Use of ACS Data to Produce SAIPE Model-Based Estimates of Poverty for Counties

December 2007

William Bell
Wesley Basel
Craig Cruse
Lucinda Dalzell
Jerry Maples
Brett O'Hara
David Powers

ACKNOWLEDGMENTS

This assessment of incorporating American Community Survey (ACS) data into the Small Area Income and Poverty Estimates (SAIPE) program has benefited greatly from the advice and support of many individuals. Roderick Little of the University of Michigan, Hal Stern of the University of California—Irvine, and Alan Zaslavsky of Harvard University all thoroughly reviewed a draft of this paper. Their comments are reflected in this final version. In addition, Donald Luery, Donald Malec, and David Waddington of the U.S. Census Bureau also provided detailed comments reflected here.

This research paper also benefited substantially from other members of the SAIPE team and the Data Integration Division. Elizabeth Huang provided an assessment of state raking factors. Tamara Lee, Marjorie Hanson, Joanna Turner, Anita Molina, and Monea Hendricks assisted with formatting and provided editorial comments.

Richard A. Denby
Assistant Division Chief for Small Area Estimates
Data Integration Division
U.S. Census Bureau
December 2007

Table of Contents

List of Figures

List of Tables

Chapter 1: Introduction and Background

1.1 INTRODUCTION

The U.S. Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program has produced median household income estimates and poverty by age group estimates for U.S. states and counties since 1993 (estimates released in 1997), and population and poverty estimates of school-aged (ages 5 - 17) children for school districts since 1995 (estimates released in 1999). The estimates are now produced annually. The state and county estimates have been produced using Fay-Herriott (1979) models fitted to direct income and poverty estimates from the Annual Social and Economic Supplement (ASEC) of the Current Population Survey (CPS). Due to a lack of data for modeling at the school district level, the school district estimates have been produced using a simple updating scheme that takes the estimated shares of the number of school-aged children in poverty among school districts within each county from the previous decennial census and multiplies these by the current year's county level model-based estimates of the number of 5 - 17 year old children in poverty.[1] The school district poverty estimates are particularly important because the U.S. Department of Education uses them to allocate Title I funds (about $13 billion dollars in 2006). The state and county estimates are of independent interest, but are also important because these estimates are constrained to be consistent across the various geographic levels. The share approach used for the school district estimates forces them to aggregate to the county estimates, but in addition, the county model-based estimates of number in poverty are raked so they add to their corresponding state model-based estimates. The latter are raked so they add to the CPS ASEC direct national estimate.

The Census Bureau has also provided direct state and national income and poverty estimates from the CPS ASEC for many years,[2] and in recent years from the American Community Survey (ACS). For several years, the ACS estimates were obtained from demonstration surveys done as part of the ACS development, but in August of 2006, estimates from the 2005 full production ACS were released. As the ACS has a much larger sample size (about 3 million addresses for full production) than the CPS ASEC (about 100,000 addresses), the ACS data will support direct survey estimates for much smaller areas than will the CPS ASEC. Thus, the 2005 ACS produced single-year direct estimates for counties and other places with populations of 65,000 or more.[3] In addition, in 2006, the Census Bureau switched the source of the official direct survey state estimates of income and poverty from the CPS ASEC to the ACS. The CPS ASEC remains the source of the official direct national estimates of income and poverty.

[1] Since some school districts cross county boundaries the estimation is made for "school district pieces," which are the parts of school districts contained within individual counties, and then the estimates for the pieces for each district are added. Also, in cases where the county consists of a single school district, the school district estimate equals the model-based estimate for that county.

[2] To reduce sampling error in the CPS ASEC direct state estimates, averages over 3 consecutive years of CPS ASEC data were provided.

[3] Ultimately, the ACS will provide direct estimates for all counties and school districts, as well as for other small geographic areas (e.g., census tracts). For areas with populations less than 65,000, these estimates will be based on 3-year or 5-year (for areas with populations less than 20,000) accumulations of ACS data.

Given the switch from CPS ASEC to ACS as the source of direct state estimates, the SAIPE program is currently planning to switch from using CPS ASEC data to using ACS data as the basis for its state and county model-based estimates, essentially for two reasons: [4]

1. Historically, SAIPE has been designed to be consistent with the Census Bureau's official direct survey estimates, in the sense of taking as its "target" what the official direct estimates are estimating. Thus, it was assumed that if the official direct income and poverty estimates were ever changed, SAIPE would accordingly change the basis for its model-based estimates.[5]
2. The much larger sample size of the ACS provides obvious advantages for small area estimation. For example, the SAIPE county level models have used 3-year averages of CPS ASEC data to reduce sampling variability in the direct estimates being modeled. With ACS data, single-year county estimates can be modeled.

Outside reviewers are now being asked for their feedback on technical issues related to this switch. The focus is primarily on comparing and evaluating alternative county level models of the number of children ages 5 - 17 in poverty.[6] Since the plan is to rake the county model-based estimates to state estimates, some results on state level models for ACS estimates are presented.[7] Feedback is appreciated on any technical issues that seem important, including the overall, but very difficult question of whether or not the switch from CPS ASEC to ACS data appears likely to produce "better" results. The related questions of how and how well the effects of this change might be assessed are also crucial.

The following questions are of particular interest.

Specific technical questions for the reviewers:

1. Various models of ACS log number 5 - 17 in poverty and of ACS log 5 - 17 poverty rates have been compared. One difficulty that has been encountered, particularly with comparing regression predictions, is that these two types of models use different dependent variables. This leads to the following two questions:
    a. Are there other ways to compare models for log number in poverty and log poverty rates that should be tried?
    b. Based on the results so far, which of the two types of models appears to produce better results, or can it not be determined? Since SAIPE has historically modeled log number in poverty at the county level, some advantages from modeling log poverty rates should first be established before making a change.
2. Are any significant inadequacies observed for any of the models?

---

[4] When ACS estimates for all school districts become available, SAIPE will consider how to make use of these in producing school district poverty estimates.
[5] The situation now is perhaps less clear than before, since the CPS ASEC still provides the official direct national income and poverty estimates, while the ACS provides the official direct state estimates.
[6] For efficiency, the commonly used term "5 – 17 poverty" will be used hereinafter to refer to children ages 5 – 17 that live below the poverty threshold.
[7] A main question at the state level is whether or not we should continue to model state estimates or simply use the ACS direct state estimates. Results relevant to this question are presented in Section 2.7 of Chapter 2.

3. Is there a clear choice of a best model?
4. Assuming that the ACS county model-based estimates of number of 5 - 17 in poverty will be raked to corresponding state estimates, is it better to rake to the direct ACS state estimates or to the corresponding predictions obtained from models for ACS state data?
5. (Research on this item will be presented in a future addendum to this document). The county models fitted here use direct ACS sampling error variance estimates. Some results are also presented on models that use sampling error variances obtained from generalized variance functions (GVFs) used to smooth the direct variance estimates. Will using the GVFs likely improve the results?
6. Are there suggestions for further research to improve the models for future years? Given the plan to switch to the use of ACS data for the SAIPE estimates to be produced this fall, it does not seem feasible to make major changes this year that are well outside the scope of the models already examined. More major changes can be considered in future years, however.

The report is organized as follows: The remainder of Chapter 1 gives additional material that may be taken as "background." Chapter 2 discusses results from several alternative county poverty models applied to 2005 ACS data. Chapter 3 provides some comparisons of results obtained from several years of ACS data, including data from the demonstration surveys (2000-2004) and the full production ACS data for 2005 and 2006.

1.2 HISTORICAL BACKGROUND

Prior to the creation of the SAIPE program, the decennial census long-form was the only source of income distribution and poverty statistics for households, families, and individuals if one needed data for "small" geographic areas, e.g., counties, cities, and other sub-state areas. The ten-year span between censuses left a large gap in information concerning the economic situations of local areas. Thus, in the 1990s, federal agencies and the Congress asked the Census Bureau to develop intercensal estimates of income and poverty. In September 1994, Congress passed the Improving America's Schools Act (PL 103-382), of which Title I specified that the distribution of federal funds be made to school districts based largely on "the number of children ages 5 - 17, inclusive, from families below the poverty level on the basis of the most recent satisfactory data ... available from the Department of Commerce." Provisions in the law allowed for modifications of this approach for school districts with fewer than 20,000 people.

This law further required the Secretary of Education to use updated Census Bureau data on school-aged children in poverty for counties starting with allocations for the 1997-98 school year, and for school districts beginning in the 1999-2000 school year, unless the Secretaries of Education and Commerce determined that the use of the updated estimates would be "inappropriate or unreliable." The law also directed the Secretary of Education to fund a National Academies of Science (NAS) panel to provide advice on the suitability of the Census Bureau's estimates for allocating funds. The NAS panel's first report (National Research Council, 1997, p. 38) recommended that allocations for the 1997-98 year be based on estimates obtained by averaging poverty rate estimates from the previous (1990) census and the SAIPE county models (for 1993), as the panel felt that further evaluations of the county models were needed. Subsequent to such additional evaluations by the Census Bureau, the NAS panel's second report (National Research Council, 1998, p. 79) recommended basing allocations for 1998-99 fully on revised county model-based estimates for 1993. The NAS panel's third report (National Research Council, 1999, pp. 79-80) then recommended use of the Census Bureau's 1995 school district estimates for making allocations for the 1999-2000 school year. In making this recommendation, the NAS panel said the following (National Research Council, 1999, p. 3):

> Although the Census Bureau's 1995 estimates of school-age children in poverty have potentially large errors for many school districts, the panel nonetheless concludes that they are not inappropriate or unreliable to use for direct Title I allocations to districts as intended by the 1994 legislation. In reaching this conclusion, the panel interprets "inappropriate and unreliable" in a relative sense. Some set of estimates must be used to distribute Title I funds to school districts. The panel concludes that the Census Bureau's 1995 estimates are generally as good as—and, in some instances, better than—estimates that are currently being used.

And in a later report (National Research Council, 2000b, p. 46) the panel said the following about their earlier recommendations regarding use of the SAIPE model-based state and county estimates:

> The internal and external evaluations of the 1993 and 1995 state and county estimates led the panel to conclude that the models are working reasonably well and that these estimates are preferable to 1990 census estimates as a basis for Title I allocations (National Research Council, 2000b, p. 46).

Appendix A gives further relevant quotes from NAS panel reports.

1.3 THE NAS PANEL'S SPECIFIC RECOMMENDATIONS AND SAIPE RESEARCH

In addition to the general recommendations regarding use of SAIPE estimates for Title I allocations, the NAS panel's reports contain many specific comments and recommendations about the SAIPE models and estimation procedures. In their report that looked ahead to future work (National Research Council, 2000b, pp. 4-9), the panel summarized their recommendations, grouping them under three headings: research and development of current models, role of survey estimates, and role of administrative records. Below is a review of these recommendations and the research SAIPE has pursued to address them.

