The Sudan 2014 Enterprise Surveys Data Set

I. Introduction

1.   This document provides additional information on the data collected in Sudan between September 2014 and February 2015 under, an initiative of the World Bank. As part of its strategic goal of building a climate for investment, job creation, and sustainable growth, the World Bank has promoted improving business environments as a key strategy for development, which has led to a systematic effort in collecting enterprise data across countries. The Enterprise Surveys (ES) are an ongoing World Bank project in collecting both objective data based on firms' experiences and enterprises' perception of the environment in which they operate.

The Enterprise Surveys currently cover over 130,000 firms in 135 countries, of which 121 have been surveyed following a standard methodology. This allows for better comparisons across countries and across time. Data are used to create statistically significant business environment indicators that are comparable across countries. The Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

The report outlines and describes the sampling design of the data, the data set structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

II. Sampling Structure

2.   The sample for Sudan was selected using stratified random sampling, following the methodology explained in the *Sampling Manual*[1]. Stratified random sampling[2] was preferred over simple random sampling for several reasons[3]:

a. To obtain unbiased estimates for different subdivisions of the population with some known level of precision.

b. To obtain unbiased estimates for the whole population. The whole population, or universe of the study, is the non-agricultural economy. It comprises: all manufacturing sectors according to the group classification of ISIC Revision 3.1: (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this definition excludes the following sectors: financial intermediation (group J), real estate and renting activities (group K, except sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

c. To make sure that the final total sample includes establishments from all different sectors and that it is not concentrated in one or two of industries/sizes/regions.

d. To exploit the benefits of stratified sampling where population estimates, in most cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)

e. Stratification may produce a smaller bound on the error of estimation than would be produced by a simple random sample of the same size. This result is particularly true if measurements within strata are homogeneous.

f. The cost per observation in the survey may be reduced by stratification of the population elements into convenient groupings.

---

[1] The complete text can be found at http://www.enterprisesurveys.org/documents/Implementation_note.pdf

[2] A stratified random sample is one obtained by separating the population elements into non-overlapping groups, called strata, and then selecting a simple random sample from each stratum. (Richard L. Scheaffer; Mendenhall, W.; Lyman, R., "Elementary Survey Sampling", Fifth Edition).

[3] Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95

3. Three levels of stratification were used in this country: industry, region, and size was not available in the sampling frame for most contacts. The original sample design with specific information of the industries and regions chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into manufacturing industries and two service sectors (retail and other services).

6. Regional stratification for the Sudan ES was defined in four regions:
- Bahri
- Khartoum
- Om Durman
- Other/Unknown

III. Sampling implementation

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and region) are required to draw the sample.

8. The international firm of Ipsos was hired to conduct the survey and they partnered with local agency ConsulSat for Information Services in Sudan.

9. For the Sudan ES, a sample frame was built using data from UNIDO as well are municipal commercial registries, import/export registries, and Chambers of Commerce.

Sudan, Sample Frame

	Size	Manufacturing	Other Services	Retail	Wholesale	Total
Bahri	1- Small (5 to 19)	44	23	20	47	134
Bahri	2- Medium (20 to 99)	32	18	3	9	62
Bahri	3- Large (100+)	23	6			29
Bahri	4-Size Unknown	126	29	64	47	266
Khartoum	1- Small (5 to 19)	357	255	178	433	1223
Khartoum	2- Medium (20 to 99)	135	120	20	97	372
Khartoum	3- Large (100+)	41	29	3	3	76
Khartoum	4-Size Unknown	649	491	673	731	2544
Om Durman	1- Small (5 to 19)	97	12	88	146	343
Om Durman	2- Medium (20 to 99)	41	3	9	20	73
Om Durman	3- Large (100+)	12	3		3	18
Om Durman	4-Size Unknown	149	53	214	140	556
Other/Unknown	1- Small (5 to 19)	26	18	20	53	117
Other/Unknown	2- Medium (20 to 99)	9	6		12	27
Other/Unknown	3- Large (100+)	9	3		3	15
Other/Unknown	4-Size Unknown	123	881	205	369	1578
	Total	1873	1950	1497	2113	7433

10. The sample design for the Sudan Enterprise Survey was generated with the aim of obtaining interviews at 720 establishments. Establishments with undefined size were included as part of this sample frame for Sudan in order to ensure a representative sample. Size information collected during the survey process can then be used to categorize these firms. Establishments with undefined location were included as part of this sample frame for Sudan in order to ensure a representative sample. Location information collected during the survey process can then be used to categorize these firms. It was found to be common that firms register a PO box in Khartoum but operate elsewhere in the country.

