


Labour Markets and Social Policy

A Comparison of the Old and New Weights of the LFS Data: 2000-2005

C. Woolard
I. Woolard

December 2005

HSRC
employment growth & development initiative
innovative employment strategies

A COMPARISON OF THE OLD AND NEW WEIGHTS OF THE LFS DATA: 2000-2005

Chris Woolard

Research Associate

Department of Chemistry
Nelson Mandela Metropolitan University

Ingrid Woolard

Chief Research Officer

Southern Africa Labour and Development Research Unit
School of Economics, University of Cape Town

HSRC
Human Sciences Research Council

employment growth & development initiative

Human Sciences Research Council

December 2005

Produced by: Chris Woolard and Ingrid Woolard

Contact: Dr Miriam Altman
Executive Director, EGDI

E-mail: maltman@hsrc.ac.za

Tel: +27 12 302 2402

Contents

Tables

Figures

Introduction

This note compares the population weights from the Labour Force Survey (LFS) data-sets as originally released by Statistics South Africa (Stats SA) with the re-weighted data-sets released in September 2005. Some conclusions are also drawn about the applicability of the data for the generation of nuanced statistics.

In general it is found that the trends in broad variables over time are much smoother using the re-weighted data than those obtained using the original data. This, however, overlies some subtle differences in the two sets of data.

The data used were from LFS2, LS4, LFS6, LFS8, LFS10 and LF11 (September 2000, September 2001, September 2002, September 2003, September 2004 and March 2005).

In the discussion that follows the original weighted data are referred to as 'old' and the re-weighted data as 'new'.

Number of households

Figure 1 – The variation in the number of households with time

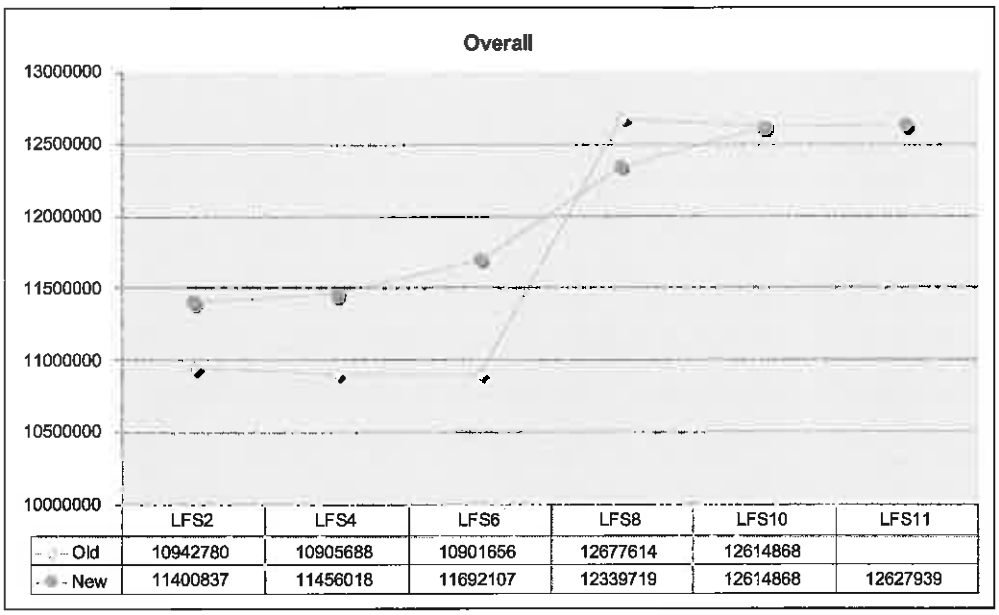
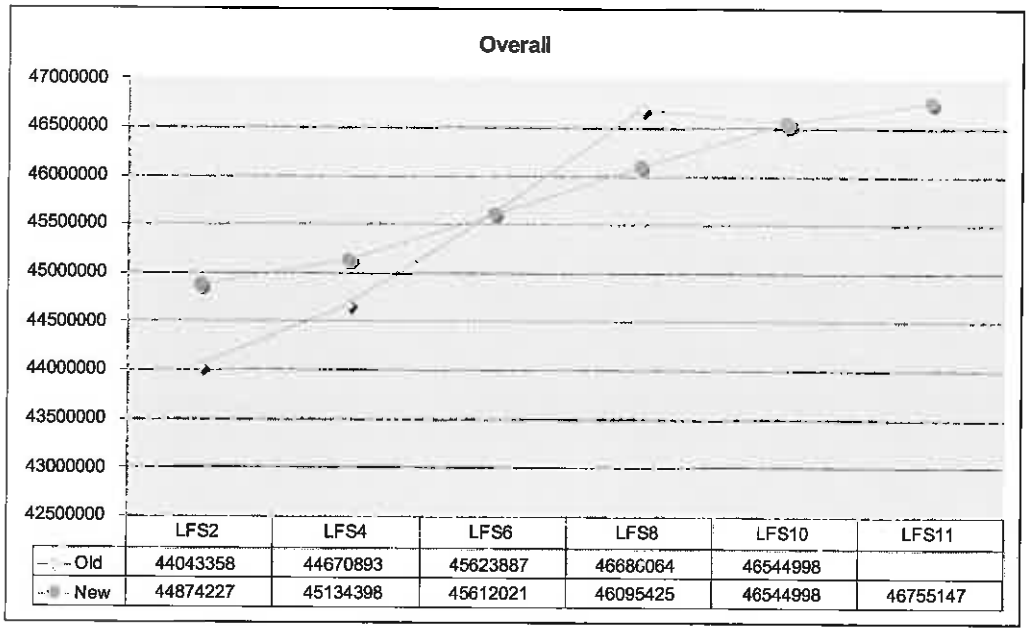


Figure 1 shows the variation in the number of households with time. The two graphs coincide from LFS10 onwards because these data-sets were weighted according to Census2001 from the outset. It is apparent that the growth in the number of households is smoother using the new data. The step discontinuity between 2002 and 2003 is avoided. The variation in provincial household numbers may be found in Appendix 1. With the exception of the Free State, change is smoother with time using the re-weighted data for all provinces. Note that some anomalies have been ameliorated if not removed. See the change in household numbers in the Eastern Cape during the period September 2000 to September 2001 to September 2002.

Number of persons

A similar situation prevails when one looks at the data at a person level. The figure below illustrates the variation in the number of persons with time.

Figure 2 – The variation in the number of persons with time



The variation in provincial numbers may be found in Appendix 2. As before the variation with time is much smoother.

The re-weighted data would suggest a movement of people from rural provinces such as Limpopo, the Free State and The Eastern Cape to the urban centres of the Western Cape and Gauteng. Based on the old data, such movements seem to have occurred primarily from the Eastern Cape and Limpopo. The trend is much less pronounced for the Free State. The table below details the percentage change in population by province for the period September 2000 to September 2004. These numbers correspond to an overall annualised growth rate over the four-year period of 1.39% using the old data and 0.92% using the new data.

Note that it is likely that the population in 2000 is overestimated in the re-weighted data. The year-on-year growth rate to 2001 is just 0.58% but near 1% thereafter, indicating that the start year population is likely too large.

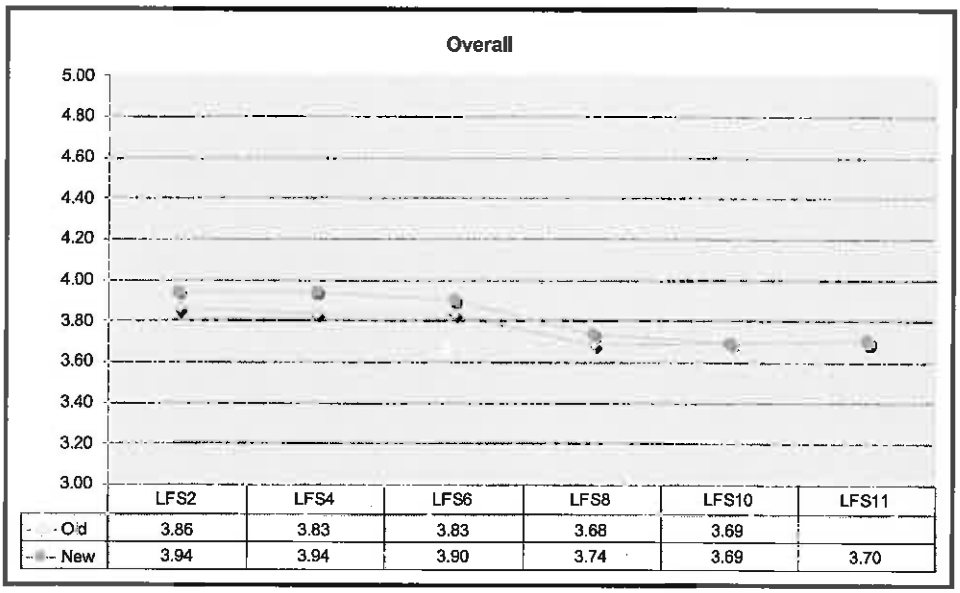
Table 1: Percentage change in population by province (September 2000 - September 2004)

Province	% change Old data	New data
Western Cape	8.3%	8.4%
Eastern Cape	1.7%	0.9%
Northern Cape	2.6%	2.6%
Free State	4.9%	1.4%
KwaZulu-Natal	5.9%	3.0%
North West	5.1%	2.1%
Gauteng	12.4%	7.6%
Mpumalanga	4.3%	2.7%
Limpopo	0.6%	2.1%
Overall	**5.7%**	**3.7%**

Household size

If one calculates a derived variable such as household size, one obtains the following trend.

Figure 3 – The variation in the weighted household size with time



What one observes is that despite the smoother population growth in the new data-sets, there is a greater change in household size implied by the re-weighting. In fact, the data still indicate the improbable step change in household size between 2002 and 2003. This clearly indicates that adjusting macro-scale variables such as population (via weighting) does not necessarily make derived variables more convincing. The provincial variation in household size is contained in Appendix 3. With the exception

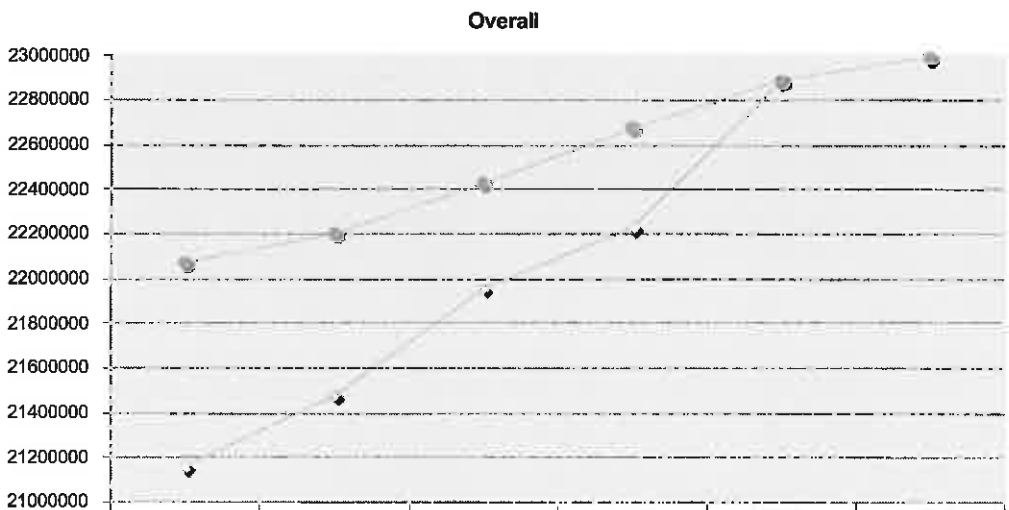
of Limpopo Province (where the variation as indicated by the original data seems more plausible), there is nothing to recommend one data-set over the other.

While the variation in aggregate variables (those in which numbers are totalled) is smoother for the re-weighted data, the variation in derived variables (especially those which are a quotient or percentage) may not necessarily be improved by re-weighting.

Number of men

The figure below shows the variation in the number of men with time.

Figure 4 – The variation in the number of men with time



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	21155966	21474423	21954522	22227448	22885754	
New	22069497	22198455	22423934	22671595	22885754	22990855

In this respect the re-weighted data is a great improvement over the original data. This is because the growth in the male population mirrors the overall population using the re-weighted data (the percentage growth of men during the period under consideration is the same as the overall growth rate). By contrast, there is a significant difference using the old data (1.98% annually for men vs. 1.39% overall). This implies that the female population is growing much slower (at 0.83% annually), which is unlikely. Using the new data, the growth rates are the same by gender.

The provincial data for men may be found Appendix 4. Although there is more random variation than the overall population, the trends are still significantly smoother using the re-weighted data compared with the old data.

The table below shows the percentage change in the male population over the period September 2000 to September 2004. The migration patterns are exaggerated.

Table 2: Percentage change in the male population by province (September 2000 - September 2004)

	% change	
Province	**Old data**	**New data**
Western Cape	9.4%	6.8%
Eastern Cape	5.3%	2.2%
Northern Cape	5.7%	-1.0%
Free State	6.3%	-0.4%
KwaZulu-Natal	9.1%	4.0%
North West	7.6%	2.6%
Gauteng	16.2%	8.0%
Mpumalanga	5.0%	1.0%
Limpopo	0.5%	1.0%
Overall	**8.2%**	**3.7%**

Although there are some discrepancies, the trends are similar to the overall population.

Population by race

The figures below illustrate the variation in population by race.

