

The Mauritania 2014 Enterprise Surveys Data Set

I. Introduction

1. This document provides additional information on the data collected in Mauritania between September 2014 to February 2015. 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 (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 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 Mauritania 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 two sectors – Manufacturing and services.

5. For the Mauritania 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 Mauritania ES was defined in two regions: Nouadhibou and Nouakchott.

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 Mauritania 2014 enterprise surveys roll out.

9. The frame was the Chamber of Commerce List from the Chambre de Commerce d'Industrie et d'Agriculture de Mauritanie (CCIAM) & Centre d'Information pour le Développement Economique et Technique (CIMDET). Appendix B provides counts of the frame. The database contained the following information

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

10. The enumerated establishments with 5 employees or more were then used as the sample frame for the Mauritania Enterprise Survey with the aim of obtaining interviews at 150 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 10.6% (36 out of 339 establishments)⁴.

Breaking down by industries and size, the following sample targets were achieved (using screener variables a3a, a4b and a6b):

Achieved sample

	Manufacturing	Services	Grand Total
Nouadhib	23	37	60
1-Small	2	24	26
2-Medium	12	11	23
3-Large	9	2	11
Nouakcho	29	61	90
1-Small	11	32	43
2-Medium	14	21	35
3-Large	4	8	12
Grand Total	52	98	150

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

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

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 “SNG” and “AF4” indicate questions specific to Mauritania or the region, 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. 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. Note that the fiscal years in Mauritania is the calendar year

V. Universe Estimates

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

22. Appendix B shows the overall estimates of the numbers of establishments in Mauritania based on the sample frame.

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*.

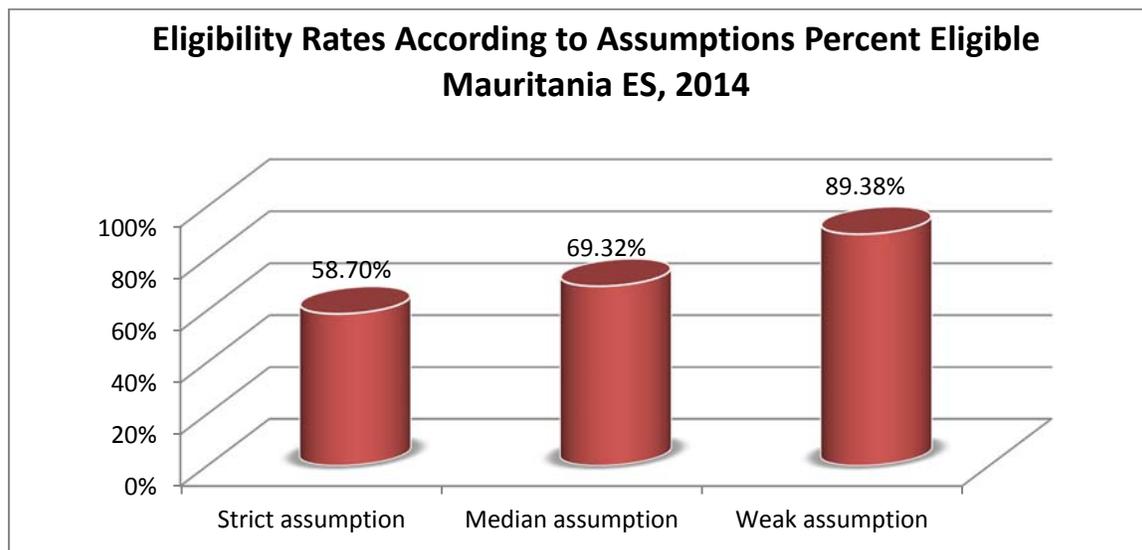
Median eligibility = (Sum of the firms with codes 1,2,3,4,16,10,11, & 13) / Total

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*.

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

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.



29. Universe estimates for the number of establishments in each industry-region-size cell in Mauritania 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.)⁵

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

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

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.

