

The 2015 Ethiopia Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Ethiopia between June 2015 and February 2016 for the 2015 Ethiopia Enterprise Survey (ES). The objective of the Enterprise Survey is to gain an understanding of what firms experience in the private sector.

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 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 standardized methodology, explained in the *Sampling Manual*¹. 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 the 2015 Ethiopia ES was selected using stratified random sampling, following the standard methodology. Stratified random sampling² was preferred over simple random sampling for several reasons³:

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

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

sub-sector 72, IT, which was added to the population under study), and all public or utilities-sectors.

- 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.
- 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.)
- 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.
- 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 D.

4. Industry stratification was designed in the way that follows: the universe was stratified into four manufacturing industries (Food and Beverages (ISIC Rev. 3.1 code 15), Textile and Garments including leather (ISIC codes 17-19), Non-metallic mineral products (ISIC code 26), and other manufacturing (ISIC Codes 16, 20-25, 27-37)) and three services sectors (Transportation (ISIC codes 60-62, 64), Retail (ISIC code 52) and Other Services (ISIC codes 45, 50, 51,55 and 72)).

5. Size stratification was defined as follows: small (5 to 19 employees), medium (20 to 99 employees), and large (more than 99 employees).

6. Regional stratification for the 2015 Ethiopia ES was done across six geographic regions: Addis Ababa and Dire Dawa city administrations, and Amhara, Oromia, SNNPR and Tigray regional states.

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.

8. WAAS International Plc was hired to implement the data collection for the 2015 Ethiopia Enterprise Survey.

9. The sample frame consisted of listings of firms from two sources. First, for panel firms, the list of 644 firms covered in the 2011 Ethiopia Enterprise Survey (i.e. “panel” firms) is used. Secondly, for fresh firms (i.e., firms that were not covered in the 2011 survey), business registry data collected from the Trade and Industry Bureaus of the six administrative regions and cities, and additional list of business registry data from the Federal Ministry of Trade and Industry were used. This list contains firms that obtained

investment license from relevant government bureaus and not necessary firms that started operations. The business registry database contained the establishment's name, address, sector of activity (manufacturing, services, and trade) and phone number if available. Further, the business registry data includes information on registered capital, but miss, for a large number of firms, information on employment size. Since the ES uses employment as the size stratification variable, information on firm's registered capital is used to impute the implied size of employment, following a concordance used by Ethiopia's Ministry of Industry; Appendix II provides this concordance. Nevertheless, there are sizable number of firms with missing information on employee number and capital. Additional size strata, "unknown", is added to the sampling stratification for size to accommodate these firms.

Overall estimates of the numbers of establishments in Ethiopia based on the sample frame is given as follows:

Table 1: Sample frame for the 2015 Ethiopia ES

Region	2:Textile and		3:Non-Metalic	4:Other		5:Retail	6:Transportation	7:Other Services	Grand Total
	1:Food	Garments	Minerals	Manufacturing					
1:AddisAba	322	250	160	1233	4909	18441	28703	54018	
1:Small	146	91	57	503	3039	12070	5750	21656	
2:Medium	94	69	25	408	1542	4706	2258	9102	
3:Large	23	36	10	91	222	188	325	895	
4:Unknown	59	54	68	231	106	1477	20370	22365	
2:Amhara	80	12	10	62	553	549	1398	2664	
1:Small	59	4	6	38	459	450	501	1517	
2:Medium	15	1		3	86	88	125	318	
3:Large	4	1		1	7	4	15	32	
4:Unknown	2	6	4	20	1	7	757	797	
3:Dredawa	22	2	2	16	5		26	73	
1:Small				2			1	3	
2:Medium	1			2	4		22	29	
3:Large					1		2	3	
4:Unknown	21	2	2	12			1	38	
4:Oromia	581	171	40	221	2735	4700	3036	11484	
1:Small	122	28	7	36	1060	2825	901	4979	
2:Medium	156	46	6	48	691	1136	589	2672	
3:Large	60	23		12	105	485	159	844	
4:Unknown	243	74	27	125	879	254	1387	2989	
5:SNNPR	551	5	15	37	69	4	220	901	
1:Small	8		8	7	6		20	49	
2:Medium	524	3	7	25	58	4	174	795	
3:Large	19	2		5	5		26	57	
4:Unknown									
6:Tigray	146	94	107	311	897	113	1014	2682	
1:Small	69	29	70	143	619	96	597	1623	
2:Medium	60	19	34	115	122	9	287	646	
3:Large	10	6	1	38	3	5	85	148	
4:Unknown	7	40	2	15	153	3	45	265	
Grand Total	1702	534	334	1880	9168	23807	34397	71822	

Source: World Bank, Ethiopia's Federal Ministry of Trade, and regional state and city administration Trade and Industry Bureaus.

