

Developing estimates for South African local municipality socio-economic models: applications of the rank-size rule 2007 to 2013

Hlabi Morudu

Public Sector Economist Forum 6th Annual Congress

Lord Charles Hotel, Somerset West, Cape Town

26 November 2014



••• your leading partner in quality statistics

Introduction

Addresses some national challenges on spatial statistics

Applies the rank-size rule (Zipf's rule) for the estimation of key variables at local municipality level

Presents estimates of key socio-economic variables at local municipality level: i.e. population size, employment, unemployment & GDP

The problem

Socio-economic planners on local municipalities do not have key data for the development and required monitoring of IDPs, SDFs & LEDs

Stats SA regularly provides data on population size, employment, unemployment and GDP during non-Census periods, but due to budgetary constraints, provides such data only at national and provincial levels.

Local municipality planners get Census data (or large sample data) once in 5 to 10 years.

They do not get data on key variables in non-Census years, i.e. population size, employment, unemployment and GDP.

Local municipality population size, employment, unemployment and GDP are central in developing, monitoring and aligning local municipality plans with the National Development Plan.

The problem

Several efforts are currently in progress within Stats SA to develop estimates at local municipality level, but are not expected to yield results in the immediate future.

The private sector does simulate some estimates of local municipality population sizes, employment, unemployment and GDP, but is not explicit with its methods

The objectives of this paper are to:

- (1) develop practical local municipality estimates; and
- (2) promote transparency in the development of local municipality estimates, viz. population size, employment, unemployment and GDP.

Approach

Propose use of Zipf's rule for the development of local municipality estimates

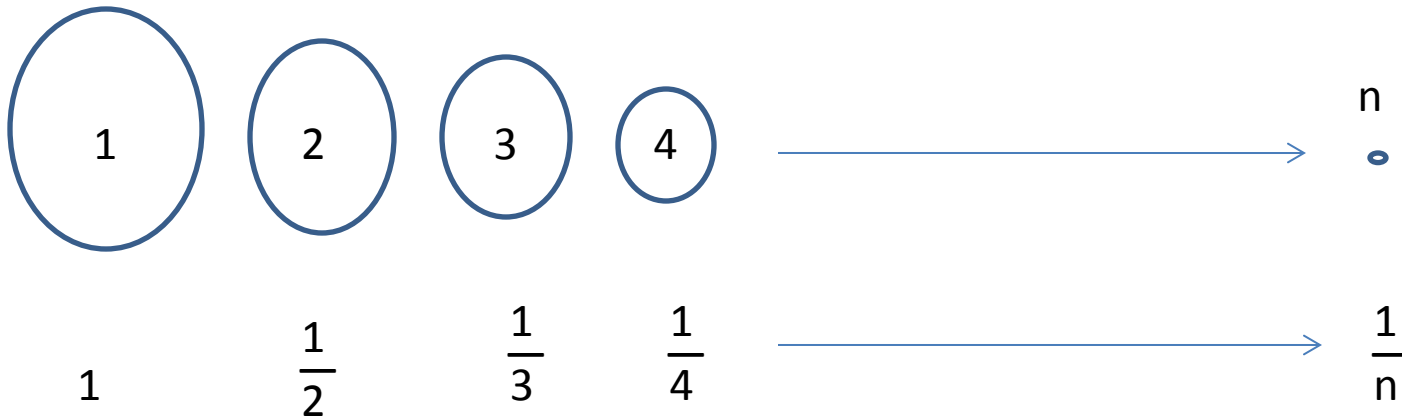
Zipf's rule is proposed because of its proven empirical performance internationally, with consistently high statistical inference properties.

Zipf's rule applies with regard to the distribution of a number of variables, e.g. geographical distributions of population sizes, employment, unemployment and GDP

Regular ratio-rule methods currently posit significant challenges for local municipality estimates:

- (a) there are very few useful censuses (1996, 2001, 2011) to can develop any reliable statistical inferences.

Approach: rank-size rule



More generally, the rank-size rule is written as:

[1] $S_i = CR_i^{-\alpha}$...when $\alpha=1$, Zipf's rule holds

or

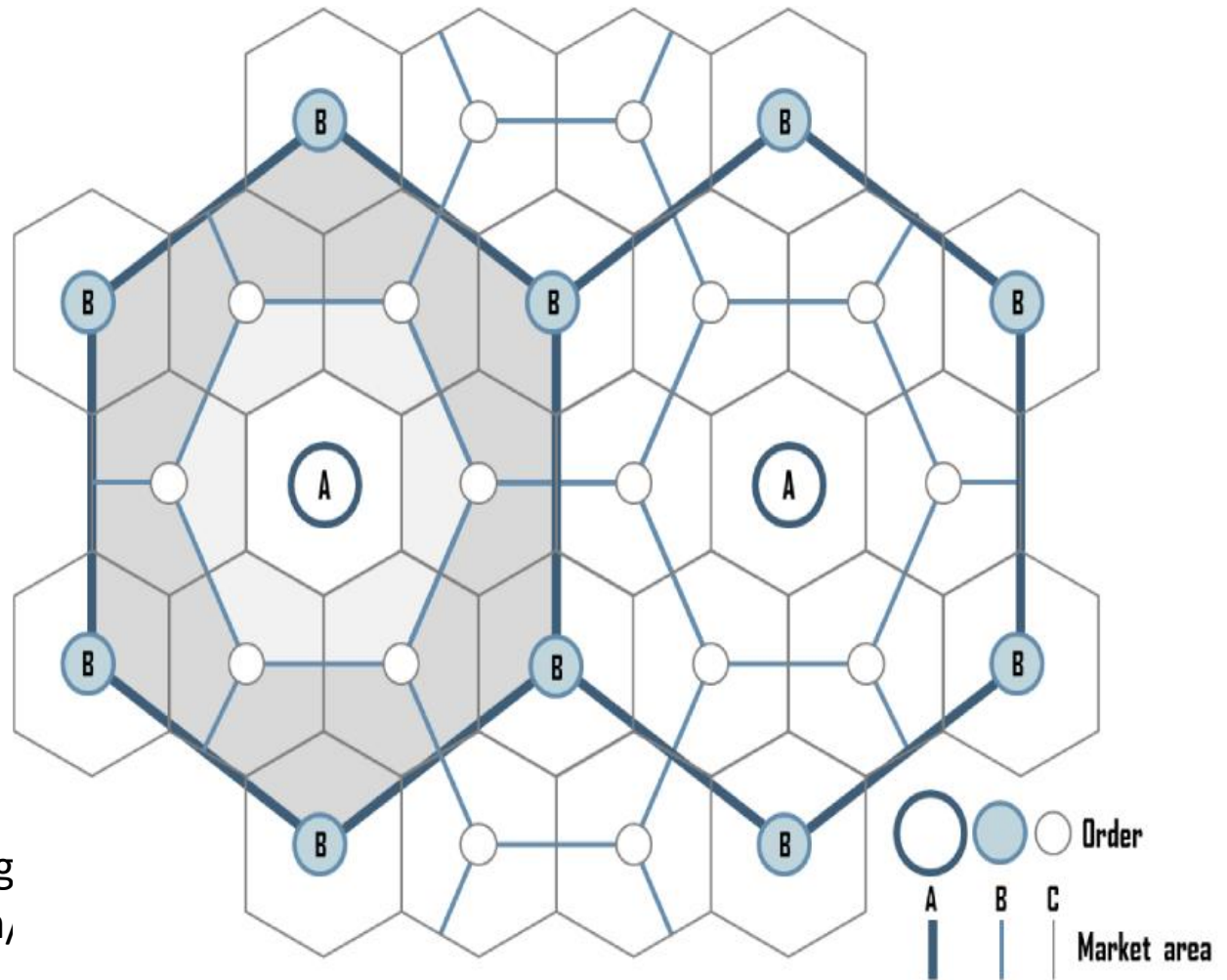
[2] $\log(S_i) = C - \alpha \log(R_i) + \epsilon_i$...in logarithmic form

Literature review: basis of the rank-size rule in geography

Rationale: based on classic works of von Thunen (1826), Christaller (1933), Losch (1954), Philbrick (1957), Berry (1964) on the role of central places, how they are distributed, and fit into a hierarchy of higher places.

Source:

