanson, 2002).

3.2 The Role of Unobservable Characteristics

Literature has also sought to disentangle the effect of unobservable characteristics in the decision to migrate. Even in the case of migrants and non-migrants who appear similar in observable characteristics such as age or education, there are potential differences in latent attributes that affect the migration decision (Nakosteen *et al*., 2008). The result of this higher innate ability and motivation is the substantial increase over time in the earnings potential of immigrants, eventually surpassing those of native born workers (Gabriel and Schmits, 1995). Findings of studies suggest that models predicting the distribution of migrant outcomes based only on observable migrant characteristics may tend to bias the nature of the selection revealed by data regardless of whether there is positive or negative selection.

Dostie and Léger (2009) tested the model proposed by Borjas (1987) and Borjas *et al*. (1992b) while considering the separate contribution of observable and unobservable characteristics in earnings. Using survey data of Canadian physicians, they found that unobservable characteristics, and their relative returns across different locations, play an important role in the migration decision of individuals and that ignoring unobservable characteristics may lead to the false rejection of the Borjas model of selection in migration (Dostie *et al*., 2009). Similarly, Borjas *et al*. (2015) found that the Roy (1951) model has more precise predictions about self-selection of migrants than previously realised when self-selection is decomposed into observable and unobservable characteristics. Making use of the Danish full population administrative data, Borjas *et al*. (2015) found that more than half of the difference between the expected earnings of migrants and non-migrants arises because of differences in unobserved characteristics.

3.3 Migration in South Africa: A Brief Overview

South Africa has a long history of migration which has been the subject of extensive research across academic disciplines. Migration flows in Southern Africa were largely steered by an institutional framework that shored up an economic system dependent on cheap labour. The apartheid state engaged in a series of interventions to mobilise and control labour while preventing the permanent urban settlement for most migrants (Posel, 2004). Migration policy under apartheid was tightly regulated and was characteristically racist in design and implementation. The state encouraged and incentivised the immigration and permanent settlement of whites from Europe while blacks from neighbouring African countries were recruited on temporary work contracts to supply labour to South African mines and commercial farms (Crush, 2014). The state restricted the permanent settlement of these migrant labourers by prohibiting them from bringing their families to South Africa and by sending the labourers home once their contracts expired (Crush, 2014).

Since 1990, the period immediately preceding democracy, South Africa has seen growing internal migration and movement of foreign migrants and refugees into the country (Posel, 2004). Cross-border migrants into the county originated primarily from South Africa's historical labour supply areas, that is, countries from Southern African Development Community (SADC). Migrants have also come from African countries further north such as Nigeria and Democratic Republic of Congo (Wentzel and Tlabela, 2006). It is commonly accepted that South Africa's democratisation was a key reason behind the increased levels of migration to the country (Wentzel and Tlabela, 2006). The exact extent of migration in and out of the country is not known and estimates vary widely. This is because much of the movement is undocumented (Segal, Elliot and Mayadas, 2006: p. 366). The absence of exact numbers has allowed the predominance of inflated estimates arrived at through questionable methodologies and these figures are often cited in anti-immigrant discourse. More realistic estimates also vary.

The United Nation's Department of Economic and Social Affairs (2015) estimates that, in 2015, the number of international migrants living in South Africa was an estimated 3,14 million, or 5.7% of the population. The 2016 Community Survey conducted by Statistics South Africa indicates that the number of international migrants was half of that figure, an estimated 1,6 million, or 2.8% of the population.

Despite disparate population estimates, the consensus is that South Africa's migrant population is growing. In light of this, South Africa has seen an alarming growth in xenophobia and hostility toward foreign migrants that often culminates in violence. Xenophobic violence in 2008 resulted in 62 deaths and the displacement of many more (Mail and Guardian, 2008). Several reasons behind the pervasive xenophobia have been advanced. According to the United Nations Human Rights Commission (2013), increased manifestations of xenophobia coincide with periods of economic hardship, election campaigns, political instability and conflict. Considering increasing poverty, high unemployment, constrained provision of housing, social welfare, education and healthcare, foreign migrants become an easy target of hostile attitudes. As the competition for employment and resources intensifies, migrants are ready 'scapegoats' for societal ills (Segal, Elliot and Mayadas, 2006: p. 368).

Prior studies of the labour market performance of migrants in South Africa suggest positive selection of migrants, particularly on entrepreneurial skill. Contrary to the experience of other sub-Saharan African countries, South Africa has a small informal sector that coincides with very high unemployment (Kollamparambil, 2017). Even so, the percentage of international migrants working in the informal sector was found to be twice as high as non-migrants (Fauvelle-Aymar, 2014). This may be because the informal sector has the lowest entry cost into the labour market but there may also be unobserved heterogeneity in worker characteristics that determines selection into a sector (Bargain and Kwenda, 2011). The importance of this unobserved heterogeneity and its influence on the self-selection of workers is that the high

incidence of informal employment of international migrants may coincide with the higher incidence of self-employment and employer status of the same. Budlender (2014) found that international migrants were far more likely to be self-employed or employers than both domestic migrants and non-migrants in South Africa. Kollamparambil (2017) points out that the higher incidence of entrepreneurship among migrants compared to locals may be out of necessity because migrants face difficulty in accessing wage employment, but it could also be the result of a self-selection process where migrants possess observable and unobservable characteristics that enable their success in self-employment compared to locals. Contrary to the prevalent narrative that foreign migrants are job-takers, research shows that a significant proportion of migrants engage in entrepreneurship, out of either election or necessity, and employ both South Africans and migrants in their businesses (Kalitanyi and Visser, 2010; Kalitanyi, 2007; Fatoki and Patswairi, 2012).

3.4 Empirical Studies

Several studies have tested self-selection in migration using of OLS regression estimation. Borjas (1991), in his analysis of the roles played by selection in both observed and unobserved characteristics, addressed age and cohort effects by pooling data from two censuses. He controlled for assimilation by imposing a restriction on the size of the period effect on migrant earnings. Gabriel and Schmitz (1994) propose a human capital model of migration that modifies the migration decision process to account for potential differences in individual labour market ability. In this model, individuals with higher ability or motivation will thus earn a higher rate of return to migration and will have a higher probability of migration. To test their model and to enable comparison over successive time intervals, Gabriel and Schmitz (1994) estimated individual OLS regressions across several interval periods to examine the wage effects of migration. Chiquiar and Hanson (2002) made use of OLS estimation across five samples of individuals to test the Borjas (1987) model among Mexican migrants to the United

States. The authors note that the results of this analysis do not differ significantly to those where sample selection bias was controlled for. Cristina (2007), in her study of self-selection among Argentinian internal migrants, makes use of the Heckman Two-Step procedure to account for self-selection bias. Other studies have turned to maximum likelihood estimation of panel data such Axelsson and Westerlund (1995), in their study of the impact of migration on income for Swedish multi-adult households and Nakosteen *et al*. (2008) in their study of the distinct effects of latent and measured characteristics on self-selection of migrants. Although studies may enable the comparison over successive time intervals, OLS estimation also fails to account for individual heterogeneity and country specific fixed effects that may affect the decision to migrate. Even after controlling for selection bias, OLS estimation may yield biased estimates due to endogeneity arising from the inclusion of variables such as education and a lagged dependent variable in the model. In the instance where selection bias is controlled for, the likely presence of endogeneity bias in the variables may lead to biased estimates. Therefore, to address these limitations, the present study will make use of the GMM method of estimation.

Empirical studies in South Africa give evidence generally pointing to the positive selection of migrants. Several findings of Fauvelle-Aymar's (2014) econometric analysis of the migration module piloted by Statistics South Africa in the third quarter (Q3) 2012 of the Quarterly Labour Force Survey, demonstrate this. Both domestic and international migrants were found to have, on average, higher levels of education than non-migrants. The higher levels of education may point to a higher level of observable skills but also to unobservable characteristics such as motivation and grit. The rate of employment of international migrants was also found to be higher than that of domestic migrants and non-migrants. This study makes use of cross-sectional OLS estimation and as aforementioned, this method does not account for individual heterogeneity or endogeneity in the explanatory variables. Furthermore, cross-sectional analysis does not give insight into changes over time. Budlender (2014), in a statistical analysis

of the same data comes to the same findings. This study, however, is limited to descriptive analysis and tabulation and does not make use of multi-variate regression analysis of key variables. Mbatha and Roodt (2014), using the 2008 and 2010 waves of the National Income Dynamics longitudinal survey, find that internal migrants on average experience higher labour-force participation rates and a higher probability not only of informal but also formal employment than non-migrants. In contrast, Kollamparambil (2017), in a district level analysis of the labour market impact of internal in-migration, found little evidence of positive self-selection among internal migrants in South Africa. In-migration of internal migrants was found to significantly and negatively impact on the informal sector self-employment rate indicating that internal migrants are less engaged in self-employment as compared to locals and international migrants. Mbatha and Roodt (2014) acknowledge that the results of their multinomial logistical model may be overestimations since labour market outcomes are not always exogenous to the decision to migrate. Kollamparambil (2017) on the other hand, makes use of system GMM to account for endogeneity bias arising from reverse causality between the choice of migration destination and labour market conditions. The study does not, however, seek to disentangle the effects of observable and unobservable characteristics in the self-selection and labour market performance of migrants.

Both Mbatha and Roodt (2014) and Kollamparambil (2017) restrict their analysis to internal migrants. The employment outcomes of internal migrants and international migrants in South Africa may differ for various reasons. The rate of assimilation of internal migrants is likely to be higher (Borjas *et al*., 1992). This is because internal migrants are more likely to possess characteristics -such as education and language- that are compatible to the destination region's labour market. Another factor likely to differ between internal and international migrants is the ease of permanent settlement in the destination region. South Africa's current immigration policies have progressively restricted access to unskilled economic migrants and have overtly

favoured skilled immigrants (Carciotto and Mavura, 2016: p. 20). These policies have been criticised as biased against African migrants (Carciotto and Mavura, 2016: p. 20). Furthermore, initiatives such as the recent police operation dubbed "Operation Fiela" targeting undocumented migrants and proposed amendments to migration legislation that include the creation of a Border Management Agency point to an increased emphasis toward migration control (Carciotto and Mavura, 2016: p. 81). These factors mean that international migrants face very different labour market conditions from internal migrants. The present study will therefore distinguish between these categories of migrants.

The contribution of the present study therefore, is to investigate the presence of self-selection among internal and international migrants in Gauteng by investigating the proposition by Borjas (1987) that positively selected migrants can be expected to outperform the native population in the country of destination while negatively selected migrants can be expected to not perform well in the country of destination's labour market. This study will make use of system GMM estimation to address endogeneity bias. In addition, the present study will seek to disentangle the effects of observed and unobserved characteristics in the self-selection of migrants by conducting Oaxaca-Blinder decomposition on relevant employment outcome variables.

The research questions follow:

1) Are there differences in the labour market performance of locals, internal migrants and international migrants with respect to the following variables?
    a. Employment overall
    b. Self-Employment
    c. Formal Employment
    d. Informal Employment

2) Are there differences in the overall well-being of locals, internal migrants and international migrants as captured in the following variables?
    a. Household income
    b. Per capita household income
    c. Household hunger
    d. Social grant recipient
    e. Life Satisfaction
    f. Debt
    g. Years of education
3) To what extent does self-selection determine employment outcomes?

4. DATA

The study makes use of the Gauteng City – Region Observatory Quality of Life (GCRO QOL) survey. This is the largest social attitudes survey conducted in the Gauteng province. The stated aim of the survey is to inform the GCRO, the Provincial Government and other role players about the perceived state of the municipalities within the Gauteng City Region footprint (Development Research Africa, 2009). The four cross-sectional surveys of individuals 18 years and older were conducted in 2009, 2011, 2013 and 2015. The sample was constructed using a multistage stratified sampling approach with 2011 local election wards as the explicit stratification variable (GCRO, 2016). A minimum of 30 respondents per ward was drawn in non-metro wards and 60 respondents in metro wards with no ceiling (GCRO, 2016). According to the GCRO (2016), the aim of was of the sampling procedure was to provide a ward-representative sample of the entire province. The sample size of each wave is 6636, 16 729, 27 490 and 30 000 respectively. The survey allows the analysis of the quality of life of inhabitants, providing insight into values and attitudes, levels of social capital and levels of alienation amongst other dimensions (GCRO, 2015). Gauteng is an appropriate unit of study

because it is home to the highest number of both internal and international migrants in South Africa by a significant margin (Statistics South Africa, 2015). This means that the results of the present study are likely to provide a reliable picture of the migration landscape in South Africa.

