

# The People's Republic of China 2012 Enterprise Surveys Data Set

## 1. Introduction

1. This document provides additional information on the data collected in China between December 2011 and February 2013. A total of 2,700 privately-owned and 148 state-owned were successfully interviewed over this period. All questions referring to the last complete fiscal year refer to FY2011. Establishments interviewed prior to the end of FY2011 were re-interviewed to update these responses.

The objective of the survey is to obtain feedback from enterprises on the state of the private sector in client countries 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 China was selected using stratified random sampling, following the methodology explained in the Sampling Manual<sup>1</sup>. Stratified random sampling<sup>2</sup> was preferred over simple random sampling for several reasons<sup>3</sup>:

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

b. To obtain unbiased estimates for the whole population of privately owned firms. The whole population, or “universe” of the study, is the non-agricultural economy of firms with at least 5 employees and positive amounts of private ownership. The non-agricultural economy 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.)

---

<sup>1</sup> The complete text can be found at [http://www.enterprisesurveys.org/documents/Implementation\\_note.pdf](http://www.enterprisesurveys.org/documents/Implementation_note.pdf)

<sup>2</sup> 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).

<sup>3</sup> 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. For the case of China, a special sample of fully state-owned establishments was included as this is an important part of the economy. Data on 148 state-owned enterprises is provided separately from the data of 2,700 private sector firms. To maintain comparability of the China Enterprise Surveys to surveys conducted in other countries, only the data set of privately sector firms should be used.

4. 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 the Annex (Excel workbook).

5. Industry stratification was designed in the way that follows: the universe was stratified into 11 manufacturing industries and 7 services industries as defined in the sampling manual. Each manufacturing industry had a target of 150 interviews. Sample sizes were inflated by about 20% to account for potential non-response cases when requesting sensitive financial data and also because of likely attrition in future surveys that would affect the construction of a panel. **Note that 100% government owned firms are categorized independently of their industrial classification.**

6. The 148 surveyed state-owned enterprises were categorized as a separate sector group to preserve the representativeness of other sector groupings for the private economy. For example, a government owned bakery would not be classified as ‘food manufacturing’ in the dataset.

7. 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.

8. Regional stratification was defined in twenty-five metro areas: **(Table 1)**

	N
Beijing (municipalities) 北京	121
Chengdu City 成都	121
Dalian City 大连	121
Dongguan City 东莞	99
Foshan City 佛山	120
Guangzhou City 广州	120
Hangzhou City 杭州	85

Hefei City 合肥	121
Jinan City 济南	121
Luoyang City 洛阳	119
Nanjing City 南京	120
Nantong City 南通	118
Ningbo City 宁波	120
Qingdao City 青岛	119
Shanghai (municipalities) 上海	37
Shenyang City 沈阳	117
Shenzhen City 深圳	121
Shijiazhuang City 石家庄	121
Suzhou City 苏州	120
Tangshan City 唐山	120
Wenzhou City 温州	120
Wuhan City 武汉	107
Wuxi City 无锡	120
Yantai City 烟台	120
Zhengzhou City 郑州	120
	<b>2,848</b>

### 3. Sampling implementation

9. 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.

10. A single frame was used for China. The sample frame containing fresh contacts used in the China was obtained by SunFaith from SinoTrust. The database contained the following information:

- Name of the firm
- Location
- Contact details
- ISIC code
- Number of employees.

11. Universe Figures for China are available in the Annex in the Excel Workbook: “China\_2012\_Implementation\_Report\_Annex.xlsx”

12. The enumerated establishments were then used as the frame for the selection of a sample with the aim of obtaining interviews at 3000 establishments with five or more employees

13. 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.

14. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments are 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 31% (6485 out of 20,616 establishments).<sup>4</sup> Breaking down by industry, the following numbers of establishments were surveyed: **(Table 2)**

	N
15 Food	147
17 Textiles	154
18 Garments	126
24 Chemicals	143
25 Plastics & rubber	149
26 Non metallic mineral products	150
27 Basic metals	89
28 Fabricated metal products	175
29 Machinery and equipment	152
31 - 32 Electronics	161
34 - 35 Motor Vehicles	135
38 Other Manufacturing	112
45 Construction Section F:	133
50 Services of motor vehicles	144
51 Wholesale	146
52 Retail	153
55 Hotel and restaurants: section H	150
60 Transport Section I: (60-64)	143
72 IT	138
100% Government Owned	148
	<b>2848</b>

**Local Agency teams involved in the study:**

Local Agency #1	Name: Mekong Economics Ltd. Location: Hanoi, Vietnam Membership of international organization: N/A Activities since: 2000
Name of Project Manager	Mr. Adam McCarty
Name and position of other key persons of the project: Local Survey	<b>MKE Project Coordinator:</b> Ms. Tran Ngoc Diep; Mr Mark Oldenbeuving

Implementation Team and corresponding supervisor and enumerator codes:	<b>Data entry Specialist:</b> Ms. Nguyen Thi Minh Hieu <b>Call-back staff:</b> Nguyen Minh Thu Huyen, Nguyen Thuy Chi, Duong Hong Trang
--	--

Local Agency #2	<b>Name:</b> SINOFAITH IP GROUP (SUNFAITH China LTD.) <b>Location:</b> 28 F, Tower B, Eton Plaza, 555 Pudong Avenue, Shanghai 200120, China <b>Membership of international organization:</b> AMA, SCIP, INTA, AmCham Shanghai <b>Activities since:</b> 1996
Name of Project Manager	Sandy Cui & Bob Gao
Name and position of other key persons of the project: Local Survey Implementation Team and corresponding supervisor and enumerator codes:	<b>QC Manager: Joe Pan</b> <b>Call-back staff: Ben Cai, Yang, He Peng cheng, Huang Chun li</b>
Other staff involved:	<b>Questionnaire entry:</b> 7 data entry staff  There are 25 local teams, spread over 25 cities, that were involved in the implementation of the enterprise survey. The total number of team members are 221, including 25 Supervisors, 30 Screeners, 137 Enumerators and 29 local quality control (QC) staff.

### Sample Frame:

Characteristics of sample frame used	Variables: Name of establishment, address, activity, telephone number, number of employees
Year:	2008
Comments on the quality of sample frame:	The Universe Table was originally to be composed using 2008 census data from National Bureau of Statistics (NBS). However, this database did not contain data on number of employees. Although the 2008 census questionnaire did contain a question on number of employees, NBS census enumerators failed to collect the information from all enterprises, and therefore NBS didn't include it in the final database. The original sample was

	therefore provided by SunFaith and drawn by Sinotrust.
Year and organism who conducted the last economic census	
Other sources for companies statistics	Oral/face-to-face interview

### Sectors included in the sample:

Original Sectors (ISIC Rev. 3.1)	Manufacturing: 15 - 37 Retail: 52 Services: 45,50,51,55,60,61,62,63,64,72 100% Government owned
-------------------------------------	--

### Sample:

Comments/ problems on sectors and regions selected in the sample	
Comments on the response rate	<p>The overall response rate of establishments was low. Response rates were especially low in certain cities such as Shanghai where the response rate ultimately led to a failure to complete the desired sample. The team in Shanghai was replaced by a high-performing team from another city and encountered the same difficulty in gaining participation. Firms often refused to participate because they either did not have time, did not want to share information, were concerned about corporate espionage and other confidentiality issues despite reassurances from the World Bank offices in Washington DC and Beijing.</p> <p>In order to maximize the response rates from first preference enterprises, efforts were made to employ the strategies that not only maximize the response rates but also fitting to working well in the local context. One strategy was to recruit enumerators who were experienced in both conducting surveys and encouraging individuals and business to participate in surveys. In particular enumerators who had a good knowledge of local businesses, were more senior and formal in appearance, and had experience conversing with or</p>

	interviewing top company managers or owners. Second, target establishments were explained the indirect and long-term benefits that they receive from participating in the survey. This included, for example, a valuable contribution of data towards gaining a greater understanding of the key restrictions to private sector growth. Furthermore, that the survey results were likely to influence China's policy makers to improve business performance and employment. As an incentive for participation, all respondents were offered a China 2003 Country Profile and a free business-related book - with a choice from a list of eight popular management books in China. Participating firms were also included in which they could win a tablet computer. In the very last effort to accelerate the interview process, screeners and enumerators were sent out to secure the participation of firms who have already refused to participate or whose contacts were invalid.
Comments on the sample design:	
Other comments:	None