Figure 5a – The variation in the African population with time

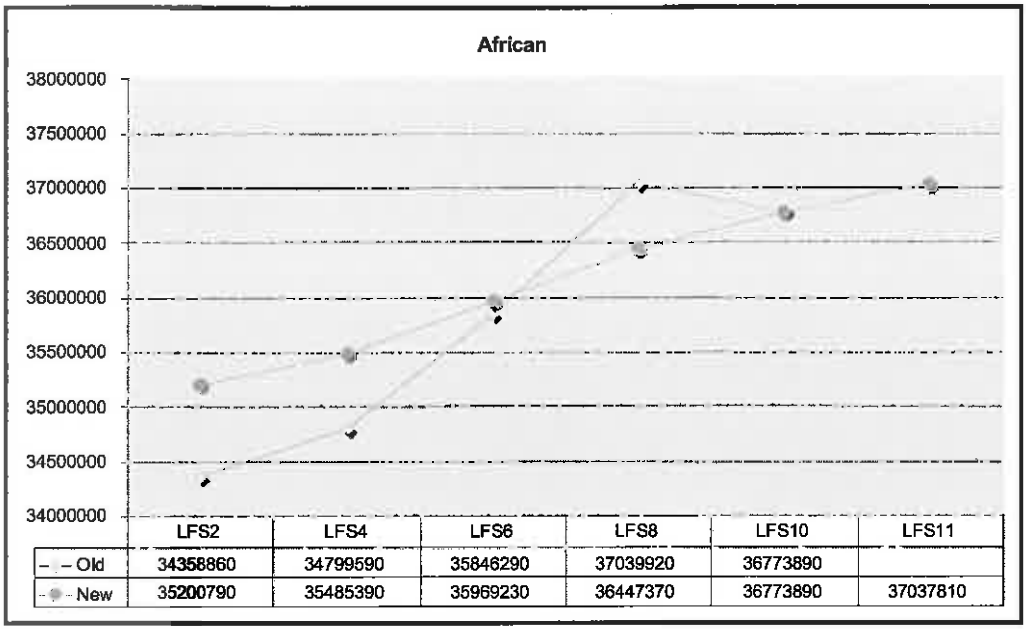


Figure 5b – The variation in the Coloured population with time

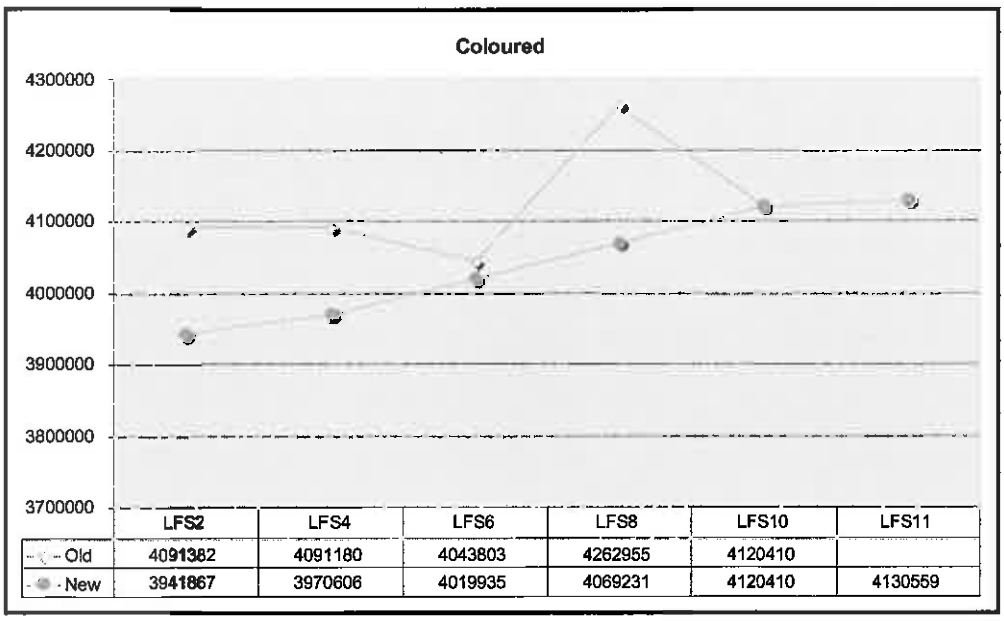
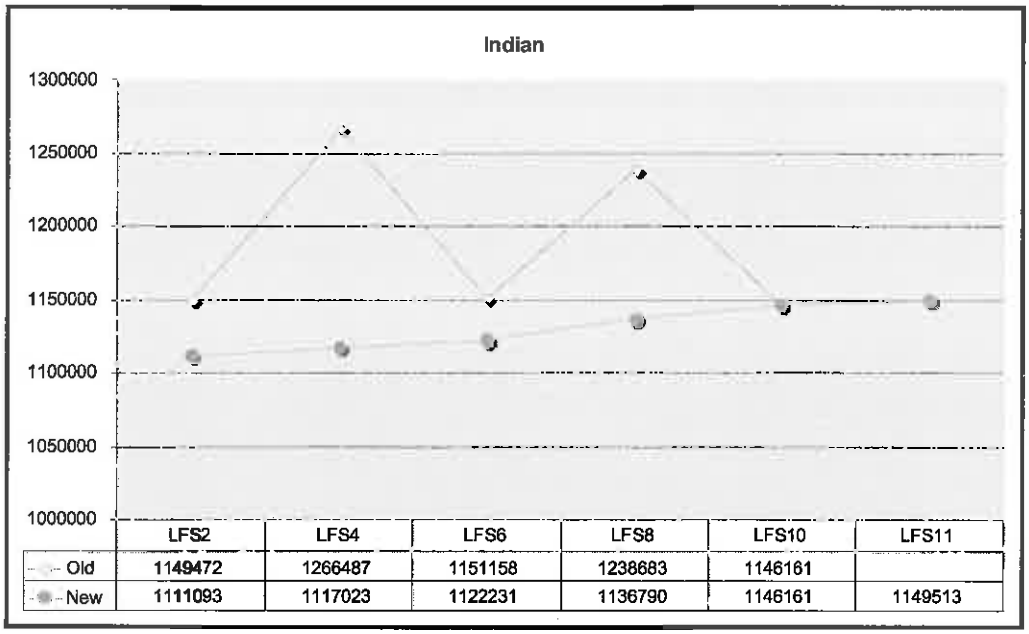
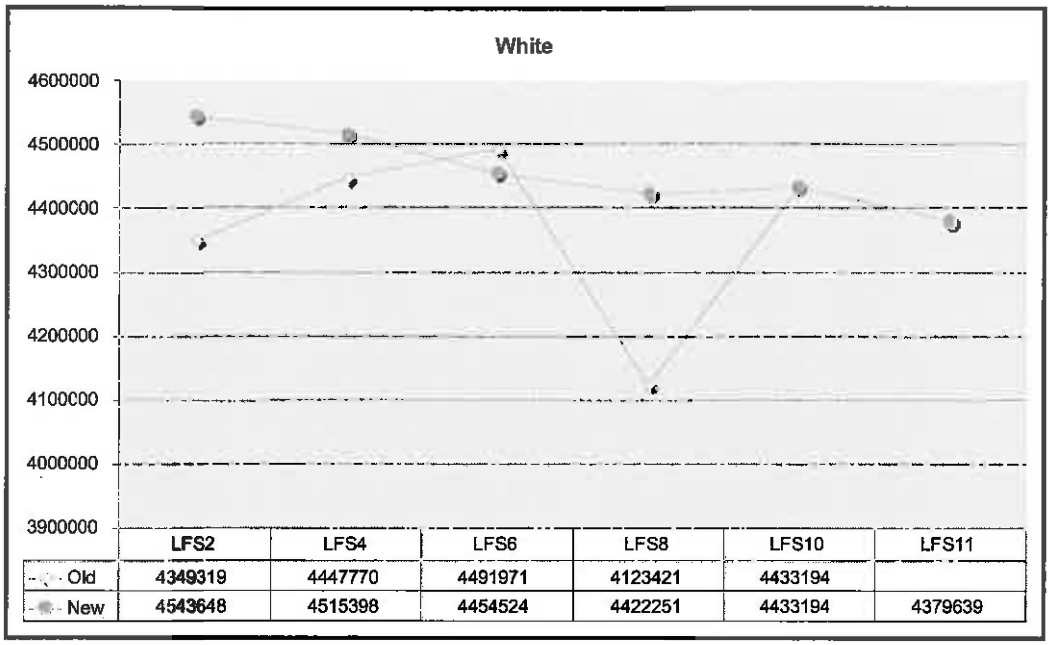


Figure 5c – The variation in the Indian population with time



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1149472	1266487	1151158	1238683	1146161	
New	1111093	1117023	1122231	1136790	1146161	1149513

Figure 5d – The variation in the White population with time



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4349319	4447770	4491971	4123421	4433194	
New	4543648	4515398	4454524	4422251	4433194	4379639

There is appreciably less variation in the different racial populations. Significantly, the White population declines in the period 2000-2002 using the re-weighted data while it increases using the old data.

Persons aged 20-24

To look at the effect of the new data on smaller groups, the population aged 20-24 was determined. Overall it would appear that the re-weighted data fluctuates less than the old data. The variation in the old data is driven by that in Gauteng. The overall picture obscures volatility in the data at provincial data. These are illustrated in Appendix 5. In some cases the trends are in fact more plausible using the old weights (KwaZulu-Natal and Mpumalanga). This indicates that as the data are disaggregated, more variation and consequent uncertainty can be expected.

Figure 6 – The variation in the population aged 20-24 with time

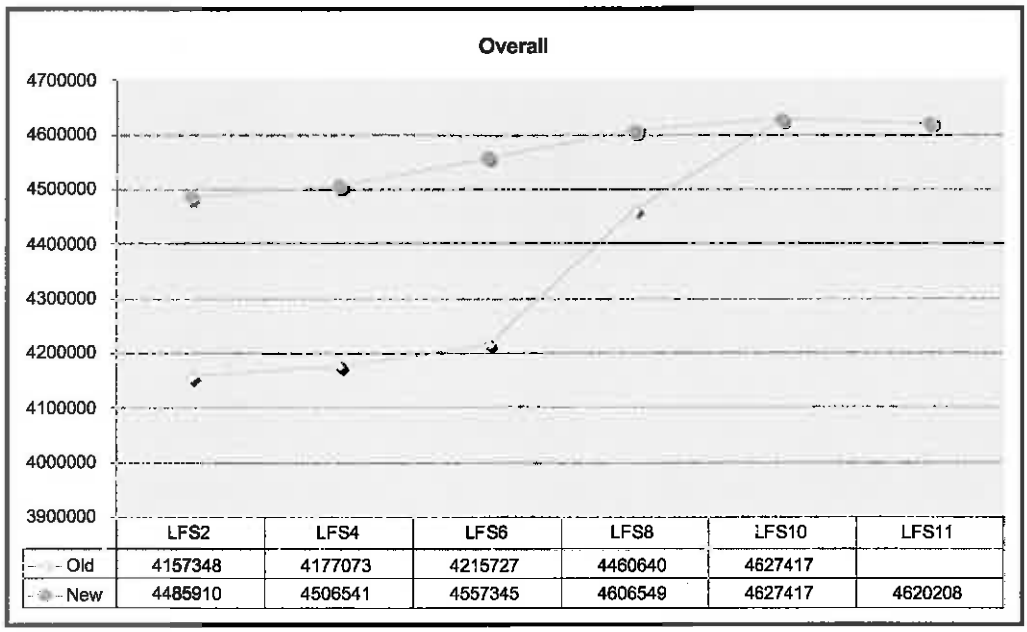


Table 3 – Percentage change in the population aged 20-24 by province (September 2000 - September 2004)

Province	% change 2000-2004 Old data	New data
Western Cape	23.3%	17.0%
Eastern Cape	-7.0%	1.6%
Northern Cape	-12.5%	-15.5%
Free State	8.0%	-5.0%
KwaZulu-Natal	7.4%	1.5%
North West	7.5%	-0.7%
Gauteng	50.9%	4.2%
Mpumalanga	4.8%	3.1%
Limpopo	-2.7%	6.5%
Overall	**11.3%**	**3.2%**

The data in the above table suggest that the discrepancies between the two data-sets are much larger when using smaller groups. For example, note the massive difference in growth of the population aged 20-24 in Gauteng using the two sets of weights. This is likely to have significant consequences for labour market studies which attempt to look at small segments of the population.

A further subtlety is the effect on the overall age distribution. Below are age pyramids for September 2003 (old and new data) and March 2005. Ironically, the age pyramid using the old weights in 2003 better matches that obtained in 2005. This indicates some discrepancy between the overall trend in population and the trends at lower levels. Note that the years (2003 and 2005) were chosen randomly. The age pyramids were not chosen to suggest that the age distribution is better using the old weights in all samples sets but rather that in particular years one or the other set may be better than the alternative.

Figure 7a – Age pyramid in September 2003 using old weights

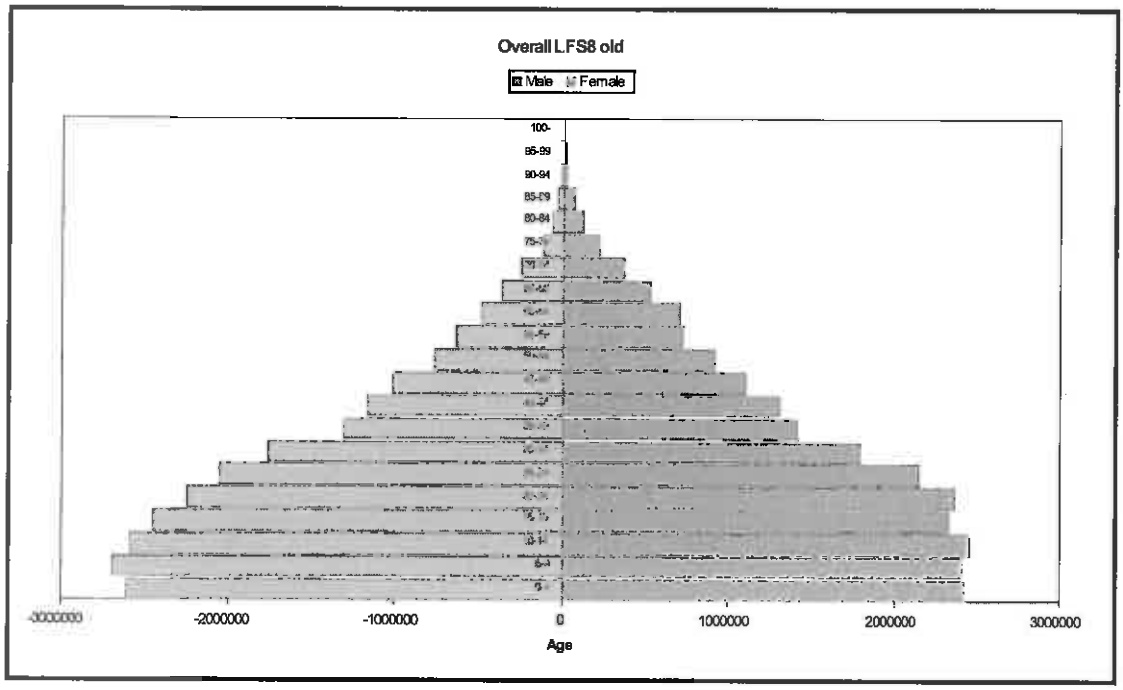


Figure 7b – Age pyramid in September 2003 using new weights

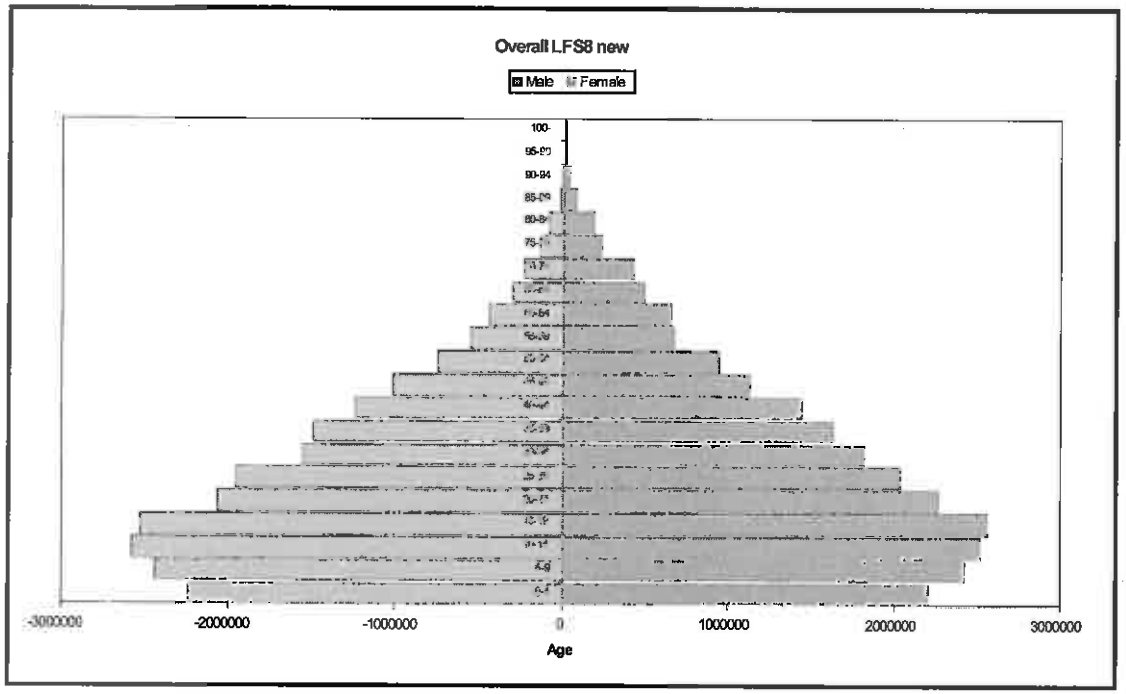
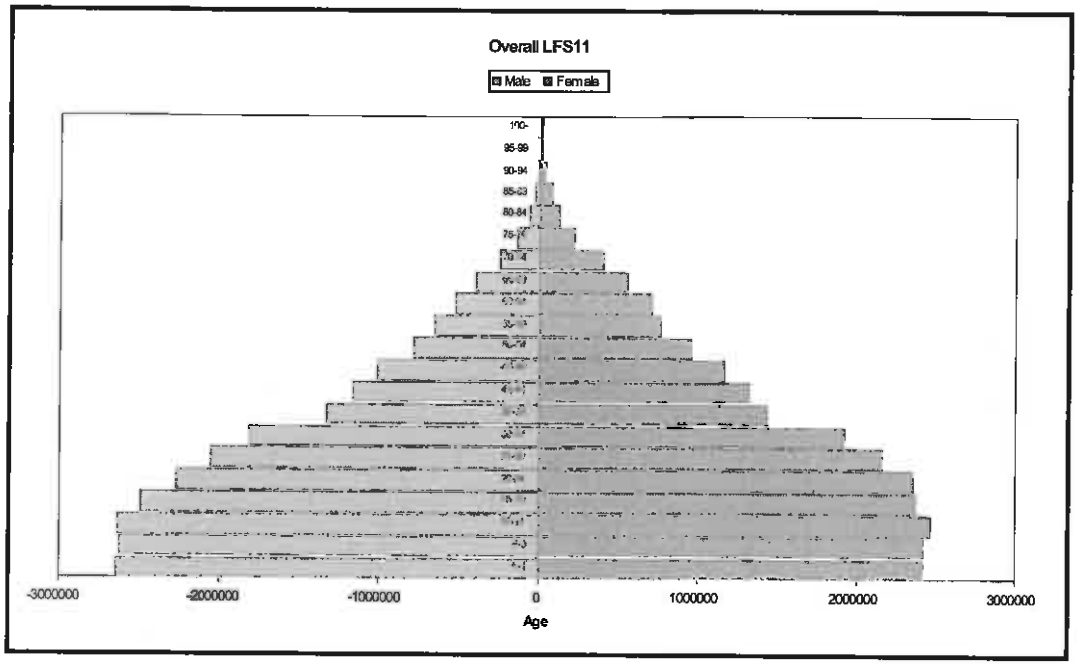


Figure 7c – Age pyramid in March 2003



The effect of re-weighting on uncertainty

One of the consequences of re-weighting the data is that the uncertainties associated with any estimation have increased. In Table 4, the mean weights are reported as an unweighted average of the household weights.

Table 4 – Mean household weights using the old and new data

		LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Original weights	**Mean**	411	398	411	472	441	
	S.D.	283	249	260	253	472	
	Min.	37	50	46	5	7	
	Max.	2,342	1,801	2,430	3,432	8,803	
Re-weighted	**Mean**	428	419	442	460	441	438
	S.D.	414	344	375	384	472	478
	Min.	9	16	9	6	7	7
	Max.	16,975	12,080	17,903	8,036	8,803	11,223

While the mean of the weights are not significantly different (this is to be expected), the standard deviations of the weights using the re-weighted data are significantly larger. This is also illustrated by the maximum weights used. The very high numbers for the maximum would not have been anticipated from the original stratification.

The mean of the person weights are reported in Table 5.

Table 5 – Mean person weights using the old and new data

		LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Original weights	**Mean**	418	420	445	473	424	
	S.D.	230	236	247	242	423	
	Min.	47	54	54	5	7	
	Max.	1,667	1,367	1,396	3,434	88,03	
Re-weighted	**Mean**	426	424	445	467	424	422
	S.D.	347	314	342	358	423	442
	Min.	9	16	9	6	7	7
	Max.	16,975	12,080	17,903	8,036	8,803	11,223

As with the household weights, the standard deviation is much larger using the re-weighted data. Note that in the re-weighted data-set, person weights are the same as household weights but this is not the case with the original data.

The consequence of the larger standard deviations is that uncertainties associated with any calculated quantities will be larger.

To illustrate this, the average household expenditure was estimated using LFS8 data. All households who did not know or refused to answer were dropped. A uniform expenditure distribution was applied to the bottom expenditure category. All other categories were assumed to have a pareto distribution and households were awarded the expected values for each category: R0-R399 (R200); R400-R799 (R577); R800-R1,199 (R987); R1,200-R1,799 (R1,480); R1,800-R2,499 (R2,131); R2,500-R4,999 (R3,607); R5,000-R9,999 (R7,214); and >R9,999 (R14,427).

The mean expenditures, standard deviations and 95% confidence intervals are reported in Table 6.

Table 6 – Mean household expenditure using the old and new data

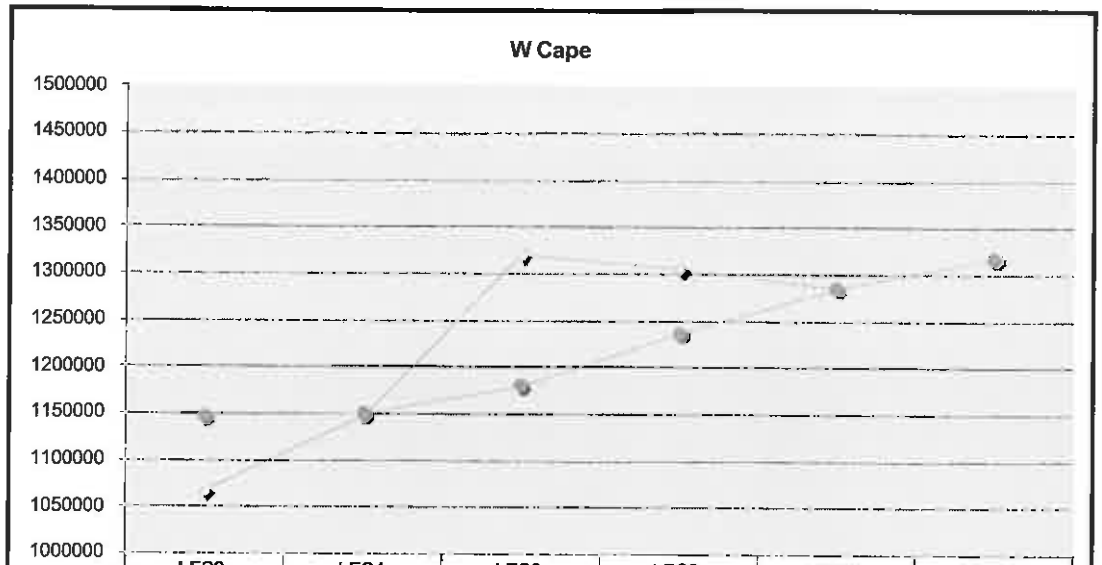
	Mean (/R)	S.D. (/R)	95% low (/R)	95% high (/R)
Old data	1,686	39	1,610	1,762
New data	1,756	50	1,659	1,853

Although the means are different, note the 25% increase in the standard deviation and the consequent increase in the size of the confidence interval.