International migrants are defined in the data as individuals who were born in a country outside of South Africa. Internal migrants are defined as those who moved to Gauteng from another province. Locals are defined as those who were born in Gauteng. Unlike the three subsequent surveys, the 2009 survey does not have a question asking respondents where they were born. Instead, the survey includes a question that asks respondents "Which province do you consider to be "home"? ". The distribution of responses to this question was in line with the distribution of respondent origin found in the subsequent surveys. Therefore, we proceeded to use this question to identify migrant origin in the 2009 survey. Several other questions relevant to the present study were altered in successive survey questionnaires. For example, in the 2009 and 2011 waves, respondents were given a single list that identified employment status. In 2013 and 2015, employment status was determined by several successive questions. This has meant that our analysis required the use of the available questions to reconstruct some variables of interest including labour force and employment. Care must therefore be exercised when using the GCRO data to compare changes of these variables over time. The GCRO survey does not have a measure of individual income or earnings but includes only total household income earned by all household members. Therefore, we truncate the dataset to include only household heads. We calculate per capita income as total household income divided by the number of individuals within the household.

A limitation of our data is the exclusion of a control for years since migration. This is due to a lack of a relevant measure of years since migration in the 2009 survey. Years since migration allows researchers to account for assimilation effects as migrants adapt to the destination

country's labour market. We partially address this limitation by placing internal migrants and international migrants in separate categories in the estimation model. The process of assimilation is likely to be faster for internal migrants than migrants from other countries (Borjas *et al*., 1992a). Internal migrants adapt more quickly because they are not subject to immigration legislation and are also more likely to possess human capital characteristics compatible with the destination region's labour market. Another limitation of the present study is small sample size. The Arellano-Bond (1991) linear generalised method of moments estimator is designed for situations with few time periods and many individuals -small T, large N (Roodman, 2006). System GMM is, however, considered the appropriate method of estimation to address issues of endogeneity and observed autocorrelation in OLS given that it is found to display the best features in terms of small sample bias and precision (Soto, 2009). To further address issues of misspecification, we include a lagged dependent variable in each of the estimated models. Table 1 details the variable definitions.

5. METHODOLOGY

To answer the research questions, the present study will take a two-pronged approach. The first is to conduct a system GMM analysis on the variables of interest. The second is to conduct Oaxaca-Blinder decomposition on the employment variables of interest to determine the influence of unobservable characteristics in determining employment outcomes. To overcome the cross-sectional nature of the GCRO QOL survey data, the present study will create a pseudo panel across all four waves of the survey. This approach allows for the tracking of cohorts across time and helps address issues arising from measurement error. Antman and Mckenzie (2006) state that the pseudo-panel approach helps deal with measurement error in two ways. The averaging process eliminates individual-level measurement error in the cross-section and since each individual is observed once, measurement errors observed in one period will be

Table 1: Variable Definitions

Employment	The ratio of employed over the total labour force of cohort.
Self-Employment	Ratio of self-employment over total employment of cohort.
Formal Employment	Ratio of formal employment over total employment of cohort.
Informal Employment	Ratio of informal employment over total employment of cohort.
Log Household Income	The QOL Survey divides household income into 16 categories of increasing magnitude. Individuals are asked to indicate total household income that is then assigned to the corresponding category. Log Household Income is the cohort average of these categories.
Log Per Capita Income	To obtain individual level household income, each respondent is assigned the median of their respective income category. To obtain per capita household income, the median is divided by household size.
Social Grant	The ratio of individuals who indicated that they receive some form of social assistance or grant (including but not limited to child grant, disability grant and old-age pension).
Life Satisfaction	The ratio of individuals who indicated that they agreed or strongly agreed to the statement: “How satisfied are you with your life AS A WHOLE these days?”. Response was according to a five-point Likert scale.

Hunger	The ratio of individuals who indicated "Yes" to either of the following: • "In the last year, has there ever been a time when you or any other adult in this household had to skip a meal because there was not enough money to buy food?" • "In the last year, has there ever been a time when there was not enough money to feed the children in the household?"
Debt	The ratio of individuals who indicated "Yes" to the following: "Do you owe money to anyone including a bank or a shop or a money lender?"
Education	The average years of education of individuals in cohort.
Race: Black	The ratio of black individuals in cohort.
Race: White	The ratio white individuals in cohort.
Gender	The ratio of females in cohort.
International	The ratio of individuals born outside of South Africa in cohort.
Internal	The ratio of individuals born outside of Gauteng in cohort.
Local	The ratio of individuals born in Gauteng in cohort.
Age	The average age of individuals in cohort.
AgeSq	The square of the average age of individuals in cohort.

different to those observed in another period. Secondly, non-random attrition poses less of a problem given that each household is observed only once.

5.1 Synthetic Panel Estimation

The typical specification of an individual-level Mincerian panel data regression model is as follows:

$$y_{i,t} = \alpha + \beta x_{i,t} + \varepsilon_{i,t} \qquad \text{with } i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \qquad (1)$$

where $y_{i,t}$ is the dependent variable representing the natural log of monthly income; $x_{i,t}$ is a row vector of explanatory variables including demographic variables such as age and race, years of education and a dummy variable indicating whether individual i is a migrant or not; β is a corresponding column vector of regression coefficients and α is a scalar. Most panel data analyses will make use of a one-way error component model:

$$\varepsilon_{i,t} = \mu_i + v_{i,t} \qquad (2)$$

where μ_i represents unobservable, time-invariant and individual-specific effects and $v_{i,t}$ represents a time-varying and individual-specific error term (Russell and Fraas, 2005). Combining equations 1 and 2 yields:

$$y_{i,t} = \alpha + \beta x_{i,t} + \mu_i + v_{i,t} \qquad \text{with } i = 1, \ldots, N \text{ and } t = 1, \ldots, T, \qquad (3)$$

The time-invariant individual effects, μ_i, include ability, motivation and grit that invariably affect an individual's income and other labour market outcomes (Himaz and Aturapane, 2012). It is likely that μ_i is correlated with years of education and, as the literature suggests, the individual's origin (whether the individual is a local, internal migrant or international migrant). The presence of these unobservable effects, μ_i, will cause an ordinary least squares estimation to be biased. Given the availability of true panel data, individual-specific effects can be accounted for in various ways including constructing instruments, transforming models to first differences or by including individual history in the model (Himaz and Aturapane, 2012).

As aforementioned, in the absence of true panel data, Deaton (1985) suggests the use of a pseudo-panel that tracks cohorts of individuals grouped by some common characteristic over repeated cross sectional surveys. A cohort, *c*, is defined as a group whose membership is fixed over time (Deaton, 1985). The cohorts are defined such that individuals are a member of only one cohort (Himaz and Aturapane, 2012). The present study makes use of birth-year and origin as the cohort defining parameters. Due to sample size limitations, we were unable to define more parameters. Cohort means within each period are treated as observations in the pseudo-panel (Verbeek, 2008). Therefore, taking the mean value of each cohort's sample in each period yields:

$$\bar{y}_{c,t} = \beta \bar{x}_{c,t} + \bar{\mu}_{c,t} + \bar{v}_{c,t} \qquad \text{with } c = 1, \ldots, C \text{ and } t = 1, \ldots, T, \qquad (4)$$

where $\bar{y}_{c,t}$ is the average of all monthly income for all individuals in cohort *c* at time *t;* $\bar{x}_{c,t}$ is now a row vector of the observed means of the explanatory variables within *c* at time *t;* β is a column vector of corresponding regression coefficients; $\bar{v}_{c,t}$ is the mean of individual-specific error terms within *c* in time *t* and $\bar{\mu}_{c,t}$ now represents the mean cohort effect in time *t*. Given that the cohort mean for each period is calculated from a different set of individuals, $\bar{\mu}_{c,t}$ is likely to be different for each period. It therefore retains the subscript, *t* (Russell and Fraas, 2005).

If μ_i is correlated with $x_{i,t}$, $\bar{\mu}_{c,t}$ is likely to be correlated with $\bar{x}_{c,t}$. Verbeek (2008) states that if this is the case, treating $\bar{\mu}_{c,t}$ as part of the error term is likely to lead to inconsistent estimators. If cohort means are based on a sufficiently large number of individuals, $\bar{\mu}_{c,t}$ can be treated as a fixed unknown parameter such that $\bar{\mu}_{c,t} = \mu_i$. This yields the pseudo panel equation:

$$\bar{y}_{c,t} = \beta \bar{x}_{c,t} + \mu_i + \bar{v}_{c,t} \qquad \text{with } c = 1, \ldots, C \text{ and } t = 1, \ldots, T, \qquad (5)$$

Russell and Fraas (2005) further state that if the cell size is large, random individual effects will tend to be eliminated in the process of estimating the cell mean, leaving only the cohort

fixed effect. The present study is unlikely to experience this benefit because of the relatively small cohort cell sizes. This provides further cause for the use of the system GMM method.

Cohorts are defined per household head origin -native born, internal migrant and international migrant- and according birth year in five-year spans, starting with those born between 1991 and 1987 and ending with those born between 1946 and 1942. Therefore, the data was truncated to include household heads who were between 18 years old and 67 years old in 2009. Several cohorts are small and this presents a limitation to the analysis. Nevertheless, to address this issue, those cohorts with a very small number of observations (10 or less) have been dropped from the analysis. Table 2 details the cohort sizes of the synthetic panel.

Table 2: Cohort Definition and Cohort Sizes

	2009			2011			2013			2015		
	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.
Birth Year												
91-87	52	48	4	288	161	38	454	482	186	1063	627	252
86-82	153	82	27	428	249	81	647	668	303	1424	799	335
81-77	292	127	30	526	273	67	779	777	345	1424	761	273
76-72	264	123	23	550	270	42	824	666	212	1349	692	205
71-67	299	109	11	557	207	37	910	605	167	1150	545	118
66-62	253	98	6	515	219	32	807	531	112	1047	434	75
61-57	235	76	3	491	157	26	782	452	79	943	364	76
56-52	164	70	2	412	135	13	677	340	57	772	309	40
51-47	165	39	1	395	130	10	641	259	34	558	181	40
46-42	148	42		268	101	16	359	188	33	385	136	27
Total	2025	814	107	4430	1902	362	6880	4968	1528	10115	4848	1441

Source: Author's calculations based on the GCRO Quality of Life Survey data

5.2 Estimation Strategy

For the sake of comparison and to check robustness, the first part of the present study will conduct pooled OLS and system general method of moments analysis. The inclusion of origin dummy variables invariant over time precludes the use of fixed effect analysis. Random effects estimation is also precluded since the assumption that the cohort specific effect is a random variable uncorrelated with the explanatory variables of all past, current and future time periods of the same cohort is unlikely to hold (Schmidley, 2016). The presence of unobserved individual heterogeneity and country-specific fixed effects that affect the decision to migrate mean the OLS estimation may yield biased estimates (Borjas, 1987). Furthermore, the assumption of strict exogeneity is not a natural restriction given likely labour market dynamism and inertia. The current study will therefore make use of GMM to account for this endogeneity.

The system GMM method of estimation is applicable to a wide range of problems in economics, particularly when the model of interest contains endogenous or predetermined explanatory variables but the processes generating these series are not completely specified (Bond, 2002). GMM estimation allows economic models to be specified while avoiding unwanted or unnecessary assumptions such as specifying a particular distribution for the errors (Sheppard, 2015). Some assumptions about the data-generating process embodied in the GMM estimators are that there may be arbitrarily distributed fixed individual effects; the process may be dynamic; some regressors may be endogenous and idiosyncratic disturbances may have individual-specific patterns of heteroscedasticity and serial correlation (Roodman, 2009).

