Lao PDR Enterprise Surveys Data Set

1. Introduction

1.   This document provides additional information on the data collected in Lao PDR between May and July 2009 as part of the Indicators Survey component of the survey East Asia and Pacific Enterprise Survey 2009.

The objective of the survey is to obtain feedback from enterprises on the state of the private sector in client countries (Lao PDR, Tonga, Samoa, Vanuatu, Timor Leste, Fiji, Federated States of Micronesia and Papua New Guinea) as well as to help in building a panel of enterprise data that will make it possible to track changes in the business environment over time, thus allowing, for example, impact assessments of reforms.

Through interviews with firms in the manufacturing and services sectors, the survey will assess the constraints to private sector growth and 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.

2. Sampling Structure

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

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

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

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

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

---

[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

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 oblast (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 23 manufacturing industries, 2 services industries -retail and IT-, and one residual sector as defined in the sampling manual.

5.   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). For stratification purposes, the number of employees was defined on the basis of reported permanent full-time workers. This seems to be an appropriate definition of the labor force since seasonal/casual/part-time employment is not a common practice, except in the sectors of construction and agriculture.

6.   Regional stratification was defined in four regions: Vientiane Capital, Champasack, Luang Prabang, and Savannakhet.

3. 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 for the Indicator Surveys.

8.   One frame was used for Lao PDR. The sample frame used in Lao PDR was obtained from the 'Preliminary Survey in the Business Sector' (2008), maintained by the National Statistic Centre, Department of Statistics under the Ministry of Planning and Investment, Government of Lao PDR. The modified and translated sample frame was used to select the sample of establishments for the full survey. This database contained the following information:
    -Name of the firm
    -Contact details
    -ISIC code
    -Number of employees.

Counts from sample frame shown below.

Universe Figures for Lao PDR

ELIGIBLE UNIVERSE					
Region	Size	Manufacturing	Retail	Other Services	Grand Total
Champasack	1-Small	74	45	141	260
	2-Medium	2		25	27
	3-Large	1			1
	Region Total	77	45	166	288
LuanPrabang	1-Small	45	28	170	243
	2-Medium	6		34	40
	3-Large			3	3
	Region Total	51	28	207	286
Savannakhet	1-Small	59	73	128	260
	2-Medium	10	1	30	41
	3-Large	9		2	11
	Region Total	78	74	160	312
VTE Capital	1-Small	347	229	653	1229
	2-Medium	62	17	169	248
	3-Large	35	1	21	57
	Region Total	444	247	843	1534
Grand Total		650	394	1376	2420

9.   The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 360 establishments with five or more employees

10.   The quality of the frame was assessed at the onset of the project through calls 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. Due to response rate and ineligibility issues, additional sample had to be extracted from the universe in order to obtain enough eligible contacts and meet the sample targets.

11.   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% (40 out of 402 establishments).[4]   Breaking down by industry, the following numbers of establishments were surveyed:

Manufacturing	125
Retail & IT	118
Other services	117

---

[4] Appendix B shows the tabulations for the Lao PDR sample of registered firms of response codes that are classified as eligible and non-eligible.

Local Agency team involved in the study:

Local Agency	Name: MB Consulting Country: Lao PDR Membership of international organization: N/A Activities since: 2004
Name of Project Manager	Mr. Phoungsouvanh Sygnavong
Name and position of other key persons of the project:	Alexander Julian: Mekong Economics Country Coordinator Mr. Xayphone Chonephetsarath: Managing Director Mr. Phouagnsouvanh Mr Khamla Ms Peurn
Enumerators involved:	Enumerators: 14 Recruiters: 14
Other staff involved:	Fieldwork Coordinators: 1 Editing: 1 Data Entry: 1

Sample Frame:

Characteristic of sample frame used:	Variables: Name of establishment, address, activity, telephone number, number of employees
Source:	Preliminary Survey in Business Sector, held by National Statistics Centre
Year:	2008
Comments on the quality of sample frame:	Good quality and coverage, and contained all the variables needed to construct a sample frame for the enterprise survey. Some addresses outdated.
Year and organism who conducted the last economic census	National Statistics Centre, Ministry of Planning and Investment, 2008
Other sources for companies statistics	All the interviews were conducted with the sample frame obtained from the blocks enumeration and no other lists were used.