1.3.1 Research and development of current models

In this section of their report, the panel identified the following areas for research and development by the Census Bureau in the near term:

Closer integration of the state and county models. The panel expressed concern that differences between the state and county models could produce inconsistencies in their results and suggested that, to address this, SAIPE examine including state effects in the county models, such as state random effects.

Censoring issue. As the county model has used the log of the CPS ASEC estimates of the number of school-age children in poverty, counties with zero school-age children in poverty in the CPS ASEC sample (implying an estimate of zero) could not be included in the regression prediction. Up to twenty percent of counties have thus been excluded from the regression, though as these counties have small CPS ASEC samples, the percentage of the overall CPS ASEC sample excluded has been much less. The panel recommended work on estimation techniques, such as generalized linear mixed models (GLMM), which would permit including all counties in the regression prediction.

Problems in estimating the variance components. For the state model, the problem was frequent estimates of zero for the model error variance (state random effects) by maximum likelihood. For the county model, the problem was that direct estimates of sampling error variances of the CPS ASEC direct county estimates were unavailable, which led the SAIPE program to estimate the county model error variance from an auxiliary census equation and estimate CPS ASEC sampling error variances when fitting the model to the CPS ASEC county poverty estimates. This required using a simple parametric assumption for the sampling error variances as a rough approximation. Diagnostic plots suggested problems with the parametric assumption that was used.

Research on use of additional survey data sources. The panel suggested investigation of modeling techniques that can use multiple years of data from the same survey or data from additional surveys or both. Mention was made of use of ACS data when it became available.

Time lag. The panel suggested giving attention to reducing the 3- to 4-year lag between the release of estimates and the income reference year. (For example, the SAIPE program released state and county estimates for 1995 in the fall of 1998.)

The SAIPE program has addressed many of these recommendations, and the planned switch to ACS data as the basis for SAIPE estimates has important implications for these issues as well. These issues are discussed below.

Research on closer integration of the state and county models. County models with fixed state effects were investigated early on (with results presented to the NAS panel.) The fixed state effects did not seem to yield meaningful improvements to the models. Tapabrata Maiti (American Statistical Association/National Science Foundation/Census Research Fellow/Iowa State University) tried to apply a linear model with state random effects to SAIPE county data, but his estimation procedure failed to converge, so he abandoned the attempt.

Research on the censoring issue. Eric Slud (Statistical Research Division of the U.S. Census Bureau/University of Maryland) compared results from various Fay-Herriot log rate county models and related generalized linear mixed models (GLMMs) fitted to several years (1990, 1994, 1996, 1998, 1999, 2000) of CPS ASEC data (Slud 2004). A primary motivation for examining GLMMs is that they allow use of data from counties with zero people in poverty in sample in the regression predictions. Slud compared these models in regard to how well they predicted the CPS ASEC direct county 5 - 17 poverty estimates and how well they predicted the (1990 and 2000) census county estimates. He summarized the results as follows.

> The besetting problem in SAIPE, that estimators must be judged both by an internal (CPS) and external (census) standard of truth, has been seen to yield Fay-Herriot type estimators which conform remarkably well to the census standard, and GLM type estimators which fit the CPS sampled data better by all reasonable loss-function measures. Both the Fay-Herriot and random-intercept logistic models (with essentially the same predictor variables) are good, and both perform comparably well on the non-CPS-sampled counties (Slud 2004, p. 15).

Earlier in the paper Slud (2004, p. 10) noted that, "… the fitted GLMM's involve rather large PSU [primary sampling unit]-effect variances[8] … and as a result their SAE's [small area estimators] give much greater weight to the CPS direct estimators at small samples sizes than do …" [the fitted Fay-Herriot models.] This probably explains why Slud's fitted GLMMs predicted the CPS ASEC direct estimates better than did the Fay-Herriot models, and also why the GLMMs predicted the census data worse than the Fay-Herriot models—the GLMM predictions were, to some extent, predicting sampling error in the CPS ASEC direct estimates. Fixing the GLMM PSU effect variance at a smaller value resulted in GLMM predictions less influenced by the CPS ASEC direct estimates, which reduced their ability to predict the CPS ASEC data but improved their ability to predict the census results. When model predictions were compared with census results for counties that were not in the CPS ASEC sample both types of models used fixed effect predictions and their performances were more similar.

The results, thus, did not demonstrate any advantage to using GLMMs for SAIPE county poverty estimation. It might be that, for the GLMMs, the PSU effect variance was being overestimated, which might have resulted from the regression predictions not taking account of the CPS ASEC survey design (beyond using sample weighted estimates of county poverty rates). Accounting for survey design when fitting GLMMs, however, is a difficult technical problem without a standard solution. On the other hand, the results from the Fay-Herriot models may have been affected by the crude way sampling error variances were parameterized and estimated as part of the regression predictions (see Section 2.1.2) due to our not having direct estimates of sampling error variances available.

Research on problems in estimating the variance components. For the state models, the problem of zero variance estimates was addressed by switching to a Bayesian approach (using a flat prior on the model error variance). This had several advantages in addition to avoiding variance estimates of zero, as is discussed in Bell (1999).

For the county models, the problems arose because sampling error variance estimates for the direct CPS ASEC county estimates were not produced due to complications of processing the CPS ASEC. The recent construction of replicate weight files should soon permit relatively

[8] The term "PSU effect" refers to what we here call model errors.

straightforward calculations of estimated sampling error variances of the CPS ASEC county estimates. However, switching to use of ACS data for the county models immediately addresses this issue, as estimates of sampling error variances are available for the ACS county estimates. In fact, SAIPE has recently been using direct estimates of ACS sampling error variances in developing county ACS models, leading to the more conventional situation of needing to estimate the model error variance in the regression predictions. The fitting of GVFs to the direct estimates of ACS sampling error variances is also being explored, in an effort to improve the sampling error variance estimates.

Research on use of additional survey data sources. At the state level, Huang and Bell (2004) developed and tested a bivariate state poverty ratio model with the ACS as the dependent variable in one equation and the CPS ASEC as the dependent variable in the second equation. The focus was on improving estimation of poverty in the CPS ASEC equation. They found little benefit from the use of an unrestricted bivariate model, but more substantial benefits (reduced posterior variances) from use of a restricted bivariate model that assumed the regression coefficients (apart from the intercept) were the same in the two equations. However, for both the restricted and unrestricted bivariate models, there were occasional instances of large posterior variance increases relative to the CPS ASEC equation univariate model. These occurred for states with large standardized residuals in the ACS equation.

In unpublished work, Bell investigated borrowing information in state models from CPS ASEC estimates for other age groups or for the previous year and found little or no benefit from doing so.

Although the focus in Huang and Bell (2004) was on borrowing information from ACS data to improve poverty estimates from the CPS ASEC equation, their bivariate model also provided results on whether borrowing information from CPS ASEC data would improve poverty estimates from the ACS equation. These latter results clearly showed negligible benefits from using the CPS ASEC data. Basically, the sample size and consequent sampling error variance differences between the ACS and CPS ASEC direct estimates are too large to allow the CPS ASEC data to provide improvements if the target poverty ratios are those that the direct ACS estimates are estimating. Thus, use of the bivariate instead of univariate ACS model showed very little difference in the results for the ACS equation. While this has not been studied directly for county level models, similar results would be expected.

At the county level, Fisher (2003) and Gee and Fisher (2004) developed a multivariate errors-in-variables model in a hierarchical Bayesian framework. This treated the direct CPS ASEC and ACS estimates, as well as other variables previously used as regression variables, as dependent variables in the model. Although initial results from this model were encouraging, with the planned switch to use of ACS data as the focus of the models, further development of the errors-in-variables model was put on hold.

Work on reducing the time lag of estimates. The time lag in the first years of SAIPE estimates was partly due to the use of revised (second-vintage) population estimates in the county model and in constructing state estimates of the number in poverty (from the model-based state estimates of poverty ratios). Starting in December 2004, the SAIPE program switched to using the first-vintage population estimates and released poverty estimates for 2001 and 2002, reducing the previous lag in release by one year. Baumgardner, Cruse, and Gee (2004) studied historical differences between first- and second-vintage population estimates. They found that the overall difference between the first and second population estimate vintages was negligible and that while some seemingly large differences between the estimate vintages occurred for particular counties, the differences appeared to be, in effect, random and still small relative to the total error of the population estimates.

1.3.2 Role of survey estimates

In their report, the panel suggested some research on the possible uses of ACS data described above. The panel also suggested using Census 2000 long-form estimates instead of 1990 Census long-form estimates in the state and county models. This was done starting with estimates for 1999. Finally, the panel suggested studies (such as exact match studies) to compare the various data sources on income and poverty: Census 2000 long-form, ACS, CPS ASEC, Internal Revenue Service (IRS) and Survey of Income and Program Participation (SIPP). ACS and CPS have intentionally non-overlapping samples, and the CPS ASEC and the SIPP presumably have an overlapping sample that is too small to produce any meaningful evaluations. The Census Bureau has conducted research on matching CPS and IRS records, but there is nothing relevant to SAIPE to report at this time.

1.3.3 Role of administrative records

The following areas were suggested for research by the panel, and SAIPE has pursued much of this research as noted.

Develop subcounty IRS data for use in school district estimates. The SAIPE staff, in collaboration with the Census Bureau's Geography Division, was able to tabulate IRS tax data for school districts. Maples (2004) and Maples and Bell (2005) then investigated use of this data in the school district poverty and population estimates. The results are promising and this approach is being considered for use in school district poverty and population estimates production.

Regularly review the quality and consistency of administrative records data and their relation to income and poverty. The panel specifically mentioned questions about the continued relevance of food stamp participation data after welfare reform was implemented in 1997. The significance of the food stamp predictor variable was monitored and was dropped from the state model after it was found to be insignificant in the state models for 1997 and 1998.[9] In recent years, it appeared that the food stamp variable was becoming consistently significant again, and it was added back to the state model for the 2004 estimates. One theory is that since welfare

[9] The food stamp predictor was insignificant for some years (at least from 1997 to some time past 2000) in the state CPS ASEC equations, but was significant in the Census 2000 equation. This was presumably due, to a large extent, to the much lower standard errors on the regression coefficients in the census equation because the census state estimates had negligible sampling error.

reform allowed states freedom to decide how they administer the food stamp program, a number of states pursued different approaches. These different approaches initially broke (or at least greatly diminished) the link between food stamp participant data and the poverty estimates. Over time the administration of the food stamp program by the states may have stabilized, restoring the food stamp variable to significance in our state models. The food stamp predictor variable in the county model remained significant through this period, and so was retained in the model. The greater geographic detail of the county level may be responsible for the relation between food stamp participation and poverty continuing to show through in the county model, despite the higher level of sampling error in the county CPS ASEC estimates and the concern about the quality of the food stamp data for counties relative to the quality of the data for states.

Also, the SAIPE program has continued to evaluate all the chosen model input data each year, and (with the exception noted of the food stamp variable in the state models) has generally continued to find them satisfactory predictors of poverty. In addition, the SAIPE program has a well-structured, careful system of quality control procedures for recognizing anomalies in the data, particularly for counties. In the very small number of cases where input data seemed extreme, after additional research, imputations have been made.

