Annual Report *of the* Center for Statistical Research and Methodology
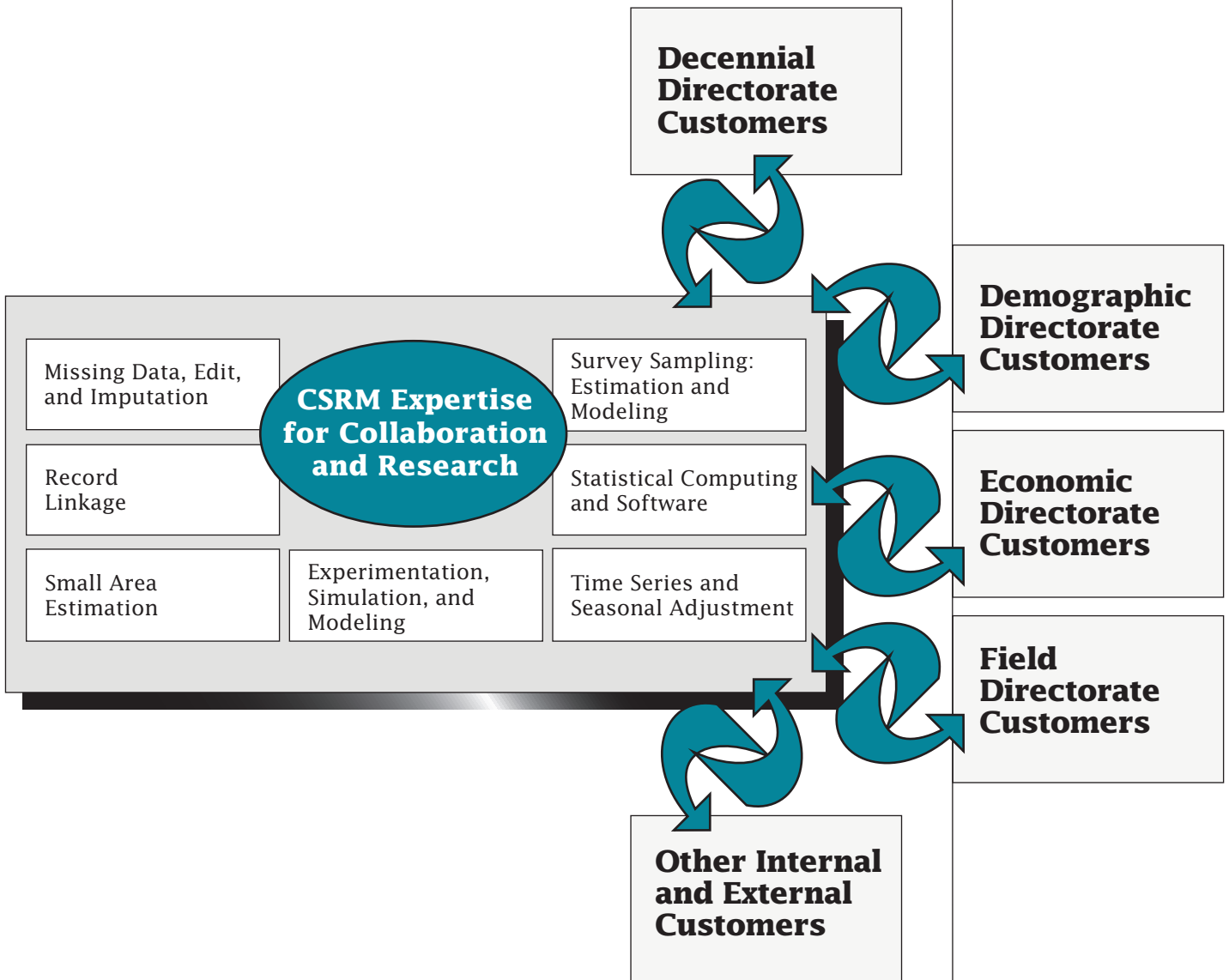
Research and Methodology Directorate

Fiscal Year 2011

Decennial Directorate Customers

Demographic Directorate Customers

Economic Directorate Customers

Field Directorate Customers

Other Internal and External Customers

CSRM Expertise for Collaboration and Research

Missing Data, Edit, and Imputation

Record Linkage

Small Area Estimation

Experimentation, Simulation, and Modeling

Survey Sampling: Estimation and Modeling

Statistical Computing and Software

Time Series and Seasonal Adjustment

Since August 1, 1933—

"… As the major figures from the American Statistical Association (ASA), Social Science Research Council, and new Roosevelt academic advisors discussed the statistical needs of the nation in the spring of 1933, it became clear that the new programs—in particular the National Recovery Administration—would require substantial amounts of data and coordination among statistical programs. Thus in June of 1933, the ASA and the Social Science Research Council officially created the Committee on Government Statistics and Information Services (COGSIS) to serve the statistical needs of the Agriculture, Commerce, Labor, and Interior departments … COGSIS set … goals in the field of federal statistics … (It) wanted new statistical programs—for example, to measure unemployment and address the needs of the unemployed … (It) wanted a coordinating agency to oversee all statistical programs, and (it) wanted to see statistical research and experimentation organized within the federal government … In August 1933 Stuart A. Rice, President of the ASA and acting chair of COGSIS, … (became) assistant director of the (Census) Bureau. Joseph Hill (who had been at the Census Bureau since 1900 and who provided the concepts and early theory for what is now the methodology for apportioning the seats in the U.S. House of Representatives) … became the head of the new Division of Statistical Research … Hill could use his considerable expertise to achieve (a) COGSIS goal: the creation of a research arm within the Bureau …"

Source: Anderson, M. (1988), *The American Census: A Social History,* New Haven: Yale University Press.

Among others and since August 1, 1933, the Statistical Research Division has been a key catalyst for improvements in census taking and sample survey methodology through research at the U.S. Census Bureau. The introduction of major themes for some of this methodological research and development where staff of the Statistical Research Division[1] played significant roles began roughly as noted—

- **Early Years (1933–1960s):** *sampling (measurement of unemployment and 1940 Census); probability sampling theory; nonsampling error research; computing; and data capture.*

- **1960s–1980s:** *self-enumeration; social and behavioral sciences (questionnaire design, measurement error, interviewer selection and training, nonresponse, etc.); undercount measurement, especially at small levels of geography; time series; and seasonal adjustment.*

- **1980s–Early 1990s:** *undercount measurement and adjustment; ethnography; record linkage; and confidentiality and disclosure avoidance.*

- **Mid 1990s–Present:** *small area estimation; missing data and imputation; usability (human-computer interaction); and linguistics, languages, and translations.*

At the beginning of FY 2011, most of the Statistical Research Division became known as the Center for Statistical Research and Methodology. In particular, with the establishment of the Research and Methodology Directorate, the Center for Survey Measurement and the Center for Disclosure Avoidance Research were separated from the Statistical Research Division, and the remaining unit's name became the Center for Statistical Research and Methodology.

---

[1]The Research Center for Measurement Methods joined the Statistical Research Division in 1980. In addition to a strong interest in sampling and estimation methodology, research largely carried out by mathematical statisticians, the division also has a long tradition of nonsampling error research, largely led by social scientists. Until the late 1970s, research in this domain (e.g., questionnaire design, measurement error, interviewer selection and training, nonresponse, etc.) was carried out in the division's Response Research Staff. Around 1979 this staff split off from the division and became the Center for Human Factors Research. The new center underwent two name changes—first, to the Center for Social Science Research in 1980, and then, in 1983, to the Center for Survey Methods Research before rejoining the division in 1994.

U.S. Census Bureau
Center for Statistical Research and Methodology
Room 5K108
4600 Silver Hill Road
Washington, DC 20233
301-763-1702

We help the Census Bureau improve its processes and products. For fiscal year 2011, this report is an accounting of our work and our results.

Center for Statistical Research & Methodology

Highlights of What We Did...

As a technical resource for the Census Bureau, each researcher in our center is asked to do three things: *collaboration/consulting*, *research*, and *professional activities and development*. We serve as members on teams for a variety of projects and/or subprojects.

Highlights of a selected sampling of the many activities and results in which the Center for Statistical Research and Methodology staff members made contributions during FY 2011 follow, and more details are provided within subsequent pages of this report:

- *Missing Data, Edit, and Imputation*: (1) delivered synthetic data for each state and Puerto Rico to DSCMO; (2) developed new imputation methods for longitudinal data with nonmonotone patterns of nonresponse; (3) worked with Director's Office in examining discrete-time hazard models with time-dependent covariates for the evaluation of field productivity through propensity scoring in the Current Population Survey; and (4) worked with SEHSD to design a simulation for the evaluation of three methods for SIPP earnings imputation.

- *Record Linkage*: (1) assisted in the application of the basic *SRD Matching* software and *BIGMATCH* for several 2010 Census matching follow-up studies; and (2) generalized *DISCRETE* Modeling/Edit/Imputation software (October 2010 version) for general modeling for analyses and production edit/imputation.

- *Small Area Estimation*: (1) developed a small area model for design-based sampling variance of poverty rates for the ACS; (2) completed report on a partition model for non-ignorable data and comparison to existing models; and (3) completed report on using random effect models to protect against model error in logistic regression models for census coverage.

- *Survey Sampling-Estimation and Modeling*: (1) carried out two evaluations of imputation and weighting methods developed to improve small area group quarters estimates in the ACS; (2) developed small area model and software for the *Voting Rights Act-Section 203* special tabulation—estimating U.S. citizenship, limited-English proficiency, and illiteracy for language minority groups; (3) used a logit model approach to determine person and housing unit characteristics related to differential response between voluntary and mandatory treatments in the ACS Voluntary Test; (4) refined previous simulation methodology to investigate the properties of two methods (M-estimation and Clarke Winsorization) which detect and treat influential values; (5) built on previous development of models to account for errors in the Census Coverage Measurement Program; and (6) investigated ranking methods based on sample survey estimates, including the ACS ranking tables.

- *Statistical Computing and Software*: (1) continued development and support of software by CSRM researchers; (2) continued development of TEA, a general purpose R package for survey processing, focusing on editing and imputation; (3) applied TEA to produce synthetic data for ACS and Decennial Census Group Quarters; and (4) demonstrated, by writing software in Python, the capability to do basic Web scraping.

- *Time Series and Seasonal Adjustment*: (1) revised and updated the software program *X-13ARIMA-SEATS* for seasonal adjustment; (2) enhanced the uSim-X13 software package by the addition of a nonparametric modeling interface featuring new methods in AM-FM empirical mode decomposition for the extraction and analysis of signals in time series; (3) derived, encoded, and tested model estimation and forecasting algorithms; and (4) implemented a fully Bayesian seasonal adjustment procedure involving long-range dependent seasonal and trend components.

How Did We[1] Do...

For the 13th year, we received feedback from our sponsors. Near the end of fiscal year 2011, our efforts on 24 of our program (Decennial, Demographic, Economic, External) sponsored projects/subprojects with substantial activity and progress and sponsor feedback (Appendix A) were measured by use of a Project Performance Measurement Questionnaire (Appendix B). Responses to all 24 questionnaires were obtained with the following results (The graph associated with each measure shows the performance measure over the last 13 fiscal years):

Measure 1.   *Overall, Work Met Expectations*

Percent of FY2011 Program Sponsored Projects/Subprojects where sponsors reported that overall work met their expectations (agree or strongly agree) (24 out of 24) …….………..... 100%
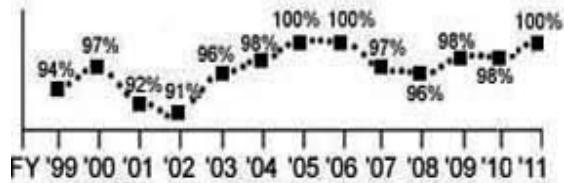

Measure 2.   *Established Major Deadlines Met*

Percent of FY2011 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met (17 out of 17 responses) ……………………... 100%
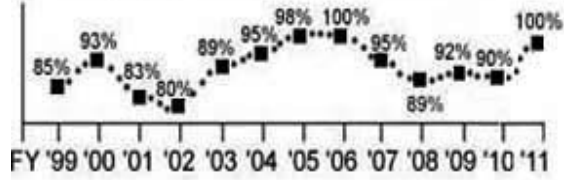

Measure 3a.   *At Least One Improved Method, Developed Technique , Solution, or New Insight*

Percent of FY2011 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight (22 out of 24 responses) ………… 92%
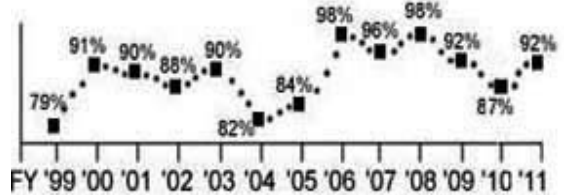

Measure 3b.   *Plans for Implementation*

Of these FY2011 Program Sponsored Projects/Subprojects reporting at least one improved method, technique developed, solution, or new insight, the percent with plans for implementation (13 out of 20 responses) ………………….. 65%
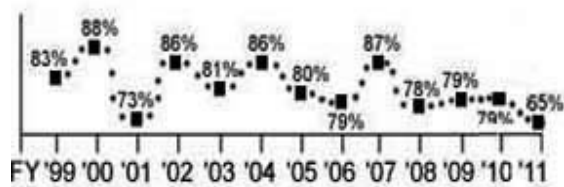
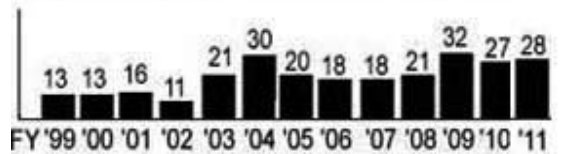*Measure 4.*   *Predict Cost Efficiencies*

Number of FY2011 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency" ………..…… 5
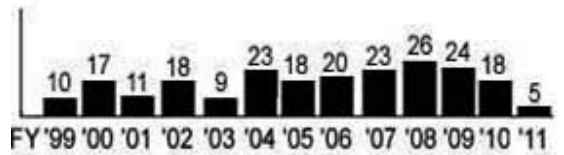
From Section 3 of this ANNUAL REPORT, we also have:

Measure 5.   *Journal Articles, Publications*

Number of peer reviewed journal publications documenting research that appeared (18) or were accepted (10) in FY2011 ……………………………………………………………... 28

Measure 6.   *Proceedings, Publications*

Number of proceedings publications documenting research that appeared in FY2011 ……………………………………..…… 8
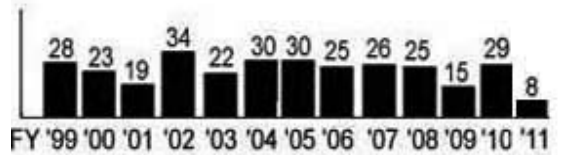
Measure 7.   *Center Research Reports/Studies, Publications*

Number of center research reports/studies publications documenting research that appeared in FY2011 ……………… 8
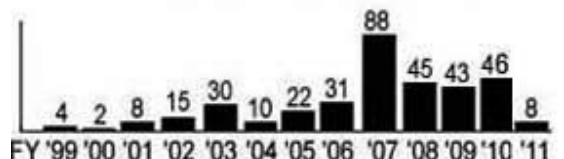
Each completed questionnaire is shared with appropriate staff to help improve our future efforts.

TABLE OF CONTENTS

1. COLLABORATION

1.1 STATISTICAL DESIGN AND ESTIMATION
(Decennial Project 5610102)

A. Decennial Record Linkage

Description: Under this project, our staff will provide advice, develop computer matching systems, and develop and perform analytic methods for adjusting statistical analyses for computer matching error, with a decennial focus.

Highlights: See Projects 0351000 and 1871000 (A), General Research – Statistical Computing Methodology.

Staff: William Winkler (x34729), William Yancey, Ned Porter

B. Census Unduplication Research

Description: The goal of this project is to conduct research to guide the development and assessment of methods for conducting nationwide matching and unduplication in the 2010 Decennial Census and future Censuses. One of the major problems is how to incorporate the effects of name frequency into the unduplication procedures. Our staff also provides assistance in specifying and reviewing output from the matching and unduplication procedures for test Censuses and for Census 2010. We began this project in May of 2004.

Highlights: In FY 2011, staff attended sessions to review data and to set cutoffs for within-response links, i.e., multiple links between a pair of housing units, from an evaluation of Census Followup Adds and from the Census Coverage Measurement survey. Two memoranda for the record were released.