<http://people.hofstra.edu/gotrans/eng/ch2en/conc2en/centralplacetheory.html>



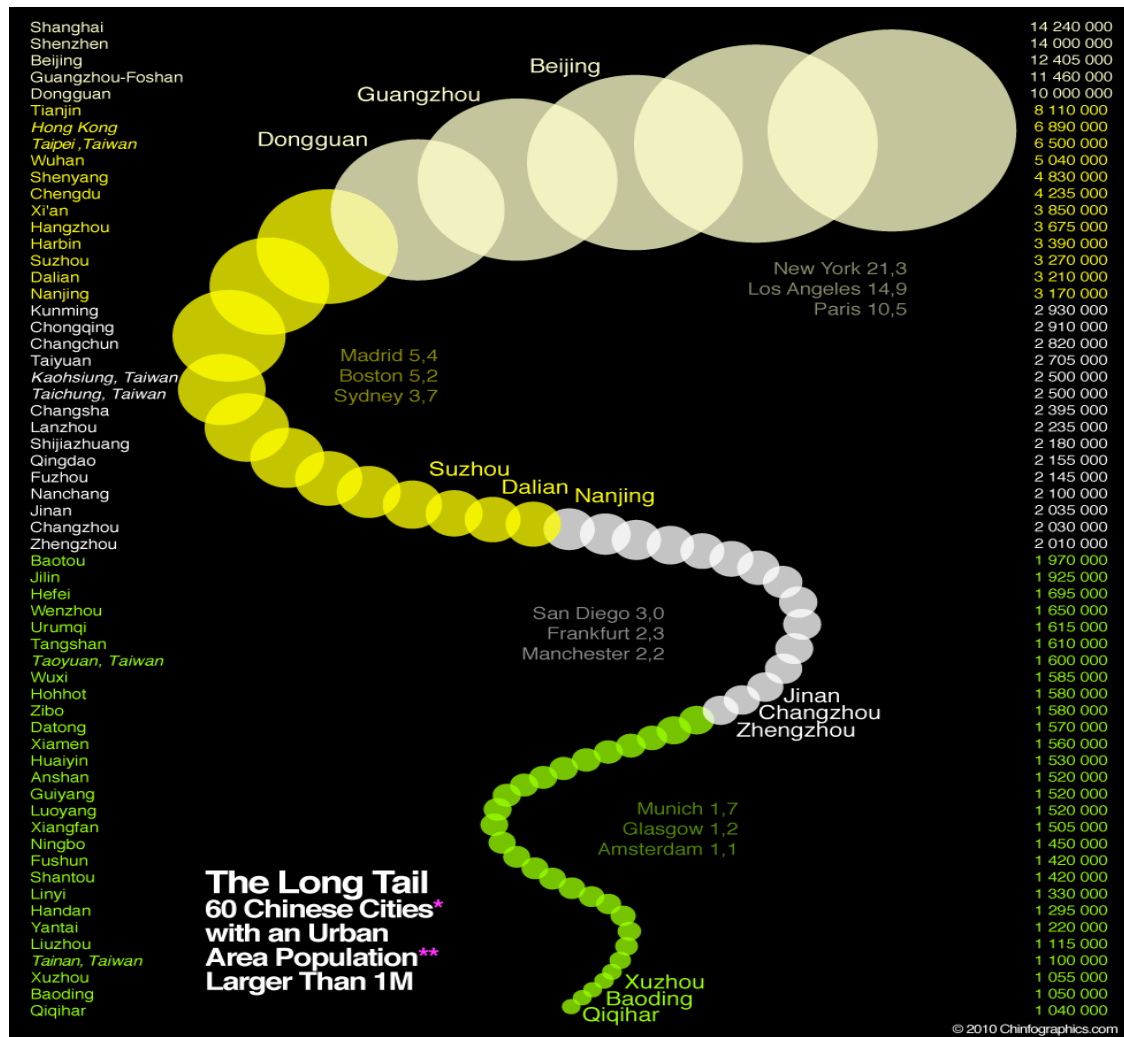
... your leading partner in quality statistics

Literature review: basis of the rank-size rule

Hsu (2008) and others, hypothesize the distribution of variables in central place theory is consistent with Zipf's rule, and proves it.

Literature review: some empirical illustrations

Source:
<http://chinatraveltrends.com/the-long-tail-60-chinese-cities-with-a-population-of-over-1-million>



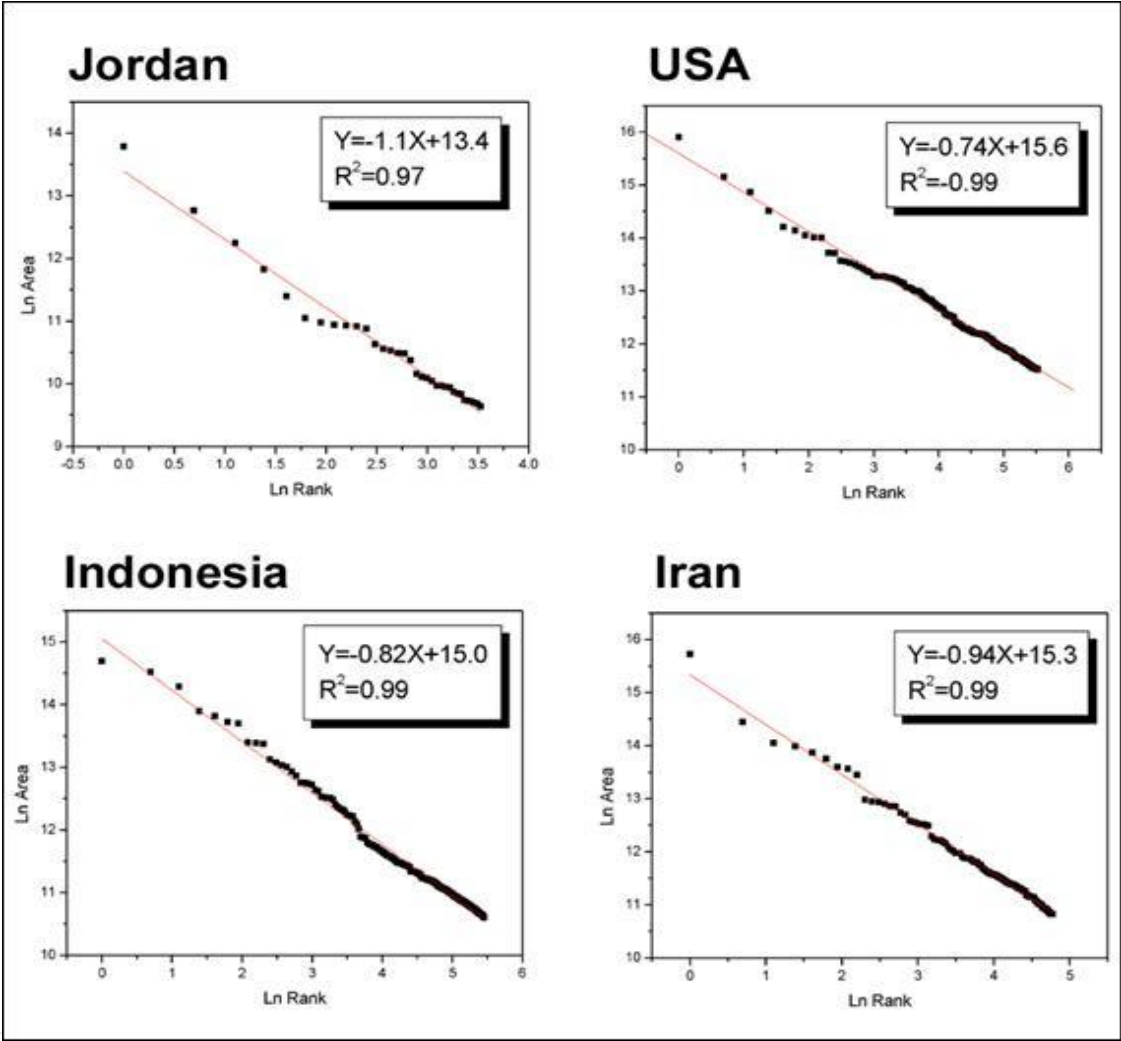
- Cities with a population larger than 10M (5)
- Cities with a population between 3-10M (12)
- Cities with a population between 2-3M (15)
- Cities with a population between 1-2M (28)

The Long Tail - 60 Chinese Cities with an Urban Area Population of 1M+
 By: Chinographics.com © 2010
 Date: 2010 04 20
 Source: www.demographia.com
 (chinese cities 2007, western mostly estimates for 2009)
 *Hong Kong and Taiwanese cities are included
 **Generally the lighted area that can be observed from an airplane at night



•••• your leading partner in quality statistics

Literature review: some empirical illustrations



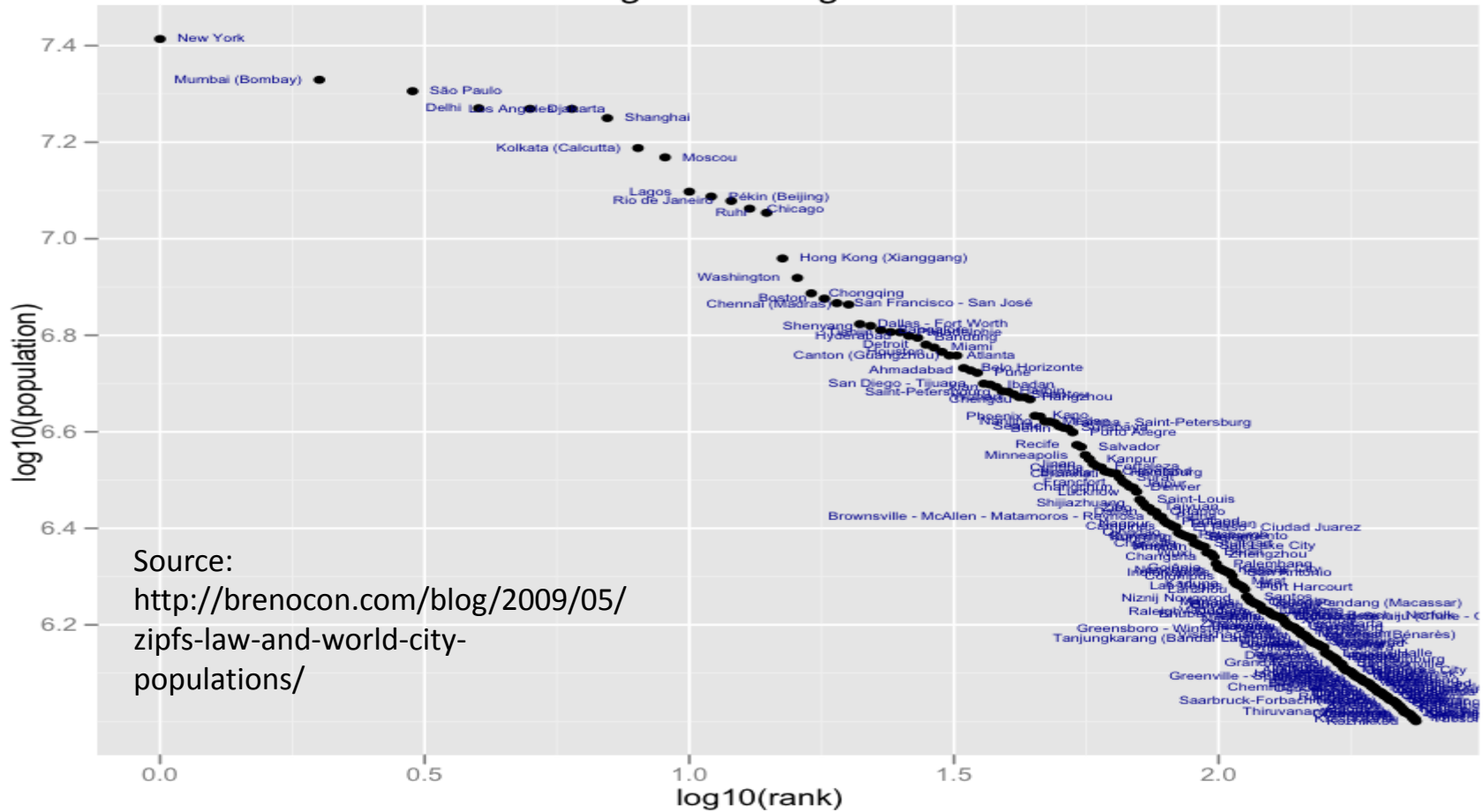
Source: <http://cgu4u-models-in-human-geography2010.wikispaces.com/The+Rank+Size+Rule>