11.   The quality of the frame was assessed at the onset of the project through visits to a random subset of firms and local contractor knowledge. The sample frame was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc. Initial sample frame was of very poor quality and lacked useful stratification information. Before fieldwork began, the local contractor performed a verification exercise to obtain as much stratification information as possible. The resulting sample design was then based off of this verified information. The local contractor had to screen the contacts by visiting them which resulted in slow fieldwork in many cases.

12.   Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. Breaking down by stratified industries, the following sample targets were achieved:

Achieved sample

	Size	Manufacturing	Other Services	Retail	Wholesale	Total
Bahri	1- Small (5 to 19)	4	1	0	0	5
Bahri	2- Medium (20 to 99)	5	0	0	0	5
Bahri	3- Large (100+)	5	0			5
Bahri	4-Size Unknown	11	2	1	2	16
Khartoum	1- Small (5 to 19)	59	32	19	41	151
Khartoum	2- Medium (20 to 99)	18	16	2	12	48
Khartoum	3- Large (100+)	3	4	0	0	7
Khartoum	4-Size Unknown	78	41	42	63	224
Om Durman	1- Small (5 to 19)	19	2	11	24	56
Om Durman	2- Medium (20 to 99)	8	0	1	2	11
Om Durman	3- Large (100+)	2	1		1	4
Om Durman	4-Size Unknown	20	8	25	15	68
Other/Unknown	1- Small (5 to 19)	0	0	1	0	1
Other/Unknown	2- Medium (20 to 99)	0	0		0	0
Other/Unknown	3- Large (100+)	0	0		0	0
Other/Unknown	4-Size Unknown	7	31	11	12	61
	Total	239	138	113	172	662

IV. Data Base Structure:

13.   The structure of the data base reflects the fact that 2 different versions of the survey instrument were used for all registered establishments. Questionnaires have common questions and respectfully additional manufacturing and services specific questions. The eligible manufacturing industries have been surveyed using the ***Manufacturing*** questionnaire (includes a common set of core variables, plus manufacturing specific questions). Eligible services have been covered using the ***Services*** questionnaire. Each variation of the questionnaire is identified by the index variable, *a0*.

14.   All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section *A*, question *1* (some exceptions apply due to comparability reasons). Variable names proceeded by a prefix *"SL"* indicate questions specific to Sudan, therefore, they may not be found in the implementation of the rollout in other countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an "x" at the end of their names. The suffix "x" denotes that the variable is alpha-numeric.

15. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling region), *a6a* (sampling establishment's size), and *a4a* (sampling sector) contain the establishment's classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

16. There are three levels of stratification: industry, size, and region. Different combinations of these variables generate the strata cells for each industry/region/size combination. A distinction should be made between the variable a4a and d1a2 (industry expressed as ISIC rev. 3.1 code). The former gives the establishment's classification into one of the chosen industry-strata, whereas the latter gives the actual establishment's industry classification (four digit code) in the sample frame.

17. All of the following variables contain information from the sampling frame. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.
     -*a2* is the variable describing sampling regions
     -*a6a*: coded using the same standard for micro, small, medium, and large establishments as defined above. The code -*9* was used to indicate units for which size was undetermined in the sample frame.
     -*a4a*: coded using ISIC codes for the chosen industries for stratification. These codes include most manufacturing industries (15 to 37), other manufacturing (2), retail (52), and (45, 50, 51, 55, 60, 63, 72) for other Services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. In the case of Sudan, this screener was administered face-to-face. Then a face-to-face interview takes place with the Manager/Owner/Director of each establishment. However, the phone numbers were unavailable in the sample frame, and thus the enumerators applied the screeners in person. The variables *a4b* and *a6b* contain the industry and size of the establishment from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

19. Note that there are variables for size (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes. Variables *l1*, *l6* and *l8* were designed to obtain a more accurate measure of employment accounting for permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

20. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results. Please note that sometimes this variable is removed due to privacy issues.

