# Impact of internal in-migration on income inequality in receiving areas: A district level study of South Africa

Umakrishnan Kollamparambil

**ERSA working paper 646**

**October 2016**

Economic Research Southern Africa (ERSA) is a research programme funded by the National Treasury of South Africa.

# Impact of internal in-migration on income inequality in receiving areas: A district level study of South Africa

Umakrishnan Kollamparambil *

October 25, 2016

**Abstract**

The impact of internal migration on regional income inequality of the receiving areas has hitherto gone largely unstudied. This dearth of literature is especially surprising because income inequality and in-migration into urban centres of growth are two issues that many developing economies are faced with and tackling these issues effectively involves understanding the interactions between these two related phenomena. This study is therefore a first attempt to analyse the impact of internal in-migration on receiving areas and is placed in the context of South Africa. Based on a conceptual analysis the study argues that In-migration into formal sector of the receiving areas will in general reduce inequality while in-migration into informal or unemployed sector increases inequality. Using individual panel data the study further tests empirically at the district level the impact of in-migration and finds that rising urban inequality in the urban areas can be attributed at least in part to rural-urban migration. This works through both the wage as well as employment channel. The employment channel can be said to have a stronger impact than the wage channel as indicated by the coefficients estimated through our system GMM regression analysis.

Keywords: internal migration, In-migration, Income inequality

JEL codes: O15, R23

## 1 Introduction

Although the Human Development Report of 2009 (UNDP 2009) highlighted the need to further explore the implications of internal and cross-border in-migration on the labour market of the destination areas, the impact of internal migration on regional income inequality of the receiving areas have largely gone unstudied. The existing studies analyzing the relationship between internal migration and

*Associate Professor in Economics, School of Economic and Business Sciences, University of Witwatersrand, Johannesburg, South Africa. Email: uma.kollamparambil@wits.ac.za

inequality have focused predominantly on either rural-urban inequality or to a more limited extend the impact on the income inequality levels within the sending areas (Lipton 1980, Ha Wei et al 2009). Studies that analyse the impact of internal migration on intra-regional inequality within the sending areas draw attention to the differences in the impact of internal and international migration on inequality in sending community. Tanja Bastia (2013) provides an overview of such studies and their contentious findings some of which argue that while internal migration mitigates inequality international migration perpetuates and increase inequality through remittances.

While the impact of cross-border in-migration on the receiving country labor market outcomes and income inequality levels has been analysed to a limited extend (Black et al 2005, Card 2009, Raphael & Smolensky 2008, Sanderson 2013), studies on the impact of internal migration on the receiving areas is surprisingly missing (Lall et al 2006). This dearth of literature is especially surprising because income inequality and in-migration into urban centres of growth are two issues that many developing economies are faced with and tackling these issues effectively involves understanding the interactions between these two related phenomena. The link between the two on intra-regional inequality however is yet to be fully explored empirically.

An analysis that looks deeper into the role of migration in the context of rising overall inequality is also warranted by studies that indicate that the offsetting trends in inequality (OTI) hypothesis, which claims that, underlying the overall inequality trend of falling rural–urban inequality there has been a tendency for rising intra-sectoral inequality (Eastwood & Lipton 2000) . Incorporating the migration dimension is expected to provide a further explanation for the OTI hypothesis.

This study is therefore a first attempt to analyse the impact of internal in-migration on receiving areas and is placed in the context of South Africa. The issue is relevant to no other country as much as to South Africa, which has one of the highest income inequalities in the world attributable to historical factors such as the discriminatory policies of the Apartheid regime (1948-1994) against women and non-whites. Despite the democratic government of South Africa instituting a series of legislations meant to ensure wage equality across race groups and genders, income inequality has been increasing post-democracy from a gini coefficient of 0.59 in 1993 to 0.65 in 2011 (World Bank 2013).

The issue of migration is also of specific interest in the South African context. Under the Apartheid regime, the movement of the vast majority of the population was restricted through the oppressive Group Areas Act and Influx control policies (Zuma, 2013). The elimination of these policies meant that in the Post-Apartheid years the country experienced accelerated urban migration (Mulcahy and Kollamparambil 2016). The intersection of inequality with migration arises from the finding of Leibbrandt, Finn and Woolard (2010) that while urban inequality has increased since 1993, rural inequality seems to have fallen. This points to the possibility of rural-urban migration playing a role in the emerging trends in inequality. Exploring the role of in-migration in explaining this phenomenon of rising urban inequality is hence of high relevance in the context of South Africa.

Furthermore, Mulcahy and Kollamparambil (2016) show that although internal migration has increased substantially the income levels of migrants in South Africa, their subjective well-being markedly declined post migration. The study attributes this partly to the increased income inequality in urban areas as compared to the rural areas to be one explanation for declining subjective well-being despite the improved income levels. This analysis is therefore an extension of the Mulcahy and Kollamparambil (2016) in that its objective is to analyse the inequality implications of internal migration within receiving areas in South Africa.

The study which is undertaken at the district level makes use of the National Income Dynamics Survey data to map the movement of individuals between districts over the period 2008-2015. The findings of the paper clearly indicate regional concentration in the direction of in-migration. Results of the multivariate analysis are unambiguously pointing to the positive contribution of in-migration to the rising overall inequality of receiving districts over the period of the study.

Rest of the paper is structured as follows: Section two undertakes a brief review of literature followed by a conceptual framework to place the empirical analysis in section 3. Data description and preliminary statistical analysis are undertaken in section 4. This is followed by a discussion on the methodology and the estimation models in Section 5. Section 6 discusses the empirical results followed by conclusions in the final section.

# 2 Review of Literature

While studies linking inequality with in-migration in South Africa does not exist to the best of the researcher's knowledge, a separate review of the extensive literature on inequality and the more limited studies on migration in South Africa can provide us with a context to our present study.

## 2.1 Inequality in South Africa

Income inequality in South Africa is predominately analysed along race lines based on the historical discrimination against the non-whites under the Apartheid regime. This continues to manifest in the labour market developments which is identified as playing a major role in the high inequality situation of the democratic South Africa. The labour market effect on inequality can be decomposed as employment effect and wage effect (ILO 2015). The high levels of unemployment rate in South Africa is well documented with narrow unemployment rate being higher than 25% and the level of unemployment among youth being over 50% (Zaakirah and Kollamparambil 2015). Employment effect is therefore uncontentiously a leading factor driving inequality in South Africa. These figures from the labour market statistics translates into interesting household-level analysis. Leibbrandt, M., Wegner, E., Finn, A. (2011) find that labour force participation rates are highest in the top income deciles, which also have the highest labour absorption rates. The study points out that although unemployment rates have fallen in the top deciles since 1993, the overall unemployment rates have trended upwardly since 1993 due to sharply rising unemployment rates of those in the bottom four deciles. This means that unemployment rates are higher among those located in the bottom income deciles, thus perpetuating inequality.

Unemployment however is not the sole cause of income inequality. Wage inequality also contributes substantially to inequality. Thus, being employed is not sufficient to eliminate inequality as a household is likely to continue in the lower deciles of the household income distribution with an earner with unskilled employment. Leibbrandt, Finn and Woolard (2010) decomposes income inequality and show that while labour market income accounted for 83 per cent of income inequality in 1993, this increased to 85 per cent in 2008. While these decompositions confirm the importance of rising unemployment as a key driver of inequality, they also emphasise the importance of the rising inequality of earnings for those households with access to labour market earnings (Leibbrandt, Woolard, McEwen, and Koep, 2010).

ILO (2015) point out that the overall level of total inequality is determined to a large extend by the wage gaps between different groups of workers like men-women, non-migrant -migrants, and workers in the formal-informal economy etc. It is therefore relevant to bring in the question of gender wage gap into our analyisis (ILO 2015, Gustafsson and Li 2000). While the studies on gender wage gap in the context of South Africa confirm the existence of gender wage gap, they lack consensus on the level and trend of it in the country. Muller (2009) found that among full-time workers the gender wage gap has narrowed between 1995 and 2006) while Ntuli (2007) found that the gap had increased between 1995 and 2004. Furthermore (Kollamparambil and Razak 2011) found the wage gap due to gender based discrimination to have declined in the period 2001-2007. This shows the changing dynamics of inequality in South Africa. These contributing factors have to be controlled for in our analysis while assessing the role of regional income inequality in the context of in-migration.