10. The enumerated establishments with 5 employees or more (fresh and panel) were then used as the sample frame for the Ethiopia Enterprise Survey with the aim of obtaining interviews of 900 establishments.

11. The quality of the frame was assessed at the onset of the project through visits to a random subset of firms and the contractor's local knowledge. The sample frame was not immune from the typical problems found in establishment surveys, particularly given the list is based on business registry data and not establishment census data. Consequently, there were positive rates of non-eligibility, repetition, non-existent units, units that didn't start operations, etc. In addition, for some of the firms, the sample frame does not contain telephone/fax numbers, so the local contractor had to screen some 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 33% (1056 out of 3447 establishments), reflecting the fact that the fresh sample frame is based on a business registry data⁴. In fact, most of the non-eligibility issue is the result of firms turning out to be micro (which is not part of our sampling universe) up on screening.

Breaking down by industry and size, the following sample targets were achieved (based on the sampling information):

Table 2: Achieved Interviews (Fresh and Panel Combined)

Region	2:Textile and		3:Non-Metalic	4:Other	5:Retail	6:Transportation	7:Other Services	Grand Total
	1:Food	Garments	Mineral	Manufacturing				
1:AddisAbaba	24	47	37	87	59	61	136	451
1:Small	8	11	12	29	29	10	47	146
2:Medium	6	11	9	33	22	23	51	155
3:Large	8	21	5	22	6	16	23	101
4:Unknown	2	4	11	3	2	12	15	49
2:Amhara	8	4	3	12	12	9	22	70
1:Small	2	1	2	9	5	3	7	29
2:Medium	4				6	3	7	20
3:Large	2			1	1	1	4	9
4:Unknown		3	1	2		2	4	12
3:Dredawa	10	1		6	2		8	27
1:Small								
2:Medium	1				1		6	8
3:Large					1		1	2
4:Unknown	9	1		6			1	17
4:Oromia	13	17	7	28	21	16	36	138
1:Small	2	3	1	6	11	10	16	49
2:Medium	4	3		13	6	4	8	38
3:Large	5	8		7	3	2	4	29
4:Unknown	2	3	6	2	1		8	22
5:SNNPR	12	2	6	8	7	2	16	53
1:Small	2		3	4	1		6	16
2:Medium	5	1	3	2	3	2	5	21
3:Large	5	1		2	3		5	16
4:Unknown	2	3	6	2	1		8	22
6:Tigray	9	15	24	14	16	5	26	109
1:Small	2	6	14	6	11		15	54
2:Medium	3	7	9	4	4	2	7	36
3:Large	3	2		3		3	2	13
4:Unknown	1		1	1	1		2	6
Grand Total	76	86	77	155	117	93	244	848

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

Table 3: Achieved Interviews (Panel Firms)

Region	2:Textile and		3:Non-Metallic	4:Other			Grand Total	
	1:Food	Garments	Minerals	Manufacturing	5:Retail	6:Transportation		7:Other Services
1:Addis Ababa	16	30	6	73	29	16	92	262
1:Small	6	7	2	25	16	4	32	92
2:Medium	4	7	2	29	10	8	40	100
3:Large	6	16	2	19	3	4	20	70
2:Amhara	2			8	7		10	27
1:Small				7	3		3	13
2:Medium	1				4		5	10
3:Large	1			1			2	4
4:Oromia	4	2	1	19	1	2	14	43
1:Small		1	1	4	1	1	8	16
2:Medium	2			10			4	16
3:Large	2	1		5		1	2	11
5:SNNPR				2	4		9	15
1:Small				2	1		4	7
2:Medium					2		2	4
3:Large					1		3	4
6:Tigray	2	3	1	4	4	1	10	25
1:Small		2		4	4		8	18
2:Medium	1	1	1				2	5
3:Large	1					1		2
Grand Total	24	35	8	106	45	19	135	372

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 (*core* module) and respectfully additional manufacturing and services 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 *Services* questionnaire (includes the *core* module plus retail specific questions) and the residual eligible services have been covered using the *Services* questionnaire (includes the *core* module). 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 “ETH” indicate questions specific to Ethiopia, 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.