33. Appendix C shows the cell weights for registered establishments in Mauritania.

VII. Appropriate use of the weights

34. 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.

35. 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

36. 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

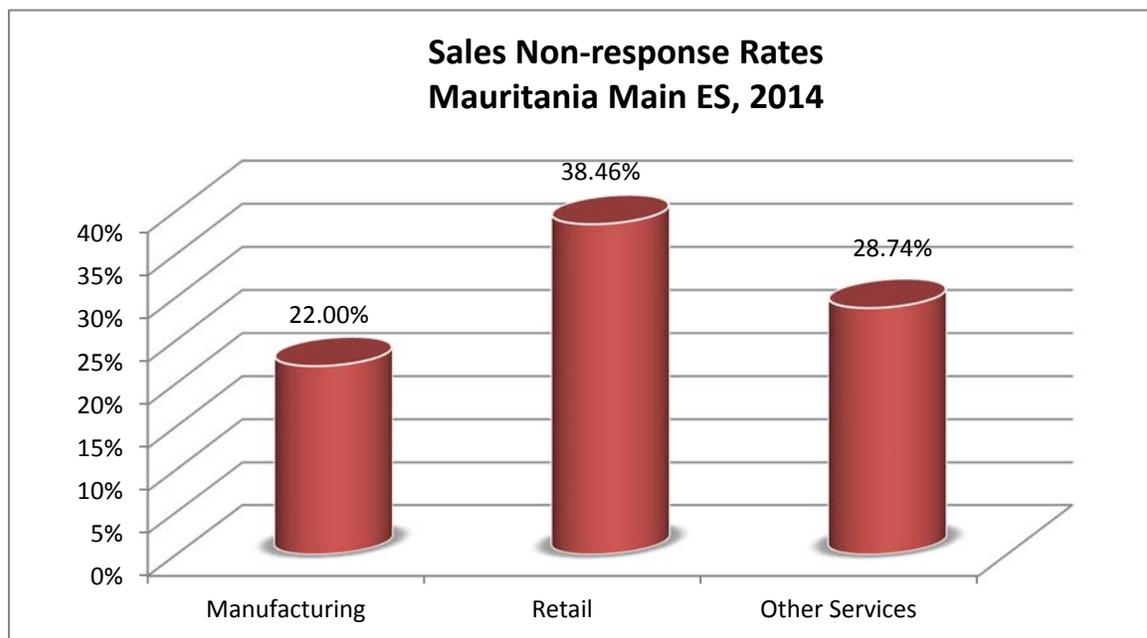
⁶ For the surveys that implemented a screener over the phone.

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.

37. 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.

38. 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 reflect both categories (DKs and NAs).