V. Universe Estimates

21. Universe estimates for the number of establishments in each cell in Sudan were produced for the strict, weak and median eligibility definitions. The estimates were the multiple of the relative eligible proportions.

23.   For some establishments where contact was not successfully completed during the screening process (because the firm has moved and it is not possible to locate the new location, for example), it is not possible to directly determine eligibility. Thus, different assumptions about the eligibility of establishments result in different adjustments to the universe cells and thus different sampling weights.

24. Three sets of assumptions on establishment eligibility are used to construct sample adjustments using the status code information.

25. Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights are included in the variable *wstrict*.

Strict eligibility = (Sum of the firms with codes 1,2,3,4,&16) / Total

26. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *wmedian*.

Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

27. Weak assumption: in addition to the establishments included in points a and b, all establishments for which it was not possible to contact or that refused the screening questionnaire are assumed eligible. This definition includes as eligible establishments with dead or out of service phone lines, establishments that never answered the phone, and establishments with incorrect addresses for which it was impossible to find a new address. Under the weak assumption only observed non-eligible units are excluded from universe projections. The resulting weights are included in the variable *wweak*.

Weak eligibility= (Sum of the firms with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

28. The indicators computed for the Enterprise Survey website use the median weights. The following graph shows the different eligibility rates calculated for firms in the sample frame under each set of assumptions.

Eligibility Rates According to Assumptions Percent Eligible

Sudan ES, 2014

	Strict Assumptions	Median Assumptions	Weak Assumptions
	31.3%	55.9%	83.8%
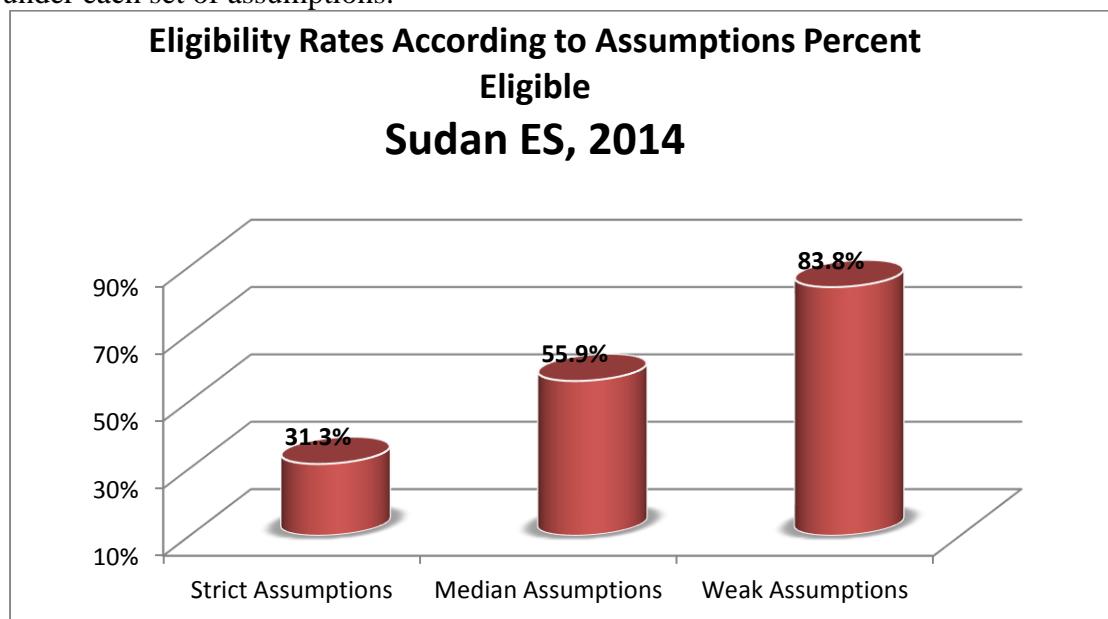
29. Universe estimates for the number of establishments in each industry-region-size cell in Sudan were produced for the strict, weak and median eligibility definitions. Appendix D shows the universe estimates of the numbers of registered establishments that fit the criteria of the Enterprise Surveys.