... your leading partner in quality statistics

Literature review: some empirical illustrations

world city populations for 8 countries
log-size vs log-rank



•••• your leading partner in quality statistics

Literature review: basis of the rank-size rule

Author	Variable	Countries	α estimates	R^2
Nota & Song (2008)	Population	US	0,895-0,925	0,989
Giesen & Suedekum (2009)	Population	Germany	0,929-1,365	0,932-0,992
Kyriakidou, Michalakelis, et al (2011)	Population	Greece	0,875-1,181	0,878-0,965
Tanaka & Hatsukano(2011)	Employees (All)	Cambodia	1,33	0,994
	Company size	Cambodia	0,92-1,31	0,917-0,997
	Sector	Cambodia	1,30-1,38	0,992-0,995
	Ownership	Cambodia	1,24-1,37	0,983-0,994
Hinloopen & Marrewijk(2007)	Ballasa trade index (1970-1997)	166 countries	0,849-1,031	0,968-0,991

Treatment of key variables: available data

Actual data from Censuses (large surveys) is used as base data for the initial ranking of local municipalities w.r.t. population size, employment and unemployment.

Annual mid-year population estimates at national & district municipality levels are used to estimate local municipality population sizes and ranking in non-Census years.

Available QLFS estimates at national & provincial levels are used to estimate (and rank) employment & unemployment at local municipality level in non-Census years.

For the GDP, data collected on local government income (excluding grants) in the annual Financial Census of Municipalities (FCM) is used as **base** data for the initial ranking of local municipalities.

Available annual GDP estimates at national and provincial levels are used to estimate (and rank) GDP at local municipality level.

Treatment of key variables: available data

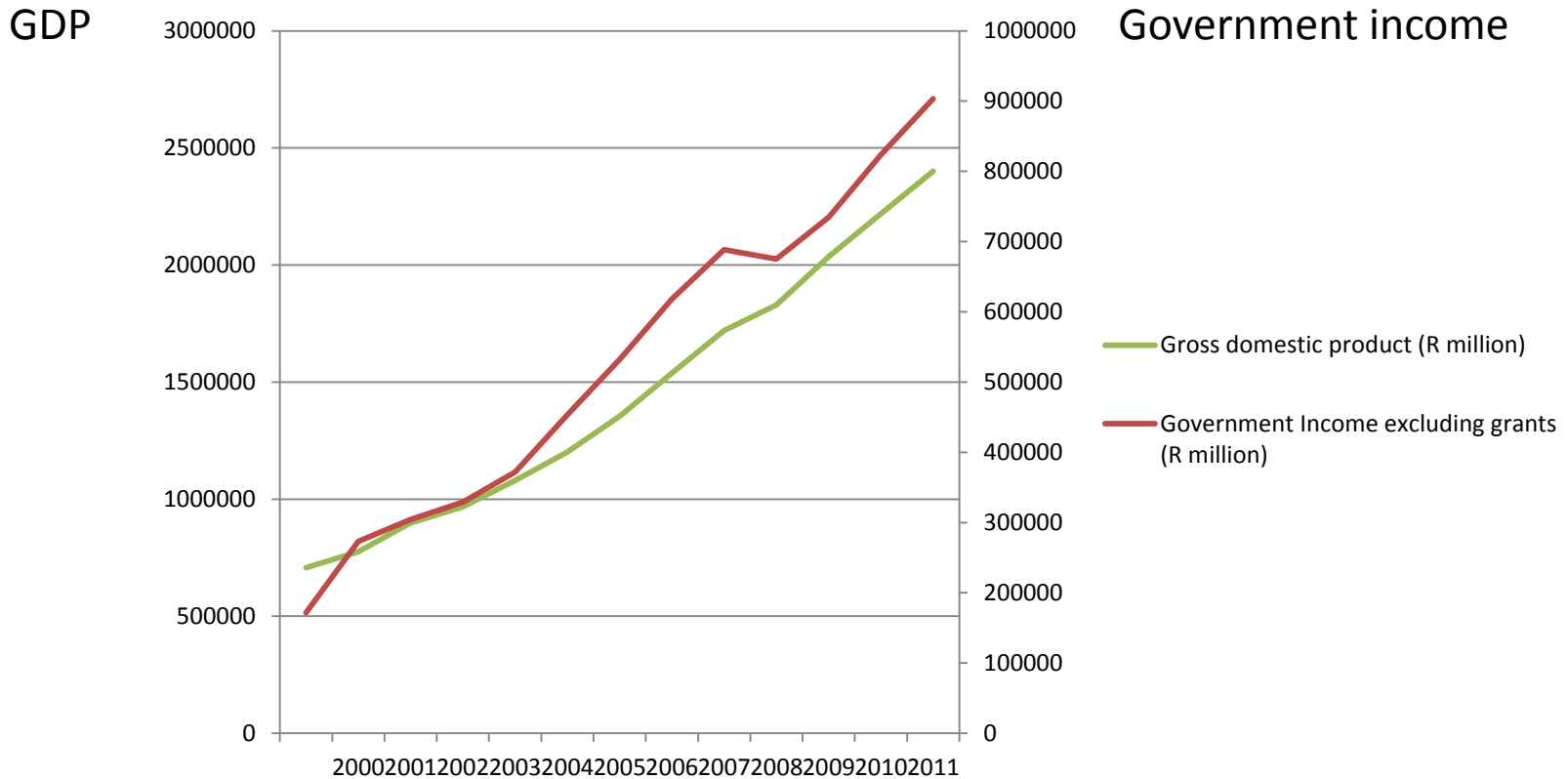
National data confirms a strong positive relationship between GDP and Government income (excluding grants), i.e.:

an increase (decrease) in GDP is associated with an increase (decrease) in government income from taxes and various other government income generating sources.

The study assumes a similarly strong positive relationship at local municipality level, and adopts the ranking of local municipalities based on municipal income.

Treatment of key variables: available data

GDP vs Government income (2000-2011): annual nominal data



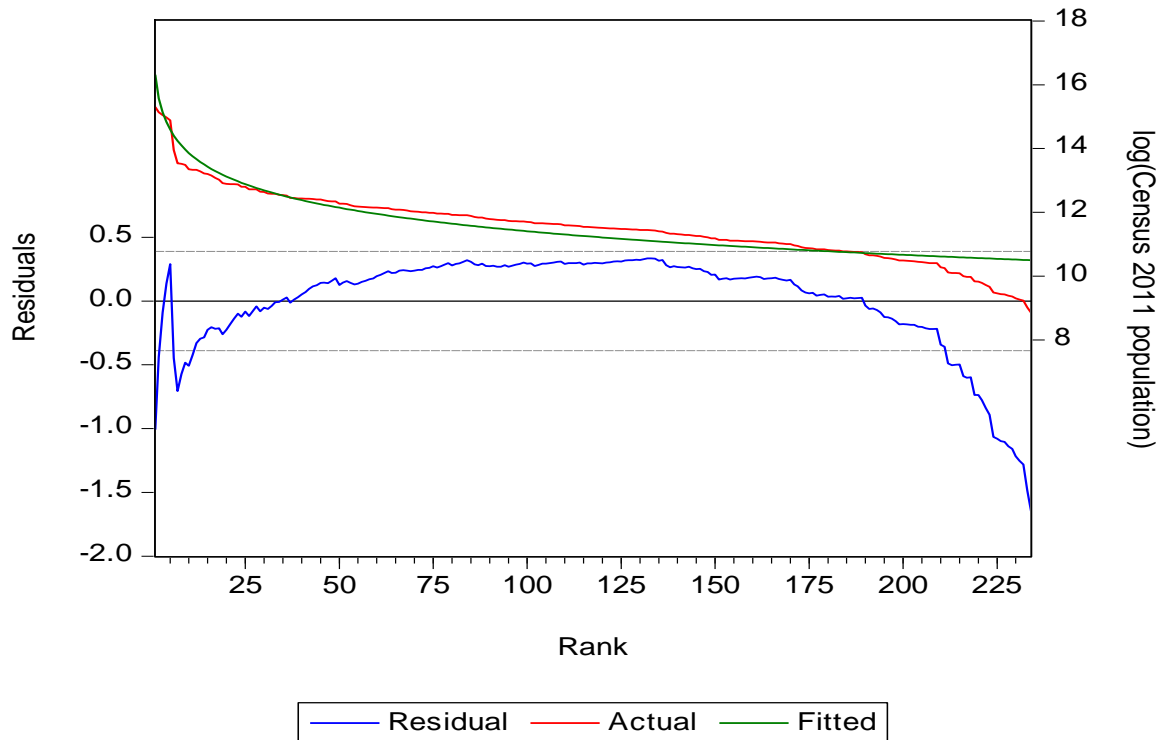
... your leading partner in quality statistics

