DataFirst Technical Papers

The mystery of South Africa's ghost workers in 1996: Measurement and mismeasurement in the Manufacturing Census, Population Census and October Household Surveys

by
Martin Wittenberg

Technical Paper Series
Number 1

About the Author(s) and Acknowledgments

The assistance of the Data First Resource Unit, University of Cape Town and in particular of Lynn Woolfrey and Matthew Welch is gratefully acknowledged. Taryn Dinkelman, Murray Leibbrandt, Charles Meth, Vimal Ranchhod, Charles Simkins and participants in the 2004 DPRU/Cornell University Forum "African Development and Poverty Reduction: The Macro-Micro Linkage"" provided useful feedback on an earlier version of this paper. I would also like to thank the editor of this journal and an anonymous referee for their comments. All errors are, of course, my exclusive responsibility.

Recommended citation

Wittenberg, M. (2004). The mystery of South Africa's ghost workers in 1996: Measurement and mismeasurement in the Manufacturing Census, Population Census and October Household Surveys. A DataFirst Technical Paper Number 1. Cape Town: DataFirst, University of Cape Town

---

DataFirst, University of Cape Town, Private Bag, Rondebosch, 7701, Tel: (021) 650 5708,
Email: info@data1st.org/support@data1st.org

THE MYSTERY OF SOUTH AFRICA'S GHOST WORKERS IN 1996: MEASUREMENT AND MISMEASUREMENT IN THE MANUFACTURING CENSUS, POPULATION CENSUS AND OCTOBER HOUSEHOLD SURVEYS

MARTIN WITTENBERG[1]

"Is there any point to which you would wish to draw my attention?"
"To the curious incident of the dog in the night-time."
"The dog did nothing in the night-time."
"That was the curious incident,"
remarked Sherlock Holmes.
(*The Memoirs of Sherlock Holmes*, Silver Blaze)

ABSENCES CAN BE AS TELLING AS PRESENCES, as Sherlock Holmes reminds us. Some times, however, it is difficult to know whether one is really dealing with an absence or not. In the case of South African labour economics some absences have attracted attention: the surprisingly small size of the informal sector, or the surprisingly small rate of job creation during the 1990s. To these

[1] School of Economics, University of Cape Town. The assistance of the Data First Resource Unit, University of Cape Town and in particular of Lynn Woolfrey and Matthew Welch is gratefully acknowledged. Taryn Dinkelman, Murray Leibbrandt, Charles Meth, Vimal Ranchhod, Charles Simkins and participants in the 2004 DPRU/Cornell University Forum "African Development and Poverty Reduction: The Macro-Micro Linkage" provided useful feedback on an earlier version of this paper. I would also like to thank the editor of this journal and an anonymous referee for their comments. All errors are, of course, my exclusive responsibility.

mysteries can be added another: the disappearance of about 300 000 manufacturing workers from the 1996 population census.

Good mystery stories ought to have a satisfying conclusion. In this case, however, that will turn out not to be the case. Nevertheless in trying to unravel this mystery we will have occasion to reflect on the quality of the labour statistics as available in the population census, the manufacturing census and the October Household Surveys. The fundamental lesson to be learned is that trying to tell a convincing story about the overall evolution of South Africa's labour market is difficult if we do not find some way of acknowledging the gaps. Good theories, as Sherlock Holmes also maintains, need to be able to deal with the inconsistencies and not just sweep them under the carpet.

1. COMPARING THE MANUFACTURING CENSUS 1996 AND THE POPULATION CENSUS 1996

The starting point for this investigation is a comparison of the manufacturing employment figures as recorded in the 1996 manufacturing census and the 1996 population census. The manufacturing census is a large-scale survey of manufacturing firms, while the population census is an enumeration of individuals at their point of residence. Since individuals were asked what sector they were working in, it becomes possible to compare the aggregates obtained from both. The substance of the problem can be seen in the last line of Table 1. The manufacturing census of 1996 reports a total of 1.4 million workers and an additional seven thousand "working proprietors". By contrast the population census of the same year only records 1.1 million people employed in manufacturing.

Table 1 provides more geographical information about the mismatch. It is clear that some localised mismatches are probably due to commuting patterns. For instance statistical region 01 (Cape Town) shows twenty-three thousand more jobs in manufacturing than are picked up in the population census. Ten thousand of these, however, live in the Kuils River magisterial district which falls into statistical region 02 (Boland). That still

leaves a surplus of over ten thousand manufacturing jobs in the greater Cape Town region.