The system GMM method has won increasing favour among researchers when dealing with dynamic panel models. Arellano and Bond (1991) developed the difference GMM estimation. This method transforms all regressors by first differencing and instruments endogenous variables with lags of their own levels (Roodman, 2009; Kollamparambil, 2017). This estimator was later augmented by Arellano and Bover (1995) and Blundell and Bond (1998) who made

the additional assumption that first differences of instrument variables are uncorrelated with the fixed effects which allowed for more instruments and greater efficiency (Roodman, 2009). This approach, called system GMM, augments the difference GMM estimator by adding original equations in levels. Strictly exogenous and endogenous variables in levels are instrumented with lags of their own first differences (Kollamparambil, 2017).

5.3 Model Specification

The current study will undertake econometric analysis of data using the following model:

$$Y_{c,t} = \alpha + \beta X_{c,t} + \varepsilon_{c,t} \quad (6)$$

where $Y_{c(t),t}$ denotes the sample mean of the dependent variable of cohort c observed in time t. According to the set of research questions, the following are the dependent variables that the model will test:

- $ln\, y_{c,t}$: the mean of log household income of cohort c observed in time t;
- $E_{c,t}$: the employment rate of cohort c observed in time t.
- $SE_{c,t}$: the self-employment rate of cohort c observed in time t.

Additional dependent variables that this paper will consider are the following:

- $D_{c,t}$: the percentage of individuals who indicated that they have some form of debt in cohort c observed in time t;
- $H_{c,t}$: the mean number of households where an adult or child has missed a meal in cohort c observed in time t
- $LS_{c,t}$ denotes the percentage of individuals who are satisfied with their lives overall. On a likert scale of 1 to 5, respondents indicate their level of satisfaction with their lives overall.
- The $X_{c,t}$ denotes a vector of the cohort mean of explanatory variables observed at time t. The control variables are years of education, age, age-squared. In addition, we include a measure of race which denotes the percentage of respondents who are black in cohort c

observed at time t and a measure of gender which denotes the percentage of respondents who are female in cohort c at time t (Mbatha and Roodt, 2014). The variables of interest describe respondent origin – these are dummy variables indicating origin (local, internal, international).

- $\varepsilon_{c,t}$ denotes the composite error term.

6. DESCRIPTIVE STATISTICS

Table 3 details the demographic profile of the sample. As aforementioned, the sample was truncated to include household heads born between 1991 and 1942. The data reveals that the international migrant population is younger than the local and internal populations. In all four waves of the GCRO QOL survey, a larger cumulative percentage of international migrants is 47 years old or younger than is the case with both internal migrants and locals. Males represent most the sample for all three groups in all four waves. This is accounted for by the fact that the unit of measurement is the household head. The share of males is notably larger for international migrants. This is in line with the literature which points to the gendered nature of migration (Kanaiaupuni, 2000). Studies indicate that factors such as civil status, the probability of marriage, number of children, the gender composition of migrant networks and wage differentials affect the propensity of women to migrate differently than they do men (Berman and Wolfe, 1984; Kanaiaupuni, 2000; Cackley, 1993; Davis and Winters, 2001). The data also indicates that an overwhelming majority of migrants into Gauteng are black. Migrant households are found to be, on average, smaller than those of locals with international migrant households being the smallest on average. Kollamparambil and Mulcahy (2016) found that South African rural-urban migrants experienced sharp reduction in average household size post migration. This may be evidence of migrants pursuing what Fan, Sun and Zheng (2011) describe as a circulation strategy that entails splitting the household into two or more places for economic betterment. South Africa has a long history of circular migration underpinned by

Table 3: Demographics by Origin and Survey Year

	2009			2011			2013			2015		
	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.
Age (%)												
18-27	10.12	15.97	28.97	11.85	16.09	22.40	7.97	11.92	15.37	7.91	10.09	13.19
28-37	27.46	30.71	49.53	23.68	28.13	36.53	21.42	29.11	42.05	27.70	31.99	42.82
38-47	27.26	25.43	15.89	25.03	23.76	19.73	25.09	25.54	23.26	25.65	26.67	24.64
48-57	19.70	17.94	4.67	21.22	17.14	12.00	23.11	19.10	11.96	19.64	16.89	11.03
58-67	15.46	9.95	0.93	15.67	12.36	7.20	18.31	11.27	5.58	14.45	10.79	6.04
Overall Mean Age												
	43	42	32	43	41	37	46	42	38	44	42	39
Gender (%)												
Male	52.25	53.81	72.90	54.56	54.47	62.40	53.08	58.43	71.68	54.44	58.37	69.47
Female	47.75	46.19	27.10	45.44	45.53	37.60	46.92	41.57	28.32	45.56	41.63	30.53
Race (%)												
African	79.95	90.54	96.26	76.64	87.07	82.67	80.00	90.84	86.73	75.58	88.24	85.70
Coloured	0.69	0.74	0.00	2.64	2.42	4.80	4.01	1.61	0.20	5.27	2.50	0.35
Indian/Asian	4.74	0.86	0.00	4.11	2.16	1.33	1.86	1.49	1.58	2.14	2.02	3.05
White	14.62	7.86	3.74	16.61	8.36	11.20	13.98	5.96	8.28	16.88	7.14	7.08
Other	0.00	0.00	0.00	0.00	0.00	0.00	0.15	0.10	3.22	0.14	0.10	3.82
Mean Household Size												
	3.20	3.19	2.30	3.56	3.39	3.18	3.28	2.82	2.36	3.14	2.97	2.50
Sample Size (N)												
	2025	814	107	4430	1902	375	6880	4968	1522	10115	4848	1441

Source: Author's calculations based on the GCRO Quality of Life Survey

a well-established system of migrant labour and apartheid policy that prevented as far as possible, the permanent settlement of labourers. Collinson, Tollman, Kahn and Clark (2006) found that high levels of circular migration between rural and urban areas continue to prevail in the country. Research suggests that circular migration may be persistent and that permanent migration may not be inevitable. Fan *et al*. (2011: p. 2165), citing research by Hugo (1982) on Indonesian migrants, states that these migrants did not perceive their mobility as a preliminary state before permanent relocation but rather, they

> "exhibit a strong and apparently long-term commitment to bi-locality, opting for the combination of activities in both rural and urban areas that a non-permanent migration strategy allows them".

Table 4 details the average years of education by birth year and origin for all four waves of the GCRO survey. The data indicates that, on average, older individuals have fewer years of education. This phenomenon is, however, less prominent among international migrants.

Table 4: Mean Years of Education by Birth Year, Origin and Survey Year

	2009			2011			2013			2015		
	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.
Birth Year												
91-82	11.41	11.4	9.42	11.06	10.73	9.99	12.63	12.46	11.46	12.82	12.55	11.18
81-72	11.15	10.65	10.27	11.33	10.49	9.46	12.68	12.1	11.28	12.67	11.93	11.16
71-62	10.19	9.25	8.63	10.71	9.26	9.55	11.64	10.3	10.64	11.89	10.87	10.54
61-52	9.04	7.9	10.67	9.17	8.14	8.69	10.08	8.6	10.18	10.56	9.15	10.03
51-42	6.95	6.37		8.24	7.56	8.96	9. 49	8.39	11.03	9.51	8.4	10.7
Overall Mean Years of Education												
	9.85	9. 34	9.79	10.26	9.56	9.56	11.41	10.9	11.11	11.9	11.3	10.96

Source: Author's calculations based on the GCRO Quality of Life Survey

Table 5 details the average employment rates by birth year and origin for all four waves of the GCRO survey. The employment rate increases for older individuals. This is true for all three origin groups. This finding, coupled with the lower mean levels of education of older individuals indicates that the probability of employment increases with age and that age exerts a greater influence on employment probability than years of education, a finding in line with Fauvelle-Aymar (2014). The data also indicates that the rate of employment for international migrants is higher than that of locals and internal migrants for all age groups and for all survey waves. In 2013 and 2015, internal migrants exhibited higher rates of employment than locals.

Table 5: Mean Employment Rate by Birth Year, Origin and Survey Year

	2009			2011			2013			2015		
	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.
Birth Year												
91-82	56.82	52.63	60	50.18	53.20	75.25	57.84	67.62	80.56	70.06	70.40	78.49
81-72	67.66	68.85	80.39	65.10	55.74	67.92	63.78	72.41	85.47	72.82	70.30	78.98
71-62	65.35	73.06	87.50	64.60	61.80	68.85	66.29	73.39	88.10	73.08	72.32	83.72
61-52	69.44	69.23	100	58.93	58.17	85.71	65.85	80.40	81.65	70.30	76.30	86.21
51-42	80.77	80.95		70.69	72.00	77.78	76.16	88.17	90.32	68.05	80.00	85.71

Source: Author's calculations based on the GCRO Quality of Life Survey

Figures 1 to 4 provide a more detailed view of the employment status of the origin groups. The higher rate of employment experienced by international migrants is driven by higher rates of informal and self-employment -that is- a greater percentage of migrants that are employed work in the informal sector or are self-employed. Figures 5 to 8 plot household income for the three origin groups across all survey years. In the survey years 2009 to 2013, the largest cumulative percentage of individuals had a household income of R3 200 or less. This was true for all three

origin groups. The survey year 2015 saw an increase in the percentage of individuals occupying higher household income brackets. This was also true for all three origin groups.

Figure 1: Mean employment rates by origin in 2009

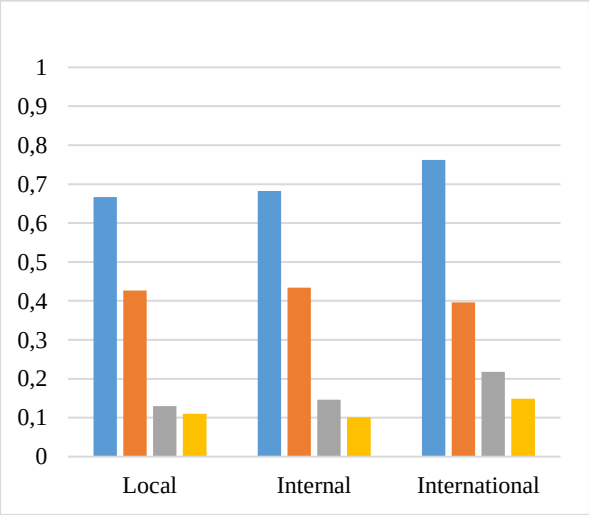


Figure 2: Mean employment rates by origin in 2011

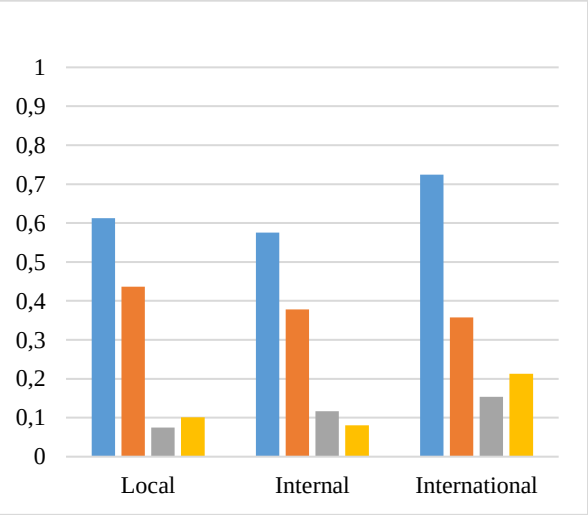


Figure 3: Mean employment rates by origin in 2013

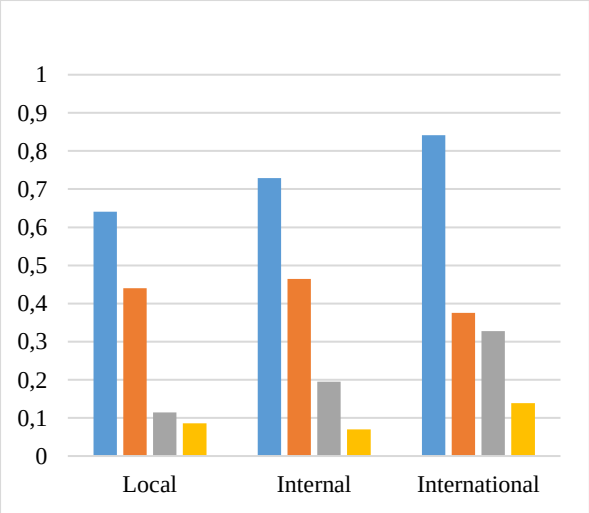
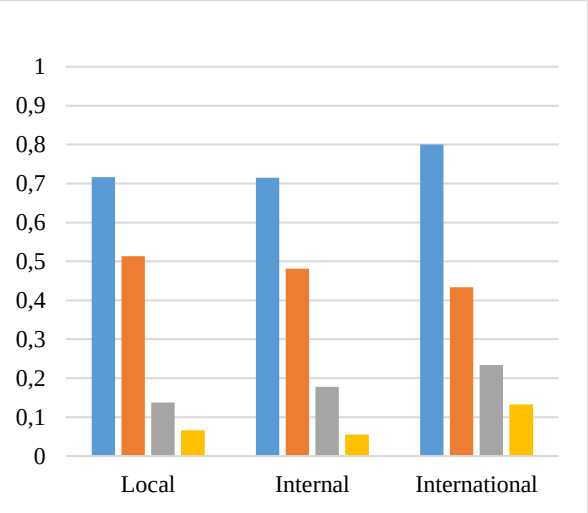