-*a4a*: coded following the stratification by sector as defined above.

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, sometimes 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. Most of the firms used July 2013 to June 2014 as their fiscal year. Also note for questions pertaining to monetary amounts, the unit is the Ethiopian Birr.

V. Universe Estimates

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

22. 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.

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

24. 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*.

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

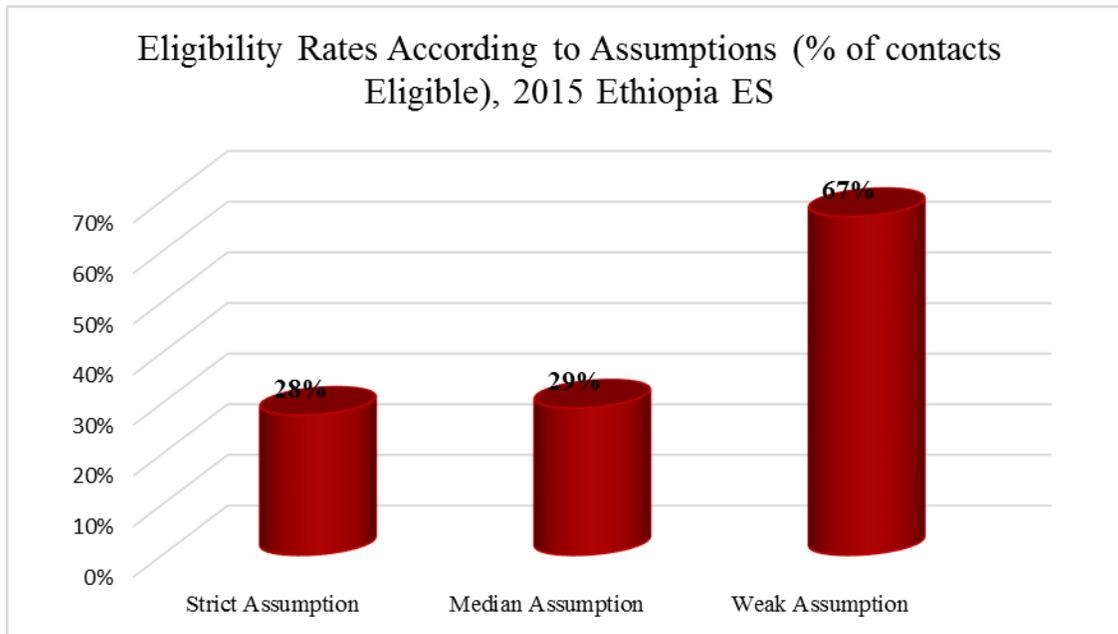
25. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *wmedian*.

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

26. 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*.

$$\text{Weak eligibility} = (\text{Sum of the firms with codes, } 1,2,3,4,16,10,11,13,91,92,93,94,12) / \text{Total}$$

27. 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.



28. Universe estimates for the number of establishments in each industry-region-size cell in Ethiopia were produced for the strict, weak and median eligibility definitions. Appendix B shows the universe estimates of the numbers of registered establishments that fit the criteria of the Enterprise Surveys.

29. 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

30. 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

31. 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, 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

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.

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

32. 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.