2.2 Internal Migration in South Africa

Analysing the post-Apartheid trends in internal migration, Posel (2004, 2010) highlight the increase in temporary labour migration and also increased feminization of migrants. Moreover according to Posel (2010), the vast majority of internal migrants originates from the rural areas. Zuberi & Sibanda (2004) find that 59 per cent of the 20–55 year old urban males in South Africa in 1996 were internal migrants. A more recent survey by the World Bank revealed that the proportion of people living in urban areas increased from 52% in 1990 to 62% in 2011 (SAIRR, 2013). The evidence therefore points to the bulk of the migration being from rural to urban areas.

Posel (2010) and Zuma (2013) associate the rural to urban movement of workers within the South African economy with the pursuit of better employment options. Studies done in the South African context generally conclude that migrants improve their incomes post migration. Cornwell and Inder (2004) as well a more recent study by (Mbatha &Roodt 2014) find that rural-urban migrants in South Africa are more likely to find formal employment than their non-migrant job seeking counterparts. Rogan et al., (2009) reports a positive correlation between migration and material well-being, measured either as the financial poverty line, an asset index, or household living standards. Mulcahy (2015) found that the increase in the real incomes of migrants was 52 per cent over the period 2008–2012, while that of non-migrants was only 19 per cent over the same period. However Mulcahy and Kollampambarabil (2016) show that this is accompanied by a decrease in the subjective well-being of migrants.

Studies have also investigated the impact of migration on the rural sending community. In one of the

earliest studies on remittances, Lucas (1987) finds that emigration to South African mines leads to an

increase in domestic plantation wages, an increase in long run cattle accumulation and crop productivity due to remittances and a decrease in short run crop production. Studying cross-border migrants, Lucas (2013) finds that through the mechanism of remittances, economic gains experienced by the South African emigrant are redirected to the rural sending countries which allows for potential decreases in poverty and inequality, and the development of rural areas.

# 3 Conceptual framework

Empirical literature is weak on the impact of migration on the inequality in receiving areas. This is despite the fact that there is considerable debate in theoretical literature surrounding the impact of migration on the overall inequality as well as inequality within receiving areas.

Kuznets (1955) illustrates an inverted U shaped relationship between inequality and stage of development. In early stages of development inequality is low but as growth is centered in urban the rural-urban wage gap increases and triggers migration from rural to urban areas. This results initially in increased wage difference between urban and rural incomes increasing inequality. Subsequently with sustained migration, rural wages also increase reducing the income inequality between regions. Another observation he makes in his seminal paper is that ".. inequality in the percentage shares within the distribution for the rural population is somewhat narrower than in that for the urban population-even when based on annual income; and this difference would probably be wider for distributions by secular income levels." The higher levels of inequalities alluded to by Kunets is not directly linked to migration by him, however Harris Todaro (1970) in their equally groundbreaking paper linked migration with urban inequality by showcasing that even with urban unemployment levels migration will continue further increasing urban income inequality. It is therefore surprising that none of the many extensions of the Harris-Todaro mode that icorporates a welfare anaylsis does not explicitly take into account the impact of in-migration on inequality (Temple 2005).

In this paper we build on the Harris-Todaro framework to explain the increasing income inequality within the urban areas as a result of migration. While migration occurs in the expectation of attaining high wages existing in the urban formal sector, a proportion of migrants fail to do so and are relegated to low wages of the urban informal sector. According to HT model the urban informal sector wage is even lower than the existing rural wages and consequently leads to rising urban inequality.

We construct a three sector model of formal, informal and unemployed sectors within the migrant receiving areas. The simplifying assumption made is that inequality in the community stems exclusively from the inter-sector income differences and that employment between sectors are non-substitutable as the levels of education and skills required are different between sectors. Inequality level in the initial period is a function of the distribution of the total population between the three sectors. Wage levels of the formal sector ($W^f$ )is higher than that of the informal sector ($W^i$ ) which in turn is higher than the zero unemployment wage ($W^u$). P is the labour force population.

$$I_0 = f(P_0^f, P_0^i, P_0^u)$$
$$W^f > W^i > W^u$$

The inequality level at the initial time period ($I^0$) can be altered by in-migration (N) that changes the distribution of labour force population between the three sectors in the next time period ($I^1$).

$$I_1 = f(N^f + P_0^f, N^i + P_0^i, N^u + P_0^u)$$
$$I_1 - I_0 = \Delta I = f(N^f, N^i, N^u)$$

The impact of migration on inequality is constructed in the four scenarios as follows:

Scenario 1:

According to the population rule of inequality measurement, cloning of the entire population will retain the inequality levels (Ray1998). In other words, although in-migration results in increased population, the inequality level will be unaltered provided each of the three sectors retain their initial proportions.

$$I_1 = f(N^f + P_0^f, N^i + P_0^i, N^u + P_0^u)$$
$$\frac{N_1^u + P_0^u}{P_1} = \frac{P_0^u}{P_0}$$
$$\frac{N_1^f + P_0^f}{P_1} = \frac{P_0^f}{P_0}$$
$$\frac{N_1^i + P_0^i}{P_1} = \frac{P_0^i}{P_0}$$
$$I_1 = I_0 = \Delta I = 0$$

Scenario 2:

Assuming that migration results in an increase in unemployment rate ($P^u$ /P), income inequality in period 2 will be higher as a result of a change in the population distribution.

$$\frac{N_2^u + P_0^u}{P_2} > \frac{P_0^u}{P_0}$$

$$\frac{N_2^u+P_0^u}{P_2} - (\frac{N_2^f+P_0^f}{P_2} + \frac{N_2^i+P_0^i}{P_2}) > \frac{P_0^u}{P_0} - (\frac{P_0^f}{P_0} + \frac{P_0^i}{P_0})$$
$$I_2 > I_0$$

Scenario 3:

Holding unemployment rate constant between time periods, an increase in the relative proportion of skilled migrants into the formal sector vis-à-vis unskilled migrants into the informal sector will result in a reduction in inequality as the population distribution shifts towards the higher wages of the formal sector.

$$\frac{N_3^u+P_0^u}{P_3} - (\frac{N_3^f+P_0^f}{P_3} + \frac{N_3^i+P_0^i}{P_3}) = \frac{P_0^u}{P_0} - (\frac{P_0^u}{P_0} + \frac{P_0^f}{P_{01}})$$
$$\frac{N_3^f+P_0^f}{P_3} - \frac{N_3^i+P_0^i}{P_3} > \frac{P_0^f}{P_0} - \frac{P_0^i}{P_0}$$
$$I_3 < I_0$$

The same outcome is expected with an increase in the proportion of the informal sector vis-à-vis the formal sector as a result of migration, holding unemployment rate constant. The average income however is likely to be lower in this case as compared to the first.

Scenario 4:

Assuming constant unemployment rate as in scenario 3, and introducing wage flexibility in the informal sector whereby wage levels respond to market forces of demand and supply of labour within the sector is considered next. We retain the assumption of the Harrod Domar model that the formal sector wages remain inflexible and is not subject to the market forces because it is determined by the labour legislations, organized labor unions as well as the desire of firms to retain credible threat of firing in order to elicit maximum productivity from its employees. A net in-migration into the formal sector ($N^f$) and informal sector ($N^i$) therefore will have very different impact on the income distributions even if the proportion of the formal and informal sectors are retained. A net in-migration into the informal sector will reduce the post migration informal sector ($W_4^i$) wages, while the formal sector wages ($W_4^f$) remain unchanged as a result of in-migration into formal sector. Outcome of scenario 1 changes even if the in-migration results in a cloning of population across the three sectors. The influx of informal sector in-migration will therefore lead to increased inequality within the receiving community.

$$\frac{N_4^u+P_0^u}{P_4} = \frac{P_0^u}{P_0}$$
$$\frac{N_4^f+P_0^f}{P_4} = \frac{P_0^f}{P_0}$$
$$\frac{N_4^i+P_0^i}{P_4} = \frac{P_0^i}{P_0}$$
$$N_4^f > 0 => W_4^f = W_0^f$$
$$N_4^i > 0 => W_4^f < W_0^f$$
$$I_4 > I_0$$

Scenario 5:

Changing scenario 4 to accommodate a change in the distribution of the population across the formal and informal sectors, it can be shown that an increase in the proportion of informal sector vis-à-vis formal sector post-migration further exacerbates the inequality situation. Inequality will further increase within the receiving areas due to the compounded population and income distributions effects.

$$\frac{N_5^i+P^0}{P_5} > \frac{N^i}{P_0}$$
$$\frac{N_5^f+P^0}{P_5} < \frac{N^f}{P_0}$$
$$\frac{N_5^u+P^0}{P_5} = \frac{N^u}{P_0}$$
$$I_5 > I_4 > I_0$$

Scenario 6:

Lastly, relaxing the constraint of constant unemployment rate, we see that the inequality outcomes of in-migration in the context of informal wage flexibility are less definite. In-migration leading to reduced unemployment rate and increased share of the informal sector vis-a-vis the formal sector can result in reduced inequality or increased inequality. The formal is likely with a marginal reduction in the informal sector wages and the latter with a substantial wage reduction.

$$\frac{N_6^i+P^0}{P_6} > \frac{N^i}{P_0}$$
$$\frac{N_6^f+P^0}{P_6} < \frac{N^f}{P_0}$$
$$\frac{N_6^u+P^0}{P_6} < \frac{N^u}{P_0}$$
$$I_6 >=< I_0$$

From a conceptual analysis it is clear that the labour market outcomes of in-migration is critical in determining the impact of in-migration on income inequality within the receiving community. A numeric illustration of the scenarios discussed above is provided in Appendix (2).