Figure 4: Mean employment rates by origin in 2015



Source: Author's own calculations from GCRO Quality of life Survey Data

Figure 5: Household Income by Origin in 2009

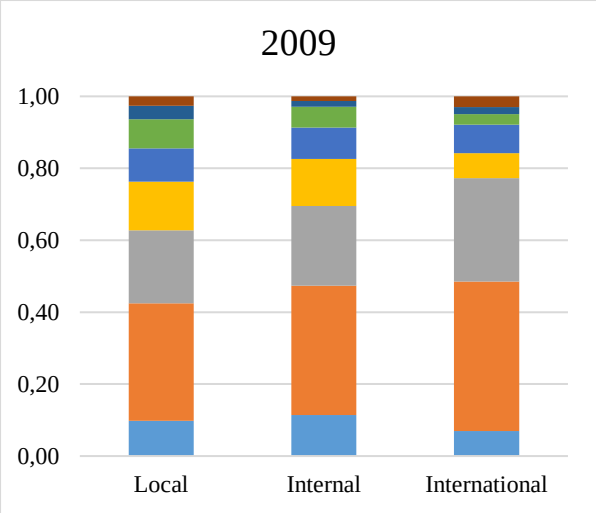


Figure 6: Household Income by Origin in 2011

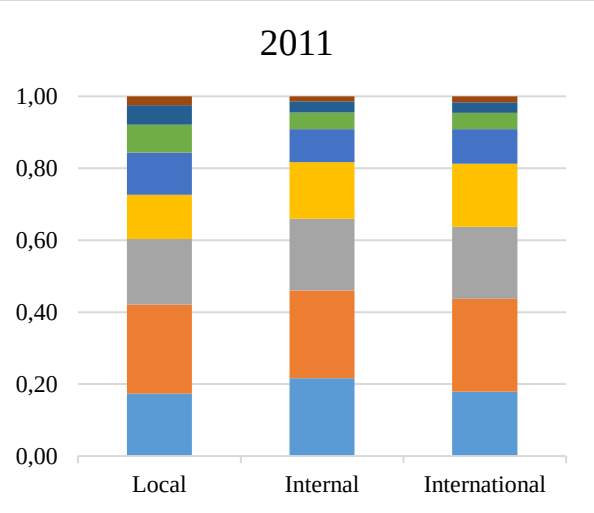


Figure 7: Household Income by Origin in 2013

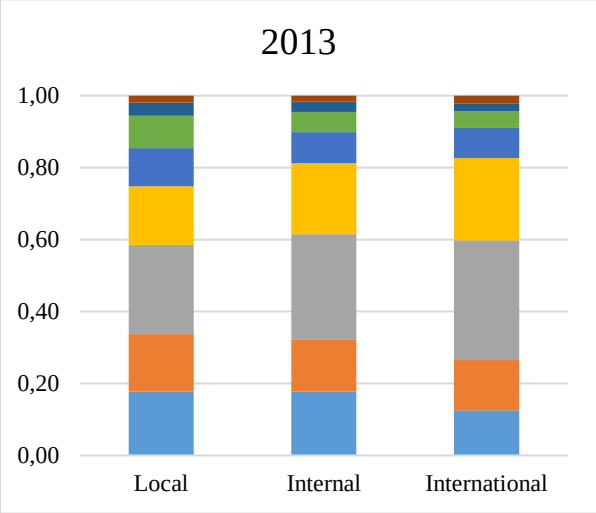
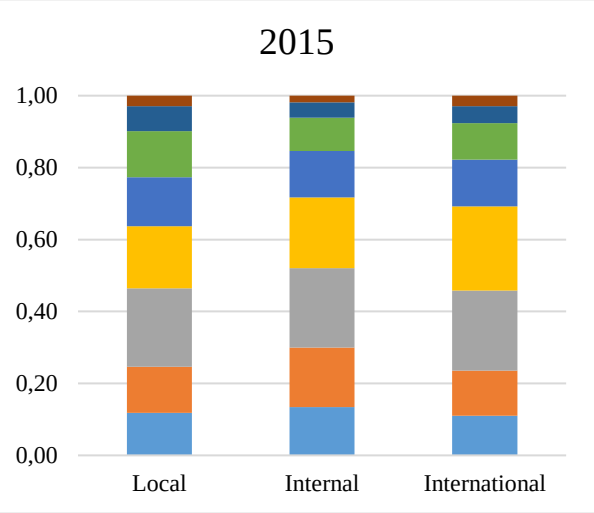


Figure 8: Household Income by Origin in 2015



R 0 - R 400	R401 - R 1 600	R 1 601 - R 3 200	R 3 201 -R 6 400
R 6 401 - R 12 800	R 12 801 - R 25 600	R 25 601 - R51 200	R 51 201 or more

Source: Author's own calculations from GCRO Quality of life Survey Data

Table 6 details the percentage of individuals who reported to have some form of debt. There is a clear life-cycle pattern across age groups. The highest percentage of individuals who reported to be in debt were aged between 28 and 57. Locals and internal migrants had similar percentages of individuals who were in debt while international migrants had a significantly lower percentage. This lower percentage may be because individuals who are not South African citizens have constrained access to credit from formal sources such as banks and micro-lenders. Most financial institutions in South Africa require South African identification or documentation from the Department of Home Affairs to accept applications for credit. As outlined in the Financial Intelligence Centre Act, these documents are required should an individual desire to open a bank account locally.

Table 6: Percentage who indicated "Yes" to the question, "Do you owe money to anyone including a bank or a shop or a money lender?"

	2009			2011			2013			2015		
	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.	Local	Internal	Intl.
Birth Year												
91-82	20.00	20.77	16.13	17.19	16.79	7.76	35.60	34.43	8.70	44.31	42.01	20.61
81-72	39.21	37.60	15.09	29.59	30.02	12.93	43.61	41.72	14.00	48.00	50.52	28.66
71-62	33.33	33.33	23.53	30.00	29.93	25.71	41.18	36.62	23.30	47.20	50.46	30.57
61-52	33.83	34.93	60.00	24.27	31.60	13.16	33.65	29.80	22.06	37.96	41.01	34.48
51-42	16.93	13.58		15.14	20.87	17.24	21.70	19.24	10.45	25.98	26.81	20.90

Source: Author's Own Calculations from GCRO Quality of Life Survey Data

7. RESULTS

The results of the OLS estimation are presented in Appendix A for comparison with the results of the system-GMM estimation in Table 7. The respective tests for autocorrelation in OLS estimation indicate that concerns about endogeneity are founded. Therefore, the system-GMM method is deemed the appropriate method of estimation.

7.1 System-GMM estimation

The influence of the variables of interest are largely as expected. Table 7 presents the results of the system GMM analysis on employment variables when Local is the base category. The regression results on overall employment confirm the results found in the descriptive analysis. International migrants have, on average, higher rates of overall employment, self-employment and informal employment than locals, results significant at the 1%, 10% and 1% level respectively. Internal migrants are found to have, on average, lower rates of self-employment than locals, a result significant at the 1% level. Furthermore, internal migrants have, on average, higher levels of formal and informal employment than locals, significant at the 1% and 10% level respectively. Appendix B1 presents the results when the base category is internal migrants. These results confirm that international migrants have, on average, higher rates of overall employment than internal migrants, significant at the 1% level. International migrants are found to have higher rates of self-employment and informal employment than internal migrants, significant at the 1% and 5% level respectively, and lower rates of formal employment, significant at the 5% level. Results in Appendix B1 confirm that locals have higher rates of self-employment than internal migrants, significant at the 1% level and lower levels of formal and informal employment, significant at the 1% and 10% level respectively.

The above findings are in line with Budlender (2014) and Fauvelle-Aymar (2014) who found that the employment rate of international migrants was higher than that of internal migrants and locals in South Africa. Furthermore, international migrants were found less likely to be employees and significantly more likely to be self-employed. Our study finds that the incidence of informal employment is higher for internal and international migrants than locals while Budlender (2014) and Fauvelle-Aymar (2014) found that international migrants were significantly more likely to be employed in the informal sector while there was no difference between locals and internal migrants. The current study's findings that international migrants

are more likely to be employed and self-employed points to the positive selection of international migrants to Gauteng.

Table 7: Results of System GMM Analysis on Employment Variables (Local omitted)

Variables	Employment	Self-Employment	Formal Employment	Informal Employment
Lag Y	-0.344* (0.193)	-0.449* (0.224)	-0.381* (0.209)	-0.406** (0.153)
Race: Black	0.684* (0.396)	0.560** (0.251)	-0.574*** (0.206)	0.059 (0.296)
Gender: Female	0.174** (0.064)	0.089 (0.069)	-0.179*** (0.050)	0.105 (0.079)
International	0.221*** (0.055)	0.072* (0.042)	0.035 (0.023)	0.127*** (0.034)
Internal	0.042 (0.033)	-0.063*** (0.021)	0.098*** (0.025)	0.048* (0.027)
Education	0.012 (0.030)	0.008 (0.024)	0.055*** (0.013)	-0.053** (0.023)
Age	0.009* (0.005)	-0.009 (0.005)	0.025*** (0.008)	0.0004 (0.005)
AgeSq	-0.000 (0.000)	0.0002* (0.000)	-0.0003*** (0.000)	-0.0000 (0.000)
Lag Log H. Income				-0.208*** (0.069)
Constant	-0.437 (0.689)	-0.429 (0.462)	0.031 (0.300)	1.045* (0.568)
Observations	54	54	54	54
No. of Instruments	21	21	21	25
Hansen (p-value)	0.114	0.254	0.190	0.239
Diff Hansen (p-value)	0.953	0.630	0.162	0.306
F-test (p-value)	0.000	0.000	0.000	0.000

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, International, Internal, Education. IV-style instruments: Race: Black, Gender: Female, Age, AgeSq.

We turn to Table 8 that presents the regression results of variables indicating overall well-being when Local is the base category and Appendix B4 when Internal Migrant is the base category. We consider both household income and per capita income to address the effect of smaller average migrant household size.

Table 8: Results of System GMM Analysis on Overall Well-Being Variables (Local omitted)

Variables	Log H. Income	Log PCI	Social Grant	Life Satisfaction	Hunger	Debt	Education
Lag Y	-0.330*** (0.057)	-0.038 (0.127)	0.366*** (0.106)	-0.232*** (0.075)	-0.930*** (0.234)	0.240* (0.126)	0.248* (0.123)
Race: Black	-0.509 (0.321)	-0.945 (0.808)	0.144 (0.287)	-0.233* (0.120)	0.126 (0.265)	0.234 (0.394)	-5.724*** (1.776)
Gender: Female	-0.177*** (0.055)	-0.843*** (0.302)	-0.068** (0.032)	0.064 (0.038)	-0.481*** (0.128)	-0.228*** (0.052)	1.005 (0.599)
International	0.060* (0.043)	0.251** (0.120)	-0.178*** (0.028)	0.031* (0.016)	0.017 (0.036)	-0.082* (0.041)	-0.557*** (0.139)
Internal	0.008 (0.036)	0.094 (0.103)	-0.024 (0.023)	-0.012 (0.011)	0.041* (0.021)	0.041* (0.022)	-0.157 (0.180)
Education	0.018 (0.033)	0.065 (0.077)	0.013 (0.030)	-0.005 (0.011)	0.017 (0.025)	0.075* (0.038)	
Age	0.030*** 0.006)	0.046** (0.019)	-0.021*** (0.005)	-0.005** (0.003)	-0.002 (0.005)	0.024*** (0.006)	-0.007 (0.034)
AgeSq	-0.0003*** (0.000)	-0.0006** (0.000)	0.0003*** (0.000)	0.0000** (0.000)	0.0000 (0.000)	-0.0002** (0.000)	-0.0009** (0.000)
Constant	1.559** (0.755)	7.844*** (2.142)	0.276 (0.630)	1.124*** (0.259)	0.362 (0.529)	-1.252 (0.868)	15.278*** (3.317)
Observations	54	54	54	54	54	54	54
No. of Instruments	17	21	17	21	17	17	13
Hansen (p-value)	0.455	0.229	0.153	0.416	0.106	0.186	0.172
Diff Hansen (p-value)	0.444	0.366	0.677	0.890	0.147	0.284	0.692
F-test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, International, Internal, Education. IV-style instruments: Race: Black, Gender: Female, Age, AgeSq.