In the first memorandum, "Initial Comparison of Matching Results from 2010 Duplicate Person Identification (DPI) Process to Research Results on the 2000 Census," staff compared available results from the research based on the 2000 Census to the results from the 2010 DPI. Staff found nearly double the amount of provisionally accepted links in the 2000 research compared with the 2010 DPI, due to the provisionally accepted within-response links within the block or tract.

In the second memorandum, "Miscellaneous Situations Encountered During Census Unduplication Research or 2010 Duplicate Person Identification (DPI) Cutoff Review" staff briefly describe each situation, the resolution of the situation, and the reasoning behind the resolution. Staff revised two related specifications: "2010 Decennial Census Coverage Followup and Census

Coverage Measurement Match Modeling Software Requirements Specification" (procedures for evaluating person links) and "2010 Decennial Census Census Coverage Followup and Census Coverage Measurement Person Matching Parameter Software Requirements Specification" (parameter settings for matching). Staff are co-authors for both specifications.

Staff provided to the Australian Bureau of Statistics (ABS) requested information on unduplication strategies used for the 2010 Census. After consultation with the Decennial Statistical Studies Division (DSSD), staff obtained and sent to ABS an electronic copy of a study plan that included how potential duplicates were resolved. Staff provided additional detail on the structure of the computer matching operation. Finally, the results of the 2010 Decennial Census Person Identification process have been used in a cognitive/qualitative project to assess the results of the matching.

Staff: Michael Ikeda (x31756), Ned Porter

C. Voting Rights Section 203 Model Based Methodology: Research, Development, and Production

Description: Section 203 of the Voting Rights Act asks for determinations relating to limited English proficiency and limited education of specified small domains (race and ethnicity groups) at a small area such as county or minor civil division (MCD) level. The research undertaken seeks to provide a small area model-based estimate derived from American Community Survey (ACS) 5-year data and 2010 Census data, which will greatly improve over the estimates which would be alternatively provided from ACS design-based estimates.

Highlights: In response to consultations with the Decennial Statistical Studies Division (DSSD) and consultations with members of the Director's Office, in FY 2011, staff began an investigation into the application of small area models relating to Section 203. Initial explorations began with exploratory data analysis and the construction of synthetic models, and then proceeded into the construction of generalized linear hierarchical models. These models use the American Community Survey (ACS) unit level weights in conjunction with a binomial distribution-based small area model. Staff also developed a regression model to create a stratification procedure which identifies small areas that are similar to each other by using informative characteristics from ACS and Census. Staff refined the methodology to deal with multiple race/ethnicity persons and the appropriate use of stratification in relation to the small area model. Staff developed and tested a program for the purpose of small area estimation and confirmed the estimates with a

parallel estimation program. Likewise, the particular stratification assignments were verified between center staff and DSSD staff. The final estimates were delivered to DSSD for the purpose of producing the Section 203 determinations.

Staff was also charged with double programming the Section 203 determinations for the proposed model form. The variances from the model-based methodology was compared to the ACS 2005-09 methodology, and it was determined that the model-based method was superior to the ACS method beyond the preset thresholds. The model-based determinations, as of the end of FY 2011, await publication in the *Federal Register* but are otherwise considered complete. Staff continues to formulate a main methodology technical report. Research for future application in 2015 is ongoing.

Staff: Patrick Joyce (x36793), Donald Malec, Ryan Janicki, Aaron Gilary, Rolando Rodriguez, Julie Tsay, Rob Creecy

D. Synthetic Decennial Microdata File
Description: In some cases, data users have an interest in the full microdata file whose disclosure is prohibited by law. We seek to produce synthetic individual records (microdata files) so that when we produce tables from them, the results are the same as the comparable publicly available tables from the original individual records. The synthetic individual records should be close to the underlying microdata while protecting confidentiality. The goal of this project is to produce synthetic microdata files from the decennial short form (now the American Community Survey) variables for block level geography. We are approaching the problem by using iterative proportional fitting and log linear models from fully cross-classified tables of short form variables and then creating synthetic microdata records by randomly sampling records using the estimated parameters.

Highlights: In FY 2011, staff created a synthetic Decennial Census data file for the state of Maryland using a three step process of: (1) using iterative proportional fitting to fit a log-linear model to year 2000 Census data for Maryland; (2) performing an allocation of the fitted values from the model in such a way which accounts for the fact that published Census tabulations provide more detailed age information at tract level than at block level geography (thus allowing the synthetic data to match these detailed tract level age margins that are published); and (3) applying a controlled unbiased rounding procedure to obtain integer valued synthetic data that match the year 2000 Census data for Maryland nearly exactly on two published high order margins.

The synthetic Decennial Census file in this case consists of a complete cross-classification of Maryland's population based on the variables: sex, age, race, Hispanic origin, relationship to head of the household,

and census block. Additionally, the large and sparse contingency tables which arise in this application have motivated a study of methods for testing dependence in such tables. For the two dimensional case, staff proposed some tests for dependence which, based on an empirical study, appear to maintain the nominal probability of type I error.

Staff: Martin Klein (x37856), Rob Creecy

1.2 COVERAGE MEASUREMENT PLANNING AND DEVELOPMENT (Decennial Project 5610103)

A. Coverage Measurement Research
Description: Staff members conduct research on model-based small area estimation of census coverage, and they consult and collaborate on modeling census coverage measurement (CCM).

Logistic Mixed Models for Small Domain Estimation
The research done in FY 2011 is as follows. Using Census 2000 revision II data on the correct enumeration rate, staff conducted an evaluation to include random effects in the logistic coverage models as a way to safeguard against model error (due to type II error). The aim of the study is to determine the possible gains in using random effects in the coverage measurement context and to determine the usefulness of available software. The resulting effect estimates demonstrate the typical compromise between their corresponding fixed effect estimate and zero. If only one random effect distribution is used for a large number of effects (many which may be zero), the added random effects contribute little due to a small estimated variance component. If however, subsets have separate random effects distributions, some effects make a larger contribution. Using a large, realistic model based on Decennial Statistical Studies Division (DSSD) coverage research, the standard SAS procedures often never converged. Both an implementation of Breslow and Lin's approximation in PROC IML and the Pinero and Bates procedure (implemented in R) performed well. Work is needed in developing data-based procedures to select multiple random effect models, as well as in developing better ways to handle the "design effects" in the data. Staff suggest that the techniques could be used, as-is, as part of a sensitivity test. Staff published a report detailing the research done in FY 2010 to the *Center for Statistical Research & Methodology Research Report Series.*

Estimation of Synthetic Estimators in the CCM Program:
The CCM Group is evaluating the performance of a synthetic estimator based on a logistic model for small areas (e.g., counties and places). Staff is comparing an average mean squared error approach (Gonzales and Waksberg 1973, Marker 1995) to a Bayesian formulation to obtain estimates of the synthetic bias (error due to

using data outside the area of interest). ASA/NSF/Census Research Fellow Partha Lahiri developed solution to negative estimates in average means squared error approach using adjusted likelihoods. Comparison of the two approaches is currently ongoing.

Empirical Best Prediction (EBP) Method to Estimate the Proportion of Erroneous Enumeration for the 2006 Travis County, Texas Test Data:
This project is to help understand how future censuses can be improved, and the main goal is to develop a robust statistical method to estimate the proportion of erroneous census enumeration, which is an important component of possible census error that the CCM group is studying. The methodology developed only depends on moment assumptions. Staff is working with Partha Lahiri to develop a Census Coverage model for race/origin, age/sex, and tenure covariates. Staff has applied a two-stage cluster sampling model to the 2006 CCM data to find variance estimates for the coverage samples, and staff will use this data into Dr. Lahiri's model for Census Coverage.

Non-ignorable Models for Coverage Component Estimation
In FY 2011, staff developed new partition model for analyzing categorical data with possibly non-ignorable non-response. The model was extended to account for the possibility that certain domains may be weakly related or unrelated to the other domains in the model, a case which our current partition model cannot handle. Our new approach combined parts of the uniform model and the partition model, where domains with a uniform component are not assumed to share parameters across cells. Early simulation results suggested this method produces accurate results when the model assumptions are close to correct, and that estimates are more robust to model misspecification than other non-ignorable models. A report documenting this work has been completed and published to the *Center for Statistical Research & Methodology Research Report Series*.

Coverage Measurement Planning and Development:
Staff used data from Census 2000 and Census Coverage Estimation to develop race-based, data-driven models for the Match and Correct Enumeration rates. These models were used to evaluate the viability of producing multiracial estimates, the success of the current race/origin domain framework, and the validity of modeling different racial subpopulations, among other questions. Staff presented recommendations several times to Decennial Statistical Studies Division (DSSD) stakeholders, in meetings of the Component Missing Data team, and the Census Coverage Measurement (CCM) Estimation team. Staff is submitting a research report for review based upon this work. Staff has also developed a SAS program to automate the recursive partitioning procedure, based on previous partitioning work.

Staff: Jerry Maples (x32873), Aaron Gilary, Ryan Janicki, Eric Slud

B. Accuracy of Coverage Measurement
Description: 2010 Census Coverage Measurement (CCM) Research conducts the research necessary to develop methodology for evaluating the coverage of the 2010 Census. This includes planning, designing, and conducting the research, as well as analyzing and synthesizing the results to evaluate their accuracy and quality. The focus of this research is on the design of the Census Coverage Measurement survey and estimation of components of coverage error, with a secondary emphasis on the estimation of net coverage error. Overcount and undercount estimation has not been done separately for previous censuses because of the difficulty of obtaining adequate data for unbiased estimates.

Highlights: In FY 2011, staff continued work on a 2010 Census Program for Evaluations and Experiments (CPEX) project on the development of the error structure for the estimates of the net census coverage error, census erroneous enumerations, and census omissions that will be produced by the Census Coverage Measurement Program (CCM). Staff developed a general form for estimators to account for nonsampling errors through adjusting the sufficient statistics for logistic regression estimation. This has been in the context of using logistic regression in dual system estimation as planned for the CCM. In addition, staff examined the data that will be available from evaluations of CCM and assessed its potential for aiding in the estimation of nonsampling errors. These results will be used in research that seeks to use sensitivity analyses to investigate the effects of nonsampling errors on the estimates of net census coverage erroneous enumeration, and omissions. In particular, the research will assess the combined effect of all the nonsampling errors and identify which sources and types of error have the largest effects.

Staff co-authored study plans for two other CPEX projects. The Matching Error Study (MES) Study Plan will provide information about the level and effect of clerical matching errors on net error and components of census coverage person estimates for the 2010 Census Coverage Measurement (CCM) survey. The methodology for the study is designed to use data from the quality control operation for 2010 CCM person matching operation to produce estimates of the residual matching error in the CCM data. The Evaluation to Assess Effect of Errors in Census Address List on Census Coverage Measurement Estimates Study Plan will examine the impact of (1) census geocoding errors and (2) errors in the creation of the initial census address list on the CCM estimates of census coverage error. The study design uses data generated by another CPEX study, the Evaluation of Address Frame Accuracy and Quality, to estimate the effect of errors in the creation of the census address list on the quality of estimates of census coverage error.

Staff designed methodology to describe the uncertainty in the estimates of foreign-born immigration based on data from the American Community Survey (ACS). The Demographic Analysis estimates of the U.S. population on April 1, 2010 included estimates of foreign-born immigration each year between 2000 and 2010 based on the data from the ACS. The main estimation method for foreign-born immigration uses the responses to two questions, one that asks citizenship and another that asks Residence One Year Ago (ROYA). Staff identified the sources of nonsampling error in the estimates of foreign-born immigration. The estimate of the foreign-born immigration may be affected by (1) errors in Population Estimates Program (PEP) estimates used as population controls, (2) ACS data errors that create error in the ACS estimate of the foreign-born who resided outside the U.S. one year ago, and (3) errors caused by an inconsistency between the ACS and the PEP in the variables used to form cells for the ratio adjustment. Staff also developed a strategy to use the error model in the design of a simulation to study the propagation of errors in the estimates of foreign-born immigration. The results of the planned simulation study will produce a sensitivity analysis that assesses the uncertainty in the form of the bias and variance in estimates of foreign-born immigration.

Staff: Mary Mulry (x31759)

1.3 EVALUATION PLANNING COORDINATION
(Decennial Project 5610106)

A. Explaining How Census Errors Occur through Comparing Census Operations History with Census Coverage Measurement (CCM) Results
Description: The goal of this project is to understand what errors tend to be associated with the different Census operations, especially for persons and housing units removed from the census. We will compare Census files to the CCM results for a subsample of CCM areas. This comparison is intended to help find patterns of errors in Census operations and provide insights into ways to avoid these errors.

Highlights: The study plan for the evaluation was released by the Decennial Management Division (DMD) as a *2010 Census Planning Memorandum.* The decision was made to remove persons deleted by the Primary Selection Algorithm (PSA), which resolves cases of multiple returns at a Census ID from the scope of the evaluation. The reason for this decision was concerns that determining which persons were deleted by PSA would compromise the confidentiality of the PSA process.

Staff wrote several drafts of the software specification for the person computer matching and processing for the evaluation and sent them out for comments. Staff

outlined proposed procedures for clerical matching based on discussions with staff from the Decennial Statistical Studies Division and the National Processing Center.

Staff: Michael Ikeda (x31756), Mary Mulry

1.4 AMERICAN COMMUNITY SURVEY (ACS)
(Decennial Project 5385160)

A. ACS Applications for Time Series Methods
Description: This project undertakes research and studies on applying time series methodology in support of the American Community Survey (ACS).

Highlights: In FY 2011, staff met with clients to discuss alternative temporal weighting schemes of raw ACS data, with the goal to reduce bias and improve timeliness. A methodology was outlined, drafted, and encoded in expectation that it be applied in the future. Staff made revisions and improvements to the methodology for assessing the usability of multi-year estimates.

Staff: Tucker McElroy (x33227)

B. ACS Variances
Description: Work under this heading this year concerned three research projects: (i) revision of a journal paper on simultaneous nonresponse adjustment, calibration, and weight trimming with new examples on Survey of Income and Program Participation (SIPP) data; (ii) revision of a journal paper on comparison of alternative methods of estimating variances for complex survey estimates with a new method, 'hybrid'+ estimates of cross-classified population totals containing design-based cell size estimates for some cross-classified categories, and model-based conditional-probability estimates within other categories; and (iii) (Slud only) initiation of a project to develop new methods to assess the validity of 0 estimates in ACS tables.

Highlights: Work in FY 2011 included additional computing and writing and giving talks on the papers under revision under headings (i) and (ii). In (iii), work consisted of literature review, drafting a research proposal, and ACS data specification for a small area estimation project based on generalized linear models which will be used to obtain upper confidence bounds for ACS table cells in which estimates are 0.

Work continued on a project and paper on simultaneous nonresponse adjustment, calibration, and weight trimming; during this quarter work was restricted to new calculations of examples based on Survey of Income and Program Participation (SIPP) data.

A second ongoing project and paper revision concerns hybrid estimation methods combining design-based

estimates with marginal table totals and model-based estimates of conditional probabilities. Work this quarter involved only rewriting earlier efforts.

A third project, directed toward assessing the validity of "zero-estimates" in ACS tables, continued actively this quarter. Both a preprint and a seminar for the Center for Statistical Research & Methodology (CSRM) given May 3 were prepared. These reviewed competing methods of providing upper confidence bounds for survey-based cell probabilities; and a model-based method was developed for interval estimation of finely cross-classified demographic cells by combining information across higher-level aggregates. The models developed are random-effects generalized linear models incorporating a higher-level synthetic predictor, which is in some cases the only significant predictor variable.

Staff: Eric Slud (x34991), Yves Thibaudeau

C. ACS Small Area Estimation for Selected Characteristics
Description: This project aims to propose, develop, and evaluate small area estimation methodologies to produce ACS estimates for selected characteristics in geographies with small populations. The characteristics of initial interest are unemployment, income, and poverty.

Highlights: No significant progress in FY 2011.

Staff: Lynn Weidman (x34902)

D. ACS Small Area Estimation for Group Quarters (GQ)
Description: This project aims to propose, develop, and evaluate small area estimation methodologies to produce ACS estimates for the GQ population (totals and characteristics) for substate geographies, including counties, places, and tracts.

Highlights: Imputation for the second evaluation using ACS data from years 2006-09 along with a derived 2010 sample data file was carried out and weights calculated. The number of donors used, how often each donor was used, and geographic and type relationships between the donor and imputed GQs were analyzed. Presentations on this analysis, the imputation procedure, and a project overview were made at a meeting of the National Academy of Sciences Panel on Statistical Methods for Measuring the Group Quarters Population in the ACS. In early March, the new ACS Change Control Board decided that implementation of this methodology will be postponed until next year. Staff in several other divisions are working on defining and implementing processing changes that will be needed for applying this methodology for the ACS data products produced in 2012.