# 4 Data and Descriptive Statistics

Before undertaking a multivariate econometric analysis based on the theoretical postulations made in the previous section, we look at the data used in the analysis in this section.

## 4.1 Data

This paper uses the National Income Dynamics Survey (NIDS) Wave 1, Wave 3 and Wave 4 datasets spread over the years 2008, 2012 and 2014 respectively. Wave 2 is not used due to high attrition rate reported in that Wave in relation to the other waves (National Planning Commission, 2013). NIDS was conducted by the South African Labour and Development Unit (SALDRU) at the University of Cape Town, in conjunction with the National Treasury. Individuals between the ages of 15 and 60 years in the year 2008 are included in the study. A migrant is defined as an individual whose district of location changed between Wave 3 (2012) and Wave 4 (2014), and who had not moved in the 10 years prior to 2012. This definition of a migrant has been informed by the South African migration context. For example, by considering a migrant to be an individual who is above 18 years of age at the time of migration, we select those who undertook migration out of their own free-will. Collinson et al (2003) have reported high prevalence of circular migration in the South African labour market and therefore by further stipulating that stayers should not have moved within the last 10 years, we make sure to exclude circular migration is avoided which otherwise may have biased our results. As the study seeks to understand the impact of in-migration as a whole, no attempt is made to differentiate between temporary labour and permanent migration.

NIDS follows a survey structure applied four times with two year intervals. NIDS follws a stratified, two-stage cluster sample design where the explicit strata in the Master Sample are the 53 district councils (Leibbrandt, Woolard, de Villiers, 2009). Hence we believe district level analysis using the sample weights to correct for non-response, attrition and household changes will yield a representative sample.

Table 1 present the sample of the study comprising of individuals within the age group of 15-59 years in 2008 who were matched between Waves 1, 3 and 4 and fell into either the stayer or migrant group. From this, we dropped all individuals with missing observations under any of the required data variable categories. The eligible sample, which is used in our analysis, reflects a complete and balanced panel of observations. As the panel weighting takes into account attrition bias and survey design bias, we are confident that the initial sample is nationally representative. Although a sample bias can arise from dropping those with missing information if they follow a systematic pattern, we have no reason to suspect that this is not random. Table 1 shows that we have a large enough sample of stayers and migrants with the latter comprising 5.22 per cent of the total sample of individuals to conclude statistically significant results. These individuals were then mapped to their location of residence at the district level. The nine provinces in South Africa are divided into metropolitan and district municipalities. The largest metropolitan areas are governed by 8 metropolitan municipalities, while the rest of the country is divided into 44 district municipalities. For the purpose of the study we collectively call them as districts. South Africa has 52 districts of which 51 are included in the sample. The district uMzinyathi district of Kwa-Zulu-Natal (KZN) province was excluded from the sample as zero population was recorded in our 2008 sample. The distribution of districts in our sample across the nine provinces of South Africa are presented in Table 2.

Table 3 presents the stayers and in-migrants at the province level. While Kwa-zulu-natal stands out as the most populous province, Gauteng province has the largest proportion of in-migrants to stayers ratio.

The top and bottom ten districts based on the criterion of in-migration is presented in table 4. The top and bottom 10 districts account for over 50% and less than 4% of total in-migration in the country.

Most studies on income inequality is based on per capita household income, however using the same approach for migration related inequality in receiving areas becomes problematic as individuals often migrate to urban areas without moving their entire original households along. Migrants, especially in the short term, tend to share accommodation with other migrants or relatives and friends in the receiving areas. This is validated by the descriptive statistics of our sample where a drastic reduction in the average household size of migrants post-migration is observed. Therefore this analysis is based on the individual income analysis of adults (15-60) years at the time of Wave 1 in 2008.

For the purpose of analyzing income inequality we calculate at the district level the gini coefficient as well as a generalized entropy measure (GE2) of individual income for broad as well as narrow income (Table 5). Inequality levels are very high by any benchmark. Narrow income inequality is understandably greater than broad income and highlights the significant role of government transfers in mitigating poverty and inequality. Although inequality is observed to be reducing over time with significant reductions observed in the period 2008-2012, these figures must not be taken as the national reality because of the specific sampling conditions imposed by restricting the sample for 2008 and 2012 to exclude all internal migration during that period. The decline in inequality ceases in the last period when including internal migrants in 2014.

While the substantial difference in the ginis based on narrow and broad income validate the redistributive function of government transfers and remittances, the declining difference between the periods points to the decreasing effectiveness of the state's redistributive actions. The declining remittance receipt as highlighted by Posel (2010) also can explain the reduction in the differences in inequality between narrow and broad incomes over time.

The GE(2), being an entropy measure, further allows a decomposition of inequality to enable further understanding of the source of inequality. It is evident from Table 6 that the source of decline in both income and wage inequality between first two periods is the within group (district) inequality. However the within-group inequality ceased to decline post migration indicating the positive contribution of in-migration to inequality. There is a marginal reduction in the between district income inequality following migration.

## 4.2 Descriptive Statistics

A brief summary of the various characteristics of our weighted sample of 7589 individuals is provided in this section. Over 75% of the migrants fall within the age category of 18-34 years at the time of migration post 2012 (Table 7). It is not surprising to note the very high share of the young among the immigrants as the young is more likely to venture out in the search for economic opportunities. Our sample has more females than males and this seems to influence the gender distribution of migrants as well. A similar 55%-45% female-male ratio among migrants was observed in the sample set used in Mulcahy and Kollamparambil (2016). A race-wise classification indicates our sample is nationally representative. The share of black race among migrants is seen to be well above what would be warranted by the population distribution. The average household size is seen to be 5 members. A drastic reduction in this is noticed among migrants post migration in 2012 which would indicate that individuals leave behind their families when they migrate. This validates our choice of individual rather than per capita household income to estimate the gini coefficient.

Table 8 shows that the share of individuals with no schooling is seen to be lower among migrants as compared to non-migrants. The share of population with matric and above qualifications has increased substantially over time for both migrants and non-migrants, however the increase is much more impressive among migrants. This clearly points to positive self-selection among migrants that differentiate them from non-migrants.

The high rate of South African unemployment is reflected in our sample (Table 9). Interesting differences in the rate of unemployment between migrants and non-migrants as well as change over time of unemployment rate among migrants can be highlighted. Firstly, migrants have a higher rate of unemployment as compared to non-migrants even post migration. Secondly, the rate in fall of unemployment is higher for migrants post migration as compared to non-migrants over the same period. The informal sector employment share of migrants and non-migrants are seen to be more or less similar. The average narrow income is higher for migrants post migration even though their income was lower than non-migrants before migration. This clearly indicates that while some migrants remain unemployed, others are able to improve their incomes dramatically post migration. Inclusion of government transfers and remittances substantially increases the average income of both stayers and migrants. The average broad income of the stayers is higher than of the migrants even post migration. The average male income is over twice that of the females. Although the gender wage ratio is lower among migrants, the differences in the gender wage ratio across migrants and non-migrants are not statistically significant.

