Comparing estimates derived from Imputation and Sampling/Design weight adjustment: Agricultural Survey, 2011 gross farming income

Karabo Stephen Sebolai

Data Quality workshop

Introduction

- Agricultural survey - annual survey
- The survey covers – Farms engaged in:

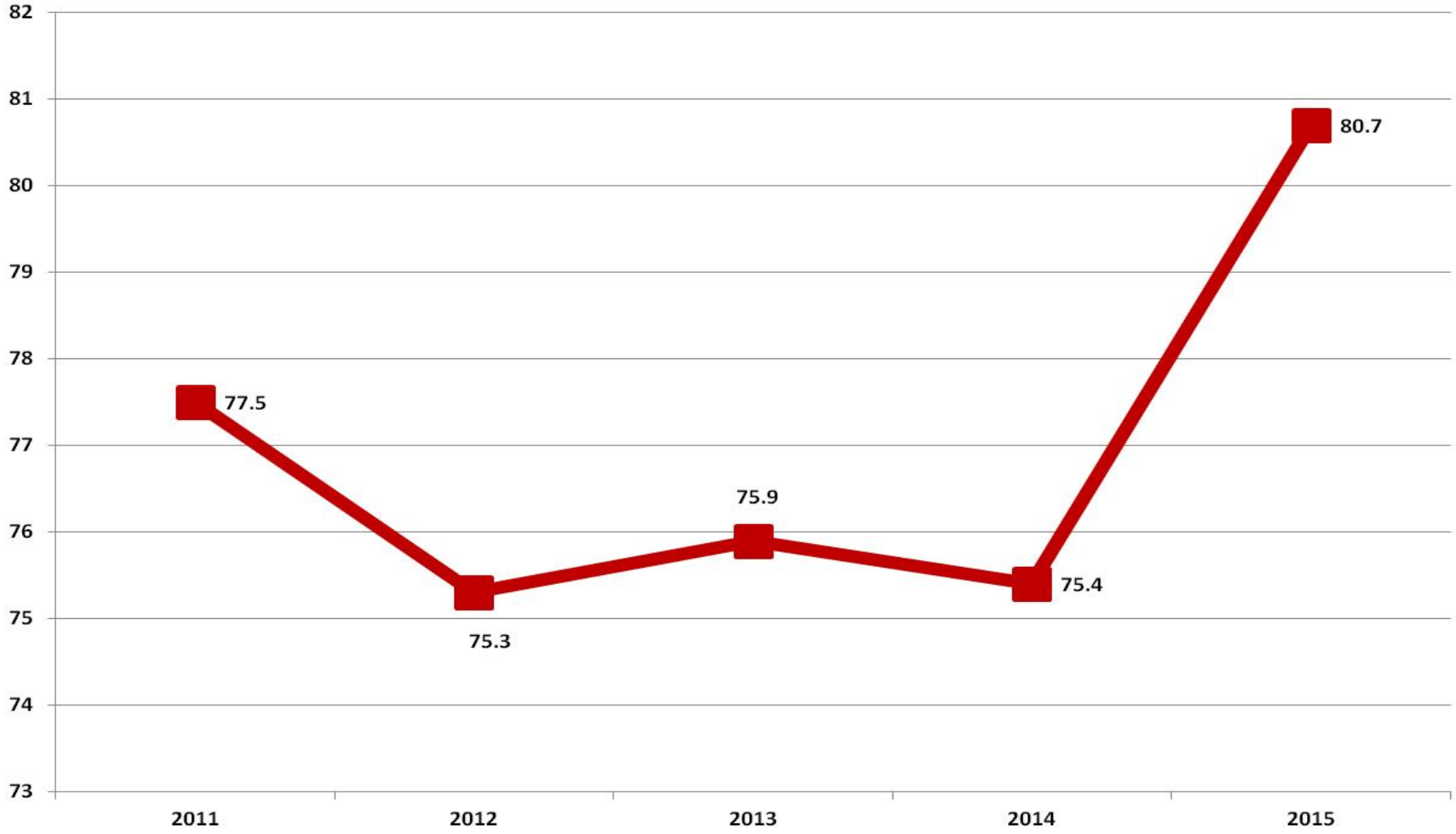
Growing of crops (large & medium v/s small & micro);

Farming of animals (large & medium v/s small & micro)

Mixed farming (large & medium v/s small & micro)

Introduction

Agricultural survey collection rate (%)



How do the survey, currently, handle unit non response?

Large & Medium Imputation

Small & Micro Reweighting

Large & Medium Imputation

Small & Micro Reweighting

Large & Medium Imputation

Small & Micro Reweighting

Imputation

How was imputation done?