30. Once an accurate estimate of the universe cell projection was made, weights for the probability of selection were computed using the number of completed interviews for each cell.

VI. Weights
31.   Since the sampling design was stratified and employed differential sampling, individual observations should be properly weighted when making inferences about the population. Under stratified random sampling, unweighted estimates are biased unless sample sizes are proportional to the size of each stratum. With stratification the probability of selection of each unit is, in general, not the same. Consequently, individual observations must be weighted by the inverse of their probability of selection (probability weights or *pw* in Stata.)[4]

32.   Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each region/industry/size stratum to account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, no tone in the phone line, answering machine, fax line[5], wrong address or moved away and could not get the new references) The information required for the adjustment was collected in the first stage of the implementation: the screening process. Using this information, each stratum cell of the universe was scaled down by the observed proportion of ineligible units within the cell. Once an accurate estimate of the universe cell (projections) was available, weights were computed using the number of completed interviews.

VII. Appropriate use of the weights
33.   Under stratified random sampling weights should be used when making inferences about the population. Any estimate or indicator that aims at describing some feature of the population should take into account that individual observations may not represent equal shares of the population.

34.   However, there is some discussion as to the use of weights in regressions (see Deaton, 1997, pp.67; Lohr, 1999, chapter 11, Cochran, 1953, pp.150). There is not strong large sample econometric argument in favor of using weighted estimation for a common population coefficient if the underlying model varies per stratum (stratum-specific coefficient): both simple OLS and weighted OLS are inconsistent under regular conditions. However, weighted OLS has the advantage of providing an estimate that is independent of the sample design. This latter point may be quite relevant for the Enterprise Surveys as in most cases the objective is not only to obtain model-unbiased estimates but also design-

---

[4] This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.
[5] For the surveys that implemented a screener over the phone.

unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)[6]

35.   From a more general approach, if the regressions are descriptive of the population then weights should be used. The estimated model can be thought of as the relationship that would be expected if the whole population were observed.[7] If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

VIII. Non-response
36.   Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to address these issues.

37.   Item non-response was addressed by two strategies:
a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as a different option from don't know (-7).
b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. However, there were clear cases of low response. The following graph shows non-response rates for the sales variable, $d2$, by sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between "Don't know" and "refuse to answer", thus the non-response in the chart below for both enterprise surveys (ES) reflect both categories (DKs and NAs). Sudan suffered a relatively high item non-response rate for sales despite callback attempts.

Sales Non-response Rates
Sudan ES, 2014

---
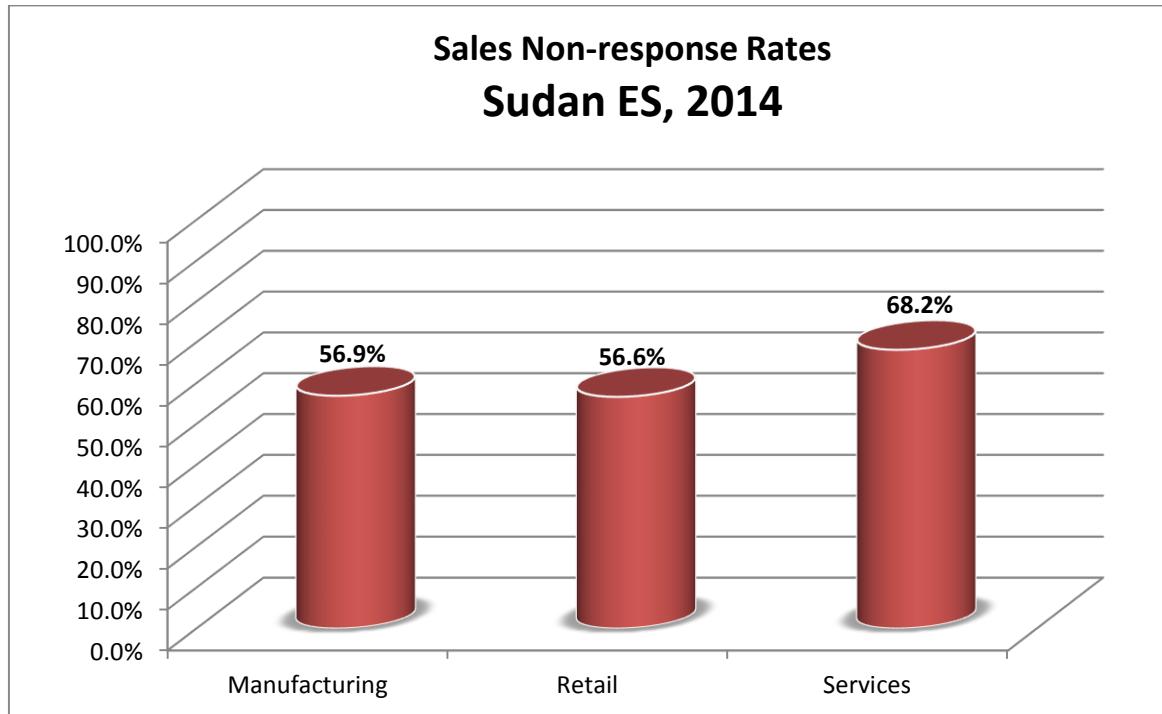
[6] Note that weighted OLS in Stata using the command regress with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands *svy* will provide appropriate standard errors.
[7] The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