# 5 Methodology

## 5.1 Estimation issues

We use data from 51 districts over three time periods (2008, 2012 and 2014) as made available by wave 1, wave 3 and wave 4 of NIDS for a multivariate econometric estimation of the impact of in-migration on income inequality within the receiving area. As a prelude, we first discuss some of the empirical and econometric issues that need has to be taken into account in the context of this study. Firstly, although OLS regression estimates is the most efficient in its class of linear estimators, the bidirectional nature of relationship between the dependent and independent variables leads to endogeneity bias using OLS estimation. This is relevant in our analysis because inequality levels within various areas will influence the destination choice of the migrant. Moreover each region has individual specific time invariant characteristics that can determine migrant choice, which if not captured entirely in the estimation model can also lead to biased results. Another econometric problem is that the time-series component in the panel data may involve autocorrelation of the disturbances. Autocorrelation might result in some or all estimated coefficients being biased, which could severely affect the interpretation of the relative impact of the various determinants on the inequality levels. OLS estimations ignoring these aspects can lead to biased results.

While the fixed effects separates out the time invariant characteristics from the rest of the explanatory variables and by differencing between time periods, removes these variables from the specification, it would continue to yield biased results in the presence of endogeneity arising from bidirectional causal relationship between variables. A GMM instrument variable approach to estimation is effective in countering this issue. Moreover, the persistent nature of inequality entails that past levels of inequality needs to be factored into estimations. Additionally, including the lagged dependent variable as a regressor in the regression equations as shown in equation (4) helps address significantly the problem of autocorrelation. By including the values of the dependent variable the econometric specification is changed to a dynamic panel. Therefore a dynamic GMM model is considered the appropriate choice for estimation in this context.

A method commonly applied to dynamic panels is the Arellano and Bond (1991) GMM estimator. Roodman (2009) recommends Arellano–Bond (Arellano and Bond 1991) and Arellano–Bover/Blundell–Bond (Arellano and Bover 1995; Blundell and Bond 1998) dynamic panel estimators for situations with 1) few time periods and many individuals (i.e. small T, Large N panels); 2) a linear functional relationship; 3) one left-hand-side variable that is dynamic, depending on its own past realizations; 4) independent variables that are not strictly exogenous, meaning they are correlated with past and possibly current realizations of the error; 5) ?xed individual e?ects; and 6) heteroskedasticity and autocorrelation within individuals but not across them. Arellano–Bond estimation (also known as the di?erence GMM) starts by transforming all regressors, usually by di?erencing, and uses the generalized method of moments (GMM) (Hansen 1982). The Arellano–Bover/Blundell–Bond estimator augments Arellano–Bond by making an additional assumption that ?rst di?erences of instrument variables are uncorrelated with the ?xed effects. The set-up of their estimator implies that the fixed effects are eliminated using first differences and an instrumental variable estimation of the differenced equation is performed. This allows the introduction of more instruments and can dramatically improve efficiency. It is known as the system GMM. This study uses the one-step system GMM estimator. The overall appropriateness of the instruments can be verified by a Difference in Hansen test for over-identifying restrictions.

Estimation models:

Based on the conceptual framework developed in Section 2, we evaluate empirically the relationship between in-migration and inequality in the receiving areas using the function below for both income inequality as well as for wage inequality among those employed:

$$Ineq_{it} = f(\Pr opmig_{it}, Employrate_{it}, Empifpop_{it})$$

$Ineq_{it}$ is the individual income inequality measure using Gini coefficient as well as Generalised Entropy index (GE(2)) in district $i$ at time $t$. Our target variables of interest, *propinmig, employrate* and *empifpop* indicate the proportion of net in-migrants into each district, the employment rate of the district, and the proportion of employed in the informal sector respectively. Based on our initial statistical analysis we expect *propinmig* to have a positive impact on inequality as over 50% of migrants are without matric level education and expected to either remain unemployed or find employment predominantly in the informal sectors of the receiving areas. The employment rate variable is expected to have a negative impact on inequality as increased level of employment is expected to decrease inequality. The coefficient of the *empifpop* may be positive or negative as discussed in Section 3.

In addition we include a set of control variables, $X_{it}$ based on our review of literature in Section 2. These include the average income level of the receiving areas (*avdistindinc*), the proportion of population with educational qualifications over matric level, (*abovMatricratio*), the proportion of population with educational qualifications with matric level (*matric*), the ratio of male average wages to female wages (*wageratio*). In addition, in order to account for the individual fixed effects we incorporate the province dummies in the OLS estimation. Furthermore a one period lagged dependent variable is included in the dynamic OLS model as the presence of autocorrelation is indicated in the OLS model. As a final model we estimate the system-GMM that incorporates the endogenous nature of some of the explanatory variables along with the persistent effect of inequality. A detailed description of the variables is provided in Appendix A1. All variables are included in log form in the econometric model to account for non-linearity.

$$Ineq_{it} = \alpha_1 + \alpha_2 \Pr opmig_{it} + \alpha_3 Employrate_{it} + \alpha_4 Empifpop_{it} + \phi X_{it} + \varepsilon_{it} \quad (1)$$

Our initial model is a multivariate ordinary least squares (OLS) regression. Model I incorporates our target variables of interest, and the control variables mentioned earlier in the section. We undertake a robust estimation to correct for heteroscedacticity in model 1. Next we run a dynamic model with lagged dependent variable included as an explanatory variable to correct for autocorrelation.

$$\begin{aligned} Ineq_{it} &= \gamma_1 + \gamma_2 Ineq_{it-1} + \gamma_3 \Pr opmig_{it} + \gamma_4 Employrate_{it} \\ &\quad + \gamma_5 Empifpop_{it} + \phi X_{it} + \varepsilon_{it} \end{aligned} \quad (2)$$

Model 2, nevertheless can potentially still suffer from endogeneity issues arising from suspected bidirectional causality as well as unaccounted for district level fixed effects.

$$\begin{aligned} \Delta Ineq_{it} &= \beta_1 \Delta \Pr opmig_{it} + \beta_2 \Delta Employrate_{it} \\ &\quad + \beta_3 \Delta Empifpop_{it} + \gamma \Delta X_{it} + \varepsilon_{it} \end{aligned} \quad (3)$$

The OLS estimates is improved upon with the fixed effects estimation (Model 3) that takes into account the three-period panel nature of the dataset. This method takes care of the endogeneity bias arising from unspecified fixed effects by separating out the time variant and invariant variables. This is achieved by undertaking the first difference between time periods, due to which the time invariant individual specific effect is effectively erased from the estimation. The fixed effects estimation however still suffer from suspected bias arising from reverse causality and autocorrelation arising from misspecification by not including the lagged dependant variable in the estimation.

$$\begin{aligned} \Delta Ineq_{it} &= \varphi_1 \Delta Ineq_{i,t-1} + \varphi_2 \Delta \Pr opmig_{it} + \varphi_3 \Delta Employrate_{it} \\ &\quad + \varphi_4 \Delta Empifpop_{it} + \lambda` \Delta X_{it} + \eta_i + \varepsilon_{it} \end{aligned} \tag{4}$$

The GMM-SYS estimation which is a system estimation of the level equation (2) and the difference equation (Model 4) effectively takes care of the issues of fixed effects, reverse causality, autocorrelation and model misspecification through omission of persistence effect of inequality. The bidirectional relationship that may exist in our specification is not just restricted to proportion of migrants and the inequality but also between the dependent variable on the one side and average income levels and educational attainment variables on the other. These variables are considered to be endogenous in the GMM-SYS estimation with the remaining variables assumed to be exogenous.

# 6 Multivariate Regression Results

Multivariate regressions were estimated to comprehend the impact of in-migration on the income and wage innequality. Results of these estimations are presented below.

## 6.1 Income inequality and In-migraton

The results of regression estimations of Models 1-4 with *Gini* as the dependent variable is presented in Tables 10 - 11.

Estimates of Model 1 and Model 3 estimates in these tables can be considered to be biased as the test for autocorrelation led to the rejection of the null hypothesis of no autocorrelation. The dynamic OLS regression (Model 2) with a statistically significant lagged dependant variable can hence be considered to be an improvement over Model 1. This indicates that inequality is persistent over time and requires to be incorporated into estimation models. Even so, since Model 2 does not take into account the endogeneity arising from reverse causality we consider Model 4 to be the better estimate. Nevertheless it must be highlighted that a comparison of models 2 and 4 yield very similar estimates with the dynamic OLS model yielding the additional effect of time invariant variable viz., province of the district. The results indicate that districts that belong to Free State, Eastern Cape, Mpumalanga, Kwa-zulu-natal and Limpopo have higher income inequality as compared to Gauteng province.

The positive and significant coefficient of the lagged dependent variable in Model 4 validates the persistent nature of regional income inequality (Table 10).