Table 1. Manufacturing Employment in the Manufacturing Census and Population Census

Subregion	Manufacturing Census 1996			Population census 1996	Difference	% difference
	Number of firms	Working proprietors	Number of workers			
01 Cape Town	3,185	926	167,880	145,000	23,806	16%
02 Boland	648	180	32,232	43,911	-11,499	-26%
03 Southern Cape	113	50	3,336	4,444	-1,058	-24%
04 George, Mossel Bay	304	120	11,888	10,655	1,353	13%
05 Oudtshoorn	51	31	2,757	2,596	192	7%
06 Worcester, Ceres	174	73	8,379	9,906	-1,454	-15%
07 Saldanha	207	51	20,814	17,763	3,102	17%
08 West Coast	39	14	1,651	1,459	206	14%
09 Beaufort West	18	14	177	637	-446	-70%
10 Namaqualand	66	22	1,341	2,081	-718	-35%
11 Sutherland	9	4	66	284	-214	-75%
12 Victoria West	6	1	141	104	38	37%
13 Richmond, Colesberg	30	14	755	674	95	14%
14 Kuruman	52	21	1,114	1,036	99	10%
15 Carnarvon	12	11	105	222	-106	-48%
16 Herbert	42	20	1,303	1,206	117	10%
17 Kimberley	115	56	4,507	3,290	1,273	39%
18 Vryburg	34	14	999	2,185	-1,172	-54%
19 Lichtenburg	194	87	9,505	8,005	1,587	20%
20 Klerksdorp	142	66	4,812	3,658	1,220	33%
21 Brits, Odi	119	26	11,562	30,907	-19,319	-63%
22 Rustenburg, Madikwe	166	60	8,146	8,474	-268	-3%
23 Mmabatho	15	6	594	2,426	-1,826	-75%
24 Boshof	16	9	178	443	-256	-58%
25 Bloemfontein	263	72	14,790	13,868	994	7%
26 Smithfield	9	10	44	224	-170	-76%
27 Welkom	191	40	5,939	4,143	1,836	44%
28 Kroonstad	180	64	5,962	6,580	-554	-8%
29 Bethlehem, Harrismith	147	54	5,180	4,823	411	9%
30 Central Free State	60	53	1,683	3,106	-1,370	-44%
31 Sasolburg	55	8	7,585	5,000	2,593	52%
32 Witsieshoek	62	13	6,230	5,962	281	5%
34 Central Transkei	45	13	5,178	6,620	-1,429	-22%
35 Aliwal North	26	14	864	901	-23	-3%
37 Queenstown, Ciskei	72	23	3,256	2,995	284	9%
38 East London	409	86	36,901	29,964	7,023	23%
39 Albany	85	44	4,273	2,163	2,154	100%
40 Umzimkulu	0	0	0	710	-710	-100%
41 Cradock	21	15	397	530	-118	-22%
42 Humansdorp	55	25	2,378	2,058	345	17%
43 Port Elizabeth	780	223	65,919	47,917	18,225	38%
33, 36 & 44 Graaf-Reinet, Pondoland, Barkly East	31	26	700	3,546	-2,820	-80%
45 Steytlerville	5	6	584	144	446	310%
46 Ndwedwe	0	0	0	1,681	-1,681	-100%
47, 48 Durban/Pinetown	3,282	1,092	179,360	152,211	28,241	19%
49, 50 Nongoma	7	0	1,006	1,275	-269	-21%

51 Pietermaritzburg	431	202	26,538	19,239	7,501	39%
52 Camperdown	151	54	13,369	14,946	-1,523	-10%
53 Port Shepstone	160	92	7,104	5,010	2,186	44%
54 Underberg, East Griqualand	32	18	854	1,282	-410	-32%
55 Natal Midlands	109	47	8,127	6,642	1,532	23%
56 Ladysmith, Estcourt	171	34	22,349	13,277	9,106	69%
57 Newcastle	219	64	22,816	17,146	5,734	33%
58 Vryheid	48	14	2,191	2,977	-772	-26%
59 Richard's Bay	394	97	37,577	29,044	8,630	30%
60 Ubombo, Ingwavuma	13	7	1,688	2,133	-438	-21%
61 Highveld Ridge	80	25	12,548	9,349	3,224	34%
62 Witbank, Middelburg	206	57	17,579	15,012	2,624	17%
63 Groblersdal, Moutse	59	15	2,902	7,445	-4,528	-61%
64 Standerton, Ermelo	182	59	10,641	8,995	1,705	19%
65 Nelspruit	115	43	6,298	2,611	3,730	143%
66 Lowveld	183	52	20,441	16,007	4,486	28%
67 Phalaborwa	64	25	3,310	4,668	-1,333	-29%
68 Soutpansberg, Venda	118	76	5,527	7,327	-1,724	-24%
69 Potgietersrus	96	36	3,327	4,021	-658	-16%
70 Thabazimbi	28	8	572	1,275	-695	-55%
71 Sekhukhuneland	128	39	7,094	6,917	216	3%
72 Pietersburg	160	102	9,088	4,114	5,076	123%
73 East Rand	3,469	687	172,927	95,376	78,238	82%
74 Heidelberg, Springs	658	193	35,753	26,508	9,438	36%
75 West Rand	1,000	211	41,048	33,413	7,846	23%
76 Bronkhorstspruit	84	15	3,912	3,752	175	5%
77 Vereeniging, Vanderbijlpark	575	161	47,236	45,282	2,115	5%
78 Pretoria	1,405	283	93,736	43,000	51,019	119%
79 Johannesburg	3,908	727	141,279	83,554	58,452	70%
Total	**25,788**	**7,200**	**1,420,302**	**1,128,109**	**299,393**	