- Imputation classes: Size group
- It is ordered

- Second time participated - Historic imputation (adjustment factor)
- First time participants - Ratio imputation

Imputation

- **Positives of imputation (Lacerda et al. (2007) and Lohr (1999)):**

- Resolves the missing data problem ; and
- Data analysis using standard statistical methods;

- **Negatives of imputation:**

- Treatment of imputed data as real;
- Potential to distort the association between variables (Mohadjer and Choudhry (2002)); and
- The estimated variance will be too small (Lohr, 1999) .

Weight adjustment/Reweighting

- Sampling/design weights: the reciprocal of the inclusion probability.
- Reweighting is a process of adjusting design weights to account for non response (role of response rate)
- Allow analysts to produce estimates of the total target population.

Analysis and findings

Imputation versus reweighting (Difference, GFI)

Stratum	R'000	% change	% point
111	-163,494	-38.2	-0.1
112	-49,844	-58.8	0.0
11 3	0	0	0.0
11 4	0	0	0.0
1111	2,175,065	6.5	1.7
1112	79,563	1.9	0.1
1113	0	0	0.0
1114	0	0	0.0
1121	**4,573,027**	**10.8**	**3.5**
1122	-125,107	-6.1	-0.1
1123	0	0	0.0
1124	0	0	0.0
1131	691,866	3.5	0.5
1132	-32,569	-1	0.0
1133	0	0	0.0
1134	0	0	0.0
Gross farming income	**7,148,507**	**5.5**	**5.5**

2nd

1st

Analysis and findings

Imputation versus reweighting (RSE)

Size	Imputation	Reweighting
Large farms	0.0	4.5
Medium farms	0.0	0.7
Small farms	2.9	3
Micro farms	10.8	10.9
Gross farming income	**0.6**	**3.4**

-Treatment of imputed data as real; and

-The estimated variance will be too small (Lohr, 1999) .

Analysis and findings

Industry concentration ratios

Stratum	Top5 >2.5%	Top 10 >5%	Top 20 >10%	Top 50 >25%	Top 100 50%
11 1	28.1	36.1	46.3	67.8	90.9
11 2	9.6	18.8	35.8	78.6	
11 3	2.1	4.1	8.0	18.8	34.0
11 4	8.2	15.6	29.1	60.0	94.2
1111	19.1	23.0	27.5	34.2	41.4
1112	0.5	1.0	1.9	4.7	9.3
1113	2.6	5.1	10.0	23.3	41.9
1114	5.5	10.8	20.8	46.6	76.6
1121	22.8	29.7	35.3	44.7	53.7
1122	0.9	1.8	3.6	9.0	17.7
1123	4.6	8.8	16.7	36.8	60.2
1124	6.3	12.2	23.2	49.5	80.4
1131	5.9	8.8	13.2	21.6	30.1
1132	0.6	1.2	2.4	5.9	11.7
1133	1.5	2.9	5.7	13.7	25.6
1134	5.7	11.1	21.4	48.0	78.6

