

The Zambia 2013 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Zambia between December 2012 and February 2014 as part of the Africa Enterprise Survey 2013, 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 the 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 Zambia was selected using stratified random sampling, following the methodology explained in the *Sampling Manual*¹. Stratified random sampling² was preferred over simple random sampling for several reasons³:

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.

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

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.

3. Three levels of stratification were used in this country: industry, establishment size, and region. 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 four manufacturing industries (food, textiles and garments, chemicals and plastics, other manufacturing) and two service sectors (retail and other services).

5. For the Zambia Enterprise Survey (ES), size stratification was defined following the standardized definition for the rollout: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees).

6. Regional stratification for the Zambia ES as well as the Zambia micro surveys was defined in five regions: Kitwe, Livingstone, Lusaka, and Ndola.

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. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. TNS Opinion was hired to implement the Africa 2013 enterprise surveys roll out. In Zambia the local subcontractor was LUSAKA PROBE.

9. For the Zambia ES, two sample frames were used. The first was supplied by the World Bank and consists of enterprises interviewed in Zambia 2007. The World Bank required that attempts should be made to re-interview establishments responding to the Zambia 2007 survey where they were within the selected geographical regions and met eligibility criteria. Due to the fact that the previous round of surveys seemed to have utilized different stratification criteria (or no stratification at all) and due to the prevalence of small firms and firms located in the capital city in the 2007 sample the following convention was used. The presence of panel firms was limited to a maximum of 50% of the achieved interviews in each cell. That sample is referred to as the Panel.

The second frame was provided by Zambia Central Statistical office. For confidentiality purposes, the Zambia Central Statistical Office randomly drew the sample of fresh establishment to be interviewed based on the sample design provided by the World Bank. The database contained the following information

- Detailed stratification variables;
- Location identifiers- address, phone number, email;
- Contact name(s).

Counts from the sample frames are shown below.

Zambia MAIN ES, Sample Frame

Source: 2010, Zambia Central Statistics Office Census of Business Establishments

Region name	Employee size	Food 15	Textile and Apparel 17-18	Metals and Machinery 27-28-29	Other Manufactu ring	Retail 52	Other Services	
Kitwe	Small 5-19	10	5	30	25	162	214	455
	Med 20-99	6	0	9	9	23	63	113
	Large 100+	5	1	4	3	2	12	27
		21	6	43	37	187	289	595
Livingstone	Small 5-19	9	1	2	2	64	84	161
	Med 20-99	3	0	0	2	14	25	44
	Large 100+	0	1	0	0	0	3	4
		12	2	2	4	78	112	209
Lusaka	Small 5-19	47	22	54	149	674	753	1700
	Med 20-99	34	10	25	90	124	228	509
	Large 100+	15	0	3	11	19	44	93
		96	32	82	250	817	1025	2302
Ndola	Small 5-19	11	4	8	32	129	204	379
	Med 20-99	7	5	13	29	15	49	111
	Large 100+	5	0	2	6	2	8	22
		23	9	23	67	146	261	512
Grand Total		152	49	150	358	1228	1687	3624

10. The enumerated establishments with 5 employees or more (fresh and panel) were then used as the sample frame for the Zambia Enterprise Survey with the aim of obtaining interviews at 720 establishments.

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. In addition, the sample frame contains no telephone/fax numbers so the local contractor had to screen the contacts by visiting them.

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. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 7.49% (115 out of 1536 establishments)⁴ for the main ES firms.

IV. Data Base Structure:

13. The structure of the data base reflects the fact that 3 different versions of the survey instrument were used for all registered establishments. Questionnaires have common questions (*core* module) and respectfully additional manufacturing and retail specific questions. The eligible manufacturing industries have been surveyed using the ***Manufacturing*** questionnaire (includes the *core* module, plus manufacturing specific questions). Retail firms have been interviewed using the ***Retail*** questionnaire (includes the *core* module plus retail specific questions) and the residual eligible services have been covered using the core module only (***Other 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 preceded by a prefix “KEN” and “A2f” indicate questions specific to some countries in Africa, 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.

⁴ Based on out of target contacts and impossible to contact establishments

-*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. 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.