Sample Frame/Universe Lao PDR:

Sectors included in the sample:

Original Sectors	Manufacturing: 15,16,17,19,20,21,22,24,25,26, 28,29,36,37 Services: 45,50,51,52,55,60,61,62,63,64,72

Sample:

Comments/ problems on sectors and regions selected in the sample:	On sectors: Approximately 90% of the sectors selected in the sample were correct. On regions: No major problems.
Comments on the response rate	The overall response rate of establishments was very high. Establishments were mostly happy to participate
Comments on the sample design:	7 establishments in the large size category (preference 1) were determined ineligible after the screener: 100% public share, school, hospital, etc., and were replaced by preference 2

Fieldwork:

Date of Fieldwork	15th May 2009- 24th June 2009
Country	Lao PDR
Interview number	Manufacturing: 125 Services: 118 Other Services: 117
Problems found during fieldwork:	Bad weather (rainy season), long distance to enterprises outside the centre, and difficulty to find the exact location of some establishments. Some establishments had moved address which required tracing. Some individuals unwilling to provide income, sales, and expenditure information

4. Data Base Structure:

12.   Only one questionnaire – the Indicator Questionnaire – was used for all sectors. This questionnaire had two versions—one for manufacturing and one for services firms.

13.   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*. Variable names preceded by a prefix *"AF"* indicate questions specific to Africa and, 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.

14.   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.

15.   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.

16.   All of the following variables contain information from the sampling frame and were defined with the sampling design. 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 (oblasts)
> -*a6a*: coded using the same standard for 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), retail (52), and (45, 50, 51, 55, 60, 63, 72) for Other Services.

17.   The surveys were implemented following a 2 stage procedure. In the first stage a screener questionnaire was applied over the phone to determine eligibility and to make appointments; in the second stage, a face-to-face interview took place with the Manager/Owner/Director of each establishment. 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.

18. Note that there are additional variables for location (*a3x*) and 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.

19. Variable *a3x* indicates the actual location of the establishment. There may be divergences between the location in the sampling frame and the actual location, as establishments may be listed in one place but the actual physical location is in another place.

20. 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.

5. Universe Estimates

21. The enumerated totals were adjusted to take account of the establishments found to be ineligible when interviews were attempted. Then ratios of the total numbers of blocks of each type to the totals enumerated were formed. Those ratios were then applied to the eligible establishments enumerated to provide universe estimates.

22. Appendix C shows the overall estimates of the numbers of establishments in Lao PDR 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 *w_strict*.
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 *w_median*.
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 *w_weak*.
Weak eligibility= (Sum of the firms with codes 1,2,3,4,16,91,92,93,10,11,12,&13) / Total

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

Eligibility Rates: Lao PDR 2009

Strict Assumptions	Median Assumptions	Weak Assumptions
89.6%	90.0%	92.3%
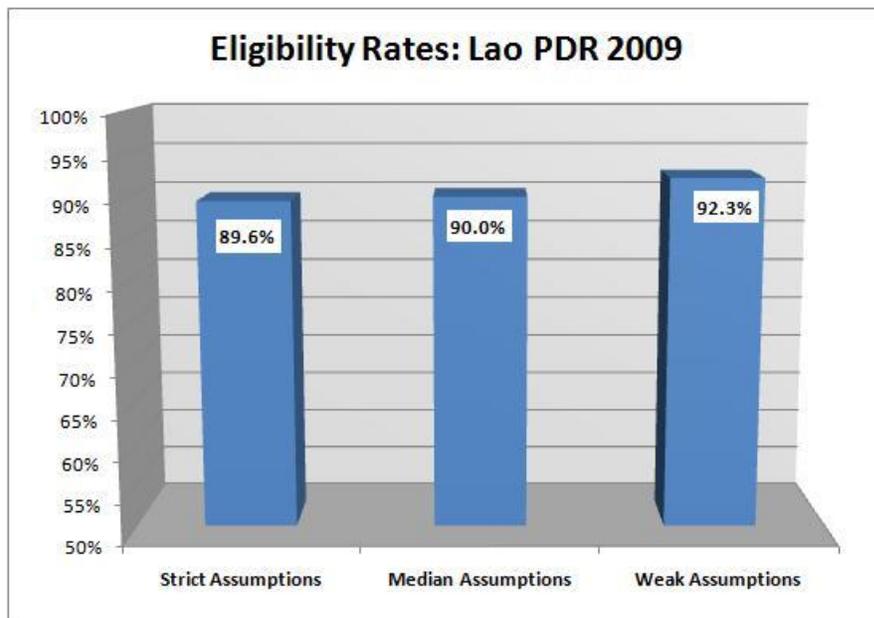
29. Universe estimates for the number of establishments in each industry-region-size cell in Lao PDR were produced for the strict, weak and median eligibility definitions. Appendix C shows the universe estimates of the numbers of registered establishments.