Staff: Lynn Weidman (x34902), Chandra Erdman, Patrick Joyce, Chaitra Nagaraja

E. ACS Data Issues
Description: Various issues related to the quality and presentation of ACS estimates were discussed and investigated by small interdivisional teams or division staff. The goal of these investigations was to make recommendations to aid in resolving the issues.

Highlights: Staff prepared files of 1000 ACS-like samples of housing units for the states of Texas and Oregon based on the Census 2000 sample data. Logit models were fitted to predict ACS mail and CATI nonrespondents (CAPI eligible cases) as a function of a few basic variables using ACS 2009 data and applied to the sample housing units to randomly flag CAPI eligible cases. In each of the 1000 samples, 1/3 of these cases were systematically selected as the CAPI subsample. These data were used by DSSD staff to carry out a research project looking at alternative statistical methods for deriving confidence intervals for very small estimates.

See Project F, "ACS Exploratory Analysis of the Differences in ACS Respondent Characteristics between the Mandatory and Voluntary Response Treatments."

Staff: Lynn Weidman (x34902), Julie Tsay, Michael Ikeda

F. ACS Exploratory Analysis of the Differences in ACS Respondent Characteristics between the Mandatory and Voluntary Response Treatments
Description: In 2002 and 2003, the Census Bureau conducted research to determine whether the ACS could be implemented as a voluntary, rather than a mandatory, survey. It showed a decrease of over twenty percent in mail response when the survey was voluntary, and the reliability of estimates was adversely impacted by the reduction in the total number of completed interviews. This project revisits the test data and uses logit models to determine person, housing unit, and household characteristics most closely related to lower response under voluntary treatment. Also studied is whether their corresponding estimates at the national level are significantly different. This project revisits the test data and uses logit models to determine person and housing unit/household characteristics most closely related to differential response between the voluntary and mandatory methods. Also studied is whether related profile estimates at the national level for the two methods are significantly different.

Highlights: All of the mandatory and voluntary treatment cases were used as a single dataset and logit models were fitted for the probability of being a mandatory vs. a voluntary respondent. Separate models were fitted for the nation and for the eight mail-back propensity segments used in the 2010 Census Integrated Communication Program. Results were examined to see which variables were the earlier selected variables in stepwise modeling and if their coefficient values were related to higher or lower response under the voluntary

method. For profile estimates related to these variables, comparisons were made between their national level estimates for the two methods to see if for those with significant differences the direction of difference was consistent with that indicated by the related variables' coefficients. If this is the case, then we say that differences in respondent characteristics are carried through to differences in the estimates. A report on the project and its results was prepared.

Staff: Lynn Weidman (x34902), Michael Ikeda, Julie Tsay

G. ACS Imputation Research and Development

Description: The American Community Survey process of editing and post-edit data-review is currently time and labor intensive. It involves repeatedly submitting an entire collection year of micro-data to an edit-enforcement program (SAS software). After each pass through the edit-enforcement program, a labor-intensive review process is conducted by a staff of analysts to identify inconsistencies and to quality problems remaining in the micro-data. Before the data are ready for public release, they have least three passes through the edit-enforcement program and three review processes by the analysts, taking upward of three months. The objective of this project is to experiment with a different strategy for editing, while keeping the same edit rules, and to assess if the new strategy can reduce the number of passes through the edit process and the duration of the review process.

Highlights: We have been preparing a version of TEA – a general edit/imputation software – for R, which is also going to be accessible from SAS. We have begun documenting the software according to professional standards. We use two related programs to do this: the R package *roxygen* and the C library *doxygen*, which allow for auto-creation of documentation from comments in the source code.

In FY 2011, TEA has been implemented for the disclosure limitation operations for the American Community Survey group quarters. The software has been expanded and can handle a variety of survey situations, including ACS housing units. A SAS interface is under development and documentation was made available.

Staff: Yves Thibaudeau *(*x31706), Ben Klemens, Rolando Rodriguez

H. ACS Item Nonresponse Rates

Description: The Census Bureau calculates and publishes item allocation rates for all data items collected in the American Community Survey as part of its data quality measures. However, these rates are actually composed of two separate components of item Nonresponse: items that can be assigned a value through the use of related items on the survey questionnaire and items that require a

statistical procedure to allocate the value of the missing item. The published allocation rates combine the two measures into one rate. It is desirable to look at both components separately to see if these rates vary across mode of data collection, individual items, geography, certain population groups, location of call centers (for data collected by CATI), and other items. In addition to the item nonresponse rates, a completeness index can be computed for the entire questionnaire and examined by the same levels as the item nonresponse rates.

Highlights: Computer programs are currently being written to calculate allocation rates, assignment rates, and total item nonresponse rates for all published questionnaire items from the 2005 and 2006 American Community Survey (ACS). The rates will also be computed by interview mode: mail, CATI, and CAPI and included the Group Quarters population for the first time for the 2006 ACS data. This project has ended because work has been completed. Our work has confirmed similar work.

Staff: Tom Petkunas (x33216)

1.5 DEMOGRAPHIC STATISTICAL METHODS DIVISION SPECIAL PROJECTS (Demographic Project TBA)

A. Tobacco Use Supplement (NCI) Small Domain Models

Description: Staff is working with Demographic Statistical Methods Division (DSMD) on a project for the National Cancer Institute (NCI), studying the relationship between smoking status and a range of geographic/ demographic covariates. Using the Tobacco Use Supplement to the Current Population Survey (TUS-CPS), staff is assisting NCI toward making estimates of smoking related behavior using county-level or state-level dependent variables (e.g., percent of males, percent Hispanic, percent below poverty level). The goal is to identify where anti-smoking funds could best be directed.

Highlights: In FY 2011, staff collaborated with CPS staff and an NCI researcher who is primarily responsible for the project, establishing which variables were needed and providing guidance for developing a general modeling framework. Staff established the general parameters for the modeling, and developed preliminary area-level and unit-level models to obtain county-level estimates and standard errors for the relevant smoking variables for all counties in the TUS-CPS sample in 2006-2007. Staff found design effects for the sampling scheme, which will be input to the modeling. Staff generated a simulated multivariate normal dataset to provide to the researcher, and collaborated with her on testing programming code developed at NCI.

Staff: Aaron Gilary (x39660), Partha Lahiri

B. Top Coding Investigation
Description: The objective of this project is to investigate novel methodologies for analyzing top coded income data that follows a log normal model of distribution.

Highlights: In FY 2011, staff investigated statistical inference for a variety of income parameters that are of practical interest (e.g., the mean, median, percentiles, Gini coefficient, fraction of incomes that are top coded in the population), based on a lognormally distributed income sample. Newly developed confidence intervals were found to exhibit excellent performance in terms of maintaining the coverage probability. The newly developed approach based on the concept of a generalized confidence interval appears to provide a unified methodology for analyzing top coded data that follow a lognormal distribution.

Staff has made significant progress in investigating noise perturbation as opposed to multiple imputation for disclosure control. A comparison of noise perturbation and multiple imputation, from the perspective of accuracy of the statistical inference, is currently under way. Noise perturbation as an alternative to top coding and multiple imputation is also being investigated.

Staff: Thomas Mathew (x35337), Aref Dajani (DSMD)

1.6 DEMOGRAPHIC SURVEYS DIVISION (DSD) SPECIAL PROJECTS
(Demographic Project 0906/7374)

A. Data Integration
Description: The purpose of this research is to identify microdata records at risk of disclosure due to publicly available databases. Microdata from all Census Bureau sample surveys and censuses will be examined. Potentially linkable data files will be identified. Disclosure avoidance procedures will be developed and applied to protect any records at risk of disclosure.

Highlights: Staff developed software to facilitate Decennial Statistical Studies Division (DSSD) testing of reidentification candidates. Staff recommended 40,921 candidates for reidentification to DSSD, and among those DSSD reidentified one, meaning that the experiment has thus far given a negative result. Staff is rechecking its methodology before definitive conclusions are drawn.

Staff discovered 87 identifiable individuals from the 2007 Public Use Microdata Sample (PUMS) file. These results are unofficial and will need to be verified. Appropriate follow-up is underway.

Staff: Ned Porter (x31798), Lisa Singh (CDAR), Rolando Rodríguez

1.7 SURVEY OF INCOME AND PROGRAM PARTICIPATION IMPROVEMENTS RESEARCH
(Demographic Project 1465444)

A. Model-Based Imputation for the Demographic Directorate
Description: Staff has been asked to review and improve ultimately all of the imputation methodology in demographic surveys, beginning with the Survey of Income and Program Participation and the Current Population Survey.

Highlights: During FY 2011, staff made a number of enhancements to code written for the imputation of earnings in the Survey for Income and Program Participation (SIPP). Model-based and hot-deck methods previously written in R were improved to run faster and with efficient memory usage. These functions were also enhanced by making use of existing R packages that allow for the transmission of files between SAS and R. In addition, CSRM staff modified SAS imputation programs provided by the Social, Economic, and Housing Statistics Division (SEHSD) so that their methodology can be used to impute earnings for data that does not have associated administrative records, and applied to simulated data sets with any subset of the earnings variables the program was originally written to impute. A simulation study to compare the various imputation methods that have been implemented, and to assess the effect of using administrative records to guide the imputation is currently underway.

Staff: Chandra Erdman (x31235), Ben Klemens, Yves Thibaudeau, Maria Garcia

1.8 SOCIAL, ECONOMIC, AND HOUSING STATISTICS DIVISION SMALL AREA ESTIMATION PROJECTS
(Demographic Project 7165000)

A. Research for Small Area Income and Poverty Estimates (SAIPE)
Description: The purpose of this research is to develop, in collaboration with Housing and Household Economic Statistics Division (HHES) – previously part of the Data Integration Division (DID) – methods to produce "reliable" income and poverty estimates for small geographic areas and/or small demographic domains (e.g., poor children age 5-17 for counties). The methods should also produce realistic measures of the accuracy of the estimates (standard errors). The investigation will include assessment of the value of various auxiliary data (from administrative records or surveys) in producing the desired estimates. Also included will be an evaluation of the techniques developed, along with documentation of the methodology.

Highlights: Staff has evaluated county level models for the poverty rate of school-aged children. The proposed poverty rate point model is a normal mixture with a nonlinear mean function and it was fitted using the poverty rate estimates from the American Community Survey (ACS). One of the inputs is the design-based variance estimate for the poverty rate. In counties with small sample size, these variance estimates can lack precision (i.e., grossly underestimate the true variance). To assist the point model, a small area variance model has been developed to use small area estimation techniques on the estimated survey variances. Staff has developed a method to specify the relative variance of the survey variance estimator. Staff has tested the feasibility of this specification in the current model framework.

Staff gave a presentation on current survey variance modeling research titled "Using Small Area Modeling to Improve Design-Based Estimates of Variance for County Level Poverty Estimates in the American Community Survey" at the 2011 Conference on Small Area Estimation in Trier, Germany, Aug. 11-13, 2011.

Staff has evaluated several bivariate models by borrowing information from Census 2000 or previous single or multi-year ACS poverty data for estimating 2009 poverty. In the current SAIPE county (state) poverty model, the Census 2000 poverty estimates have been used as one of the predictors by ignoring its sampling error. However, the variance estimates of the Census 2000 poverty estimates for small counties are large and can not be ignored. One way to take account of the sampling error in the Census 2000 poverty estimate is to treat Census 2000 county poverty estimate as a dependent variable in the 2nd equation of the bivariate model. We also considered borrowing information from the previous year ACS data or multi-year ACS data in the bivariate models because the Census 2000 data are out of date. We used the average of ACS 2005 to 2008 poverty data as the multi-year ACS poverty data. We used the regression variables (omit the Census 2000 data) in the SAIPE production model as the regression variables in the bivariate models. For estimating 2009 county log poverty level or county poverty rate, or state poverty ratio by age groups (ages 0-4, 5-17, 18-64, and 65+), the evaluation results showed that using multi-year ACS estimates is the preferred model than using the outdated Census 2000 or the previous year ACS estimates in the bivariate model. The evaluation results are documented in a paper entitled "An Empirical Study on Using Previous American Community Survey Data Versus Census 2000 Data in Models for SAIPE Poverty Estimates" by Elizabeth T. Huang and William R. Bell.

Staff: Elizabeth Huang (x34923), Jerry Maples, William Bell (Research and Methodology Directorate)

B. Small Area Health Insurance Estimates (SAHIE)
Description: At the request of staff from the Housing and Household Economic Statistics Division (HHES), our staff will review current methodology for making small area estimates for health insurance coverage by state and poverty level. Staff will work on selected topics of SAHIE estimation methodology, in conjunction with HHES.

Highlights: <u>Bayesian Benchmarking of Estimates from Distinct Geographic Models</u>
Currently, estimates for states and counties are made independently from two distinct models. Consequently, county estimates do not add up to the corresponding state estimates. *Ad hoc* adjustments to the county-level estimates produce final estimates that are based on two different, overlapping models making posterior variance determination problematic.

In FY 2011, staff developed three new methods for estimating parameters at different geographic levels while maintaining the benchmarking property. The first method involves fitting multivariate Gaussian distributions, with constraints on the parameters, to the output of the current SAHIE MCMC chain. The second method uses linear transformations to modify the full conditional posterior distributions at lower geographic levels to achieve the benchmarking constraint. The final method involves constructing a new posterior distribution through the use of Lagrange multipliers and exponential tilting, which is close to the original posterior distribution, but includes the benchmarking constraint. Staff began simulation studies to compare the different methods to each other and to the currently used method.

Staff: Ryan Janicki (x35725)

1.9 EDITING METHODS DEVELOPMENT
(Economic Project 2320154)

A. Investigation of Selective Editing Procedures for Foreign Trade Programs
Description: The purpose of this project is to develop selective editing strategies for the U.S. Census Bureau foreign trade statistics program. The Foreign Trade Division (FTD) processes more than five million transaction records every month using a parameter file called the Edit Master. In this project, we investigate the feasibility of using selective editing for identifying the most erroneous records without the use of parameters.

Highlights: During FY 2011, staff continued work on the application of selective editing to the Census Bureau foreign trade data and developed score functions for selective editing of these data. The scores have two components: a measure of how suspicious an incoming record is and a measure of the impact changes in the record may have on publication totals within a particular set of commodity groupings. Staff wrote a new version of the selective editing program using the SAS macros language. The program produces an output file using

SAS PROC RANK. This final procedure assigns to each record a priority ranking according to its score. The ranked output list provides a more efficient target of records for review identifying highly influential records requiring manual intervention prior to Edit Master processing. Staff completed a feasibility study using available test data sets for four consecutive months of export data records. The feasibility study showed that the relative errors in totals decrease as the number of records marked for follow-up are corrected.

Staff: Maria Garcia (x31703), Yves Thibaudeau, Christopher Grieves (FTD)

1.10 TIME SERIES RESEARCH
(Economic Project 2320152)

A. Seasonal Adjustment Support
Description: This is an amalgamation of projects whose composition varies from year to year but always includes maintenance of the seasonal adjustment and benchmarking software used by the Economic Directorate.

Highlights: Seasonal adjustment and X-12-ARIMA support was provided to users both within and outside the Census Bureau in FY 2011, including but not limited to Standard and Poors, Nomura Holdings, Lazard Capital Markets, Bank of America, SAS, UCL Energy Institute, Integrative Statistics, Forecast Miami, Real Estate Consulting, Fiscal Policy Institute, John Deere Company, Macroconsulting, Global Market Strategies Group, IHS Global Insight, Tiverton Trading, Future Fund (Australia), Department of Housing and Urban Development, Department of Transportation, Congressional Budget Office, U. S. Bureau of Economic Analysis, U. S. Bureau of Labor Statistics, U. S. Department of Agriculture, U. S. Senate, Federal Reserve Board, Oklahoma Employment Security Commission, Estadisticas Puerto Rico, Wisconsin Department of Revenue, Pennsylvania Department of Labor & Industry, the state government of Oregon, the Australian Bureau of Statistics, Norwegian Labour and Welfare Administration, Statistics Norway, Eurostat, Office of National Statistics (UK), Statistics Canada, Cabinet Office of Japan, Statistics New Zealand, INSEE, Statistics and Census Service of Macao, Ministry of Manpower (Singapore), Statistical Office of Scotland, National Bank of Belgium, Central Bank of Argentina, Reserve Bank of Australia, European Bank for Reconstruction and Development, Croatian National Bank, Oesterreichische Nationalbank, Princeton University, Lehigh University, University of Carlos III (Spain), University of Groningen (Netherlands), and the Brazilian Institute for Geography and Statistics.