21. Please note that the monetary values collected during the interviews were in “old” Kwacha. The new Kwacha was introduced in January 2013: 1,000,000 Kwacha (One Million Kwacha Old Currency) is equivalent to 1,000 Kwacha (One Thousand Kwacha New Currency)

V. Universe Estimates

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

23. Appendix B shows the overall estimates of the numbers of establishments in Zambia based on the sample frame.

24. 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.

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

26. 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*.

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

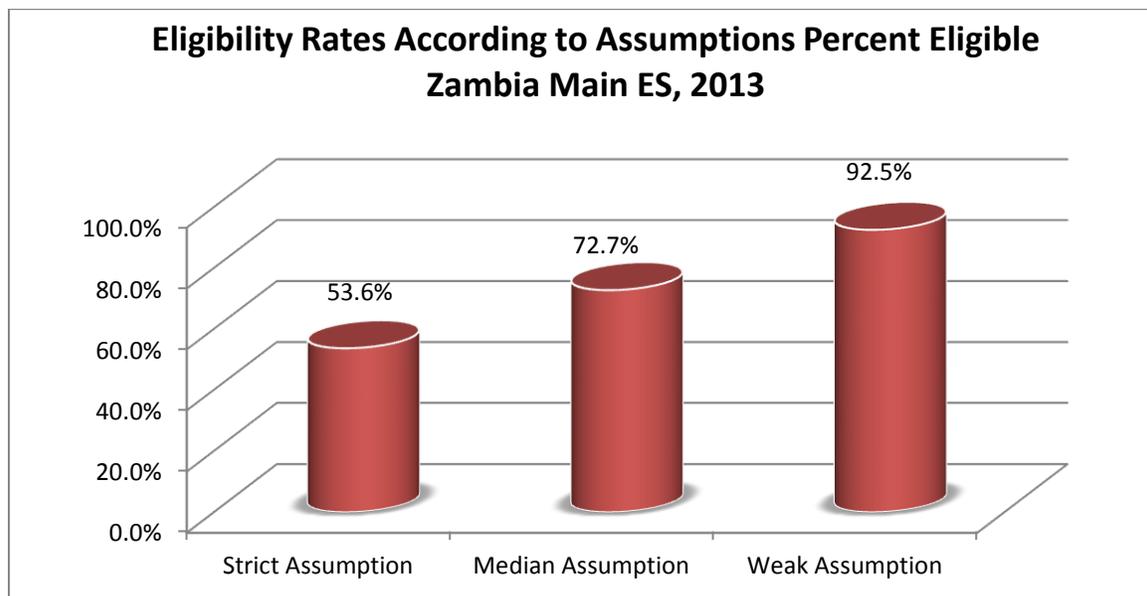
27. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *wmedian*.

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

28. 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*.

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

29. 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. The eligibility rates for micro firms are also provided below.



30. Universe estimates for the number of establishments in each industry-region-size cell in Zambia 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.

31. 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

32. 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.)⁵

33. 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 line⁶, 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.

34. Appendix C shows the cell weights for registered establishments in Zambia.

VII. Appropriate use of the weights

35. 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.

36. 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

⁵ This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

⁶ For the surveys that implemented a screener over the phone.

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-unbiased estimates (see also Cochran, 1977, pp 200 who favors the used of weighted OLS for a common population coefficient.)⁷

37. 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.⁸ 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