Notes:

1. Subregions correspond to the "Statistical Regions" used in the Manufacturing Census except for some in the North West province. The magisterial districts in the manufacturing census and the population census did not match up in this province (Wittenberg 2004). To get maximum comparability magisterial district 616 was incorporated with statistical region 22. This means that Rustenburg was excised from statistical region 21 and Koster from statistical region 19. Statistical region 47 was added to region 48, since Umlazi and Umbumbulu are functionally integrated with Durban. The other "joined" regions (*e.g.* 33, 36 and 44) were grouped together in the manufacturing census itself.
2. The population census data were extracted from the published "Community profile databases", Version 3. (Statistics South Africa, 1996a)

Similarly, the extra twenty thousand workers recorded by the population census in region 21 (Brits and Odi) will be workers commuting to jobs in Pretoria. These still do not offset Pretoria's total excess of fifty thousand.

These local deficits aside, in most cases the manufacturing census records substantially more manufacturing employment than the population census does. The biggest mismatches occur

in Gauteng. The East Rand seems to have 78 000 more manufacturing jobs than workers and Johannesburg's excess is about 58 000. The percentage difference (as percentage of the population census) is large in virtually all regions. What may account for these differences? There would seem to be several possible explanations:

Explanation 1 - Measurement error in the Manufacturing Census
There are several reasons why one might suspect measurement error in the manufacturing census. Firstly, as noted above, the manufacturing census is not a complete enumeration of all manufacturing firms. Instead it is a large-scale survey whose frame is provided by the business register[2]. Furthermore the business register is widely acknowledged to have been out of date. Secondly, there is always the possibility of employers providing inaccurate information.

There is little doubt that there would have been measurement error in the manufacturing census. Would one have expected these errors to be such that the employment estimates would have been overstated by 27 per cent? A number of factors would strongly militate against this assumption:

(a) As far as the first source of error is concerned, the manufacturing census is least likely out of all the available survey instruments to have captured some of the small scale manufacturing firms which may have come into existence in the 1990s. Workers in these firms should have been captured by the population census. Furthermore the business register would certainly not contain informal manufacturing enterprises. Informal sector employment is, however, covered to some extent by the census.

(b) As far as reporting error by employers is concerned, in the context of increasingly stringent labour legislation (and attendant

[2] Apparently the survey is carried out in a stratified manner with an effort being made to get 100 per cent coverage of the largest firms (personal communication, C. Meth).

skills levies) it seems more plausible that employers would understate employment to quasi-official sources than overstate it.

(c.) Moreover, in an environment in which many firms began to outsource parts of their operations (due *inter alia* to changes in the labour legislation), one would expect firms to report on their "core" workforce and not the *de facto* workforce, some of which would now consist of subcontracted employees.

In short, one would expect the manufacturing census figures to be subject to considerable error: but probably in the opposite direction. One might expect the total employment estimate in the manufacturing census to be an *underestimate* rather than an overestimate. The mystery of the disappearing manufacturing workers in fact deepens.

Explanation 2 - Perhaps the workers were not enumerated in the population census

If this were the case, it would be extremely troubling, since it would bear on the accuracy of all data sets at our disposal. The October Household Surveys, for instance, are based on sampling frames and weights derived ultimately from the population census. Nevertheless, one needs to acknowledge that there are some grounds for worry about the coverage of the 1996 population census. In the first instance, the prior information available to demarcate enumerator areas was just not of the appropriate quality. Secondly, there were large undercounts in some areas. The figures reported throughout this article, however, are the figures adjusted for the undercount.