38.   Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) was suggested for interview. Survey non-response did occur but substitutions were made in order to potentially achieve strata-specific goals. Further research is needed on survey non-response in the Enterprise Surveys regarding potential introduction of bias.

39.   As the following graph shows, the number of interviews per contacted establishments was 0.26[8]. This number is the result of two factors: explicit refusals to participate in the survey, as reflected by the rate of rejection (which includes rejections of the screener and the main survey) and the quality of the sample frame, as represented by the presence of ineligible units. The number of rejections per contact was 0.25.

Rejection rate and Interviews per Contact
Sudan ES, 2014
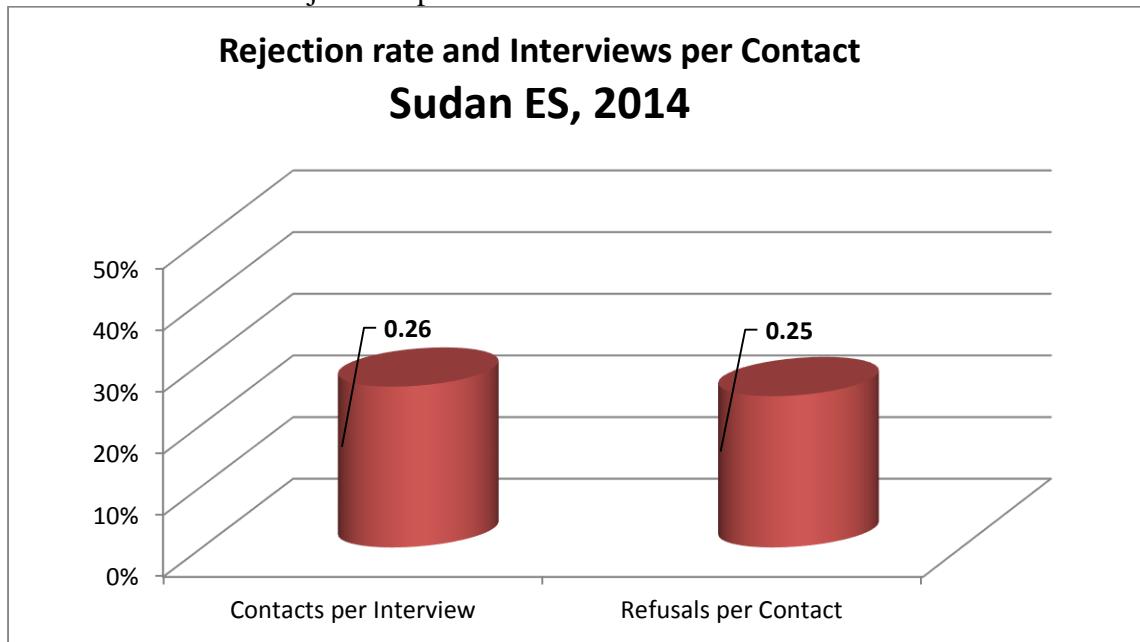
40.   Details on the rejection rate, eligibility rate, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to Sudan. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

References:
Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Samping: Design and Techniques, 1999.