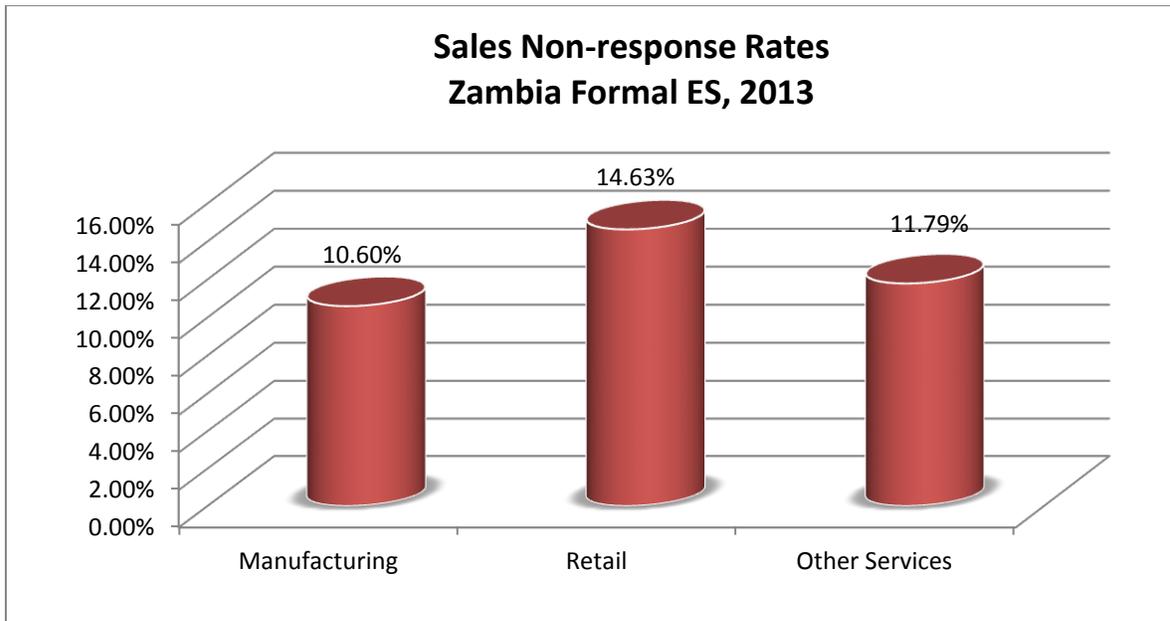
38. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to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

39. 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 charts below for both enterprise surveys (ES) and micro firms reflect both categories (DKs and NAs).

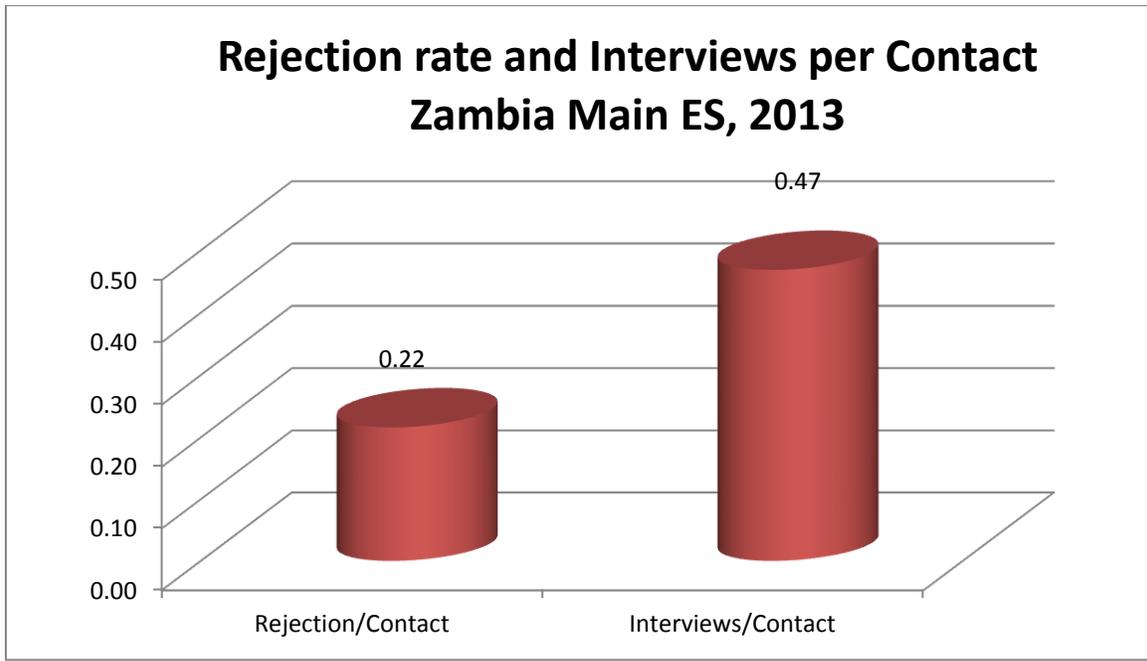
⁷ Note that weighted OLS in Stata using the command `regress` with the option of weights will estimate wrong standard errors. Using the Stata survey specific commands `svy` will provide appropriate standard errors.