Staff analyzed 26 inventory series from Taiwan and found that the regARIMA models for many of the series were improved by including stock calendar regressors for trading day and Chinese New Year.

Staff from the Time Series Research Staff and the Time Series Methods Staff of the Office of Statistical Methods and Research for Economic Programs (OSMREP) met with analysts from the Department of Agriculture to discuss issues related to time series modeling, forecasting, and seasonal adjustment. Staff responded to requests for information on seasonal and trading day adjustments from reporters of the *Wall Street Journal.* Staff met with analysts from the Economic Directorate to discuss missing value options in X-12-ARIMA and how they could be used in the case of CIR series.

Staff: Brian Monsell (x31721), David Findley (Consultant)

B. Seasonal Adjustment Software Development and Evaluation
Description: The goal of this project is a multi-platform computer program for seasonal adjustment, trend estimation, and calendar effect estimation that goes beyond the adjustment capabilities of the Census X-11 and Statistics Canada X-11-ARIMA programs, and provides more effective diagnostics. This fiscal year's goals include: (1) finishing a version of the X-13ARIMA-SEATS program with accessible output and improved performance so that, when appropriate, SEATS adjustments can be produced by the Economic Directorate; (2) developing software system that provides a simulation environment for X-13 seasonal adjustments called USim-X13; and (3) incorporating further improvements to the X-12-ARIMA/X-13A-S user interface, output and documentation. In coordination and collaboration with the Time Series Methods Staff of the Office of Statistical Methods and Research for Economic Programs (OSMREP), the staff will provide internal and/or external training in the use of X-12-ARIMA and the associated programs, such as X-12-Graph, when appropriate.

Highlights: In FY 2011, staff released an updated version of X-12-ARIMA, Build 192 of Version 0.3, to the Economic Directorate for their testing and then to the general public. Staff compared adjustments from this version of the software to the last released version of X-12-ARIMA (Build 188) and found in most cases no differences in the adjustments, and small differences between compilers on Linux-based machines. Staff also provided support for analysts and programmers in the Economic Directorate in their testing of the new release. Before release of the software, staff repaired a defect in X-12-ARIMA to ensure that the same starting date was provided for revision history analysis for each of the components of a revisions history analysis of the indirect seasonally adjusted series.

Staff incorporated new model based seasonal adjustment diagnostics coded in Fortran and C++ into a version of X-

13ARIMA-SEATS. This version was shared with developers at the Bank of Spain and the Bank of Belgium for their research and development purposes. In addition, an updated likelihood testing regime for trading day regressors that includes 1 coefficient trading day regressors for both flow and stock series was finished and tested for the X-13ARIMA-SEATS prototype. Staff also repaired a defect in the X-13ARIMA-SEATS prototype related to saving information on the spectral diagnostics, and completed work to convert the output of X-13ARIMA-SEATS to accessible HTML output in preparation for testing the program by the Economic Directorate and release to the public.

Staff enhanced the uSim-X13 software package by the addition of a nonparametric modeling interface featuring new methods in AM-FM empirical mode decomposition for the extraction and analysis of signals in time series, incorporated regCMPNT routines, and included access to routines for likelihood testing of calendar regressors.

Finally, staff developed Fortran routines to generate accessible HTML output for the regCMPNT program, and added additional options to save program output into a directory for use with graphics software.

Staff: Brian Monsell (x31721), Christopher Blakely, Natalya Titova, David Findley (Consultant)

C. Research on Seasonal Time Series - Modeling and Adjustment Issues

Description: The main goal of this research is to discover new ways in which time series models can be used to improve seasonal and calendar effect adjustments. An important secondary goal is the development or improvement of modeling and adjustment diagnostics.

This fiscal year's projects include: (1) continuing research on goodness of fit diagnostics (including signal extraction diagnostics and Ljung-Box statistics) to better assess time series models used in seasonal adjustment; (2) studying the effects of model based seasonal adjustment filters; (3) studying multiple testing problems arising from applying several statistics at once; (4) determining if information from the direct seasonally adjusted series of a composite seasonal adjustment can be used to modify the components of an indirect seasonal adjustment, and more generally investigating the topics of benchmarking and reconciliation for multiple time series; (5) studying alternative models of seasonality, such as Bayesian and/or long memory models and/or heteroskedastic models, to determine if improvement to seasonal adjustment methodology can be obtained; (6) studying the modeling of stock holiday and trading day on Census Bureau time series; (7) studying methods of seasonal adjustment when the data is no longer univariate or discrete (e.g., multiple frequencies or multiple series); (8) studying alternative seasonal adjustment methods that may reduce revisions or have alternative properties; (9) studying nonparametric methods for estimating

regression effects, and their behavior under long range dependence and/or extreme values.

Highlights: During FY 2011, staff: (a) Extended the Empirical Mode Decomposition approach to seasonal adjustment to perform decompositions based on salient frequencies in the data, and improved it by utilizing adaptive filtering rather than splines. The method uses X-12-ARIMA to handle regression effects, and thus can be applied to nonstationary time series (Project 8).

(b) Studied the performance of benchmarking methods with graphics developed by Statistics Canada, and compared the performance of the proportional Denton benchmarking procedure on Census Bureau series to that of the Causey-Trager method used at the Census Bureau (Project 4).

(c) Extended formulas and code that allow estimation of forecasts and signal estimates for mixed frequency stock and flow data, such as mixed monthly and quarterly data. Staff tested the method on synthetic and real data satisfactorily. The eventual goal is to produce X-12-ARIMA seasonal adjustment estimates for such mixed frequency data (Project 7).

(d) Developed a sequential testing paradigm for controlling Type I error when dependent statistical tests are applied serially. The method was derived for the sequence of Ljung-Box statistics, for which the joint asymptotic distribution was previously derived, allowing for a correct statement of p-values (Projects 1 and 3).

(e) Developed full distribution theory and power curves for studentized sample mean statistics, when the time series data has long-range dependence, anti-persistence, or is the difference of a stationary time series. Staff explored two viable methods for consistently estimating the quantiles of the limit distribution, namely a plug-in approach and the subsampling methodology (Project 9).

(f) Studied the performance of common tail index estimators in the presence of long-range dependence, with the finding that heavy serial correlation has no impact on quantile-based statistics (Project 9).

(g) Developed asymptotic results on the sample autocorrelations and the mixing properties of heavy-tailed long-range dependent time series, and the behavior of studentized test statistics in this context (Project 9).

(h) Continued research on an alternative model fitting criterion based on multi-step ahead forecasting performance, and compared empirical performance of competing methods on a sub-group of series from a forecasting competition. The criterion that uses 12-step ahead forecasting error performs about as well as the classical 1-step ahead criterion, across most monthly time series that were considered. Staff demonstrated that

empirical and theoretical properties are in concordance, it terms of in-sample forecasting optimization (Project 5).

(i) Continued research on multivariate signal extraction and seasonal adjustment for multiple time series, developing code to fit common trends and common cycles models – as well as produce signal extraction estimates and multivariate frequency response graphs – and developing empirical examples (Project 7).

(j) Derived formulas to forecast a continuous-time causal process, and also obtained concurrent signal extraction formulas. Developed an algorithm for spectral factorization of continuous time ARMA processes (Project 7).

(k) Provided additional empirical analysis and justification for the dynamic matching seasonal adjustment method, which provides seasonal adjustments with correlation patterns matching those of the original signal. Over dozens of series, staff showed a remarkable decline in negative residual seasonal autocorrelation in seasonally adjusted data, and a decrease in revisions (Project 8).

(l) Developed code allowing estimation of hysteric economic time series models, where correlation is allowed between seasonal, trend, and irregular. Estimation of the model utilizes a model parameterization that allows unconstrained optimization, such that a positive definite matrix for the covariance of the three innovation processes is always guaranteed. Such models, if found to be a better description of a series, offer a quite different seasonal adjustment method. Empirical examples developed, and for some series the new models were superior (Project 8).

(m) Developed methods and code for fitting and studying constrained bivariate VAR(1) models, and accompanying test statistics to assess one-step ahead forecast performance for competing models. It is shown that the optima of these constrained models to not minimize the component functions of the asymptotic mean squared forecast error matrix, as is the case with unconstrained VAR models, and hence do not satisfy the Yule-Walker equations. Also developed test statistics to assess forecast performance of these models (Projects 1 and 7).

(n) Completed revisions to work studying the impact of prior seasonal adjustment on the estimation of business cycles, and the correct quantification of their statistical error. Final results indicate that using a full cycle-trend-seasonal-irregular can be reliably estimated, and does the best job of providing accurate signal extraction mean squared errors (Projects 2 and 5).

(o) Implemented a fully Bayesian procedure for computing seasonal adjustments via long-range dependent unobserved component models. Final results indicated that trend and seasonal long memory

parameters could be reliably estimated when integration was properly accounted for. The resulting seasonal adjustments were qualitatively similar to those of other popular models, but with a fully Bayesian quantification of uncertainty (Projects 5 and 8).

(p) Developed a variance decomposition for filtered nonstationary time series, that allows one to interpret phase delay and gain appropriately. Code is partially developed for visualization (Project 2).

(q) Studied the estimation of a new heavy-tailed long-range dependent time series model, which may be useful for addressing the presence of extreme values. Staff developed an algorithm for computing the likelihood of a new time series model for serially correlated heavy-tailed variables (Project 5).

(r) Developed an algorithm for computing the covariance function of a cepstral random field model, which facilitates maximum likelihood and Bayesian estimation (Project 10).

(s) Developed a feasible model and computational algorithm for generalizing the Exponential time series model to the vector time series case, also allowing for modeling of long memory (Projects 5 and 7).

(t) Began writing survey paper on the Direct Filter Approach, connecting concurrent filter optimization with various forecasting and signal extraction problems.

(u) Completed a note describing why the sample autocorrelations of moving average processes have oscillatory behavior.

(v) Started research on comparisons of direct and iterated forecasting, demonstrating that the approaches are identical when a semi-infinite past is utilized.

Staff: Tucker McElroy (x33227), Christopher Blakely, Brian Monsell, Ekaterina Sotiris, William Bell (Research and Methodology Directorate), David Findley (Consultant)

D. Supporting Documentation and Software for X-12-ARIMA and X-13A-S
Description: The purpose of this project is to develop supplementary documentation and utilities for X-12-ARIMA and X-13A-S that enable both inexperienced seasonal adjustors and experts to use the program as effectively as their backgrounds permit. This fiscal year's goals include improving the X-13ARIMA-SEATS documentation, rendering the output from X-13A-S accessible, further developing the Usim-X13 software and documentation, and exploring the use of component and Java software developed at the National Bank of Belgium.

11

Highlights: In FY 2011, staff updated the X-12-ARIMA site in conjunction with the release of Build 192 of X-12-ARIMA, and updated the sites of the Win X-12 and X-12-Graph programs developed by the Time Series Methods Staff (OSMREP). Staff also released updated version of the Genhol utility to the public, with the ability to generate change of regime holiday regressors. The Seasonal Adjustment Papers website was also updated with new papers. Maintenance of these sites continued to ensure that they follow standards established by the Census Bureau.

Staff developed training materials and examples for a seasonal adjustment class taught in Taiwan using X-13ARIMA-SEATS, Genhol, and other software. Staff is developing a user manual for the uSimX13 software package that details the current methods used from X-13-A-S as well as the nonparametric modeling efforts. Staff prepared documents for the Economic Directorate describing differences between the output of the accessible version of X-12-ARIMA and previous versions of the software, and separate documents that showed the enhancements of the X-12-ARIMA software being prepared for public release. Additionally, staff developed an updated list of input options for the regCMPNT program for eventual release to the public. Staff developed a list of undocumented options in the X-12-ARIMA program.

Staff: Brian Monsell (x31721), Christopher Blakely, David Findley (consultant)

1.11 SURVEY OF RESEARCH AND DEVELOPMENT IN INDUSTRY, IMPUTATION AND SAMPLING RESEARCH AND SOFTWARE DESIGN (Economic Project TBA)

Description: This project undertakes research on the imputation of unreported mandatory items in the Survey of Research and Development in Industry, sponsored by the National Science Foundation. It also examines what estimators are more appropriate under alternative sampling plans; in particular, it evaluates using calibration estimators to compensate for missing data. The possibility of extending calibration to new sampling plans, such as balanced sampling, is investigated. Both traditional linear regression techniques and nonparametric regression techniques are examined.

Highlights: Staff proposed new imputation methods for longitudinal data with non-monotone patterns of nonresponse. The proposed methods are based on an assumption that the nonresponse mechanism depends on past observed or unobserved values of the study variable. A parametric model is not required on the joint distribution of the study variable across time points or on the nonresponse mechanism. Under the assumption of past-value-dependent nonresponse, these proposed methods lead to unbiased or nearly unbiased estimators of the total or mean of the study variable at a given time point. Bootstrap methods have been proposed to obtain standard errors and confidence intervals that account for additional uncertainty due to imputation. Three of the proposed imputation methods are adopted for further analysis: 1) linear regression imputation, 2) kernel regression imputation, and 3) one-dimensional index kernel regression imputation. Staff applied and compared these imputation methods to data from the Survey of Industrial Research and Development (SIRD). Staff conducted simulation studies to further investigate the properties of the proposed imputation procedures; the simulation settings include the case of a normal population, and the case of a population constructed from SIRD data. Other imputation methods from the literature have also been considered in the empirical studies. Staff prepared a paper based upon this work. Staff presented the paper in two professional settings, submitted the paper for publication, and are currently working on a revision.

Staff has started investigating another technique in variance estimation, the Bayesian bootstrap.

Staff: Yves Thibaudeau (x31706), Martin Klein, Jun Shao

1.12 GOVERNMENTS DIVISION PROJECT ON DECISION-BASED ESTIMATION (Economic Project TBA)

Description: This project involves joint work with Governments Division on point and variance estimation for total government employment and payrolls in the Survey of Public Employment and Payroll, within a 'decision-based' method stratumwise GREG estimation after collapsing substrata of small versus large units according to the results of hypothesis tests on equality of regression slopes.

Highlights: In FY 2011, staff collaborated with Jun Shao to develop the large-sample behavior of decision-based estimators, after lengthy efforts to learn the bias and variance properties of such estimates through Monte Carlo and Bootstrap simulations in R. Ongoing work includes study of the method's moderate-sample behavior when regression models are misspecified and Monte Carlo assessment of Bootstrap variance estimators for this decision-based estimation method.

Staff wrote two short preprints in preparation for an Invited JSM talk on the topic, one on the MSE behavior of one-substratum versus two- substratum estimators exploring when there may be advantages of the single-stratum pooled estimator (and therefore also of the decision-based estimator), and one addressing theoretical

obstacles to estimation of rejection probabilities by bootstrap methods.

Staff: Eric Slud (x34991), Yang Cheng (GOVS), Carma Hogue (GOVS)

1.13 PROGRAM DIVISION OVERHEAD
(Census Bureau Project 0381000)

A. Center Leadership and Support
This staff provides ongoing leadership and support for the overall collaborative consulting, research, and operation of the center.

Staff: Tommy Wright (x31702), Robert Creecy, Eric Slud, Michael Hawkins, Sarah James, Michael Leibert, Gloria Prout, Esan Sumner, Sarah Wilson, Kelly Taylor

B. Research Computing
Description: This ongoing project is devoted to ensuring that Census Bureau researchers have the computers and software tools they need to develop new statistical methods and analyze Census Bureau data.

Highlights: During FY 2011, center staff, in cooperation with the IT Directorate, completed the blade migration, replacing the SGI Altix 3700 bx2 (research1) with a new cluster of IBM blade servers (research2). The new system went into production in January, and the old system was subsequently decommissioned. We continued to work as members of the Data Management Requirements and Evaluation Team (DMRET) on the development of a pilot system ("DMC Pilot") which demonstrated an automated process of authorizing projects using agency data and the provisioning of computing resources and data for those projects. The pilot system was developed by a contractor (Dovel Technologies) and was delivered at the end of March. After the initial delivery, the DMRET members performed acceptance testing of the system on the Center for Applied Technologies lab network, after which the pilot was implemented on the Census internal network for testing by a wider range of users. The report documenting the results of the pilot is expected early in FY 2012. Staff also assisted CSVD with the implementation of PBSPro, a job scheduler that is currently used on research2, on the CES research cluster (research1).