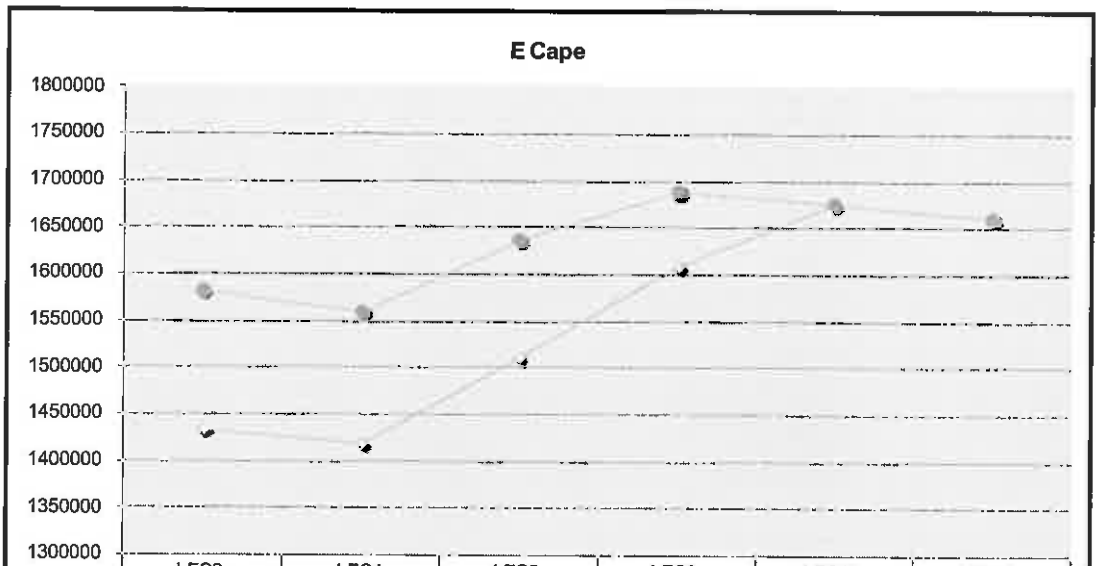
Conclusion

The trends in derived variables are much more plausible when the re-weighted data-set is used. This is because all changes occur much more smoothly. This re-weighting does, however, come with the consequence that the confidence intervals are significantly enlarged.

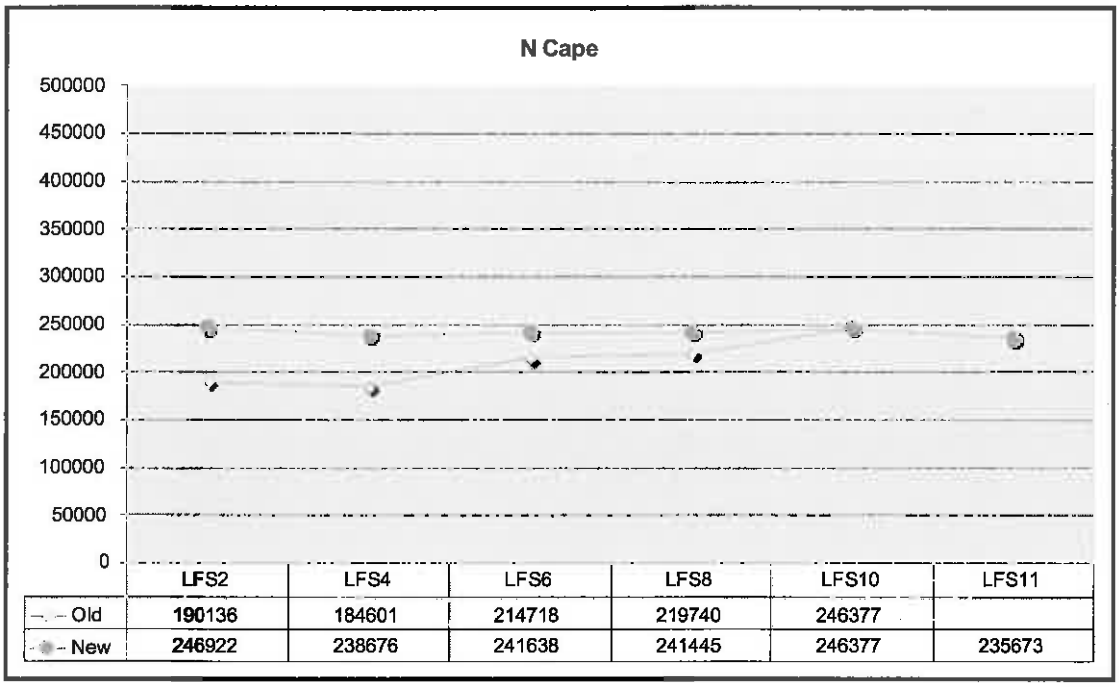
Appendix 1 – Variation in household numbers by province



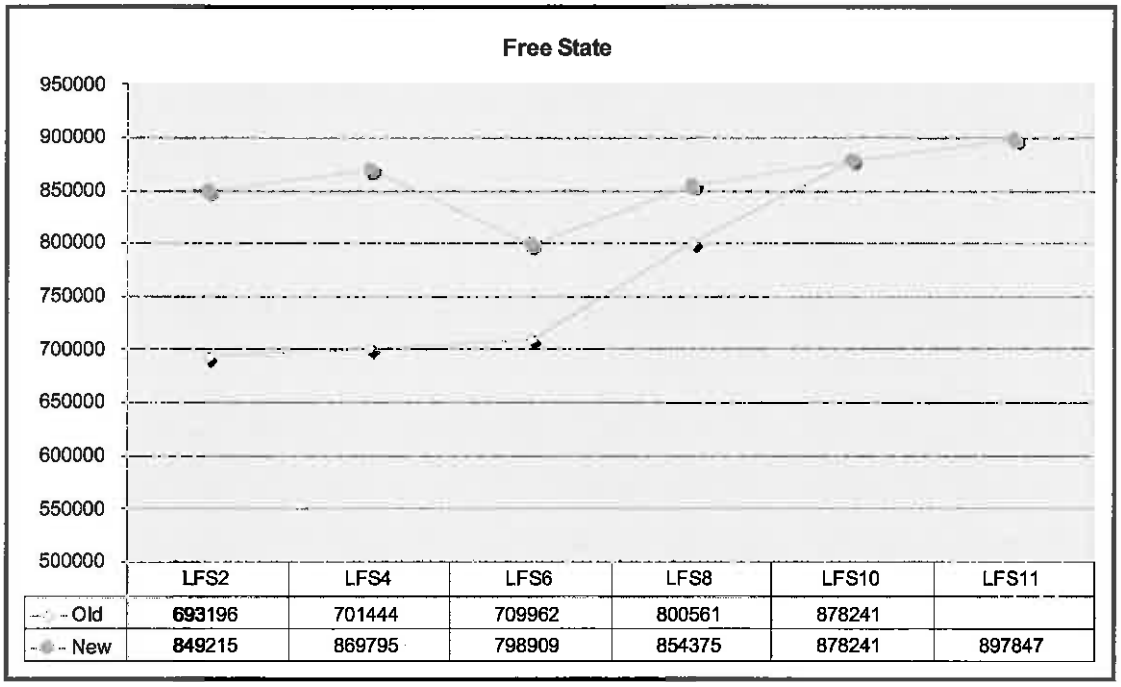
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1066652	1148598	1320218	1305189	1284974	
New	1147198	1149727	1180440	1236037	1284974	1317246



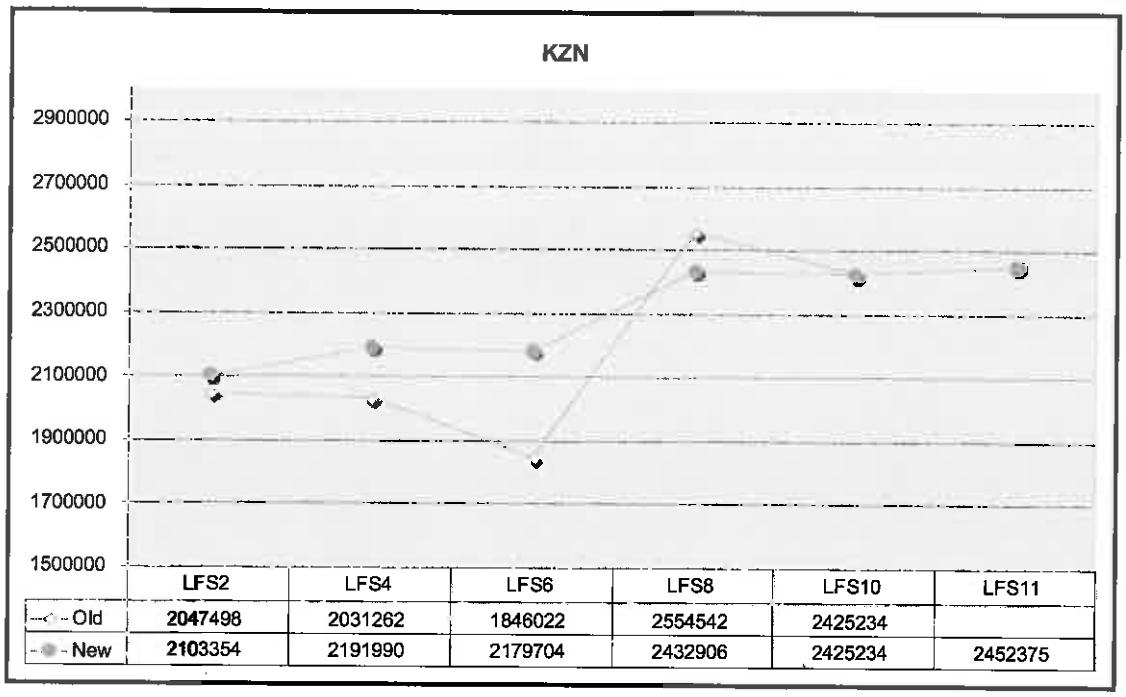
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1433908	1418056	1510264	1608190	1675283	
New	1581413	1558687	1636108	1688609	1675283	1659611



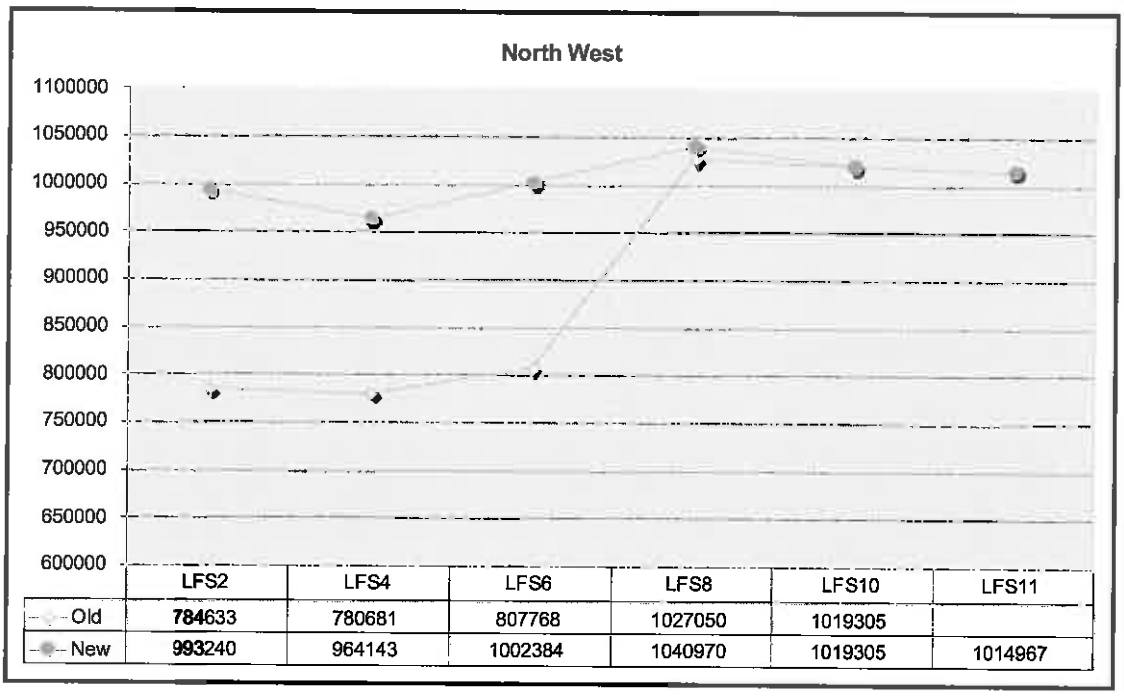
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	190136	184601	214718	219740	246377	
New	246922	238676	241638	241445	246377	235673



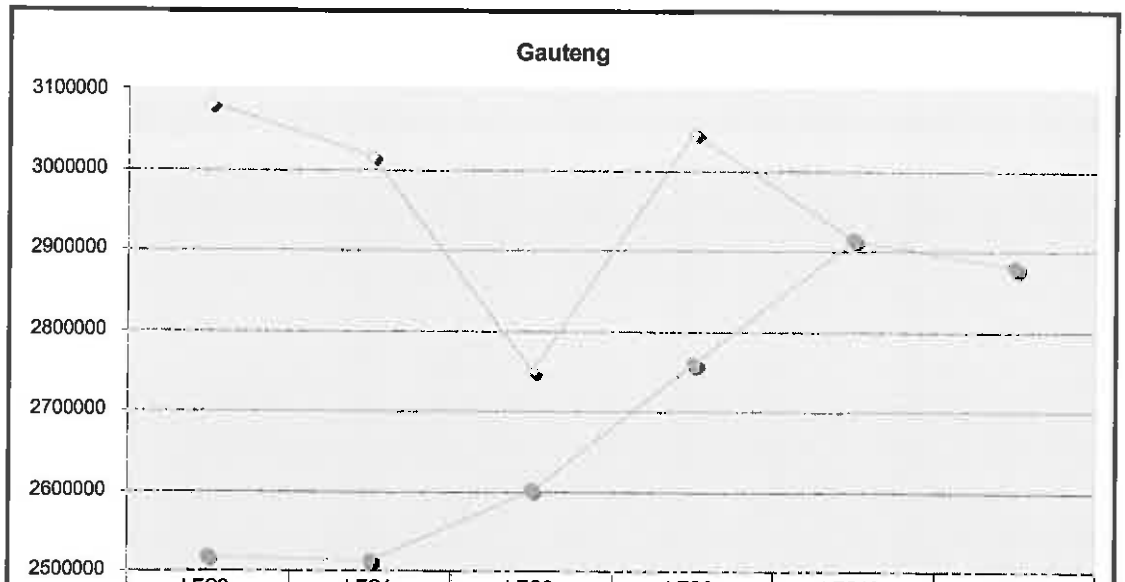
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	693196	701444	709962	800561	878241	
New	849215	869795	798909	854375	878241	897847



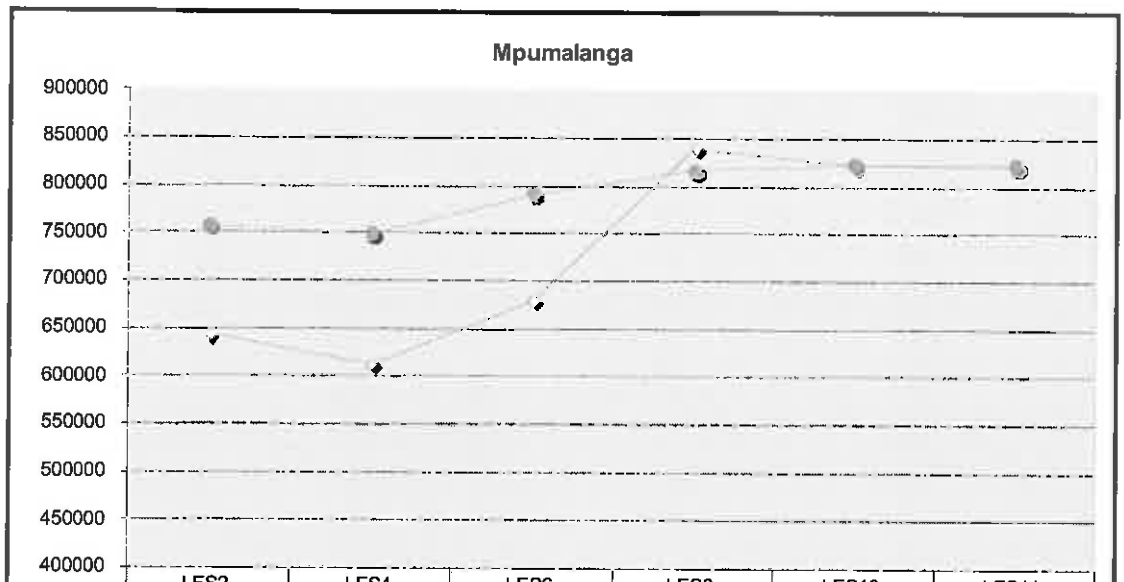
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	2047498	2031262	1846022	2554542	2425234	
New	2103354	2191990	2179704	2432906	2425234	2452375