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

34. 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

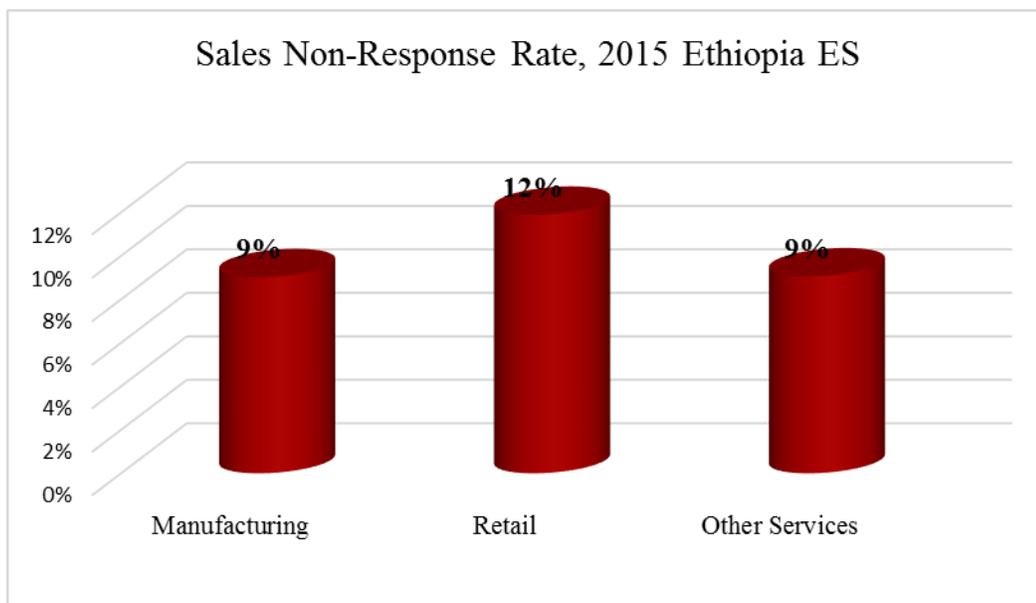
35. Survey non-response must be differentiated from item non-response. The former refers to refusals to participate in the survey altogether whereas the latter refers to the refusals to answer some specific questions. Enterprise Surveys suffer from both problems and different strategies were used to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

36. 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

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.

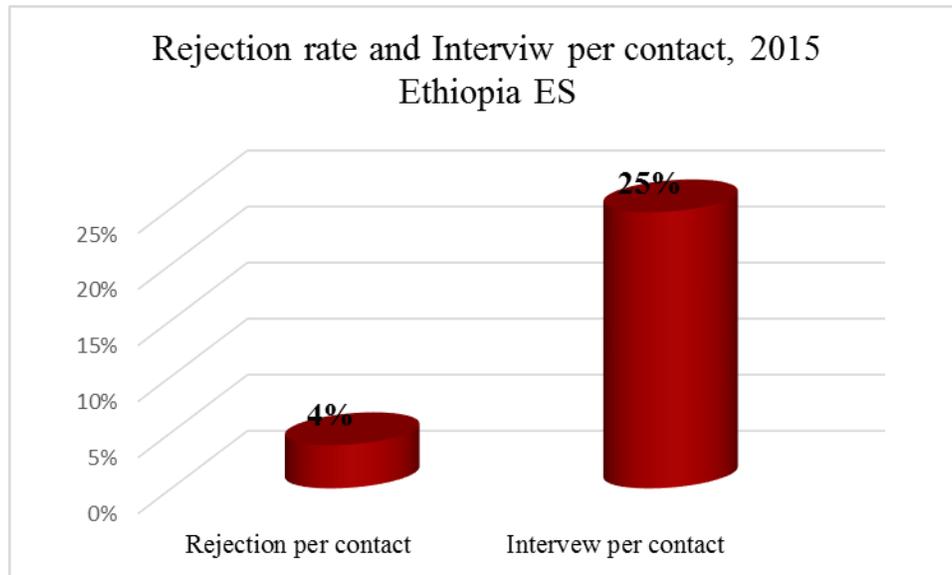
low response. The following graph shows non-response rates for the sales variable, *d2*, by sector. Please, note that the coding utilized in this dataset does not allow us to differentiate between “Don’t know” and “refuse to answer”, thus the non-response in the charts below reflect both categories (DKs and NAs).



37. 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.

38. As the following graph shows, the number of interviews per contacted establishments was 0.25.⁹ 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 share of rejections per contact was 0.04.

⁹ The estimate is based on the total number of firms contacted including ineligible establishments.



39. Details on the rejection rate, eligibility rate, and item non-response are available at the level of 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 Ethiopia. All sample based firm-level 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):