Staff: Chad Russell (x33215)

2. RESEARCH

2.1 – 2.2 GENERAL RESEARCH AND SUPPORT TOPICS
(Census Bureau Projects 0351000, 1871000)

Statistical Methodology

A. Disclosure Avoidance for Microdata
Description: Our staff investigates methods of microdata masking that preserves analytic properties of public-use microdata and avoid disclosure.

Highlights: Staff refereed one paper on microdata confidentiality for the *Journal of Official Statistics* and another paper for *Statistical Data Protection 2011*. Staff gave a Center for Statistical Research and Methodology (CSRM) seminar titled *General Modeling/Edit/ Imputation Methods and Their Application to Producing Synthetic Data with Reduced/Eliminated Re-identification Risk.*

Staff sent information related on general confidentiality and differential privacy to Avi Singh of the National Opinion Research Center (NORC). Staff provided some pointers and comments about papers on cryptographic privacy to staff at Microsoft Research and colleagues at Carnegie-Mellon University.

Staff provided modeling/edit/imputation software, confidentiality software, and record linkage software to the IAB institute in Nuremberg, Germany. The software had been used during two short courses at the University of London in September 2010 and a CSRM seminar in October 2010.

Staff made detailed comments on a draft paper on microdata confidentiality for three individuals in the Center for Disclosure Avoidance Research (CDAR). Staff made comments on a draft report on microdata confidentiality of ACS transportation data that had been done for the Transportation Board at the National Academies of Science. The comments related to the lack of description of the specific randomization and sampling procedures, to whether the joint distributions of the masked data satisfied valid analytic properties, to the lack of controlled rounding procedures for assuring that the microdata added to published tables from the original, unmasked microdata, and to the extreme difficulty to write software implementing the ideas.

Staff developed some new algorithms for exceedingly fast pps sampling of models for producing synthetic microdata with valid analytic properties. One component of the methods mimics binary search. Another component uses sophisticated indexing to search an enormous contingency table for cells with appropriate properties and almost instantly build the structure for the binary search. The methods are somewhat related to, but much less sophisticated than, methods given by Eric Slud in a CSRM seminar. Professor Slud needs additional analytic properties that are not needed in the synthetic-data-generation situation.

Staff presented the talk "Providing valid analytic properties in a synthetic version of confidential discrete data while nearly eliminating re-identification risk" at the 4th Workshop on Privacy and Confidentiality in Nuremberg, Germany. Staff provided the latest version of the modeling/edit/imputation software to researchers at the University of Duisburg-Essen, Cornell University, and the Census Bureau. Cornell University was interested in the software because of its extreme speed and power in comparison to commercial software (500,000 cells with epsilon $<10^{-12}$ in 200 iterations in less than 60 seconds; 500 million cells in 1000 minutes).

Staff provided comments to faculty at the University of Southampton related to ways to research methods on controlled rounding and the associated computational algorithms in 3 or more dimensions. The issues are related to having tables of synthetic microdata that, when rounded, agree with published tables. Although there has been considerable research for controlled rounding procedures in three or more dimensions, the only methods currently are in *Center for Statistical Research Report Series #RRS2009-08*. The subtle rounding issues in the algorithms for the capacitated lp procedures in the paper are highly non-trivial.

Staff: William Winkler (x34729), William Yancey

B. Seasonal Adjustment
[See Economic Project 2320052]

C. Household Survey Design and Estimation
[See Project 5385160 (D), Decennial Directorate – American Community Survey (ACS)]

D. Survey Productivity and Cost Analysis
Description: The Survey Productivity and Cost Analysis (SPCA) Group has been established as a cross-directorate analytic team to conduct methodological research toward the goal of continuous improvement in survey operational efficiency. The group will both initiate and respond to issues related to survey performance indicators including cost, data quality, and data collection progress, as they relate to survey design. Our Center is represented on this team along with staff from the Research and Methodology Directorate, the Demographic Programs Directorate, the Decennial Directorate, the Center for Economic Studies, the Field Directorate, and the Center for Survey Measurement.

Highlights: In FY 2011, the SPCA team gathered Contact History Instrument observations from the Current Population Survey and merged these with Cost and Response Management Network records, Windows Field Representative Earnings Data, and demographics from

the Tract-Level Planning Database with Census 2000 data. With this collection of information, we developed prototype charts and tables for survey supervisors to use in the management of field operations. In addition, we have examined a number of discrete semi-parametric hazard models for estimating the probability that an interview is obtained on a particular visit.

Staff: Chandra Erdman (x31235), Julie Tsay

E. Sampling and Estimation Methodology: Economic Surveys
Description: The Economic Directorate of the Census Bureau encounters a number of issues in sampling and estimation in which changes might increase the accuracy or efficiency of the survey estimates. These include estimates of low-valued exports not currently reported, alternative estimation for the Quarterly Financial Report, and procedures to address nonresponse and reduce respondent burden in the surveys. Further, general simulation software might be created and structured to eliminate various individual research efforts. An observation is considered influential if the estimate of total monthly revenue is dominated by its weighted contribution. The goal of the research is to find methodology that uses the observation but in a manner that assures its contribution does not dominate the estimated total or the estimates of period-to-period change.

Highlights: Staff collaborated in FY 2011 with staff from the Economic Directorate in the implementation of a newly designed simulation to investigate the properties of two methods, M-estimation and Clarke Winsorization, which detect and treat influential values. The team designed the simulated population to look like 20 consecutive months of the population for the Monthly Retail Trade Survey. The simulations for the first scenario induced one influential value in Month 4 and then drew replicates until 200 contained the influential value. This strategy permitted unconditional and conditional analyses to examine the statistical properties of both methods. Staff presented the results of the first scenario where the influential value was unusually high in "Detecting and Treating Influential Values in the Monthly Retail Trade Survey via M-estimation & Clarke Winsorization." at the Statistics Canada/U.S. Census Bureau Interchange held in Suitland, MD on March 28-29. Further work is underway to investigate the statistical properties of the methods in other scenarios for the occurrence of influential values. Staff is also working with a team in Governments Division to investigate methodology for detecting and treating influential values in their surveys.

Staff: Mary Mulry (x31759)

F. Research and Development Contracts
Description: The Research and Development Contracts are indefinite delivery, indefinite quantity task order contracts for the purpose of obtaining contractor services in highly technical areas to support research and development activities across all Census Bureau programs. The contracts provide a pool of contractors to assist the Census Bureau in conducting research on all survey and census methods and processes to improve our products and services. The prime contractors include educational institutions, university supported firms, and privately owned firms that concentrate in sample survey research, methodology, and applications to create a pool of specialists/experts to tackle some of the Census Bureau's most difficult problems through research. Many of the prime contractors are teamed with one or more organizations and/or have arrangement with outside experts/consultants to broaden their ability to meet all of the potential needs of the Census Bureau. These five-year contracts allow Census Bureau divisions and offices to obtain outside advisory and assistance services to support their research and development efforts quickly and easily.

R&D 2007 Contracts
During FY 2011, twenty-eight (28) modifications were awarded and nine were completed. To date, there have been ninety-six (96) task orders awarded under the R&D 2007 contracts, with a monetary value of over $128 million. Eighty-eight (88) task orders have been completed and one task order terminated, leaving seven active tasks.

R&D 2014 Contracts
During FY 2011, sixteen (16) new task orders were awarded, forty-four (44) were modified and seventeen (17) were completed. To date, there have been forty-seven (47) task orders awarded under the R&D2014 contracts with a monetary value of $38 million. Nineteen (19) task orders have been completed, leaving twenty-eight (28) active task orders.

Staff: Ann Dimler (x34996), Christina Cooper

G. Small Area Estimation
Description: Methods will be investigated to provide estimates for geographic areas or subpopulations when sample sizes from these domains are inadequate.

Highlights:
Aggregate Level Small Area Estimation Modeling
Staff continued to explore new ways to estimate the probability of correct enumeration in the census using small area techniques. This work involves incorporating random effects into the model and accounting for clustering in the sample. Staff further investigated how to modify the design weights to construct a new quasi-likelihood with certain desirable properties, and to avoid potential problems, such as modified weights being set to zero or tending toward infinity. Staff has derived first order Taylor approximations for the replicate weight variance estimator used in the ACS. These approximations give an insight on how to build models which require specifications of means and variances of

the survey variance estimators. Prior specification of the variance of the survey variance estimator was determined by empirical results. This approach leads to a more theory driven model. The staff has started an evaluation to determine how well this approach works for areas with small sample sizes since the Taylor approximations are typically more accurate for large sample sizes. The results from this work will feed into the SAIPE county-level poverty rate model research.

Estimates from Distinct Geographic Small Area Models
Many small area programs use distinct models to make estimates at different geographic levels; for example in the SAHIE and SAIPE programs, estimates for states and counties are made independently from two distinct models. When this method is used, it is often assumed that certain estimates from different models are independent, or uncorrelated. Staff investigated ways to measure whether this assumption is valid. Staff has been investigating several different model approaches and a new approach has been developed where both the smaller level area estimates and the larger area estimate are shrunk towards each other.

Prior Distributions for Small Area Models
Small area models make use of area specific covariates and random effects which attempt to explain variation beyond that which is explained by auxiliary variables. Since there is rarely substantial prior information, non-informative or diffuse priors on the random effects are often used. However, some non-informative priors can lead to improper posterior distributions. Staff investigated the impact of different priors for a wide class of multivariate generalized linear mixed models for small area estimation. This class of models includes the multivariate logistic regression model, which is similar to the model used in the SAHIE program. Necessary and sufficient conditions were found for the propriety of the posterior distribution. Future work involves fitting the model to data sets and conducting simulation studies. Also, it is of interest to test the effect of using vague, but proper prior distributions.

This research was presented "Multivariate Generalized Linear Models for Small-Area Estimation" as a talk and proceedings paper at the Joint Statistical Meetings in Miami Beach, Florida, July 30-August 4, 2011.

Staff: Jerry Maples (x32873), Aaron Gilary, Elizabeth Huang, Ryan Janicki

H. Web Scraping Feasibility Investigation
Description: The goal of this project is to investigate the feasibility of developing and implementing a Web scraping tool. This tool will collect publicly available information posted by businesses. Knowledge of this auxiliary information may be useful in improving estimates with economic data.

Highlights: In FY 2011, staff built a time series sampled weekly of a weighted basket of products closely comprising the list of particular goods used to compute the Consumer Price Index. However, this approach utilizes Web scraping to extract the sampled data. Staff has written software in Python to scrape sample outlets that have Web sites. Staff wrote a web scraper in Python for three different online grocery stores in the US for extracting prices of any good in the store. Staff also prepared and presented a talk with the Bureau of Labor Statistics on the project.

Staff: Chris Blakely (x31722)

Statistical Computing Methodology

A.   Record   Linkage   and   Analytic   Uses   of Administrative Lists
Description: Under this project, staff will provide advice, develop computer matching systems, and develop and perform analytic methods for adjusting statistical analyses for computer matching error.

Highlights: In FY 2011, staff distributed a new version of the *BigMatch* software to professors and graduate students at the University of London, University of Michigan, UCLA, Berkeley and New York University. Professors John McDonald (London), Trivellore Raghunathan (Michigan), and Jim Pitman (Berkley) and their students are looking at some of the methods. While maintaining the accuracy of earlier record linkage software, *BigMatch* continues as 40+ times as fast as experimental parallel software from Stanford, Pennsylvania State University and Australia National University. Staff wrote a short document on background and issues for Professor Jim Pitman of the University of California   Berkeley   Statistics   Department.   The background included a list of papers on record linkage, some of the subtleties of the data being considered in Pitman's project, and a suggestion to contact two of the CS professors at Berkeley who had worked on a very similar class of problems.

Staff visited with Professor Leicester Gill of Oxford University to discuss record linkage issues, where newly developed software methods and two very large PCs (each with two core I7 processors and 32 megabytes of RAM) for processing files were demonstrated. With the Linux machine, he was able to match the blinded (one way encrypted versions) of the UK National Health Files against themselves (80 million x 80 million) in less than 12 hours. Staff provided comments related to string comparators and file preprocessing to Professor Anna Ferrante of the University of Western Australia who is directing a large comparative study of record linkage software   in   the   health   sciences   area.   The   string comparators are used for dealing with typographical error. The file preprocessors were the current versions of

the Geography Division address standardizer and the SRD name standardizer.

Staff gave the talk "Record Linkage: Introductory Overview" at the UMBC Statistics Seminar on February 11, 2011. Staff met with three individuals from HHES to discuss issues with very large record linkage with *BigMatch*. HHES is considering adding information from additional files to their LEHD data base that is widely available to approved researchers in the RDCs. Staff had earlier provided background material and comments to the HHES-proposed linkage project. Staff provided a number of background documents, advice, and software to Selvaratnam Sridharma and David Chapman in the Economic Directorate. Most of the advice was related to nuances of cleaning up and standardizing addresses prior to using them in the matching software.

Staff continued investigating generalizations of the Birthday and Collision Problems, and staff continued studying general parallel algorithms and the more specific algorithms needed for fast computation using graphics processors. Staff studied Markov Chain Monte Carlo methods in the context of possible extensions of error analysis and correction of analyses in sets of files that have been combined via record linkage. Staff also began learning R-software that had been used and disseminated by Ray Chambers.

Staff distributed *BigMatch* software to the University of Duisburg-Essen and the Institut für Arbeitsmarkt- und Berufsforschung. Staff also gave the talk "Record Linkage: Introductory Overview" and a *BigMatch* demonstration at the Institute. Staff presented the paper "Record Linkage" in the session "Best from Wiley Interdisciplinary Studies: Computational Statistics" at the Interface Conference on June 3. Staff served as the discussant for three papers at the Conference on Person Validation and Entity Resolution in Washington, D.C. in May.

Staff completed the paper "Machine Learning and Record Linkage" for an invited paper session at the International Statistical Institute meetings. Staff completed the paper "Cleaning and using administrative lists: Enhanced practices and computational algorithms for record linkage and modeling/editing/imputation" and associated slides for an invited JSM (Joint Statistical Meetings) session on administrative lists. Staff wrote the chapter/entry "Record Linkage" for *Encyclopedia of Environmetrics*.

Staff met with Dr. Ivan Fellegi regarding record linkage in the Fellegi-Sunter model and edit/imputation in the Fellegi-Holt model.

Staff: William Winkler (x34729), William Yancey, Ned Porter

B.1 Editing
Description: This project covers development of methods for statistical data editing. Good methods allow us to produce efficient and accurate estimates and higher quality microdata for analyses.

Highlights: During FY 2011, staff continued research on developing selective editing strategies for our foreign trade data. Staff developed score functions that are a combination of two terms: one term measuring how suspicious a record is and a second term measuring the impact the suspicious record has on cell estimates. The measure of suspicion is adapted from the Hidiroglou-Berthelot method for detecting outlying observations in periodic historical data. The measure of impact is calculated as the effect the difference between the reported value and an estimate of the true value had on the total value within the domain (e.g., commodities and commodity groupings). Staff completed a feasibility study on the application of selective editing to the checking and correction phase of foreign trade data processing. Staff used the absolute pseudo-bias as defined by Latouche and Berthelot (1992) for evaluating the results of the feasibility study. Plots of the pseudo-bias for selected testing cells show the pseudo-bias rapidly decreasing as the percentages of records flagged for review increases. This implies that we could stop reviewing records when the pseudo-bias approaches zero and the estimated selective editing total approaches the final publication total.

Staff: María García (x31703)

B.2 Editing and Imputation
Description: Under this project, our staff provides advice, develops computer edit/imputation systems in support of demographic and economic projects, implements prototype production systems, and investigates edit/imputation methods.

Highlights: Staff is reviewing one paper for publication in a scientific journal. Another one is being submitted for publication. In one paper, staff documented the new methodology for imputing group quarters on the ACS sampling frame for counties that do not have group quarters in sample. In another paper staff presented a method to improve cross-sectional finite population predictions in longitudinal surveys by merging information from different cross-section, even though the prediction is only for one of the cross-section.

Staff: Yves Thibaudeau (x31706), Chandra Erdman, María García, Martin Klein, Ben Klemens, Rolando Rodriguez

C. Missing Data and Imputation: Multiple Imputation Feasibility Study
Description: Methods for imputing missing data are closely related to methods used for synthesizing sensitive items for disclosure limitation. One method currently

applied to both issues is multiple imputation. Although the two issues may be addressed separately, techniques have been developed that allow data users to analyze data in which both missing data imputation and disclosure limitation synthesis have been accomplished via multiple imputation techniques (e.g., synthetic data). This project ascertains the effectiveness of applying multiple imputation to both missing data and disclosure limitation in the American Community Survey (ACS) group quarters data. Statistical models are used to generate several synthetic data sets for use within the multiple-imputation framework.