Examine possible use of data from the free and reduced price lunch (FRPL) program. Cruse and Powers (2006) assessed whether there was predictive power in FRPL counts at the school district level. They found that the FRPL data reflected many reporting problems and inconsistent program enrollment and thus bore a weak relation to Census 2000 poverty. In unpublished work, Bell tried including a state level participation rate in the state models and found that models with the FRPL variable were almost never preferred in Akaike information criterion (AIC) comparisons (in models for 2000-2004). The FRPL variable could reasonably substitute for one of the other variables in the state model (such as the food stamp participation rate variable for some years since 1997), but due to strong correlation with some of the other variables in the model, it did not add much new information.

In addition to research on the above panel suggestions, the SAIPE program has considered incorporating other administrative data into its models. For the county SAIPE models, Powers (2005) tested whether there was utility in including Medicaid participant totals, and Basel and O'Hara (2006) tested whether there was utility in including data on the Earned Income-Tax Credit (EITC). Neither the Medicaid data nor the EITC data were found to be sufficiently helpful to use in refining the SAIPE program's estimates of poverty.

1.4 COMPARING ACS AND CPS ASEC AS SOURCES OF INCOME AND POVERTY DATA

As noted earlier, the SAIPE program is planning to use ACS instead of CPS ASEC data as the basis for its state and county poverty estimates. Potential use of ACS data at the school district level is some years off. This section compares the two surveys, discussing various differences between their measurements of income and poverty.

1.4.1 Some implications for SAIPE regarding ACS and CPS ASEC sample size differences

As noted earlier, the CPS ASEC sample is from roughly 100,000 addresses each year, whereas the ACS sample is from roughly 3,000,000[10] addresses each year. The larger sample size of ACS has the following implications for the modeling of county poverty estimates:

The ACS covers all 3,141 counties in the United States, while the CPS ASEC typically samples about 1,100 counties. For SAIPE, estimates of counties with no CPS ASEC sample were obtained as pure regression predictions, rather than as (empirical Bayes) weighted averages of a direct survey estimate and a regression prediction.

While all counties in the 2005 ACS are self-representing, approximately 720 of the typically 1,100 counties sampled in the CPS ASEC are self-representing. Thus, to construct direct county estimates from the CPS ASEC data, SAIPE had to adjust the survey weights to make all counties in sample self-representing.

A significant problem in modeling the CPS ASEC county estimates of log number in poverty has been the substantial number of counties with zero school-age children in poverty in the CPS ASEC sample, leading to direct estimates of zero for these counties. Typically about 200 counties (18 percent of the 1,100 in sample) had estimates of zero. Since logs of these zero estimates cannot be taken, when using these models, SAIPE was forced to drop these counties from the regression predictions. With the larger sample size of ACS, this problem is diminished. In the 2005 ACS, 169 (5 percent of the 3,141 total counties) had zero estimated school-age children in poverty.

For its county models, SAIPE used three-year averages of CPS ASEC data to (*i*) reduce the high level of sampling error variance in the single-year CPS ASEC direct county estimates, and (*ii*) reduce the number of counties with zero estimates of school-age children in poverty. The ACS sample is large enough to use single-year ACS estimates for the county poverty models.[11]

1.4.2 Some methodological differences between the ACS and CPS ASEC income and poverty estimates

In addition to the large sample size difference between the ACS and the CPS ASEC, various methodological differences can affect the income and poverty estimates from the two surveys. Nelson (2006) gives a valuable discussion of this subject, and somewhat more detail (with a

---

[10] This figure is for the full production ACS sample, first implemented in 2005. For the demonstration period surveys conducted in 2000-2004, the sample sizes varied, but were on the order of 800,000 addresses.

[11] The official published direct ACS county estimates are single-year estimates only for sufficiently large counties (greater than 65,000 people); three-year or five-year accumulations of ACS data will be used in constructing estimates for smaller counties. Since modeling produces estimates with reduced sampling error, we feel we can use single-year ACS estimates for all counties in our models. We also feel it is important to do so since primary uses of the SAIPE estimates (e.g., their use in Title I allocations) effectively involve comparing poverty estimates across places. For such uses, having all the estimates on a common basis is important, so that if we wanted to use multi-year ACS estimates for small counties, we should probably also use them for the large counties.

slightly different focus) is given by Bishaw and Stern (2006). The following section briefly summarizes some of the main differences.

The CPS ASEC collects survey responses mostly in March of each year, asking questions about income in the prior calendar year. In contrast, the ACS collects survey responses every month, asking questions about income in the prior twelve months. The annual ACS estimates combine results from the corresponding twelve monthly surveys, starting in January and ending in December, each of which refers back twelve months. A single-year ACS poverty estimate thus uses income reports that cover a total of twenty-three months of reference, from January of the prior year through November of the current year. Resulting timing differences between the CPS ASEC and ACS estimates are discussed in the next section.

The CPS ASEC collects data primarily through telephone interviews; while the ACS collects data successively in three stages of mailed paper questionnaires, telephone follow-up of mail nonrespondents, and personal visit follow-up of a subsample of the remaining nonrespondents. Income reports could differ across these different modes of data collection.

The CPS ASEC asks over fifty questions about income, including asking about particular income sources such as interest income, dividend income, and pension income. The ACS combines many sources of income into eight income questions. Though the income and poverty measures used by SAIPE depend only on total reported income for households, not on income by particular types, asking the income questions in these different ways can affect the total income reported.

The CPS ASEC and the ACS differ in how they cover the population. The CPS ASEC covers the civilian, non-institutionalized population of the United States, including residents of civilian, non-institutional group quarters (GQ). The 2005 ACS covered only people living in housing units, but ACS started collecting data from GQ residents in 2006. Also, the CPS ASEC counts all people in a housing unit who consider the unit as their usual residence or who have no residence elsewhere; the ACS counts all people in a housing unit living or staying in the unit for more than two months.

The CPS ASEC gathers more information than the ACS about the relationships of people in sample housing units. The CPS ASEC can thus consider individuals related to each other but not to the householder as a distinct family unit, and can determine the poverty status of such "unrelated subfamilies." The ACS generally considers such families as separate individuals, which affects both the poverty threshold that applies and whether or not the person is even in the poverty universe.

As an example of the implications of the last two points, consider the college dormitory population. The CPS ASEC includes college dormitory residents in the poverty universe and generally counts them at their parents' home addresses, but the ACS does not include college dormitory residents in the poverty universe. For example, a family of four with one child living in a college dormitory would have its poverty threshold under the ACS computed for a family with three members, but according to the CPS ASEC for a family with four members.

1.4.3 Timing differences between ACS and CPS ASEC income and poverty estimates

As discussed above, the single-year ACS income and poverty estimates combine results from data collected over the twelve months of the year with rolling twelve-month income reference periods. As a result, the ACS income and poverty estimates span twenty-three months of reference and have an income reference period effectively centered on mid-December of the prior year. For example, the 2005 ACS collected data throughout 2005 and produced estimates with an income reference period centered on December 15, 2004. In contrast, the (national) CPS ASEC income and poverty estimates based on the data collected in February, March, and April of 2005 refer to income earned in calendar year 2004, so the resulting income reference period effectively was centered on July 1, 2004.

Late last year the SAIPE program produced state estimates from modeling CPS ASEC direct state estimates for income year 2004 and county estimates from modeling a three-year average of direct CPS ASEC county estimates for income years 2003-2005. In both cases, the center of the income reference period was July 1, 2004. If ACS data had been modeled last year, the 2004 ACS estimates would have been used, so the center of the income reference period would have been December 15, 2003. This is approximately 6.5 months earlier than that of the SAIPE production estimates.

The timing of the ACS and CPS ASEC estimates is also affected by the population controls for which the survey weights are adjusted. For example, the 2005 ACS estimates utilized as population controls the July 1, 2005 population estimates from the Census Bureau's Population Estimates Program (PEP), which is 6.5 months later than the effective center of the ACS income reference period. The CPS ASEC estimates for income year 2004 utilized as population controls the projected March 2005 population, based on an eight-month projection from the July 1, 2004 population estimates. These controls are thus eight months later than the center of the CPS ASEC income reference period. However, since CPS ASEC population controls apply at the state, not the county, level, they should less directly affect the timing of the CPS ASEC income and poverty estimates. Population controls for ACS apply to individual counties or small groups of counties, and thus have a more direct effect on the timing of the county income and poverty estimates.

The information discussed above is summarized below in Table 1.1, which also indicates the timing of the direct ACS estimates that will (eventually) be published for small (≤ 20,000 population) and medium (> 20,000 but ≤ 65,000 population) size counties. The timing shown for the 3- and 5-year ACS estimates is hypothetical and illustrative, since multi-year estimates from the ACS have not yet been released. These dates are shown in reference to dates for estimates released in 2006 only for concreteness, to illustrate what their timing would have been in relation to the estimates that actually were released.

Table 1.1: Timing of income and poverty estimates for CPS ASEC and ACS (assuming ACS was in full production)[12]

Estimate level	Source	Release date	Population controls and income reference period
National	CPS ASEC	8/06	Pop: Mar 06; Income: calendar year 05, midpoint 7/1/05.
	ACS	8/06	Pop: Jul05; Income: 23 mo. wtd avg, centered on 12/15/04.[13]
State	CPS ASEC	8/06	Average of 04-06 surveys. Pop: March 05; Income: Midpoint 7/1/04.
	ACS	8/06	Pop: July 05; Income: 23 mo. wtd avg, centered on 12/15/04.
	SAIPE – CPS ASEC	11/06	Pop: March 05; Income: calendar year 04, midpoint 7/1/04.
State	SAIPE – ACS	11/06	Pop: July 04; Income: 23 mo. wtd avg centered on 12/15/03.
Large Counties	ACS	8/06	Pop: July 05; Income: 23 mo. wtd avg, centered on 12/15/04.
Medium Counties[14]	ACS	8/06	Pop: Avg of July 03-July 05 Pop, midpoint 7/1/04; Income: 47 mo. wtd avg centered on 12/15/03.
Small Counties	ACS	8/06	Pop: Avg of July 01-July 05, midpoint 7/1/03; Income: 71 mo. wtd avg centered on 12/15/02.
All Counties	SAIPE – CPS ASEC	11/06	Pop: March 05; Income: Average of 04-06 surveys, midpoint 7/1/04.
All Counties	SAIPE – ACS	11/06	Pop: July 04; Income: 23 mo. wtd avg centered on 12/15/03.

[12] The original version of this document was produced by William Bell for the Census Bureau's Income Estimates Workshop in May 2006. Information has been added on the timing of the population controls used in the CPS ASEC and the ACS.

[13] Any single-year ACS income definition is a weighted average over a 23-month period, but the geographic distribution of the population totals (from the controls) is still a single point in time. An accumulated 5-year ACS estimate for a "small" county is effectively like a weighted average based on 71 months of respondent income data, while a 3-year ACS estimate for a "medium" sized county is like a weighted average based on 47 months of respondent data.