---

[8] The estimate is based on the total no. of firms contacted including ineligible establishments.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., Elementary Survey Sampling, Fifth Edition, 1996.

Appendix A

Status Codes:

	Sudan
Sample Target	**720**
Complete interviews **(Total)**	662
Incomplete interviews	80
Elegible in process	56
Refusals	12
Out of target	258
Impossible to contact	730
Ineligible - coop.	116
Refusal to the Screener	625
Total	**2539**

Response rate	**51%**
Out of target + impossible to contact	**39%**
Impossible to contact	**29%**

Eligibles	1.Elegible establishment (Correct name and address)	779
	2. Elegible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	1
	3. Elegible establishment (Different name but same address - the firm/establishment changed its name)	1
	4. Elegible establishment (Wrong address - the firm/establishmen has changed address and the address could be found)	14
Ineligibles	5. The establishment has less than 5 permanent full time employees	4
	6. The firm discontinued businesses	101
	7. Not a business: private household	147
	8. Ineligible activity: education, agriculture, finances, governments…	6
Unobtainable	91. No reply *(after having called in different days of the week and in different business hours)*	319
	92. Line out of order	80
	93. No tone	309
	94. Phone number does not exist	0
	10. Answering machine	0
	11. Fax line - data line	1
	12. Wrong address/ moved away and could not get the new references	21
	13. Refuses to answer the screener	625
	14. In process *(the establishment is being called/ is being contacted - previous to ask the screener)*	**16**
	151. Out of target - outside the covered regions, firm moved abroad	112
	152. Out of target - firm moved abroad	4
	153. Out of target - Not registered with SAT	0
	Total	**2540**

Appendix C Weights
Strict Weights

	Row Labels	Manufacturing	Other Services	Retail	Wholesale
Bahri	1- Small (5 to 19)	2.4	2.8		
Bahri	2- Medium (20 to 99)	1.6			
Bahri	3- Large (100+)	1.2			
Bahri	4-Size Unknown	1.7	1.2	5.2	2.2
Khartoum	1- Small (5 to 19)	1.9	1.4	1.7	2.2
Khartoum	2- Medium (20 to 99)	2.7	1.6	2.0	1.9
Khartoum	3- Large (100+)	5.1	1.5		
Khartoum	4-Size Unknown	1.8	1.5	1.9	1.6
Om Durman	1- Small (5 to 19)	2.0	1.3	1.7	1.5
Om Durman	2- Medium (20 to 99)	2.3		2.3	2.9
Om Durman	3- Large (100+)	2.8	1.0		1.0
Om Durman	4-Size Unknown	2.0	1.0	1.3	1.5
Other/Unknown	1- Small (5 to 19)			1.3	
Other/Unknown	2- Medium (20 to 99)				
Other/Unknown	3- Large (100+)				
Other/Unknown	4-Size Unknown	1.4	1.3	1.0	1.6

Median Weights

	Row Labels	Manufacturing	Other Services	Retail	Wholesale
Bahri	1- Small (5 to 19)	4.4	6.6		
Bahri	2- Medium (20 to 99)	2.3			
Bahri	3- Large (100+)	1.6			
Bahri	4-Size Unknown	3.4	3.1	15.2	5.9
Khartoum	1- Small (5 to 19)	2.8	2.7	3.5	4.2
Khartoum	2- Medium (20 to 99)	3.1	2.3	3.3	2.8
Khartoum	3- Large (100+)	5.6	2.1		
Khartoum	4-Size Unknown	2.9	3.0	4.5	3.4
Om Durman	1- Small (5 to 19)	2.4	2.0	3.0	2.4
Om Durman	2- Medium (20 to 99)	2.1		3.0	3.5
Om Durman	3- Large (100+)	2.4	1.0		1.0
Om Durman	4-Size Unknown	2.6	1.7	2.4	2.6
Other/Unknown	1- Small (5 to 19)			4.3	
Other/Unknown	2- Medium (20 to 99)				
Other/Unknown	3- Large (100+)				
Other/Unknown	4-Size Unknown	3.6	4.2	3.0	5.3