Both household income and per capita household income was found to be higher, on average, for international migrant than locals, significant at the 10% and 5% level respectively. International migrants less likely to access some form of social grant or assistance. This result is significant at the 1% level when local and internal migrant are the respective base category. The results pertaining to Life Satisfaction indicate that international migrants experience higher levels of Life Satisfaction, on average, than both locally born individuals and internal migrants, significant at the 10% level respectively while there is no difference between locals and internal migrants.

Internal migrants are more likely to have a child or adult in the household go hungry in the recent period than locals and international migrants, significant at the 10% level respectively. International migrants are less likely to have some form of debt than both locals and internal migrants, significant at the 10% and 1% level respectively. Internal migrants are more likely to have some form of debt than both locals and international migrants, significant at the 10% and 1% level respectively. In summary, international migrants, on average, have higher total and per capita household income and are less likely to receive social assistance. They also exhibit higher life satisfaction and lower likelihood of having some form of debt.

To explain the success of migrants in the labour market, we consider years of education. We find that international migrants have lower years of education on average than locals, significant at the 1% level. Internal migrants are also found to have lower average years of education than locals, significant at the 5% level when internal migrant is the base category. These results indicate that the higher levels of employment and self-employment of international migrants is determined by factors other than education. If we assume that years of education are a measure of observable skills, we are required to turn to unobservable

characteristics to explain the difference in the rates of employment between international migrants and their South African-born counterparts.

As aforementioned in the literature, self-selection in migration and labour market outcomes are likely to depend not only on observable characteristics such as years of education but also on unobservable, innate characteristics. Even in the case of migrants and non-migrants who appear similar in observable characteristics such as age or education, there are potential differences in latent attributes that affect the migration decision (Nakosteen *et al*., 2008). The next section will make use of the Oaxaca-Blinder Decomposition method to investigate the extent to which these unobservable characteristics explain the labour market success of international migrants in Gauteng.

7.2 Oaxaca – Blinder Decomposition

The regression-based decomposition developed by Oaxaca (1973) and Blinder (1973) will be used to divide the differential in the outcomes of interest -employment and self-employment rates- between native born individuals and migrants into explained and unexplained variation. Oaxaca-Blinder method decomposes the gap in the means of an outcome variable between the groups of interest into an explained part that is due to group differences in the magnitudes of a set of measured predictor variables and an unexplained part that researchers use as measure of discrimination but it can also be understood to contain the effects of group differences in unobserved predictors (Jann, 2008). Discrimination in the labour market occurs when individuals with similar productive characteristics experience differential treatment. The Oaxaca-Blinder method is a useful analytical framework to investigate discrimination because it allows researchers to quantify the extent to which differences in observed productive characteristics account for differentials in labour market outcomes -including employment

rates and wages- and the extent to which these differentials are explained by other unobserved factors which may include discrimination.

For the present study, we consider the following employment determination model:

$$E_{c,t} = \alpha + \beta_t X_{c,t} + \varepsilon_{c,t} \quad (7)$$

where $E_{c,t}$ denotes the employment rate of cohort c observed in time t, $X_{c(t),t}$ denotes a vector of the cohort means of explanatory variables of cohort c observed at time t. These variables are years of education, age, age-squared, gender and race, $\beta_{c,t}$ denotes the regression coefficients and $\varepsilon_{c,t}$ denotes the error term. To account for the higher employment and self-employment rates of international migrants, separate employment determination equations are estimated for locals, internal migrants and international migrants respectively. Separate decomposition estimations are then conducted comparing international migrants to locals and internal migrants respectively. Decomposition estimations are also conducted comparing locals to internal migrants. We demonstrate the Oaxaca-Blinder decomposition as follows.

The gap can be viewed as the difference in the predicted means of the employment rates of the groups in question:

$$E^{FM}_{c,t} - E^{L}_{c,t} = \alpha^{FM} + \beta_t^{FM} X^{FM}_{c,t} - \alpha^{L} - \beta_t^{L} X^{L}_{c,t} - u_t \quad (8)$$

where $u_t = \varepsilon^{L}_{c,t} + \varepsilon^{FM}_{c,t}$. and where the superscripts refer to international migrants (FM) and locals (L). The resultant Oaxaca – Blinder decomposition equation is as follows:

$$E^{FM}_{c,t} - E^{L}_{c,t} = \beta^{FM} (X^{FM}_{c,t} - X^{L}_{c,t}) + (\alpha^{FM} - \alpha^{L}) + (\beta^{FM} - \beta^{L}) X^{L}_{c,t} \quad (9)$$

The first term on the right-hand side of equation (9) is interpreted as the explained component and the sum of the final two terms is the unexplained component (Elder, Goddeeris and Haider, 2010). The presence of this unexplained component points to the presence of unobserved

factors that determine the differential in the rate of employment and self-employment between the respective groups. Table 9 summarises the results of the Oaxaca-Blinder decomposition. More explanatory variables to control for differences in other observable characteristics between the three origin groups may alter the extent of discrimination and this factor calls for further research (Kollamparambil and Razak, 2016). However, the influence of unobservable characteristics may still be significant given that respondents were drawn from the same region if Gauteng, controlling for factors relating to location and economy. Furthermore, the analysis includes a lagged dependent variable to address issues of misspecification and missing variables. Further details on the respective regression results is detailed in Appendix D.

The unobserved factors explain 127.21% and 148.97% of the difference in employment rates between international migrants and locals and between international and internal migrants respectively. Differences in observed human capital characteristics are more important in accounting for the difference in the rate of employment of locals over internal migrants. Unobserved factors explain only 63.28% of this difference. This follows given that both locals and internal migrants are South African-born and therefore are not subject to differences in unobserved factors relating to country of origin. When interpreting the results of the respective decompositions, we note that the unexplained portion may be attributable to several factors other than self-selection. Lee (2012: p. 466) points out that locals and migrants may "receive different returns on their productive characteristics because the variables used to capture the productive characteristics are imperfect". For example, locals may receive a higher return to education than international migrants or rural migrants because the education obtained by locals is more relevant and valuable to the destination labour market. The unexplained portion may also capture the effect of employer bias toward international migrant workers; that is, there may be positive discrimination toward international migrant workers. Various factors may

Table 9: Oaxaca–Blinder Decomposition of Employment and Self-Employment

	Employment Rate			Self – Employment Rate		
	International and Local	International and Internal	Local and Internal	International and Local	International and Internal	Local and Internal
Explained Part	-0.085	-0.094	-0.044	0.028	-0.123	0.007
Unexplained Part	0.395	0.286	-0.075	1.004	1.153	0.778
Total Predicted Gap	0.310	0.192	-0.119	1.032	1.030	0.785
% Explained	-27.21	-48.97	36.72	2.74	-11.90	0.94
% Unexplained	127.21	148.97	63.28	97.26	111.90	99.06

Source: Author's own calculations from the GCRO QOL data.

explain this including how employers may perceive the cost of international migrant labour to be lower. Research indicates that the rate of union membership amongst international migrants is less than half that of South African-born employees (Budlender, 2014). International migrants also had a noticeably lower rate of access to employment benefits than South African-born employees (Budlender, 2014). Undocumented migrants also do not benefit from labour protection legislation and policies. We therefore turn to the decomposition of the self-employment rate given that employer discrimination is of less concern here. Unobserved factors account for 97.26% and 111.90% of the difference in the rate of self-employment between locals and international migrants and internal and international migrants respectively. Unobserved factors account for 99.06% of the difference in the rate of self-employment between locals and internal migrants. These results indicate that unobserved factors other than employer discrimination explain the difference in the self-employment rates of locals, internal migrants and international migrants and that international migrants may possess unobserved characteristics that determine their higher rates of employment and self-employment. The results of the Oaxaca-Blinder analysis provide further evidence of the positive self-selection of international migrants.

8. CONCLUSION

The present study sought to investigate the presence of self-selection among internal and international migrants in Gauteng by investigating the proposition by Borjas (1987) that positively selected migrants can be expected to outperform the native population in the country of destination while negatively selected migrants can be expected to not perform well in the country of destination's labour market. In addition, the present study sought to disentangle the effects of observed and unobserved characteristics in the self-selection of migrants by conducting Oaxaca-Blinder decomposition on overall employment and self-employment outcome variables. Preliminary descriptive statistics indicated that international migrants

experienced markedly higher levels of employment than both locals and internal migrants driven by higher rates of informal and self-employment. System GMM analysis of synthetic panel data confirmed these results and showed that international migrants had a higher likelihood of employment, self-employment and informal employment than locals and internal migrants. Internal migrants were more likely to be formally employed. International migrants were also found to outperform their South African-born counterparts on various variables indicative of well-being. International migrants experienced higher total household income and per capita household income. This group was also, on average, less likely to receive social assistance. They also exhibited higher life satisfaction, and a lower likelihood of having some form of debt. Oaxaca-Blinder decomposition provided further evidence of the positive selection of international migrants to Gauteng on unobservable characteristics.

The policy implications of this result highlight the positive effect on job creation and economic growth when migrant entrepreneurship is enabled and supported. There is also a need for the development of skills to facilitate and grow entrepreneurship amongst locals. The results of this study indicate that international migrant possess differentiable skills that position them for success in self-employment and informal employment. These skills are likely to be country specific since most international migrants come from SSA countries with larger informal sectors.

LIST OF REFERENCES

Akee, R. 2010. Who Leaves? Deciphering Immigrant Self-Selection from a Developing Country. Economic Development and Cultural Change, Vol. 58, No. 2 (January 2010), pp. 323-344

Antman, F. and Mckenzie, D. 2007. Earnings Mobility and Measurement Error: A Pseudo-Panel Approach. Economic Development and Cultural Change, Vol 56 (1), pp. 125-161

Bargain, O. and Kwenda, P. 2011. Earnings Structures, Informal Employment, and Self-Employment: New Evidence from Brazil, Mexico, And South Africa. Review of Income and Wealth Series 57, Special Issue

Batista, C. 2014. Return Migration, Self-Selection and Entrepreneurship in Mozambique. IZA Discussion Paper No. 8195

Bond, S. 2002. Dynamic Panel Data Models: A Guide to Microdata Methods and Practice. The Institute for Fiscal Studies Department of Economics, UCL working paper CWP09/02

Blinder, A. S. 1973. Wage Discrimination: Reduced Form and Structural Estimates. Journal of Human Resources, 18, 436-455.

Borjas, G. J., 1987. Self-Selection and the Earnings of Immigrants. The American Economic Review, Vol. 77, No. 4. (Sep., 1987), pp. 531-553.

Borjas, G. J. 1991. Immigration and Self-Selection. Online. (Available: http://www.nber.org/chapters/c6663)

Borjas, G. J., Bronars, S. G. and Trejo, S. J. 1992a. Assimilation and the Earnings of Young Internal Migrants. The Review of Economics and Statistics 74(1): 170-75.

- 1992b. Self-Selection and Internal Migration in the United States. Journal of Urban Economics 32:159-85.

Brucker, H. and Trubswetter, P. 2004. Do the Best Go West? An Analysis of the Self-Selection of Employed East-West Migrants in Germany. IZA Discussion Paper No. 986

Budlender, D. 2014. Migration and employment in South Africa: Statistical analysis of the migration module in the Quarterly Labour Force Survey, third quarter 2012. MiWORC Report No. 5. June 2014

Carciotto, S. and Mavura, M. 2016. The Evolution of Migration Policy in Post-Apartheid South Africa: Emerging Themes and New Challenges. Cape Town: The Scalabrini Institute for Human Mobility in Africa.