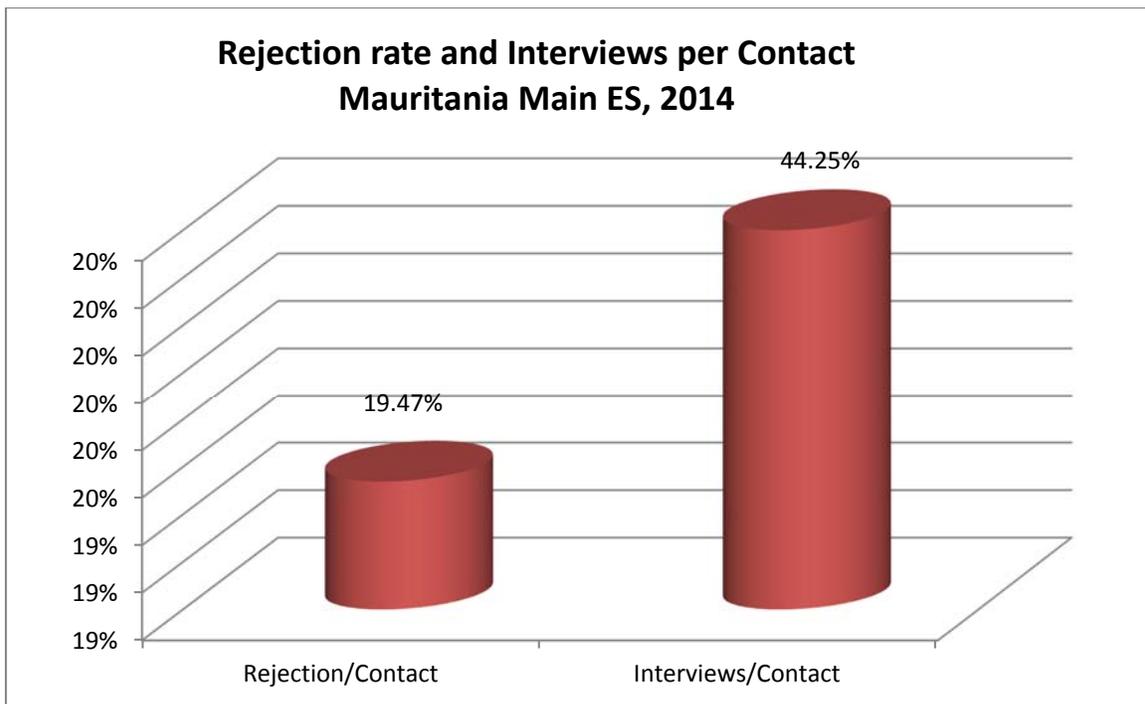


39. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) 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.

40. As the following graph shows, the percentage of interviews per contacted establishments was 44.25%⁹. This number is the result of two factors: explicit refusals to

⁹ The estimate is based on the total no. of firms contacted including ineligible establishments.

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 percentage of rejections per contact was 19.47%.



41. 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 Mauritania. 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):

	Mauritania
1. Eligible establishment (Correct name and address)	194
2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	3
3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	0
4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	2
5. The establishment has less than 5 permanent full time employees	14
6. The firm discontinued businesses	0
7. Not a business: private household	0
8. Ineligible activity: education, agriculture, finances, governments...	1
91. No reply (<i>after having called in different days of the week and in different business hours</i>)	9
92. Line out of order	1
93. No tone	6
94. Phone number does not exist	0
10. Answering machine	3
11. Fax line - data line	0
12. Wrong address/ moved away and could not get the new references	52
616. The firm discontinued business - (Establishment went bankrupt)	1
618. The firm discontinued business - (Original establishment disappeared and is now a different firm)	1
620. The firm discontinued business - (It was impossible to determine for whatever reason)	14
13. Refuses to answer the screener	33
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	0
152. Out of target - moved abroad	5
	339

Response Outcomes Enterprise (ES) Survey:

	Mauritania
Sample target	150
Complete interviews (Total)	150
Incomplete interviews	5
Elegible in process	11
Refusals	33
Ineligible	31
Impossible to contact	71
Ineligible - coop.	5
Refusal to the Screener	33
(Screener) in Process	0
	339

Appendix B

Universe Mauritania:

Source: Chamber of Commerce List

	Manufacturing	Services	Grand Total
Nouadhibou	72	56	128
1-Small	17	26	43
2-Medium	43	21	64
3-Large	12	9	21
Nouakchott	115	276	391
1-Small	64	188	252
2-Medium	41	67	108
3-Large	10	21	31
Grand Total	187	332	519

Appendix C

Strict Cell Weights Mauritania:

	Manufacturing	Services
Nouadhibou		
1-Small	1.25	1.09
2-Medium	2.26	1.70
3-Large	1.45	2.17
Nouakchott		
1-Small	3.30	2.33
2-Medium	1.36	2.63
3-Large	1.33	1.00

Median Cell Weights Mauritania

	Manufacturing	Services
Nouadhibou		
1-Small	1.23	1.24
2-Medium	2.16	1.88
3-Large	1.51	2.62
Nouakchott		
1-Small	3.94	3.23
2-Medium	1.58	3.54
3-Large	1.69	1.28

