

# **Congo-Brazzaville Indicator Survey Data Set**

## **1. Introduction**

This document provides additional information on the data collected in Congo from 15 September 2008 to 13 February 2009 as part of the Indicator Survey, an initiative of the World Bank.

The objective of the Indicator Surveys is to obtain feedback from enterprises in client countries on the state of the private sector as well as 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.

Through interviews with firms in the manufacturing and services sectors, the Indicator Survey data provides information on the constraints to private sector growth and is used to create statistically significant business environment indicators that are comparable across countries.

The report describes the sampling design of the survey, the structure of the dataset and additional information that may be useful when using the data, including information on non-response rates, the calculation of sample weights and country-specific factors that may have affected survey implementation.

## **2. Survey Target Population**

The whole population, or the universe, covered in the Indicator Surveys is the non-agricultural economy. It comprises: all manufacturing sectors according to the ISIC Revision 3.1 group classification (group D), construction sector (group F), services sector (groups G and H), and transport, storage, and communications sector (group I). Note that this population 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.

For Congo, the sectors included in the sample by two-digit ISIC code are as follows:

Manufacturing:

15,16,17,19,20,21,22,24,25,26, 28,29,36

Services:

45,50,51,52,55,60,61,62,63,64,72

The original aim was to obtain 75 interviews with manufacturing establishments and 75 with establishments in the services sectors. Only 73 manufacturing establishments were found in the enumerated blocks. Therefore, interviewers were issued with instructions to interview as many as possible. A total of 285 eligible establishments were enumerated from the services sectors. They were divided into three preferences where available. In total 204 services establishments were issued. Interviewers were to attempt interviews with the first preferences before they could move on to subsequent preferences.

In the Indicator Surveys, the requirements for registration are defined on a country-by-country basis using the information collected by Doing Business and information from the in-country contractors. In Congo, registered firms were defined as firms that had a Tax Identification Card (carte d'identification fiscale).

### 3. Sampling for Registered Establishments

The sample for registered establishments in Congo was selected using stratified random sampling, following the methodology explained in the *Sampling Manual*.<sup>1</sup> As discussed in greater detail in the *Sampling Manual*, stratified random sampling was preferred over simple random sampling in the Indicator Surveys for several reasons:<sup>2,3</sup>

- a. To obtain unbiased estimates for different subpopulations within the economy with some known level of precision.
- b. To obtain unbiased estimates for the whole population.
- c. To ensure that the final sample includes establishments from all relevant sectors in the country 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 many cases, will be more precise than using a simple random sampling method (i.e., lower standard errors, other things being equal.)
- e. Stratification may also produce a smaller bound on the estimation errors 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 of collecting the survey data may be reduced by stratifying the population elements into convenient groupings.

Three levels of stratification were used in the Congo sample: firm sector, firm size, and geographic region.

Industry stratification was designed as follows: the universe was stratified into one manufacturing industry and one services industry (retail). The initial sample design had a target of 75 interviews each in the services and manufacturing categories, though this sample design was later adjusted to reflect the low prevalence of establishments in Congo.

Size stratification was defined following the standardized definition used for the Indicator 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.

---

<sup>1</sup> The complete text of the *Sampling Manual* 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> See Cochran, W., 1977, pp. 89; Lohr, Sharon, 1999, pp. 95.

Regional stratification was defined in terms of the geographic regions with the largest commercial presence in the country: Brazzaville and Pointe-Noire.

#### 4. Sampling implementation

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.

It was not possible to obtain a usable sample frame for Congo-Brazzaville, and so a blocks enumeration methodology was used: Detailed maps of Brazzaville and Pointe-Noire were obtained (i.e., aerial mappings were projected to a usable scale). They served as the basis of a multi-stage approach. First, each city was divided into 'blocks', and then the blocks were classified into strata defined by the predominant spatial use. The classifications used for the blocks included industrial, commercial, residential, and commercial/residential (mixed) coding. The accuracy of the classification was tested using 29 pilot blocks. When that test proved to be successful, 200 more blocks were selected and 170 of those were enumerated (building by building, floor by floor). Each separate unit was identified and classified. Further details, such as employee numbers, activity, name, and phone number, were collected for business establishments. Universe totals were projected from this enumeration of 199 total blocks. The establishments enumerated 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.

