

# The Venezuela 2010 Enterprise Surveys Data Set

## I. Introduction

1. This document provides additional information on the data collected in Venezuela between May 2010 and April 2011 as part of the Latin America and Caribbean (LAC) Enterprise Survey 2010, an initiative of the World Bank.

The Enterprise Surveys, through interviews with firms in the manufacturing and services sectors, capture business perceptions on the biggest obstacles to enterprise growth, the relative importance of various constraints to increasing employment and productivity, and the effects of a country's business environment on its international competitiveness. They are used to create statistically significant business environment indicators that are comparable across countries. The Enterprise Surveys are also used to build a panel of enterprise data that will make it possible to track changes in the business environment over time and allow, for example, impact assessments of reforms.

The report outlines and describes the sampling methodology, the sample structure as well as additional information that may be useful when using the data, such as information on non-response cases and the appropriate use of the weights.

## II. Sampling Structure

2. The sample for Venezuela was selected using stratified random sampling, following the methodology explained in the *Sampling Note*<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 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/locations.

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. Three levels of stratification were used in this country: industry, establishment size, and location. The original sample design with specific information of the industries and locations chosen is described in Appendix E.

4. Industry stratification was designed in the way that follows: the universe was stratified into 1 manufacturing industry, 1 service industry -retail -, and 1 residual sector as defined in the sampling manual. The manufacturing industry, service industry, and residual sectors had a target each of 120 interviews.

5. Size stratification was defined following the standardized definition for the Enterprise Surveys: 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 three locations (city and the surrounding business area): Caracas, Maracay, and Valencia.

### **III. Sampling implementation**

7. Given the stratified design, sample frames containing a complete and updated list of establishments as well as information on all stratification variables (number of employees, industry, and location) are required to draw the sample. Great efforts were made to obtain the best source for these listings. However, the quality of the sample frames was not optimal and, therefore, some adjustments were needed to correct for the presence of ineligible units. These adjustments are reflected in the weights computation (*see below*).

8. TNS Opinion was hired to implement the LAC 2010 enterprise surveys roll out. In Venezuela the local subcontractors were Corporación Request, C.A. and StatMark Group, S.A.

9. For Venezuela, two sample frames were used. The first was supplied by the World Bank and consists of enterprises interviewed in Venezuela 2006. The World Bank required that attempts should be made to re-interview establishments responding to the Venezuela 2006 survey where they were within the selected geographical locations and met eligibility criteria. That sample is referred to as the Panel. The second sample frame was produced by StatMark, using census materials from 2006, including efforts made to update frame information of un-contacted firms from the previous round of the surveys. Each database contained the following information:

- Coverage;

- Up to datedness;
- Availability of detailed stratification variables ;
- Location identifiers- address, phone number, email;
- Electronic format availability;
- Contact name(s).

Counts from sample frames are shown below.

### Panel sample counts

Region	Employees	Manufacturing	52	Other Services	Grand Total
CARACAS	5-19	102	52	39	193
	20-99	38	8	11	57
	100+	18	1	7	26
CARACAS Total		158	61	57	276
MARACAY	5-19	9	2	6	17
	20-99	10	1	1	12
	100+	6		2	8
MARACAY Total		25	3	9	37
VALENCIA	5-19	74	20	39	133
	20-99	23	6	14	43
	100+	3	1	7	11
VALENCIA Total		100	27	60	187
Grand Total		283	91	126	500

## Sample Frames

Source: Block Enumeration 2006

Region	Employees	Manufacturing	52	Other Services	Grand Total
CARACAS	5-19	102	528	353	983
	20-99	56	62	111	229
	100+	13	2	15	30
CARACAS Total		171	592	479	1,242
MARACAY	5-19	21	22	41	84
	20-99	9	5	16	30
	100+			2	2
MARACAY Total		30	27	59	116
VALENCIA	5-19	79	249	250	578
	20-99	53	22	75	150
	100+	16		11	27
VALENCIA Total		148	271	336	755
Grand Total		349	890	874	2,113