Zipf's rule results

Variable	α	t-value	R ² adjusted	SER
Actual base data				
Population 2011	1.0651	40.1625	0.8737	±0.388254
Employment 2011	1.1706	62.8670	0.9443	±0.272596
Unemployment 2011	1.1985	45.3137	0.8980	±0.387218
FCM 2011	1.7708	44.0025	0.8925	±0.589156
Generated estimates				
Population 2013	1.0651	40.1625	0.8874	±0.388254
Employment 2012	1.1706	62.8646	0.9443	±0.272607
Unemployment 2012	1.1985	45.3095	0.8980	±0.387258
GDPM 2011	1.7708	44.0025	0.8925	±0.589156
Generated estimates adjusted per district municipality or provincial data releases				
Population 2013	1.0666	40.0966	0.8734	±0.389414
Employment 2012	1.1404	59.2016	0.9376	±0.282018
Unemployment 2012	1.1594	52.6938	0.9226	±0.322124
GDPM 2011	1.7657	38.2582	0.8626	±0.675642

Rank-size rule results

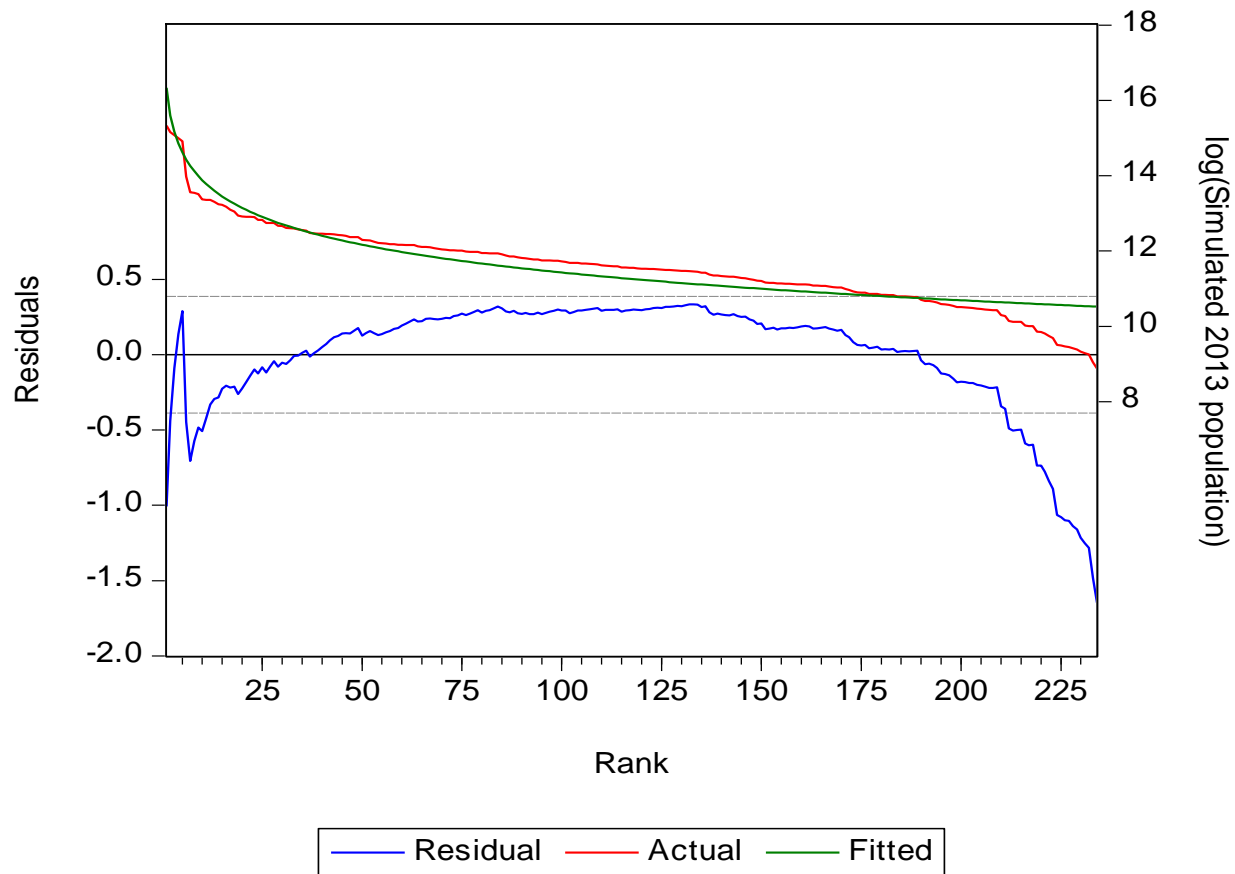
Figure 5.1(i): Census 2011 population



The bottom tail of the curve is usually removed from the sample because of the insignificance and randomness of such centres. They are retained in this study for exploration and completeness.

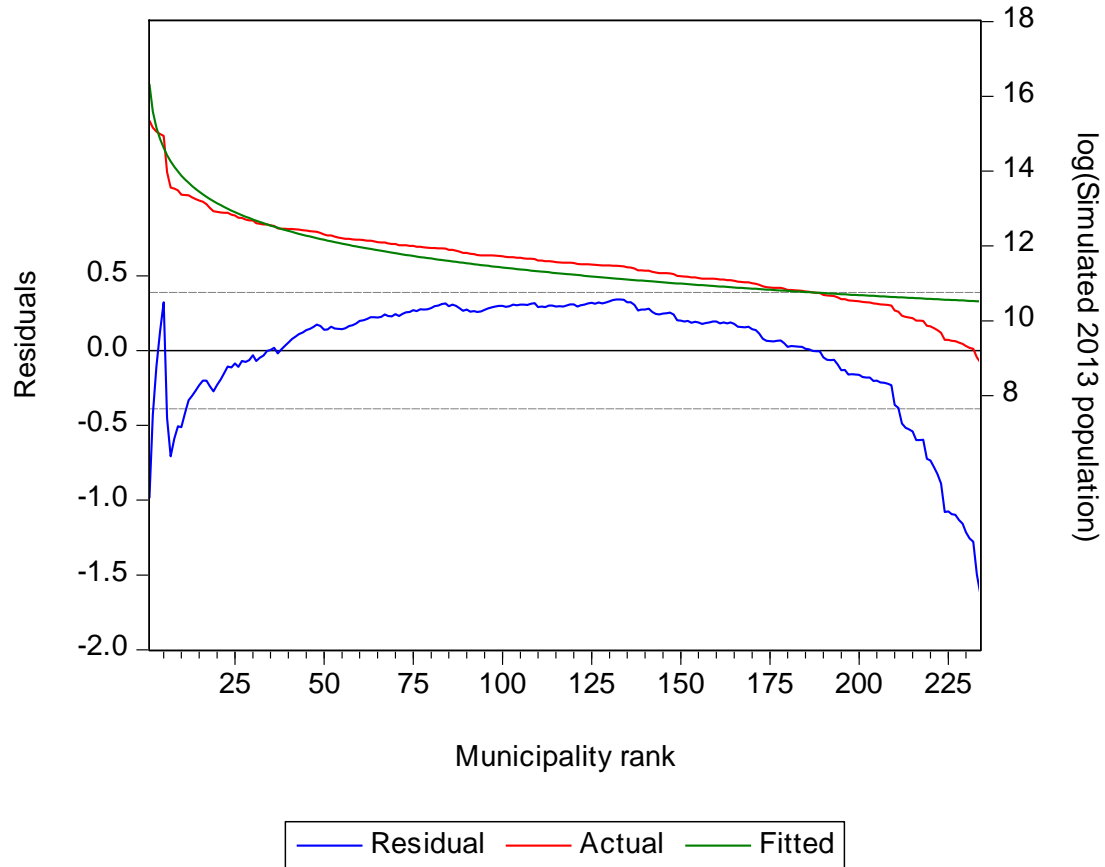
Rank-size rule results

Figure 5.1(ii): Simulated 2013 population



Rank-size rule results

Figure 5.1(iii): Simulated 2013 population adjusted per 2013 district municipality data



Ex ante performance

How realistic are the estimates?