#### **Fieldwork:**

Date of Fieldwork	Dec 2011- February 2013
Locations	See table 1 above for details
Interview number	2848 --- See details in table 2 above
Problems found during fieldwork	Poor response rates
Other observations:	None

#### **4. Data Base Structure:**

15. The structure of the data base reflects the fact that 3 different versions of the questionnaire were used. The basic questionnaire, the Core Module, includes all common questions asked to all establishments from all sectors (manufacturing, services and IT). The second expanded variation, the Manufacturing Questionnaire, is built upon the Core Module and adds some specific questions relevant to the sector. The third expanded variation, the Services Questionnaire, is also built upon the Core Module and adds to the core specific questions relevant to either retail or IT. Each variation of the questionnaire is identified by the index variable, *a0*.

16. 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 "CN" indicate questions specific to China 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.

17. 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.

18. 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.

19. 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, 100% government owned.

20. 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.

21. 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.

22. 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. **Variable *a3x* may not be available for some datasets in order to protect the confidentiality of respondents.**

23. 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



24. 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.

25. The Annex (Excel Workbook) shows the overall estimates of the numbers of establishments based on data from NBS.

26. 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.

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

28. 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

29. 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

30. 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

31. 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.



32. Universe estimates for the number of establishments in each industry-region-size cell in China were produced for the strict, weak and median eligibility definitions.

33. 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

34. 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  $p_w$  in Stata.)<sup>5</sup>

35. 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

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

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.

36. The Annex (Excel workbook) the cell weights for registered establishments in China. **Note that when making comparisons to other countries it is important to exclude the 100% state-owned firms.**

## **7. Appropriate use of the weights**

37. 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.

38. 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 use of weighted OLS for a common population coefficient.)<sup>6</sup>

39. 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.<sup>7</sup> 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**

40. 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.

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

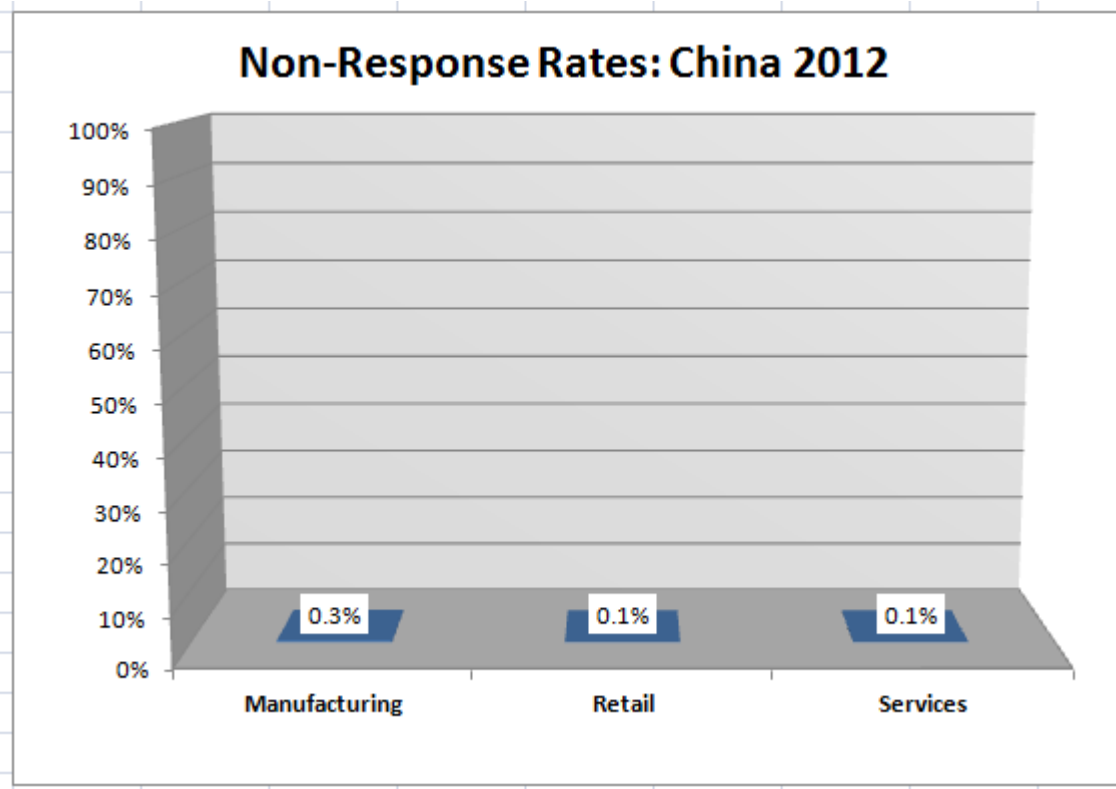
- a- For sensitive questions that may generate negative reactions from the respondent, such as corruption or tax evasion, enumerators were instructed to collect the refusal to respond as a different option from don't know (-7).