Highlights: During FY 2011, staff made improvements to the R package 'TEA' in support of several sponsored projects. Improvements include an improved interface for defining consistency rules (structural zeroes) and support for more statistical models. Staff utilized these improvements in two completed projects: disclosure avoidance for group quarters in the 2010 Decennial Census and disclosure avoidance for group quarters for the ACS 2010 sample. Later in the year, improvements to the package's data fingerprinting functions were made, and are currently being applied to test code for disclosure avoidance for group quarters for the 2010 Census of Island Areas.

Staff: Rolando Rodríguez (x31816), Ben Klemens, Yves Thibaudeau

D. Modeling, Analysis, and Quality of Data
Description: Our staff investigates methods of the quality of microdata primarily via modeling methods and new software techniques that accurately describe one or two of the analytic properties of the microdata.

Highlights: In FY 2011, staff continued working on methods of analysis on sets on files that had been linked with record linkage error. Staff refereed one paper for *Biometrika*.

Staff distributed a new version of the modeling/edit/imputation software to students from the course *Cleaning Administrative Data: Improving Quality with Edit and Imputation* at the University of London and to professors and their graduate students at the University of Michigan, University of London, and Carnegie-Mellon University. Staff provided the same software and additional background information with pointers to related research papers to several individuals in CSRM. The background included a pointer to the details of the algorithms for computing the likelihood in *SRD Report Series: Statistical Research Report #RRS2010-08*, additional details of the new computational algorithms, and some issues related to preserving joint distributions while satisfying edits.

Staff provided background information on text categorization to the Associate Director and Michael Kornbau of the Economic Directorate. Text categorization methods are used to convert textual material to NAIC codes for companies and job-classification codes for individuals.

Staff provided a short review and pointers to papers related to text categorization comments from the Associate Director to Michael Kornbau. Staff felt that the methods of feature extraction and use (from transferring general text to dictionaries) and dimensionality reduction introduced by Kornbau were likely to be as useful as alternative methods in the machine learning literature. Staff attended a meeting with members of HHES and Michael Kornbau and his colleagues related to a text categorization system that they are developing for HHES.

Staff sent background information on text categorization and support vector machines to Ric Clarke the Australia Bureau of Statistics who will participate in an invited International Statistical Institute (ISI) session on machine learning methods the staff member. In particular, staff sent a very widely-referenced but difficult-to-obtain tutorial on Solaris Volume Manager (SVM) by Christopher Burgess of Microsoft Research, written when he was at Bell Labs. Staff sent background material related to text categorization and a talk that had been given at the Australia Bureau of Statistics (ABS) in April 2005 to Dennis Farrell and Ric Clarke of the ABS.

Staff sent a number of detailed comments to an Associate Director of the National Institute of Statistical Science regarding methods for edit/imputation of economic data and on certain aspects of microdata confidentiality. The comments on editing were related to the efficacy of the SPEER routines in the Plain Vanilla Edit System used during the last three Economic Censuses (same quality as Statistics Canada's General Edit and Imputation System (GEIS), 60 times as fast (allowing us to run the largest Economic Census in eight hours in 1997), and the portability of the software (demonstrated on three different surveys on three different machines: VAX, Unix workstation, Windows PC) in two hours by changing parameter files and edit tables successively, copying, recompiling, and re-running on a second machine and finally repeating the process on a third machine. Also, until very recently, the Structured Programs for Economic Editing and Referrals (SPEER) routine was the only method that would assure that the 'corrected' records would satisfy edits (both ratio edits and simple linear inequality edits) and balance equations (allowing items to add to totals). Systems at Statistics Canada, Statistics Netherlands, and Italy's Istituto Nazionale di Statistica (ISTAT) were unable to assure that items added to totals, although LP methods for assuring the restraints were known to the respective agencies but not the much faster algorithms developed at the Census Bureau. Staff mentioned that the new modeling/edit/imputation methods assure that edited data satisfy both edit restraints and joint distributions simultaneously and allow generation of synthetic microdata with suitable analytic

properties and reduced/eliminated re-identification risk. Key aspects of the new methods are the ability to scale aggregates from the microdata to external totals from benchmarks, to place upper and lower bounds on all cells to reduce re-identification risk, the ability to still preserve analytic properties, and exceedingly fast algorithms for imputation, searching contingency tables for matching cells, and pps sampling.

At the request of Nancy Gordon, staff sent considerable background information on the Fellegi-Holt model of edit/imputation and related computational methods to Gilberto Cavillo who heads a research group in Mexico. The material was a subset of the material mentioned in earlier paragraphs.

At the request of Michael Hidiroglou, staff sent our Winkler-Hidiroglou paper on developing analytic programming capability to Jean-Francois Beaumont of Statistics Canada. Staff also sent a short document on issues with developing exceedingly fast software to Hidiroglou and Beaumont. In particular, the document describes the development of (1) new record linkage parameter estimation methods for the 1990 Decennial Census that were 220+ times as fast as the original SAS version and that allowed processing in less than twenty minutes in each of the 457 areas being matched, (2) the SPEER algorithms that produced equal quality results to Statistics Canada's GEIS and were sixty times as fast allowing production editing in less than eight hours in 1997, (3) the 2010 *BigMatch* that is more than forty times as fast as recent parallel software that allows us to process 10^{17} pairs (300 million x 300 million) on forty CPUs on an SGI Linux machine in less than thirty hours and is the minimum speed needed for Decennial production processing, and (4) new modeling/edit/imputation software that has several algorithms that are one hundred times as fast as those in commercial software and experimental university software. The fast algorithms allow us to do modeling and production in situations that were previously impossible.

Staff provided comments to statistical analysis on very large data sets and software development issues to two individual in the Governments Division. In particular, staff described naming conventions (and the types of changes associated with them) for various versions of generalized software. Generalized software can be used on differing files by changing parameter files. Examples are SAS regression, SRD record linkage, *BigMatch*, SPEER, the generalized name and address standardization software, and the general modeling/edit/imputation software. The specifics were based on methods used in the development of the production record linkage software for the 1990 Decennial Census and versions that we have given out internationally since January 1990. Staff provided background on machine learning to an individual in the Decennial Directorate. In particular, staff referenced the 2007 monograph that greatly generalizes some of the graphical methods used

by statisticians for local computations with probabilities on graphical structures and their application to expert systems.

Staff made a number of comments related to maximum overlap sampling to staff in DSMD. In particular, staff made comments on using bootstrap for estimating variances. Larry Ernst in a review of Rottach's work had mentioned that they needed to estimate variances. In any early version of the multi-way and multi-purpose paper, staff had used bootstrap but later managed a more elegant solution (and computationally much faster solution) using theory due to J. N. K. Rao's *Unbiased Variance Estimation for Multistage Designs.* Staff mentioned the total unsuitability of Horvitz-Thompson estimators for consideration in the Rottach's work or in the multi-sampling paper. The unsuitability of Horvitz-Thompson methods (including substantial bias and negative variances in some situations) are known to a number of sampling experts.

Staff sent a short document on issues with developing exceedingly fast software to researchers at various universities and government research institutions. The document describes the development of the 2010 *BigMatch* that is more than 40 times as fast as recent parallel software that allows us to process 10^{17} pairs (300 million x 300 million) on 40 CPUs on an SGI Linux machine in less than 30 hours and is the minimum speed needed for Decennial production processing, and new modeling/edit/imputation software that has several algorithms that are 100 times as fast as those in commercial software and experimental university software. The fast algorithms allow us to do modeling and production in situations that were previously impossible. The exceedingly fast methods, for the first time, allow processing large administrative list situations and will be presented in the JSM invited paper session on administrative lists.

Staff: William Winkler (x34729), Rob Creecy, William Yancey, María García

E. Ranking Methodology Development and Evaluation
Description: This project undertakes research into the development and evaluation of statistical procedures for using sampled data to rank several populations with respect to a characteristic of interest. The research includes an investigation of methods for quantifying and presenting the uncertainty in an estimated ranking of populations. As an example, a series of ranking tables are released from the American Community Survey in which the fifty states and the District of Columbia are ordered based on estimates of certain quantities of interest.

Highlights: In FY 2011, staff: (1) proposed several Bayesian procedures for obtaining a complete ranking of

normal populations with respect to the unknown means when the variances are known; (2) developed efficient computational methods for implementation of each procedure; (3) used simulation to compare the procedures with the Census Bureau's current methodology for producing ranking tables based on data from the American Community Survey; (4) applied each procedure to a data set from the American Community Survey and compared the results; and (5) used a parametric bootstrap to estimate a frequentist-based probability that can be used to quantify the uncertainty in a released ranking. Staff published an article based upon this work, and presented the work in some professional meetings and academic settings. Staff interacted with students in The REU Site: Interdisciplinary Program in High Performance Computing in the Department of Mathematics and Statistics, University of Maryland, Baltimore County. This interaction resulted in a study of the application of parallel computational methods to the proposed ranking procedures.

Staff: Martin Klein (x37856), Tommy Wright

F. Synthetic Survey and Processing Experiments
Description: To improve operational efficiencies and reduce costs of survey processing, this project will simulate a survey, in which an artificial team of interviewers seek out an artificial set of respondents, to test alternative methods of allocating resources in the field and to test alternatives for the post-processing of the gathered survey data.

Highlights: In FY 2011, staff developed a simulation for the process of operating a survey in the field and post-processing the survey to produce output statistics. We use TEA, a system developed under Missing Data and Imputation: Multiple Imputation Feasibility Study [see project in Research, Statistical Computing Methodology (C)] to impute values for the surveys that could not be gathered. For each question on the survey, the simulation outputs measures of standard errors and bias relative to the true value for each question, and the experiment also outputs measures of total survey costs. These outputs can be used to compare different methods for allocating interviewers and different methods of post-processing.

Staff: Ben Klemens (x36864)

Research Support and Assistance
This staff provides substantive support in the conduct of research, research assistance, technical assistance, and secretarial support for the various research efforts.

Staff: Sarah James, Gloria Prout, Esan Sumner, Kelly Taylor

3. PUBLICATIONS

3.1 JOURNAL ARTICLES, PUBLICATIONS

Alexandrov, T., Bianconcini, S., Dagum, E., Maass, P., and McElroy, T. (In Press). "The Review of Some Modern Approaches to the Problem of Trend Extraction." *Econometric Reviews*.

Andridge, R. H. and Little, R. J. (2011). "Proxy Pattern-Mixture Analysis for Survey Nonresponse." *Journal of Official Statistics* 27, 2, 153-180.

Blakely, C. (In Press). "Extracting Intrinsic Modes in Stationary and Nonstationary Time Series Using Reproducing Kernels and Quadratic Programming." *International Journal of Computational Methods*, 8, 3.

Cain, K.C., Harlow, S., Little, R.J., Nan, B., Yosef, M., Taffe, J., and Elliott, M. R. (2011). "Bias Due to Left Truncation and Left Censoring in Longitudinal Studies of Developmental Processes." *American Journal of Epidemiology* 173, 9, 1078-1091.

Findley, D. F., Monsell, B. C., and Hou, C.-T. (In Press). "Stock Series Holiday Regressors Generated from Flow Series Holiday Regressors." *Statistica Sinica.*

Giusti, C. and Little, R.J. (2011). "A Sensitivity Analysis of Nonignorable Nonresponse to Income in a Survey with a Rotating Panel Design." *Journal of Official Statistics,* 27, 2, 211-229.

Guo, Y. and Little, R.J. (2011). "On Using Summary Statistics from an External Calibration Sample to Correct for Covariate Measurement Error." To appear in *Epidemiology.*

Guo, Y. and Little, R.J. (2011). "Regression Analysis Involving Covariates with Heteroscedastic Measurement Error." *Statistics in Medicine,* 30, 18, 2278-2294.

Herzog, T. H., Scheuren, F., and Winkler, W. E. (2011). "Data Quality: A Case Study," *Wiley Interdisciplinary Reviews: Computational Statistics,* 3 (1), January/February, 12-21.

Jach, A., McElroy, T., and Politis, D. (In Press). "Subsampling Inference for the Mean of Heavy-Tailed Long Memory Time Series," *Journal of Time Series Analysis.*

Klein, M. and Wright, T. (2011). "Ranking Procedures for Several Normal Populations: An Empirical Investigation," *International Journal of Statistical Sciences,* 11, 37-58.

Li, Y., Taylor, J. M. G., and Little, R. J. (2011). "A Shrinkage Approach for Estimating a Treatment Effect Using Surrogate Marker Data in Clinical Trials." *Biometrics*.

Little, R. J. (2011). "Calibrated Bayes, for Statistics in General, and Missing Data in Particular (with Discussion and Rejoinder)." *Statistical Science,* 26, 2, 162-186.

Little, R. J., Yosef, M., Nan, B., and Harlow, S. (2011). "A Method for the Longitudinal Prospective Evaluation of Markers of a Subsequent Event (with Discussion and Rejoinder)." *American Journal of Epidemiology,* 173, 12, 1380-1387.

Little, R. J. and Zhang, N. (2011). "Subsample Ignorable Likelihood for Regression Analysis with Missing Data." *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 60, 591-605.

Ma, Y., Hart, J. D., Janicki, R., Carroll, R. J. (2011). "Local and Omnibus Goodness-of-Fit Tests in Classical Measurement Error Models." *Journal of the Royal Statistical Society, Series B*, 73, 81-98.

McElroy, T. and Holan, S. (In Press). "On the Estimation of Autocovariances for Generalized Gegenbauer Processes," *Statistica Sinica.*

McElroy, T. (2011). "A Nonparametric Method for Asymmetrically Extending Signal Extraction Filters." *Journal of Forecasting,* 30, 597-621.

McElroy, T. and Jach, A. (2011). "Tail Index Estimation in the Presence of Long Memory Dynamics." *Computational Statistics and Data Analysis,* 56, 266-282.

McElroy, T. and Politis, D. (2011). "Fixed-B Asymptotics for the Studentized Mean for Long and Negative Memory Time Series." *Econometric Theory,* 28, 1-11.

McElroy, T. and Trimbur, T. (2011). "On the Discretization of Continuous-Time Filters for Nonstationary Stock and Flow Time Series." *Econometric Reviews,* 30, 475-513.

Mulry, M. and Olson, T. (2011). "Analyses for the U.S. 2010 Census Partnership Program." *Social Marketing Quarterly,* 17, 1, 27-55.

Slud, E. and DeMissie, M. (In Press). "Validity of Regression Meta-analyses Versus Pooled Analyses of Mixed Effect Linear Models." *Mathematics in Engineering, Science and Aerospace: (Special Issue on Recent advances in Biostatistics, Survival Analysis, Medical Statistics and Related Topics).*

Slud. E. and Maiti, T. (2011). "Small-Area Estimation Based on Survey Data from a Left-Censored Fay-Herriot Model," *Journal of Statistical Inference and Planning,* 141, 3520-3535.

Slud. E. (2011). "Discussion of "Impact of Frequentist and Bayesian Methods on Survey Sampling Practice: A Selective Appraisal" by J. N. K. Rao" *Statistical Science,* 26, 262-265.

Wang, C., Little, R. J., Nan, B., and Harlow, S. (2011). "A Hot-Deck Multiple Imputation Procedure for Gaps in Longitudinal Recurrent Event Histories." *Biometrics.*

Zhang, G. and Little, R. J. (2011). "A Comparative Study of Doubly-Robust Estimators of the Mean with Missing Data." *Journal of Statistical Computation and Simulation, iFirst,* 2011, 1-20.

Zhang, N. and Little, R. J. (2011). "A Pseudo-Bayesian Shrinkage Approach to Regression with Missing Covariates." To appear in *Biometrics*.

3.2 BOOKS/BOOK CHAPTERS

McElroy, T. and Holan, S. (In Press). "The Error in Business Cycle Estimates Obtained from Seasonally Adjusted Data," in *Economic Time Series: Modeling and Seasonality.* Chapman-Hall.

Holan, S. and McElroy, T. (In Press). "On the Seasonal Adjustment of Long Memory Time Series," in *Economic Time Series: Modeling and Seasonality.* Chapman-Hall.

Winkler, W. E. (In Press), "Record Linkage," in *Encyclopedia of Environmetrics*, John Wiley.

3.3 PROCEEDINGS PAPERS

Joint Statistical Meetings, American Statistical Association, Vancouver, Canada, July 31 – August 5, 2010.
- Elizabeth Huang and William Bell, "Further Simulation Results on the Distribution of Some Survey Variance Estimators."
- Martin D. Klein and Robert H. Creecy, "Steps Toward Creating a Fully Synthetic Decennial Census Microdata File."
- Tucker McElroy, "Fixed-Bandwidth Asymptotics for the Studentized Mean for Long and Negative Memory Time Series."