Generated estimates for 2011 would have been provided around February/March 2012 to policymakers

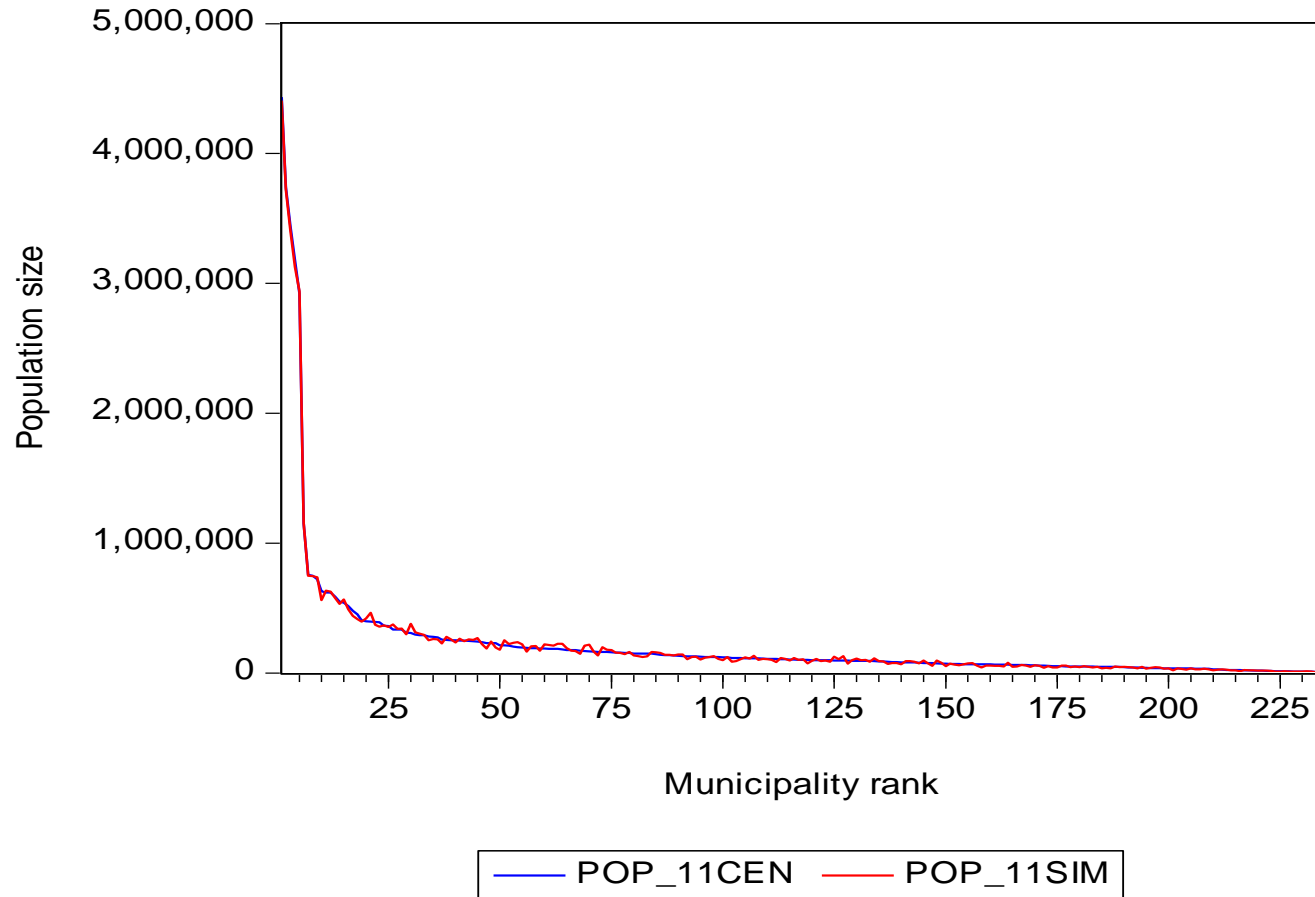
Census 2011 results were released in October 2012



Would 2011 generated estimates closely resemble Census 2011 results?

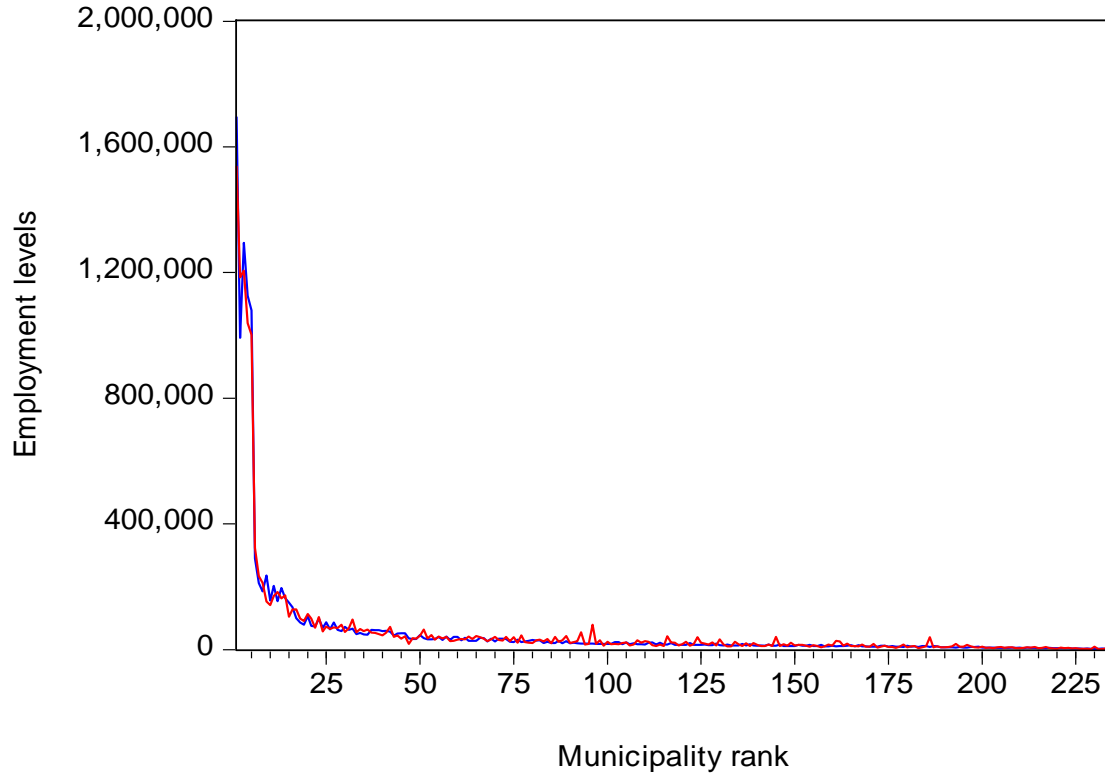
Ex ante performance

Figure 6.1(i): Census 2011 population versus Simulated 2011 population



Ex ante performance

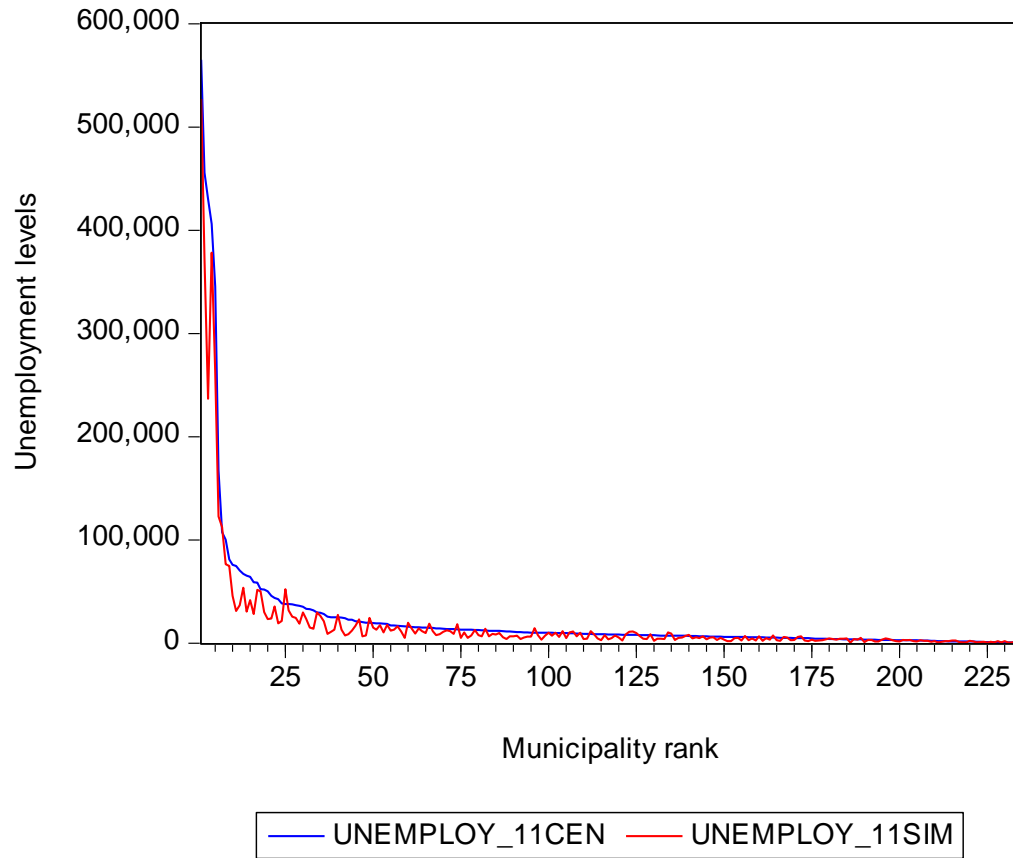
Figure 6.1(ii): Census 2011 employment vs Simulated 2011 employment