⁸ 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.



40. 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.

41. As the following graph shows, the number of interviews per contacted establishments was 0.47. 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 0.22 for ES firms.



42. 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 Zambia. 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. Sampling: Design and Techniques, 1999.

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

Appendix A

Status Codes Enterprise Survey (ES) Total:	ELIGIBLES	
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	809
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	7
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	6
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	2
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	31
	6. The firm discontinued businesses	71
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	10
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	128
	92. Line out of order	34
	93. No tone	0
	94. Phone number does not exist	21
	10. Answering machine	8
	11. Fax line - data line	14
	12. Wrong address/ moved away and could not get the new references	121
	13. Refuses to answer the screener	271
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	1
	152. Out of target - firm moved abroad	1
	153. Impossible to find	1
	Total	1536

Response Outcomes Enterprise (ES) Survey Total:

Complete interviews (<i>Total</i>)	720
Incomplete interviews	29
Eligible in process	14
Refusals	61
Out of target	112
Impossible to contact	326

Ineligible - coop.	3
Refusal to the Screener	271
Total	1536

Status Codes Enterprise Survey (ES) Fresh:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	642
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	5
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	2
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	2
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	31
	6. The firm discontinued businesses	54
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	4
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	72
	92. Line out of order	15
	93. No tone	0
	94. Phone number does not exist	18
	10. Answering machine	7
	11. Fax line - data line	13
	12. Wrong address/ moved away and could not get the new references	82
	13. Refuses to answer the screener	204
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	1
	152. Out of target - firm moved abroad	1
	153. Impossible to find	1
	Total	1154

Response Outcomes Enterprise Survey (ES) Fresh:

Complete interviews (<i>Total</i>)	568
Incomplete interviews	22
Eligible in process	8
Refusals	53
Out of target	89

Impossible to contact	207
Ineligible - coop.	3
Refusal to the Screener	204
Total	1154

Status Codes Enterprise Survey (ES) Panel:

ELIGIBLES		
Eligible	1. Eligible establishment (<i>Correct name and address</i>)	167
	2. Eligible establishment (<i>Different name but same address - the new firm/establishment bought the original firm/establishment</i>)	2
	3. Eligible establishment (<i>Different name but same address - the firm/establishment changed its name</i>)	4
	4. Eligible establishment (<i>Wrong address - the firm/establishment has changed address and the address could be found</i>)	0
	16. Panel firm - now less than five employees	0
Ineligible	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	17
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	6
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	56
	92. Line out of order	19
	93. No tone	0
	94. Phone number does not exist	3
	10. Answering machine	1
	11. Fax line - data line	1
	12. Wrong address/ moved away and could not get the new references	39
	13. Refuses to answer the screener	67
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	153. Impossible to find	0
	Total	382

Response Outcomes Enterprise Survey (ES) Panel:

Complete interviews (<i>Total</i>)	152
Incomplete interviews	7
Eligible in process	6
Refusals	8
Out of target	23
Impossible to contact	119
Ineligible - coop.	0
Refusal to the Screener	67
Total	382

Appendix B

Universe Estimates as in the Zambia Central Statistics Office Census of Business Establishments:

Region name	Employee size	Food 15	Textile and Apparel 17-18	Metals and Machinery 27-28-29	Other Manufactu ring	Retail 52	Other Services	
Kitwe	Small 5-19	10	5	30	25	162	214	455
	Med 20-99	6	0	9	9	23	63	113
	Large 100+	5	1	4	3	2	12	27
		21	6	43	37	187	289	595
Livingstone	Small 5-19	9	1	2	2	64	84	161
	Med 20-99	3	0	0	2	14	25	44
	Large 100+	0	1	0	0	0	3	4
		12	2	2	4	78	112	209
Lusaka	Small 5-19	47	22	54	149	674	753	1700
	Med 20-99	34	10	25	90	124	228	509
	Large 100+	15	0	3	11	19	44	93
		96	32	82	250	817	1025	2302
Ndola	Small 5-19	11	4	8	32	129	204	379
	Med 20-99	7	5	13	29	15	49	111
	Large 100+	5	0	2	6	2	8	22
		23	9	23	67	146	261	512
Grand Total		152	49	150	358	1228	1687	3624