Cattaneo, C. 2007. The Self-Selection in the Migration Process: What Can We Learn? Liuc Papers n. 199, Serie Economia e Impresa, 52, febbraio 2007

Chiquiar and Hanson. 2002. International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. Working Paper 9242, National Bureau of Economic Research

Chiswick, B. 2000. Are Immigrants Favourably Self-Selected? An Economic Analysis. IZA Discussion Paper Series, No. 131

Collinson, M., Tollman, S., Kahn, K. and Clark, S. 2006. Highly prevalent circular migration: Households, mobility and economic status in rural South Africa. Conference on African Migration in Comparative Perspective, Johannesburg, South Africa, 4-7 June, 2003

Constant, A. 2002. Self-Selection, Earnings, and Out-Migration: A Longitudinal Study of Immigrants to Germany. IZA Discussion Paper No. 672

Cristina. A. D. 2007. Self-Selection and Migration in Argentina. Online (Available: http://www.aaep.org.ar/anales/works/works2007/cristina.pdf)

Deaton, A. 1985. Panel Data from Time Series of Cross-Sections. Journal of Econometrics. 30, pp. 109-126

Dostie, B. and Léger, P. T. 2006. Self-Selection in Migration and Returns to Unobservable Skills. IZA Discussion Paper No. 1942 Favourable Self-Selection and the Internal Migration of Young White Males in the United States. The Journal of Human Resources, Vol. 30, No. 3 (Summer, 1995), pp. 460-471

Elder, T., Goddeeris, J. and Haider, S. 2010. Unexplained Gaps and Oaxaca-Blinder Decompositions. Labour Economics 17, pp. 284-290

Fan, C., Sun, M. and Zheng, S. 2011. Migration and Split Households: A Comparison of Sole, Couple and Family Migrants in Beijing, China. Environmental and Planning A, 43 (9): 2164-85

Fatoki, O. and Patswawairi, T. 2012. The Motivations and Obstacles to Immigrant Entrepreneurship in South Africa. J Soc Sci, 32(2): 133-142

Fauvelle-Aymar, C. 2014. Migration and employment in South Africa: An econometric analysis of domestic and international migrants (QLFS (Q3) 2012). MiWORC Report No. 6. August 2014

Gabriel, P. and Schmitz, S. 1995. Favourable Self-Selection and the Internal Migration of Young White Males in the United States. The Journal of Human Resources, Vol. 30, No. 3 (Summer, 1995), pp. 460-471

Gauteng City Region Observatory. 2016. Quality of Life in the GCR: QOL Survey 2015 Launch Presentation. Online. Available: www.gcro.ac.za/media/redactor_files/2016.06.28_QoL_topline_for_launch.pdf&ved=0ahUK

EwjbrvTxjM3VAhUmC8AKHe4oBFoQFggjMAA&USG=AFQjCNGSEbC6ek9juoBvRDvksVqp_pWcjw

Heintz, J. and Posel, D. 2007. Revisiting segmentation and informal employment in the South Africa Labour Market. South African Journal of Economics, 76 (1): 26-44

Himaz, R.and Arurapne, H. 2012. Returns to Education in Sir Lanka: A Pseudo Panel Approach. Oxford University Department of Economics Discussion Paper 615

Human Development Report. 2009. Overcoming Barriers: Human Mobility and Development. New York: United Nations Development Programme

Jann, B. 2008. The Blinder-Oaxaca Decomposition for Linear Regression Models. The Stata Journal 8(4), pp, 453-479

Kalitanyi, V. 2007. Evaluation of Employment Creation by African Immigrant Entrepreneurs for Unemployed South Africans in Cape Town. (Online) Accessed: 21 March 2017: http://etd.uwc.ac.za/xmlui/bitstream/handle/11394/2398/Kalitanyi_MCOM_2007.pdf?sequence=1

Kalitanyi, V. and Visser, K. 2010. African Immigrants in South Africa: Job Takers or Job Creators? South African Journal of Economic Management Science 13 (4), pp. 376-390

Kaestner, R. and Malamud, O. 2011. Self-Selection and International Migration: New Evidence from Mexico. National Bureau of Economic Research

Kanaiaupuni, S. 2000. Reframing the Migration Question: An Analysis of Men, Women and Gender in Mexico. Social Forces 78 (4): 1311-1347

Kollamparambil, U. 2017. Labour Market Impact of Internal In-Migration: A District level analysis of South Africa. Economic Research Southern Africa Working Paper 667

Kollmparambil, U. and Razak, A. 2016. Trends in Gender Wage Gap and Discrimination in South Africa: A Comparative Analysis across Races. Indian Journal of Human Development 10 (1), pp. 1-15

Mbatha, C. and Roodt, J. 2014. Recent Internal Migration and Labour Market Outcomes: Exploring The 2008 And 2010 National Income Dynamics Study (NIDS) Panel Data in South Africa. SAJEMS NS 17 (5), pp. 6653-672

Mulcahy, K. and Kollamparambil, U. 2016. The Impact of Rural-Urban Migration on Subjective Well-Being in South Africa, The Journal of Development Studies

Nakosteen, R., Westerlund, O. and Zimmer, M. 1980. Migration and Income: The Question of Self-Selection. Southern Economic Journal 46 (3), pp. 840-851

Nakosteen, R., Westerlund, O. and Zimmer, M. 2008. Migration and Self-Selection: Measured Earnings and Latent Characteristics. Journal of Regional Science, Vol. 48, No. 4, 2008, pp. 769–788

Oaxaca, R. 1973. Male-Female Wage Differentials in Urban Labor Markets. International Economic Review, 14, 693-709.

Posel, D. 2004. Have migration patterns in post-apartheid South Africa changed? Journal of Interdisciplinary Economics, 15, 277-292

Roodman, D. 2006. How to do xtabond2: An introduction to difference and system GMM in Stata. The Stata Journal 9 (1), pp. 86-136

Roy. A. 1951. Some Thoughts on the Distribution of Earnings. Oxford Economic Papers New Series 3 (2), pp. 135-146

Russell, J. and Fraas, J. 2005. An Application of Panel Regression to Pseudo Panel Data. Multiple Linear Regression Viewpoints, Vol. 31 (1)

Schmidley, K. 2016. Panel Data: Fixed and Random Effects. University of Basel.

Segal, U., Elliot, D. and Mayadas, N. 2010. Immigration Worldwide: Policies, Practices and Trends. New York: Oxford University Press, Inc.

Sheppard, K. 2015. Generalised Method of Moments. Online. Available: https://www.kevinsheppard.com/images/5/55/Chapter6.pdf

Soto, M. 2009. System GMM Estimation with a small sample. Barcelona Economics Working Paper Series Working Paper 395

United Nations Human Rights Commission. 2013. Xenophobia. (Online) Accessed: 22 March 2017. Available: http://unicpretoria.org.za/files/2015/04/Xenophobia-final.pdf

Verbeek, M. 2008. Pseudo-Panels and Repeated Cross-Sections. In: L. M´aty´as, P. Sevestre (eds.), The Econometrics of Panel Data. Heidelberg: Springer-Verlag

Wentzel, M. and Tlabela, K. 2006. Historical background to South African migration. In: Kok, P., Gelderblom, D., Oucho, J. O. and Van Zyl, J. (eds). Migration in South and southern Africa: Dynamics and Determinants. Cape Town: HSRC. Pp 71-96.

Zsohar, P. 2012. Short Introduction to the Generalised Method of Moments. Hungarian Statistical Review, Special Number 16. (Online) Available: http://www.ksh.hu/statszemle_archive/2012/2012_K16/2012_K16_150.pdf

APPENDICES

Appendix A: Regression Analysis when Omitting Internal Migrant

Appendix A1: Results of OLS Analysis on Employment Variables (Local Omitted)

Variables	Employment	Self-Employment	Formal Employment	Informal Employment
Lag Y	-0.055 (0.144)	0.034 (0.137)	-0.171 (0.107)	-0.199 (0.147)
Race: Black	-0.036 (0.127)	0.109 (0.089)	-0.319 (0.120)	0.352** (0.169)
Gender: Female	0.227** (0.087)	0.059 (0.091)	-0.223 (0.086)	0.194** (0.072)
International	0.149*** (0.025)	0.037** (0.017)	0.012 (0.018)	0.140*** (0.025)
Internal	0.036* (0.019)	-0.036** (0.014)	0.054* (0.021)	0.046** (0.021)
Education	-0.023** (0.009)	-0.017*** (0.006)	0.043*** (0.007)	0.009 (0.008)
Age	0.007 (0.005)	-0.005 (0.003)	0.022*** (0.006)	-0.0003 (0.004)
AgeSq	-0.0000* (0.000)	0.0001* (0.000)	-0.0002*** (0.000)	0.0000 (0.000)
Lag H. Income				-0.192*** (0.055)
Constant	0.575 (0.259)	0.220* (0.121)	-0.084 (0.158)	-0.128 (0.260)
Observations	84	84	84	84
R-Squared	0.501	0.510	0.578	0.567
F (8, 29)	35.59***	23.42***	26.48***	18.69***
Wooldridge autocorrelation test	6.584**	16.981***	0.120	1.441

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis.

Appendix A2: Results of OLS Analysis on Overall Well-Being Variables (Internal Migrant Omitted)

Variables	Log H. Income	Log PCI	Social Grant	Life Satisfaction	H. Hunger	Debt	Education
Lag Y	-0.127** (0.059)	0.059 (0.089)	0.106 (0.120)	0.185*** (0.043)	-0.313 (0.216)	0.157* (0.089)	0.6611*** (0.096)
Race: Black	-1.052*** (0.171)	-1.264* (0.715)	0.703*** (0.129)	-0.051 (0.086)	0.097 (0.143)	0.338** (0.161)	-3.039* (1.510)
Gender: Female	-0.239*** (0.065)	-0.433** (0.206)	-0.175*** (0.041)	0.024 (0.042)	-0.566*** (0.090)	-0.212*** (0.030)	2.887*** (0.847)
International	-0.006 (0.023)	0.189*** (0.072)	-0.168*** (0.031)	0.014 (0.010)	0.012 (0.025)	-0.079*** (0.019)	-0.143 (0.098)
Internal	-0.005 (0.029)	0.094 (0.087)	-0.034 (0.027)	-0.017* (0.009)	0.037* (0.018)	0.033 (0.019)	-0.093 (0.119)
Education	-0.051*** (0.010)	0.079 (0.057)	0.072*** (0.008)	0.003 (0.007)	0.016 (0.012)	0.065*** (0.006)	
Age	0.027*** (0.005)	0.075*** (0.020)	-0.028*** (0.005)	-0.006*** (0.002)	-0.0008 (0.004)	0.027*** (0.004)	-0.009 (0.020)
AgeSq	-0.0004*** (0.000)	-0.0008*** (0.000)	0.0005*** (0.000)	0.0001*** (0.000)	-0.0000 (0.000)	-0.0002*** 0.000	-0.0005** (0.000)
Constant	2.824*** (0.292)	6.286*** (1.637)	-0.776*** (0.210)	0.645 *** (0.130)	0.269 (0.209)	-1.306*** (0.172)	5.800** (2.558)
Observations	84	84	84	84	84	84	84
R-Squared	0.644	0.429	0.910	0.719	0.541	0.822	0.830
F (8, 29)	17.75***	17.53***	78.25***	28.37***	15.12***	78.13***	283.96***
Wooldridge autocorrelation test	0.281	1.357	20.939***	29.327***	16.454***	10.651***	0.549

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis.