— Census 2011 employment
— Simulated 2011 employment

Ex ante performance

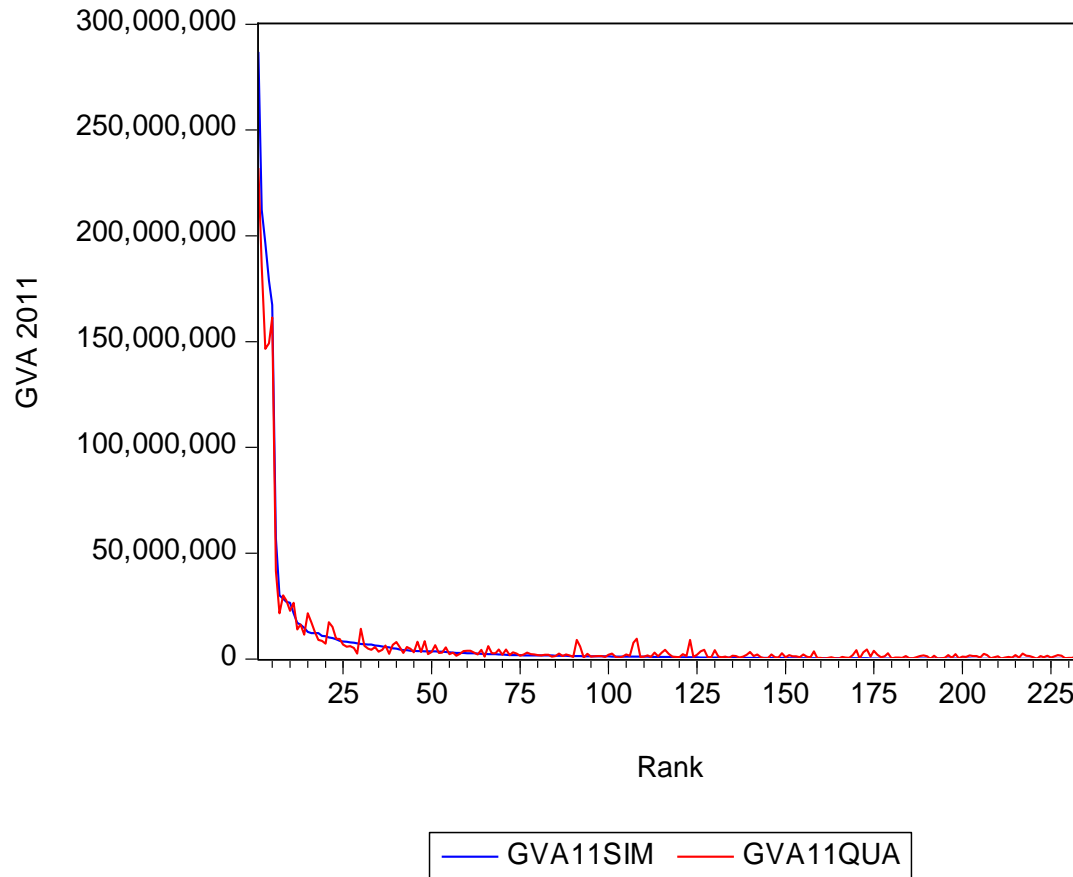
Figure 6.1(iii): Census 2011 unemployment versus Simulated 2011 unemployment



Ex ante performance

$H_0: \mu_1 = \mu_2$	t-test	Probability	ANOVA F-test	Probability
Population	0,016980	0,9865	0,000288	0,9865
Employment	-0,056487	0,955	0,003191	0,955
Unemployment	1,035364	0,3010	1,071979	0,3010

On GDP: contrast with private sector estimates



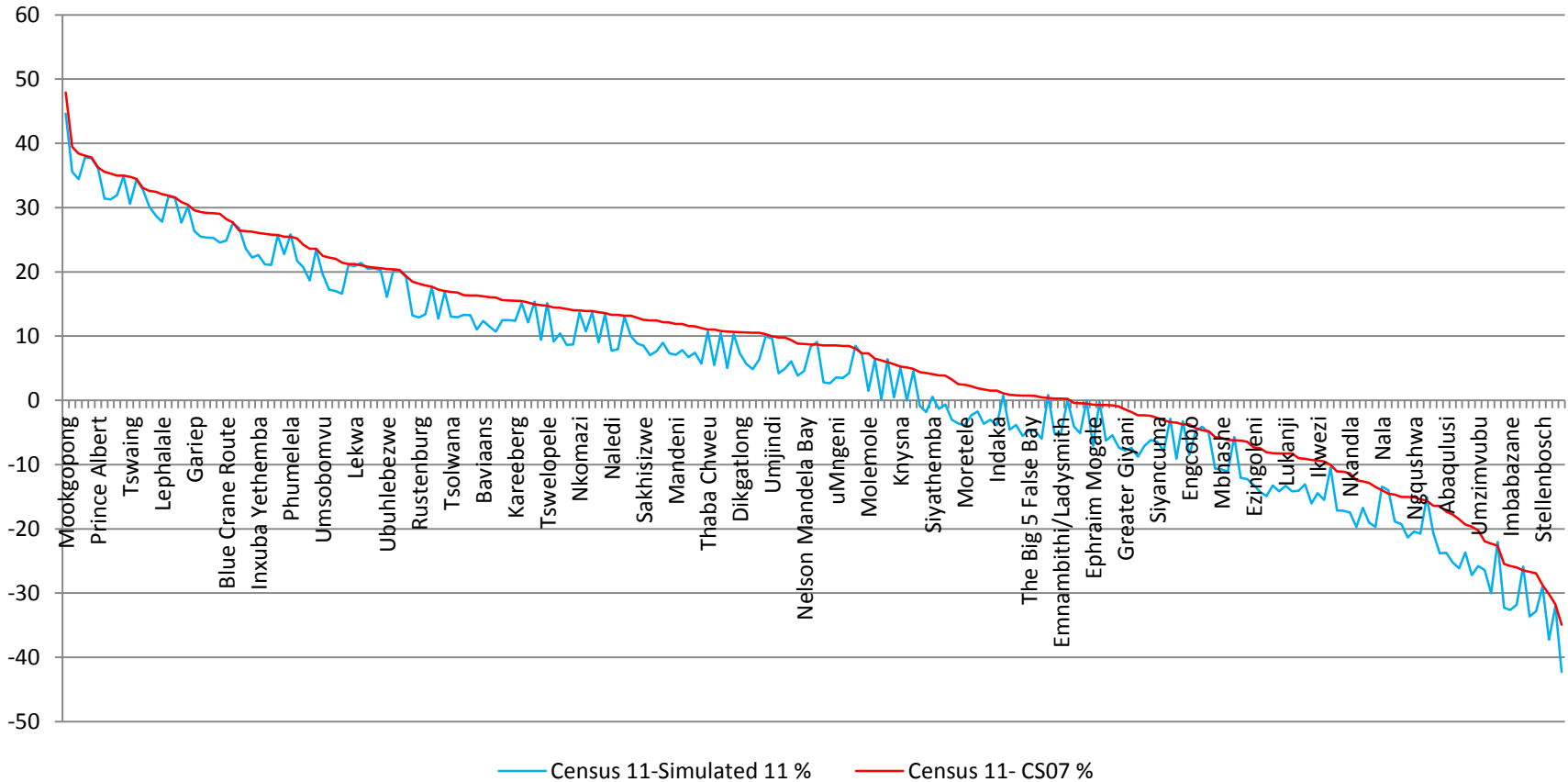
The source of Quantec's 2011 national GVA (at R1 690 166 051 825) remains unclear, but it approximates Stats SA's 2011 GVA of R1 692 724 000 000 at basic prices in constant 2005 prices.

On GDP: contrast with private sector estimates

$H_0: \mu_1 = \mu_2$	t-test	Probability	ANOVA F-test	Probability
GVA 2011	0.004145	0.9967	0.000017	0.9967

Ex ante performance: micro look at deviations

Population: simulated vs actual changes



... your leading partner in quality statistics

Ex ante performance: micro look at deviations

Major reasons for deviations

Use of Community Survey 2007 as a base, despite its deficiencies as a sample. Experiments with prior Censuses (2001) yields better results.

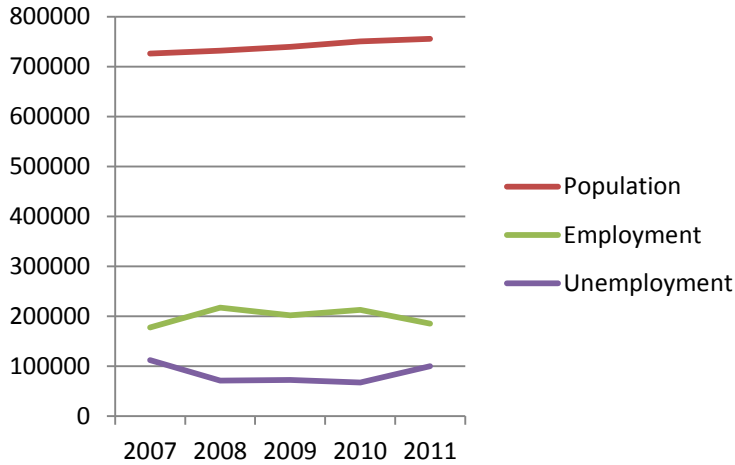
This study has included all local municipalities. Rank-size rule studies usually focus on the top-end of the curve, and completely discard the bottom-end from the sample due to its insignificance and randomness.