It must be pointed out that the results obtained by both models 2 and 4 are consistent in indicating a positive and significant impact of in-migration on narrow income inequality *(Gini_ Narrow)* within the receiving areas. As per our model of choice, a 1 % increase in in-migration (*propinmig*) into a district increases individual income inequality by 0.02%. The level of employment (*empratio*) is also seen in both the models to be critical at 99% confidence level in determining the income inequality levels. Increased employment rate will effectively reduce inequality by 0.11%. Districts with higher proportion of population with education level above matriculation (*abovMatricratio*) is seen to have higher levels of inequality at 5% significance level.

Comparing the above results with the regression estimation using broad income gini coefficient (*Gini_ Broad*) as the dependent variable (Table 11) yield interesting observations. While the proportion of in-migration (*propinmig*) continues to be positive and significant as in the earlier estimates (Table 10), employment rates (*empratio)* is no longer seen to be statistically significant in determining the inequality rates. This is an expected result as the government welfare transfers and remittances is expected to mitigate the effect of an absence of wage income through unemployment. Inequality as measured by the Generalised entroy measure of GE (2) indicates in-migration to have a positive and signifcant impact on inequality along the same lines as Gini measure of inequality (Table A3).

## 6.2 Wage inequality and In-migration

In order to better comprehend the labour market channel through which in-migration impacts on inequality in the receiving districts, we next investigate the role of in-migration in determing wage inequality. The dependant variable here is the Gini coefficient of wages (*Gini_ Wages*) among the employed, capturing the wage inequality (Table 12).

The insignificant coefficient of *propinmig* on wage inequality yields interesting insights into the labour market mechanism that determines the impact of in-migration on income inequality. It is now clear that in-migration does not contribute significantly to wage inequality and therefore its contribution to income inequality can be construed to be predominantly through the unemployment effect rather than through formal/informal sector employment. These conclusions are in consonance with the observations made from the descriptive statistics that show that unemployment rates are higher among migrants as compared to stayers, while informal sector share of migrants and stayers are not seen to be different from one another. However, it needs to be highlighted that, as expected, districts with higher share of informal sector employment have higher wage inequality. Wage inequality is seen to be higher more district thtat have a higher average individual income.

# 7 Conclusion

The paper addresses the issue of in-migration and individual income inequality in the receiving areas within South Africa in the Harris Todaro model framework. The analysis developed a three sector model which highlighted the role of labour market outcomes of in-migration in determining the impact on regional income inequality within the receiving areas. Evidence from the data shows that although migration results in lower unemployment rates among the migrants, they continue to have a higher rate of unemployment as compared to those who do not migrate

Findings from the multivariate analysis shows that rising urban income inequality in the urban areas as indicated by Leibbrandt, Finn and Woolard (2010) can be attributed at least in part to rural-urban migration. This is predominantly due to the higher unemployment rate among migrants. Other findings of the study are that while unemployment rate contributes to narrow income inequality, it does not contribute signficantly to broad income inequality. This highlghts the relevance of government welfare transfers as well as migrant remmittances. Districts with higher average individual income as well as with higher proportion of labour force with above matriculation education levevls are seen to have higher income inequality. While informal sector employment is not seen to be a significant factor in determining income inequality, it contributes significantly to wage inequality.

The policy implications of the study are clear in highlighting the need to address unemployment issues in general, and among migrants in particular, in making a dent on income inequality. Improving the employability of migrants is critical in reducing the inequality, therefore interventions to reduce income inequality in migrant receiving areas cannot ignore education and capacity building within the migrant source areas. Based on the findings of Posel (2010) that the largest group of migrants are rural-urban in nature, it may be argued that addressing education and skill formation in rural sector will improve the quality of migrants that can be absorbed in the urban formal sector without increasing inequality. In conclusion, it can be said that the urban bias in policy formulation that ignores the rural sector cannot be successful as it can only lead to the opposite effect as indicated by the Harris Todaro model.

One limitation of the analysis emanates from the bias in the sample pertaining to gender distribution. This study is nevertheless a starting point for empirical literature to further analyse the impact of in-migration on the inequality of receiving areas taking into account unemployment as well as the wage gaps between migrant and non-migrants, formal and informal sectors, wage income and self-employment income.

# References

[1] Arellano, M. & Bond, S.(1991). Some tests of specification for panel data: Monte Carlo evidence and an application to employment equations, *Review*

*of Economic Studies* 58, 277–297.

[2] Arellano, M. & Bover, O.(1995). Another look at the instrumental variable estimation of error-components models, *Journal of Econometrics* 68 (1995) 29-51.

[3] Bastia Tanja (2013). *Migration and Inequality*, Routledge.

[4] Black, R., Natali, C. & Skinner, J. (2005) *Migration and inequality*, Washington , World Bank.

[5] Blundell, R., & Bond, S. (2000). GMM estimation with persistent panel data: an application to production functions, *Econometric reviews*, 19(3), 321-340.

[6] Card, David (2009) Immigration and Inequality, *American Economic Review*: Papers and Proceeding, 99:2, 1-21.

[7] Collinson, M., Tollman, S., Kahn, K., & Clark, S. (2003). *Highly prevalent circular migration: Households, mobility and economic status in rural South Africa*. Conference on African Migration in Comparative Perspective. Johannesburg: Agincourt Health and Population Unit: University of the Witwatersrand.

[8] Cornwell, K., & Inder, B. (2004). *Migration and unemployment in South Africa: When motivation surpasses the theory*. Monash University, Department of Econometrics and Business Statistics: Monash Econometrics and Business Statistics Working Papers 2/04.

[9] De Villiers, L., Brown, M., Woolard, I., Daniels, R.C., & Leibbrandt, M. (Eds). (2013). *National income dynamics study wave 3 user manual*. Cape Town: Southern Africa Labour and Development Research Unit.

[10] Eastwood R & Lipton M (2000*) Rural Urban Dimensions of Inequality Change*, working paper no.200, UNU-WIDER, Finland.

[11] Gustafsson, Björn & Shi Li Economic transformation and the gender earnings gap in urban China, *Journal of Population Economics*, July 2000, Volume 13, Issue 2, pp 305-329.

[12] Ha, Wei & Yi, Junjian & Zhang, Junsen, 2009. *Internal Migration and Income Inequality in China: Evidence from Village Panel Data*, MPRA Paper 16896, University Library of Munich, Germany.

[13] Hansen, L. P. 1982. Large Sample Properties of Generalized Method of Moments Estimators, *Econometrica* 50: 1029-1054.

[14] Harris John R. and Michael P. Todaro (1970) Migration, Unemployment and Development: A Two-Sector Analysis, *The American Economic Review*, Vol. 60, No. 1 (1970), pp. 126-142.

[15] ILO (2015) *Global Wage Report* -2014/15: Wages and income inequality, Geneva.

[16] Kollamparambil U & Razak A (2011) *Trends in Gender wage gap and Discrimination in South Africa*, conference paper presented at Fourth meeting of the Society for the Study of Economic Inequality (ECINEQ), Catania (Italy), July 18-20, 2011.

[17] Kuznet Simon (1955) Economic Growth and Income Inequality, *The American Economic Review*, Volume XlV March, 1955 Number One.

[18] Lall, S.V. Harris, S. & Shalizi, Z (2006) Rural-urban migration in developing countries: A survey of theoretical predictions and empirical findings. World Bank Policy Research Workng Paper, 3915 May.

[19] Leibbrandt, M., Woolard, I., de Villiers, L. (2009) *Methodology Report on NIDS Wave 1*, Technical paper no.1, National Income Dynamics Survey, South Africa.

[20] Leibbrandt, M., Wegner, E., Finn, A. (2011). *The Policies for Reducing Income Inequality and Poverty in South Africa,* A Southern Africa Labour and Development Research Unit Working Paper Number 64. Cape Town: SALDRU, University of Cape Town.

[21] Leibbrandt, M., Finn, A. and I. Woolard (2010) *"Describing and decomposing post-apartheid inequality in South Africa: An analysis based on income and expenditure data from the 1993 Project for Statistics on Living Standards and Development and the 2008 base wave of the National Income Dynamics Study".* Conference paper, Overcoming Inequality and Structural Poverty in South Africa: Toward inclusive growth and development, Birchwood Hotel, Johannesburg, 20–22 September 2010.

[22] Leibbrandt, M., Finn, A., Argent, J., & Woolard, I. (2010). http://reference.sabinet.co.za/sa_epublication_article/bersee_v34_n3_a3 Changes in Income Poverty over the Post-Apartheid Period: An Analysis Based on Data from the Project for Statistics on Living Standards and Development and the 2008 Base Wave of the South African National Income Dynamics Study. *Studies in Economics and Econometrics*, 34(3), 25-43.