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.

6. 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]

---

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

32.   Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each region/industry/size stratum to account for the presence of ineligible units (the firm discontinued businesses or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, out of order, no tone in the phone line, answering machine, fax line, 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 D shows the cell weights for registered establishments in Lao PDR.

7. 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.)[6]

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.[7] If the models are developed as structural relationships or behavioral models that may vary for different parts of the population, then, there is no reason to use weights.

8. Non-response

---

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

[7] The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

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 table below reflects both categories (DKs and NAs).
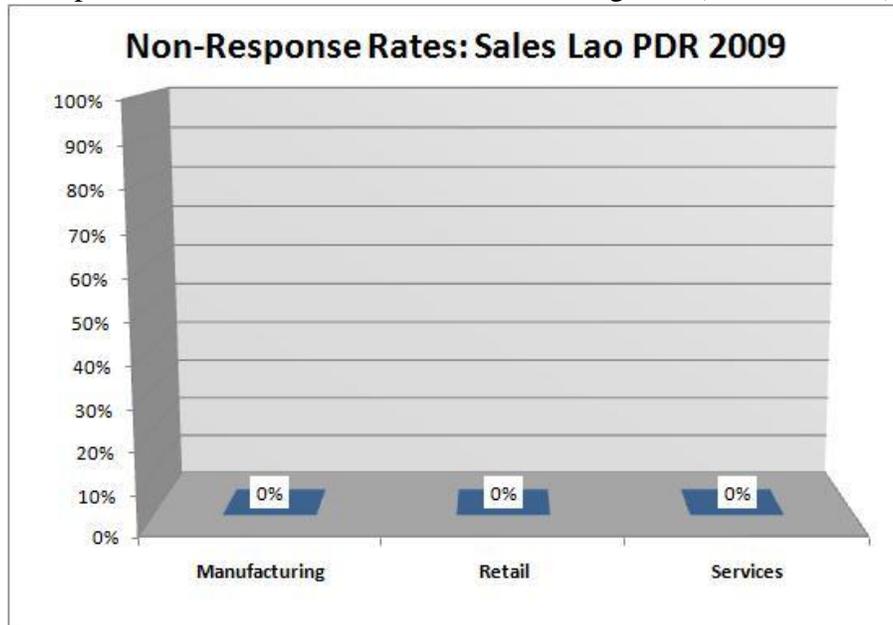
39.   Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics) 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 number of contacted establishments per realized interview was 1.12. 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 0.005.