10. The two sample frames were then used for the selection of a sample with the aim of obtaining interviews with 360 establishments with five or more employees

11. The quality of the frame was assessed at the outset of the project through visits to a random subset of firms and local contractor knowledge. The sample frame was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc. In addition, the sample frame contains no telephone/fax numbers so the local contractor had to screen the contacts by visiting them. Due to response rate and ineligibility issues, additional sample had to be extracted by the World Bank in order to obtain enough eligible contacts and meet the sample targets.

12. Given the impact that non-eligible units included in the sample universe may have on the results, adjustments may be needed when computing the appropriate weights for individual observations. The percentage of confirmed non-eligible units as a proportion of the total number of sampled establishments contacted for the survey was 12.17% (112 out of 920 establishments)<sup>4</sup>.

### IV. Data Base Structure:

13. 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, retail, and other services). 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

---

<sup>4</sup> Based on out of target contacts and impossible to contact establishments

adds to the core specific questions relevant to either retail or IT. Each variation of the questionnaire is identified by the index variable, *a0*.

14. All variables are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. Variable names preceded by a prefix “*LAC*” indicate questions specific to LAC, therefore, they may not be found in the implementation of the rollout in other countries. All other suffixed variables are global and are present in all country surveys over the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” denotes that the variable is alpha-numeric.

15. There are 2 establishment identifiers, *idstd* and *id*. The first is a global unique identifier. The second is a country unique identifier. The variables *a2* (sampling location), *a6a* (sampling establishment’s size), and *a4a* (sampling sector) contain the establishment’s classification into the strata chosen for each country using information from the sample frame. The strata were defined according to the guidelines described above.

16. There are three levels of stratification: industry, size and location. Different combinations of these variables generate the strata cells for each industry/location/size combination. A distinction should be made between the variable *a4a* and *d1a2* (industry expressed as ISIC rev. 3.1 code). The former gives the establishment’s classification into one of the chosen industry-strata, whereas the latter gives the actual establishment’s industry classification (four digit code) in the sample frame.

17. All of the following variables contain information from the sampling frame. They may not coincide with the reality of individual establishments as sample frames may contain inaccurate information. The variables containing the sample frame information are included in the data set for researchers who may want to further investigate statistical features of the survey and the effect of the survey design on their results.

- a2* is the variable describing sampling locations

- 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), other manufacturing (2), retail (52), and (45, 50, 51, 55, 60, 63, 72) for other Services.

18. The surveys were implemented following a 2 stage procedure. Typically first a screener questionnaire is applied over the phone to determine eligibility and to make appointments. Then a face-to-face interview takes place with the Manager/Owner/Director of each establishment. However, the phone numbers were unavailable in the sample frame, and thus the enumerators applied the screeners in person. The variables *a4b* and *a6b* contain the industry and size of the establishment

from the screener questionnaire. Variables *a8* to *a11* contain additional information and were also collected in the screening phase.

19. Note that there are additional variables for location size by population (*a3*) and firm size by number of workers (*l1*, *l6* and *l8*) that reflect more accurately the reality of each establishment. Advanced users are advised to use these variables for analytical purposes.

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.

21. Variables *a17x* gives interviewer comments, including problems that occurred during an interview and extraordinary circumstances which could influence results. Please note that sometimes this variable is removed due to privacy issues.

## **V. Universe Estimates**

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

23. Appendix B shows the overall estimates of the numbers of establishments in Venezuela based on the sample frame.

24. 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.

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

26. Strict assumption: eligible establishments are only those for which it was possible to directly determine eligibility. The resulting weights, which include adjustments applied to panel firms (see below), are included in the variable *w\_strict\_panadj*.