Weak Cell Weights Mauritania:

	Manufacturing	Services
Nouadhibou		
1-Small	1.46	1.43
2-Medium	2.21	1.86
3-Large	1.51	2.53
Nouakchott		
1-Small	6.43	5.10
2-Medium	2.21	4.80
3-Large	2.31	1.69

Appendix D

Strict Universe Estimates

	Manufacturing	Services	Grand Total
Nouadhibou	63	40	103
1-Small	13	16	29
2-Medium	41	17	58
3-Large	10	7	17
Nouakchott	60	120	179
1-Small	30	75	104
2-Medium	24	34	59
3-Large	5	11	16
Grand Total	123	160	283

Median Universe Estimates

	Manufacturing	Services	Grand Total
Nouadhibou	62	45	107
1-Small	12	19	31
2-Medium	39	19	58
3-Large	11	8	18
Nouakchott	71	163	234
1-Small	35	103	139
2-Medium	28	46	74
3-Large	7	14	21
Grand Total	132	209	341

Weak Universe Estimates

	Manufacturing	Services	Grand Total
Nouadhibou	65	48	113
1-Small	15	21	36
2-Medium	40	19	58
3-Large	11	8	18
Nouakchott	107	244	351
1-Small	58	163	221
2-Medium	40	62	102
3-Large	9	19	28
Grand Total	172	292	464

Appendix E

Original Sample Design, Mauritania:

	Manufacturing	Services	Grand Total
Nouadhibou	52	23	75
1-Small	12	3	15
2-Medium	28	11	39
3-Large	12	9	21
Nouakchott	23	52	75
1-Small	5	29	34
2-Medium	8	4	12
3-Large	10	19	29
Total	75	75	150

Completed Interviews, Mauritania:

	Manufacturing	Services	Grand Total
Nouadhib	23	37	60
1-Small	2	24	26
2-Medium	12	11	23
3-Large	9	2	11
Nouakcho	29	61	90
1-Small	11	32	43
2-Medium	14	21	35
3-Large	4	8	12
Grand Total	52	98	150

Appendix F

Local Agency team involved in the study:

Local Agency	Name: TNS RMS Senegal Country: Senegal Activities since: 2005
Enumerators involved:	Enumerators: 22 Recruiters: 22
Other staff involved:	Fieldwork Coordinators: 4 Data Entry: None, CAPI was used Data Editing and Processing: 6

Sample Frame:

Characteristic of sample frame used:	All registered firms
Source:	National Statistics Office Mauritania Chamber of Commerce and Industry
Year:	2013
Additional list	None

Sectors included in the Sample:

Original Sectors	Manufacturing, Services
Added (top up) Sectors	None

Fieldwork and country situation:

Date of Fieldwork	29 September 2014 – 11 February 2015
Country	Mauritania
Use of CAPI	<ul style="list-style-type: none"> • Yes
Problems found during fieldwork:	<ul style="list-style-type: none"> ▪ The initial data base provided for the study was outdated. Some firms in the list have shut down. ▪ Some of the firms have changed locations or changed their address so it was difficult to locate their new venues. ▪ There were numerous labor agitations during the time of the fieldwork that made some appointments to be rescheduled continually. ▪ Companies with landline numbers were difficult to contact, as most of the numbers have changed over time and this made it difficult to get in touch with these firms. ▪ Many of the respondents complained that the questionnaire is voluminous and thus takes their productive time when responding.
Country specific situation	<ul style="list-style-type: none"> • The numerous taxes imposed on firms and new customs regulations applied on enterprises makes some respondents hesitant to provide very accurate financials with the view that such information may be leaked to government authorities in charge of taxes and economic policies. • There were numerous labor agitations making some appointments to be rescheduled continuously. • During the “biological rest” (September to November 2014) in the fishing industry in Nouadhibou many firms closed down temporarily and the others that were still continuing operations did not have their managers or owners in the office as they travelled in this period.