[14] The timing shown for the 3- and 5-year ACS estimates is hypothetical and illustrative, since multi-year estimates from the ACS have not yet been released. These dates are shown in reference to dates for estimates released in 2006 only for concreteness, and to illustrate what their timing would have been in relation to the estimates that actually were released.

1.4.4 Differences between ACS and CPS ASEC direct poverty estimates

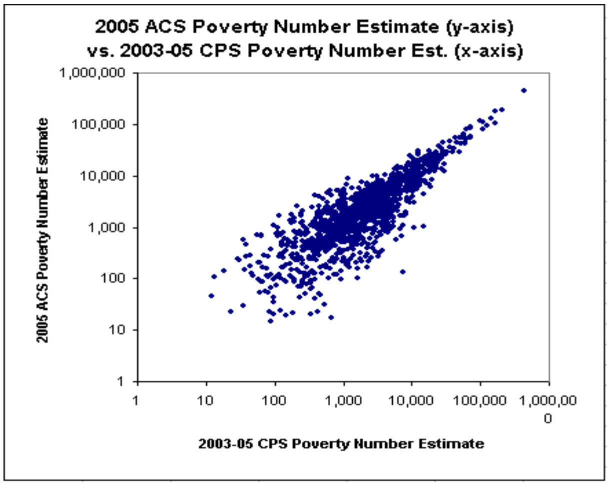
Child poverty rate estimates from the ACS tend to be higher than child poverty rate estimates from the CPS ASEC. This can be seen in the national child poverty rate estimates from the two surveys, which are given in Table 1.2. In the table, the 2005 ACS data is lined up horizontally with the 2005 CPS ASEC data (for income and poverty in 2004); the 2004 ACS data is lined up vertically with the 2004 CPS ASEC data (for income and poverty in 2003); and so forth. The difference between the CPS ASEC and the ACS was statistically significant at the 10 percent level for survey years 2001–2005 (income years (IY) 2000-2004). The difference was not statistically significant for survey year 2000 (IY 1999). The ACS estimates for survey years 2001-2004 are from the ACS demonstration surveys, while the survey year 2005 estimate is from the full production ACS.

Table 1.2: Child poverty rate and number (ages 0 - 17) CPS ASEC and ACS

Year		Child poverty rate		Child poverty number (in millions)	
CPS ASEC	ACS	CPS ASEC	ACS	CPS ASEC	ACS
2000 CPS ASEC (for IY 1999)	C2SS	17.1%	17.3%	12.28	12.21
2001 CPS ASEC (for IY 2000)	2001 ACS	16.2%	16.9%	11.59	11.96
2002 CPS ASEC (for IY 2001)	2002 ACS	16.3%	17.6%	11.73	12.52
2003 CPS ASEC (for IY 2002)	2003 ACS	16.7%	17.7%	12.13	12.67
2004 CPS ASEC (for IY 2003)	2004 ACS	17.6%	18.4%	12.87	13.25
2005 CPS ASEC (for IY 2004)	2005 ACS	17.8%	18.5%	13.04	13.36

In addition to the higher child poverty rate estimates observed in the ACS, the geographic distribution of child poverty across states may also be different. Nelson (2006) provides results of chi-squared tests showing a statistically significant difference in the geographic distribution of all-ages poverty across states in the 2005 ACS compared with the two-year average of CPS ASEC estimates for 2004 and 2005. Tests were run for various definitions of "equal geographic distributions of poverty," including equal state shares of national poverty and equal deviations of state poverty rates from the national poverty rates. The resulting chi-squared statistics were statistically significant at the 1 percent level, providing evidence that states collectively have a different geographic distribution of poverty in the 2005 ACS than in the two-year average of the CPS ASEC estimates for 2004 and 2005. Similar chi-squared tests were also run for the prior year, comparing the 2004 ACS with the two-year average of CPS ASEC estimates for 2003 and 2004. These tests also provide evidence of differing geographic distributions of poverty, though in this case at the 5 percent level of significance. (The 2004 ACS had a smaller sample size than the 2005 ACS.)

Figures 1.1 and 1.2 plot the 2005 ACS ages 5 - 17 related child poverty estimates against the 2003-05 CPS ASEC ages 5 - 17 related child poverty estimates. In both figures, the x- and y-axes are shown in log scale but labeled in linear scale. Figure 1.1 shows the correlation between the county poverty number estimates for related children ages 5 - 17 for the CPS ASEC and ACS surveys. Figure 1.2 plots the same concept but for poverty rates rather than poverty numbers. The positive correlation between ACS and CPS ASEC county poverty is more diffuse when poverty is observed in rate form as opposed to levels form.



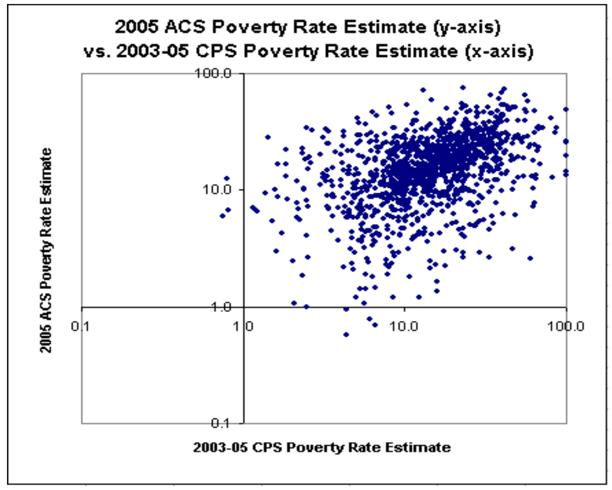


Figure 1.1 (left): County-level 2005 ACS number survey estimates, ages 5 - 17 related against 2003-05 CPS ASEC number survey estimates, ages 5 – 17 related.

Figure 1.2 (right): County-level 2005 ACS rate survey estimates, ages 5 – 17related against 2003-05 CPS ASEC rate survey estimates, ages 5 – 17 related.

In both figures the x- and y-axes are shown in log scale but labeled in linear scale.

Chapter 2: ACS County Poverty Models

2.1 INTRODUCTION

This chapter discusses small area models applied to single-year survey estimates from the ACS for county level poverty estimates of school-age (5 - 17) children. The starting point is the current SAIPE production model that has been applied to county CPS ASEC survey estimates (from 3-year averages of CPS ASEC data) of the (log) number of children ages 5 - 17 in poverty[15]. Section 2.2 examines application of this model with single-year 2005 ACS survey estimates replacing the three-year average CPS ASEC survey estimates used in previous years. This model is referred to as the log-level model, since it involves modeling estimates of the log of the number in poverty. Section 2.2 also examines some alternative forms of log-level models.

Another general type of model considered in the original SAIPE county model evaluations was the log-rate model (National Research Council 1998). In this model, the dependent variable is the logarithm of a direct survey estimate of the county poverty rate of children ages 5 - 17.[16] In its basic form, the regression variables are also all log-rate.[17] Section 2.3 examines log-rate models applied to ACS county poverty rate estimates for children ages 5 - 17. This section also considers some variations of log-rate models that involve adding additional regression variables to the model, such as the logarithm of the 5 - 17 population. These variations relate to forming a connection between log-level and log-rate models. Note that the log poverty rate can be broken into the logarithm of the number in poverty (from the numerator) minus the logarithm of the population (from the denominator). Hence, a log-rate model can be rewritten to imply a model for the log number in poverty, and if the log-rate regression variables are similarly broken down, a log-rate model can be rewritten with only log-level as regressors. Such a model only differs in terms of the dependent variable. Comparing fits of the basic and alternative versions of the log-rate model gives indications of whether a pure log-rate regression predicts the data as well as a log-level model.

For SAIPE production, the county estimates of number 5 - 17 are rescaled within each state (raked) to force the estimates to aggregate to the state estimate of number 5 - 17 in poverty

---

[15] More precisely, the model uses CPS ASEC county estimates of the number of 5 – 17-year-old related children in families in poverty. The ACS county estimates we use here are defined similarly. There are, however, some differences in how the ACS and the CPS ASEC define resident population and determine the number in poverty, as discussed in Section 1.4.

[16] The SAIPE state models are rate models, but instead of actual poverty rates (defined as number of 5 – 17-year-old related children in families in poverty over number of 5 – 17-year-old children in the poverty universe), poverty "ratios" (defined as number of children ages 5 - 17 in poverty over total 5 - 17 population) are used. This facilitates converting model-based estimates of the ratios to model-based estimates of number of 5 - 17 children in poverty—we multiply the model-based estimate of the ratio by a demographic estimate of the 5 - 17 population, avoiding the need to construct an estimate of the poverty universe. The same could be done at the county level, i.e., model county poverty ratios instead of county poverty rates.

[17] In the basic county poverty rate model, the regression variables are the logarithms of a pseudo-IRS child poverty rate, a rate of filing of tax returns, a food stamp program participation rate, and the previous census's 5 - 17 poverty rate.

obtained from the state model results. Most results in this chapter are for unraked regression predictions and shrinkage estimates. Section 2.7 briefly discusses raking factors, examining the amount of raking adjustment required to force model-based county estimates from 2005 ACS data to aggregate to ACS survey and model-based state estimates. With the larger sample of ACS the need for modeling at the state level is not obvious, so both the survey and model-based state estimates are considered. Also, because of the importance of the state estimates for the raking of the county estimates, Section 2.7 presents some results on applying the SAIPE state 5 - 17 poverty ratio model to ACS data.

The remainder of the chapter considers three additional topics. Section 2.5 considers choice of the timing of the regression variables in log-level models. This is pertinent because, as discussed there and in Section 1.4.3, the ACS poverty estimates do not relate to income reports strictly for a calendar year, but for a period spanning nearly two years—thus, there are two alternative choices for which reference year to use in determining the regression variables (which generally refer to a specific calendar year). Section 2.6 discusses estimation of the sampling error variances of the ACS estimates used here, which is an important part of the model development. Section 2.7 discusses the state model and the resulting raking factors.

2.1.1 Fay-Herriot models

The county models examined here, both log-level and log-rate models, and the state poverty ratio model used in Section 2.7, reflect the general form suggested by Fay and Herriot (1979) written below. Because the county models considered here are all variations on the log-level and log-rate forms, the general model below is written explicitly in terms of the logs of the direct survey estimates (of number in poverty or poverty rate) and the logs of the corresponding true population quantities. The model is

$$\log(y_i) = \log(Y_i) + e_i \quad \text{where} \quad e_i \sim ind.\ N(0, \mathrm{v}_i) \tag{1}$$

$$\log(Y_i) = x_i'\beta + u_i \quad \text{where} \quad u_i \sim i.i.d.\ N(0, \sigma_u^2) \tag{2}$$

where, for county i,

y_i = ACS survey estimate of 5 - 17 poverty (number in poverty or poverty rate)
Y_i = true population value of 5 - 17 poverty (number in poverty or poverty rate)
$e_i = \log(y_i) - \log(Y_i)$ = sampling error in $\log(y_i)$ as an estimate of $\log(Y_i)$
x_i = vector of regression variables
β = vector of regression parameters
u_i = random model error (county random effect).