#### Universe Figures for Congo

|                    |           | Sector        |            |             |
|--------------------|-----------|---------------|------------|-------------|
| City               | Employees | Manufacturing | Services   | Grand Total |
| BRAZZAVILLE        | 5-19      | 19            | 86         | 105         |
|                    | 20-99     | 13            | 33         | 46          |
|                    | 100+      | 1             | 1          | 2           |
| BRAZZAVILLE Total  |           | 33            | 120        | 153         |
| POINTE-NOIRE       | 5-19      | 13            | 115        | 128         |
|                    | 20-99     | 24            | 47         | 71          |
|                    | 100+      | 3             | 3          | 6           |
| POINTE-NOIRE Total |           | 40            | 165        | 205         |
| <b>Grand Total</b> |           | <b>73</b>     | <b>285</b> | <b>358</b>  |

Source: Block enumeration conducted by the local agency.  
Year: 2008

The quality of the frame was assessed at the onset of the project and was not immune from the typical problems found in establishment surveys: positive rates of non-eligibility, repetition, non-existent units, etc. 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 22.74% (63 out of 277 establishments for the sample).<sup>4</sup>

## 5. Database Structure

Only one questionnaire – the Indicator Questionnaire – was used for all sectors.

All variables in the database are named using, first, the letter of each section and, second, the number of the variable within the section, i.e. *a1* denotes section A, question 1. Variable names preceded by a prefix “AF” indicate questions specific to the Africa region; these questions may not have been asked in Indicator Surveys conducted in countries in other regions. All other variables are global and are present in all country surveys conducted throughout the world. All variables are numeric with the exception of those variables with an “x” at the end of their names. The suffix “x” indicates that the variable is alpha-numeric.

The variable *idstd* uniquely identifies each establishment at the global level.

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. These variables generate the strata cells for each industry/region/size combination. 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 the sampling regions

- a6a*: coded using the definition for micro, small, medium, and large establishments as discussed 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 industries that comprise the manufacturing, services, and residual categories used in the stratification. These codes include most manufacturing industries (15 to 37), and retail, and IT for services (52, and 72 respectively). All establishments within the ‘other manufacturing’ stratum were coded with *a4a*=2.

- id*: each firm is given a unique ID number at the country level

Note that these variables may not coincide with reality for some establishments as sample frames may contain information that is later found to be inaccurate.

The surveys were implemented following a two stage procedure. In the first stage a screener questionnaire was administered over the phone to determine sampled establishment’s eligibility for the survey 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 that was collected in the screening phase.

---

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

The main questionnaire contains variables for location (*a3x*), industry (*d1a2*), and number of employees (*l1*, *l6* and *l8*) that more accurately reflect describe the characteristics of establishments than the information provided on these variables in the sample frame or the screener.

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 industry-strata based on information available from the sample frame, whereas variable *d1a2* indicates the actual ISIC code of the main output of the establishment as answered by the interviewee. This variable is probably the most accurate variable with which to classify establishments by activity.

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 on the sample frame but the actual physical location is in another place.

Variables *l1*, *l6* and *l8* provide a more accurate measure of employment and account for both permanent and temporary employment. Special efforts were made to make sure that this information was not missing for most establishments.

## 6. Universe Estimates

Special care is given to the correct computation of universe estimates and weights in the Indicator Surveys. Considering the varying quality of sample frames across countries, it is important to accurately adjust the universe totals within each region/industry/size stratum to account for the presence of ineligible units in the sampling frame.<sup>5</sup> The information collected during the screening process is used to scale down the universe estimate for each cell by the observed proportion of ineligible units within the cell.

Only one set of weights was calculated for the data set.

Appendix C shows the overall estimates of the numbers of establishments in Congo based on the block ratios.

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.