The implications of having missed 300 000 manufacturing workers would, however, be truly staggering. Out of a population of about 40 million only about 1.4 million are manufacturing workers. If the underenumeration of these workers was representative of the entire population, one might come to the conclusion that the entire population was 27 per cent larger. All available demographic information would go counter to such a position.

One would therefore need to posit a process by which it is

specifically *manufacturing* workers that got missed in the population census. There are perhaps two circumstances in which one might expect a manufacturing worker not to be enumerated:
(a) If the worker was living as a one-person household in an urban area (*e.g.* in someone's back yard) and was not captured at census time.
(b) If the worker was an illegal immigrant and did not want to be enumerated.

The fact that the biggest mismatches recorded in Table 1 all happen in the big metropolitan areas may increase the likelihood that some of these factors may be at work. It is extremely unlikely, however, that these can explain the very large deficit. Firstly we know from other work that individuals with stable jobs are more likely to start families and become heads of households than other individuals (Klasen and Woolard 2001). This should increase rather than decrease the probability of these individuals becoming captured in the census or in standard household surveys. Secondly, if a worker was in the country illegally we would expect that person's employer not to report his or her presence either. Illegal workers can therefore not explain the large differences between the manufacturing census and the population census. In any event it would be startling if such a large proportion of the formal manufacturing work force were to consist of illegal immigrants.

In short, we would find it surprising if the workers were not enumerated. The most plausible assumption is that most manufacturing workers were, in fact, captured in the census but not recorded as being manufacturing workers.

2. COMPARING THE MANUFACTURING CENSUS 1996 TO THE OCTOBER HOUSEHOLD SURVEYS 1995-1997

Indeed, if there was a systematic tendency to "miss" manufacturing workers at their place of residence, we would expect to see a similar mismatch between the manufacturing census and the October Household Surveys. As Table 2 shows,

this is not the case.

The bottom line of that Table is more in line with the expectations outlined above in relation to the manufacturing census. According to the 1995 and 1997 OHS it appears as though total manufacturing employment should be of the order of 1.5 million. This would suggest that the manufacturing census missed about eighty thousand workers. The estimate from the 1996 OHS comes in under the manufacturing census total, but that survey was smaller than the typical OHS. Indeed the reason for reporting the estimates from the two years around 1996 is that there are concerns about how good the 1996 OHS could have been, given the demands of the census in that year.

One should note that the 1995 OHS estimates were based on the revised weights released after the 1996 census. The population aggregates are therefore broadly consistent with each other. Indeed all of the OHS estimates are dependent on weights derived ultimately from the population census. The major reason for the mismatches is the fact that the OHSs find higher *proportions* of the sampled workforce and population involved in manufacturing.

In Table 2 we also report provincial totals. We observe that these also correspond much more closely with the provincial totals from the manufacturing census. The most notable exception is North-West, which has much lower recorded employment in the census than in any of the household surveys. The 1996 census total is not in any of the 95 per cent confidence intervals for the household surveys. The reason for this may be cross-border commuting from the Winterveld area.

In each survey, however, there are other mismatches which are less easy to rationalise. In 1995 the estimated level of manufacturing employment in the Eastern Cape is significantly below the 1996 manufacturing census figures. In 1996 it is Gauteng that is significantly below. In 1997 the Free State and Limpopo are significantly above the 1996 benchmark. If one believes that even in the provincial breakdowns the manufacturing census is likely to give an underestimate,

Table 2. Manufacturing employment by province in the Manufacturing Census and the October Household Surveys

Province	Manufacturing Census 1996	October Household Surveys											
		1995	95% confidence interval		In band?	1996	95% confidence interval		In band?	1997	95% confidence interval		In band?
Western Cape	250,573	279,759	225,244	334,274	Yes	250,912	186,876	314,948	Yes	269,024	224,387	313,661	Yes
Eastern Cape	120,925	91,218	74,567	107,868	**No**	154,299	105,623	202,975	Yes	128,496	98,141	158,851	Yes
Northern Cape	9,481	12,203	6,174	18,231	Yes	15,669	5,315	26,023	Yes	15,407	9,253	21,561	Yes
Free State	47,914	42,612	32,646	52,578	Yes	54,036	34,391	73,681	Yes	67,740	50,425	85,055	**No**
KwaZulu Natal	324,700	336,865	283,799	389,931	Yes	295,505	220,035	370,975	Yes	333,660	286,700	380,620	Yes
North West	35,877	67,084	49,281	84,886	**No**	88,863	53,437	124,289	**No**	93,546	73,193	113,899	**No**
Gauteng	538,168	556,001	474,709	637,292	Yes	412,431	356,247	468,616	**No**	495,165	438,415	551,915	Yes
Mpumalanga	70,660	86,775	56,368	117,183	Yes	86,487	51,193	121,781	Yes	80,190	62,085	98,295	Yes
Limpopo	29,204	29,826	20,532	39,121	Yes	44,689	27,022	62,356	Yes	45,479	33,024	57,934	**No**
Total	**1,427,502**	**1,502,342**	1,324,661	1,680,023	Yes	**1,402,891**	1,181,226	1,624,556	Yes	**1,528,707**	1,444,715	1,612,699	Yes