Different versions of the ACS county model are defined by choosing y_i to be either the ACS estimate of the county number of 5 - 17 in poverty or the 5 - 17 county poverty rate, and by different choices of the regression variables contained in x_i. The latter includes an intercept term, regression predictors from administrative sources, and the previous census (2000) estimate corresponding to y_i. These models are discussed in Sections 2.2 and 2.3.

The sampling error variances, v_i, of the $\log(y_i)$, are treated as known. In reality, they are estimated by replication methods, as discussed in Section 2.6. Generalized variance functions (GVFs) for smoothing these estimates are also being investigated, but not included in this report.

The model defined by equations (1) and (2) is estimated by maximum likelihood, the unknown parameters being β and σ_u^2. Given σ_u^2, the maximum likelihood estimation of β is obtained by weighted least squares regression of $\log(y_i)$ on x_i, as follows:

$$\hat{\beta} = (X'V^{-1}X)^{-1}X'V^{-1}\mathbf{y} \tag{3}$$

$$\mathrm{Var}(\hat{\beta}) = (X'V^{-1}X)^{-1} \tag{4}$$

where X is a matrix whose rows are given by the row vectors of regression variables x_i', $\mathbf{y} = (y_1,\ldots,y_m)'$ is the vector of ACS survey estimates for all m counties used in the regression prediction, and V is a diagonal matrix with elements given by the total variances $\sigma_u^2 + v_i$. Thus, maximization of the likelihood can be performed by iterative weighted least squares, iterating between computing β by weighted least squares for given σ_u^2, and maximizing the likelihood over σ_u^2 for given β.

Having estimated the model by maximum likelihood, shrinkage estimates (empirical best predictions) in the log scale, i.e., predictions of log (Y_i), and the corresponding prediction error variances are computed by plugging the parameter estimates into the following standard formulas (Bell 1999):

$$\widehat{\log(Y_i)} = (1 - w_i)\log(y_i) + w_i(x_i'\hat{\beta}) \quad \text{where} \quad w_i = v_i / (\hat{\sigma}_u^2 + v_i) \tag{5}$$

$$\mathrm{Var}[\log(Y_i) - \widehat{\log(Y_i)}] = w_i\hat{\sigma}_u^2 + w_i^2\left(x_i'\mathrm{Var}(\hat{\beta})x_i\right) \tag{6}$$

The variances given by (6) treat the model error variance, σ_u^2, as known. Asymptotic corrections to this formula to account for the error in estimating σ_u^2 by maximum likelihood are given by Datta and Lahiri (1995). These asymptotic corrections have been found to be minor in the CPS ASEC county model, and likely the same is true in the ACS county models. This will be examined in future research.

Predictions of Y_i on the original scale (not logged), and associated variances, are defined from the above results assuming a log-normal distribution.

$$\hat{Y}_i = \exp\left(\widehat{\log(Y_i)}\right)\exp\left[\mathrm{Var}\left(\log(Y_i) - \widehat{\log(Y_i)}\right)/2\right] \tag{7}$$

$$\mathrm{Var}(\hat{Y}_i) = \hat{Y}_i^2 \left[\exp\left(\mathrm{Var}[\log(Y_i) - \widehat{\log}(Y_i)]\right) - 1\right] \qquad (8)$$

For some counties $y_i = 0$, and these counties are excluded from the regression prediction because the log of zero cannot be taken. As noted in Section 1.4.1, however, this occurs much less frequently with ACS data than with CPS ASEC data, and so is much less of a problem. For such counties, the pure regression predictions are used for the shrinkage estimates, as would be the case if a county had no sample data.[18] In these cases, equation (5) reduces to $\widehat{\log}(Y_i) = x_i'\hat{\beta}$, and equation (6) to $\mathrm{Var}[\log(Y_i) - \widehat{\log}(Y_i)] = \hat{\sigma}_u^2 + \left(x_i'\mathrm{Var}(\hat{\beta})x_i\right)$; that is, effectively $w_i = 1$ in both equations.

2.1.2 County estimation methodology with CPS ASEC data

For comparison purposes and for historical reference, this section summarizes how county poverty models have been applied to CPS ASEC data. As noted in Chapter 1, models were applied to three-year average CPS ASEC county estimates of the log(number of 5 - 17 related children in poverty). The model is in the form of equations (1) and (2), with the predictor variables as described in Section 2.2 on the county log-level models for ACS data. The models have been estimated by maximum likelihood, with predictions of the number in poverty and associated variances obtained from equations (5)-(8).

Two main complications arose in applying the models to the CPS ASEC data. The first complication is that direct estimates of the sampling error variances of the CPS ASEC estimates have not been available.[19] To address this problem, the model error variance, σ_u^2, was first estimated by fitting a model in the form of (1) and (2) (by maximum likelihood) to the previous census (2000) estimates of the number of 5 - 17 in poverty, with regression variables defined for 1999, the census "income year."[20] In this "census equation," the sampling error variances were treated as known and set equal to available estimates of sampling error variances of the Census 2000 long-form estimates (using Taylor series linearization to account for the log transformation). The fitting of the census equation was used solely to estimate σ_u^2. Next, the model error variances in the census and the CPS ASEC model equations were assumed to be the same. Then, a simple parametric form for the sampling error variances, v_i , was postulated. Originally, $v_i = \gamma_e/n_i$ was used where γ_e is a parameter to be estimated and n_i is the CPS ASEC sample size for county i expressed in terms of the number of households in sample (cumulated over the three years of data used). Variance diagnostics applied to model residuals showed some problems–dependence of variances of the "standardized" model residuals on both sample size

---

[18] Most (about 2/3) of U.S. counties have no CPS ASEC sample, but all counties in the U.S. have ACS sample. So the use of the pure regression predictions from the ACS county model occurs only for the counties with ACS sample but with $y_i = 0$.

[19] As noted earlier, this situation should change shortly with the recent construction of replicate weight files for the CPS ASEC data.

[20] Except that the 2000 census equation included 1990 (not 2000) census estimates as a regression variable.

and previous census poverty rate. Thus, later the parametric form of the sampling error variances was switched to $v_i = \gamma_e/(n_i)^{.5}$, which produced a better regression prediction (Fisher and Asher 2000). It was recognized, though, that obtaining sampling error variances from GVFs fitted to direct CPS ASEC county sampling error variance estimates would be desirable.

The second complication involves dealing with counties with no 5 – 17-year-olds in poverty in the CPS ASEC sample (implying $y_i = 0$). This has been dealt with as discussed above for ACS data, but has been much more of a problem because of its much more frequent occurrence due to the small CPS ASEC county sample sizes. The seemingly related issue of many counties (about two-thirds) having no CPS ASEC sample does not actually pose a methodological complication. These counties have no data to contribute to the regression prediction, and the use of pure regression predictions for these counties is clearly (under the model assumptions) the appropriate procedure.

2.2 LOG-LEVEL MODELS

This section contains detailed results and tests for the log-level model structure. The section starts by defining the variables in the model and reporting summary statistics and relations. Regression prediction results for the version of this model currently used for SAIPE production are given next, followed by analysis of residuals and of shrinkage estimates. The section concludes by comparing several alternative versions of the log-level model, with discussion of the structural meaning of the restrictions required for each. The results of these comparative tests reinforce the validity of focusing on the log-level type of model used for past SAIPE production.

2.2.1 Variable definitions

Table 2.1 lists variable definitions for the models examined in this section. The dependent variable in all the models is the log of the direct 2005 ACS county estimates of the number of related children ages 5 - 17 in poverty. As discussed in Section 1.4.3, the 2005 ACS represents a population distribution (controls) from the U.S. Census Bureau's Population Estimates Program (PEP) for July 1, 2005, while the poverty estimates use income reports covering 12-month spans that start as early as January 2004 and end as late as November 2005. The right-hand side variables in the base model (LL6) are those from the model used to produce the official 2004 SAIPE county poverty estimates. The tax data are for income year 2004, the Food Stamp participants are for July 2004, and the Census 2000 estimates refer to income year 1999. Population estimates for July 1, 2005 are used to correspond with the ACS survey controls and to correspond approximately with the population distribution of tax filings, which are mostly made between January and May of the year following the income year.

Table 2.1: Variable definitions for the log-level models (LL6 and LL8)

Short Name	Description
Dependent Variable	
Log (ACS poor, ages 5-17 related)	Log estimated county number of 5 - 17 related children in poverty from the 2005 ACS.
Regressors for LL6	
Log (IRS child tax-poor exemptions)	Log number of county tax-poor child exemptions from IRS administrative records, where tax-poor is defined as Adjusted Gross Income (AGI) below the poverty level for a household size defined by the total number of exemptions on the return.
Log (Food Stamp Program participants)	Log number of county Food Stamp Program participants reported in July (data from the USDA Food and Nutrition Service), raked to a control total obtained from state Food Stamp participant data.
Log (PEP population, ages 0-17)	Log county population, ages 0 - 17, as of July 1, 2005, from the Census Bureau's Population Estimates Program (PEP) of intercensal demographic estimates.
Log (IRS child tax exemptions)	Log total number of county child exemptions from IRS administrative records.
Log (Census 2000 poor, ages 5-17 related)	Log estimated county number of 5 - 17 related children in poverty from Census 2000.
Additional Regressors for LL8	
Log (PEP population, all ages)	Log county population, all ages, as of July 1, 2005, from the PEP intercensal demographic estimates.
Log (Census 2000 poverty universe, ages 5-17 related)	Log number of county 5 - 17 related children from Census 2000.
Further information about these input data is available on the SAIPE program's webpage, http://www.census.gov/hhes/www/saipe/techdoc/inputs/datintro.html.	

The ACS provides estimates for every county, as noted earlier.[21] For some counties with small samples, the estimate of the number of related children ages 5 – 17 in poverty is zero by random chance. Since logs cannot be taken of these estimates, such counties are excluded from the regression prediction. Of the 3,140 counties in the SAIPE universe (Kalawao, Hawaii is excluded), for the 2005 ACS this led to the exclusion of 165 counties from the modeling. A further 3 counties were omitted for the complementary boundary issue of having poverty rate estimates of 100% (which leads to unreasonable direct variance estimates of zero). Thus, for all results in this report, the observations in models for the 2005 ACS estimates represent 2,972 counties.

Sampling error variances for the county log number in poverty are estimated directly using a set of replicate weights included in the survey micro-data file. Section 2.6 gives details on the variance estimation. Table 2.2 gives summary statistics for the variance estimates. Note that the standard deviation corresponding to the median of the variance estimates is $(.186)^{.5} = .43$. Since standard deviations for the log estimates can approximate coefficients of variation (CV) on the original scale (though for standard deviations this large, the approximation is not so good), this value reflects substantial uncertainty in the direct ACS poverty estimates for many counties.