## 7. Weights

---

<sup>5</sup> For example, ineligible units could include: firms that discontinued businesses, education or government establishments, establishments with less than 5 employees in the ES sample, establishments where there was no reply after having called in different days of the week and in different business hours, the number was out of order, no tone in the phone line, answering machine, fax line, wrong address or moved away and could not obtain new contact information.

Since the sampling design was stratified and employed differential sampling of the strata, 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>6</sup>

One set of weights for each cell was computed. The set of estimates calculated proportions using the raw sample count for each cell.

Appendix D shows the cell weights for registered establishments in Congo.

## **8. Appropriate use of weights**

As discussed above, 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.

However, there is some discussion on the proper 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 Indicator 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).

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, there is no reason to use weights.

## **9. Non-response**

---

<sup>6</sup> This is equivalent to the weighted average of the estimates for each stratum, with weights equal to the population shares of each stratum.

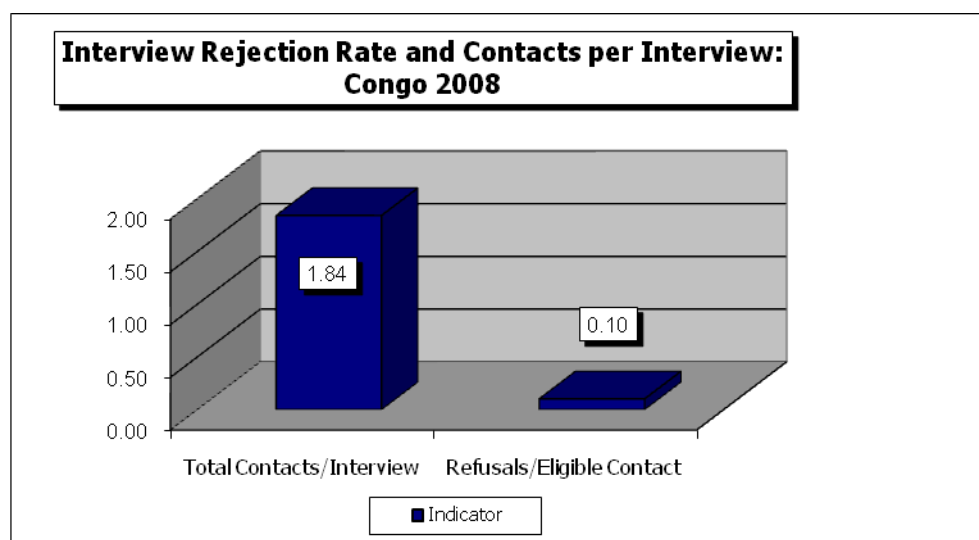
<sup>7</sup> The use weights in most model-assisted estimations using survey data is strongly recommended by the statisticians specialized on survey methodology of the JPSM of the University of Michigan and the University of Maryland.

The Indicator Surveys, along with all other surveys, suffer from both survey non-response and 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. Different strategies were used to address these issues.

Survey non-response was addressed by maximizing efforts to contact establishments that were initially sampled. When the survey frame was extracted from the sampling frame, several establishments with the same strata characteristics were randomly selected for each interview and each establishment was assigned a preference number.<sup>8</sup> Substitutions of replacement establishments were made in order to help achieve targets on the number of interviews for each stratum. Extensive efforts were made to complete interviews with each first preference establishment before contact with a replacement establishment was allowed. At least four attempts were made to contact each sampled establishment for an interview at different times/days of the week before a replacement establishment was allowed to be contacted for an interview.

Further research is needed on survey non-response in the Indicator Surveys regarding the potential introduction of bias through substitution and non-response.

As the following graph shows, the number of contacted establishments per realized interview was 1.84. 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 (e.g., establishments that closed or were in ineligible sectors). Refusal rates are also shown in the graph below. For each establishment eligible for an interview, 0.10 refused to participate.