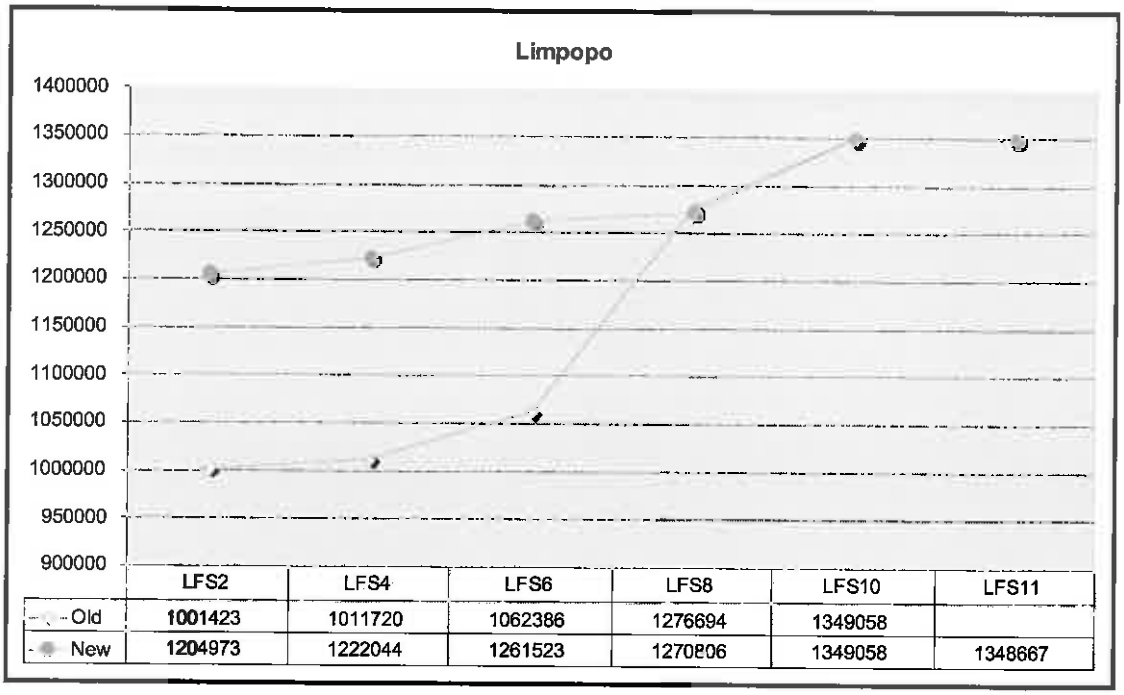
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	784633	780681	807768	1027050	1019305	
New	993240	964143	1002384	1040970	1019305	1014967



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3082113	3017006	2750596	3046939	2914926	
New	2518256	2512788	2600730	2759188	2914926	2880269

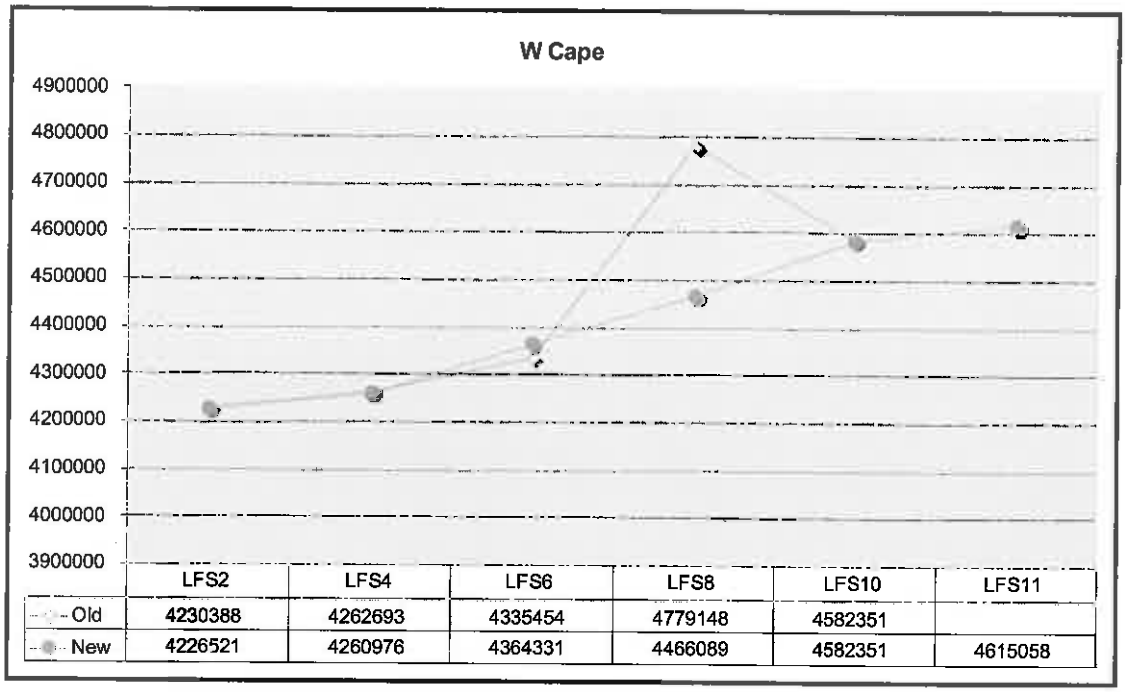


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	643221	612320	679722	838709	821470	
New	756266	748168	790672	815384	821470	821285

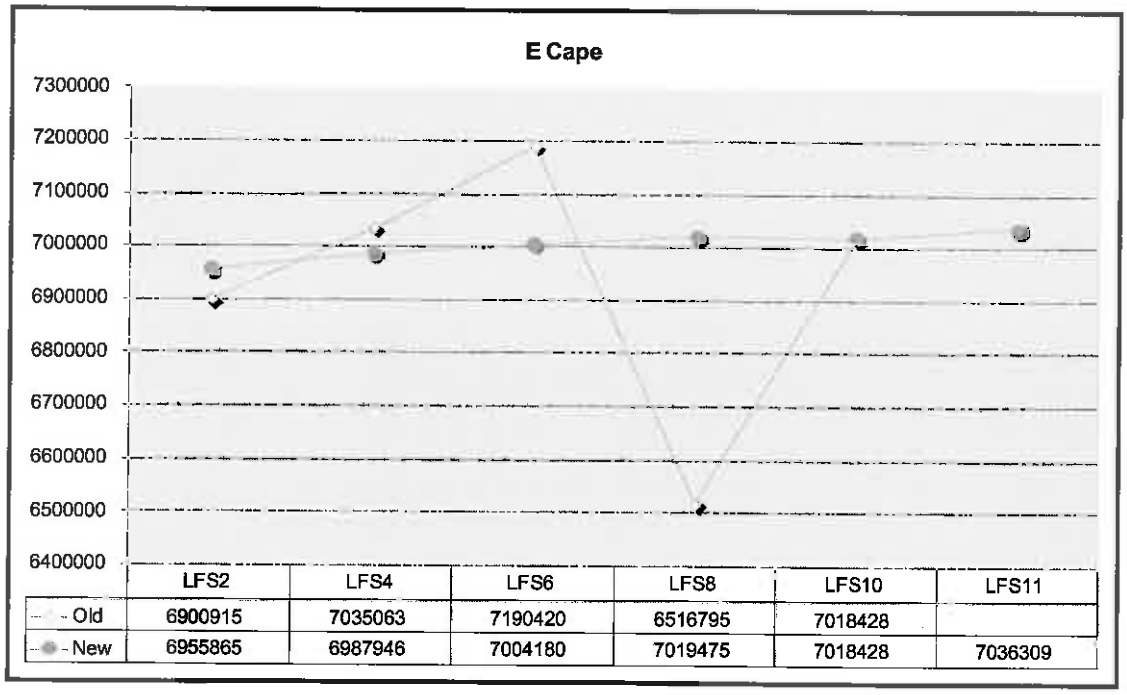


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1001423	1011720	1062386	1276694	1349058	
New	1204973	1222044	1261523	1270806	1349058	1348667

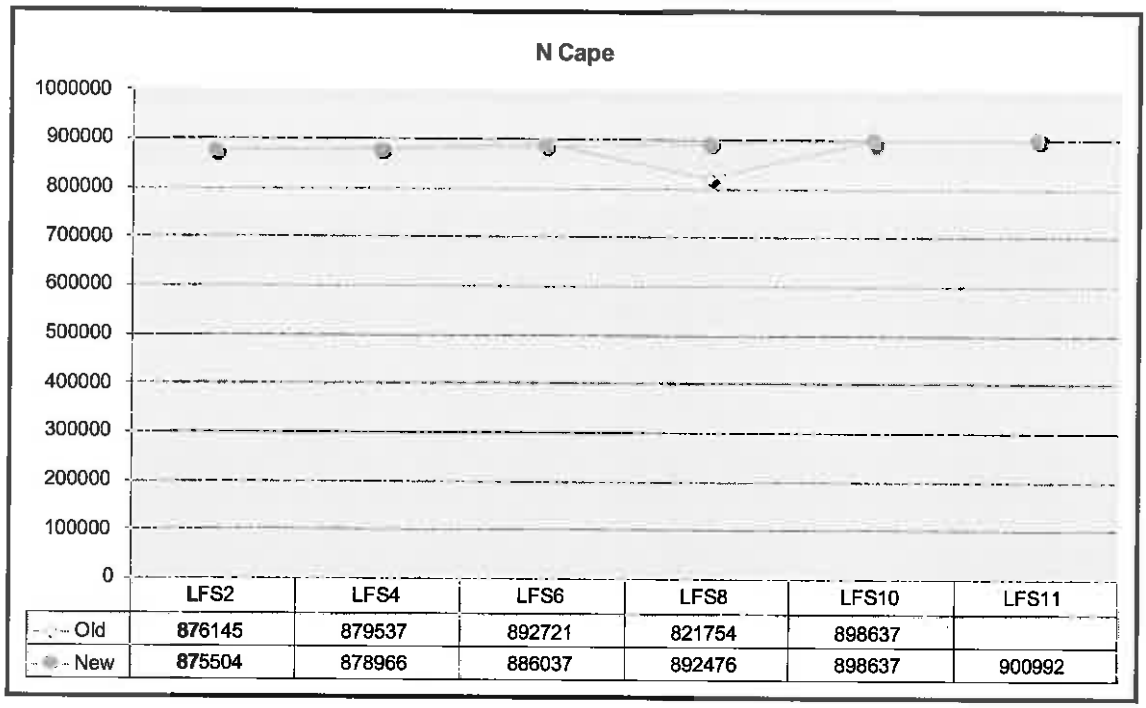
Appendix 2 – Variation in person numbers by province



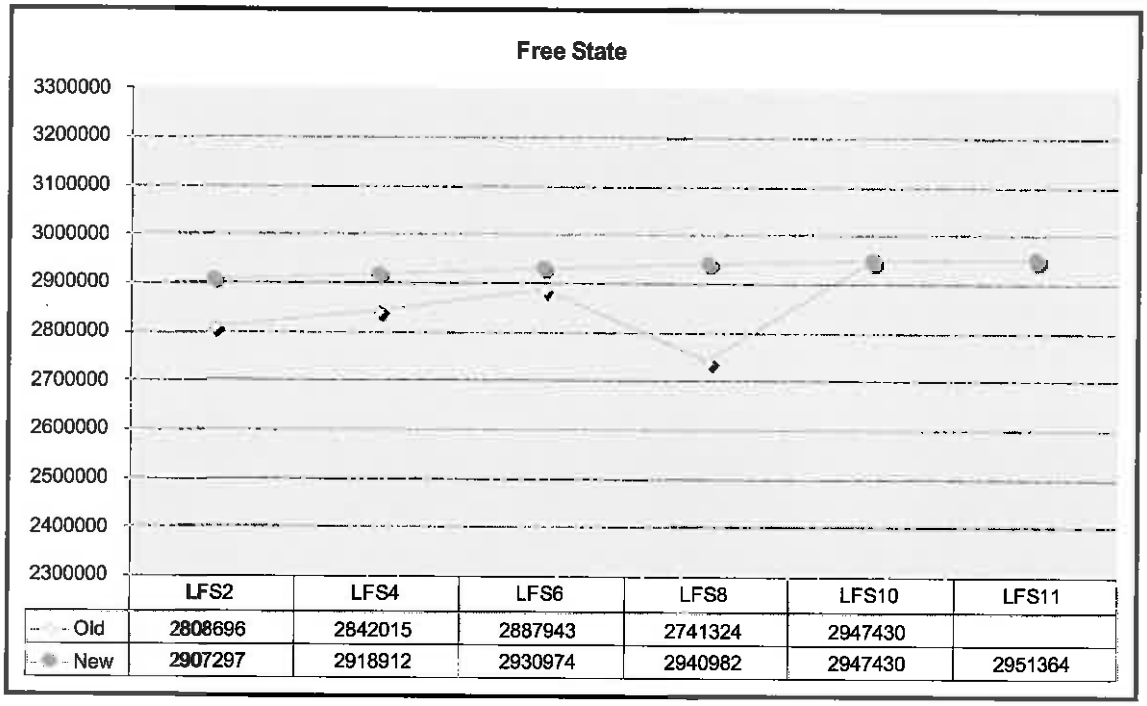
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4230388	4262693	4335454	4779148	4582351	
New	4226521	4260976	4364331	4466089	4582351	4615058



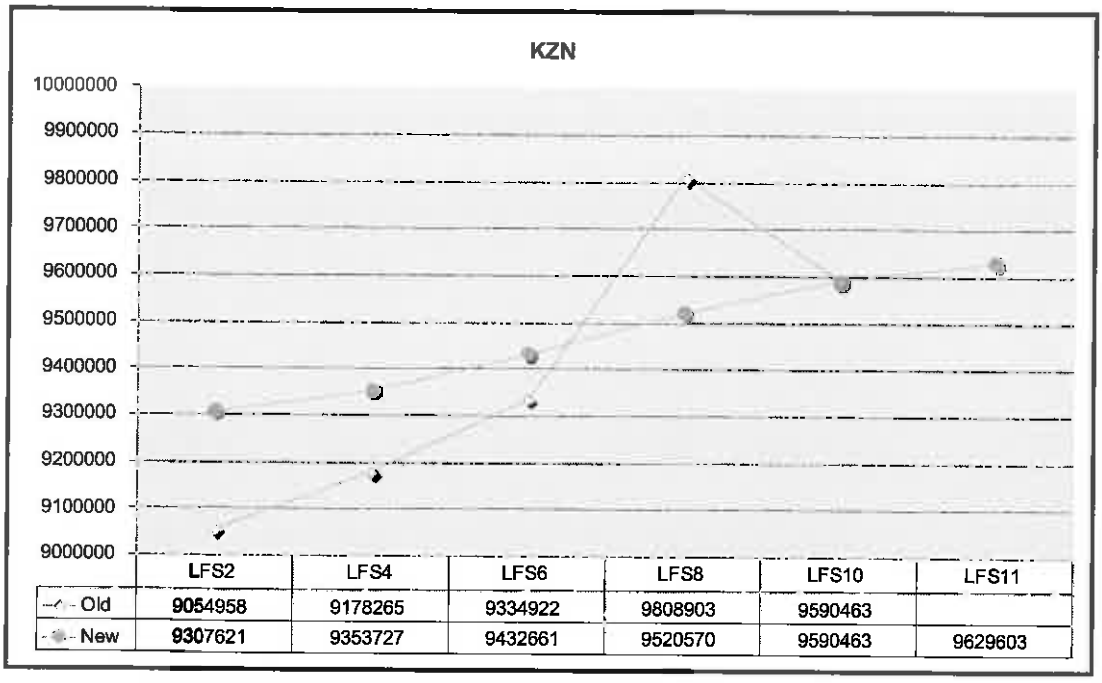
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	6900915	7035063	7190420	6516795	7018428	
New	6955865	6987946	7004180	7019475	7018428	7036309



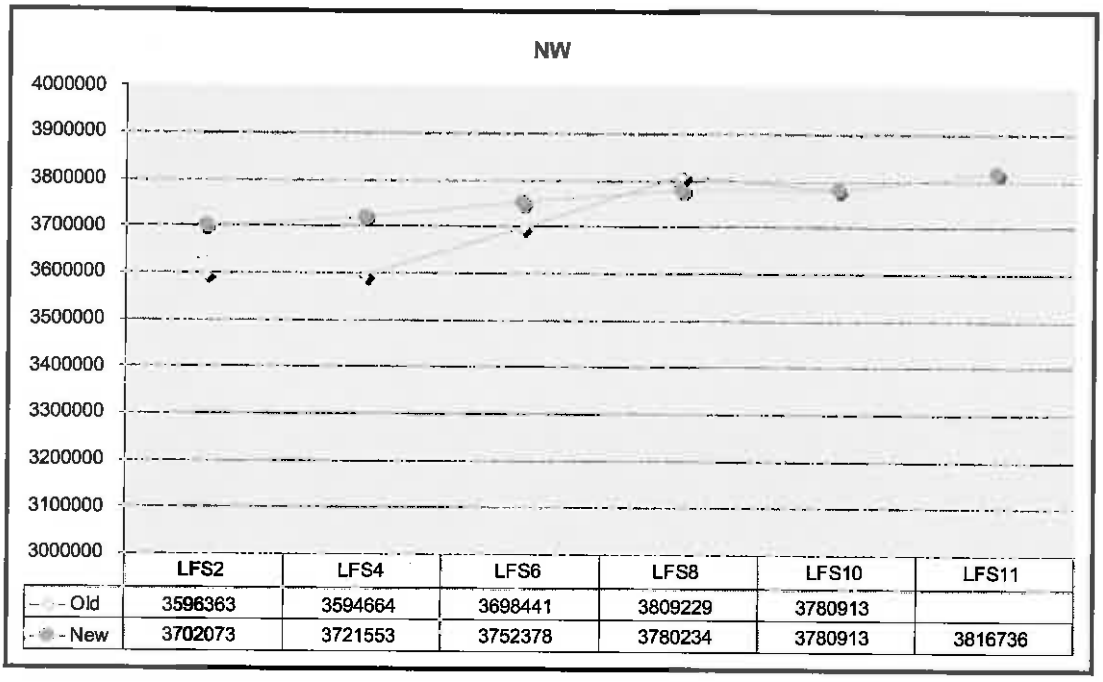
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	876145	879537	892721	821754	898637	
New	875504	878966	886037	892476	898637	900992