Interview Rejection Rate and Contacts per Interview: Lao PDR 2009

- Contacts per Interview: 1.117
- Refusals per Contact: 0.005
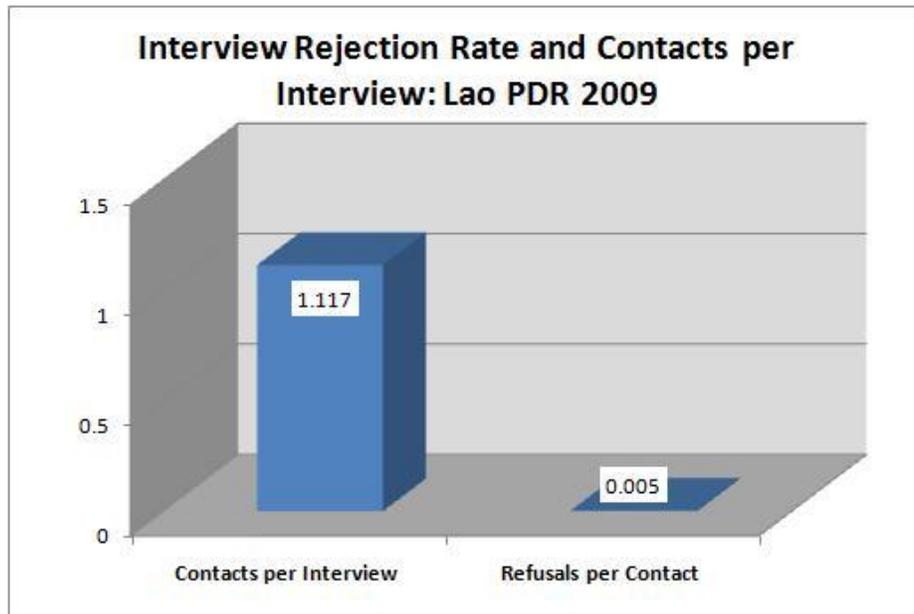
41.   Details on the rejection rate, eligibility rate, and item non-response are available at the level strata. This report summarizes these numbers to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and faulty sampling frames are not unique to Lao PDR. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

Appendix A

Questionnaires:

Problems for the understanding of questions (write question number)	Problems with understanding questions K.1, K2, D.3, and M.1 were found during the pilot survey. The survey manager and supervisors then allocated additional time for training to resolve any remaining understanding issues.
Problems found in the navigability of questionnaires (for example, skip patterns).	No problems
Comments on questionnaire length:	A number of respondents commentated that the questionnaire was too long, although in most cases the questionnaire was completed taking under one hour.
Suggestions or other comments on the questionnaire:	Make question wording shorter and more concise

Database:

Comments on the data entry program:	Data entry program chosen: CSPro Comments: The software program required some training by the MKE Country Coordinator but was overall easy to use
Comments on the data cleaning:	Checking for data errors and inconsistencies was conducted by MKE and a quality control report and list of corrections was provided to the Lao data entry staff

Country Situation:

General aspects of the economic, political or Social situation in your territory that could affect the results of the survey:	Lao PDR is a one party communist state and the political and social situation in the country is currently stable. Lao PDR is classified as a least Development Country (LDC), with an agricultural based economy and over 80% of the population residing in rural areas. Non primary industries are at a relatively early stage of development. As a result there were few large scale establishments in the manufacturing and service sector to be surveyed.
Relevant local events occurred during fieldwork:	Lao PDR New Year festival from 13th to 17th April

Appendix B

Status Codes:

	Eligibles	
Eligibles	1.Elegible establishment (Correct name and address)	360
	2. Elegible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	0
	3. Elegible establishment (Different name but same address - the firm/establishment changed its name)	0
	4. Elegible establishment (Wrong address - the firm/establishmen has changed address and the address could be found)	0
Ineligibles	5. The establishment has less than 5 permanent full time employees	0
	6. The firm discontinued businesses	15
	7. Not a business: private household	0
	8. Ineligible activity: education, agriculture, finances, governments...	16
Unobtainable	91. No reply (after having called in different days of the week and in different business hours)	5
	92. Line out of order	1
	93. No tone	0
	10. Answering machine	0
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	3
	13. Refuses to answer the screener	2
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
	151. Out of target - outside the covered regions, firm moved abroad	0
	152. Out of target - firm moved abroad	0
	Total	402

Response Outcomes:

Complete interviews	360
Incomplete interviews	0
Eligible in process	0
Refusals	0
Out of Target	31
Impossible to contact	9
Ineligible - Coop.	0
Refusal to the Screener	2
Total	402