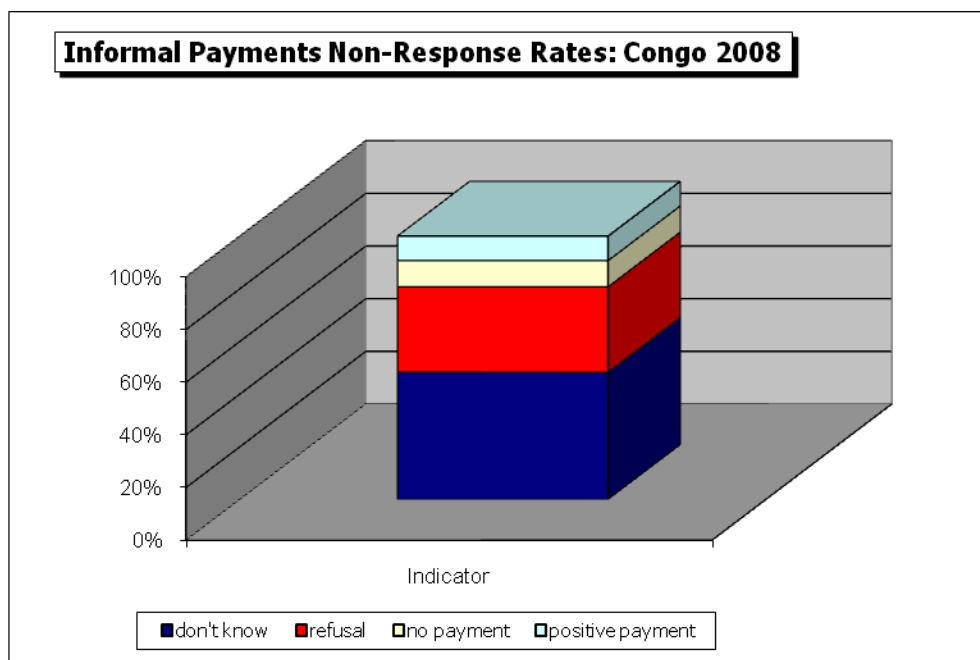


<sup>8</sup> In cases where the number of contacts initially drawn from the sample frame are insufficient to obtain an interview with the targeted number of establishments in a given strata, additional contacts for that strata may be drawn from the sampling frame. If all establishments in that strata have already been contacted and the sample target has not been reached, the sample design may be adjusted to allow additional interviews in other strata.

In completed surveys, 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 (-7) as a different option from don't know (-9).

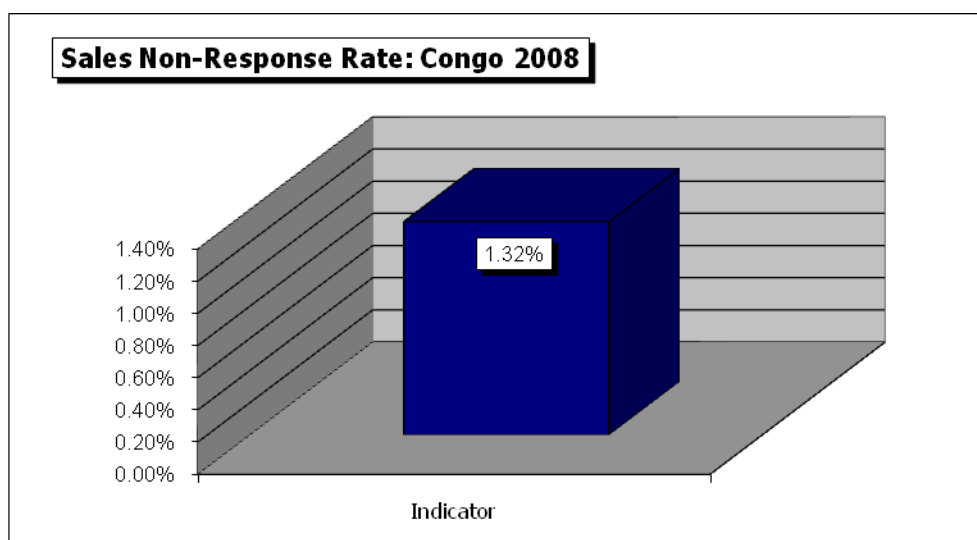
The following graph shows the breakdown of answers about the total amount of informal payments made annually (variable *j7b*) by questionnaire type.



b- Establishments with incomplete information on important questions including total sales, cost figures and employment levels were re-contacted in order to complete this information. However, re-contacts did not fully eliminate low response rates for some items.