[21] The estimates used in this report are constructed from a single year of ACS data for all counties. Official published estimates for smaller counties (with populations less than 65,000) will be constructed from 3-year or 5-year (for counties with population less than 20,000) accumulations of ACS data.

Table 2.2: Estimated sampling error variances for log (2005 ACS number of children in poverty, ages 5 - 17 related)

Mean	0.359	Median	0.186
Std Deviation	0.447	Variance	0.200
Skewness	2.062	Kurtosis	3.796
Percentiles			
100%	2.230	50%	0.186
99%	1.895	25%	0.076
95%	1.632	10%	0.033
90%	0.913	5%	0.019
75%	0.440	1%	0.006
Coefficient of Variation (CV)			
Mean	0.093	Median	0.064
Min	0.002	Max	0.852

Figures 2.1 through 2.5 plot county-level 2005 ACS poverty estimates of related children ages 5 – 17 (the dependent variable) against each of the independent variables from the base model (LL6). In the figures, both the x- and y-axes are shown in log scale but labeled in linear scale. This use of log scaling will be seen in most of the graphs given in this report when the data involve population-scale numbers, e.g., population or poverty estimates, tax exemptions, etc. Also, counties with zero ACS poverty are not included in many figures.

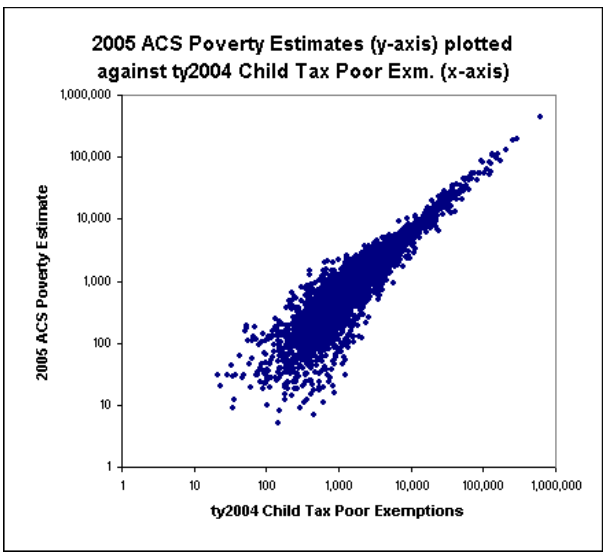
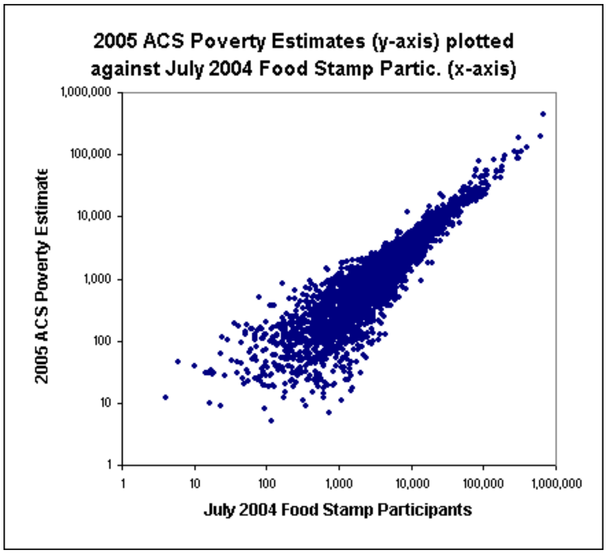


Figure 2.1: 2005 ACS number survey estimates, ages 5-17 related in poverty against 2004 IRS child tax-poor exemptions.
The x-axis is shown in log scale but labeled in linear scale.



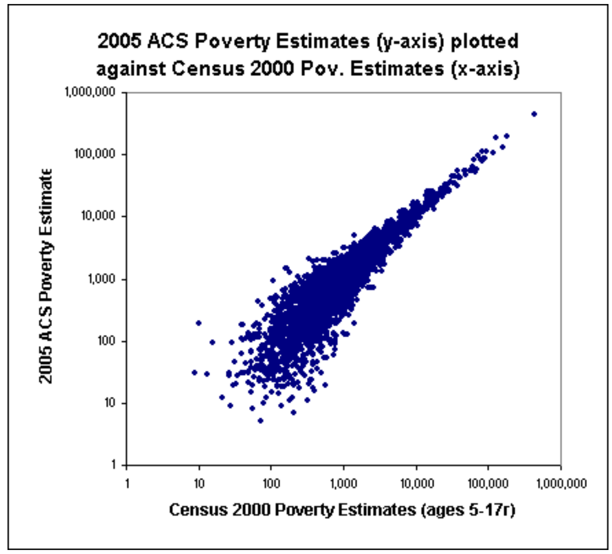
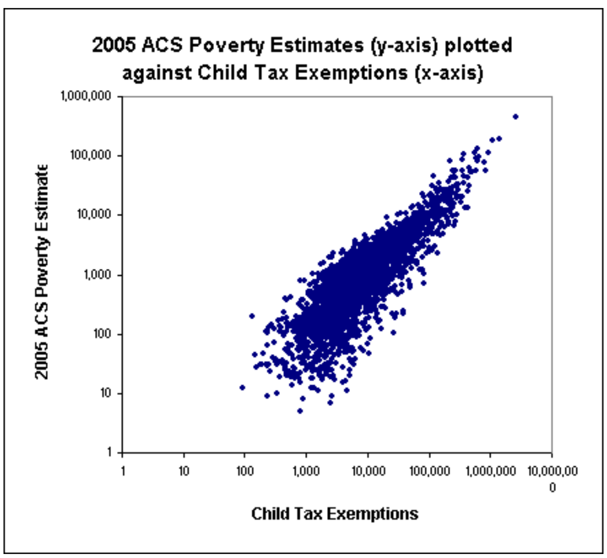


Figure 2.2 (left): 2005 ACS number survey estimates, ages 5-17 related in poverty against July 2004 Food Stamp Program participants.

Figure 2.3 (right): 2005 ACS number survey estimates, ages 5-17 related in poverty against Census 2000 poor, ages 5 - 17 related.



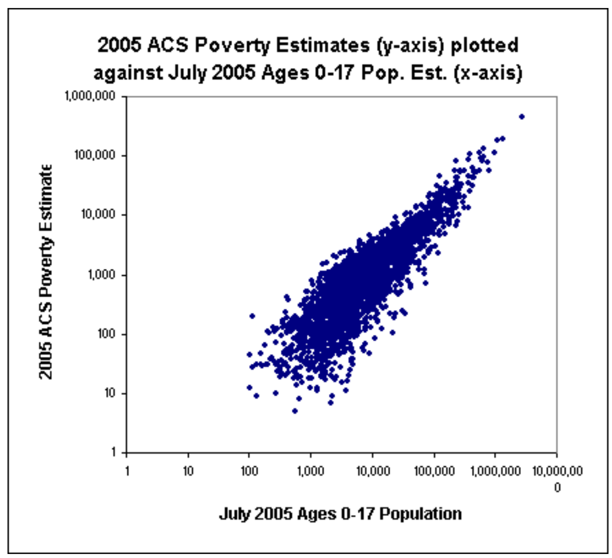


Figure 2.4 (left): 2005 ACS number survey estimates, ages 5-17 related in poverty against 2004 IRS child tax exemptions.

Figure 2.5 (right): 2005 ACS number survey estimates, ages 5-17 related in poverty against 2005 PEP population, ages 0 - 17.

Each of these graphs shows a strong positive correlation between the ACS estimate of log number in poverty and each of the independent variables. Including IRS child tax exemptions and PEP population, ages 0 – 17 in the model has in the past led to coefficient estimates with magnitudes that were not statistically significantly different, but with a negative coefficient for the log (IRS child tax exemptions) variable and a positive coefficient for the log (PEP population, ages 0-17) variable. This has been interpreted as providing a proxy for a log tax "filers" rate [= log (IRS child tax exemptions) - log (PEP population, ages 0-17)], for which a negative relation to poverty is expected.

2.2.2 Regression prediction results

Table 2.3 shows regression prediction results for the 2005 ACS log-level model with six 2004 regressors, with no restrictions on the coefficients (LL6u). The correlation matrix for the regression coefficients is given in Table 2.4. Discussion of models with the expanded regressor set and/or scale restrictions on the coefficients is given in Section 2.2.5. Model estimation was by maximum likelihood via an iterated weighted least squares algorithm. Convergence occurred in less than 10 iterations in all cases.

The coefficient estimates on all variables bear the expected sign, with the coefficient estimates positive except for the coefficient on the IRS child tax exemptions variable, which is negative as expected. The results can be summarized as follows: (*i*) the coefficient estimates are each individually highly significant, (*ii*) the log (IRS child tax-poor exemptions) variable is the most important (highest t-statistic), (*iii*) the R-squared is large, indicating that a lot of the variation in the dependent variable (mostly due to county population size) is being explained, and (*iv*) the sum of the coefficients on the log (PEP population, ages 0-17) and log (IRS child tax exemptions) variables is not statistically different from zero. This reinforces the interpretation of the net effect of these variables as relating to the effect of a log (tax filer rate) variable.

Table 2.3: Regression prediction results for unrestricted log-level model with six regressors (LL6u)

These results are based on direct ACS sampling error variances of dependent variable log (2005 ACS poor, ages 5-17 related).

Variable	Regression Coefficient	Standard Error	t	Pr>|t|
Intercept	-0.421	0.057	-7.39	<0.0001
Log (IRS child tax-poor exemptions)	0.548	0.045	12.26	<0.0001
Log (Food Stamps)	0.173	0.022	7.90	<0.0001
Log (PEP population, ages 0-17)	1.050	0.122	8.62	<0.0001
Log (IRS child tax exemptions)	-1.037	0.114	-9.13	<0.0001
Log (Census 2000 poor, ages 5-17 related)	0.268	0.030	8.90	<0.0001
Degrees of freedom = 2966				
Model error variance = 0.022				
R-squared = 0.936				

Table 2.4 reports the correlation matrix for the regression coefficient estimates. The correlation for the coefficients on the IRS child tax exemptions and PEP population, ages 0-17 variables is nearly negative one. Since the estimate values are equal and have opposite signs as well, this is further evidence that the two terms could be combined without appreciably changing the model predictions or standard errors.