Appendix C

Universe Estimates, Lao PDR:

ELIGIBLE UNIVERSE					
Region	Size	Manufacturing	Retail	Other Services	Grand Total
Champasack	1-Small	74	45	141	260
	2-Medium	2		25	27
	3-Large	1			1
	Region Total	77	45	166	288
LuanPrabang	1-Small	45	28	170	243
	2-Medium	6		34	40
	3-Large			3	3
	Region Total	51	28	207	286
Savannakhet	1-Small	59	73	128	260
	2-Medium	10	1	30	41
	3-Large	9		2	11
	Region Total	78	74	160	312
VTE Capital	1-Small	347	229	653	1229
	2-Medium	62	17	169	248
	3-Large	35	1	21	57
	Region Total	444	247	843	1534
Grand Total		650	394	1376	2420

Appendix D

Strict Cell Weights Lao PDR:

STRICT Weights				
Region	Size	Manufacturing	Retail	Other Services
Champasack	1-Small	10.70	2.05	43.57
	2-Medium	0.87		2.49
	3-Large	1.02		
LuanPrabang	1-Small	9.05	2.08	39.15
	2-Medium	1.30		2.69
	3-Large			0.93
Savannakhet	1-Small	12.07	2.57	39.97
	2-Medium	1.10	0.91	2.01
	3-Large	1.03		0.94
VTE Capital	1-Small	40.59	6.52	32.92
	2-Medium	1.22	0.94	3.36
	3-Large	0.94	0.98	1.01

Median Cell Weights Lao PDR:

MEDIAN Weights				
Region	Size	Manufacturing	Retail	Other Services
Champasack	1-Small	10.65	2.07	43.38
	2-Medium	0.87		2.48
	3-Large	1.01		
LuanPrabang	1-Small	9.01	2.11	38.97
	2-Medium	1.29		2.68
	3-Large			0.92
Savannakhet	1-Small	12.01	2.61	39.79
	2-Medium	1.09	0.92	2.00
	3-Large	1.02		0.93
VTE Capital	1-Small	40.80	6.67	33.09
	2-Medium	1.22	0.96	3.38
	3-Large	0.94	0.99	1.01

Weak Cell Weights Lao PDR:

WEAK Weights				
Region	Size	Manufacturing	Retail	Other Services
Champasack	1-Small	10.67	2.06	44.98
	2-Medium	0.89		2.63
	3-Large	1.03		
LuanPrabang	1-Small	8.83	2.05	39.51
	2-Medium	1.29		2.78
	3-Large			0.95
Savannakhet	1-Small	11.94	2.57	40.92
	2-Medium	1.11	0.93	2.11
	3-Large	1.03		0.98
VTE Capital	1-Small	41.32	6.70	34.68
	2-Medium	1.27	0.99	3.63
	3-Large	0.97	1.01	1.07

Appendix E

Original Sample Design, Lao PDR:

The original aim was to obtain 120 interviews with manufacturing establishments 120 interviews with retail establishments and 120 interviews with establishments in the services sectors.

A total of 2420 eligible establishments were enumerated within the manufacturing, retail and services. They were divided into four preferences where available. In total 642 establishments were issued (only preferences one and two). Interviewers were to attempt interviews with the first preferences before they could move on to subsequent preferences.

Completed Interviews, Lao PDR:

Region	Size	Manufacturing	Retail	Other Services	Grand Total
Champasack	1-Small	7	23	3	33
	2-Medium	2		8	10
	3-Large	1			1
	Region Total	10	23	11	44
LuanPrabang	1-Small	5	14	4	23
	2-Medium	4		10	14
	3-Large			3	3
	Region Total	9	14	17	40
Savannakhet	1-Small	5	30	3	38
	2-Medium	8	1	12	21
	3-Large	9		2	11
	Region Total	22	31	17	70
VTE Capital	1-Small	8	34	17	59
	2-Medium	41	15	37	93
	3-Large	35	1	18	54
	Region Total	84	50	72	206
Grand Total		125	118	117	360