- Eric Slud, Yang Cheng, and Carma Hogue, "Variance estimation for Decision-Based Estimators with Application to the Annual Survey of Public Employment and Payroll."
- Eric Slud, J. Suntornchost, and R. Wei, "Modeling U.S. Cause-specific Mortality Rates Using an Age-segmented Lee-Carter Model."
- Natalya Titova, David Findley, and Brian Monsell, "Comparing the Causey-Trager Method to the Dagum-Cholette Regression Method of Benchmarking Sub-annual Data to Annual Benchmarks."
- Ryan Janicki, "Estimation and Hypothesis Testing in Submodels using Fisher Estimating Functions."

2011 International Statistical Institute Proceedings, Dublin, Ireland, August 21-26, 2011.
- William E. Winkler, "Machine Learning and Record Linkage."
- McElroy, T. and Monsell, B. "Extension of X-12-ARIMA to Seasonally Adjust Mixed Frequency Stock and Flow Time Series."

3.4 CENTER FOR STATISTICAL RESEARCH & METHODOLOGY RESEARCH REPORTS
<http://www.census.gov/srd/www/byyear.html>

RR (Statistics #2010-08): William R. Bell. "Unit Root Properties of Seasonal Adjustment and Related Filters," December 21, 2010.

RR (Statistics #2010-09): Chandra Erdman and Chaitra H. Nagaraja. "Imputation Procedures for American Community Survey Group Quarters Small Area Estimation," December 21, 2010.

RR (Statistics #2011-01): Donald J. Malec, Julie H. Tsay, and Elizabeth T. Huang. "A Random Effect Approach to Protection Against Model Error in Logistic Models of Census Coverage," March 17, 2011.

RR (Statistics #2011-02): Jerry J. Maples. "Using Small Area Modeling to Improve Design-Based Estimates of Variance for County Level Poverty Rate Estimates in the American Community Survey," March 17, 2011.

RR (Statistics #2011-03): Ryan Janicki and Donald Malec. "A Partition Model for Analyzing Categorical Data Subject to Non-ignorable Non-response," March 17, 2011.

RR (Statistics #2011-04): Aaron Gilary. "Recursive Partitioning for Racial Classification Cells," March 31, 2011.

RR (Statistics #2011-05): Michael Ikeda. "Developing Guidelines Based on CVs for when Three-Year Estimates Can Be Used Instead of Five-Year Estimates in the American Com," April 20, 2011.

RR (Statistics #2011-06): Tucker S. McElroy and Scott H. Holan. "On the Computation of Autocovariances for Generalized Gegenbauer Processes," April 20, 2011.

3.5 OTHER REPORTS

Donald Brown, Chandra Erdman, Kirsten Ling, and Laurie Santos (2010), "Revealed Preferences for Risk and Ambiguity," *Cowles Foundation Discussion Paper, 1774*: 1-22.

4. TALKS AND PRESENTATIONS

The Washington Statistical Society, Washington, D.C., October 26, 2010.
- Eric Slud, "Simultaneous Calibration and Nonresponse Adjustment."

2010 International Methodology Symposium , Ottawa, Ontario, Canada, October 26-29, 2010.
- Bill Yancey, "The EM Algorithm and Record Linkage."

National Academy of Sciences Sixth Meeting of the Panel on Statistical Methods for Measuring the Group Quarters Population in the American Community Survey (ACS), Washington, D.C., February 17-18, 2011.
- Chandra Erdman, "Imputation Procedures for Group Quarters Small Area Estimation."
- Lynn Weidman, " Overview of Small Area Imputation Research for ACS Group Quarters" and "Donor and Imputation Counts for 5-year Estimates from the Group Quarters Test with ACS Data."

Statistics Canada, U.S. Census Bureau Methodology Interchange, Washington, D.C., March 29, 2011.
- Jerry Maples, "Using Small Area Modeling to Improve Design-Based Estimates of Variance for County Level Poverty Rate Estimates in the American Community Survey."

Statistics Department, University of Maryland at Baltimore County, February 11, 2011.
- William E. Winkler, "Record Linkage: Introductory Overview."

Statistics Department, University of Missouri, Columbia, April 7, 2011.
- Tucker McElroy, "Noah-Joseph Models for Economic Time Series: Incorporating Heavy Tails and Long Range Dependence."

Statistics Seminar, Department of Mathematics and Statistics, University of Maryland, Baltimore County, April 8, 2011.
- Martin Klein, "Imputation for Nonmonotone Nonresponse in the Survey of Industrial Research and Development."

The International Statistical Association (IISA) Conference on Probability, Statistics, and Data Analysis, Raleigh, North Carolina, April 21-24, 2011.
- Tommy Wright, "An Early Step in Assessing an Estimate's Accuracy in the Presence of Competing Estimates."

5th Annual Probability and Statistics Day at University of Maryland Baltimore County, Baltimore, Maryland, April 23, 2011.
- Martin Klein, "Ranking Procedures for Several Normal Populations: An Empirical Investigation."

Person Validation and Entity Resolution Conference, L'Enfant Plaza Hotel, Washington, DC, May 16, 2011.
- William E. Winkler, discussant of three invited papers in the session "Dynamic Processing."

2011 Interface Conference, Session: Best of Wiley Interdisciplinary Reviews: Computational Statistics, Cary, North Carolina, June 3, 2011.
- William E. Winkler, "Record Linkage."

39th Annual Meeting of the Statistical Society of Canada, Wolfville, Nova Scotia, June 12-15, 2011.
- Martin Klein, "Ranking Procedures for Several Normal Populations: An Empirical Investigation."

2011 International Total Survey Error Workshop. Quebec, Quebec, Canada, June 19-22, 2011.
- Mary H. Mulry, "Overview of Error Model for Estimates of Foreign-born Immigration Using Citizenship and Residence One Year Ago from the American Community Survey."

The REU Site: Interdisciplinary Program in High Performance Computing, Department of Mathematics and Statistics at University of Maryland Baltimore County, June 30, 2011.
- Martin Klein, "Ranking Procedures for Several Normal Populations: An Empirical Investigation." (Invited talk)

Institut für Arbeitsmarkt- und Berufsforschung, Nuremberg, Germany, June 29, 2011.
- William E. Winkler, "Record Linkage: Introductory Overview."

Institut für Arbeitsmarkt- und Berufsforschung, 4th Workshop on Privacy and Confidentiality, Nuremberg, Germany, June 30, 2011.
- William E. Winkler, "Providing Valid Analytic Properties in a Synthetic Version of Confidential Discrete Data while Nearly Eliminating Reidentification Risk."

Statistics 2011 Canada / IMST 2011 - FIM XX, Montreal, Canada, July 4, 2011.
- Ryan Janicki, "A Bayesian Model Averaging Approach to Analyzing Categorical Data Subject to Non-ignorable Nonresponse."

Joint Statistical Meetings, American Statistical Association, Miami Beach, Florida, July 30-August 4, 2011.
- Aaron Gilary, "Recursive Partitioning for Racial Classification Cells."
- Ryan Janicki, "Selection of Prior Distributions for Multivariate Small Area Models with Application to Small Area Health Insurance Estimates."
- Martin Klein and Jun Shao, "Imputation for Nonmonotone Past-Value-Dependent Nonresponse in Longitudinal Studies with Application to the Survey of Industrial Research and Development."
- Donald Malec and Ryan Janicki, "Bayesian Predictive Inference for Small Areas with Uncertain Unit-Level Models."
- Mary Mulry, "Designing Estimators of Nonsampling Errors in Components of Census Coverage Error."
- Eric Slud, "Resampling Variance Estimation Methods for Decision-Based Stratified Regression Estimates."
- Yves Thibaudeau and Eric Slud, "Simultaneous Calibration and Nonresponse Adjustment with Explicit Variance Formulas."
- William Winkler, "Cleaning and Using Administrative Lists: Enhanced Practices and Computational Algorithms for Record Linkage and Modeling/Editing/Imputation."

Mathematics Department, Institute for Data Analysis and Process Design, Winterthur, Switzerland, June 11-19, 2011.
- Tucker McElroy, "Signal Extraction for Nonstationary Multivariate Time Series with Applications to Trend Inflation."

Conference on Small Area Estimation, Trier, Germany, August 11-13, 2011.
- Jerry Maples, "Using Small Area Modeling to Improve Design-based Estimates of Variance for County Level Poverty Estimates in the American Community Survey."

World Congress of Statistics, International Statistical Institute, Dublin, Ireland, August 21-26, 2011.
- William Winkler, "Machine Learning and Record Linkage."
- Tucker McElroy, "Model Estimation, Prediction, and Signal Extraction for Nonstationary Stock and Flow Time Series Observed at Mixed Frequencies."

2011 NBER-NSF Time Series Conference, East Lansing, Michigan, September 17, 2011.
- Tucker McElroy, "Distribution Theory for the Studentized Mean for Long, Short, and Negative Memory Time Series."

5. CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY SEMINAR SERIES

William Winkler, CSRM, U.S. Census Bureau, "General Modeling/Edit/Imputation Methods and Their Application to Producing Synthetic Data with Reduced/Eliminated Re-Identification Risk," October 13, 2010.

Roderick Little, U.S. Census Bureau, "Bayes, Models and Design-Based Inference from Complex Surveys," October 26, 2010.

Andrew Gelman, Columbia University, "Why We (Usually) Don't Worry about Multiple Comparisons," October 29, 2010.

Eric Slud, University of Maryland, College Park and CSRM, U.S. Census Bureau, "Symmetric 'Rejective' Probability Proportional to Size Sampling," November 2, 2010.

Adrian Raftery, University of Washington, "Probabilistic Projection of HIV/AIDS Prevalence Using Bayesian Melding," November 15, 2010.

Kimberly Henry, Internal Revenue Service, "Using Statistics of Income (SOI) Sample Data to Reduce Nonsampling Error in State-Level IRS Estimates," November 16, 2010.

Chris Blakely, CSRM, U.S. Census Bureau, "uSim-X13: A Graphical User-Interfaced Time Series Modeling and Simulation Environment Using X-13," December 9, 2010.

Eitan Greenshtein, Center of Statistical Bureau (Israel), "Empirical Bayes in the Presence of Explanatory Variables with Application to Spatio-Temporal Data and Small Area Estimation," December 13, 2010.

Yves Thibaudeau, CSRM, U.S. Census Bureau, "Merging Information from Auxiliary Variables and Survey Responses through Modeling to Improve Finite Population Prediction," February 16, 2011.

Roderick Little, U.S. Census Bureau, "A Modeler's View on Total Survey Error," February 23, 2011.

Roberto Rigobon, MIT Sloan & NBER, "The Distribution of the Size of Price Changes," March 8, 2011.

Daniel Adrian, Iowa State University, "Statistical Analysis of Complex-Valued Functional MRI Data," March 30, 2011.

Peter Meyer, U.S. Bureau of Labor Statistics, "Statipedia: A Platform for Collaboration Across Statistical Agencies," March 31, 2011.

Roderick Little, U.S. Census Bureau, "Statistical Analysis with Missing Data, Planned and Unplanned," April 6, 2011.

Julian Chan, University of Utah, "Detecting Changes in Panel Data," April 12, 2011.

Peter Miller, Northwestern University, "From Rapport to Friendly Machines: Survey Measurement in Context," April 13, 2011.

Joseph Schafer, The Pennsylvania State University, "Average Causal Effects from Longitudinal Surveys when the Treatment Is a Latent Class," May 2, 2011.

Eric Slud, University of Maryland, College Park & CSRM, U.S. Census Bureau, "Quality Assessment of Zeroes in American Community Survey Tables", May 3, 2011.

Jon Krosnick, Stanford University & CSM, U.S. Census Bureau, "Measuring Intent to Participate and Participation in the 2010 Census and Their Correlates and Trends: Comparisons of RDD Telephone and Non-Probability Sample Internet Survey Data," May 18, 2011.

Jae Kwang Kim, Iowa State University, *SUMMER AT CENSUS*, "A Model-Assisted Approach to Combining Data from Two Independent Surveys," May 19, 2011.

Bimal Sinha, University of Maryland, Baltimore County & CDAR, U.S. Census Bureau, "Privacy Protection and Quantile Estimation from Noise Multiplied Data," May 25, 2011.

Snigdhansu Bhusan Chatterjee, University of Minnesota, *SUMMER AT CENSUS*, "Towards a Nonparametric Small Area Framework: The Extended Fay-Herriot Model and Beyond," May 31, 2011.

Taniecea Arceneaux, Princeton University, "The Global Resilience of Social Networks with Multiple Relations," June 1, 2011.

Dawn V. Nelson, Mathematica Policy Research, "Will Less be More? One Organization's Journey from Multiple Customized Surveys to Integrated Continuous Measurement," June 6, 2011.

Benjamin Kedem, University of Maryland, College Park, *SUMMER AT CENSUS*, "Integration of Information from Multiple Sources," June 7, 2011.

Stephen Guptill, U.S. Geological Survey (Retired), *SUMMER AT CENSUS*, "Quantifying the Quality of the MAF/TIGER Database, June 9, 2011.

Zhen-Qing Chen, University of Washington, *SUMMER AT CENSUS*, "Floating and Sinking?" June 13, 2011.

Zhaowei Hua, University of North Carolina at Chapel Hill, "Bayesian Analysis of Varying Coefficient Models and Applications to Biostatistics," June 14, 2011.

Peter Gottschalk, Boston College, *SUMMER AT CENSUS*, "Some Evidence on the Importance of Sticky Wages," June 16, 2011.

Derek Young, Bechtel Bettis, Inc. & The Pennsylvania State University, "Semiparametric Mixtures of Regressions and the "mixtools" Package," June 24, 2011.

Nan Astone, Johns Hopkins University Bloomberg School of Public Health, *SUMMER AT CENSUS*, "Work Marriage and Fatherhood: Findings from the NLSY79," July 11, 2011.

Bruce Meyer, University of Chicago, *SUMMER AT CENSUS*, "Errors in Survey Reporting and Imputation and Their Effects on Estimates of Food Stamp Program Participation," July 11, 2011.

Michael Omi, University of California, Berkeley, *SUMMER AT CENSUS*, "Colorblindness and the Contradictions of Racial Classification," July 12, 2011.

Howard Winant, University of California, Santa Barbara, *SUMMER AT CENSUS*, "Toward a New Racial Studies: A University of California System-wide Initiative," July 12, 2011.

Michael Omi, University of California, Berkeley; Howard Winant, University of California, Santa Barbara, *SUMMER AT CENSUS*, "Measuring Race & Ethnicity Discussion Panel," July 13, 2011.

Bikas Sinha, Indian Statistical Institute, *SUMMER AT CENSUS*, "Understanding Species' Abundance," July 14, 2011.

Thomas Louis, Johns Hopkins University, *SUMMER AT CENSUS*, "Bayesian Ranks, Histogram and Triple-Goal Estimates," July 18, 2011.

Carl Schmertmann, Florida State University, *SUMMER AT CENSUS*, "Improving ACS Estimates with Bayesian Models and Demographic Priors," July 19, 2011.

William Frey, Brookings Institution; John DeWitt, University of Michigan, *SUMMER AT CENSUS*, "An Overview of CensusScope.org - A Lookup Tool for Census and ACS Data Suitable for Educators and the General Public," July 20, 2011.

Howard Chernick, Hunter College & Graduate Center; Andrew Reschovsky, University of Wisconsin-Madison, *SUMMER AT CENSUS*, "Revenue Diversification, the Housing Crisis, and the Financing of U.S. Cities," July 21, 2011.

Mary Waters, Harvard University, *SUMMER AT CENSUS*, "The Second Generation in Young Adulthood: Identity, Culture and Socioeconomic Success," July 26, 2011.

Mary Waters, Harvard University, *SUMMER AT CENSUS*, "Five Years After Hurricane Katrina: What Recovery Looks Like for Survivors," July 27, 2011.

Osbert Pang, The Wharton School, University of Pennsylvania, "On the Implementation and Extension of BART," July 28, 2011.

Mick Couper, University of Michigan, *SUMMER AT CENSUS*, "Using Paradata to Understand Measurement Error in Surveys," August 1, 2011.

David Cowen, University of South Carolina, *SUMMER AT CENSUS*, "Use of Parcel Data to Update and Enhance Census Bureau Geospatial Data," August 4, 2011.

Gauri Datta, University of Georgia, *SUMMER AT CENSUS*, "Estimation of Mean Squared Error of Model-Based Small Area Estimators," August 9, 2011.