10	Screening in process	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	10
953	Eligible	1. Eligible establishment (Correct name and address) 2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment) 3. Eligible establishment (Different name but same address - the firm/establishment changed its name) 4. Eligible establishment (Moved and traced) 16. Eligible establishment (Panel Firm - now less than five employees; this code applies only to panel firms.)	933 3 2 14 1
46	Screener refusal	13. Refuses to answer the screener	46
1056	Ineligible	5. The establishment has less than 5 permanent full time employees 616. The firm discontinued businesses - (Establishment went bankrupt) 617. 618. The firm discontinued businesses - (Original establishment disappeared and is now a different firm) 619. The firm discontinued businesses - (Establishment was bought out by another firm) 620. The firm discontinued businesses - (It was impossible to determine for what reason) 621. The firm discontinued businesses - (Other) 7. Not a business: Private household 8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	769 2 0 2 1 235 6 8 33
85	Out of target	151. Out of target - outside the covered regions 152. Out of target - moved abroad 153. Out of target - Not registered with Statistical Authority 154. Out of target - establishment is HQ without production or sales of goods or services 155. Out of target - establishment was not in operation for the entirety of last fiscal year 156. Duplicated firm within the sample	12 2 0 5 22 44
1307	Unobtainable	91. No reply after having called in different days of the week and in different business hours 92. Line out of order 93. No tone 94. Phone number does not exist 10. Answering machine 11. Fax line- data line 12. Wrong address/ moved away and could not get the new references	30 178 5 6 2 1 1085
3447	Total contacted		

Response Outcomes : 2015 Ethiopia ES

Target and totals	Sample target	900
	Sample target completion rate	94.3%
	Total contacts available in frame	71822
	Total contacts issued	3734
	Total contacts contacted	3447
Screening phase	Screening in process	10
	Eligibles	953
	Screener refusal	46
	Ineligible + out of target	1141
	Unobtainable	1307
Interview phase (only if eligible)	Complete interviews without extra module	849
	Complete interviews with extra module	0
	Eligible in process + incomplete interviews	0
	Interview refusal	87
Percent breakdown (relative to total contacted)	Screening in process rate	0.3%
	Screener refusal rate	1.3%
	Ineligible + out of target rate	33.1%
	Unobtainable rate	37.9%
	Interview conversion rate	24.6%
	Eligible in process + incomplete interviews rate	0.0%
	Interview refusal rate	2.5%
	Total	27.2%

Appendix B: Universe Estimate Based on Sampling Weights

Appendix B-1: Universe Estimate based on Weak Weight

Region	1:Food	2:Textile and Garments	3:Non-Metalic Minerals	4:Other Manufacturing	5:Retail	6:Transportation	7:Other Services	Grand Total
1:AddisAbaba	190	115	69	728	2357	8245	18899	30603
1:Small	75	30	22	247	1296	4877	2956	9502
2:Medium	61	34	11	262	861	2474	1503	5207
3:Large	11	21	5	47	143	100	240	568
4:Unknown	44	30	32	172	57	793	14200	15328
2:Amhara	36	4	3	23	213	193	786	1259
1:Small	29	1	2	13	173	156	231	606
2:Medium	5				39	34	73	151
3:Large	2			1	1	1	7	12
4:Unknown		3	1	9		2	476	491
3:Dredawa	11	1		6	2		8	28
1:Small								
2:Medium	1				1		6	8
3:Large					1		1	2
4:Unknown	10	1		6			1	18
4:Oromia	475	92	7	151	1680	2397	2310	7112
1:Small	76	8	1	17	542	1330	544	2517
2:Medium	131	24		24	461	711	470	1820
3:Large	48	8		7	79	356	148	645
4:Unknown	220	52	6	103	599		1148	2129
5:SNNPR	392	2	6	21	28	2	134	585
1:Small	3		3	4	1		6	17
2:Medium	384	1	3	15	24	2	111	539
3:Large	5	1		2	3		17	28
4:Unknown								
6:Tigray	93	22	48	215	466	5	670	1519
1:Small	40	10	35	82	298		344	809
2:Medium	48	10	12	91	75	2	219	457
3:Large	4	3		35		3	75	120
4:Unknown	1		1	8	94		31	134
Grand Total	1198	236	134	1144	4746	10842	22806	41106