Weak Weights

	Row Labels	Manufacturing	Other Services	Retail	Wholesale
Bahri	1- Small (5 to 19)	10.9	22.2		
Bahri	2- Medium (20 to 99)	6.4			
Bahri	3- Large (100+)	4.6			
Bahri	4-Size Unknown	11.3	13.9	62.1	22.8
Khartoum	1- Small (5 to 19)	5.9	7.6	9.0	10.1
Khartoum	2- Medium (20 to 99)	7.4	7.2	9.7	7.9
Khartoum	3- Large (100+)	13.3	6.9		
Khartoum	4-Size Unknown	8.1	11.3	15.3	11.1
Om Durman	1- Small (5 to 19)	4.9	5.6	7.5	5.7
Om Durman	2- Medium (20 to 99)	5.0		8.6	9.5
Om Durman	3- Large (100+)	5.7	2.8		2.8
Om Durman	4-Size Unknown	7.1	6.2	8.0	8.2
Other/Unknown	1- Small (5 to 19)			16.8	
Other/Unknown	2- Medium (20 to 99)				
Other/Unknown	3- Large (100+)				
Other/Unknown	4-Size Unknown	14.9	23.5	15.6	25.6

Appendix D

Strict Universe Estimates Sudan

	Size	Manufacturing	Other Services	Retail	Wholesale	Total
Bahri	1- Small (5 to 19)	9	3	20	47	79
Bahri	2- Medium (20 to 99)	8	18	3	9	38
Bahri	3- Large (100+)	6	6			12
Bahri	4-Size Unknown	18	2	5	4	30
Khartoum	1- Small (5 to 19)	113	46	31	88	279
Khartoum	2- Medium (20 to 99)	49	25	4	23	101
Khartoum	3- Large (100+)	15	6	3	3	27
Khartoum	4-Size Unknown	140	60	81	102	383
Om Durman	1- Small (5 to 19)	38	3	19	37	96
Om Durman	2- Medium (20 to 99)	18	3	2	6	29
Om Durman	3- Large (100+)	6	1		1	8
Om Durman	4-Size Unknown	40	8	32	24	103
Other/Unknown	1- Small (5 to 19)	26	18	1	53	98
Other/Unknown	2- Medium (20 to 99)	9	6		12	27
Other/Unknown	3- Large (100+)	9	3		3	15
Other/Unknown	4-Size Unknown	10	40	11	19	81
	Total	**515**	**248**	**213**	**430**	**1407**

Median Universe Estimates Sudan

	Size	Manufacturing	Other Services	Retail	Wholesale	Total
Bahri	1- Small (5 to 19)	18	7	20	47	91
Bahri	2- Medium (20 to 99)	11	18	3	9	41
Bahri	3- Large (100+)	8	6			14
Bahri	4-Size Unknown	38	6	15	12	71
Khartoum	1- Small (5 to 19)	167	86	66	171	490
Khartoum	2- Medium (20 to 99)	56	36	7	34	133
Khartoum	3- Large (100+)	17	9	3	3	31
Khartoum	4-Size Unknown	228	125	188	217	758
Om Durman	1- Small (5 to 19)	45	4	32	57	139
Om Durman	2- Medium (20 to 99)	17	3	3	7	30
Om Durman	3- Large (100+)	5	1		1	7
Om Durman	4-Size Unknown	52	13	59	41	167
Other/Unknown	1- Small (5 to 19)	26	18	4	53	101
Other/Unknown	2- Medium (20 to 99)	9	6		12	27
Other/Unknown	3- Large (100+)	9	3		3	15
Other/Unknown	4-Size Unknown	25	129	33	63	250
	Total	**731**	**470**	**433**	**731**	**2365**