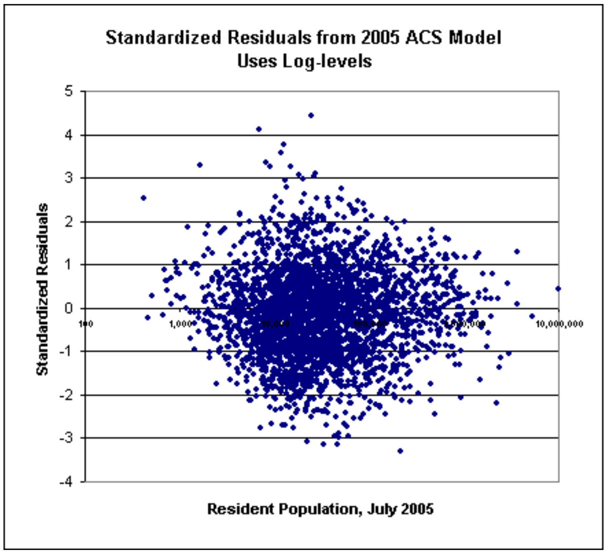
Table 2.4: Correlation matrix of coefficient estimates (LL6u)

	Intercept	Log (IRS child tax-poor exemptions)	Log (Food Stamps)	Log (PEP population, ages 0-17)	Log (IRS child tax exemptions)	Log (Census 2000 poor, ages 5-17 related)
Intercept	1.000					
Log (IRS child tax-poor exemptions)	0.185	1.000				
Log (Food Stamp Program participants)	-0.496	-0.329	1.000			
Log (PEP population, ages 0-17)	0.125	-0.098	-0.041	1.000		
Log (IRS child tax exemptions)	-0.201	-0.053	0.076	-0.983	1.000	
Log (Census 2000 poor, ages 5-17 related)	0.217	-0.609	-0.353	0.295	-0.234	1.000

2.2.3 Analysis of residuals

Figures 2.6 through 2.10 plot the standardized residuals from the LL6u model against the values of the following variables: PEP county total resident population, 2005 ACS county sample sizes, LL6u regression predictions, Census 2000 county poverty rates for all ages, and Census 2000 county percent rural.

These figures contain a small number of standardized residuals greater than three. The large, positive residuals all represent counties with high ACS sample poverty levels compared with the census poverty levels or tax-poor numbers. From Figure 2.7, it appears that most of these come from counties with fairly small ACS sample sizes. One possible explanation for these outliers is that the sampling distribution of the estimates from counties with small ACS samples has wider tails than the normal distribution. Another possible explanation is that these outliers appear because the direct sampling error variance estimates used here are not very reliable for counties with small ACS samples. This may be addressed by replacing the direct variance estimates with smoothed estimates from fitted GVFs (something that we are currently investigating.)



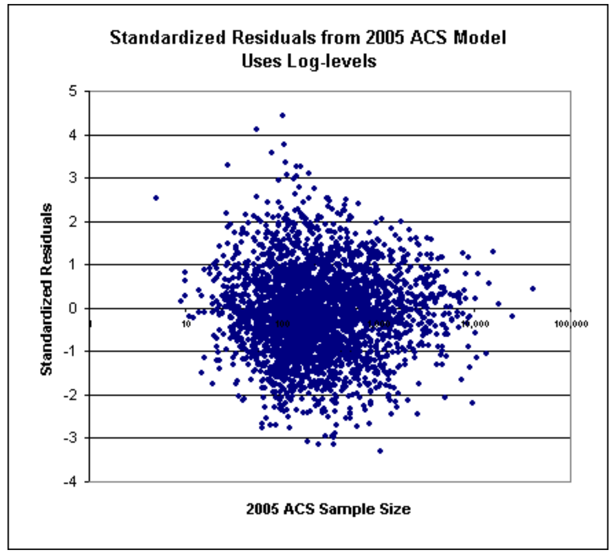


Figure 2.6 (left): Standardized residuals from 2005 ACS log-level regression against 2005 PEP county total resident population.
Figure 2.7 (right): Standardized residuals from 2005 ACS log-level regression against 2005 ACS sample size.

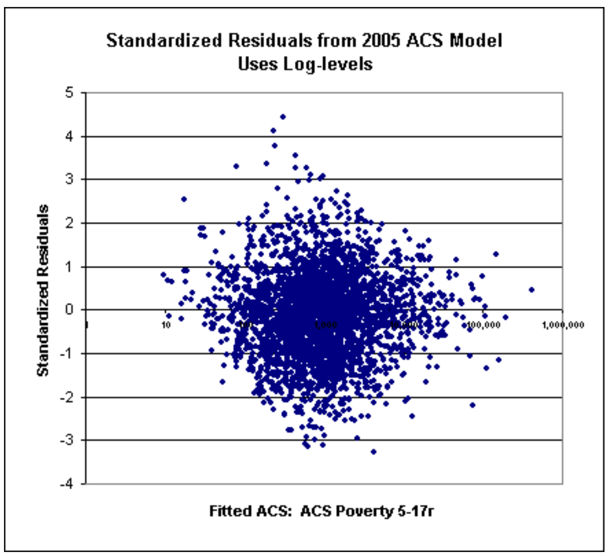
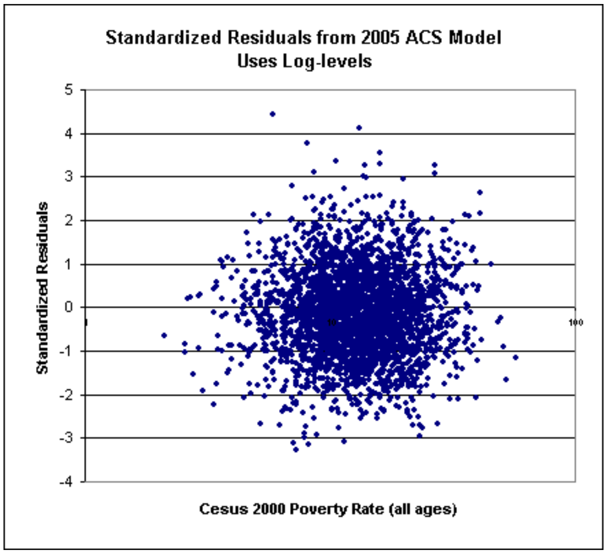


Figure 2.8: Standardized residuals from 2005 ACS log-level regression against 2005 ACS regression predictions.



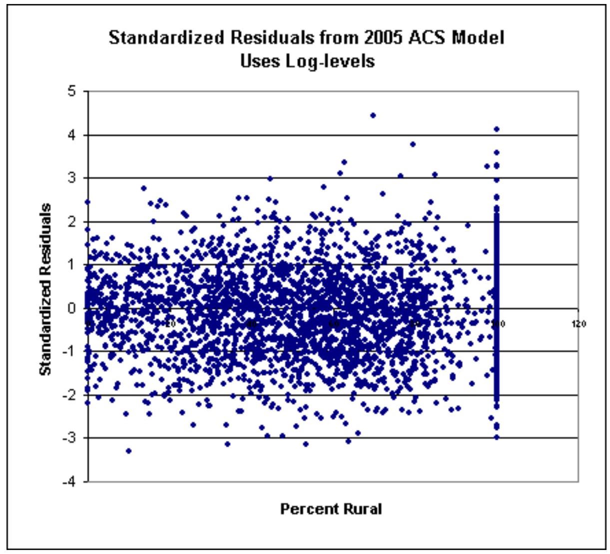


Figure 2.9 (left): Standardized residuals from 2005 ACS log-level regression against Census 2000 poverty rate, all ages.
Figure 2.10 (right): Standardized residuals from 2005 ACS log-level regression against Census 2000 percent rural.

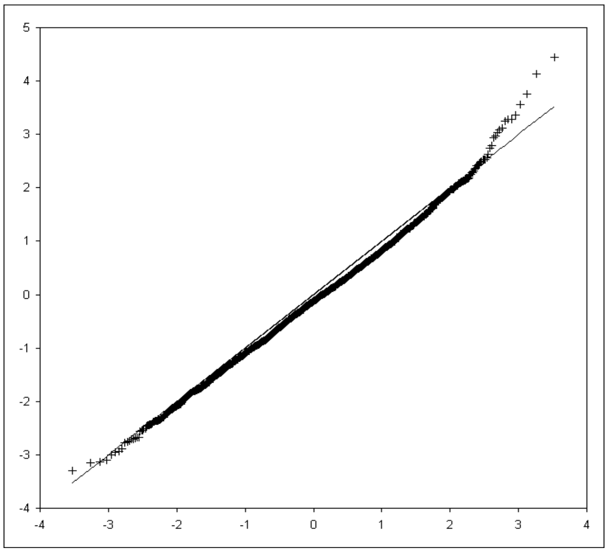
Box-whisker plots and additional scatter plots of the standardized residuals are included in Appendix B.1. These box-whisker plots group the categorization variables into quintiles. Each quintile contains 594 counties of the 2,972 counties from regression prediction. The categorization variables include the following concepts: B.1 Census 2000 total population, B.2 Census 2000 percent in poverty, B.3 Population growth Census 1990 - Census 2000, B.4 Population growth Census 2000 – 2005 PEP total population, B.5 Census 2000 percent Black, B.6 Census 2000 percent Hispanic, B.7 Census 2000 percent Asian, B.8 Census 2000 percent group quarters. The Appendix B.1 figures show mild model overprediction for counties within the middle quintile of population Asian. The figures also show some small model overprediction for counties within the lowest quintile of percent population Black. There were no discernible patterns with respect to the other classification variables.

Spearman's rank correlation coefficients between the squared standardized residuals and four covariates, selected because of their expected importance, are reported in Table 2.5, along with asymptotic t-statistics. Significant correlations would be taken as evidence of possible heteroskedasticity related to the relevant variable. The Spearman's statistics reported in Table 2.5, however, confirm the general patterns, or rather lack of patterns, seen in the plots. These tests fail to reject the assumption of homogeneous variances along the directions of any of these covariates.

Table 2.5: Spearman's rank correlation coefficients for squared standardized residuals of model LL6u

Covariates	Spearman's ρ	Asymptotic t-statistic
2005 PEP total county resident population, all ages (log)	-0.015	-0.84
2005 ACS sample size, household count (log)	-0.016	-0.83
Census 2000 poverty rate, all ages	-0.001	-0.08
Census 2000 percent rural	0.021	1.17

Figures 2.11 and 2.12 describe the distribution of the standardized residuals for the LL6u model. Figure 2.11 clearly displays a substantially higher number of large, positive residuals than expected. The standard normal distribution predicts approximately four standardized residuals should be higher than three. In this regression, there are eleven such large, positive residuals. Large negative residuals total four, which equals expectations.



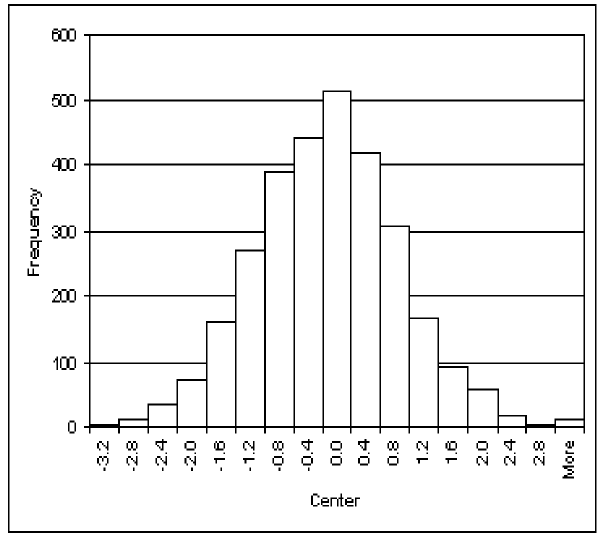


Figure 2.11 (left): Quartiles of standardized residuals for LL6u vs. standard normal quartiles.
Figure 2.12 (right): Frequency distribution of standardized residuals for LL6u model.