Appendix B: Regression Analysis when Omitting Internal Migrant

Appendix B1: Results of System GMM Analysis on Employment Variables (Internal Migrant Omitted)

Variables	Employment	Self - Employment	Formal Employment	Informal Employment
Lag Y	0.344* (0.193)	-0.447* (0.224)	-0.382* (0.209)	-0.397** (0.166)
Race: Black	0.683* (0.396)	0.558** (0.252)	-0.574*** (0.206)	0.013 (0.293)
Gender: Female	0.174*** (0.064)	0.089 (0.069)	-0.179*** (0.050)	0.105 (0.082)
Local	-0.422 (0.033)	0.063*** (0.021)	-0.098*** (0.025)	-0.047* (0.026)
International	0.179*** (0.042)	0.135*** (0.036)	-0.063** (0.024)	0.075** (0.032)
Education	0.012 (0.030)	0.007 (0.024)	0.055*** (0.013)	-0.059** (0.024)
Age	0.009* (0.005)	-0.009 (0.005)	0.025*** (0.008)	0.001 (0.004)
AgeSq	-0.0001 (0.000)	0.0002* (0.000)	-0.000*** (0.000)	-0.000 (0.000)
Lag H. Income				-0.220*** (0.076)
Constant	-0.393 (0.699)	-0.489 (0.461)	0.69 (0.307)	1.222* (0.607)
Observations	54	54	54	54
No. of Instruments	21	21	21	27
Hansen (p-value)	0.112	0.256	0.198	0.383
Diff Hansen (p-value)	0.957	0.631	0.214	0.519
F-test (p-value)	0.000	0.000	0.000	0.000

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, Local, International, Education. IV-style instruments: Race: Black, Gender: Female, Age, AgeSq.

Appendix B2: Results of OLS Analysis on Employment Variables (Internal Migrant Omitted)

Variables	Employment	Self-Employment	Formal Employment	Informal Employment
Lag Y	-0.055 (0.144)	0.034 (0.137)	-0.171 (0.142)	-0.199 (0.147)
Race: Black	-0.036 (0.127)	0.109 (0.089)	-0.319 (0.131)	0.352** (0.169)
Gender: Female	0.227** (0.087)	0.059 (0.091)	-0.223 (0.067)	0.194** (0.072)
Local	-0.036*** (0.019)	-0.036** (0.014)	-0.054** (0.021)	-0.046** (0.021)
International	0.113*** (0.023)	-0.074*** (0.016)	-0.041** (0.018)	0.094*** (0.025)
Education	-0.022** (0.009)	-0.017*** (0.006)	0.043*** (0.007)	0.009 (0.008)
Age	0.007* (0.005)	-0.005 (0.003)	0.022*** (0.006)	-0.0003 (0.004)
AgeSq	-0.0001** (0.000)	0.0001* (0.000)	-0.0002*** (0.000)	0.0000 (0.000)
Lag H. Income				-0.192*** (0.055)
Constant	0.611** (0.258)	0.183 (0.120)	-0.030 (0.155)	-0.082 (0.264)
Observations	84	84	84	84
R-Squared	0.501	0.510	0.578	0.567
F (8, 29)	35.59***	23.42***	26.48***	18.69***
Wooldridge autocorrelation test	6.584**	16.981***	0.120	1.441

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis.

Appendix B3: Results of System GMM Analysis on Overall Well-Being Variables (Internal Migrant Omitted)

Variables	Log H. Income	Log PCI	Social Grant	Life Satisfaction	H. Hunger	Debt	Education
Lag Y	-0.329***	-0.127	0.369***	0.031	-0.942***	0.242*	0.249*
	(0.057)	(0.115)	(0.106)	(0.091)	(0.232)	(0.127)	(0.124)
Race: Black	-0.511	-1.227	0.116	-0.351**	0.140	0.237	-5.722***
	(0.321)	(1.100)	(0.285)	(0.132)	(0.254)	(0.392)	(1.775)
Gender: Female	-0.178***	-0.787***	-0.068**	0.103**	-0.479***	-0.227***	1.010
	(0.055)	(0.286)	(0.032)	(0.045)	(0.129)	(0.052)	(0.603)
Local	-0.008	-0.109	0.025	0.027	-0.041*	-0.041*	0.157**
	(0.036)	(0.105)	(0.023)	(0.014)	(0.021)	(0.022)	(0.180)
International	0.052	0.151	-0.155***	0.009*	-0.023	-0.123***	-0.400
	(0.033)	(0.142)	(0.033)	(0.010)	(0.029)	(0.042)	(0.194)
Education	0.018	0.060	0.010	-0.023	0.018	0.075	
	(0.033)	(0.117)	(0.029)	(0.016)	(0.025)	(0.038)	
Age	0.030***	0.052***	-0.021***	-0.003	-0.003	0.023***	-0.008
	(0.006)	(0.019)	(0.005)	(0.002)	(0.005)	(0.006)	(0.034)
AgeSq	-0.0004***	-0.0007**	0.0003***	0.0000	-0.0001	-0.0002**	-0.0009**
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	0.000	(0.000)
Constant	1.574**	8.810***	0.326	1.239***	0.378	-1.217	15.119***
	(0.752)	(3.054)	(0.638)	(0.314)	(0.514)	(0.867)	(3.374)
Observations	54	54	54	54	54	54	54
No. of Instruments	17	17	17	17	17	17	13
Hansen (p-value)	0.445	0.215	0.116	0.182	0.103	0.184	0.147
Diff Hansen (p-value)	0.451	0.308	0.562	0.161	0.158	0.204	0.618
F-test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, Local, International, Education. IV-style instruments: Race: Black, Gender: Female, Age, AgeSq.

Appendix B4: Results of OLS Analysis on Overall Well-Being Variables (Internal Migrant Omitted)

Variables	Log H. Income	Log PCI	Social Grant	Life Satisfaction	H. Hunger	Debt	Education
Lag Y	-0.127** (0.059)	0.059 (0.089)	0.106 (0.120)	0.185*** (0.043)	-0.313 (0.216)	0.157* (0.089)	0.6611*** (0.096)
Race: Black	-1.052*** (0.171)	-1.264* (0.715)	0.703*** (0.129)	-0.051 (0.086)	0.097 (0.143)	0.338** (0.161)	-3.039* (1.510)
Gender: Female	-0.239*** (0.065)	-0.433** (0.206)	-0.175*** (0.041)	0.024 (0.042)	-0.566*** (0.090)	-0.212*** (0.030)	2.887*** (0.847)
Local	0.005 (0.029)	-0.094 (0.087)	0.034 (0.027)	0.170* (0.009)	-0.037* (0.018)	-0.033 (0.019)	0.093 (0.119)
International	-0.0007 (0.026)	0.095 (0.105)	-0134*** (0.034)	0.031*** (0.011)	-0.025 (0.028)	-0.112*** (0.020)	-0.050 (0.122)
Education	-0.051*** (0.010)	0.079 (0.057)	0.072*** (0.008)	0.003 (0.007)	0.016 (0.012)	0.065*** (0.006)	
Age	0.027*** (0.005)	0.075*** (0.020)	-0.028*** (0.005)	-0.006*** (0.002)	-0.0008 (0.004)	0.027*** (0.004)	-0.009 (0.020)
AgeSq	-0.0004*** (0.000)	-0.0008*** (0.000)	0.0005*** (0.000)	0.0001*** (0.000)	-0.0000 (0.000)	-0.0002*** 0.000	-0.0005** (0.000)
Constant	2.819*** (0.299)	6.380*** (1.661)	-0.810*** (0.223)	0.628 *** (0.134)	0.307 (0.210)	-1.306*** (0.172)	5.707** (2.600)
Observations	84	84	84	84	84	84	84
R-Squared	0.644	0.429	0.910	0.719	0.541	0.822	0.830
F (8, 29)	17.75***	17.53***	78.25***	28.37***	15.12***	78.13***	283.96***
Wooldridge autocorrelation test	0.281	1.357	20.939***	29.327***	16.454***	10.651***	0.549

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis.

Appendix C: Regression Analysis when Omitting International Migrant

Appendix C1: Results of System GMM Analysis on Employment Variables (International Migrant omitted)

Variables	Employment	Self-Employment	Formal Employment	Informal Employment
Lag Y	-0.344* (0.192)	-0.437* (0.218)	-0.381* (0.209)	-0.398** (0.166)
Race: Black	0.680* (0.397)	0.550** (0.248)	-0.574*** (0.205)	0.012 (0.294)
Gender: Female	0.174** (0.064)	0.089 (0.069)	-0.180 (0.050)	0.105 (0.082)
Local	-0.221*** (0.055)	-0.071* (0.041)	-0.034 (0.023)	-0.122*** (0.035)
Internal	-0.179*** (0.042)	-0.134*** (0.036)	0.063 (0.024)	-0.075** (0.032)
Education	0.012 (0.030)	0.006 (0.023)	0.055 (0.013)	-0.059** (0.024)
Age	0.009* (0.005)	-0.009 (0.005)	0.025 (0.008)	-0.001 (0.004)
AgeSq	-0.0000 (0.000	0.0002* (0.000)	-0.0003*** (0.000)	-0.0000 (0.000)
Lag H, Income				-0.221*** (0.077)
Constant	-0.208 (0.678)	-0.335 (0.426)	0.013 (0.295)	1.303** (0.598)
Observations	54	54	54	54
No. of Instruments	21	21	21	27
Hansen (p-value)	0.115	0.234	0.198	0.388
Diff Hansen (p-value)	0.960	0.558	0.253	0.527
F-test (p-value)	0.000	0.001	0.000	0.000

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, Local, Internal, Education. IV-style instruments: Race: Black, Gender: Female, Age, AgeSq.

Appendix C2: Results of OLS Analysis on Employment Variables (International Migrant Omitted)

Variables	Employment	Self-Employment	Formal Employment	Informal Employment
Lag Y	-0.055 (0.144)	0.034 (0.137)	-0.171 (0.142)	-0.199 (0.147)
Race: Black	-0.036 (0.127)	0.109 (0.089)	-0.319** (0.131)	0.352** (0.169)
Gender: Female	0.227** (0.087)	0.059 (0.091)	-0.223*** (0.067)	0.194** (0.072)
Local	-0.149*** (0.025)	-0.037** (0.17)	-0.012 (0.018)	-0.140*** (0.025)
Internal	-0.113*** (0.023)	-0.074*** (0.016)	0.041** (0.018)	-0.094*** (0.025)
Education	-0.022** (0.009)	-0.017*** (0.006)	0.043*** (0.007)	0.009 (0.008)
Age	0.007* (0.005)	-0.005 (0.003)	0.022*** (0.006)	-0.0003 (0.004)
AgeSq	-0.0001** (0.000)	0.0001* (0.000)	-0.0002*** (0.000)	0.0000 (0.000)
Lag H. Income				-0.192*** (0.055)
Constant	0.724*** (0.250)	0.257** (0.117)	-0.072 (0.152)	0.012 (0.263)
Observations	84	84	84	84
R-Squared	0.501	0.510	0.578	0.567
F (7, 9)	35.59***	23.42***	26.48***	18.69***
Wooldridge autocorrelation test	6.584**	16.981***	0.120	1.441

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis.

Appendix C3: Results of System GMM Analysis on Overall Well-Being Variables (International Migrant omitted)

Variables	Log H. Income	Log PCI	Social Grant	Life Satisfaction	H. Hunger	Debt	Education
Lag Y	-3.330*** (0.056)	-0.135 (0.115)	0.366*** (0.108)	0.034 (0.090)	-1.159*** (0.244)	0.273** (0.130)	0.248* (0.124)
Race: Black	-0.516 (0.321)	-1.315 (1.109)	0.101 (0.290)	-0.342** (0.136)	0.416 (0.262)	0.214 (0.385)	-5.723*** (1.778)
Gender: Female	-0.178*** (0.055)	-0.793*** (0.286)	-0.071** (0.034)	0.105** (0.045)	-0.448*** (0.137)	-0.220*** (0.051)	1.008 (0.617)
Local	-0.060 (0.043)	-0.254 (0.159)	0.182*** (0.028)	-0.019 (0.015)	0.019 (0.035)	0.080* (0.041)	0.400 (0.194)
Internal	-0.052 (0.033)	-0.145 (0.145)	0.157*** (0.033)	-0.028* (0.015)	-0.023 (0.040)	0.120*** (0.042)	0.557 (0.138)
Education	0.017* (0.033)	0.051 (0.116)	0.008 (0.030)	-0.022 (0.017)	0.046 (0.029)	0.071* (0.038)	
Age	0.030*** (0.006)	0.053*** (0.019)	-0.021*** (0.005)	-0.003 (0.003)	-0.002 (0.006)	0.022*** (0.005)	-0.008 (0.000)
AgeSq	-0.0004*** (0.000)	-0.0007*** (0.000)	0.0003*** (0.000)	0.0000 (0.000)	0.0001 (0.000)	-0.0002** (0.000)	-0.0009 (0.000)
Constant	1.639** (0.724)	9.207*** (2.966	0.211 (0.630)	1.241*** (0.319)	-0.250 0.583	-1.242 (0.822)	14.722 (3.267)
Observations	54	54	54	54	54	54	54
No. of Instruments	17	17	17	17	15	17	13
Hansen (p-value)	0.469	0.237	0.146	0.192	0.169	0.278	0.142
Diff Hansen (p-value)	0.383	0.120	0.380	0.170	0.430	0.141	0.610
F-test (p-value)	0.000	0.000	0.000	0.000	0.000	0.000	0.000

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, Local, Internal, Education. IV-style instruments: Race: Black, Gender: Female, Age, AgeSq.