The following graph shows non-response rates for the sales variable, *d2*, by type of questionnaire.<sup>9</sup>

<sup>9</sup> Please note that the question on total sales does not have a “refuse to answer” option, thus the non-response rates in the graph above reflect DKs and NAs as well as any missing values.



This report summarizes statistics on rejection rates, eligibility rates, and item non-response to alert researchers of these issues when using the data and when making inferences. Item non-response, selection bias, and imperfect sampling frames are not unique to Congo or the Indicator Surveys. All surveys suffer from these issues although they may not be made explicit.

## 10. Country specific comments

### Local Agency team involved in the study:

|  |  |
|--|--|
| Local Agency   | Name: MB Consulting<br>Country: Congo-Brazzaville<br>Membership of international organization: N/A<br>Activities since: 2004 |
| Name of Project Manager                                | Saturnin Wilfrid MAYENGUE  |
| Name and position of other key persons of the project: | Dassin Rhelly BOYAHOU, Research Executive<br>Maxim GOUARI, Research Executive  |
| Enumerators involved:                                  | Enumerators: 25<br>Recruiters: 4   |
| Other staff involved:                                  | Fieldwork Coordinators: 4<br>Editing: 4<br>Data Entry: 4   |

### Comments on sample frame:

|                                      |   |
|--------------------------------------|---|
| Characteristic of sample frame used: | Block Enumeration done before the survey and collected the following: <ul style="list-style-type: none"> <li>- Name of the firm</li> <li>- Addresses</li> <li>- Telephone/Fax number</li> </ul> |
|--------------------------------------|---|

|  |  |
|--|--|
|  | <ul style="list-style-type: none"> <li>- ISIC CODES</li> <li>- Main Activity</li> <li>- Region and city</li> </ul>         |
| Source:                                  | Block Enumeration (April 2008)   |
| Year:                                    | 2008   |
| Comments on the quality of sample frame: | Part of the block enumeration had to be repeated due to the wrong initial classification of the enumerated establishments. |
| Other sources for companies statistics   | All the interviews were conducted with the sample frame obtained from the blocks enumeration and no other lists were used. |

### Comments on sample:

|   |  |
|---|--|
| Comments/ problems on sectors and regions selected in the sample: | <p>On sectors:<br/>Approximately 99% of the sectors selected in the sample were correct.</p> <p>On regions: No major problems.</p> |
|---|--|

### Comments on Fieldwork:

|                                  |  |
|----------------------------------|--|
| Date of Fieldwork                | 15 October through 12 December 2008  |
| Country                          | Congo-Brazzaville  |
| Interview number                 | Manufactures: 37<br>Services: 114  |
| Problems found during fieldwork: | Collecting financial information was a challenge since most respondents were reluctant to give the complete requested information and figures. |

### Questionnaires:

|  |  |
|--|--|
| Problems for the understanding of questions (write question number)                | None   |
| Problems found in the navigability of questionnaires (for example, skip patterns). | There should be a specific questionnaire for each economic sector: for Manufacturing, for Services, etc.   |
| Comments on questionnaire length:  | The respondents complained that the questionnaire was too long.  |
| Suggestions or other comments on the questionnaire:                                | The questionnaire seemed to be badly designed. There should be a specific questionnaire for each activity sector. Sometimes the respondents were surprised when they had to answer questions related |

|  |  |
|--|--|
|  | to exportation, knowing that they deal with local trade only, as in the cases of restaurants, hotels, etc. |
|--|--|

### **Database:**

|                                     |  |
|-------------------------------------|--|
| Comments on the data entry program: | Data entry program chosen: an Excel Template was provided to Steadman Group, following their template.<br><br>Comments: The excel template matched with the questionnaire. |
| Comments on the data cleaning:      | An excel work sheet was used to check the validity of data.  |

### **Country situation:**

|  |   |
|--|---|
| General aspects of economic, political or social situation of the country that could affect the results of the survey: | Civil war and political troubles from 10 years ago and also bad management and corruption has forced many firms to close.<br><br>The low purchase power does not allow companies to have a significant turn over. For this reason some firms classified as “big” do not have an annual turn over in line with their size. |
|--|---|

