FINSCOPE SMALL BUSINESS SURVEY 2006 - SAMPLING METHODOLOGY
Compiled by Prof D J Stoker and Dr Clive Corder

A "Mixed Household and Enterprise Survey" sampling approach outlined in the brief was used. Conceptually a representative primary sample of adults is obtained and from this the people who qualify as small businesses are identified. Adults identified in the primary universe are grossed up to population estimates and the estimate of the number of small businesses is then derived from this.

Universe
The universe was defined as adults aged 16 years and over living in dwellings in Gauteng, but excluding live-in domestics and gardeners (see also below).

Selecting EAs – Compiled and done by Prof D J Stoker
Target area: The Gauteng province.
Primary sampling unit (PSU): The census 2001 enumerator area (EA).
Sampling frame: An independently constructed census 2001 EA sampling frame as discussed in the attachment below, updated to the StatsSA released 2005 midyear estimates (cf. attachment).
Explicit strata: The final sampling frame used for the drawing of the sample of PSU's (i.e. EA's) was obtained from the above general sampling frame by considering only EA's in Gauteng and then deleting the EA types recreational areas, industrial areas and special institutions (such as prisons, hospitals, old age homes and military camps) as well as all EA's with less than 25 households (with the view not to draw very small EA's). The remaining 11836 EA's were explicitly stratified into three strata, viz. farms, small holdings and the remainder (urban formal, urban informal and hostels). The following numbers of EA's were drawn: 22 (farms), 36 (small holdings) and 542 (remainder areas). In total 600 EA's were drawn. Within each explicit stratum the EA's were ordered according to the sampling frame variables municipality code, main place code, geography type (basically urban formal and urban informal), sub place code and EA-number. The required number of EA's were then drawn independently per explicit stratum using pps (with probability proportional to size) systematic sampling with the number of households in the EA as given on the sampling frame as MOS (measure of size).

Selecting dwellings in each EA
The sample design was for 10 dwellings to be systematically selected in each of the 600 EAs without substitution. In order to allow for possible non-contact the number of dwellings in each EA was increased to 12.

Maps of each EA were supplied by the Human Sciences Research Council (HSRC). Interviewers were instructed to count all the dwellings or occupied stands/plots. Interviewers where then requested to divided the total number of identified dwellings by 12 in order to obtain a sampling interval. After a random start, which formed the first sample point, interviewers identified subsequent points by taking a route through the EA and counting dwellings as per the sampling interval.

Identifying households
At each sample point interviewers first of all established the number of families who arranged their own food, and lived in the household or backyard, but excluding domestic workers and live-in gardeners. In the event that there was more than one household then a listing was made of all the households, with the main one being number 1. A random selection grid was then used to identify the household to be sampled.

Identifying adults in household and SMEs
At the selected household an interview was sought with the head of the household or spouse. The household was further defined as an individual or group of individuals who occupy a common dwelling, or backyard, for at least 4 nights in a week.

Information about each adult member aged 16 years and over, was obtained with regard to their population group and employment status. Those who were working for themselves, or running their own business for a living, were identified as qualifying for the study of Small Businesses, provided that they did not employ more than 200 people.

Comment on differences between survey results and initial proposal
In the survey a total of 5,039 (70%) households was contacted out of the potential of 7,200. In these households a total of 12,422 adults (2.5 per household) were listed, of whom 2,001 (16%) qualified as being self-employed or running their own business.

In the original proposal for the study it was anticipated that with a contact sample of 10 dwellings in each of 600 EAs a total of 16,200 adults would be identified (2.7 per household) and that 12% of these would be self-employed. This was based on the incidence of AMPS of 11.9% self-employed. It was thus estimated that in the region of 2000s SMEs would be found. It was however pointed out that with more detailed probing of informal sector activities it was likely that this proportion could be higher. In the actual survey the higher incidence of qualifiers compensated for the lower number of households contacted and a sample of 2,001 SMEs was achieved.

Weighting – Compiled and done by Prof D J Stoker
The EA weight and respondent weight were calculated in each stratum as follows:

EA-weight = sum of the HH-totals for all EA's on the sampling frame which fall within the stratum divided by the product of the HH-total of the EA and the number of EA's drawn (i.e. realised) in the stratum.

Respondent-weight = EA-weight multiplied by the total number of persons 16 years and older in the EA divided by the number of respondents drawn (i.e. realised) in the EA. (The latter quotient was used due to the lack of any feedback information on the number of persons 16 years and older in the drawn HH's from which the respondents were randomly selected.)

Benchmarking: The respondents weight were finally marginally benchmarked in such a way that the race totals, the age group totals and the gender totals were equal to the midyear estimates as released by StatsSA for the year 2005.