

Samoa (Upolu) Enterprise Surveys Data Set

1. Introduction

1. This document provides additional information on the data collected in Samoa between July and October 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.

This 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 Samoa was selected using stratified random sampling, following the methodology explained in the Sampling Manual¹. 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.

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. Two levels of stratification were used in this country: industry and establishment size. The original sample design with specific information of the industries chosen is described in Appendix E.
 4. Industry stratification was designed in the way that follows: the universe was stratified into 23 manufacturing industries, and one services 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 did not take place as only the island of Upolu, containing the capital city of Apia, was surveyed. Of the two islands that make up the majority of Samoa, Upolu has the largest population.

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. Due to limited data sources available in Samoa on registered businesses, the final sample frame was obtained from a combined dataset obtained from the Samoa National Provident Fund (SNPF). The list provided by the SNPF was limited to including information on the sector and location of enterprises, with no details on the number of employees. Therefore, original sample counts were not able to be stratified by enterprise size. The combined sample frame was then reviewed and duplicate establishments or establishments with ineligible characteristics (industry sector, number of employees, geographic location) removed from the list. The modified 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
 - Location
 - ISIC code

Counts from sample frame shown below.

Universe Figures for Samoa

ELIGIBLE UNIVERSE				
Region	Size	Manufacturing	Services	Grand Total
Samoa	All Sizes	141	696	837
Grand Total		141	696	837

9. The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 150 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 50% (416 out of 835 establishments).⁴ Breaking down by industry, the following numbers of establishments were surveyed:

Manufacturing	24
Services	85

Local Agency team involved in the study:

Local Agency	Name: KVA Consult Ltd. Location: Apia, Samoa Membership of international organization: N/A Activities since: 1990
Name of Country Team Leader	Mrs. Rachel Vai
Local Survey Implementation Team	Mrs. Audrey Brown-Pereira Mrs. Rachel Vaai

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

	Mr. Ken Tuioti (3 other Enumerators also on team)
Other staff involved:	Ms Kristin Smart – South Pacific Regional Coordinator

Sample Frame:

Characteristics of sample frame used	Variables: Name of establishment, address, activity, telephone number, number of employees
Sources:	Samoa National Provident Fund
Year:	2008 / 2009
Comments on the quality of sample frame:	Poor quality with limited coverage and out-dated information on operational business and contact information. No details on number of employees (firm size) available.
Year and organism that conducted the household and income survey (HIES) report census	Samoa Statistics Department, Ministry of Finance (2002)
Other sources for companies statistics	Not available

Sectors included in the sample:

Original Sectors	<ul style="list-style-type: none"> • Manufacturing (ISIC Sector D codes 15 to 37 inclusive), • Services (ISIC Sectors F, G, H, I, and Sector K code 72)
Added (top up) Sectors	(none)

Sample:

Comments/ problems on sectors and regions selected in the sample	Sectors and region for sample would have been enough to meet target of 150 if the original sample lists had been of higher quality (information on number of employees, updated, etc.). Given that this was not the case, expanding the sample region to include the other main island in Samoa would have allowed the country team to meet the 150 target for completed interviews.
--	--

Comments on the response rate	Response was high for eligible enterprises with updated contacted details, although there was a high rate of refusal to answer the screener and would have been better addresses by ensuring enumerator team were further trained in the screening / calling process.
Comments on the sample design:	None
Other comments:	Large proportion of enterprises on the sample list was either ineligible or impossible to contact.

Fieldwork:

Date of Fieldwork	25 th of May to 9 th of October, 2009
Location	Upolu, Samoa
Interview number	Manufacturing: 26 Services: 83
Problems found during fieldwork	Country team was often a little delayed in the submission of the Progress Reports and the extension of the survey past the originally planned 2 months meant that the team began to lose energy for the survey by September.
Other observations:	Country team also was simultaneously managing larger projects directly for the WB and other multi-lateral organizations are were not able to prioritize this survey.

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 “*EA*” indicate questions specific to East Asia and Pacific 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 two levels of stratification: industry and size. 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 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 (*11*, *16* and *18*) 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 *11*, *16* and *18* 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 Samoa 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