---

<sup>6</sup> 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.

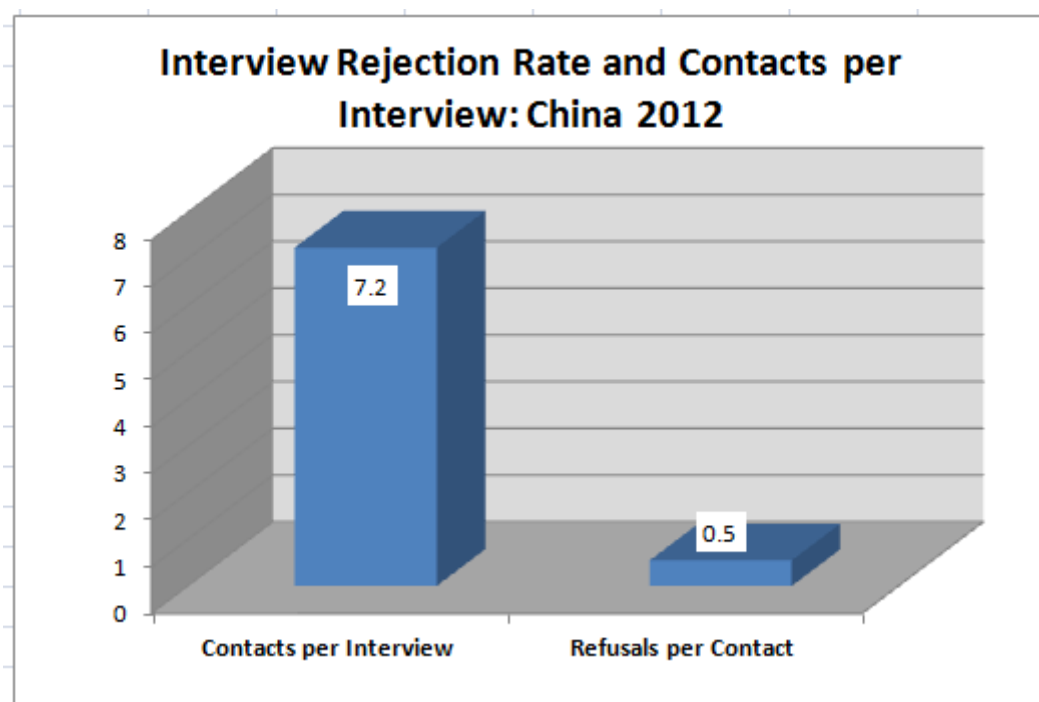
<sup>7</sup> The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

b- Establishments with incomplete information were re-contacted in order to complete this information, whenever necessary. 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 chart below reflects both categories (DKs and NAs). Incomplete information was not much of a problem in China however, refusal to participate in the survey at all was significant.



42. 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.

43. As the following graph shows, the number of contacted establishments per realized interview was 7.24. 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 0.55.



44. 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 China. All enterprise surveys suffer from these shortcomings, but in very few cases they have been made explicit.

## 9. Callbacks

45. A total of 1119 callbacks were made. The purpose of calling back respondents is to determine whether the content was accurately recorded and to verify that the interview actually took place. These call-backs have been conducted on a periodical basis, upon the receipt of each batch of completed questionnaires. The time interval between call-backs depended on the total number of questionnaires received within a certain period. The call-backs were implemented in Shanghai during the period of December 2011 – February 2013, while a smaller number of call-backs were made from Hanoi during the period of November 2011 - April 2013, both of which were conducted via telephone. Eighteen fraudulent interviews were discovered by callback during fieldwork and replacement interviews were conducted. Thirteen fraudulent interviews were discovered by callback after fieldwork had ended – these thirteen interviews were dropped from the final dataset.

