APPENDIX B.
Accuracy of the Sample Data

CONTENTS

INTRODUCTION

The tabulations in this report are based on a survey sample or a subset of the population of the Virgin Islands. The data are therefore estimates of the actual figures that would have been obtained from a 100-percent enumeration. Estimates derived from this sample are expected to be different from 100-percent figures because they are subject to sampling and nonsampling errors. Sampling error in data arises from the selection of persons and housing units to be included in the sample. Nonsampling error also affects survey data. Both types of error are discussed below.

SAMPLE DESIGN

Census Bureau geography divides each island in the US Virgin Islands into non-overlapping regions called Block Numbering Areas (BNAs). Each BNA is subdivided into smaller areas called Block Groups (BGs), and each BG is subdivided into even smaller areas called blocks. A two-stage sample design was used for this survey where BGs constitute the Primary Sampling Units (PSUs). The design called for the selection of an area sample in two stages, using the method of paired selection of PSUs with probabilities proportional to size measures (PPS). All the BGs were ordered geographically and then divided into 40 incipient strata. A pair of BGs was randomly selected from each stratum using systematic sampling.

It was decided that the sample values should be within two percentage points of the corresponding population values with a confidence level of 95 percent. The sample that was required to achieve estimates which do not exceed this maximum error due to sampling error only and without the adjustment for finite population sampling was 2,401.

The primary sampling unit for the 1995 survey was the housing unit, including all occupants. After adjusting for refusals, vacancies, and other situations that result in non-interviews, the sample size increased to 4,078. From prior surveys conducted in the USVI, it was estimated that both the completed interview rate and the coverage rate were about 85 percent. Also, the 1990 Census showed that approximately 18.5 percent of the households in the USVI were vacant. This figure was used as an estimate for the vacancy rate in calculating the final sample size.

Estimation of the tabulated values was derived by first applying weights to the sample values. The weight for each island is: St. Croix, 9.71796; St. John, 15.20741; and St. Thomas, 13.46035.

CONFIDENTIALITY OF THE DATA

Every effort has been made to protect the confidentiality of the 1995 U.S. Virgin Islands Survey data, and to make sure that published data do not disclose information about specific individuals, households, or housing units. All full-time or part-time employees of the University who handle the questionnaires are

required to swear to the same oath of confidentiality that is administered by the Bureau of the Census in the collection of census data.

EDITING OF UNACCEPTABLE DATA

The objective of the processing operation was to produce a set of data that describes the population as clearly and accurately as possible. To meet this objective, questionnaires were reviewed and edited during field data collection operations by field supervisors for consistency, completeness, and acceptability. Questionnaires were also reviewed by edit clerks in the Survey office for omissions, certain inconsistencies, and improper sample selection. For example, write-in entries such as "Don't know" or "NA" were considered unacceptable in certain quantities and/or in conjunction with other data omissions.

As a result of this review operation, a follow-up telephone or personal visit was made to obtain missing information. Potential coverage errors were included in the follow-up, as well as questionnaires with omissions or inconsistencies beyond the completeness and quality tolerances specified in the review procedures.

Subsequent to field operations, remaining incomplete or inconsistent information on the questionnaires was assigned using imputation procedures during the final automated edit of the collected data. Allocations—computer assignments of acceptable data in place of unacceptable entries or blanks—were needed most often when an entry for a given item was lacking, or when the information reported for a person or housing unit on that item was inconsistent with other information for that same person or housing unit. As in previous surveys, the general procedure for changing unacceptable entries was to assign an entry for a person or housing unit that was consistent with entries for persons or housing units with similar characteristics. The assignment of acceptable data in place of blanks or unacceptable entries enhanced the usefulness of the data.

Another way to make corrections during the computer editing process is substitution. Substitution is the assignment of a full set of characteristics for a person or housing unit. Because of the detailed field operations, substitution was not needed for the 1995 Survey.

SOURCES OF ERRORS

In any large-scale statistical operation, such as the 1995 U.S. Virgin Islands Population and Housing Survey, human- and machine-related errors do occur. These errors are commonly referred to as nonsampling errors. Such errors include not interviewing every household or every person in the survey, not obtaining all required information from the respondents, obtaining incorrect or inconsistent information, and recording information incorrectly. In addition, errors can occur during the field review of the interviewer's work, during clerical handling of the survey questionnaires, or during the electronic processing of the questionnaires.

To reduce various types of nonsampling errors, a number of techniques were implemented during the planning, data collection, and data processing activities. Quality assurance methods were used throughout the data collection and processing phases of the survey to improve the quality of the data.