[23] Leibbrandt, M., Woolard, I., McEwan, H. & Koep, C. (2010) Better Employment to Reduce Inequality Further in South Africa, Chapter 5 in OECD, *Tackling Inequalities in Brazil, India, China and South Africa: the Role of Labour Market and Social Policies.* OECD Publishing.

[24] Lucas, R. E. (1987). Emigration to South Africa's mines. *The American Economic Review*, 77(3), 313–330.

[25] Lucas, R. E. (2013). *African migration.* Boston: Boston University.

[26] Lipton, M. (1980). Migration from rural areas of poor countries: the impact on rural productivity and income distribution. *World Development*, 8, 1-24.

[27] Mbatha Cyril & Roodt Joan (2014), Recent Internal Migration And Labour Market Outcomes: Exploring The 2008 And 2010 National Income Dynamics Study (Nids) Panel Data In South Africa, *South African Journal of Economic & Management Sciences*, 17, No 5: 653-672.

[28] Muller, C. 2009. *Trends in the gender wage gap and gender discrimination among part-time and full-time workers in post-apartheid South Africa.* Development Policy Research Unit, School of Economics, University of Cape Town.

[29] Mulcahy, K. (2015). *Rural-urban migration in South Africa: An analysis of subjective and economic well-being* (Masters Research Report). University of Witwatersrand, Johannesburg.

[30] Mulcahy, K & Kollamparambil, U (2016): The Impact of Rural-Urban Migration on Subjective Well-Being in South Africa, *The Journal of Development Studies*, DOI: 10.1080/00220388.2016.1171844

[31] National Planning Commission, 2013 "*Nation Income Dynamics Study: Overview 2013*, South Africa.

[32] Ntuli. M. 2007. *Exploring Gender Wage "Discrimination" in South Africa, 1995-2004: A Quantile Regression Approach.* INTERNATIONAL POLICY CENTER Gerald R. Ford School of Public Policy, University of Michigan, IPC Working Paper Series Number 56.

[33] Posel, D. (2004) "Have migration patterns in post-apartheid South Africa changed?" *Journal of Interdisciplinary Economics* 15: 277-292.

[34] Posel, D. (2010) "Households and labour migration in post-apartheid South Africa*", Studies in Economics and Econometrics* 34(3): 129-141.

[35] Raphael Steven & Eugene Smolensky (2008*) Immigration and Poverty in the United States,* Institute for Research on Poverty, Discussion Paper no. 1347-08.

[36] Rogan, M., Lebani, L., & Nzimande, N. (2009*). Internal migration and poverty in KwaZulu-Natal: Findings from censuses, labour force surveys and panel data.* Cape Town: Southern Africa Labour and Development Research Unit, UCT.

[37] Roodman, D. (2006). *How to do xtabond2: An introduction to difference and system GMM in Stata.* Center for Global Development working paper, (103).

[38] Sanderson 2013, Immigration and global inequality: A cross-national analysis, in Tanja Bastia ed *Migration and Inequality*, Routledge.

[39] Temple Jonathan (2005) Growth and wage inequality in a dual economy, *Bulletin of Economic Research,* Volume 57, Issue 2, pages 145–169, April 2005.

[40] UNDP 1990 , *Human Development Report Overcoming barriers: Human mobility and development.* Geneva.

[41] World Bank (2013) Data sourced from http://data.worldbank.org/indicator/SI.POV.GINI?page=4).

[42] Zaakirah I and Kollamparambil U (2015) Youth Unemployment Duration and Competing Exit States: What Hides Behind Long Spells of Black Youth Unemployment in South Africa? African Development Review, Volume 27 Issue 3, pages 30–314.

[43] Zuberi, T., & Sibanda, A. (2004, Winter). How do migrants fare in a post-apartheid South African labour market. *International Migration Review,* 38(4), 1462–1491. doi:10.1111/j.1747-7379.2004.tb00244.x

[44] Zuma, N. (2013). *Rural-urban migration in South Africa. Growth, transformation, reform: Emerging economies in the next decade.* Haikou, China: Economic Policy Forum.

## Table 1: Sample size by group

| | Stayers | Migrants | Total |
|---|---|---|---|
| Unweighted (number) | 7163 | 426 | 7589 |
| % | 94.39 | 5.61 | 100 |
| Weighted (number) | 7191.25 | 397.75 | 7589 |
| % | 94.76 | 5.24 | 100 |

## Table 2: Distribution of districts across Provinces

| Province | No. of districts | Districts |
|---|---|---|
| Eastern Cape (EC) | 8 | Alfred Nzo, Amathole, Buffalo City, Cacadu, Chris Hani, Joe Gqabi Nelson Mandela Bay, O.R.Tambo |
| Free State (FS) | 5 | Fezile Dabi, Lejweleputswa, Mangaung, Thabo Mofutsanyane, Xhariep |
| Gauteng (GP) | 5 | City of Johannesburg, City of Tshwane, Ekurhuleni, Sedibeng, West Rand |
| Kwa-Zulu-Natal (KZN) | 10 | Amajuba, Sisonke, uMgungundlovu, Ugu, uMkhanyakude, uThukela, uThungulu, Zululand, eThekwini, iLembe |
| Limpopo (LP) | 5 | Capricorn, Greater Sekhukhune, Vhembe, Mopani, Waterberg |
| Mpumalanga (MP) | 3 | Ehlanzeni, Gert Sibande, Nkangala |
| North West (NW) | 4 | Bojanala, Dr Kenneth Kaunda, Dr Ruth Segomotsi, Ngaka Modiri Molema. |
| Northern Cape (NC) | 5 | Frances Baard, John Taolo Gaetsewe, Namakwa, Pixley ka Seme, Siyanda |
| Western Cape (WC) | 6 | Cape Winelands, Central Karoo, Eden, Overberg, West Coast, City of Cape Town |

## Table 3: Internal migration into Provinces

| Province | Stayers | Net In-Migrants | Total | Province share of total migration | Proportion of net in-migrants |
|---|---|---|---|---|---|
| E.Cape | 862.33 | 58.55 | 920.85 | 15.11 | 6.79 |
| Free State | 497.67 | 12.64 | 510.31 | 3.26 | 2.54 |
| Gauteng | 1990.22 | 131.54 | 2121.76 | 33.90 | 0.06 |
| KZN | 1280.97 | 69.86 | 1350.83 | 18.03 | 5.45 |
| Limpopo | 637.70 | 34.83 | 672.53 | 8.98 | 0.05 |
| Mpumalanga | 590.92 | 25.56 | 616.48 | 6.59 | 0.04 |
| North West | 367.93 | 12.02 | 379.95 | 3.10 | 3.26 |
| N. Cape | 181.26 | 2.65 | 183.91 | 0.68 | 1.46 |
| W.Cape | 782.48 | 39.79 | 822.27 | 10.27 | 5.08 |
| Total | 7191.53 | 397.75 | 7589.00 | 5.38 | 5.38 |

**Table 4: District-wise distribution of in-migration**

| Top 10 districts | Province | migrant to stayer ratio % | Share of district in total migration % | Bottom 10 districts | Province | migrant to stayer ratio % | Share of district in total migration % |
|---|---|---|---|---|---|---|---|
| City of Johannesburg | GP | 6.9 | 12.56 | Central Karoo | WC | 0.0 | 0 |
| City of Tshwane | GP | 10.22 | 10.58 | Namakwa | NC | 0.0 | 0 |
| Ekurhuleni | GP | 7.04 | 6.56 | Lejweleputswa | FS | 0.18 | 0.05 |
| eThekwini | KZN | 3.88 | 6.17 | Siyanda | NC | 1.06 | 0.86 |
| Eden | WC | 24.8 | 5.5 | Xhariep | FS | 2.64 | 0.11 |
| Alfred Nzo | EC | 25.9 | 4.6 | Pixley ka seme | NC | 1.98 | 0.12 |
| West Rand | GP | 10.1 | 4.01 | Zululand | KZN | 0.74 | 0.13 |
| City of Cape Town | WC | 2.94 | 3.57 | Overberg | WC | 1.38 | 0.14 |
| Nkangala | MP | 7.8 | 3.47 | Frances Baard | NC | 0.97 | 0.13 |
| Waterberg | LP | 11.04 | 2.8 | Sedibeng | GP | 1.06 | 0.86 |