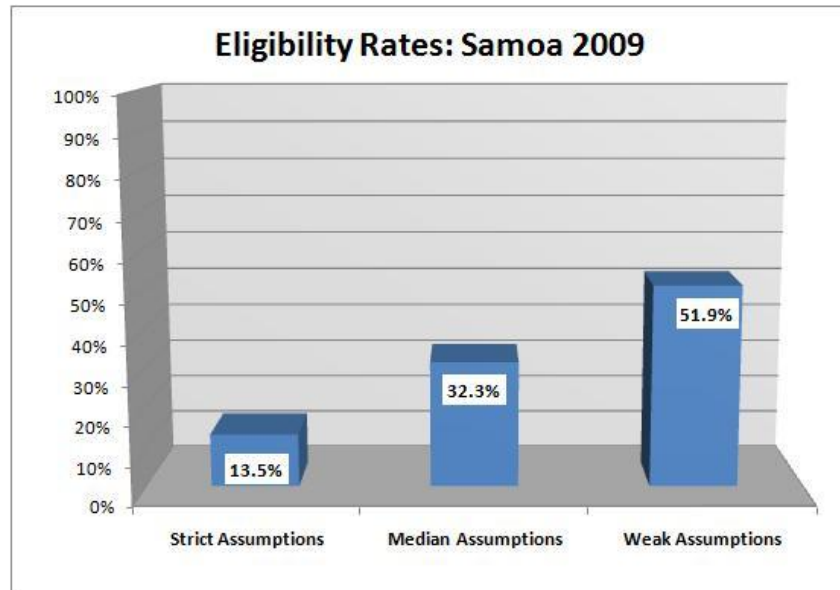
26. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w_median*.

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.



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

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

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

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 Samoa.

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

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.

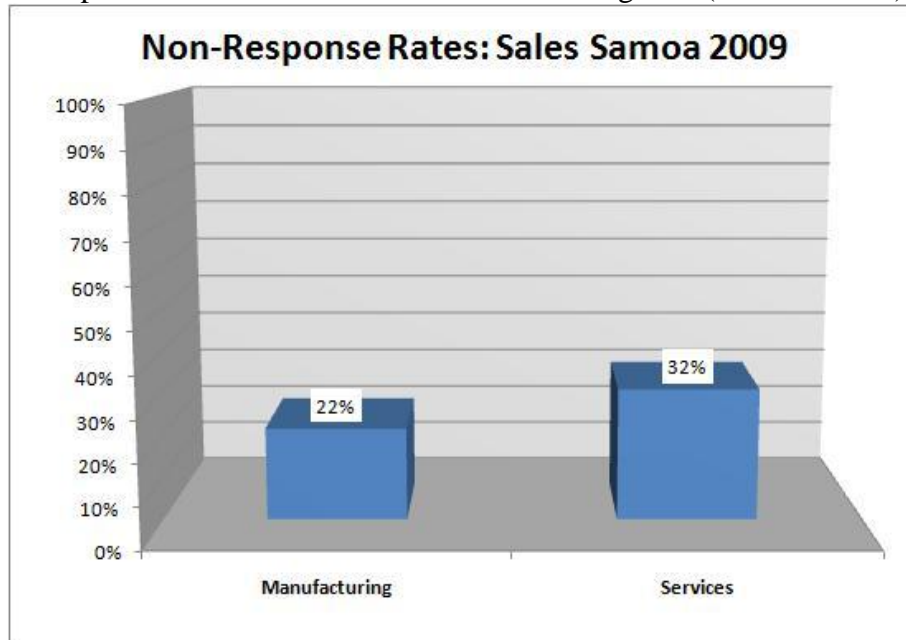
38. Item non-response was addressed by two strategies:

- 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).

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.

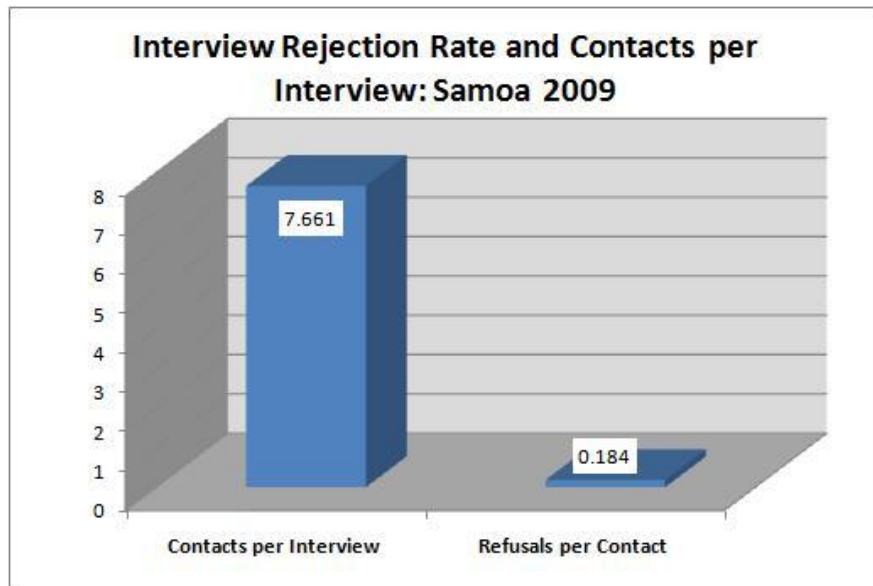
⁷ The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

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 7.66. 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 0.18.