### **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

### Original Sample Design

*The original aim was to obtain 75 interviews with manufacturing establishments and 75 with establishments in the services sectors.*

*Only 73 manufacturing establishments were found in the enumerated blocks. Therefore, interviewers were issued with instructions to interview as many as possible.*

*A total of 285 eligible establishments were enumerated from the services sectors. They were divided into three preferences where available. In total 204 services establishments were issued. Interviewers were to attempt interviews with the first preferences before they could move on to subsequent preferences.*

### Completed Interviews, Congo-Brazzaville:

|              | Sector        |            |            |
|--------------|---------------|------------|------------|
| Employees    | Manufacturing | Services   | Total      |
| 5-19         | 9             | 69         | 78         |
| 20-99        | 26            | 44         | 70         |
| 100+         | 2             | 1          | 3          |
| <b>Total</b> | <b>37</b>     | <b>114</b> | <b>151</b> |

## Appendix B

### Indicator Survey Sample

#### Status Codes

| ELIGIBLES    |  |            |
|--------------|--|------------|
| Eligible     | 1. Eligible establishment ( <i>Correct name and address</i> )  | 209        |
|              | 2. Eligible establishment ( <i>Different name but same address - the new firm/establishment bought the original firm/establishment</i> ) | 0          |
|              | 3. Eligible establishment ( <i>Different name but same address - the firm/establishment changed its name</i> )                           | 0          |
|              | 4. Eligible establishment ( <i>Wrong address - the firm/establishment has changed address and the address could be found</i> )           | 0          |
|              | 16. Panel firm - now less than five employees  | 1          |
| Ineligible   | 5. The establishment has less than 5 permanent full time employees   | 19         |
|              | 6. The firm discontinued businesses  | 12         |
|              | 7. Not a business: private household   | 2          |
|              | 8. Ineligible activity: education, agriculture, finances, governments...   | 4          |
| Unobtainable | 91. No reply ( <i>after having called in different days of the week and in different business hours</i> )                                | 1          |
|              | 92. Line out of order  | 2          |
|              | 93. No tone  | 0          |
|              | 10. Answering machine  | 0          |
|              | 11. Fax line - data line   | 0          |
|              | 12. Wrong address/ moved away and could not get the new references   | 13         |
|              | 13. Refuses to answer the screener   | 5          |
|              | <b>14. In process</b> ( <i>the establishment is being called/ is being contacted - previous to ask the screener</i> )                    | <b>0</b>   |
|              | 151. Out of target - outside the covered regions, firm moved abroad  | 2          |
|              | 152. Out of target - firm moved abroad   | 7          |
| <b>Total</b> |  | <b>277</b> |

#### Response Outcomes

|                                      |            |
|--------------------------------------|------------|
| Complete interviews ( <i>Total</i> ) | 151        |
| Incomplete interviews                | 0          |
| Eligible in process                  | 42         |
| Refusals                             | 17         |
| Out of target                        | 37         |
| Impossible to contact                | 16         |
| Ineligible - coop.                   | 9          |
| Refusal to the Screener              | 5          |
| <b>Total</b>                         | <b>277</b> |

## Appendix C

### Universe Estimates, Congo-Brazzaville:

|              | Sector        |              |              |
|--------------|---------------|--------------|--------------|
| Employees    | Manufacturing | Services     | Total        |
| 5-19         | 76            | 1,218        | 1,294        |
| 20-99        | 155           | 392          | 547          |
| 100+         | 8             | 6            | 13           |
| <b>Total</b> | <b>238</b>    | <b>1,616</b> | <b>1,854</b> |

## Appendix D

### Cell Weights Congo-Brazzaville:

|                  | <b>Sector</b>        |                 |
|------------------|----------------------|-----------------|
| <b>Employees</b> | <b>Manufacturing</b> | <b>Services</b> |
| 5-19             | 8                    | 18              |
| 20-99            | 6                    | 9               |
| 100+             | 4                    | 6               |