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

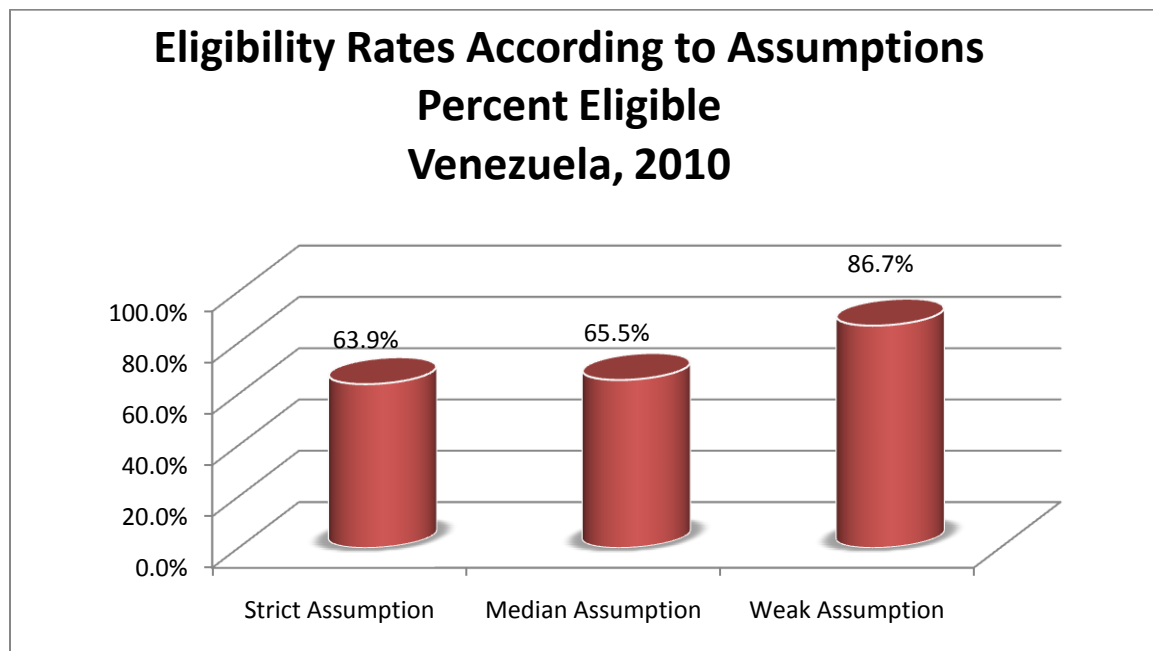
27. Median assumption: eligible establishments are those for which it was possible to directly determine eligibility and those that rejected the screener questionnaire or an answering machine or fax was the only response. The resulting weights are included in the variable *w\_median\_panadj*.

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

28. 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\_panadj*.

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

29. 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.



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

31. Once an accurate estimate of the universe cell projection was made, weights for the probability of selection were computed using the number of completed interviews for each cell.

## **VI. Weights**

32. 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.)<sup>5</sup>

33. Special care was given to the correct computation of the weights. It was imperative to accurately adjust the totals within each location/industry/size stratum to account for the presence of ineligible units (the firm discontinued business or was unattainable, education or government establishments, establishments with less than 5 employees, no reply after having called in different days of the week and in different business hours, no tone in the phone line, answering machine, fax line<sup>6</sup>, 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.

The selection of panel firms required additional adjustments to account for varying probabilities of selection between fresh and panel sample universes. For additional information on this methodology, please refer to Enterprise Survey documentation of weighting methodology.

34. Appendix C shows the cell weights for registered establishments in Venezuela.

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

35. 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.

36. 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

---

<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.

<sup>6</sup> For the surveys that implemented a screener over the phone.

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>7</sup>

37. 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>8</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.

### **VIII. Non-response**

38. 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.