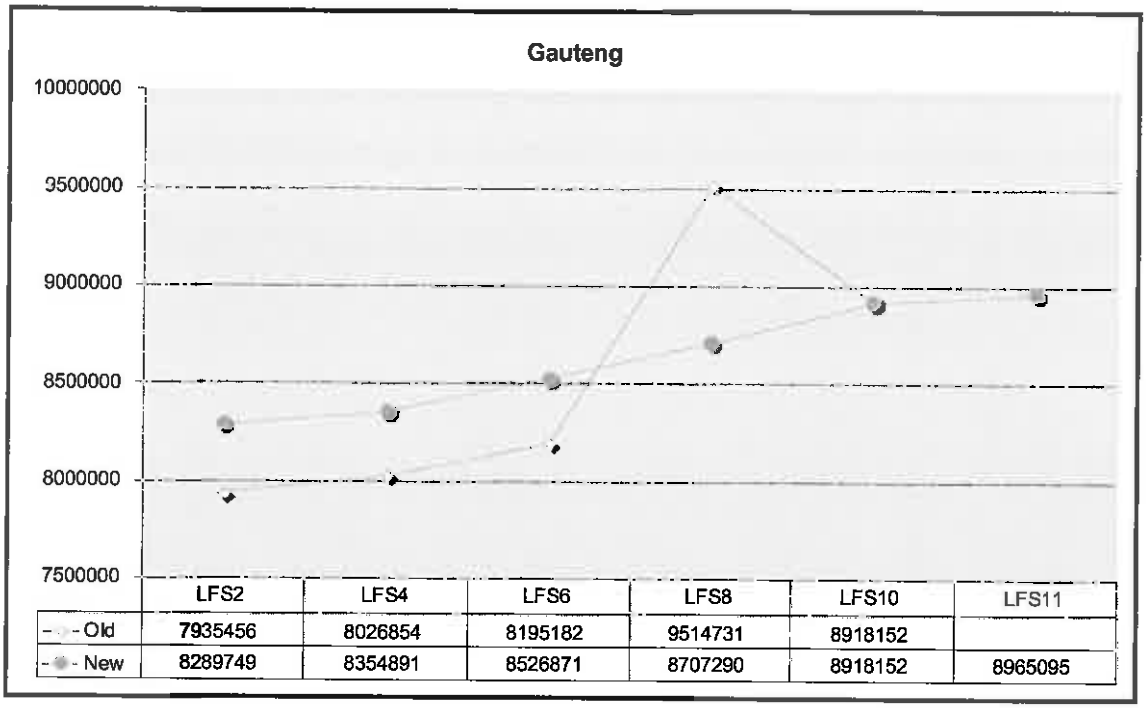
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	2808696	2842015	2887943	2741324	2947430	
New	2907297	2918912	2930974	2940982	2947430	2951364



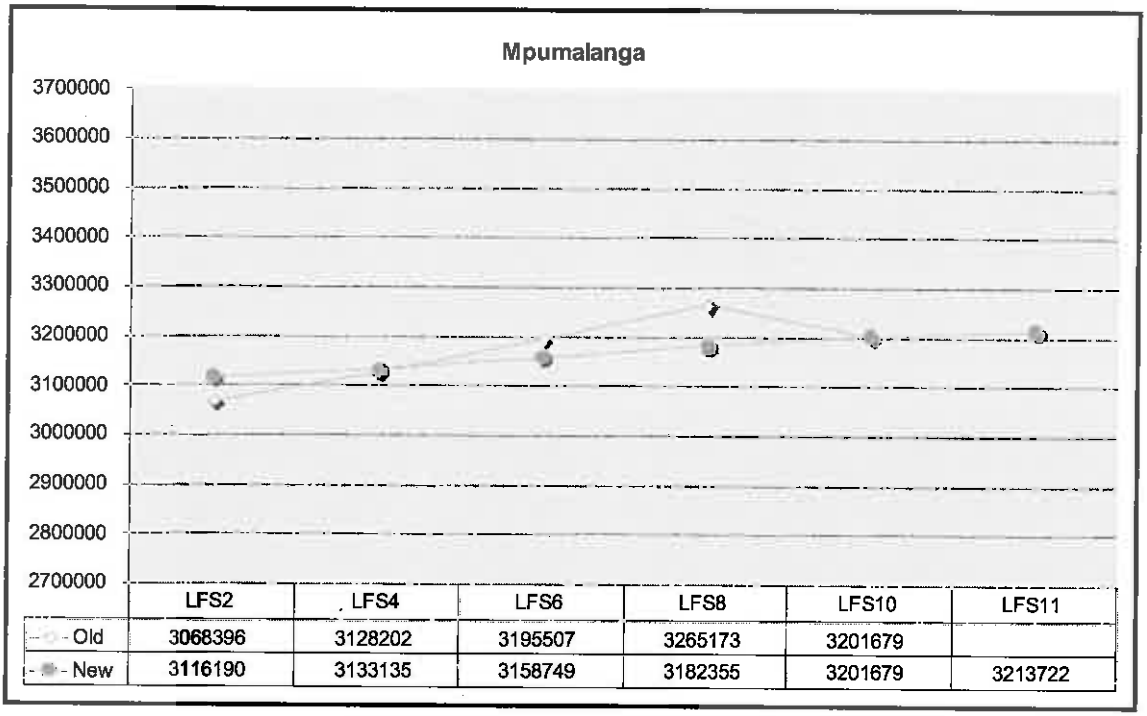
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	9054958	9178265	9334922	9808903	9590463	
New	9307621	9353727	9432661	9520570	9590463	9629603



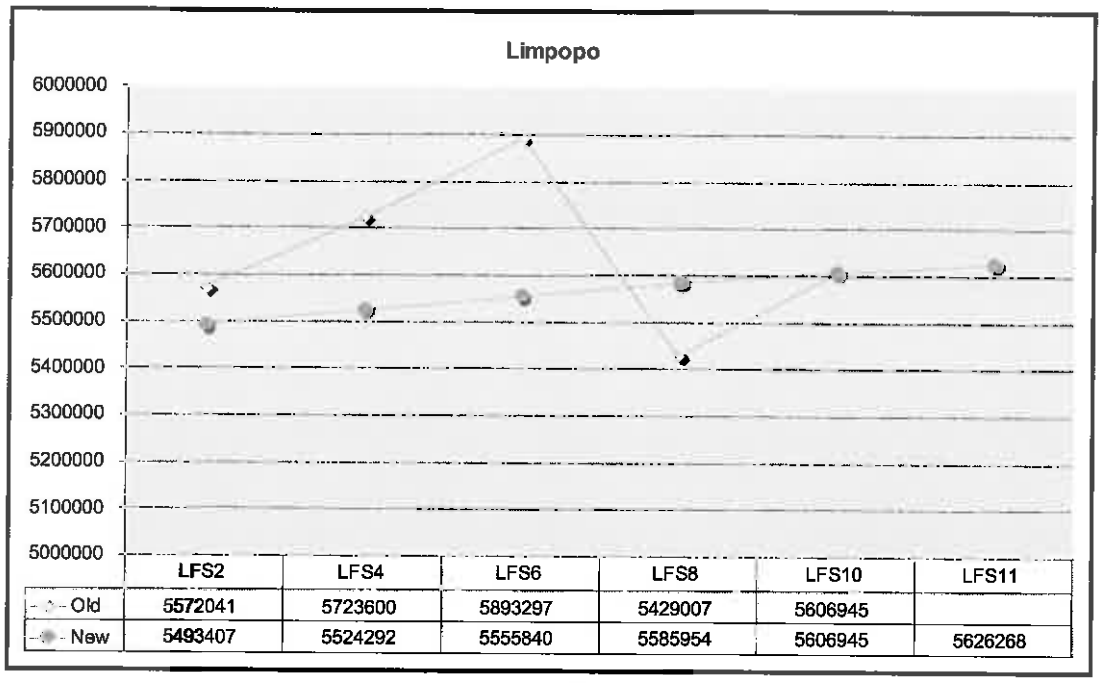
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3596363	3594664	3698441	3809229	3780913	
New	3702073	3721553	3752378	3780234	3780913	3816736



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	7935456	8026854	8195182	9514731	8918152	
New	8289749	8354891	8526871	8707290	8918152	8965095

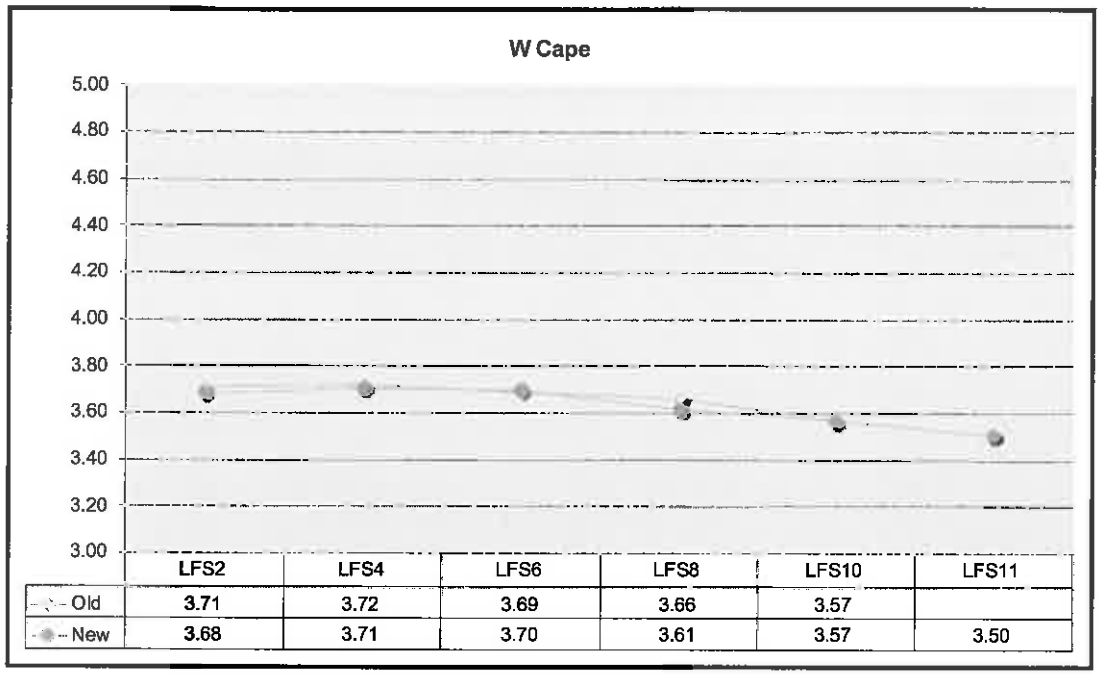


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3068396	3128202	3195507	3265173	3201679	
New	3116190	3133135	3158749	3182355	3201679	3213722

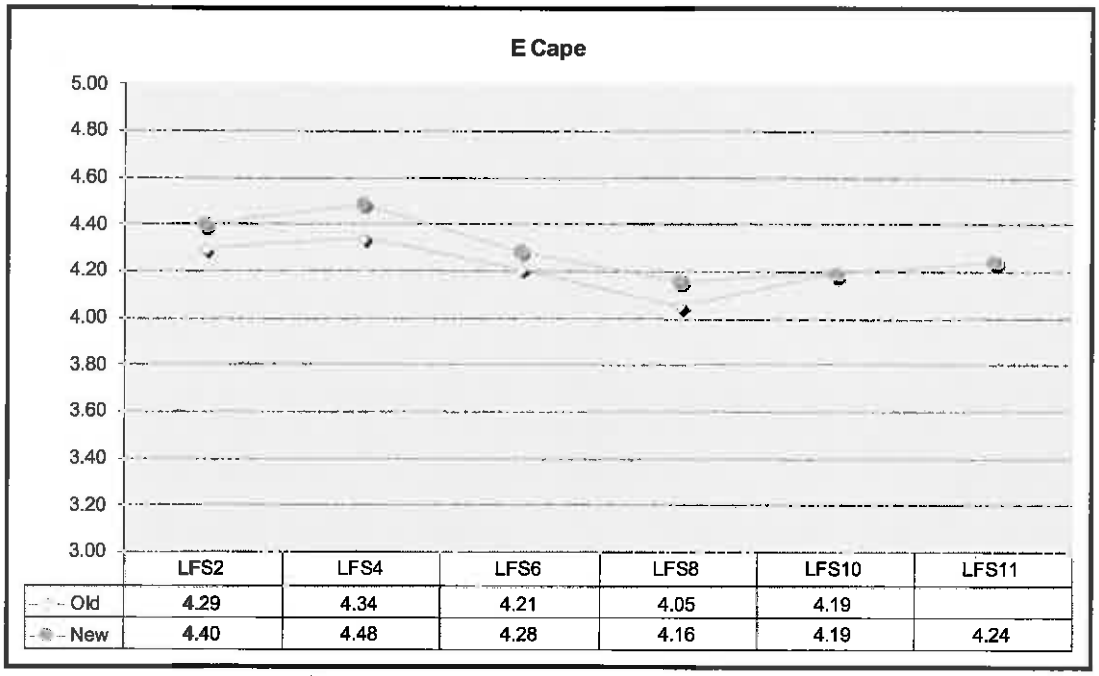


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	5572041	5723600	5893297	5429007	5606945	
New	5493407	5524292	5555840	5585954	5606945	5626268

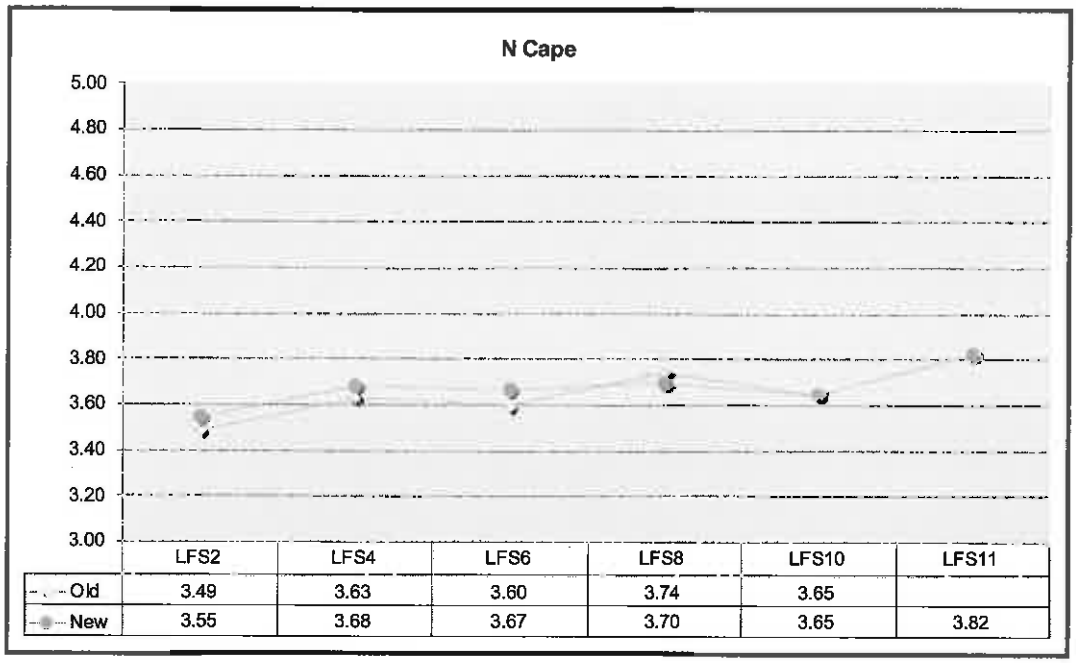
Appendix 3 – Variation in household size by province



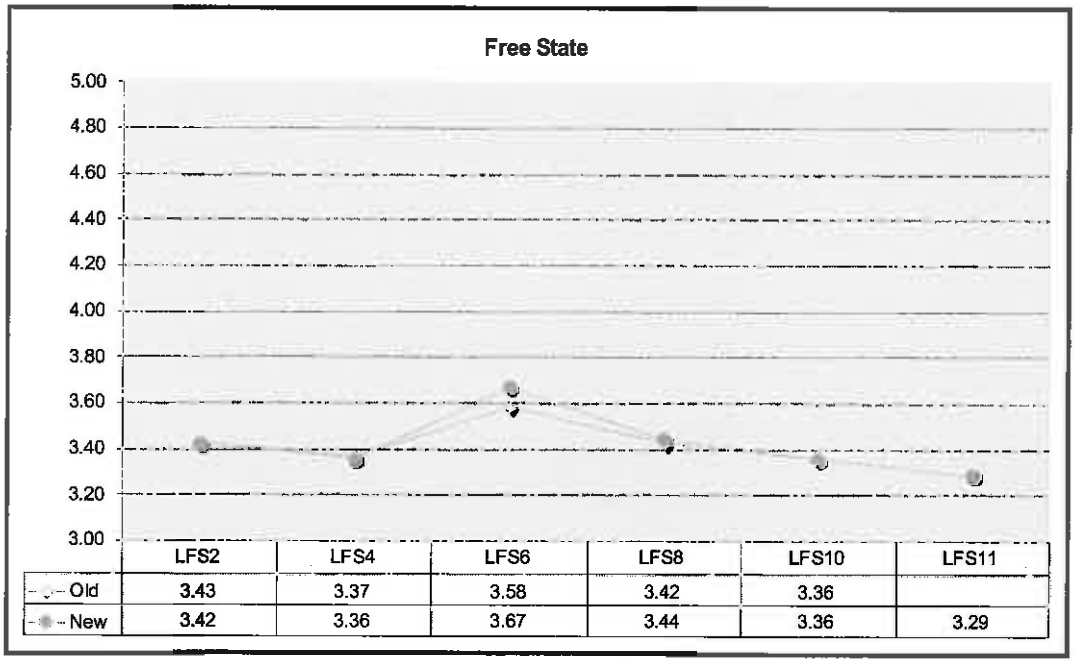
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3.71	3.72	3.69	3.66	3.57	
New	3.68	3.71	3.70	3.61	3.57	3.50



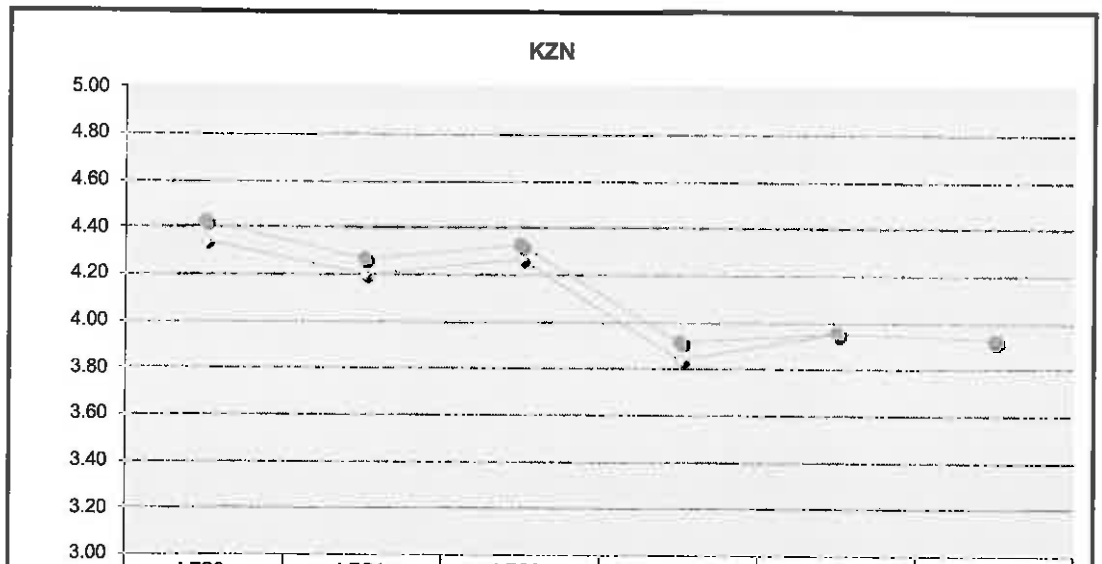
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4.29	4.34	4.21	4.05	4.19	
New	4.40	4.48	4.28	4.16	4.19	4.24