Appendix C

Strict Cell Weights Zambia:

Fresh Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services
Kitwe						
Small (5 to 19)	1.16	1.58	1.00	1.48	83.39	3.90
Medium (20 to 99)	1.89	0.00	1.00	1.00	1.00	18.36
Large (100+)	0.00	1.00	1.00	1.00	1.00	1.40
Livingstone						
Small (5 to 19)	1.00	1.00	0.00	1.98	1.21	9.67
Medium (20 to 99)	1.00	0.00	0.00	1.00	1.00	1.00
Large (100+)	0.00	0.00	0.00	1.00	0.00	1.03
Lusaka						
Small (5 to 19)	10.16	1.00	10.45	1.01	10.56	9.22
Medium (20 to 99)	3.89	1.00	1.00	1.00	40.00	96.25
Large (100+)	1.00	0.00	1.00	1.10	1.00	10.08
Ndola						
Small (5 to 19)	1.99	2.71	1.17	1.69	57.61	2.82
Medium (20 to 99)	1.00	2.76	1.00	1.16	1.46	1.34
Large (100+)	0.00	0.00	0.00	0.00	0.00	1.00

Panel Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services
Kitwe						
Small (5 to 19)	0.00	0.00	1.03	1.90	1.11	1.00
Medium (20 to 99)	1.42	0.00	1.00	1.00	1.00	0.00
Large (100+)	0.00	0.00	0.00	1.00	0.00	0.00
Livingstone						
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00
Lusaka						
Small (5 to 19)	1.63	1.00	1.00	1.19	1.15	1.00
Medium (20 to 99)	1.47	1.71	1.26	1.09	1.11	2.47
Large (100+)	1.00	0.00	1.00	1.54	0.00	1.09
Ndola						
Small (5 to 19)	1.00	0.00	0.00	1.00	1.38	1.00
Medium (20 to 99)	1.00	1.53	0.00	2.45	0.00	1.00
Large (100+)	1.00	0.00	0.00	1.38	0.00	0.00

Median Cell Weights Zambia:

Fresh Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services
Kitwe						
Small (5 to 19)	1.33	1.46	1.18	1.44	105.84	4.20
Medium (20 to 99)	2.55	0.00	1.00	1.02	1.30	23.32
Large (100+)	0.00	1.00	1.00	1.00	1.00	1.77
Livingstone						
Small (5 to 19)	1.00	1.00	0.00	2.01	1.61	10.93
Medium (20 to 99)	1.00	0.00	0.00	1.00	1.54	1.00
Large (100+)	0.00	0.00	0.00	1.05	0.00	1.37
Lusaka						
Small (5 to 19)	14.05	1.07	17.29	1.19	16.17	11.99
Medium (20 to 99)	6.33	1.00	1.61	1.11	72.23	147.56
Large (100+)	1.29	0.00	1.36	1.51	1.00	15.36
Ndola						
Small (5 to 19)	2.67	2.92	1.87	1.92	85.39	3.56
Medium (20 to 99)	1.42	3.50	1.85	1.55	2.56	1.99
Large (100+)	0.00	0.00	0.00	0.00	0.00	1.13

Panel Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services
Kitwe						
Small (5 to 19)	0.00	0.00	1.31	1.96	1.19	1.00
Medium (20 to 99)	2.43	0.00	1.33	1.00	1.04	0.00
Large (100+)	0.00	0.00	0.00	1.12	0.00	0.00
Livingstone						
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00
Lusaka						
Small (5 to 19)	2.36	1.10	1.00	1.36	1.37	1.00
Medium (20 to 99)	2.81	2.79	2.32	1.64	1.73	3.85
Large (100+)	1.29	0.00	1.00	2.24	0.00	1.65
Ndola						
Small (5 to 19)	1.09	0.00	0.00	1.00	1.80	1.00
Medium (20 to 99)	1.00	2.76	0.00	4.05	0.00	1.63
Large (100+)	1.34	0.00	0.00	2.22	0.00	0.00

Weak Cell Weights Zambia:

Fresh Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services
Kitwe						
Small (5 to 19)	1.30	1.68	1.51	1.52	133.50	5.13
Medium (20 to 99)	2.61	0.00	1.26	1.13	1.71	29.73
Large (100+)	0.00	1.00	1.00	1.00	1.00	2.12
Livingstone						
Small (5 to 19)	1.00	1.00	0.00	1.83	1.75	11.48
Medium (20 to 99)	1.00	0.00	0.00	1.00	1.75	1.10
Large (100+)	0.00	0.00	0.00	1.00	0.00	1.41
Lusaka						
Small (5 to 19)	14.69	1.31	23.63	1.34	21.76	15.62
Medium (20 to 99)	6.92	1.25	2.29	1.31	101.58	200.76
Large (100+)	1.32	0.00	1.82	1.67	1.24	19.65
Ndola						
Small (5 to 19)	2.66	3.41	2.44	2.06	109.52	4.41
Medium (20 to 99)	1.48	4.27	2.51	1.74	3.42	2.58
Large (100+)	0.00	0.00	0.00	0.00	0.00	1.38

Fresh Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services
Kitwe						
Small (5 to 19)	0.00	0.00	3.15	3.92	2.36	1.39
Medium (20 to 99)	5.13	0.00	3.31	1.74	2.15	0.00
Large (100+)	0.00	0.00	0.00	2.18	0.00	0.00
Livingstone						
Small (5 to 19)	0.00	0.00	0.00	0.00	0.00	0.00
Medium (20 to 99)	0.00	0.00	0.00	0.00	0.00	0.00
Large (100+)	0.00	0.00	0.00	0.00	0.00	0.00
Lusaka						
Small (5 to 19)	3.25	1.65	1.00	1.83	1.84	1.14
Medium (20 to 99)	3.99	4.34	3.90	2.29	2.41	5.19
Large (100+)	1.71	0.00	1.05	2.94	0.00	2.09
Ndola						
Small (5 to 19)	1.54	0.00	0.00	1.19	2.48	1.27
Medium (20 to 99)	1.17	4.40	0.00	5.81	0.00	2.25
Large (100+)	1.83	0.00	0.00	2.98	0.00	0.00

Appendix D

Strict Universe Estimates

Fresh Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	9	4	24	26	96	196	355
Small (5 to 19)	7	3	16	19	83	152	281
Medium (20 to 99)	2	0	5	5	11	37	60
Large (100+)	0	1	3	2	2	7	15
Livingstone	11	1	0	5	37	80	134
Small (5 to 19)	8	1	0	2	30	58	99
Medium (20 to 99)	3	0	0	2	7	20	32
Large (100+)	0	0	0	1	0	2	3
Lusaka	24	15	29	146	318	522	1054
Small (5 to 19)	10	7	21	86	264	406	794
Medium (20 to 99)	4	8	7	55	40	96	210
Large (100+)	10	0	1	4	14	20	50
Ndola	11	5	8	35	63	151	273
Small (5 to 19)	8	3	4	22	58	124	218
Medium (20 to 99)	3	3	4	13	6	21	50
Large (100+)	0	0	0	0	0	5	5
Grand Total	55	26	60	212	515	949	1816

Panel Enterprise Survey

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	1	0	7	5	7	1	21
Small (5 to 19)	0	0	1	2	6	1	9
Medium (20 to 99)	1	0	6	2	1	0	10
Large (100+)	0	0	0	1	0	0	1
Livingstone	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Lusaka	25	18	5	33	37	17	134
Small (5 to 19)	10	16	1	15	32	11	85
Medium (20 to 99)	7	2	3	11	4	5	32
Large (100+)	8	0	1	6	0	1	16
Ndola	7	2	0	9	4	3	25
Small (5 to 19)	3	0	0	3	4	1	11
Medium (20 to 99)	2	2	0	5	0	2	10
Large (100+)	2	0	0	1	0	0	3
Grand Total	34	19	12	47	47	21	179