Notes:

1. The figures for the 1995 OHS were calculated with 1996 weights.

3. The confidence intervals were calculated taking into consideration the cluster design of the surveys, but disregarding stratification.

then relying exclusively on the household surveys one would have made serious errors in at least 1995 and 1996. The fundamental point is that estimating totals from survey evidence is likely to be a tricky business, given the fact that the estimates will be sensitive to the weights used. For instance, an underestimate of 10 per cent in the weights, provided it is across the board, will not affect any of the proportions or means estimated on a sample, but will result in all totals being out by 10 per cent.

Furthermore given the clustered nature of the data, the standard errors around the totals are likely to be large. This is very evident in the 95 per cent confidence intervals given in Table 2. Indeed looking across the three surveys it is clear that not only are the confidence intervals large, but the point estimates jump around a lot. It is hard to believe that these are "real" changes. Given that the household surveys are supposedly designed to give reasonably accurate provincial breakdowns of variables, Table 2 would make one hesitant about placing too much weight on trends in the estimate of aggregates across different household surveys.

To return to the main topic of this article, however, we note that there is no evidence to suggest that household based data gathering exercises would tend to miss manufacturing workers. If anything the data from the surveys suggest that the manufacturing census estimate of 1.42 million workers is probably an underestimate.

Explanation 3 – Misclassification of employment status in the Population Census

One possibility is that manufacturing workers may have been erroneously classified as unemployed or not economically active. *A priori* this seems unlikely. The manufacturing census is likely to capture mainly stable, formal sector jobs. Such workers are the most likely to be reported as employed in the population census. Indeed, the aggregate employment estimate from the population census (9.2 million) is in the same ballpark as the aggregates from the three household surveys (10 million, 9.2 million and 9.2 million

respectively). One would expect that if 300 000 formal manufacturing jobs were reassigned to the non-employed this would happen across a wider spectrum of formal sector work. This would imply that we should see much higher rates of unemployment or non-employment in the population census than in the household surveys.

A more subtle possibility is that many manufacturing workers were enumerated under the "institutional" population[3]. This is a category designed for prisons, hospitals or similar dwellings. Their residents were not asked questions about their employment status or industry. About 700 000 South Africans were enumerated under this category. It is conceivable that some other communal dwelling places (such as hostels) might have been erroneously enumerated under this category. On closer scrutiny, however, this possibility cannot solve the puzzle. The area in which the greatest deficit is recorded, Gauteng, only has an institutional population of 159 758. Once allowance is made for Gauteng's hospitals and prisons, there is little leeway left to explain the manufacturing employment deficit.

Explanation 4 - Misclassification of industrial sector in the Population Census

This is perhaps the most likely suspect. The classification of workers into industries was based on the written answers to the question:

"Describe the MAIN INDUSTRY, economic activity, produce or service of the person's employer or company, *e.g.* gold mining, road construction, supermarket, police service, hairdressing, banking; OR activity of the person, if self-employed *e.g.* subsistence farming"

Since the census questionnaire was self-administered it is hardly surprising that the quality of the answers to this question turned out to be rather poor. Indeed the population census contains 1.1 million individuals in the category "Industry not elsewhere classified or unspecified". Given this, there is room to believe that the disappearance of the manufacturing workers may be

[3] I am indebted to Charles Simkins for pointing out this possibility.

more apparent than real.

As Sherlock Holmes would have warned, however, it is wise to subject even the most obvious explanations to further scrutiny. If the manufacturing workers have simply been re-recorded under the "Industry NEC classified or unspecified" category, then the sectoral and regional distributions should be consistent with this story.