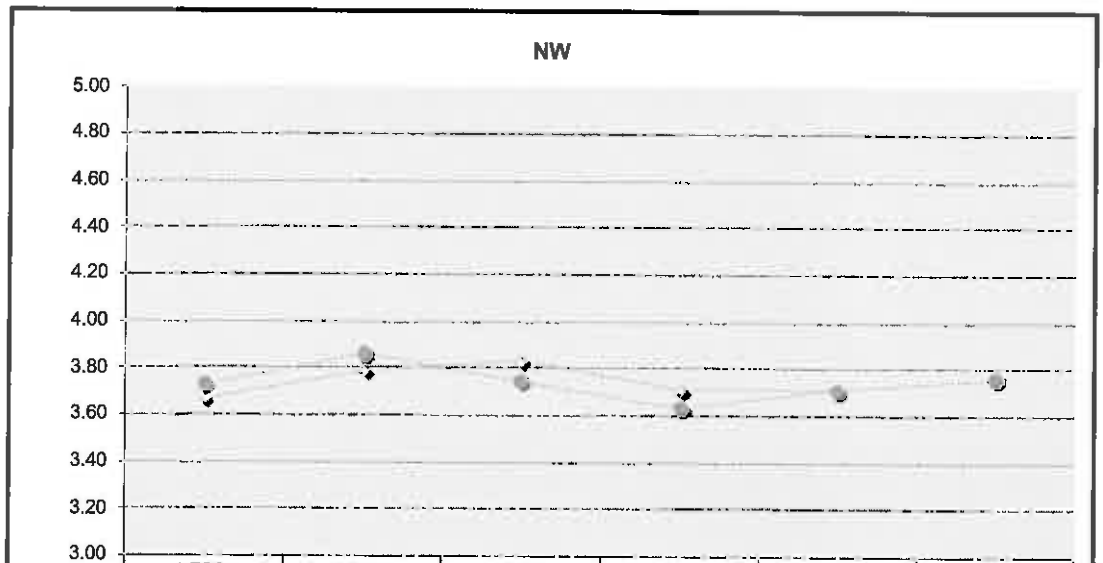
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3.49	3.63	3.60	3.74	3.65	
New	3.55	3.68	3.67	3.70	3.65	3.82



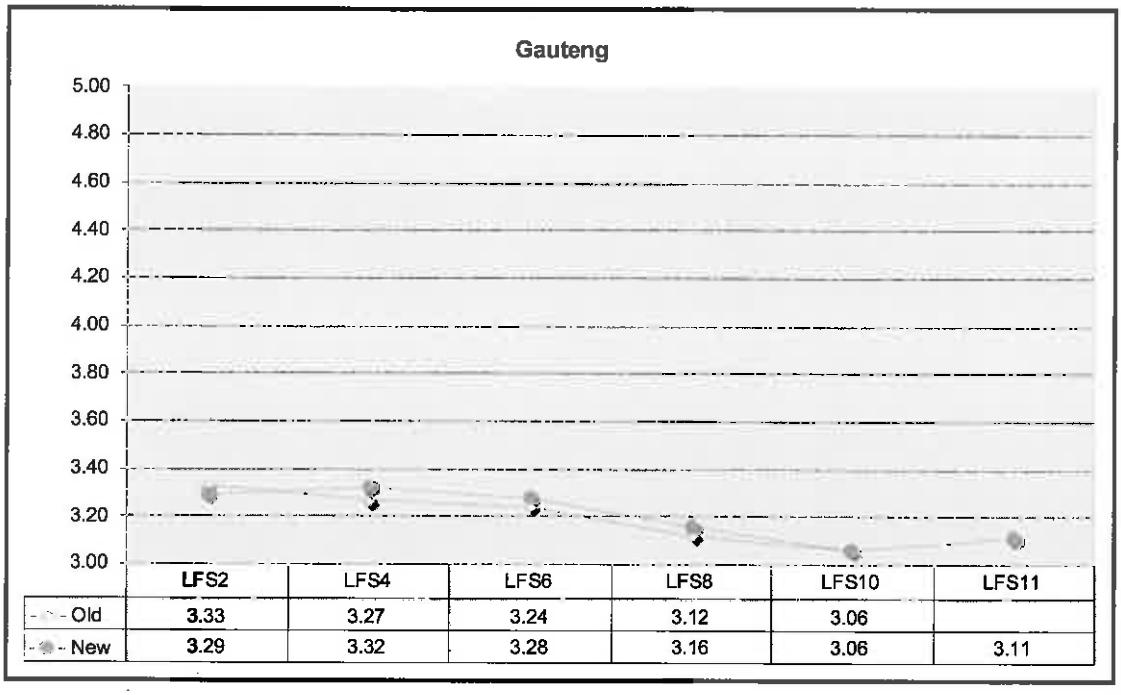
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3.43	3.37	3.58	3.42	3.36	
New	3.42	3.36	3.67	3.44	3.36	3.29



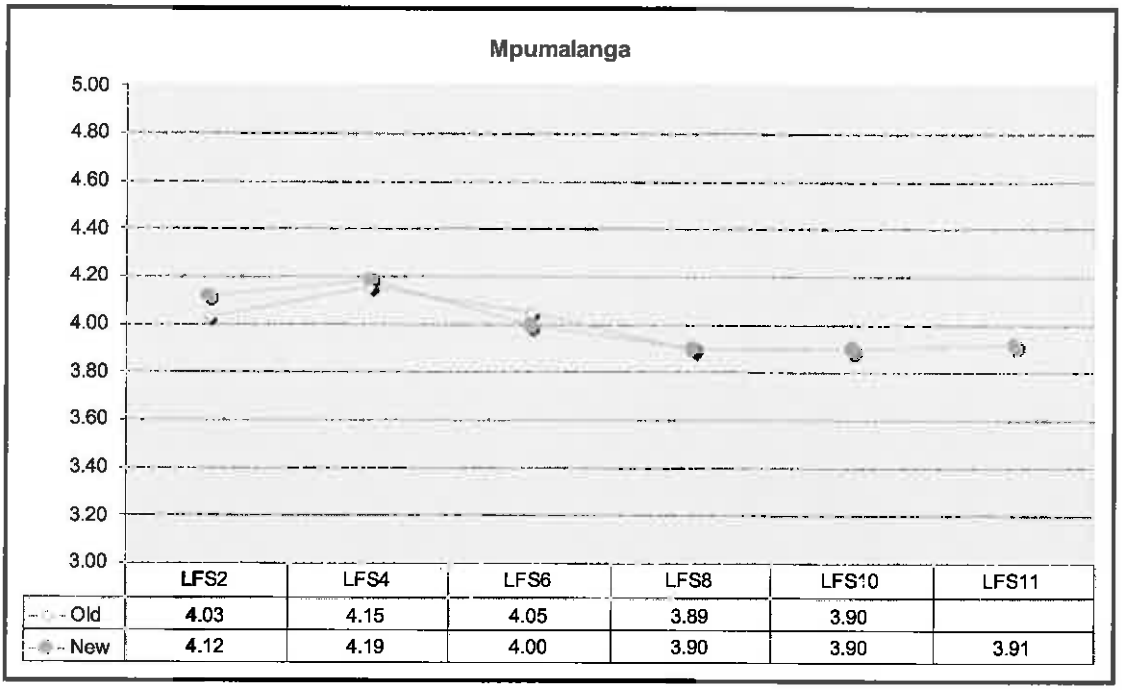
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4.34	4.20	4.27	3.84	3.95	
New	4.43	4.27	4.33	3.91	3.95	3.93



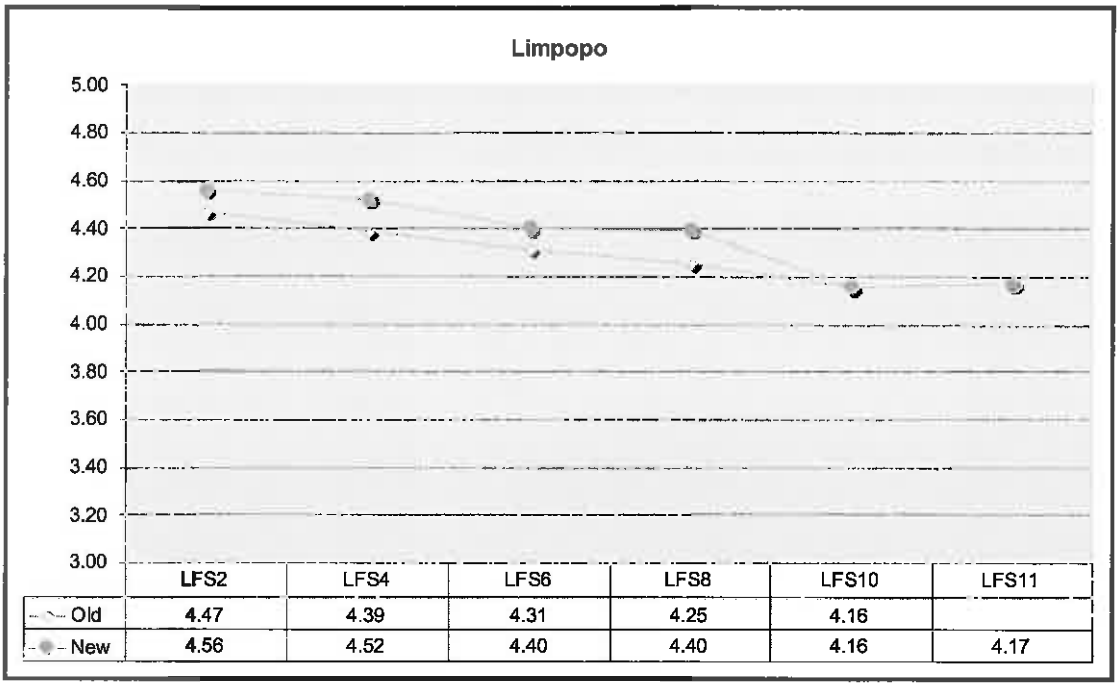
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3.67	3.79	3.83	3.71	3.71	
New	3.73	3.86	3.74	3.63	3.71	3.76



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3.33	3.27	3.24	3.12	3.06	
New	3.29	3.32	3.28	3.16	3.06	3.11

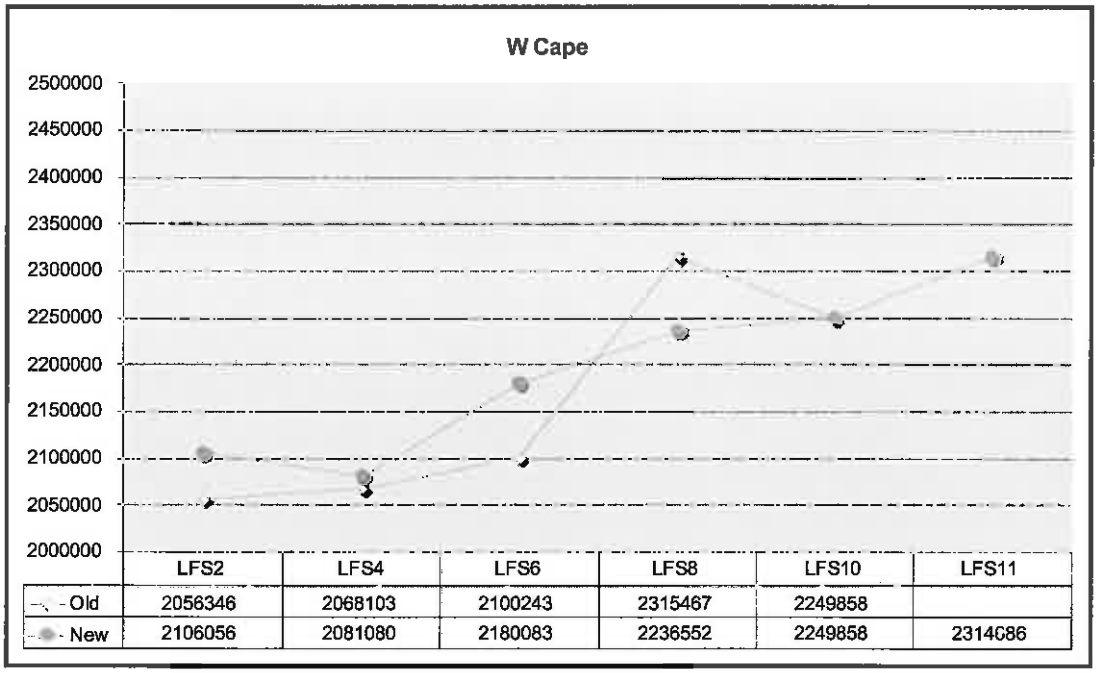


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4.03	4.15	4.05	3.89	3.90	
New	4.12	4.19	4.00	3.90	3.90	3.91

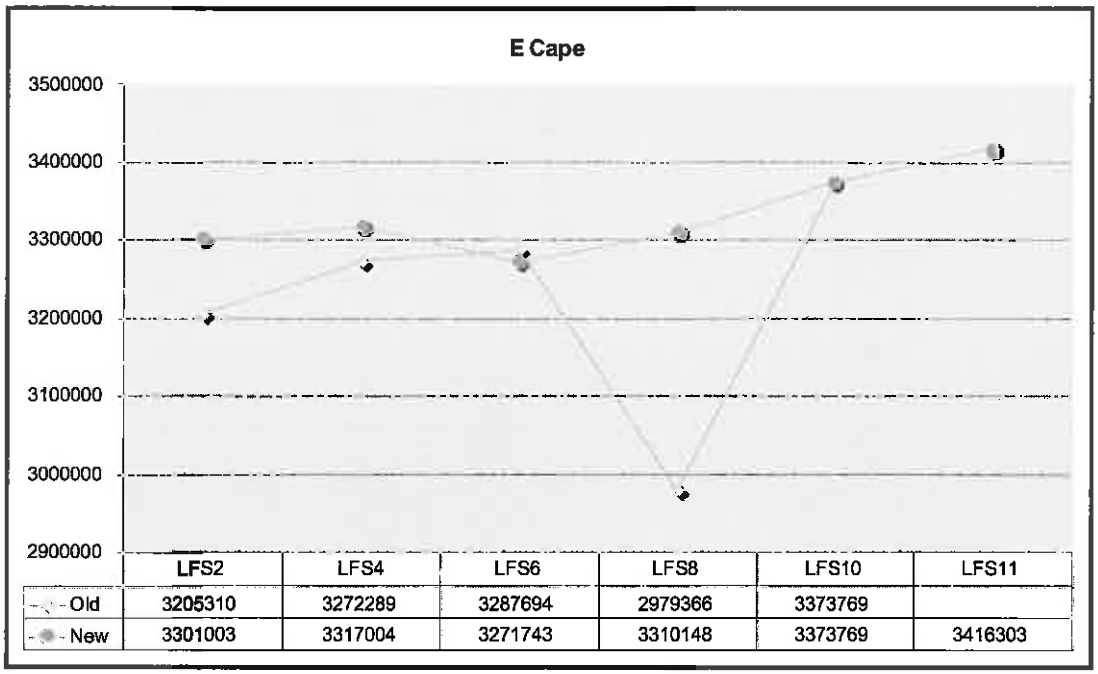


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4.47	4.39	4.31	4.25	4.16	
New	4.56	4.52	4.40	4.40	4.16	4.17

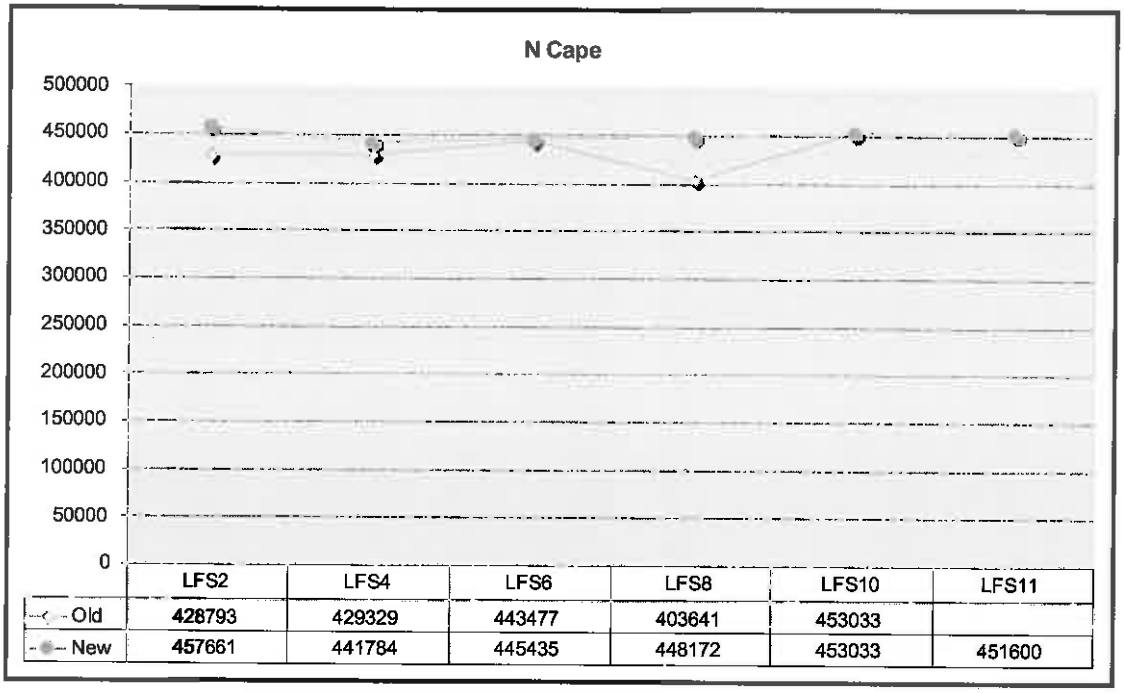
Appendix 4 – Variation in the male population by province



