

## **Sampling Procedure Cambodia ICA**

### *Available lists of firms*

The following lists of firms are available:

1. PBES (March-April 2006): census of 41,775 firms with at most 50 workers in the provincial capital and the next 3 most populous areas in the 10 economically most important provinces (note: it actually also includes 58 firms with more than 50 workers). The PEBS contains information on the name, address, phone number, location, sector, and number of regular workers of the firm.
2. Customs Department (updated June 2007): list of 1,403 firms that have registered with customs. The Customs Department list contains information on the name, address, phone number, and location of the firm. Information on the sector and number of regular workers is missing.
3. Ministry of Commerce (1998-2007): list of 14,432 firms that have formally registered with the Department of Commerce in the period 1998-2007 in Cambodia. (Our understanding is that it was not possible for firms to register with MoC before 1998.) The Department of Commerce list contains information on the name, address, location of the firm and the date of registration. There is also information on the sectors, but this information appears often uninformative. Information on the phone number and number of regular workers is missing.
4. Ministry of Commerce (2006): list of 985 registered firms which have provided a declaration profile for 2006. Registered firms are required to file a declaration profile each year by March 15<sup>th</sup>, and this profile is added to the company file and also entered in an electronic database. This list contains information on the name, registration code, address and date the profile was submitted. Additional information on sector and main contacts within the company are included on the declaration profile and may be available (though not electronically).
5. Ministry of Commerce (2006): list of approximately 2500 firms in Phnom Penh registered with a business permit in 2006. Apart from formal registration (which allows a firm to do import/export transactions), a firm can also opt for a one-year renewable business permit. Formally registered companies have to pay income tax, while firms with a business permit only pay for the business permit. Also a formal procedure is required to close down a formally registered firm (because of tax liabilities), while a firm with a business permit can simply stop operations (without renewing the permit for another year). The business permit can be obtained at the municipal/governor office or the local department of the Ministry of Commerce and the list of firms with business permits is kept locally. The electronic database we have been given includes firms that registered for permits in 2006 and provides the name of the business, address and sector, as well as the owner's name and year of birth.
6. GMAC (June 2007): list of 404 garments firms of which 302 are active members. The GMAC list contains information on the name, address, phone number, location, sector, and number of regular workers of the firm. GMAC estimates that the member list includes almost 100% of garment manufacturers that are exporting. They further estimate that there are 50 – 60 garment manufacturers that are not members. These firms are not considered part of the “formal” garment sector by GMAC because they are not exporting (they are usually subcontracted by “formal” firms that are exporting).

7. ICA (2003): list of 445 firms interviewed in the Investment Climate Assessment Survey of 2003, except for 19 firms interviewed in Phnom Penh, 3 in Kampong Cham, and 38 (all) firms in Battambang, and 2 extra in Siem Reap. The ICA list contains information on the name, address, location and type of company (sector), and phone numbers are available for most firms. The contact sheet does not include information on the size of the firm (i.e., the number of workers), but it should be possible to identify this information in the ICA database (which does not include the name and address of the firms).

### *Sampling frame*

In terms of sampling frame, the PEBS provides a census of firms with at most 50 workers in the most important economic areas in Cambodia. Therefore, it does not include any of the firms with more than 50 workers (except for 58 firms somehow included). However, it is plausible that all firms with more than 50 employees are formally registered, and therefore included in the list of formally registered firms with the Department of Commerce. One complication could be that some firms with at least 50 workers have only a business permit, but it is not possible to determine this with the current files we have.

We therefore assume that the combination of the ISC 2003, PBES and the Department of Commerce lists (firms registered by April 2006, and the profile list for 2006) provides us with a complete sampling frame.

However, this sampling frame has two important shortcomings. First, because registered firms with 50 or fewer employees are included in both the PBES and the Department of Commerce list of formally registered firms, the sampling frame will consist of many duplicates. Second, because the Department of Commerce list does not include (good information on) the sector and number of workers, the sampling frame does not provide the population shares across sector and firm size. These population shares across sector and firm size are crucial, however, for rebalancing (post-stratifying) the sample of panel firms from the ICA 2003. Although the ICA 2003 was based on a random sample of firms from the IRL company and SME database (as well as from business association membership lists and the WBG list of electric power providers), the subsequent finding of a far larger number of firms in the PBES suggests that this list was significantly incomplete. In addition, further biases in the sample of panel firms can be expected if there will be significant and non-random attrition from the ICA 2003 sample.