41. Details on the rejection rate, eligibility rate, and item non-response are available at the strata level. 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 Samoa. 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)	Some problems among enterprises for question d1a2 that found choosing a code number that best described their main product / service difficult. Question k3, about working capital, was also difficult as most respondents ran small family owned businesses. Similar problems arose for questions related to collateral (k13, 14 and 15a) and fixed assets (k5) as concepts generally unfamiliar for smaller businesses.
Problems found in the navigability of questionnaires (for example skip patterns)	Problem with skip pattern for L.10 in the Manufacturing version of the questionnaire. Question L.6 instructed skip to L.30 when should have said skip to L.10. Data on L.10 missing for Manufacturing firms. Took enumerators a few days to get hang of the skip patterns, a longer pilot interview period would be suggested for future surveys.
Comments on questionnaire length:	Respondents often complained questions were lengthy and some said they would prefer to fill out questionnaire themselves. Length of questionnaire meant that by the end respondents were ready to finish the interview and speed through the last few questions, bringing to question the quality of the data provided at the end.
Suggestions or other comments on the questionnaire:	Clear up skip patterns and formatting to make more favorable to enumerator.

Database:

Comments on the data entry program	Data entry program chosen: CSPro The software programme required some training by the MKE Data Entry Team coordinator but was overall easy to use. The data entry program allowed for mistakes in skip patterns to be entered by the data entry team if they were did not use CSPro correctly. It would be helpful to use a different data entry program in the future that allows for stronger limits on data entry team errors.
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 data entry staff The call backs process was complicated by the amount of time between the shipment of surveys leaving Samoa, data entry by the Hanoi team and cleaning process by MKE and the WB. By the time MKE requested call backs of the Samoa country team the team had been disassembled and difficult to organize.

Country Situation:

General aspects of the economic, political or Social situation in your territory that could affect the results of the survey:	The business community in Samoa is made up mainly of family businesses thus there was difficulties in getting responses to sensitive financial questions, partially because of a lack of formal accounting systems and also because of uncertainty among respondents of what details were appropriate to share and which to keep within the family.
Relevant local events occurred during fieldwork:	Tsunami (29 th September) which wiped out 2 enterprises that had already completed the survey and made those who were in progress to complete an interview reluctant to participate.
Other aspects:	Main problems were related to getting responses to financial questions and preventing respondents from getting angry / annoyed with these questions. The country team suggested that asking for percentages (ex. Sales in comparison to 5 years ago, etc.) might be better than actual dollar terms.

Appendix B

Status Codes:

Eligibles	1. Eligible establishment (Correct name and address)	106
	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	6
	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	1
	4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	0
Ineligibles	5. The establishment has less than 5 permanent full time employees	79
	6. The firm discontinued businesses	134
	7. Not a business: private household	20
	8. Ineligible activity: education, agriculture, finances, governments...	17
Unobtainable	91. No reply (<i>after having called in different days of the week and in different business hours</i>)	58
	92. Line out of order	21
	93. No tone	38
	10. Answering machine	3
	11. Fax line - data line	0
	12. Wrong address/ moved away and could not get the new references	46
	13. Refuses to answer the screener	154
	14. In process (<i>the establishment is being called/ is being contacted - previous to ask the screener</i>)	0
	151. Out of target - outside the covered regions, firm moved abroad	152
	152. Out of target - firm moved abroad	0
		835

Response Outcomes:

Complete interviews (Total)	109
Incomplete interviews	0
Elegible in process	4
Refusals	0
Out of target	250
Impossible to contact	166
Ineligible - coop.	152
Refusal to the Screener	154
Total	835

Appendix C

Universe Estimates, Samoa:

ELIGIBLE UNIVERSE				
Region	Size	Manufacturing	Services	Grand Total
Samoa	All Sizes	141	696	837
Grand Total		141	696	837

Appendix D

Strict Cell Weights Samoa:

STRICT Weights			
Region	Size	Manufacturing	Services
Samoa	All Sizes	1.00	1.05

Median Cell Weights Samoa:

MEDIAN Weights			
Region	Size	Manufacturing	Services
Samoa	All Sizes	2.17	2.57

Weak Cell Weights Samoa:

WEAK Weights			
Region	Size	Manufacturing	Services
Samoa	All Sizes	3.50	4.12

Appendix E

Original Sample Design, Samoa:

The original aim was to obtain 75 interviews with manufacturing establishments and 75 interviews with establishments in the services sectors. However, upon inspection of the available sampling frame it became clear that this was an unrealistic goal.

A total of 835 eligible establishments were enumerated within manufacturing and services. Due to the size of the economy and Samoa's reliance on tourism we found very few manufacturing establishments on this list. Additionally, it was found that half of all the establishments on the list were either ineligible or unobtainable. The 141 manufacturing enterprises were split into two preferences while the 696 services establishments were divided into nine preferences. In total, all 837 establishments were issued. Interviewers were to attempt interviews with the first preferences before they could move on to subsequent preferences. Once it was found that it would be impossible to complete 75 manufacturing interviews, the team moved on to completing as many interviews in the services sector as possible.

Completed Interviews, Samoa:

Region	Size	Manufacturing	Services	Grand Total
Samoa	All Sizes	24	85	109
Grand Total		24	85	109