Appendix B-2: Universe Estimate based on Median Weight

Region	2:Textile and 3:Non-Metalic			4:Other	5:Retail	6:Transportation	7:Other Services	Grand Total
	1:Food	Garments	Minerals	Manufacturin				
1:AddisAbaba	153	78	53	629	886	2580	8494	12873
1:Small	50	16	15	172	394	1329	1267	3243
2:Medium	61	26	11	264	378	975	931	2646
3:Large	13	21	5	73	97	61	228	497
4:Unknown	30	16	22	119	17	216	6068	6488
2:Amhara	19	4	3	14	46	39	227	350
1:Small	13	1	2	9	33	27	63	149
2:Medium	4				11	9	29	52
3:Large	2			1	1	1	5	10
4:Unknown		3	1	4		2	130	140
3:Dredawa	10	1		6	2		8	27
2:Medium	1				1		6	8
3:Large					1		1	2
4:Unknown	9	1		6			1	17
4:Oromia	118	24	7	47	177	252	339	965
1:Small	15	3	1	6	48	106	68	248
2:Medium	38	5		13	59	82	86	283
3:Large	21	8		7	16	63	41	156
4:Unknown	44	8	6	21	53		144	277
5:SNNPR	639	2	6	32	21	2	145	848
1:Small	4		3	4	1		6	18
2:Medium	622	1	3	24	17	2	113	782
3:Large	13	1		4	3		26	48
6:Tigray	40	15	24	100	73	5	178	436
1:Small	13	6	14	27	44		71	176
2:Medium	23	7	9	44	16	2	66	167
3:Large	3	2		26		3	35	69
4:Unknown	1		1	3	14		6	25
Grand Total	979	125	93	829	1205	2878	9391	15499

Appendix B-3: Universe Estimate based on Strict Weight

Region	2:Textile and		3:Non-Metalic	4:Other		6:Transportation	7:Other Services	Grand Total
	1:Food	Garments	Minerals	Manufacturing	5:Retail			
1:Addis Ababa	146	73	47	566	816	2294	7489	11432
1:Small	48	14	13	154	358	1172	1128	2887
2:Medium	59	24	10	242	353	882	851	2420
3:Large	12	21	5	66	89	54	206	453
4:Unknown	28	14	19	105	15	186	5304	5672
2:Amhara	19	4	3	14	46	38	222	347
1:Small	13	1	2	9	34	27	63	148
2:Medium	4				11	9	29	53
3:Large	2			1	1	1	5	10
4:Unknown		3	1	4		2	126	136
3:Dredawa	10	1		6	2		8	27
1:Small								
2:Medium	1				1		6	8
3:Large					1		1	2
4:Unknown	9	1		6			1	17
4:Oromia	123	24	7	46	177	247	332	956
1:Small	16	3	1	6	48	103	67	244
2:Medium	40	5		13	61	82	86	287
3:Large	22	8		7	16	62	41	156
4:Unknown	45	8	6	20	52		138	270
5:SNNPR	671	2	6	32	21	2	143	878
1:Small	4		3	4	1		6	18
2:Medium	653	1	3	24	17	2	112	812
3:Large	14	1		4	3		25	48
4:Unknown								
6:Tigray	43	15	24	101	74	5	178	439
1:Small	14	6	14	27	44		70	175
2:Medium	25	7	9	45	16	2	67	171
3:Large	3	2		26		3	35	69
4:Unknown	1		1	3	14		6	24
Grand Total	1012	119	87	765	1137	2586	8372	14079

Appendix C: Concordance between Capital on Registration and Manpower

Level	Sector	Man power	Initial or End Capita
Micro	Manufacturing	<=5	<=100,000 Birr
	Service	<=5	<=100,000 Birr
Small	Manufacturing	6 - 30	<=1.5 million Birr
	Service	6 - 30	<=500,000 Birr
Medium	Manufacturing	31 - 100	<=20 million Birr
	Service	31 - 100	<=10 million Birr
Large	Manufacturing	More than 100	Above 20 million Birr
	Service	More than 100	Above 10 million Birr

Source: Updated Strategy Document for Manufacturing and Service development to be implemented GTP II, Ministry of Industry, Federal Democratic Republic of Ethiopia

Note: Birr is Ethiopia's currency

Appendix E: Original Sample Design

Original sample design for Fresh firms

Region	Size	Food	Textiles & Garments	Non-Metallic Minerals	Other Manufacturing	Retail	Transport	Other Services	Total
Addis Ababa	1:Small	2	4	12	4	15	15	15	67
	2:Medium	2	4	7	4	13	15	15	59
	3:Large	2	5	3	3	3	11	3	30
	4:Unknown	2	3	13	3	2	12	15	50
Amhara	1:Small	2	1	2	2	4	4	5	20
	2:Medium	3	0	0	0	2	2	2	9
	3:Large	1	0	0	0	1	1	2	5
	4:Unknown	0	2	1	2	0	2	7	14
Dredawa	1:Small	0	0	0	0	0	0	0	0
	2:Medium	1	0	0	0	1	0	6	8
	3:Large	0	0	0	0	1	0	1	2
	4:Unknown	9	1	0	6	0	0	1	17
Oromiya	1:Small	2	3	0	2	9	15	8	39
	2:Medium	2	3	0	2	6	9	5	29
	3:Large	2	7	0	2	3	5	2	21
	4:Unknown	3	4	2	2	8	3	11	32
SNNPR	1:Small	2	0	3	2	0	0	3	10
	2:Medium	7	1	3	3	3	2	4	22
	3:Large	5	1	0	2	2	0	3	13
	4:Unknown	0	0	0	0	0	0	0	0
TIGRAY	1:Small	2	3	15	2	6	2	5	36
	2:Medium	2	3	7	2	3	2	3	22
	3:Large	2	1	0	2	1	2	2	10
	4:Unknown	1	3	1	2	2	1	2	12
Total		54	50	69	47	85	103	121	528