Median Universe Estimates

Fresh Main ES

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	11	4	27	26	122	219	409
Small (5 to 19)	8	3	19	19	106	164	318
Medium (20 to 99)	3	0	5	5	14	47	74
Large (100+)	0	1	3	2	2	9	17
Livingstone	11	1	0	5	51	88	156
Small (5 to 19)	8	1	0	2	40	66	117
Medium (20 to 99)	3	0	0	2	11	20	36
Large (100+)	0	0	0	1	0	3	4
Lusaka	33	15	47	168	490	706	1461
Small (5 to 19)	14	7	35	101	404	528	1089
Medium (20 to 99)	6	8	11	61	72	148	307
Large (100+)	13	0	1	6	14	31	65
Ndola	15	6	13	42	96	194	366
Small (5 to 19)	11	3	6	25	85	156	286
Medium (20 to 99)	4	3	7	17	10	32	74
Large (100+)	0	0	0	0	0	6	6
Grand Total	70	27	87	241	759	1208	2392

Panel Main ES

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	2	0	9	5	7	1	25
Small (5 to 19)	0	0	1	2	6	1	10
Medium (20 to 99)	2	0	8	2	1	0	13
Large (100+)	0	0	0	1	0	0	1
Livingstone	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Lusaka	39	20	7	43	45	20	174
Small (5 to 19)	14	18	1	18	38	11	100
Medium (20 to 99)	14	3	5	16	7	8	52
Large (100+)	10	0	1	9	0	2	22
Ndola	8	3	0	13	5	4	34
Small (5 to 19)	3	0	0	3	5	1	13
Medium (20 to 99)	2	3	0	8	0	3	16
Large (100+)	3	0	0	2	0	0	5
Grand Total	49	23	16	61	58	26	232

Weak Universe Estimates

Fresh Main ES

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	10	4	34	27	154	270	500
Small (5 to 19)	8	3	24	20	133	200	389
Medium (20 to 99)	3	0	6	6	19	59	93
Large (100+)	0	1	3	2	2	11	19
Livingstone	11	1	0	5	56	94	166
Small (5 to 19)	8	1	0	2	44	69	123
Medium (20 to 99)	3	0	0	2	12	22	39
Large (100+)	0	0	0	1	0	3	4
Lusaka	35	19	65	193	663	927	1902
Small (5 to 19)	15	9	47	114	544	687	1416
Medium (20 to 99)	7	10	16	72	102	201	407
Large (100+)	13	0	2	7	17	39	78
Ndola	15	8	17	46	123	242	452
Small (5 to 19)	11	3	7	27	110	194	352
Medium (20 to 99)	4	4	10	19	14	41	93
Large (100+)	0	0	0	0	0	7	7
Grand Total	71	32	116	271	996	1533	3020

Panel Main ES

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	5	0	23	10	14	1	53
Small (5 to 19)	0	0	3	4	12	1	20
Medium (20 to 99)	5	0	20	3	2	0	31
Large (100+)	0	0	0	2	0	0	2
Livingstone	0	0	0	0	0	0	0
Small (5 to 19)	0	0	0	0	0	0	0
Medium (20 to 99)	0	0	0	0	0	0	0
Large (100+)	0	0	0	0	0	0	0
Lusaka	53	31	10	58	61	25	238
Small (5 to 19)	19	26	1	24	51	13	135
Medium (20 to 99)	20	4	8	23	10	10	75
Large (100+)	14	0	1	12	0	2	29
Ndola	11	4	0	18	7	6	46
Small (5 to 19)	5	0	0	4	7	1	17
Medium (20 to 99)	2	4	0	12	0	5	23
Large (100+)	4	0	0	3	0	0	7
Grand Total	69	35	33	86	82	32	338

Appendix E

Original Sample Design, Zambia:

Region name	Employee size	Food 15	Textile and Apparel 17-18	Metals and Machinery 27-28-29	Other Manufact uring	Retail 52	Other Services	
Kitwe	Small 5-19	10	5	30	25	5	3	78
	Med 20-99	6	0	9	9	8	3	35
	Large 100+	5	1	4	3	2	12	27
		21	6	43	37	15	18	140
Livingstone	Small 5-19	9	1	2	2	15	4	33
	Med 20-99	3	0	0	2	14	24	43
	Large 100+	0	1	0	0	0	3	4
		12	2	2	4	29	31	80
Lusaka	Small 5-19	3	13	3	71	51	51	192
	Med 20-99	9	10	7	90	3	3	122
	Large 100+	15	0	3	11	14	3	46
		27	23	13	172	68	57	360
Ndola	Small 5-19	8	4	7	32	3	3	57
	Med 20-99	7	5	13	29	3	3	60
	Large 100+	5	0	2	6	2	8	23
		20	9	22	67	8	14	140
Grand Total		80	40	80	280	120	120	720