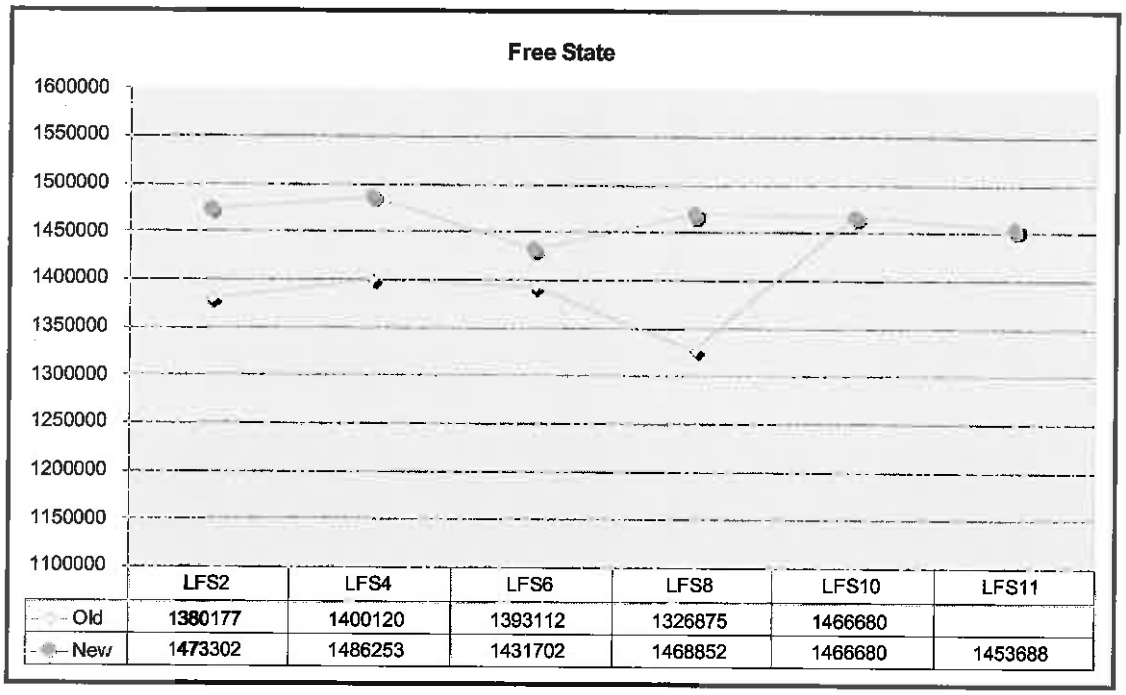
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	2056346	2068103	2100243	2315467	2249858	
New	2106056	2081080	2180083	2236552	2249858	2314086



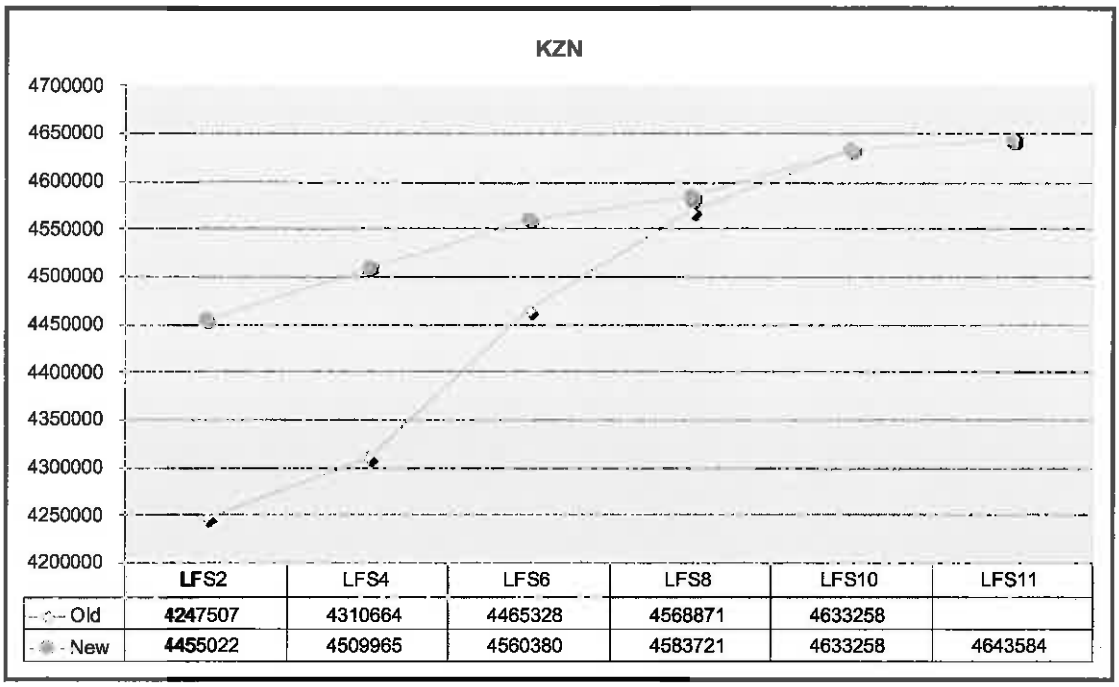
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	3205310	3272289	3287694	2979366	3373769	
New	3301003	3317004	3271743	3310148	3373769	3416303



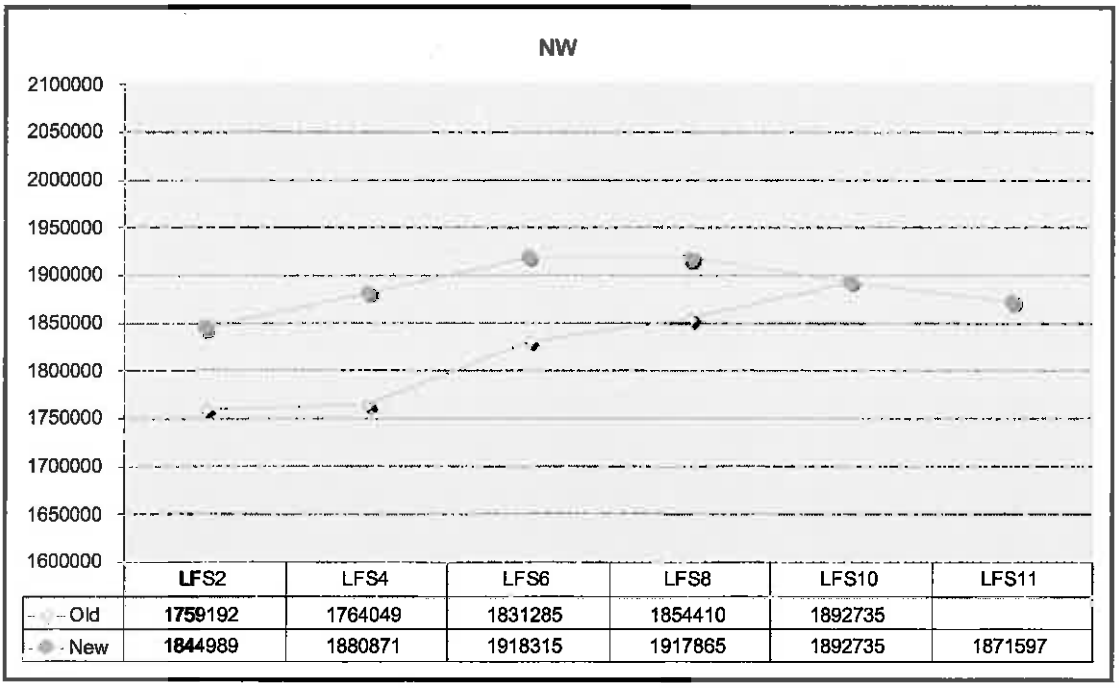
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	428793	429329	443477	403641	453033	
New	457661	441784	445435	448172	453033	451600



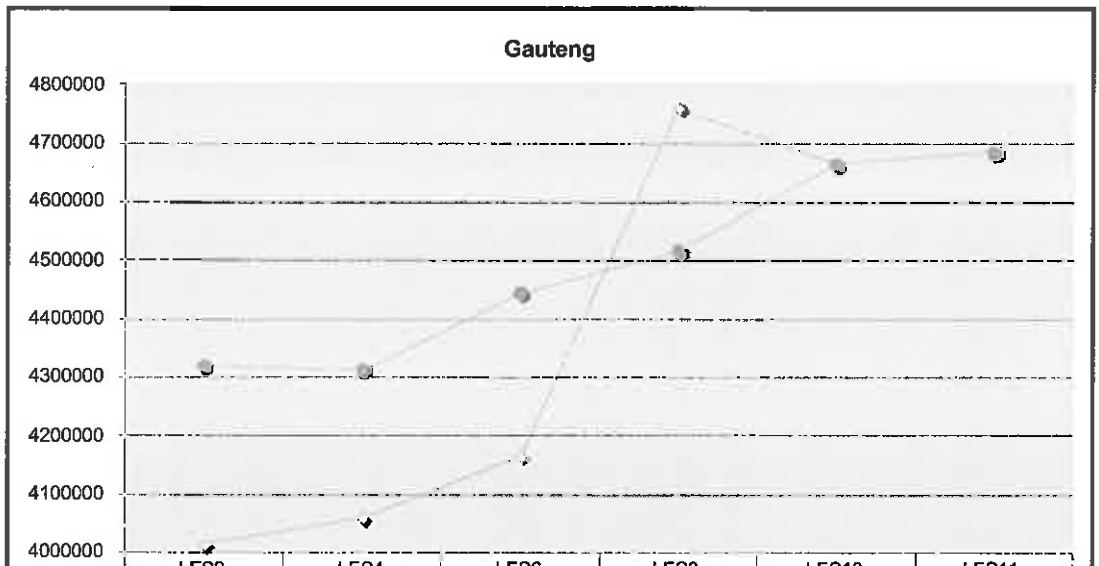
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1380177	1400120	1393112	1326875	1466680	
New	1473302	1486253	1431702	1468852	1466680	1453688



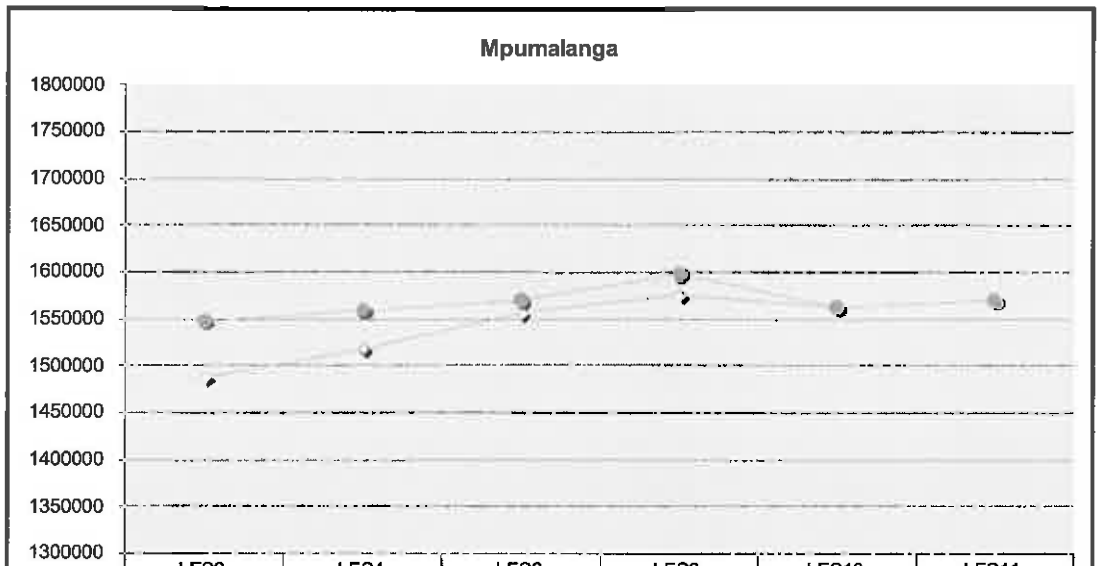
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4247507	4310664	4465328	4568871	4633258	
New	4455022	4509965	4560380	4583721	4633258	4643584



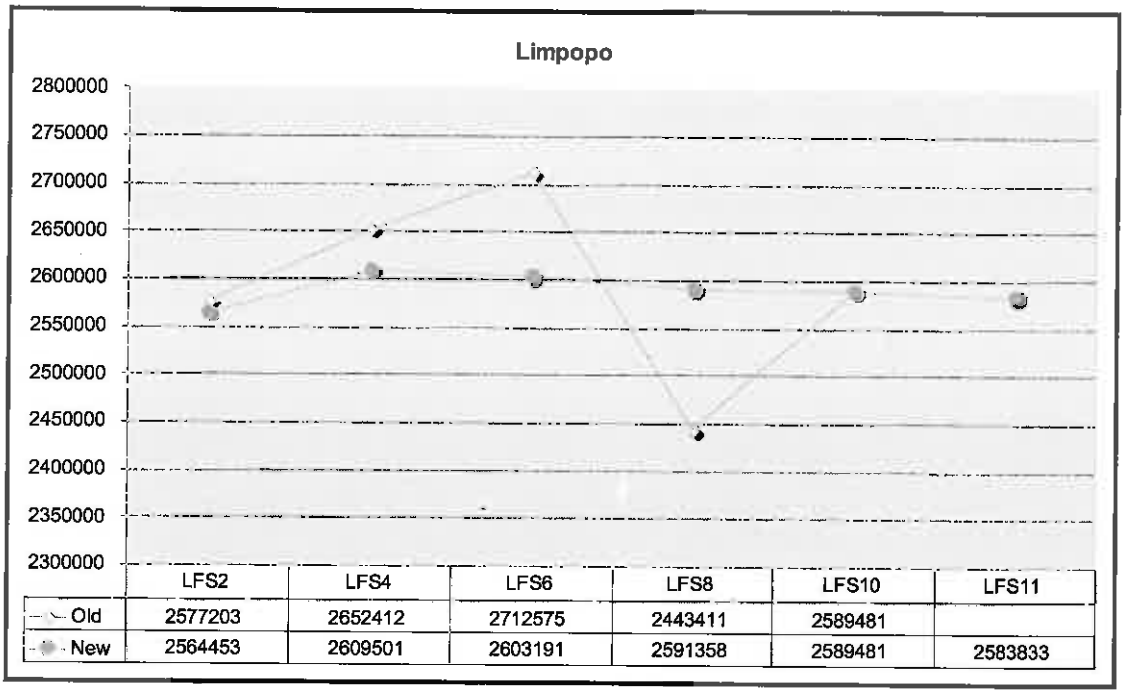
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1759192	1764049	1831285	1854410	1892735	
New	1844989	1880871	1918315	1917865	1892735	1871597



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	4013727	4059168	4164649	4760090	4664723	
New	4319578	4312878	4443643	4516225	4664723	4685481

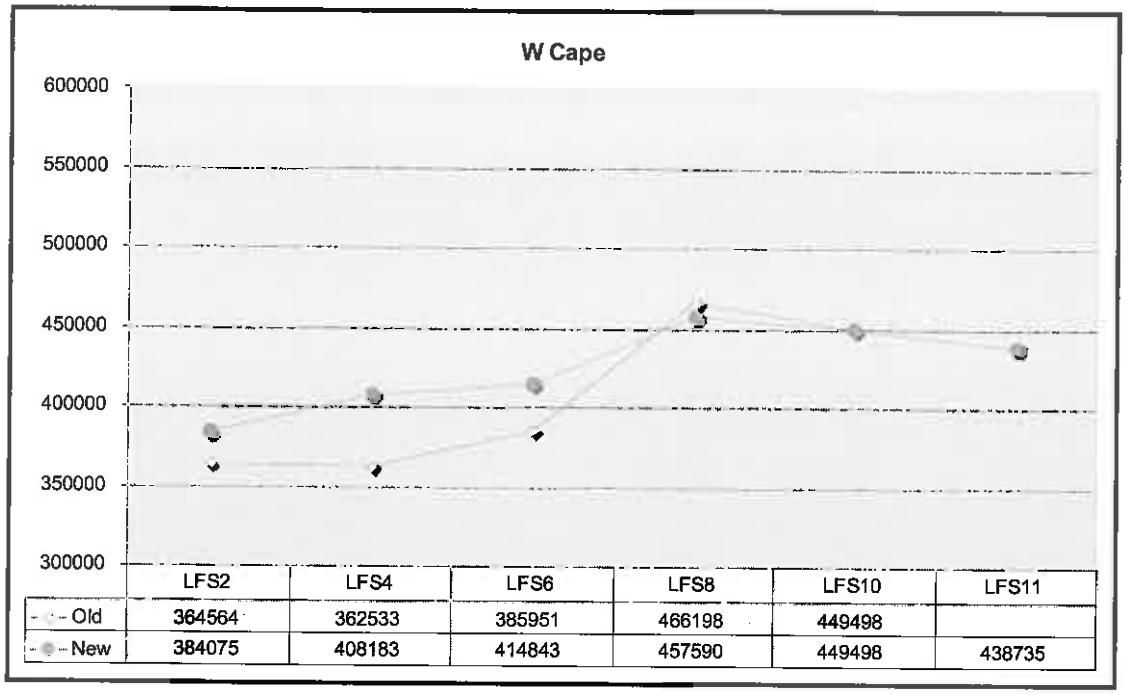


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	1487711	1518289	1556159	1575317	1562217	
New	1547433	1559119	1569442	1598702	1562217	1570083

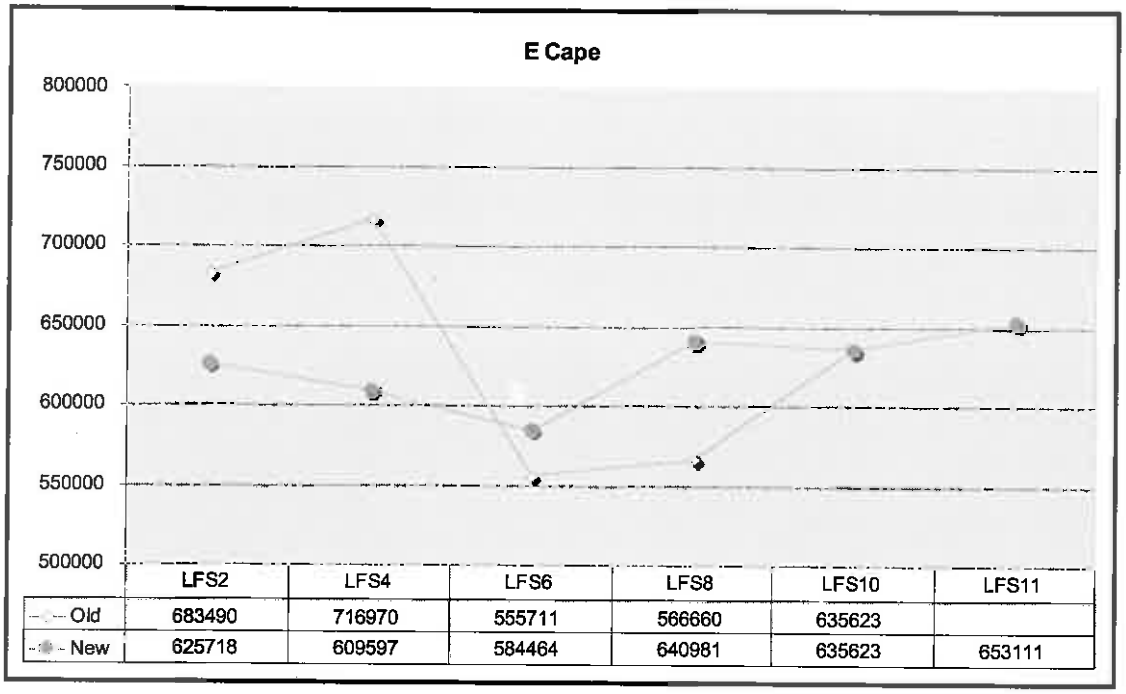


	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	2577203	2652412	2712575	2443411	2589481	
New	2564453	2609501	2603191	2591358	2589481	2583833