Appendix C4: Results of OLS Analysis on Overall Well-Being Variables (International Migrant omitted)

Variables	Log H. Income	Log PCI	Social Grant	Life Satisfaction	H. Hunger	Debt	Education
Lag Y	-0.127** (0.059)	0.059 (0.089)	0.106 (0.120)	0.185*** (0.043)	-0.313 (0.216)	0.157* (0.089)	0.661*** (0.096)
Race: Black	-1.052*** (0.171)	-1.264* (0.715)	0.703*** (0.129)	-0.051 (0.086)	0.097 (0.143)	0.338** (0.161)	-3.039* (1.510)
Gender: Female	-0.239*** (0.065)	-0.433** (0.206)	-0.175*** (0.041)	0.024 (0.042)	-0.566*** (0.090)	-0.212*** (0.030)	2.887*** (0.847)
Local	0.006 (0.023)	-0.189** (0.072)	0.168*** (0.031)	-0.014 (0.010)	-0.012 (0.025)	0.079*** (0.019)	0.143 (0.098)
Internal	0.0007 (0.026)	-0.095 (0.105)	0.134*** (0.034)	-0.031*** (0.011)	0.025 (0.028)	0.112*** (0.020)	0.050 (0.121)
Education	-0.051*** (0.010)	0.079 (0.057)	0.072*** (0.008)	0.003 (0.007)	0.016 (0.012)	0.065*** (0.006)	
Age	0.027*** (0.005)	0.075*** (0.020)	-0.028*** (0.005)	-0.006*** (0.002)	-0.0008 (0.004)	0.027*** (0.004)	-0.009 (0.020)
AgeSq	-0.0004*** (0.000)	-0.0008*** (0.000)	0.0005*** (0.000)	0.0001*** (0.000)	-0.0000 (0.000)	-0.0002*** 0.000	-0.0005** (0.000)
Constant	2.818*** (0.293)	6.475*** (1.643)	-0.944*** (0.209)	0.659*** (0.126)	0.281 (0.199)	-1.386*** (0.171)	5.657** (2.559)
Observations	84	84	84	84	84	84	84
R-Squared	0.644	0.429	0.910	0.719	0.541	0.822	0.830
F (8, 29)	17.75***	17.53***	78.25***	28.37***	15.12***	78.13	283.96
Wooldridge autocorrelation test	0.281	1.357	20.939***	29.327***	16.454***	10.651***	0.549

Notes: *, ** and *** refer to a significance level of 10, 5 and 1%, respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM.

Appendix D: System GMM Analysis by Origin and Oaxaca – Blinder Decomposition

Appendix D1: Sample means, regression estimates and decomposition results for International Migrants and Locals on dependent variable, Employment Rate

Variables	Coefficient International	Coefficient Local	Mean International	Mean Local	% Contribution to Explained Gap
Lag Y	-0.888 (0.596)	0.880** (0.273)	0.700	0.555	152.31
Race: Black	1.974 (1.136)	2.820** (1.046)	0.790	0.794	-52.55
Race: White	2.222* (1.119)	3.516** (1.573)	0.169	0.149	9.20
Gender: Female	0.151 (0.126)	-0.584* (0.278)	0.496	0.492	-0.80
Education	-0.0427 (0.041)	0.026 (0.017)	10.070	10.348	-8.84
Age	0.016 (0.023)	-0.008 (0.006)	43.261	44.486	22.72
AgeSq	-0.0002 (0.000)	0.0001 (0.000_	2072.954	2184.590	22.04
Constant	0.637 (1.533)	-2.547** (1.052)			
Observations	24	30			
No. of Instruments	10	9			
Hansen (p-value)	0.115	0.109			
Diff Hansen (p-value)	0.619	0.109			
F-test (p-value)	0.001	0.000			
Total Explained Gap	-0.085				
Total Unexplained	0.395				
Total Predicted Gap	0.310				

Notes: *. ** and *** refer to a significance level of 10. 5 and 1%. respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, IV-style instruments: Race: Black, Race: White, Gender: Female, Age, AgeSq, Education.

Appendix D2: Sample means, regression estimates and decomposition results for International Migrants and Internal Migrants on dependent variable, Employment Rate

Variables	Coefficient International	Coefficient Internal	Mean International	Mean Internal	% Contribution to Explained Gap
Lag Y	-0.888 (0.596)	0.155 (0.618)	0.700	0.607	88.25
Race: Black	1.974 (1.136)	1.022 (1.769)	0.790	0.883	-199.22
Race: White	2.222* (1.119)	-0.011 (1.789)	0.169	0.084	194.99
Gender: Female	0.151 (0.126)	0.325*** (0.103)	0.496	0.503	1.07
Education	-0.0427 (0.041)	-0.035** (0.018)	10.070	9.570	14.31
Age	0.016 (0.023)	-0.004 (0.007)	43.261	44.546	21.45
AgeSq	-0.0002 (0.000)	0.0000 (0.000)	2072.954	2190.391	-20.86
Constant	0.637 (1.533)	-0.216 (1.899)			
Observations	24	30			
No. of Instruments	10	10			
Hansen (p-value)	0.115	0.180			
Diff Hansen (p-value)	0.619	0.180			
F-test (p-value)	0.001	0.001			
Total Explained Gap	-0.094				
Total Unexplained	0.286				
Total Predicted Gap	0.192				

Notes: *. ** and *** refer to a significance level of 10. 5 and 1%. respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, IV-style instruments: Race: Black, Race: White, Gender: Female, Age, AgeSq, Education.

Appendix D3: Sample means, regression estimates and decomposition results for Locals and internal migrants on dependent variable, Employment Rate

Variables	Coefficient Local	Coefficient Internal	Mean Local	Mean Internal	% Contribution to Explained Gap
Lag Y	0.880** (0.273)	0.155 (0.618)	0.555	0.607	104.10
Race: Black	2.820** (1.046)	1.022 (1.769)	0.794	0.883	-517.54
Race: White	3.516** (1.573)	-0.011 (1.789)	0.149	0.084	574.29
Gender: Female	-0.584* (0.278)	0.325*** (0.103)	0.492	0.503	-14.90
Education	0.026 (0.017)	-0.035** (0.018)	10.348	9.570	-45.97
Age	-0.008 (0.006)	-0.004 (0.007)	44.486	44.546	-1.17
AgeSq	0.0001 (0.000_	0.0000 (0.000)	2184.590	2190.391	1.20
Constant	-2.547** (1.052)	-0.216 (1.899)			
Observations	30	30			
No. of Instruments	9	10			
Hansen (p-value)	0.109	0.180			
Diff Hansen (p-value)	0.109	0.180			
F-test (p-value)	0.000	0.001			
Total Explained Gap	-0.044				
Total Unexplained	-0.075				
Total Predicted Gap	-0.119				

Notes: *. ** and *** refer to a significance level of 10. 5 and 1%. respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, IV-style instruments: Race: Black, Race: White, Gender: Female, Age, AgeSq, Education.

Appendix D4: Sample means, regression estimates and decomposition results for International Migrants and Locals on dependent variable, Self-Employment Rate

Variables	Coefficient International	Coefficient Local	Mean International	Mean Local	% Contribution to Explained Gap
Lag Y	0.026 (0.206)	0.492*** (0.038)	0.156	0.113	3.92
Race: Black	1.105 (1.268)	0.672* (0.353)	0.790	0.794	78.17
Race: White	2.612** (1.380)	0.794* (0.390)	0.169	0.149	-36.42
Gender: Female	-0.122 (0.081)	-0.085 (0.119)	0.496	0.492	-1.94
Education	0.010 (0.031)	-0.002 (0.005)	10.070	10.348	-9.65
Age	0.042** (0.018)	-0.002 (0.002)	43.261	44.486	-182.90
AgeSq	-0.0006** (0.000)	0.0000* (0.000)	2072.954	2184.590	248.82
Constant	-1.672 (1.536)	-0.516 (0.363)			
Observations	24	30			
No. of Instruments	10	9			
Hansen (p-value)	0.238	0.338			
Diff Hansen (p-value)	0.951	0.338			
F-test (p-value)	0.005	0.000			
Total Explained Gap	0.028				
Total Unexplained	1.004				
Total Predicted Gap	1.032				

Notes: *. ** and *** refer to a significance level of 10. 5 and 1%. respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, IV-style instruments: Race: Black, Race: White, Gender: Female, Age, AgeSq, Education.

Appendix D5: Sample means, regression estimates and decomposition results for International Migrants and Internal Migrants on dependent variable, Self-Employment Rate

Variables	Coefficient International	Coefficient Internal	Mean International	Mean Internal	% Contribution to Explained Gap
Lag Y	0.026 (0.206)	-0.074 (0.113)	0.156	0.101	-1.164
Race: Black	1.105 (1.268)	1.263 (0.912)	0.790	0.883	-75.91
Race: White	2.612** (1.380)	1.513* (0.752)	0.169	0.149	197.85
Gender: Female	-0.122 (0.081)	-0.117 (0.132)	0.496	0.503	-0.67
Education	0.010 (0.031)	-0.007 (0.005)	10.070	9.570	-4.00
Age	0.042** (0.018)	-0.008 (0.007)	43.261	44.546	44.23
AgeSq	-0.0006** (0.000)	0.0001 (0.000)	2072.954	2190.391	-60.34
Constant	-1.672 (1.536)	-0.906 (0.279)			
Observations	24	30			
No. of Instruments	10	11			
Hansen (p-value)	0.238	0.627			
Diff Hansen (p-value)	0.951	0.850			
F-test (p-value)	0.005	0.000			
Total Explained Gap	-0.123				
Total Unexplained	1.153				
Total Predicted Gap	1.030				

Notes: *. ** and *** refer to a significance level of 10. 5 and 1%. respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, IV-style instruments: Race: Black, Race: White, Gender: Female, Age, AgeSq, Education.

Appendix D6: Sample means, regression estimates and decomposition results for Locals and Internal Migrants on dependent variable, Self-Employment Rate

Variables	Coefficient Local	Coefficient Internal	Mean Local	Mean Internal	% Contribution to Explained Gap
Lag Y	0.492*** (0.038)	-0.074 (0.113)	0.113	0.101	-82.08
Race: Black	0.672* (0.353)	1.263 (0.912)	0.794	0.883	-689.60
Race: White	0.794* (0.390)	1.513* (0.752)	0.149	0.149	808.53
Gender: Female	-0.085 (0.119)	-0.117 (0.132)	0.492	0.503	-12.81
Education	-0.002 (0.005)	-0.007 (0.005)	10.348	9.570	25.49
Age	-0.002 (0.002)	-0.008 (0.007)	44.486	44.546	-1.63
AgeSq	0.0000* (0.000)	0.0001 (0.000)	2184.590	2190.391	52.10
Constant	-0.516 (0.363)	-0.906 (0.279)			
Observations	30	30			
No. of Instruments	9	11			
Hansen (p-value)	0.338	0.627			
Diff Hansen (p-value)	0.338	0.850			
F-test (p-value)	0.000	0.000			
Total Explained Gap	0.007				
Total Unexplained	0.778				
Total Predicted Gap	0.785				

Notes: *. ** and *** refer to a significance level of 10. 5 and 1%. respectively. Standard errors in parenthesis. All regressions are small sample one-step system GMM. GMM style instruments: The relevant lagged variable, IV-style instruments: Race: Black, Race: White, Gender: Female, Age, AgeSq, Education.