Gauri Datta, University of Georgia, *SUMMER AT CENSUS*, "Short Course on Small Area Estimation, Part I, II, III," August 10, 11, 12, 2011.

Nathan Yau, University of California, Los Angeles, *SUMMER AT CENSUS*, "Data Visualization Stories," August 16, 2011.

Warren Brown, Cornell University, *SUMMER AT CENSUS*, "Case Studies in State and Local Government Use of the American Community Survey," August 17, 2011.

Peter Christen, Australia National University, "Techniques for Unsupervised Record Linkage Classification and Indexing Techniques for Real-time Record Linkage," September 12, 2011.

Jorre Tibor Alexander Vannieuwenhuyze, Katholieke Universiteit Leuven (Belgium), *SUMMER AT CENSUS*, "Evaluating Relative Mode Effects in Mixed-Mode Surveys: Three Methods to Disentangle Selection and Measurement Effects," September 14, 2011.

William Winkler, CSRM, U.S. Census Bureau, "Cleaning and Using Administrative Lists: Enhanced Practices and Computational Algorithms for Record Linkage and Modeling/Editing/Imputation," September 15, 2011.

David Draper, University of California, Santa Cruz, *SUMMER AT CENSUS*, "Workshop: Part I & II - Bayesian Model Specification: toward a Theory of Applied Statistics," September 19 & 20, 2011.

J.N.K. Rao, Carleton University (Canada), *SUMMER AT CENSUS*, "Small Area Estimation under Informative Sampling," September 26, 2011.

J.N.K. Rao, Carleton University (Canada), *SUMMER AT CENSUS*, "Robust Small Area Estimation," September 27, 2011.

J.N.K. Rao, Carleton University (Canada), *SUMMER AT CENSUS*, "Estimating Equations Approach to Inference from Complex Survey Data," September 28, 2011.

6. PERSONNEL ITEMS

6.1 HONORS/AWARDS/SPECIAL RECOGNITION

Bronze Medal Award, U.S. Bureau of the Census

- **Chad Eric Russell (Team Award)** – "This team successfully applied their considerable pooled creativity and drive to demonstrate a new concept for a processing environment at the U.S. Census Bureau to manage, control, share, and work on internal datasets."

6.2 SIGNIFICANT SERVICE TO PROFESSION

Chris Blakely
- Refereed a paper for *Econometrics Review.*

Chandra Erdman
- Refereed a paper for *Research Methods and Analytics.*
- Organizer, Mini-Symposium: "Applications from the EDGE," International Congress on Industrial and Applied Mathematics.

Maria Garcia
- Member, Organizing Committee, UN/ECE Work Session on Statistical Data Editing.
- Session organizer and discussant, "Macro editing and selective editing", UN/ECE Work Session on Statistical Data Editing, May 2011.

Patrick Joyce
- Refereed a paper for *Survey Methodology.*

Martin Klein
- Refereed papers for the *International Journal of Statistical Sciences* and *Sankhya, Series B.*

Jerry Maples
- Refereed a paper for *Computational Statistics.*

Thomas Mathew
- Refereed papers for *Journal of the American Statistical Association, Computational Statistics and Data Analysis, Journal of Statistical Planning and Inference, Quality and Reliability Engineering International, Communications in Statistics,* and *Statistics and Probability Letters.*
- Associate Editor, *Journal of the American Statistical Association.*
- Associate Editor, *Journal of Statistical Planning and Inference*.
- Associate Editor, *Statistical Methodology.*
- Associate Editor, *Sankhya, Series B.*
- Member, Committee on Youden Award for Interlaboratory Testing, American Statistical Association.

Tucker McElroy
- Refereed papers for *Journal of the American Statistical Association, Journal of Business and Economic Statistics, Annals of Statistics,* and *International Statistical Review.*

Mary H. Mulry
- Vice President, American Statistical Association.
- Associate Editor, *Journal of Official Statistics*.
- Member, Program Committee, International Conference on Methods for Surveying and Enumerating Hard-to-Reach Populations, October 31 – November 3, 2012, New Orleans, LA.

Rolando Rodriguez
- Refereed a paper for *Survey Methodology.*

Eric Slud
- Associate Editor, *Journal of the Royal Statistical Society, Series B (Methodological).*
- Associate Editor, *Lifetime Data Analysis.*
- Guest Co-Editor (with Partha Lahiri) of Special Issue of *Statistical Science* containing invited papers and discussion from a May 2008 Univ. of Maryland Workshop on Bayesian Methods that Frequentists Should Know.

Lynn Weidman
- Refereed a paper for *Journal of Official Statistics.*

Bill Winkler
- Refereed a paper for *QDB 2011.*
- Refereed papers for *Biometrika, Journal of Official Statistics*, and *Quality in Databases.*
- Associate Editor, *Journal of Privacy Technology*.
- Associate Editor, *Journal of Privacy and Confidentiality*.
- Associate Editor, *Transactions on Data Privacy.*
- Member, Program Committee for QDB 2011 at the 2011 Very Large Database Conference in Singapore.
- Reviewer, National Science Foundation Proposal.

Tommy Wright
- Associate Editor, *The American Statistician.*
- Member, Fellows Committee, American Statistical Association.
- Member, Advisory Board, Mathematics & Statistics Department, Georgetown University.
- Member, Cochran-Hansen Prize Committee, International Association of Survey Statisticians.
- Member, Morris Hansen Lecture Committee, Washington Statistical Society.

Bill Yancey
- Refereed a paper for *Information Fusion*.

6.3 PERSONNEL NOTES

Natalia Titova returned to New York City.

Sarah James joined the Human Resources Division and is on a temporary assignment to our center through the Mixed-Tour Program.

Michael Leibert joined our center as a Program Analyst.

Don Malec accepted a position with the National Center for Health Statistics.

Chaitra Nagaraja accepted a faculty position at Fordham University.

Esan Sumner (sophomore in business management at Prince George's Community College) joined our center as an intern.

Peter Linton (graduate student in statistics at University of Maryland, Baltimore County) joined our Missing Data Methods Research Group as an intern.

Andy Vesper (Ph.D. candidate in statistics at Harvard University) joined our Small Area Estimation Research Group as an intern.

Christopher Lee (graduate student in mathematics and statistics at Georgetown University) joined our Missing Data Methods Research Group as an intern.

Ryan McMillin (graduate student in mathematics and statistics at Georgetown University) joined our Missing Data Methods Research Group as an intern.

Joint Program in Survey Methodology Junior Fellows
- Emily Bartha (junior in computational mathematics at Rochester Institute of Technology)

Eric Slud joined our center as Area Chief for Mathematical Statistics.

Joe Schafer joined our center as Principal Researcher.

Derek Young joined our Experimentation, Simulation, and Modeling Research Group.

APPENDICES

APPENDIX A	Center for Statistical Research and Methodology FY 2011 Program Sponsored Projects/Subprojects With Substantial Activity and Progress and Sponsor Feedback (Basis for PERFORMANCE MEASURES)		
Project #	Project/Subproject Sponsor(s)	CSRM Contact	Sponsor Contact
5610102	**DECENNIAL** Statistical Design and Estimation *1. Census Unduplication Research..*	Michael Ikeda ....Charisse Jones	
	2. Voting Rights Section 203 Model Based Methodology: Research, Development, and Production	Patrick Joyce...............................Alfredo Navarro	
	3. Synthetic Decennial Microdata File...	Martin Klein .... Dan Weinberg	
5610103	Coverage Measurement Planning and Development *4. Coverage Measurement Research...*	Jerry Maples .... Tom Mule	
	5. Accuracy of Coverage Measurement...	Mary Mulry ....Pat Cantwell	
5610106	Evaluation Planning Coordination *6. Explaining How Census Errors Occur through Comparing Census Operations History with Census Coverage Measurement (CCM) Results ....*	Michael Ikeda ....Magdelena Ramos	
5385160	American Community Survey (ACS) *7. ACS Applications for Time Series Methods................................*	Tucker McElroy .... Alfredo Navarro	
	8. ACS Variances...	Eric Slud.................................... Alfredo Navarro	
	9. ACS Small Area Estimation for Group Quarters	Lynn Weidman .... Alfredo Navarro	
	10. ACS Data Issues	Lynn Weidman .... Alfredo Navarro	
	11. ACS Exploratory Analysis of the Differences in ACS Respondent Characteristics between the Mandatory and Voluntary Response Treatments..	Lynn Weidman ....Alfredo Navarro	
TBA	**DEMOGRAPHIC** Demographic Statistical Methods Division Special Projects *12. Tobacco Use Supplement (NCI) Small Domain Models...............*	Aaron Gilary.....................................Benmei Liu	
	13. Top Coding Investigation	Thomas Mathew ....Aref Dajani	
0906/7374	Demographic Surveys Division Special Projects *14. Data Integration...*	Ned Porter ....Marie Pees	
1465444	Re-Engineered Survey of Income and Program Participation (RE-SIPP) Research *15. Model-Based Imputation for the Demographic Directorate ....*	Chandra Erdman......................... Martha Stinson	
7165000	Social, Economic, and Housing Statistics Division (SEHSD) Small Area Estimation Projects *16. Research for Small Area Income and Poverty Estimates (SAIPE)*	Elizabeth Huang ....Wesley Basel	
	17. Small Area Health Insurance Estimates (SAHIE)	Ryan Janicki .... Donald Luery	
2320154	**ECONOMIC** *18. Editing Methods Development (Investigation of Selective Editing Procedures for Foreign Trade Programs) ....*	María García....................................Ryan Fescina	
2320152	Time Series Research *19. Seasonal Adjustment Support..*	Brian Monsell ....Kathleen McDonald-Johnson	
	20. Seasonal Adjustment Software Development and Evaluation	Brian Monsell ....Kathleen McDonald-Johnson	
	21. Research on Seasonal Time Series - Modeling and Adjustment Issues	Tucker McElroy ...Kathleen McDonald-Johnson	
	22. Supporting Documentation and Software for X-12-ARIMA and X-13A-S...	Brian Monsell ....Kathleen McDonald-Johnson	
TBA	*23. Survey of Research and Development in Industry, Imputation and Sampling Research and Software Design.................................*	Yves Thibaudeau...............................Jeri Mulrow	
TBA	*24. Governments Division Project on Decision-Based Estimation ......*	Eric Slud..Yang Cheng	

FY 2011 PROJECT PERFORMANCE MEASUREMENT QUESTIONNAIRE CENTER FOR STATISTICAL RESEARCH AND METHODOLOGY

Dear

In a continuing effort to obtain and document feedback from program area sponsors of our projects or subprojects, the Center for Statistical Research and Methodology will attempt for the eleventh year to provide *seven measures of performance,* this time for the fiscal year 2011. For FY 2011, the *measures of performance* for our division are:

Measure 1. Overall, Work Met Expectations: Percent of FY 2011 Program Sponsored Projects/Subprojects where sponsors reported that work met their expectations.

Measure 2. Established Major Deadlines Met: Percent of FY 2011 Program Sponsored Projects/Subprojects where sponsors reported that all established major deadlines were met.

Measure 3a. At Least One Improved Method, Developed Technique, Solution, or New Insight: Percent of FY 2011 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight.

Measure 3b. Plans for Implementation: Of the FY 2011 Program Sponsored Projects/Subprojects reporting at least one improved method, developed technique, solution, or new insight, the percent with plans for implementation.

Measure 4. Predict Cost Efficiencies: Number of FY 2011 Program Sponsored Projects/Subprojects reporting at least one "predicted cost efficiency."

Measure 5. Journal Articles, Publications: Number of journal articles (peer review) and publications documenting research that appeared or were accepted in FY 2011.

Measure 6. Proceedings Publications: Number of proceedings publications documenting research that appeared in FY 2011.

These measures will be based on response to the five questions on this form from our sponsors as well as from members of our division and will be used to help improve our efforts.

To construct these seven measures for our division, we will combine the information for all of our program area sponsored projects or subprojects obtained during November 28 thru December 9, 2011 using this questionnaire. Your feedback is requested for:

Project Number and Name: _____
Sponsoring Division(s): _____

After all information has been provided, the CSRM Contact _____ will ensure that the signatures are obtained in the order indicated on the last page of this questionnaire.

We very much appreciate your assistance in this undertaking.

Tommy Wright   Date
Chief, Center for Statistical Research and Methodology

Brief Project Description (CSRM Contact will provide from Division's Quarterly Report):

Brief Description of Results/Products from FY 2011 (CSRM Contact will provide):

(over)

TIMELINESS:
 Established Major Deadlines/Schedules Met

 1(a). Were all established major deadlines associated with this project or subproject met? **(Sponsor Contact)**

   □ Yes   □ No   □ No Established Major Deadlines

 1(b). If the response to 1(a) is No, please suggest how future schedules can be better maintained for this project or subproject. **(Sponsor Contact)**

QUALITY & PRODUCTIVITY/RELEVANCY:
 Improved Methods / Developed
 Techniques / Solutions / New Insights

 2. Listed below are at most 2 of the top improved methods, developed techniques, solutions, or new insights offered or applied on this project or subproject in FY 2011 where an CSRM staff member was a significant contributor. Review "a" and "b" below **(provided by CSRM Contact)** and make any additions or deletions as necessary. For each, please indicate whether or not there are plans for implementation. If there are no plans for implementation, please comment.

 □ No improved methods/techniques/solutions/new
     insights developed or applied.

 □ Yes as listed below. (See a and b.)

                                                 Plans for
                                                 Implementation?
 a. _____   Yes □   No □
    _____
    _____
    _____
    _____

 b. _____   Yes □   No □
    _____
    _____
    _____
    _____

 Comments (Sponsor Contact):

COST:
 Predict Cost Efficiencies

 3. Listed **(provided by CSRM Contact)** below are at most two research results or products produced for this project or subproject in FY 2011 that predict cost efficiencies. Review the list, and make any additions or deletions as necessary. Add any comments.

     □ No cost efficiencies predicted.
     □ Yes as listed below. (See a and b.)

  a.

  b.

 Comments (Sponsor Contact):

OVERALL:
 Expectations Met/Improving Future Communications

 4. Overall, work on this project or subproject by CSRM staff during FY 2011 met expectations. **(Sponsor Contact)**

     □ Strongly Agree
     □ Agree
     □ Disagree
     □ Strongly Disagree

 5. Please provide suggestions for future improved communications or any area needing attention on this project or subproject. **(Sponsor Contact)**

 (CSRM Contact will coordinate first two signatures as noted and pass to CSRM Chief.)

First_____
       Sponsor Contact Signature   Date

Second_____
       CSRM Contact Signature

 (CSRM Chief will coordinate last two signatures as noted.)

Third_____
       Sponsor Division Chief Signature   Date

Fourth_____
       CSRM Division Chief Signature

Center for Statistical Research and Methodology

Research & Methodology Directorate

STATISTICAL COMPUTING AREA

Robert Creecy
  Joe Schafer
  Sarah James (HRD)

Machine Learning & Computational Statistics Research

Bill Winkler
  William Yancey

Statistical Computing Applications & Data Visualization Research

Robert Creecy (Acting)
  Taniecea Arceneaux (Postdoc)
  Tom Petkunas
  Ned Porter

Missing Data Methods Research

Yves Thibaudeau
  Chandra Erdman
  Maria Garcia
  Martin Klein
  Ben Klemens
  Peter Linton (S)
  Rolando Rodriguez
  Jun Shao (U. of WI)
  VACANT

Research Computing Systems

Chad Russell
  VACANT

MATHEMATICAL STATISTICS AREA

Eric Slud
  Gloria Prout

Sampling & Estimation Research

Lynn Weidman
  Mike Ikeda
  Patrick Joyce
  Mary Mulry
  Julie Tsay
  VACANT

Small Area Estimation Research

Eric Slud (Acting)
  Gauri Datta (U. of GA)
  Aaron Gilary
  Elizabeth Huang
  Ryan Janicki
  Partha Lahiri (F)
  Jerry Maples

Time Series Research

Brian Monsell
  Chris Blakely (Postdoc)
  David Findley
  VACANT
Tucker McElroy
  VACANT

Experimentation, Simulation, & Modeling Research

Tommy Wright (Acting)
  Thomas Mathew (UMBC)
  Derek Young

Tommy Wright, Chief
  Kelly Taylor
  Ann Dimler
  Michael Leibert
  Michael Hawkins
  Christina Cooper (S)
  Esan Summer (S)
  Sarah Wilson (S)

(F) ASA/NSF/Census Research Fellow
(S) Student

September 30, 2011