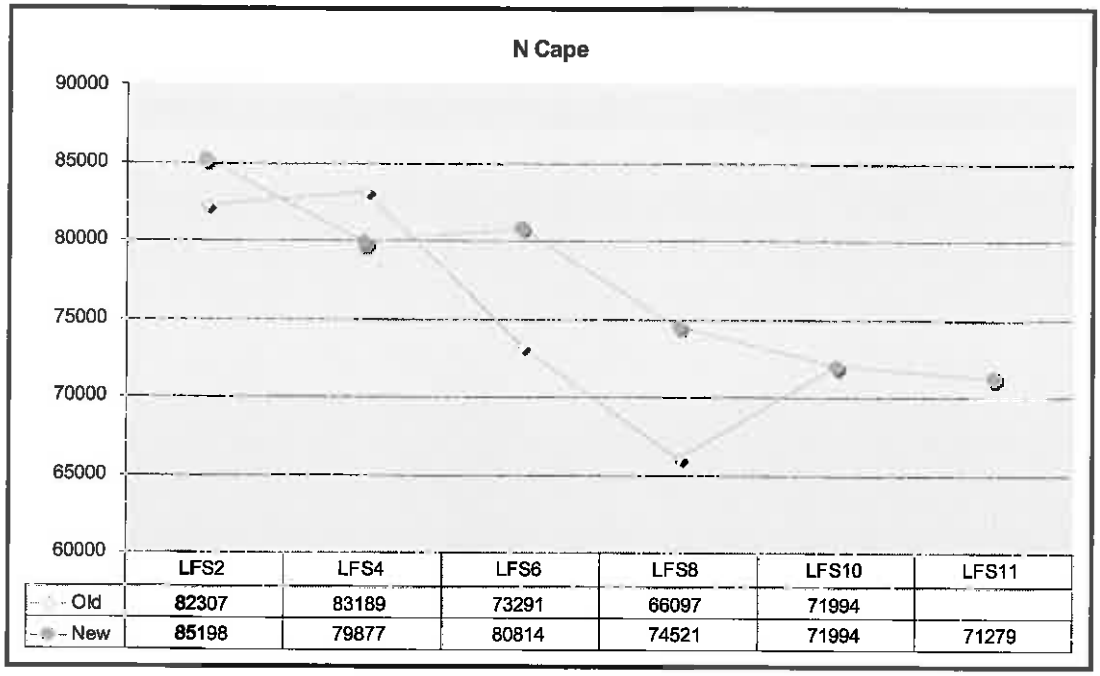
Appendix 5 – Variation in the population aged 20-24 by province



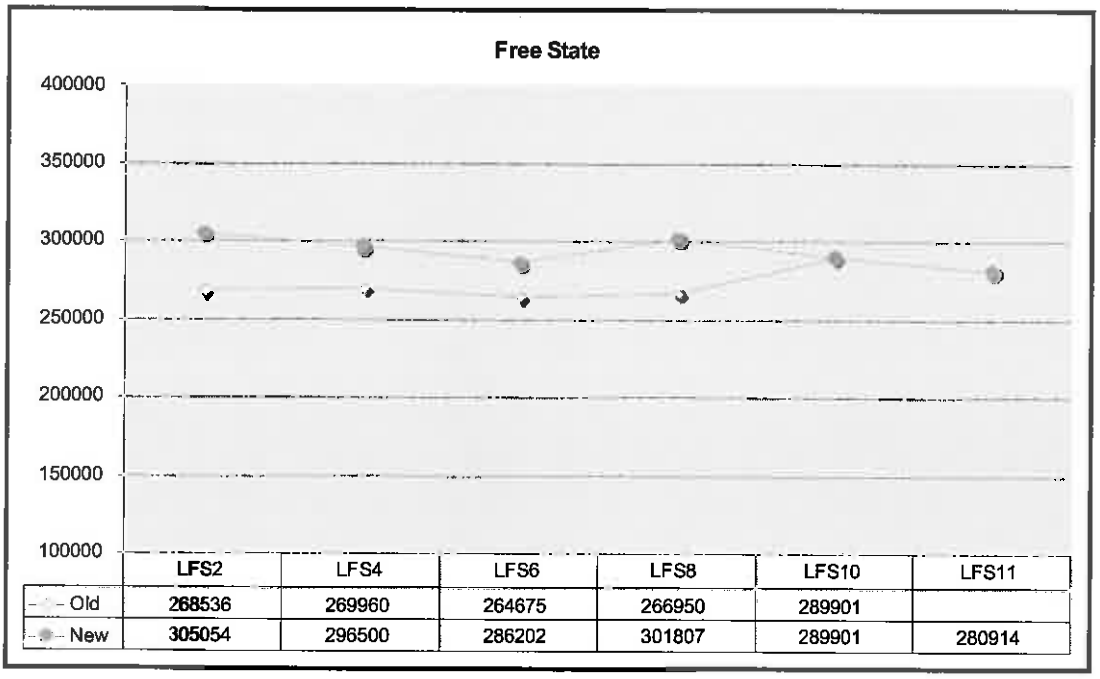
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	364564	362533	385951	466198	449498	
New	384075	408183	414843	457590	449498	438735



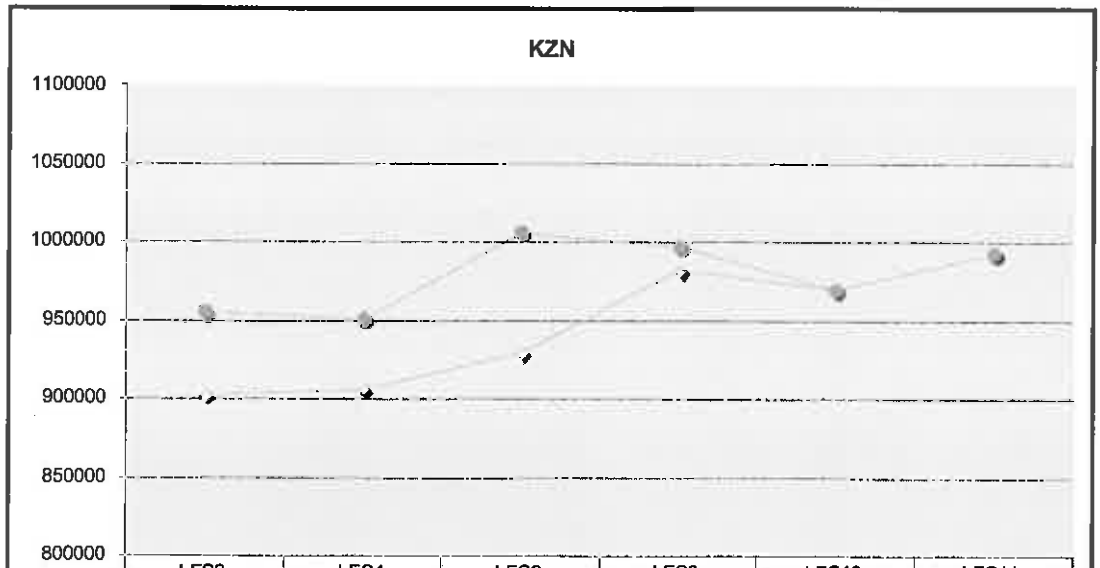
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	683490	716970	555711	566660	635623	
New	625718	609597	584464	640981	635623	653111



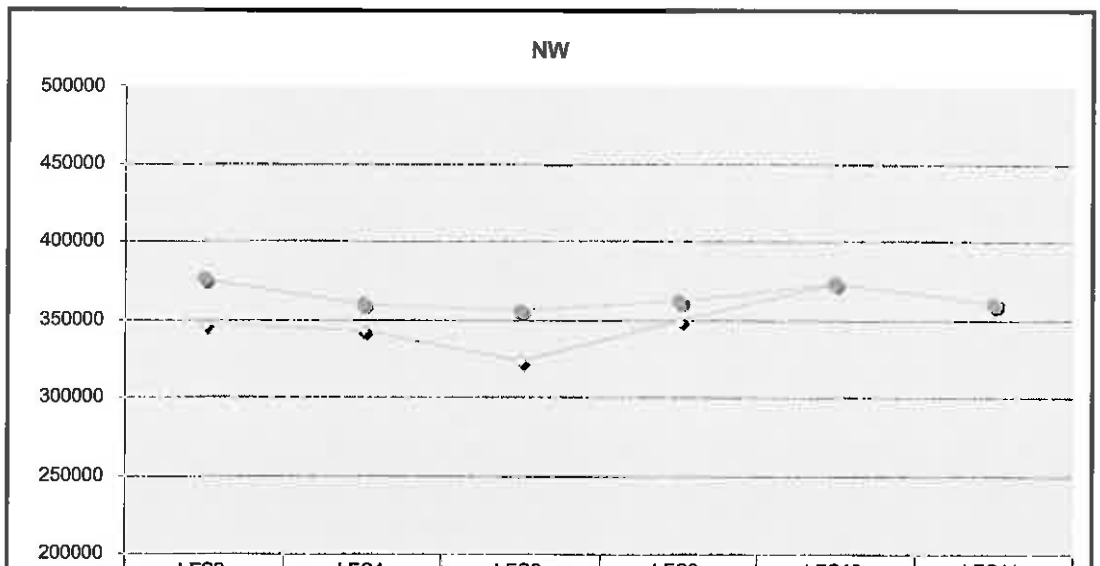
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	82307	83189	73291	66097	71994	
New	85198	79877	80814	74521	71994	71279



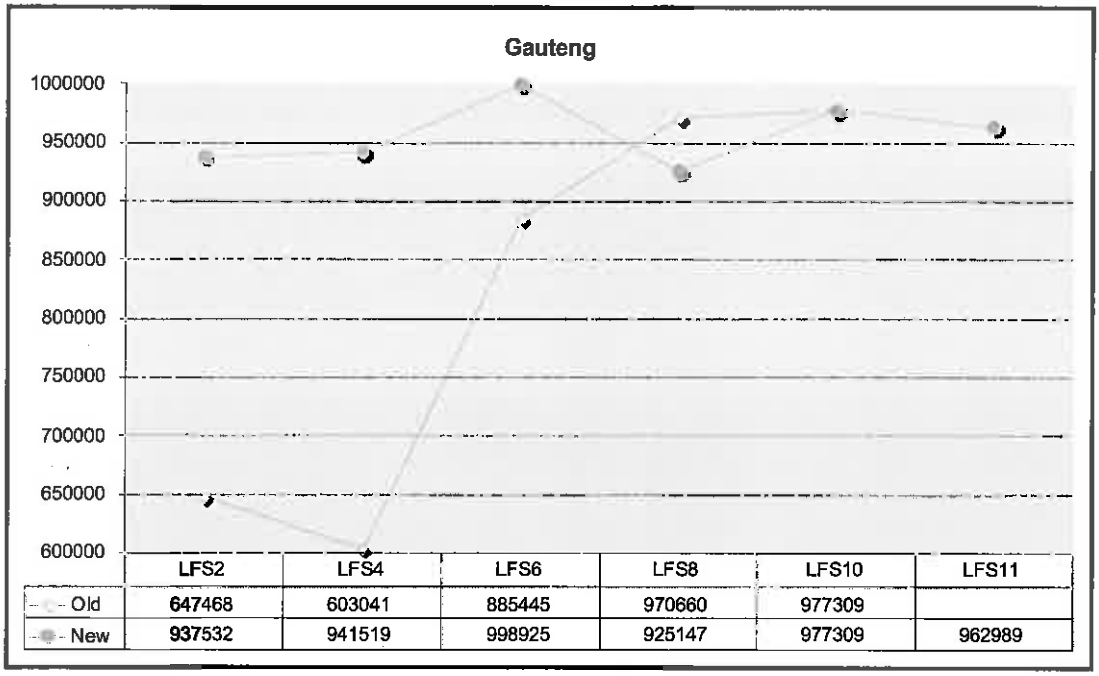
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	268536	269960	264675	266950	289901	
New	305054	296500	286202	301807	289901	280914



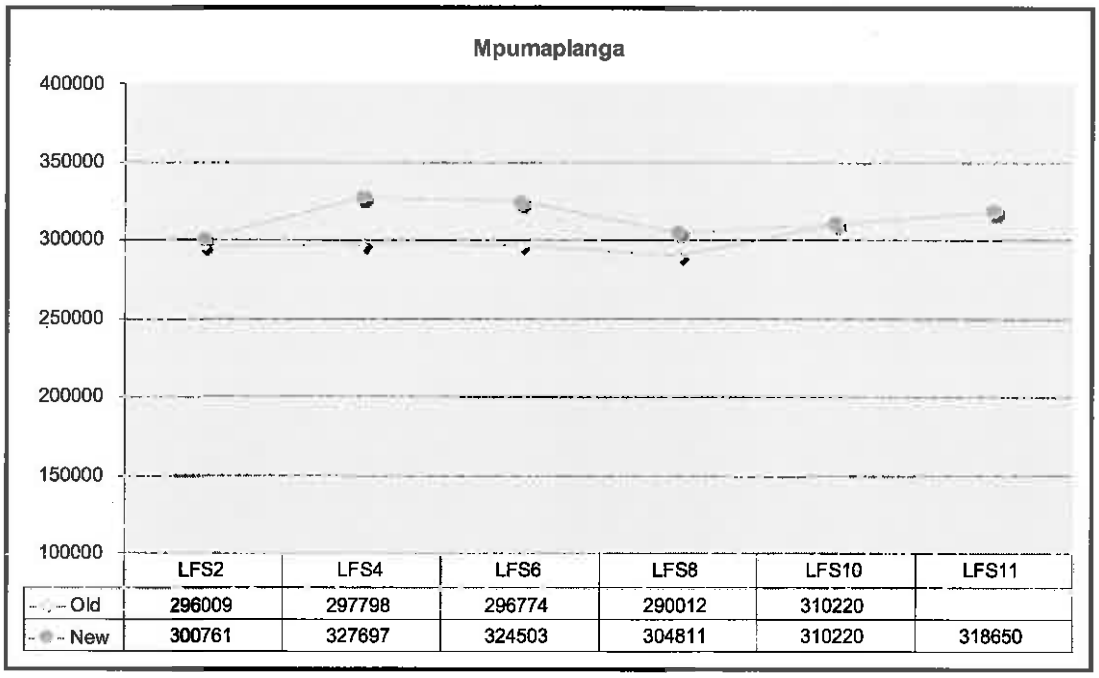
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	902211	906206	929131	981770	969310	
New	954923	950948	1006111	996898	969310	992556



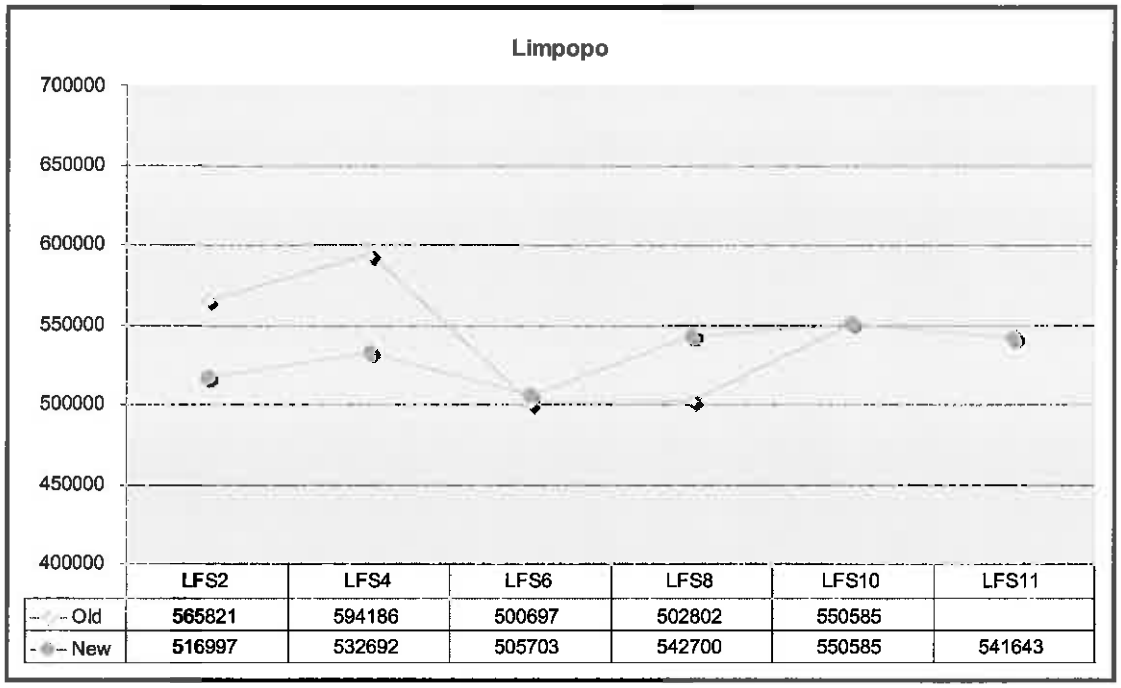
	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	346942	343191	324052	349491	372977	
New	375651	359529	355781	362094	372977	360331



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	647468	603041	885445	970660	977309	
New	937532	941519	998925	925147	977309	962989



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	296009	297798	296774	290012	310220	
New	300761	327697	324503	304811	310220	318650



	LFS2	LFS4	LFS6	LFS8	LFS10	LFS11
Old	565821	594186	500697	502802	550585	
New	516997	532692	505703	542700	550585	541643