Performance over time

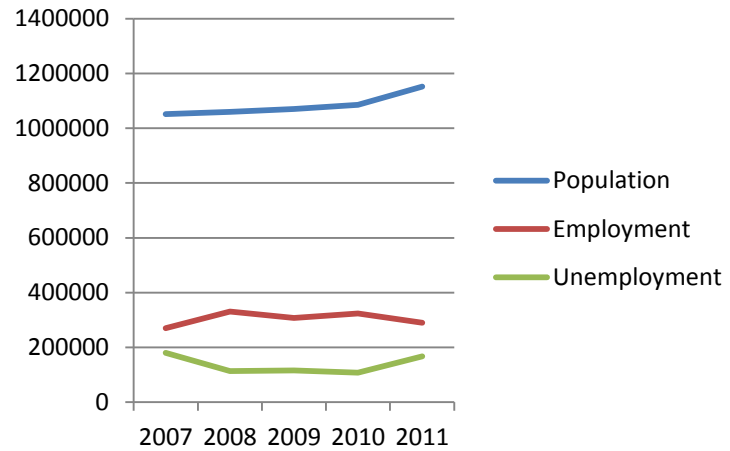
Variable	α	t-value	R ² adjusted	SER
Population 2007 CS	1.1	39.79	0.87	0.41
Population 2008 simulated	1.1	39.71	0.87	0.4
Population 2009 simulated	1.1	39.71	0.87	0.4
Population 2010 simulated	1.1	39.71	0.87	0.41
Population 2011 Census	1.1	40.16	0.87	0.39
Population 2012 simulated	1.1	40.12	0.87	0.39
Population 2013 simulated	1.1	40.14	0.87	0.39
Employment 2007 CS	1.2	50.88	0.92	0.34
Employment 2008 simulated	1.2	51.16	0.92	0.33
Employment 2009 simulated	1.2	50.44	0.92	0.34
Employment 2010 simulated	1.2	49.87	0.91	0.34
Employment 2011 Census	1.2	62.87	0.94	0.27
Employment 2012 simulated	1.1	59.20	0.94	0.28
Unemployment 2007 CS	1.2	42.35	0.88	0.42
Unemployment 2008	1.2	45.16	0.9	0.39
Unemployment 2009	1.2	49.33	0.91	0.36
Unemployment 2010	1.2	50.85	0.91	0.35
Unemployment 2011 Census	1.2	45.31	0.9	0.39
Unemployment 2012	1.2	52.69	0.92	0.32
GDPM 2007 simulated	1.8	37.5	0.86	0.69
GDPM 2008 simulated	1.8	33.69	0.83	0.77
GDPM 2009 simulated	1.8	34.67	0.84	0.76
GDPM 2010 simulated	1.8	36.3	0.85	0.72
GDPM 2011 simulated	1.8	38.26	0.86	0.68