**Table 5: Income and Wage inequality**

| | Income inequality | | | | Wage Inequality | |
|---|---|---|---|---|---|---|
| | Gini | | GE(2) | | Gini | GE(2) |
| | Narrow | Broad | Narrow | Broad | Narrow | |
| 2008 | 0.85 | 0.80 | 4.56 | 3.70 | 0.53 | 0.85 |
| 2012 | 0.79 | 0.72 | 2.94 | 2.33 | 0.52 | 0.73 |
| 2014 | 0.79 | 0.72 | 2.93 | 2.33 | 0.52 | 0.74 |

**Table 6: Generalised Entropy Index (GE(2)) Decomposition**

| | Narrow Income | | Broad Income | | Wage | |
|---|---|---|---|---|---|---|
| | Within Group | Between Group | Within Group | Between Group | Within Group | Between Group |
| 2008 | 4.39 | 0.17 | 3.57 | 0.13 | 1.37 | 0.10 |
| 2012 | 2.77 | 0.17 | 2.20 | 0.13 | 1.08 | 0.09 |
| 2014 | 2.77 | 0.16 | 2.20 | 0.12 | 1.08 | 0.09 |

## Table 7: Individual & Household Characteristics

| | All | Stayers | Migrants |
|---|---|---|---|
| Age in 2008 % | | | |
| 15-30 | 44,33 | 42,44 | 76,12 |
| 31-45 | 31,57 | 32,59 | 14,42 |
| 46-61 | 24,10 | 24,96 | 9,46 |
| Gender | % | | |
| Male | 36,58 | 36,09 | 44,68 |
| Female | 63,42 | 63,91 | 55,32 |
| Race | % | | |
| African | 80,70 | 79,99 | 92,72 |
| Coloured | 15,70 | 16,34 | 4,93 |
| Indian/Asian | 1,20 | 1,27 | 0,00 |
| White | 2,40 | 2,40 | 2,35 |
| Average Household size(numbers) | | | |
| 2008 | 5.0 | 5.0 | 4.6 |
| 2012 | 4.9 | 4.9 | 4.2 |
| 2014 | 4.7 | 4.9 | 2.4 |

## Table 8: Level of Education (%)

| | All | Stayers | Migrants |
|---|---|---|---|
| No schooling | | | |
| 2008 | 8.6 | 8.9 | 3.1 |
| 2012 | 8.1 | 8.5 | 2.6 |
| 2014 | 8.0 | 8.3 | 2.3 |
| Below Matric Schooling | | | |
| 2008 | 67.6 | 67.4 | 71.1 |
| 2012 | 62.1 | 62.2 | 60.2 |
| 2014 | 58.3 | 58.7 | 50.2 |
| Matric | | | |
| 2008 | 15.9 | 15.6 | 20.7 |
| 2012 | 15.5 | 15.1 | 22.8 |
| 2014 | 14.3 | 13.7 | 23.9 |
| Above Matric | | | |
| 2008 | 7.9 | 8.1 | 5.2 |
| 2012 | 14.3 | 14.3 | 14.4 |
| 2014 | 19.5 | 19.2 | 23.5 |

## Table 9: Labour Market Characteristics

| | All | Stayers | Migrants |
|---|---|---|---|
| Employment rate (broad measure %) | | | |
| 2008 | 39.7 | 39.8 | 37.8 |
| 2012 | 45.8 | 45.9 | 43.1 |
| 2014 | 53.3 | 53.3 | 52.5 |
| Informal sector employment (%) | | | |
| 2008 | 23.9 | 24 | 23.1 |
| 2012 | 20.9 | 20.9 | 20.9 |
| 2014 | 20.8 | 20.8 | 21 |
| Average individual Narrow income (Rands) | | | |
| 2008 | 1816.9 | 1823.7 | 1695.4 |
| 2012 | 2382.5 | 2351.8 | 1867.8 |
| 2014 | 2979.6 | 2967.2 | 3205.1 |
| Average individual Broad income (Rands) | | | |
| 2008 | 2013.6 | 2020.8 | 1884.4 |
| 2012 | 2714.5 | 2689.1 | 1998.7 |
| 2014 | 3764.7 | 3777.9 | 3512.7 |
| Gender Wage ratio | | | |
| 2008 | 2.38 | 2.39 | 2.21 |
| 2012 | 2.15 | 2.16 | 1.86 |
| 2014 | 2.08 | 2.08 | 2.09 |

**Table 9: Regression results, Dependent variable log (Gini: Narrow income)**

| VARIABLES | OLS (1) | Dynamic OLS (2) | Fixed effects (3) | GMM-SYS (4) |
|---|---|---|---|---|
| Log propinmig | 0.0004 | 0.011** | 0.002 | 0.016*** |
| | (0.006) | (0.004) | (0.005) | (0.005) |
| Log avdistindinc | -0.016 | -0.033** | -0.065*** | -0.019 |
| | (0.014) | (0.015) | (0.019) | (0.015) |
| Log empratio | -0.148*** | -0.056** | -0.074*** | -0.111** |
| | (0.027) | (0.027) | (0.024) | (0.055) |
| Log empif | 0.0138 | 0.020 | 0.028** | 0.013 |
| | (0.014) | (0.014) | (0.011) | (0.015) |
| Log matric | 0.246** | 0.0006 | 0.481 | -0.052 |
| | (0.107) | (0.0065) | (0.617) | (0.031) |
| Log abovMatricratio | -0.252** | 0.0058 | -0.345 | 0.069** |
| | (0.109) | (0.076) | (0.603) | (0.034) |
| Log wageratio | 0.0125 | 0.0093 | 0.003 | 0.011 |
| | (0.008) | (0.006) | (0.008) | (0.008) |
| ecdummy | 0.061*** | 0.038* | | |
| | (0.018) | (0.021) | | |
| fsdummy | 0.057*** | 0.043** | | |
| | (0.018) | (0.027) | | |
| kzndummy | 0.0468** | 0.016 | | |
| | (0.018) | (0.019) | | |
| lpdummy | 0.869*** | 0.422** | | |
| | (0.016) | (0.021) | | |
| mpdumy | 0.606** | 0.025 | | |
| | (0.025) | (0.026) | | |
| nwdummy | 0.022 | 0.011 | | |
| | (0.026) | (0.032) | | |
| ncdummy | -0.003 | -0.009 | | |
| | (0.021) | (0.025) | | |
| wcdummy | -0.0097 | 0.000 | | |
| | (0.024) | (0.019) | | |
| L. Log Gini_Narrow | | 0.581*** | | 0.594*** |
| | | (0.100) | | (0.108) |
| Constant | -0.303** | 0.106 | 0.885 | 0.022 |
| | (0.148) | (0.144) | (0.553) | (0.162) |
| Observations | 153 | 102 | 153 | 102 |
| F | 26.7*** | 39.5*** | 9.54*** | |
| R-squared | 0.63 | 0.806 | 0.24 | |
| Number of Districts | | | 51 | 51 |
| Woolridge | 18.12*** | | 16.23*** | |
| Hansen test | | | | 8.73 |
| Prob > chi2 | | | | 0.12 |
| Diff Hansen | | | | 8.25 |
| Prob > chi2 | | | | 0.22 |