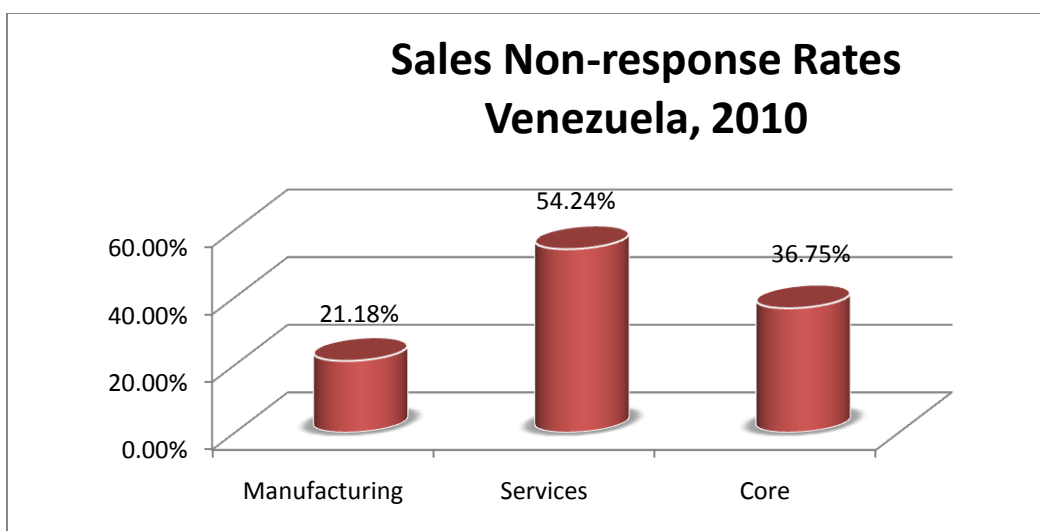
39. Item non-response was addressed by two strategies:

- 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 chart below reflects both categories (DKs and NAs).

---

<sup>7</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>8</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.

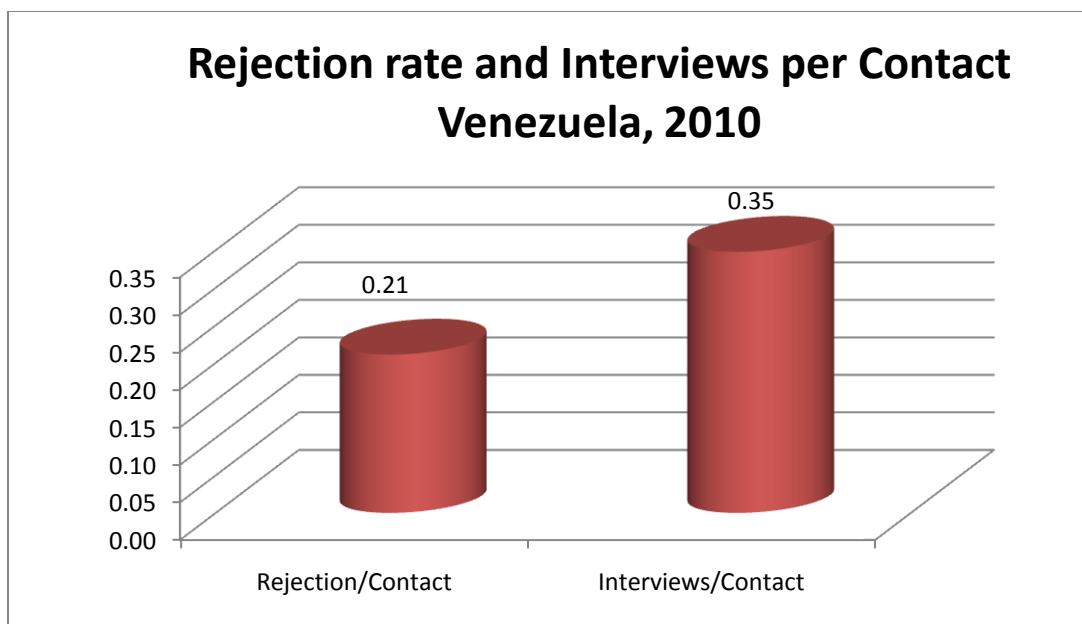


40. Survey non-response was addressed by maximizing efforts to contact establishments that were initially selected for interview. Attempts were made to contact the establishment for interview at different times/days of the week before a replacement establishment (with similar strata characteristics)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

41. As the following graph shows, the number of realized interviews per contacted establishment was 0.35<sup>9</sup>. 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 0.21.