Original sample design for Panel firms

Region	Size	Food	Textiles & Garments	Non-Metallic Minerals	Other Manufacturing	Retail	Transport	Other Services	Total
Addis Ab	1:Small	12	16	1	29	29	7	29	123
	2:Medium	5	10	2	28	28	8	28	109
	3:Large	9	19	1	28	8	6	21	92
	4:Unknown	0	0	0	0	0	0	0	0
Amhara	1:Small	0	1	0	9	7	0	4	21
	2:Medium	2	1	0	0	7	0	4	14
	3:Large	1	0	0	3	3	0	0	7
	4:Unknown	0	0	0	0	0	0	0	0
Dredawa	1:Small	0	0	0	0	0	0	0	0
	2:Medium	0	0	0	0	0	0	0	0
	3:Large	0	0	0	0	0	0	0	0
	4:Unknown	0	0	0	0	0	0	0	0
Oromiya	1:Small	3	1	1	9	12	1	10	37
	2:Medium	4	0	0	13	2	1	5	25
	3:Large	6	2	0	8	0	1	2	19
	4:Unknown	0	0	0	0	0	0	0	0
SNNPR	1:Small	0	0	0	2	4	0	3	9
	2:Medium	0	0	0	0	2	0	2	4
	3:Large	1	0	0	0	1	0	3	5
	4:Unknown	0	0	0	0	0	0	0	0
TIGRAY	1:Small	1	3	0	7	8	0	6	25
	2:Medium	1	1	0	4	0	0	2	8
	3:Large	0	0	0	1	0	1	0	2
	4:Unknown	0	0	0	0	0	0	0	0
Total		45	54	5	141	111	25	119	500

Appendix E

Local Agency team involved in the study:

Local Agency	Name: WAAS International Plc Country: Ethiopia Activities since: 1993
Enumerators involved:	Enumerators: 17 Recruiters: 4 full-time dedicated recruiters, and enumerators also participated in recruiting.
Other staff involved:	Fieldwork Coordinators: 1 Data Entry: None, CAPI was used Data Processing: 1

Sample Frame:

Characteristic of sample frame used:	There were two sources of sample frame: A. Panel sample frame, based on contacts from the 2011 Ethiopia ES B. Fresh sample frame, compiled from the Federal Ministry of Trade and Industry, and from regional trade and industry bureau.
Source:	The 2011 Ethiopia ES (for panel), and Ministry of Trade and regional (and zonal) Trade and Industry bureaus (for fresh firms).
Year:	The fresh sample frame was collected around the end of 2014, and the lists compiled were the most up-to-date available business registry data as of the time of compilation.
Additional list	None

Sectors included in the Sample:

Original Sectors	Seven sectors.
Added (top up) Sectors	None

Fieldwork and country situation:

Date of Fieldwork	June 2015 to February 2016
Country	Ethiopia
Use of CAPI	<ul style="list-style-type: none">• Yes
Problems found during fieldwork:	<p>The following were some of the challenges faced during the field work:</p> <ul style="list-style-type: none">- Unavailability of complete contact address for a large number of firms in the fresh sample frame. The problem was particularly the case for Oromia region where screening and recruitment for interview had to largely be done by searching the physical location of selected firms in person.- Information on firm size was missing for many firms in the fresh list, and a large number of firms turned out to be micro enterprises and micro enterprises are not part of the ES universe; this was particularly the case for firms in the transport sector.- Some of the respondents were not willing to give out financial related information.
Country specific situation	<p>A month before the end of the field work, there was security issue in most of the Oromia region and this has delayed the timely completion of the fieldwork in the region.</p>