Completed Interviews, Zambia:

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	8	3	31	24	20	47	133
1 - Small	6	2	17	14	6	40	85
2 - Medium	2		11	7	12	2	34
3 - Large		1	3	3	2	5	14
Livingstone	11	1		4	32	28	76
1 - Small	8	1		1	25	6	41
2 - Medium	3			2	7	20	32
3 - Large				1		2	3
Lusaka	31	32	14	171	72	61	381
1 - Small	7	23	3	98	53	55	239
2 - Medium	6	9	9	65	5	3	97
3 - Large	18		2	8	14	3	45
Ndola	14	3	7	30	8	68	130
1 - Small	7	1	3	16	4	45	76
2 - Medium	5	2	4	13	4	18	46
3 - Large	2			1		5	8
Grand Total	64	39	52	229	132	204	720

Fresh

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	7	3	24	20	14	46	114
Small (5 to 19)	6	2	16	13	1	39	77
Medium (20 to 99)	1		5	5	11	2	24
Large (100+)		1	3	2	2	5	13
Livingstone	11	1		4	32	28	76
Small (5 to 19)	8	1		1	25	6	41
Medium (20 to 99)	3			2	7	20	32
Large (100+)				1		2	3
Lusaka	12	15	10	144	40	47	268
Small (5 to 19)	1	7	2	85	25	44	164
Medium (20 to 99)	1	8	7	55	1	1	73
Large (100+)	10		1	4	14	2	31
Ndola	7	2	7	24	5	65	110
Small (5 to 19)	4	1	3	13	1	44	66
Medium (20 to 99)	3	1	4	11	4	16	39
Large (100+)						5	5
Grand Total	37	21	41	192	91	186	568

Panel

Row Labels	Food	Textile and Apparel	Metals	Other Manufacturing	Retail	Other Services	Grand Total
Kitwe	1		7	4	6	1	19
Small (5 to 19)			1	1	5	1	8
Medium (20 to 99)	1		6	2	1		10
Large (100+)				1			1
Lusaka	19	17	4	27	32	14	113
Small (5 to 19)	6	16	1	13	28	11	75
Medium (20 to 99)	5	1	2	10	4	2	24
Large (100+)	8		1	4		1	14
Ndola	7	1		6	3	3	20
Small (5 to 19)	3			3	3	1	10
Medium (20 to 99)	2	1		2		2	7
Large (100+)	2			1			3
Grand Total	27	18	11	37	41	18	152

Appendix F

Local Agency team involved in the study:

Local Agency	Name: LUSAKA PROBE Country: Zambia Activities since: 2003
Enumerators involved:	Enumerators: 35 Recruiters: 3
Other staff involved:	Fieldwork Coordinators: 4 Data Entry: 7 Data Processing: 1

Sample Frame:

Characteristic of sample frame used:	Panel: List from the 2007 Enterprise Survey Fresh: Zambia Census Of Business Establishments
Source:	Zambia Central Statistics Office
Year:	2010

Sectors included in the Sample:

Original Sectors	<p>The manufacturing sector comprises all manufacturing establishments as mentioned in group D</p> <p>The service sector includes Group F (construction), Groups G, Group H (hotels and restaurants), Group I (transport, storage, and communications) and subsector 72 from Group K</p>
Added (top up) Sectors	None

Fieldwork and country situation:

Date of Fieldwork	December 2012 to February 2014
Country	Zambia
Problems found during fieldwork:	<ul style="list-style-type: none">▪ In some cases the sample frame did not provide establishment's contact details. In these cases, firms were screened by going to the establishment physical location.▪ Respondents were expecting incentives.▪ Respondents were in general inclined to think the study aimed to get tax and social security defaulters. This might have affected the reliability of answers related to revenue and costs.
Country specific situation	<ul style="list-style-type: none">▪ The monetary values collected during the interviews were in "old" Kwacha. The new Kwacha was introduced in January 2013: 1,000,000 Kwacha (One Million Kwacha Old Currency) is equivalent to 1,000 Kwacha (One Thousand Kwacha New Currency)