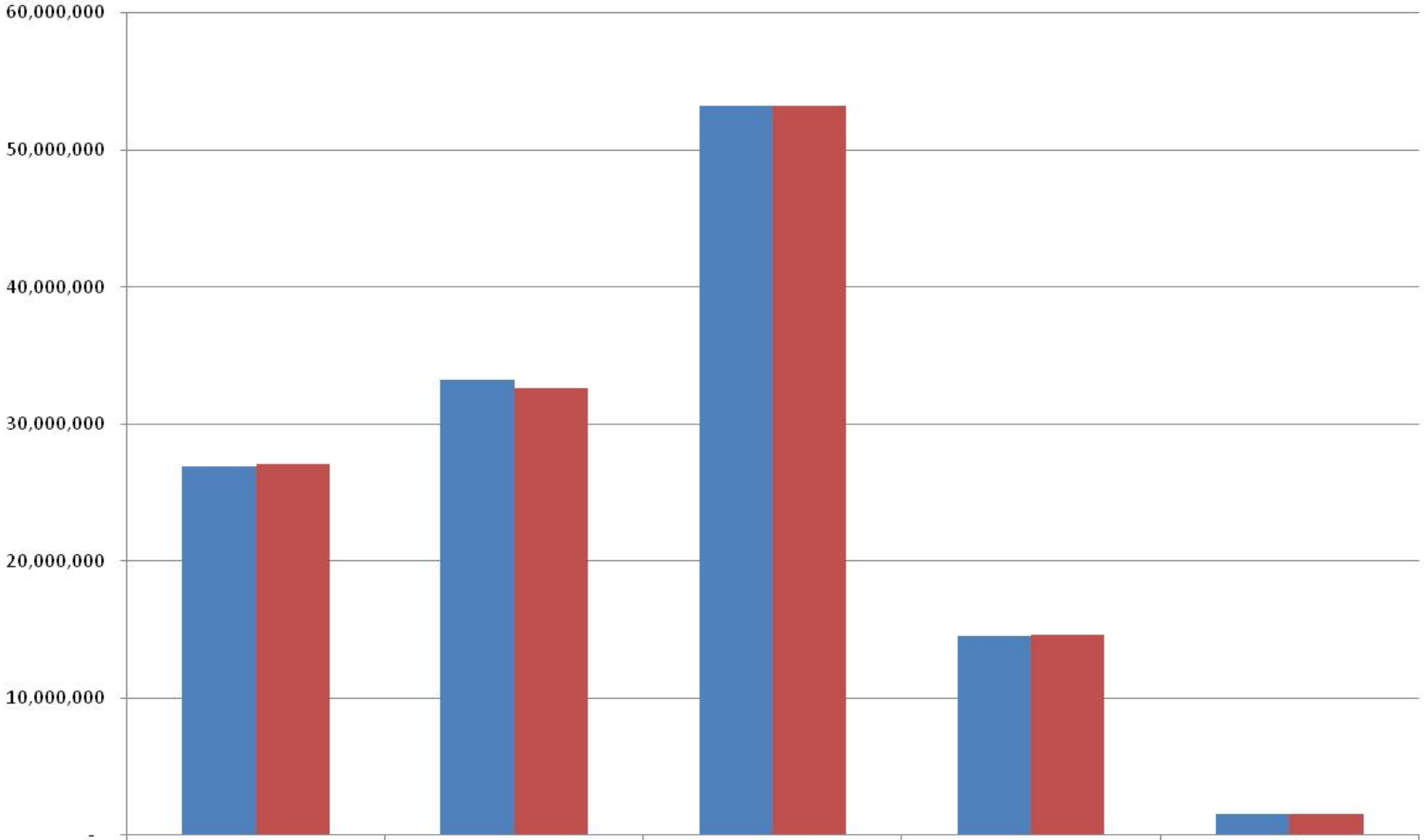
Analysis and findings

Adjusted reweighting (concentration factored) RSE of 2.9%

Stratum	R'000	% change	%point
111	-167,165	-39.1	-0.1
112	-49,026	-57.9	0.0
11 3	-122,733	-33.3	-0.1
11 4	-30,418	-25	0.0
1111	715,710	2.2	0.6
1112	82,426	2	0.1
1113	-845,929	-10.7	-0.7
1114	-226,285	-10.8	-0.2
1121	1,393,016	3.3	1.1
1122	-130,496	-6.3	-0.1
1123	-255,327	-6.5	-0.2
1124	-63,058	-3.7	0.0
1131	180,072	0.9	0.1
1132	-38,472	-1.2	0.0
1133	-477,993	-8.3	-0.4
1134	-266,980	-14.5	-0.2
Gross farming income	**-302,661**	**-0.2**	**-0.2**

Analysis and findings

Estimation of other farming income variables



	FC	Hrt	Ani	Ap	Oth
Imputatuion	26,901,534	33,204,546	53,187,040	14,531,468	1,540,058
Adj Reweighting	27,059,288	32,582,875	53,236,590	14,637,480	1,545,753

Analysis and findings

Comparison with other sources- Sampling frame

- Our expected GFI after investigations - R126.1 billion
- Agricultural method estimation –R129.4 billion (2.6%)
- Reweighting - R136.5 billion (8.2%)
- Adjusted reweighting - R129.1 billion (2.3%)

Analysis and findings

Comparison with DAFF

Source	DAFF	Imputation	Reweight	Adjusted Reweighting
	R'million			
Field Crops	27,610	26,902	28,378	27,059
Horticulture	35,490	33,205	34,138	32,583
Animals and animal products	68,599	67,719	72,373	67,874
Gross farming income	**131,699**	**129,365**	**136,513**	**129,062**

Conclusion

- The nature of South African economy makes it bit challenging to just reweight without considering concentration.
- As expected the estimation of totals is not far different that what is estimated using adjusted reweighting method.
- However we should be more concern about:

- The estimation of RSEs (0.6 v/s 2.9).
- Treatment of imputed data as real;
- Potential to distort the association between variables.
- Imputation method is bit difficult to apply than adjusted reweighting method.

Recommendation

- A move to adjusted reweighting method should be seriously considered.
- Imputation should be reserved for item non response only not unit non response.
- Adjusted reweighting method should be tested on products before can be formally adopted.