Weak Universe Estimates

	Size	Manufacturing	Other Services	Retail	Wholesale	Total
Bahri	1- Small (5 to 19)	44	22	20	47	133
Bahri	2- Medium (20 to 99)	32	18	3	9	62
Bahri	3- Large (100+)	23	6			29
Bahri	4-Size Unknown	125	28	62	46	260
Khartoum	1- Small (5 to 19)	349	242	171	415	1177
Khartoum	2- Medium (20 to 99)	134	116	19	94	363
Khartoum	3- Large (100+)	40	27	3	3	73
Khartoum	4-Size Unknown	631	464	642	697	2434
Om Durman	1- Small (5 to 19)	93	11	83	137	325
Om Durman	2- Medium (20 to 99)	40	3	9	19	71
Om Durman	3- Large (100+)	11	3		3	17
Om Durman	4-Size Unknown	142	49	201	131	523
Other/Unknown	1- Small (5 to 19)	26	18	17	53	114
Other/Unknown	2- Medium (20 to 99)	9	6		12	27
Other/Unknown	3- Large (100+)	9	3		3	15
Other/Unknown	4-Size Unknown	105	728	171	308	1311
	Total	**1813**	**1745**	**1401**	**1976**	**6935**

Appendix E

Original Sample Design, Sudan:

Original Design	Size	Manufacturing	Other services	Retail	Wholesale	Total
Bahri	1- Small (5 to 19)	15	1	7	4	27
Bahri	2- Medium (20 to 99)	11	5	1	3	20
Bahri	3- Large (100+)	8	2	0	0	10
Bahri	4-Size Unknown	35	1	7	1	43
Khartoum	1- Small (5 to 19)	52	8	13	30	102
Khartoum	2- Medium (20 to 99)	40	4	7	6	56
Khartoum	3- Large (100+)	14	10	1	1	26
Khartoum	4-Size Unknown	75	24	55	52	206
Om Durman	1- Small (5 to 19)	30	1	7	4	42
Om Durman	2- Medium (20 to 99)	14	1	3	1	19
Om Durman	3- Large (100+)	4	1	0	1	6
Om Durman	4-Size Unknown	28	1	13	1	43
Other/Unknow	1- Small (5 to 19)	9	1	1	1	12
Other/Unknow	2- Medium (20 to 99)	3	1	0	1	5
Other/Unknow	3- Large (100+)	3	1	0	1	5
Other/Unknow	4-Size Unknown	20	59	6	14	98
	Total	360	120	120	120	720

Appendix F

Local Agency team involved in the study:

Local Agency	ConsulSat for Information Services
Name of Project Manager	Mohamed Hamza
Name and position of other key persons of the project: Local Survey Implementation Team and corresponding supervisor and enumerator codes:	Co Project Manager: Lubna Elsafi 10 enumerators (inclduing 2 recruiters), 4 supervisors

Sample Frame:

Characteristics of sample frame used	Variables: name of establishment, address, sector, region, size, telephone number (for around half of all records)
Year:	2013, 2014
Comments on the quality of sample frame:	Initial sample frame was of very poor quality and lacked useful stratification information. Before fieldwork began, the local contractor performed a verification exercise to obtain as much stratification information as possible. The resulting sample design was then based off of this verified information.
Year and organism who conducted the last economic census	UNIDO
Other sources for companies statistics	Chambers of Commerce, UNIDO, Importers/Exporters registry, local contractor verification of raw frame

Sample:

Comments/ problems on sectors and regions selected in the sample	Selected regions were geographically clustered around the greater Khartoum metro area.
Comments on the response rate	Response rate were relatively low (49%), but this was primarily due to sample quality. Response rates among eligible establishments were generally much higher.
Comments on the sample design:	The sample design was followed for the opening stages of fieldwork. However, due to low completion counts, cell targets were abandoned and a census was conducted on the frame.
Other comments:	None

Fieldwork and country situation:

Date of Fieldwork	September 2014 - February 2015
Locations	Bahri, Omdurman, Khartoum
Interview number	662
Problems found during fieldwork	Due to issues with obtaining a suitable sampling frame, fieldwork started very late. However, once in field it ran relatively smoothly. There as a slight misunderstanding about the innovation module experiment, but was rectified fairly soon after first raised.
Other observations:	PAPI used for data collection. Good performance on behalf of the local contractor to achieve as many interviews as they did from the limited and poor quality sample.