4. COMPARING THE POPULATION CENSUS 1996 TO THE OCTOBER HOUSEHOLD SURVEYS 1995-1997

A look at Table 3 starts to raise serious questions about this easy explanation. The first point to note is that it is not only manufacturing employment that is significantly below the point estimates emanating from the Household Surveys. Wholesale and Retail Trade and Services are also much lower. Indeed if we average the figures for the three household surveys and compare them with the population census, we note that there is a deficit of about 500 000 retail and wholesale workers and about 460 000 service workers. Given that the "residual" category is never less than 200 000, these two deficits are sufficient to explain the entire residual category in the population census!

Looking at the sectoral breakdowns one is inclined to become much more sceptical about the prospects of doing any aggregate analysis on any of these data. The fact that mining employment in the population census is significantly higher than it is in any of the surveys speaks in favour of the census. It is well known (Klasen and Woolard 1999) that the surveys were badly designed to pick up mining employment.

The fact that the census picks up much more employment in construction and in private households is worrying more for the quality of the household surveys than for that of the census. It is hard to see how manufacturing sector workers could erroneously have been placed into these categories. Instead one can envisage how the household surveys may have missed some domestic workers and small-scale construction workers. In particular years, there are other startling differences.

Table 3. Employment by industry according to the Population Census 1996 and the October Household Surveys

Sector	Population Census	October Household Surveys											
	1996	1995	95% confidence interval		In band?	1996	95% confidence interval		In band?	1997	95% confidence interval		In band?
Agriculture	832,911	1,285,322	1,063,529	1,507,115	**No**	782,128	484,411	1,079,845	Yes	775,251	692,312	858,190	Yes
Mining	543,297	452,619	366,544	538,695	**No**	252,702	173,187	332,217	**No**	395,731	338,642	452,820	**No**
Manufacturing	1,128,109	1,502,342	1,324,661	1,680,023	**No**	1,402,891	1,181,226	1,624,556	**No**	1,528,707	1,444,715	1,612,699	**No**
Electricity	110,196	87,326	67,968	106,685	**No**	126,820	93,877	159,763	Yes	115,627	96,079	135,175	Yes
Construction	560,124	471,137	406,757	535,516	**No**	435,650	349,962	521,338	**No**	521,764	484,839	558,689	**No**
Trade	1,109,110	1,765,157	1,554,080	1,976,234	**No**	1,435,414	1,206,115	1,664,713	**No**	1,604,333	1,526,327	1,682,339	**No**
Transport	487,474	492,081	425,001	559,161	Yes	483,956	402,946	564,966	Yes	530,669	490,206	571,132	**No**
Finance	687,646	604,795	512,596	696,994	Yes	757,211	624,676	889,746	Yes	732,705	669,475	795,935	Yes
Services	1,594,542	2,247,757	2,009,593	2,485,921	**No**	2,030,902	1,722,554	2,339,250	**No**	1,894,277	1,802,105	1,986,449	**No**
Private households	1,065,052	847,394	735,598	959,190	**No**	817,770	621,756	1,013,784	**No**	765,973	716,006	815,940	**No**
Industry NEC		2,749	1,072	4,427									
Foreign govts	3,776												
NEC+Unspecified	1,108,484	200,624	166,840	234,407		625,892	528,063	723,721		360,743	326,023	395,463	
Total	**9,230,721**	**9,959,302**				**9,151,336**				**9,225,780**			

Note:

1. Industry classifications used in the October Household Surveys were the derived classifications supplied by Statistics South Africa
2. Estimates were weighted using the person weights supplied by Statistics South Africa (1996 weights were used on the 1995 OHS).
3. Confidence intervals were based on standard errors corrected for clustering, but not stratification.

The 1995 OHS picks up about four hundred thousand more agricultural workers than the census, or the other two household surveys, for that matter. It is hard to avoid the conclusion that the sampling frame for the 1995 OHS may have been less than perfect.

Similarly, the confidence intervals around the services sector figures for 1995 and 1997 do not overlap. One may conclude that employment in services collapsed between 1995 and 1997. With the increase in outsourcing by formal sector firms during that period, this would be surprising, to say the least.

In short, the sectoral breakdowns do not support the simple hypothesis "Household Surveys Good, Population Census Bad". It seems much more likely that there are strengths and weaknesses in either. The strength of the census is its coverage: there is much less room for doubt about whether the sampling frames captured particular types of individuals and households. One also does not have to be so concerned about whether the final results just happen to be driven by the particular sets of weights imposed by Statistics South Africa. The weakness of the census is that there would have been much less probing and checking of responses.