---

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



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

#### **References:**

Cochran, William G., Sampling Techniques, 1977.

Deaton, Angus, The Analysis of Household Surveys, 1998.

Levy, Paul S. and Stanley Lemeshow, Sampling of Populations: Methods and Applications, 1999.

Lohr, Sharon L. Sampling: Design and Techniques, 1999.

Scheaffer, Richard L.; Mendenhall, W.; Lyman, R., Elementary Survey Sampling, Fifth Edition, 1996.

## Appendix A

### Status Codes Fresh:

	ELIGIBLES	
Eligible	1. Eligible establishment (Correct name and address)	304
Eligible	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	3
Eligible	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	9
Eligible	4. Eligible establishment (Moved and traced)	12
		0
Ineligible	5. The establishment has less than 5 permanent full time employees	9
Ineligible	6 The firm discontinued businesses	31
Ineligible	7. Not a business: Private household	0
Ineligible	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	22
Ineligible	151 Out of target - outside the covered regions	0
Ineligible	152. Out of target - moved abroad	0
Unobtainable	91. No reply after having called in different days of the week and in different business hours	38
Unobtainable	92. Line out of order	26
Unobtainable	93. No tone	0
Unobtainable	10. Answering machine	1
Unobtainable	11. Fax line- data line	0
Unobtainable	12. Wrong address/ moved away and could not get the new references	39
	13. Refuses to answer the screener	1
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	10
	Total	505

### Response Outcomes Panel:

Target	
Complete interviews	171
Incomplete interviews	0
Elegible in process	29
Refusals	128
Out of target	62
Impossible to contact	104
Refusal to the Screener	1

495

### Status Codes Panel:

	ELIGIBLES	
Eligible	1. Eligible establishment (Correct name and address)	244
Eligible	2. Eligible establishment (Different name but same address - the new firm/establishment bought the original firm/establishment)	0
Eligible	3. Eligible establishment (Different name but same address - the firm/establishment changed its name)	5
Eligible	4. Eligible establishment (Moved and traced)	4
Eligible	16. Panel firm - now less than five employees	7
Ineligible	5. The establishment has less than 5 permanent full time employees	0
Ineligible	616 The firm discontinued businesses - (Establishment went bankrupt)	19
Ineligible	618 The firm discontinued businesses - (Original establishment disappeared and is now a different firm)	1
Ineligible	619 The firm discontinued businesses - (Establishment was bought out by another firm)	0
Ineligible	620 The firm discontinued businesses - (It was impossible to determine for what reason)	14
Ineligible	621 The firm discontinued businesses - (Other: SPECIFY in COMMENTS)	2
Ineligible	7. Not a business: Private household	3
Ineligible	8. Ineligible activity: Education, Agriculture, Finances, Government, etc.	8
Ineligible	151 Out of target - outside the covered regions	3
Ineligible	152. Out of target - moved abroad	0
Unobtainable	91. No reply after having called in different days of the week and in different business hours	16
Unobtainable	92. Line out of order	15
Unobtainable	93. No tone	0
Unobtainable	10. Answering machine	3
Unobtainable	11. Fax line- data line	0
Unobtainable	12. Wrong address/ moved away and could not get the new references	61
	13. Refuses to answer the screener	10
	14. In process (the establishment is being called/ is being contacted - previous to ask the screener)	0
	Total	415

### Response Outcomes Panel:

Panel	
Complete interviews	149
Incomplete interviews	2
Elegible in process	54
Refusals	55
Out of target	50
Impossible to contact	95
Refusal to the Screener	10