With respect to the first shortcoming, it should be possible to inspect the duplicates in the combined PBES and Department of Commerce list of formally registered firms and eliminate them. This is, however, an arduous task, given that the same firm may be listed under (slightly) different names in both lists, and given that the total number of firms in the combined lists exceeds 17,000 firms. Moreover, even if all duplicates would be removed, the second shortcoming of missing information on sector and number of workers still remains.

In principle cluster sampling would be a possible solution to the second shortcoming. First a number of locations (clusters) can be randomly selected (such as neighborhoods in urban areas). Second, within each randomly selected cluster, all firms from the Department of Commerce list (but not in the PBES list) can be contacted to provide information on the sector and number of workers. Third, a random sample (possibly stratified) of firms is selected in each cluster. Given that each cluster has been selected randomly, the combined list across all selected clusters will be representative of the sampling frame, but with the added advantage of having information on sector and number of workers for *all* firms.

However, cluster sampling requires that all firms in each selected cluster are contacted to collect information on sector and number of workers, while only a sample of contacted firms is actually interviewed. This is problematic, because it is often relatively costly to locate and contact firms and to convince them to provide information. Therefore it is preferable to include any firm in the sample that already has been located, contacted and found to be cooperative. But in that case it is also preferable to do random sampling rather than cluster sampling, as random sampling is more precise in the presence of intra-cluster correlation (see Deaton 1997).

*Sampling procedure*

Based on the above considerations, we therefore apply simple random sampling to *estimate* the population shares across sectors and firm sizes but *correcting* for duplicates. These estimated population shares can be used to post-stratify the final sample, which will be a combination of panel firms from the ICA 2003 and newly selected firms. Post-stratification will be done with respect to sector and firm size, and corrects for the sampling bias in the ICA 2003 and the attrition bias in the panel data.

Steps in the sampling procedure:

1. combine lists to create a sampling frame (including ICS 2003, PBES, and MoC list of firms registered between 1998 and April 2006).
2. exclude firms outside 5 urban areas and/or with less than 5 regular workers
3. draw random sample.
4. check for duplicates. If duplicate, then give weight equal to 1/#duplicates
5. if firm is not in PBES, collect data on sector and number of regular workers at time of PBES (April 2006), and exclude firm if outside 5 urban areas and/or with less than 5 regular workers
6. if firm started after April 2006, then exclude firm
7. estimate the population shares across strata. Strata are defined by 4 sectors as well as 3 firm sizes.
8. interview the panel firms and new firms selected from the random sample such that 125 firms are interviewed in each sector. Also to have adequate coverage across firm size, one should oversample large firms within each strata such that the number of firms within each firm size class is at least 120. Furthermore, the number of exporters should be at least 120, and if necessary strata with a higher proportion of exporters should be oversampled. The following distribution of the sample is therefore suggested:

Sector \ Size	Small (5-19 employees)	Medium (20-99 employees)	Large (100+ employees)	Total
1	≥30	≥30	≥30	125
2	≥30	≥30	≥30	125
3	≥30	≥30	≥30	125
4	≥30	≥30	≥30	125
Total	≥120	≥120	≥120	500

9. estimate post-stratification weights

Notes:

- a. The population shares are estimated for all randomly selected firms for which information on sector and number of regular workers at time of PBES are available, *even* if the firm refuses to cooperate subsequently with the survey.
- b. The selected sectors are (1) garments, (2) retail/wholesale trade, (3) tourism (hotel and restaurants, other services such as travel agencies, tour operators), and (4) other.
- c. In order to have enough large firms in the random sample, disproportionate sampling across firm size can be applied. However, if firm size is only known for PBES sample, this is not more efficient given that sector and firm size information are already known for these firms.
- d. In case random sampling of new firms does not provide for enough firms in some of the sector/firm size strata (less than 30 firms in total), then a lower number of firms can be targeted (but only after additional effort is made to increase the response rate within these strata).
- e. Given the small sample size of the survey, post-stratification can be done by sector or by firm size (as the population shares are estimated with sufficient precision only for sector and for firm size), but not for any of these variables combined. This implies that the post-stratification weights are different for analyses across sectors and across firm sizes.
- f. For each type of post-stratification there will be three sets of post-stratification weights, namely for the sample of panel firms, for the sample of new firms, and for the total sample.
- g. It is no problem to prioritize garment firms from the start, as garments will form a stratum in the final sample. However, it will be important to have enough garment firms of different firm sizes.
- h. It is useful to monitor the sectoral distribution of the randomly drawn firms – in particular, it is important to check that garments, trade, and tourism are indeed the largest sectors.