Performance of selected Eastern Cape local municipalities

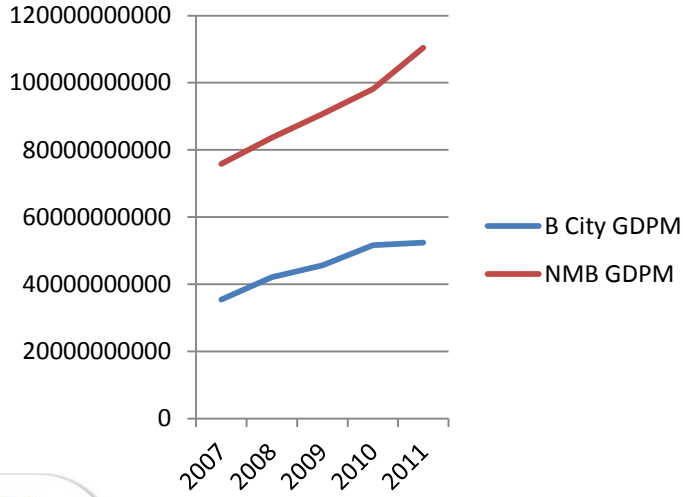
Buffalo City



Nelson Mandela Bay



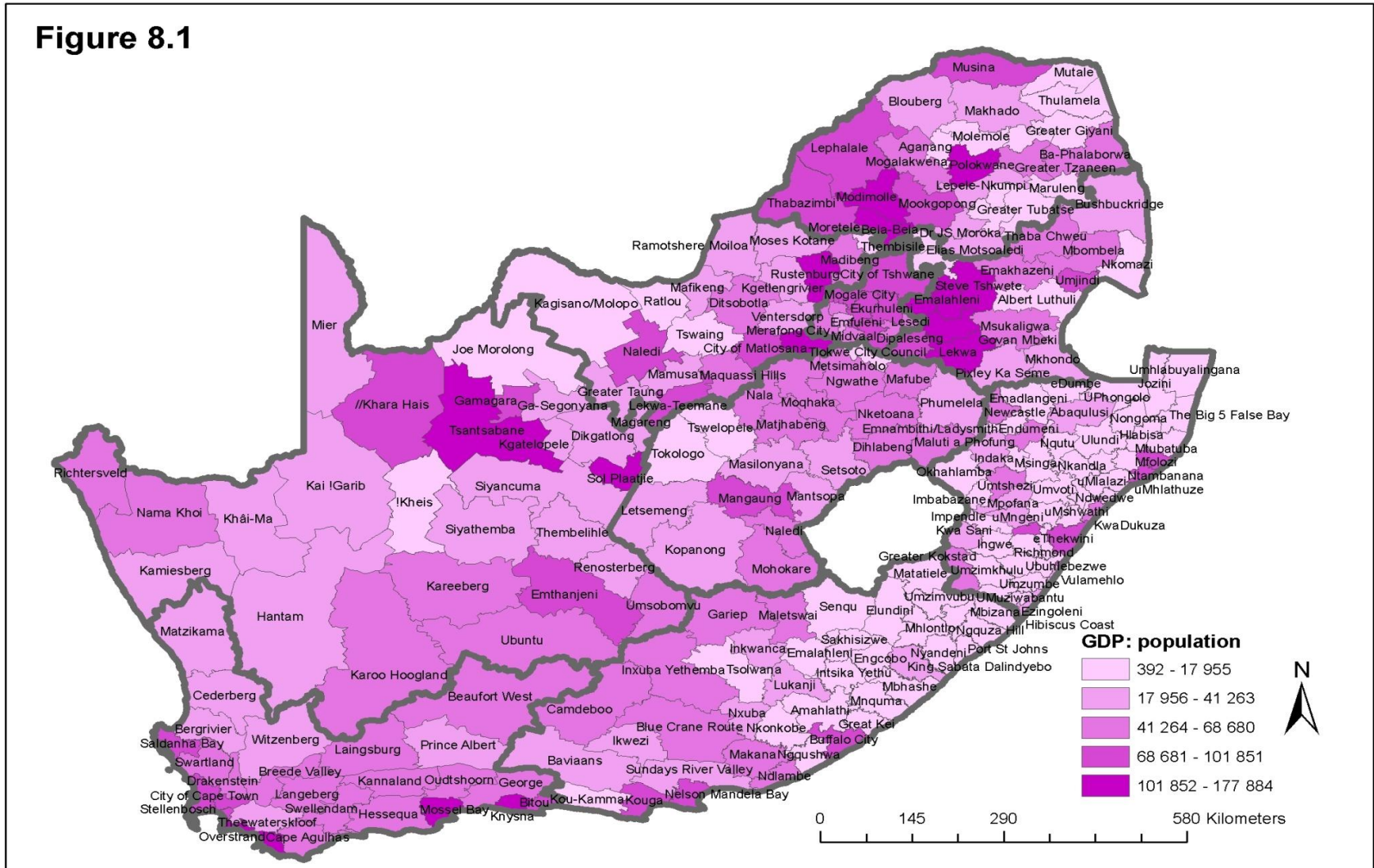
Nelson Mandela Bay & Buffalo City



... your leading partner in quality statistics

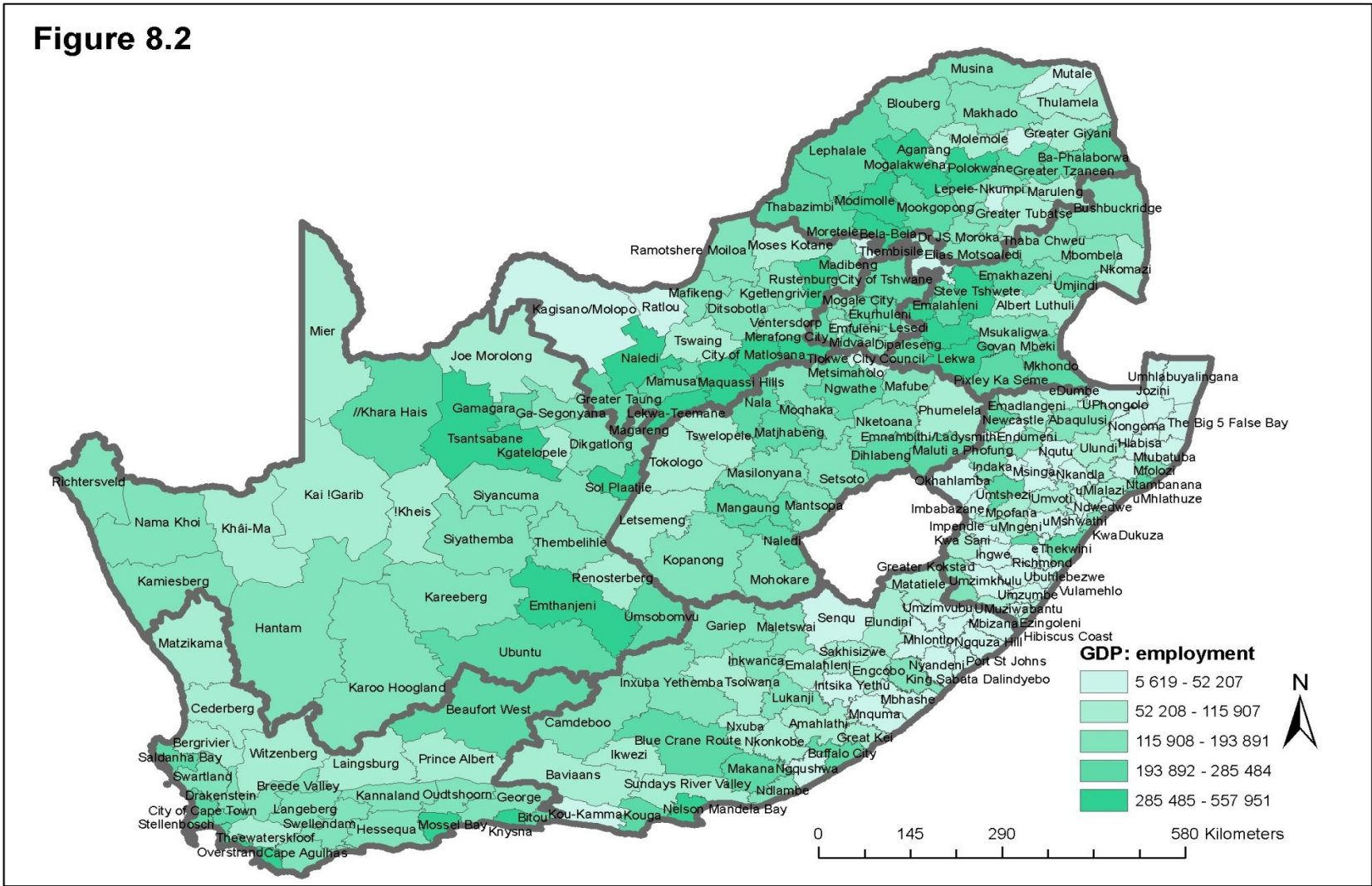
Understanding South Africa

Figure 8.1



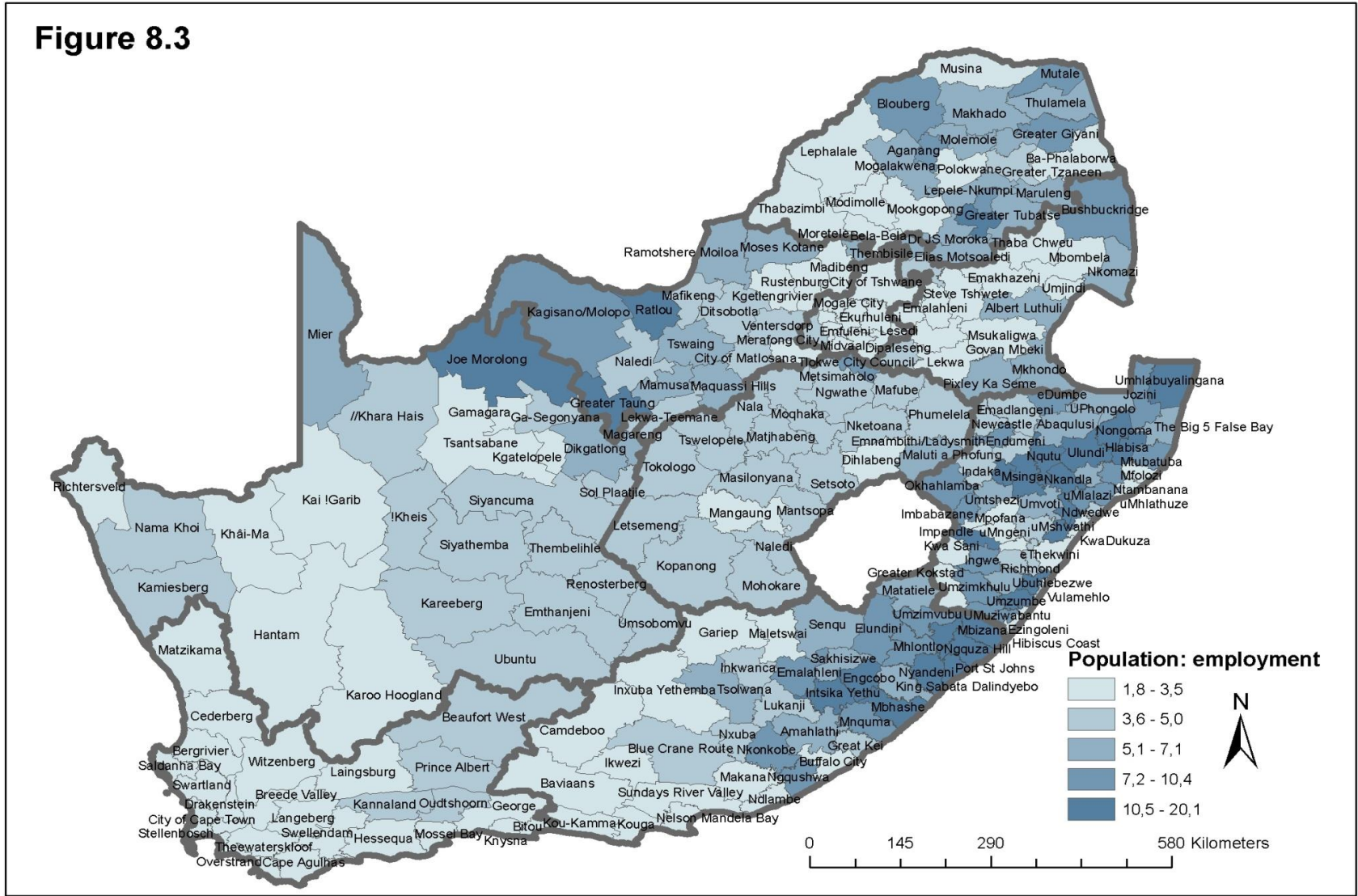
Understanding South Africa

Figure 8.2



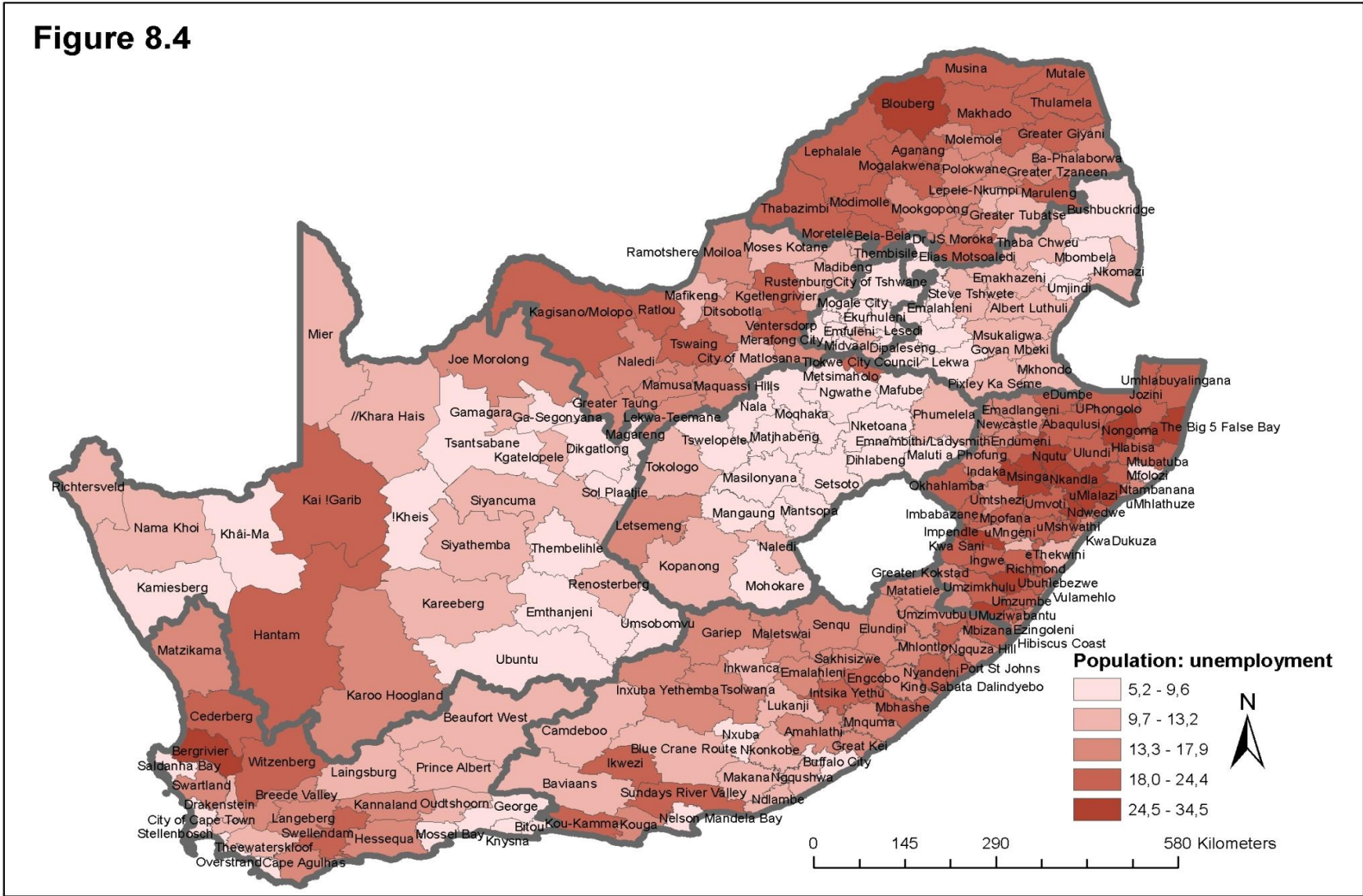
Understanding South Africa

Figure 8.3



Understanding South Africa

Figure 8.4



... your leading partner in quality statistics

Recommendations

Adoption of the approach in the interim, given current budgetary constraints and challenges posed to other small area estimation efforts at Stats SA.

Regular benchmarking and revision after every Census or Community Survey.

The study be conducted from Census 2001 to date, to better comprehend major deviations from Censuses or Community Surveys.

Conduct continuous validation, particularly of outliers, of actual observations on the ground by various experts – economists, demographers, etc.

Estimates be supplied to local municipality planners with an explicit record of estimation errors from prior Censuses or Community Surveys.

Conclusions

This presentation has demonstrated Zipf's rule can help generate useful estimates to bridge the statistics gap in the development of local municipality socio-economic models – i.e. IDPs, SDFs and LEDs.

The framework is robust: it is applicable to a wide range of variables (population size, employment, unemployment, GDP, etc.). Other variables considered crucial for the implementation of the National Development Plan could be explored.

The framework provides estimates that make it possible for policy planners to analyse complex local municipality patterns in the country, i.e. enabling an analysis of more complex relationships among different municipalities

Thank you very much!



Statistics
South Africa



•••• your leading partner in quality statistics