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

**Table 10: Regression results, Dependent variable log (Gini: Broad income)**

| VARIABLES | OLS (1) | Dynamic OLS (2) | Fixed Effects (3) | SYS-GMM (4) |
|---|---|---|---|---|
| Log propinmig | -0.001 | 0.025*** | 0.002 | 0.028*** |
| | (0.007) | (0.006) | (0.006) | (0.006) |
| Log avdistindinc | 0.074*** | 0.017 | -0.117*** | -0.021 |
| | (0.022) | (0.023) | (0.024) | (0.034) |
| Log empratio | -0.172*** | -0.009 | -0.096*** | 0.019 |
| | (0.037) | (0.031) | (0.030) | (0.044) |
| Log abovMatricratio | 0.021 | 0.258*** | -0.599 | 0.475*** |
| | (0.019) | (0.059) | (0.773) | (0.128) |
| Log matric | -0.804 | -0.277*** | 0.684 | -0.466*** |
| | (0.261) | (0.053) | (0.792) | (0.127) |
| Log wageratio | 0.004 | 0.011 | 0.012 | 0.021 |
| | (0.010) | (0.008) | (0.011) | (0.016) |
| Log empif | 0.029 | 0.029 | 0.034** | |
| | (0.019) | (0.018) | (0.014) | |
| ecdummy | 0.044 | 0.053** | | |
| | (0.027) | (0.027) | | |
| fsdummy | 0.027 | 0.038 | | |
| | (0.026) | (0.027) | | |
| kzndummy | -0.033 | 0.003 | | |
| | (0.027) | (0.023) | | |
| lpdummy | 0.103*** | 0.057** | | |
| | (0.026) | (0.029) | | |
| mpdumy | 0.066** | 0.020 | | |
| | (0.032) | (0.029) | | |
| nwdummy | 0.004 | 0.012 | | |
| | (0.026) | (0.031) | | |
| ncdummy | -0.011 | 0.013 | | |
| | (0.026) | (0.029) | | |
| wcdummy | -0.014 | -0.006 | | |
| | (0.033) | (0.026) | | |
| L. Log Gini_Broad | | 0.608*** | | 0.753*** |
| | | (0.088) | | (0.078) |
| Constant | -1.113*** | -0.409* | 0.859 | 0.128 |
| | (0.219) | (0.231) | (0.709) | (0.315) |
| Observations | 153 | 102 | 153 | 102 |
| F | 8.20*** | 27.9*** | 13.28*** | 20.54*** |
| R-squared | 0.377 | 0.704 | 0.639 | |
| Number of Districts | | | 51 | 51 |
| Wooldridge test for | 10.54*** | | 12.82 | |
| Hansen test Chi-sq(13) | | | | 19.23 |
| Prob > chi2 | | | | 0.113 |
| Diff in Hansen test Chi- | | | | 10.70 |
| Prob > chi2 | | | | 0.141 |

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

**Table 10: Regression results, Dependent variable log (Gini: Wages)**

| VARIABLES | OLS (1) | Dynamic OLS (2) | Fixed Effects (3) | SYS-GMM (4) |
|---|---|---|---|---|
| Log propinmig | 0.006 | 0.007 | 0.007 | 0.118 |
| | (0.012) | (0.010) | (0.483) | (0.010) |
| Log avdistindinc | 0.114*** | 0.027 | 0.130*** | 0.097** |
| | (0.025) | (0.036) | (0.039) | (0.040) |
| Log empifpop | 0.052* | 0.062* | 0.077*** | 0.063* |
| | (0.028) | (0.031) | (0.023) | (0.038) |
| Log matric | -0.002 | 0.126 | -0.682 | 0.135 |
| | (0.108) | (0.202) | (1.35) | (0.097) |
| Log abovMatricratio | 0.007 | -0.085 | 0.742 | -0.103 |
| | (0.108) | (0.203) | (1.321) | (0.103) |
| L. Log Gini_Broad | | 0.597*** | | 0.114 |
| | | (0.097) | | (0.114) |
| Constant | -1.39*** | -0.174 | -1.198 | -1.049*** |
| | (0.209) | (0.313) | (1.24) | (0.361) |
| Observations | 153 | 102 | 153 | 102 |
| F | 14.98*** | 14.58*** | 3.98*** | 3.98*** |
| R-squared | 0.103 | 0.48 | 0.11 | |
| Number of Districts | | | 51 | 51 |
| Wooldridge test for | 10.543*** | | 12.824*** | |
| Hansen test Chi-sq(13) | | | | 14.13 |
| Prob > chi2 | | | | 0.118 |
| Diff in Hansen test Chi- | | | | 5.58 |
| Prob > chi2 | | | | 0.472 |

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

# APPENDIX
## Table A1

| Variable Name | Definition |
|---|---|
| *GINI_NARROW* | Gini coefficient estimated (among both employed and unemployed) in each district from individual income (adjusted for inflation, base year 2012) which includes wage income and investment income |
| *GINI_BROAD* | Gini coefficient estimated (among both employed and unemployed) in each district from individual income (adjusted for inflation, base year 2012) which includes wage income, investment income and transfer income |
| *GINI_WAGE* | Gini coefficient estimated (among employed) in each district from wage income of the employed (adjusted for inflation, base year 2012) |
| *GE _NARROW* | Generalised Entropy Index (GE (2)) estimated (among both employed and unemployed) in each district from individual income (adjusted for inflation, base year 2012) which includes wage income and investment income |
| *GE _BROAD* | Generalised Entropy Index (GE (2)) estimated (among both employed and unemployed) in each district from individual income (adjusted for inflation, base year 2012) which includes wage income, investment income and transfer income. |
| *GE_WAGE* | Generalised Entropy Index (GE (2)) (among employed) estimated in each district from wage income (adjusted for inflation, base year 2012). |
| *PROPINMIG* | Ratio of in-migrants over stayers in each district |
| *AVDISTINDINC* | Average individual income in each district |
| *EMPRATIO* | Ratio of employed over total district population estimated using a broad definition of unemployment. |
| *ABOVMATRICRATIO* | Ratio of individuals with above matriculation qualifications over the total district population |
| *MATRIC* | Ratio of individuals with matriculation level qualifications over the total district population |
| *WAGERATIO* | Ratio of average Male wages over female wages in each district |
| *ECDUMMY* | Dummy variable, equals 1 for districts of Eastern Cape Province, 0 otherwise. |
| *FSDUMMY* | Dummy variable, equals 1 for districts of Free State Province, 0 otherwise |
| *KZNDUMMY* | Dummy variable, equals 1 for districts of Kwa-Zulu-Natal Province, 0 otherwise |
| *LPDUMMY* | Dummy variable, equals 1 for districts of Limpopo Province, 0 otherwise |
| *MPDUMY* | Dummy variable, equals 1 for districts of Mpumalanga Province, 0 otherwise |
| *NWDUMMY* | Dummy variable, equals 1 for districts of North West Province, 0 otherwise |
| *NCDUMMY* | Dummy variable, equals 1 for districts of Northern Cape Province, 0 otherwise |
| *WCDUMMY* | Dummy variable, equals 1 for districts of Western Cape Province, 0 otherwise |

**Table A2: Numerical illustrations of in-migration impact scenarios**

| Labour | Pre-migration | 1 | 2 | 3a | 3b | 4 | 5 | 6a | 6b |
|---|---|---|---|---|---|---|---|---|---|
| 1 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 2 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 | 5 |
| 3 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 2 |
| 4 | 3 | 3 | 3 | 3 | 3 | 2 | 2 | 1 | 2 |
| 5 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 6 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 | 0 |
| 7 | | 5 | 0 | 0 | 0 | 0 | 0 | 1 | 2 |
| 8 | | 3 | 0 | 5 | 3 | 5 | 2 | 1 | 2 |
| 9 | | 0 | 0 | 5 | 3 | 2 | 2 | 1 | 2 |
| Gini | 0.416 | 0.416 | 0.611 | 0.401 | 0.414 | 0.476 | 0.481 | 0.518 | 0.388 |

Robust standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1

**Table A3: SYS GMM Regression results, Dependent variable log GE (2)**

| VARIABLES | Narrow Income (1) | Broad Income (2) | Wages (3) |
|---|---|---|---|
| Log propinmig | 0.077*** | 0.097*** | -0.002 |
| | (0.027) | (0.028) | (0.038) |
| Log avdistindinc | -0.091 | 0.064 | 0.181 |
| | (0.098) | (0.132) | (0.136) |
| Log empratio | -0.369 | -0.196 | 0.311 |
| | (0.367) | (0.375) | (0.371) |
| Log empif | -0.050 | 0.117 | 0.156 |
| | (0.097) | (0.113) | (0.143) |
| Log matric | 0.409 | -0.063 | 1.677** |
| | (0.325) | (0.349) | (0.785) |
| Log abovMatricratio | -0.287** | 0.162 | -1.59** |
| | (0.347) | (0.366) | (0.826) |
| Log wageratio | 0.124 | 0.149* | 0.222 |
| | (0.088) | (0.088) | (0.179) |
| L. Dep variable | 0.579*** | 0.593*** | -0.037 |
| | (0.105) | (0.126) | (0.073) |
| Constant | 1.129 | -0.079 | -1.038 |
| | (0.965) | (1.165) | (1.462) |
| Observations | 102 | 102 | 102 |
| F | 106.7*** | 12.9*** | 3.89*** |
| Number of Districts | 51 | 51 | 51 |
| Hansen test | 16.69 | 18.4 | 18.4 |
| Prob > chi2 | 0.214 | 0.143 | 0.143 |
| Diff Hansen | 8.81 | 10.55 | 1.96 |
| Prob > chi2 | 0.184 | 0.103 | 0.742 |