The key question in relation to our main puzzle is therefore whether one would expect misclassification of manufacturing workers to drive the "gap" that we observed between the population census and the manufacturing census. The first point we have already noted: we can explain virtually all of the "not elsewhere classified and unspecified" workers in the population census in terms of wholesale and retail trade and services. It seems much more plausible that such workers would be misclassified given the lack of probing in the population census. The first cross-check of the misclassification story against the data therefore already raises doubts as to whether the mystery can be so easily solved. We now turn to another consistency check of the "misclassification" story.

5. COMPARING THE POPULATION CENSUS 1996 TO THE MANUFACTURING CENSUS (AGAIN)

The second, and more searching, test of the explanation is provided by the provincial breakdowns given in Table 4. The biggest deficit in manufacturing employment census is in Gauteng. We would require fully 75 per cent of everyone who was recorded under "industry not elsewhere classified and unspecified" in that province to have been a manufacturing worker, in order to get the population census and manufacturing census figures into alignment. In the other provinces, by contrast, we would require only small percentages of the residual category to be manufacturing workers. In short we would require different processes of misclassification to be operating in Gauteng than in the other regions.

Table 4. Manufacturing Employment in the manufacturing census and population census by province

Province	Manufacturing Census	Population Census	Difference	Population Census: Industry NEC or unspecified
Western Cape	250,573	236,371	14,202	130,096
Eastern Cape	120,925	97,548	23,377	128,254
Northern Cape	9,481	8,897	584	19,650
Free State	47,914	44,149	3,765	60,146
KwaZulu Natal	324,700	266,863	57,837	294,240
North-West	35,877	55,655	-19,778	61,990
Gauteng	538,168	330,885	207,283	277,193
Mpumalanga	70,660	59,419	11,241	62,068
Limpopo	29,204	28,322	882	74,847
	1,427,502	1,128,109		

The simplest explanation that is congruent with these figures is the one advanced above: the residual category consists *largely* of individuals engaged in trade or services. It may perhaps contain a few misclassified manufacturing workers. It seems unlikely, however, that it will come close to making up the deficit of 300 000.

Explanation 5 - The author's data are corrupt and/or he is incapable of analysing them correctly

This explanation has much to recommend it. Indeed as Sherlock

Holmes remarks, once all other possibilities have been eliminated, the one that remains has to be the truth. Regrettably this explanation would doom a few other analysts also. The sectoral figures given in Table 3 are similar to figures reported by Klasen and Woolard for 1995 (with different weights) and earlier years. The aggregate employment figures for the economy as a whole square with those given by Bhorat (2003) and those for total manufacturing employment are very similar to figures reported by Bhorat (2001). If there is a simple error, it is in the way the 1996 population census data have been extracted, courtesy of the SuperCross software. That seems unlikely, however.

Fewer analysts have worked on the manufacturing census, but since those figures are more easily reconciled with the October Household Surveys, there seems at this stage less cause for doubting them.

Perhaps the "error" is deeply embedded in the sampling frames and weights used by Statistics South Africa. That would let this author off the hook, but would imply that most work relying on these data sets needs to be treated with caution. Without further information about the inside workings of these data sets, we cannot be sure.

Effectively this means that the mystery remains: where did at least three hundred thousand manufacturing workers disappear to in the 1996 population census? Some of them (maybe as much as a hundred and fifty thousand) may have been misclassified under "unspecified". It still leaves a lot of unaccounted workers. Even more so, if one supposes that the true total should have been closer to 1.5 million.

How important is this mystery? My intuition is that it is extremely important. If we cannot reconcile the aggregates in one year, *viz.* 1996, we have absolutely no chance of talking meaningfully about changes in the aggregates from year to year. As it stands we have three different pictures of what happened in that year: the manufacturing census, the population census, and a household survey. By comparing them we should get much more

"traction" than we do for other years. The fact that we can't reduce the uncertainty is troubling, to say the least. The evidence seems unambiguous that the population census number is far too low. It is much more difficult to pin down where those workers have gone to.

6. CONCLUSION

The disappearance of 300 000 workers would seem to be a mystery worthy of investigation by Sherlock Holmes himself. Unfortunately, at the end of the day this author is incapable of solving it. Nevertheless a number of useful insights have emerged from this investigation:

1. The Manufacturing Census figure of 1,427,502 employees and working proprietors seems like a realistic lower bound on manufacturing employment figures for 1996.

2. The sectoral aggregates for manufacturing, wholesale and retail trade and services recorded in the population census seem too low. The shortfalls in the latter two categories can probably be largely explained in terms of misclassification.

3. The sectoral totals estimated from the household surveys record too few mining workers and possibly too few construction workers and workers employed by private households.