415

## Appendix B

### Universe Estimate, Venezuela:

#### Universe

Region	Employees	Manufacturing	52	Other Services	Grand Total
CARACAS	5-19	918	4095	4103	9116
	20-99	98	408	819	1325
	100+	16	10	141	167
CARACAS Total		1032	4513	5064	10608
MARACAY	5-19	174	415	952	1541
	20-99	104	122	305	530
	100+	34	0	39	74
MARACAY Total		312	537	1296	2145
VALENCIA	5-19	418	2083	2660	5161
	20-99	166	139	378	683
	100+	35	0	68	103
VALENCIA Total		618	2222	3106	5947
Grand Total		1962	7272	9466	18699

## Appendix C

### Strict Cell Weights Venezuela :

#### Panel

Average Strict Cell Weights, Panel firms

Location	Firm Size	Manuf.	Retail	Services
Caracas	Small	4.11	1.68	1.83
	Medium	2.41	1.32	1.75
	Large	2.42	1.00	1.56
Maracay	Small	2.03		2.00
	Medium	1.36	1.00	1.00
	Large	1.36		1.00
Valencia	Small	2.62	1.71	2.34
	Medium	1.71	1.69	1.56
	Large	1.22		1.67

#### Fresh

Strict Cell Weights, Fresh Firms

Location	Firm Size	Manuf.	Retail	Services
Caracas	Small	104.90	120.68	891.36
	Medium	10.61	14.32	26.36
	Large	12.54		125.89
Maracay	Small	30.49	49.37	114.25
	Medium	62.01	82.20	23.11
	Large			16.30
Valencia	Small	114.38	117.02	207.54
	Medium	11.48	13.98	17.88
	Large	2.88		29.03

## Weak Cell Weights Venezuela:

### Panel

Average Weak Cell Weights, Panel firms

Location	Firm Size	Manuf.	Retail	Services
Caracas	Small	9.22	2.13	3.00
	Medium	3.95	1.40	2.00
	Large	3.62	1.00	1.86
Maracay	Small	4.00		3.00
	Medium	2.67	1.00	1.00
	Large	1.36		1.00
Valencia	Small	3.94	1.87	3.56
	Medium	1.92	2.20	2.03
	Large	1.51		2.00

### Fresh

Weak Cell Weights, Fresh Firms

Location	Firm Size	Manuf.	Retail	Services
Caracas	Small	171.33	183.95	1273.15
	Medium	15.26	19.18	33.14
	Large	15.77		138.45
Maracay	Small	45.79	69.51	150.04
	Medium	81.84	101.71	26.67
	Large			16.36
Valencia	Small	171.83	164.82	272.66
	Medium	15.16	17.30	20.64
	Large	3.32		29.32

## Median Cell Weights Venezuela :

### Panel

Average Median Cell Weights, Panel firms

Location	Firm Size	Manuf.	Retail	Services
Caracas	Small	4.89	1.68	1.83
	Medium	2.52	1.32	1.75
	Large	2.42	1.00	1.56
Maracay	Small	2.50		3.00
	Medium	1.36	1.00	1.00
	Large	1.36		1.00
Valencia	Small	2.68	1.79	2.34
	Medium	1.71	1.95	1.72
	Large	1.51		2.00

### Fresh

Median Cell Weights, Fresh Firms

Location	Firm Size	Manuf.	Retail	Services
Caracas	Small	104.90	121.96	891.36
	Medium	10.61	14.45	26.36
	Large	12.54		125.89
Maracay	Small	30.86	50.73	115.62
	Medium	62.66	84.33	23.35
	Large			16.36
Valencia	Small	114.59	119.02	207.92
	Medium	11.48	14.19	17.88
	Large	2.88		29.03