46. Generally, the call-backs followed three core steps, which are presented as below:

**Firstly**, contact with the interviewed companies was made by calling the fixed-line phone number provided in the sample frame. The QC team would call the respondents enterprise on this fixed-line to determine if the respondent was working at the enterprise or not.

**Secondly**, direct contact was made with this respondent through his/her mobile phone or business telephone number, making it clear that the call was being made by a 2011 Word Bank Enterprises Survey team member. The call-back staff member would then request for further

confirmation of the name, position, and company name of the respondent. If the information proved to be consistent, a number of specific questions would be posed to them, including:

*1. Do you remember if you participated in any enterprise survey?*

[(Y/N), if interviewee answered “No”, the call-back staff member would give more details about the survey.]

*2. How long did the whole interview last?*

*3. Would you mind me asking ....*

[Asking a random selection of specific questions taken from of the questionnaires such as the year of the establishment of the enterprise, number of employees, major products, etc.]

**Lastly**, an assessment of the quality of questionnaire responses, based on specific criteria was made.

Staff members experienced certain difficulties with the call-backs themselves due to various reasons, which are listed below:

1. The interviewee refused to answer any questions. Some interviewees felt as though they were being hassled because the local firm or Sunfaith made several call-backs as a result of poor availability of the interviewee to complete the call-back interview.

2. The interviewee was too busy and found it difficult to remember all of the exact answers to the questions in the questionnaire.

3. Interviewees felt nervous when receiving a phone call from representatives of the “World Bank” and consequently, they refused to provide information.

### **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

### Questionnaires:

Problems for the understanding of questions (write question number)	No problems
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:	

### Database:

Comments on the data entry program	Data entry took place at Sunfaith's head office in Shanghai concurrently during the survey by designated data entry staff. This allowed the survey manager to identify any inconsistencies in how questions were being asked and interpreted or other errors before the survey was completed. Data was entered into a database using a data entry programme 'CSPro'
Comments on the data cleaning	The MKE statistical support team checked for the presence of data entry errors and inconsistencies. If any errors were found during data entry, the data entry team checked the information by referring to the hard copy questionnaire and contacting establishments to verify the information if necessary (with call-backs)

## Appendix B

### Status Codes:

Eligibles	1. Eligible establishment (Correct name and address)	12784
	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	75
	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	110
	4. Eligible establishment (Wrong address - the firm/establishment has changed address and the address could be found)	221
Ineligibles	5. The establishment has less than 5 permanent full time employees	38
	6. The firm discontinued businesses	374
	7. Not a business: private household	209
	8. Ineligible activity: education, agriculture, finances, governments...	37
Unobtainable	91. No reply (after having called in different days of the week and in different business hours)	2416
	92. Line out of order	639
	93. No tone	266
	94. Phone number does not exist	1371
	10. Answering machine	202
	11. Fax line - data line	192
	12. Wrong address/ moved away and could not get the new references	726
	13. Refuses to answer the screener	941
	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	14
	152. Out of target - firm moved abroad	1
	153. Out of target - Not registered with SAT	0
	<b>Total</b>	<b>20616</b>

### Response Outcomes:

<b>Sample Target</b>	<b>3000</b>
Complete interviews (Total)	2848
Incomplete interviews	10
Elegible in process	0
Refusals	10332
Out of target	658
Impossible to contact	5812
Ineligible - coop.	15
Refusal to the Screener	941
<b>Total</b>	<b>20616</b>



## **Appendix C**

### **Original Sample Design, China: (See Excel file annex for design details)**

*The original aim was to obtain 3000 interviews consisting of; 150 interviews within each of the sectors.*

*A total of 20,616 establishments were enumerated. The World Bank drew sample and issued replacements. The contacts were then issued and given preferences from 1 to 15. The implementing agency was instructed to exhaust contacts with preference-1 establishments before moving on to subsequent preferences. The World Bank tightly controlled the release of additional preferences. The implementing agency had to first demonstrate that the previous preference was exhausted before additional contacts were issued. For some cells, the contacts were exhausted before completing the sample.*

### **Completed Interviews, China: (See Excel file annex for details on completed interviews)**