4. The estimates of totals derived from the household surveys can be extremely noisy. It looks as though standard techniques may often give an *underestimate* of the true variability, even if one takes into consideration the clustered design. The additional variability possibly arises from:

(a) Coverage problems

Household surveys inevitably leave some sections of the population out. This is obvious in the case of homeless people, but it may also be true of single-person households and other fairly mobile parts of the population.

(b) Weighting issues

The weights used in the household surveys are not true sampling

weights. Instead they are designed to reproduce the demographic structure of the population. It is not clear that these weights are ideal for labour force analysis.

(c) Fieldwork errors and other types of non-sampling errors.

Although the household surveys are repeated in similar ways across the years it is not clear that the fieldwork is always performed to the same standard.

These are potentially extremely important points, since many analysts wish to compare totals between years. Indeed few of the exercises that I have seen have attempted to estimate even simple standard errors for the totals.

(5) Census information is potentially an extremely useful check on the survey data, particularly since one can do much more disaggregated analyses.

(6) Being able to cross-check information from household survey type instruments against employer provided information (such as in the Manufacturing Census) can be extremely valuable. Regrettably the 1996 census of manufacturing is the most recent available data set.

We have identified the absence. Now we just need to find those missing workers....

REFERENCES

BHORAT, HAROON (2001). "Explaining Employment Trends in South Africa: 1993 -1998", mimeo, Johannesburg: Friedrich Ebert Stiftung.

____ (2003). "The Post-Apartheid Challenge: Labour Demand Trends in the South African Labour Market, 1995-1999", DPRU Working Paper 03/82, Cape Town: Development Policy Research Unit, University of Cape Town.

CONAN DOYLE, SIR ARTHUR (1981). *The Penguin Complete Sherlock Holmes*, Harmondsworth: Penguin Books. First published 1930.

KLASEN, STEPHAN AND INGRID WOOLARD (1999). "Levels, Trends and Consistency of Employment and Unemployment Figures in South Africa", *Development Southern Africa,* 16: 3-35.

____ (2001). "Surviving Unemployment without State Support. Unemployment and Household Formation in South Africa". CESifo Working Paper No. 533.

STATISTICS SOUTH AFRICA (1996a). *Population Census, 1996. Community profile databases,* Version 3. Report 03-01-30 (1996), Pretoria: Statistics South Africa.

_____ (1996b). *Census of Manufacturing 1996: Principal Statistics on a regional basis, Western Cape,* Report 30-01-04 (1996), Pretoria.

____ *Census of Manufacturing 1996: Principal Statistics on a regional basis, Eastern Cape*, Report 30-01-05 (1996), Pretoria.

_____ *Census of Manufacturing 1996: Principal Statistics on a regional basis, Northern Cape*, Report 30-01-06 (1996), Pretoria
_____*Census of Manufacturing 1996: Principal Statistics on a regional basis, Free State,* Report 30-01-07 (1996), Pretoria.
_____*Census of Manufacturing 1996: Principal Statistics on a regional basis, KwaZulu/Natal*, Report 30-01-08 (1996), Pretoria.
_____ *Census of Manufacturing 1996: Principal Statistics on a regional basis, North-West*, Report 30-01-09 (1996), Pretoria.
_____*Census of Manufacturing 1996: Principal Statistics on a regional basis, Gauteng*, Report 30-01-10 (1996), Pretoria.
_____*Census of Manufacturing 1996: Principal Statistics on a regional basis, Mpumalanga*, Report 30-01-11 (1996), Pretoria.
_____*Census of Manufacturing 1996: Principal Statistics on a regional basis, Northern Province*, Report 30-01-12 (1996), Pretoria.
WITTENBERG, MARTIN (2004). "The 1996 Manufacturing Census statistics imputed by municipality and district", mimeo, Cape Town: School of Economics, University of Cape Town.

About DatatFirst

DataFirst is a research unit at the University of Cape Town engaged in promoting the long term preservation and reuse of data from African Socioeconomic surveys. This includes:

- the development and use of appropriate software for data curation to support the use of data for purposes beyond those of initial survey projects
- liaison with data producers - governments and research institutions - for the provision of data for reanalysis
- research to improve the quality of African survey data
- training of African data managers for better data curation on the continent
- training of data users to advance quantitative skills in the region.

The above strategies support a well-resourced research-policy interface in South Africa, where data reuse by policy analysts in academia serves to refine inputs to government planning.

DataFirst

www.datafirst.uct.ac.za
Level 3, School of Economics Building, Middle Campus, University of Cape Town
Private Bag, Rondebosch 7701, Cape Town, South Africa
Tel: +27 (0)21 650 5708