## Appendix D

### Strict Universe Estimates

#### Venezuela

Strict Universe Estimates (Fresh + Panel)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
Caracas	Small	561.51	2744.60	2694.16	6000.27
	Medium	85.39	308.57	613.25	1007.20
	Large	22.22	1.00	130.56	153.77
		669.11	3054.17	3437.97	7161.25
Maracay	Small	95.55	197.50	575.25	868.30
	Medium	66.08	83.20	208.96	358.25
	Large	4.08		33.60	37.68
		165.71	280.70	817.81	1264.22
Valencia	Small	270.73	1307.80	1681.37	3259.89
	Medium	125.56	102.90	277.55	506.01
	Large	29.54		63.06	92.61
Regional Total		425.83	1410.69	2021.98	3858.51
Grand Total		1260.65	4745.56	6277.76	12283.98

## Weak Universe Estimates

### Venezuela

Weak Universe Estimates (Fresh + Panel)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
Caracas	Small	939.63	4167.35	3852.44	8959.43
	Medium	127.06	411.17	770.20	1308.43
	Large	30.27	1.00	144.04	175.31
		1096.95	4579.52	4766.69	10443.17
Maracay	Small	145.37	278.05	756.18	1179.61
	Medium	89.84	102.71	241.03	433.58
	Large	4.08		33.72	37.80
		239.29	380.77	1030.93	1650.98
Valencia	Small	406.66	1835.45	2213.28	4455.39
	Medium	161.41	127.70	321.82	610.93
	Large	34.43		64.63	99.06
Regional Total		602.50	1963.16	2599.73	5165.39
Grand Total		1938.74	6923.45	8397.35	17259.54

## Median Universe Estimates

### Venezuela

Median Universe Estimates (Fresh + Panel)

Location	Firm Size	Manuf.	Retail	Services	Grand Total
Caracas	Small	568.51	2774.09	2694.16	6036.76
	Medium	86.37	311.29	613.25	1010.91
	Large	22.22	1.00	130.56	153.77
		677.09	3086.38	3437.97	7201.44
Maracay	Small	97.58	202.92	584.10	884.59
	Medium	66.73	85.33	211.13	363.19
	Large	4.08		33.72	37.80
		168.39	288.24	828.95	1285.58
Valencia	Small	272.03	1330.75	1684.39	3287.17
	Medium	125.59	105.19	278.56	509.33
	Large	30.43		64.06	94.49
Regional Total		428.04	1435.94	2027.01	3890.99
Grand Total		1273.52	4810.56	6293.93	12378.01

## Appendix E

### Original Sample Design, Venezuela:

Region	Employees	Manufacturing	52	Other Services	Grand Total
CARACAS	5-19	26	49	20	95
	20-99	22	27	28	77
	100+	4	1	4	9
CARACAS Total		52	77	52	181
MARACAY	5-19	4	3	15	22
	20-99	4	3	7	14
	100+	7		3	10
MARACAY Total		15	6	25	46
VALENCIA	5-19	20	30	15	65
	20-99	22	6	21	49
	100+	11	1	7	19
VALENCIA Total		53	37	43	133
Grand Total		120	120	120	360

### Completed Interviews, Venezuela:

Region	Employees	Manufacturing	52	Other Services	Grand Total
CARACAS	5-19	9	63	23	95
	20-99	10	18	21	49
	100+	3	3	3	9
CARACAS Total		22	84	47	153
MARACAY	5-19	5	6	10	21
	20-99	7		4	11
	100+	4		4	8
MARACAY Total		16	6	18	40
VALENCIA	5-19	17	21	25	63
	20-99	13	5	17	35
	100+	17	2	10	29
VALENCIA Total		47	28	52	127
Grand Total		85	118	117	320

## Appendix F

*Note:* Fieldwork was conducted by two agencies

### Local Agency 1 team involved in the study:

Local Agency	Corporación Request, C.A.
Enumerators involved:	11

### Local Agency 2 team involved in the study:

Local Agency	StatMark Group, S.A.
Enumerators involved:	Approx. 30 throughout the study

### Sample Frame:

Characteristic of sample frame used:	StatMark Partial Census of establishments conducted in 2006/2007
Source:	
Year:	2006
Comments on the quality of sample frame:	Many manufacturing companies have their own distribution system, with the warehouse and transportation centre next to the establishment. In several cases these warehouses were classified as a different establishment.  We found duplicates between Fresh and Panel.
Year and organism who conducted the last economic census	Not available (economic census does not exist in Venezuela)
Other sources for companies statistics	Not available

### Sectors included in the Sample:

Original Sectors	Manufacturing, Service and Retail
Added Sectors	None

**Sample:**

Comments/ problems on sectors and regions selected in the simple. Agency 1	The sample drawn at the beginning of fieldwork should have at least 4 contacts per interview to speed up data collection.
Agency 2	Valencia and Maracay are cities where it's generally more difficult to conduct studies than other areas of the country.
Comments on the response rate. Agency 1	The country's political and economical situation determines the low response rate. No branch could offer information without the main office or headquarters consent. Also, there is fear of governmental retaliation. There is also limited knowledge of the World Bank's activities or objectives.
Agency 2	High exit rates of companies in Venezuela; Information from 2006 was not updated hence a lot of establishments were no longer in business. A Lot of establishments refused to participate (also because of fear of information being passed on to the local government)
Comments on the sample design. Agency 1	It should be based on companies not establishments.
Agency 2	Initially, only allowing priority one establishments to participate caused higher efforts with a lot of rejection, there were not sufficient establishments available to be visited.
Other comments. Agency 1	None
Agency 2	None

**Fieldwork:**

Date of Fieldwork Agency 1	Maracay / Valencia 08/10/2010 – 03/16/2011 Caracas 03/04/2011 – 04/16/2011
Agency 2	May 2010 – February 2011
Problems found during fieldwork. Agency 1	Same as mentioned above
Agency 2	As mentioned above: High exit rates of establishments and high rejection rate.
Other observations.	None

Agency 1	
Agency 2	None

### Questionnaires:

Problems for the understanding of questions (write question number). Agency 1	<p>B6b. The question should specify “independently of any name change”</p> <p>JRB8. In case the respondent is not the owner how can they know which was the reason, the question should be more specific.</p> <p>C3. It should specify “with the company that supplies the energy”</p> <p>C.29 Most of the respondents said that the phone line is partially paid by the company.</p> <p>D12. Some respondents buy foreign products from national companies so they are not sure about the origin, it should be specified.</p> <p>M1. For many companies the barrier for investment in Venezuela is the exchange and price controls. So the real problem is not reflected.</p> <p>N2-N7 Questions regarding Productivity figures should be avoided. In Venezuela the companies tend to keep these data confidential.</p>
Agency 2	None
Problems found in the navigability of questionnaires. Agency 1	None
Agency 2	None
Comments on questionnaires length. Agency 1	None
Agency 2	Questionnaire is regarded as very long.
Suggestions or other comments on the questionnaire. Agency 1	None
Agency 2	None

### Database

Comments on the data	None
----------------------	------

map Agency 1 and 2	
Comments on the data processing Agency 1 and 2	None

### Country situation

General aspects of economic, political or social situation of the country that could affect the results of the survey. Agency 1	Many respondents are afraid of prices/exchange controls and expropriations; which leads to non-response or dishonest answers, especially when bribe is mentioned.
Agency 2	Venezuela has gone through a lot of changes since the last study, including activities of the tax authority (SENIAT) and the price control authority (INDEPABIS). There have been a number of expropriations as well.
Relevant country events occurred during fieldwork. Agency 1	None
Agency 2	A series of expropriations apart from that the “normal” Venezuelan circumstances such as a high crime rate, extensive controls from government authorities, restricted access to Dollars etc.
Other aspects. Agency 1	None
Agency 2	None