2.2.4 Analysis of shrinkage estimates

This section compares the shrinkage estimates and their standard errors obtained from the LL6u model with results for the single-year 2005 ACS direct survey estimates.

Precision – Standard errors and CV reduction

The computation of the standard errors of the ACS shrinkage estimates for poverty on the original (not logged) scale, [the square roots of the prediction error variances] is via equations (6) and (8) of Section 2.1. The standard errors for the ACS direct survey estimates are the square roots of the direct sampling error variance estimates (with limitations for counties with small samples as noted earlier) obtained as discussed in Section 2.6. The CV for the direct survey estimate is computed as the standard error of the direct survey estimate divided by the direct survey estimate, and the CV of the shrinkage estimate is computed as the standard error of the shrinkage estimate divided by the shrinkage estimate.

Figure 2.13 shows the frequency distribution of the ratio of the standard errors of the ACS shrinkage estimates to the standard errors of the ACS direct survey estimates. Much of the density appears roughly between ratios of 0.3 and 0.5.

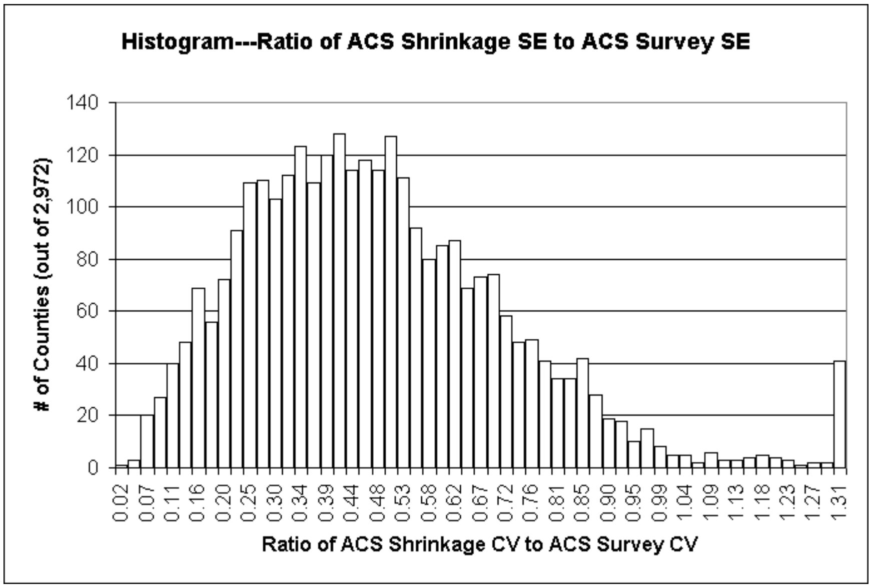
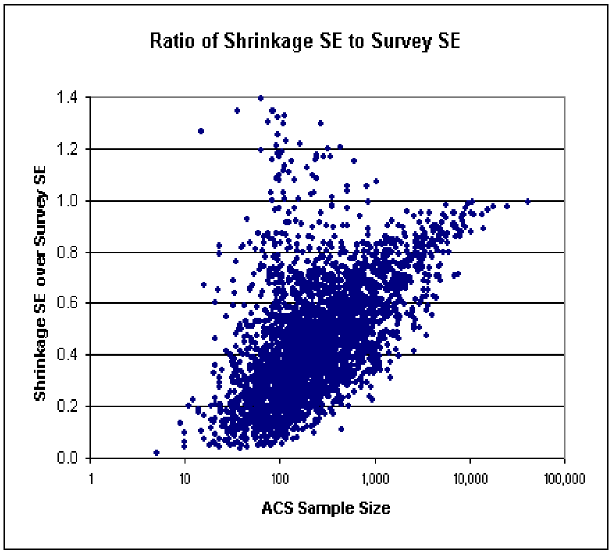


Figure 2.13: Frequency distribution of the ratios of the standard errors of the 2005 ACS shrinkage estimates to the standard errors of the 2005 ACS direct survey estimates.
The numbers shown on the x-axis are the lower bound cutoff for each bar. The rightmost bar contains all points higher than 1.31.

Figure 2.14 shows the ratios of the standard errors of the ACS shrinkage estimates to the standard errors of the ACS direct survey estimates against ACS sample size. The median ratio is 0.44 (i.e., 44 percent), which corresponds with the model producing a 56-percent decline in the SE of the direct ACS county poverty estimates. Figure 2.15 shows the ratios of the CVs of the ACS shrinkage estimates to the CVs of the ACS direct survey estimates against population. The median ratio is 0.35 (i.e., 35 percent), which corresponds with the model producing a 65-percent decline in the CV of the direct ACS county poverty estimates. The direct estimate standard errors and CVs are reduced by modeling more for counties with smaller ACS sample sizes than those with larger ACS sample sizes. Counties with large sample size have relatively precise direct survey estimates that receive a high weight in the shrinkage estimate. In Figure 2.14, the ratios of standard error greater than one occur for counties in which the shrinkage estimate is much larger than the direct survey estimate. The standard errors for the shrinkage estimates are accordingly higher. For these counties, despite reduction in the CV from modeling, their standard errors are higher from modeling due to the higher shrinkage estimates.



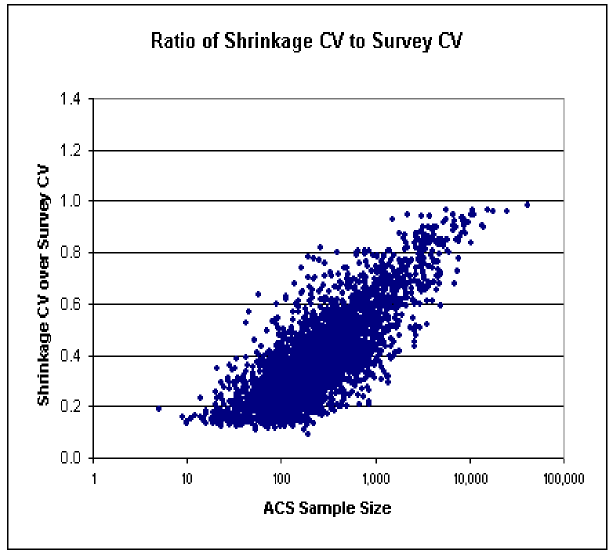


Figure 2.14 (left): Ratios of the standard errors of the 2005 ACS shrinkage estimates to the standard errors of the 2005 ACS direct survey estimates against 2005 ACS sample size.
Figure 2.15 (right): Ratios of CVs of the 2005 ACS shrinkage estimates to the CVs of the 2005 ACS direct survey estimates against 2005 ACS sample size.

Figure 2.16 is a scatter plot of the CVs of the ACS shrinkage estimates against the CVs of the ACS direct survey estimates. The CVs of the ACS shrinkage estimates are uniformly below 17 percent, while the CVs of the ACS survey estimates range up to 160 percent. We see a clear improvement from modeling for the majority of counties (which typically are small to medium size counties).

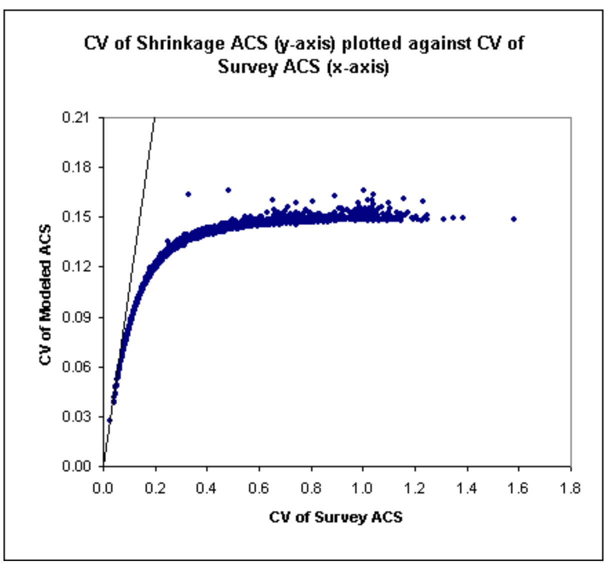


Figure 2.16: CVs of the 2005 ACS shrinkage estimates against the CVs of the 2005 ACS direct survey estimates.

Weights on direct survey estimates

Figure 2.17 shows the frequency distribution of the weights given to the direct survey estimates by the shrinkage estimates. These are the quantities $1 - w_i$ where w_i is given in conjunction with equation (5) of Section 2.1. Most counties have weights on their direct survey estimates of less than 11 percent, the median weight being 0.106. Counties with direct survey estimates of zero poor, ages 5 – 17 related are not included in the histogram.

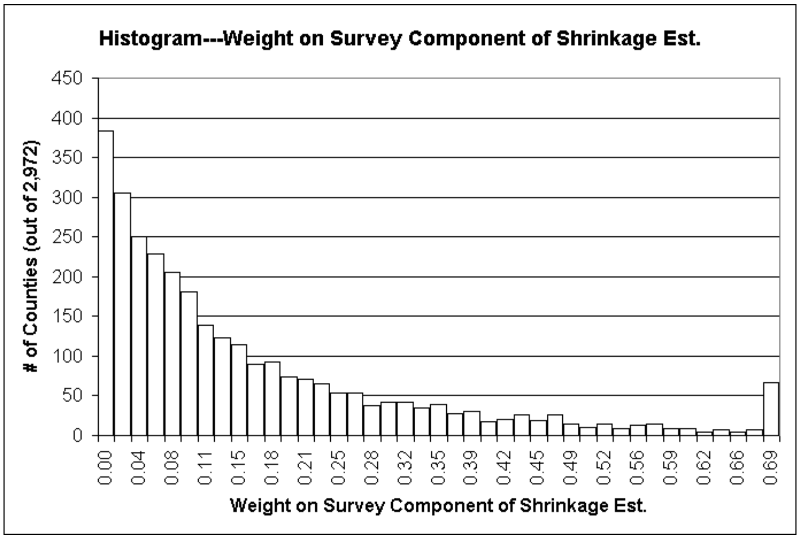


Figure 2.17: Frequency distribution of the weights on the survey component of the shrinkage estimates.
The numbers shown on the x-axis are the lower bound cutoff for each bar. The rightmost bar contains all points higher than 0.69.

Figure 2.18 plots the weights given to the direct survey estimates by the shrinkage estimates against the ACS sample sizes. Many counties with sample sizes greater than 1,000 have weights on the direct survey estimates greater than 60 percent. The points along the 1.0 line are counties with ACS 5 - 17 poverty estimates of zero.

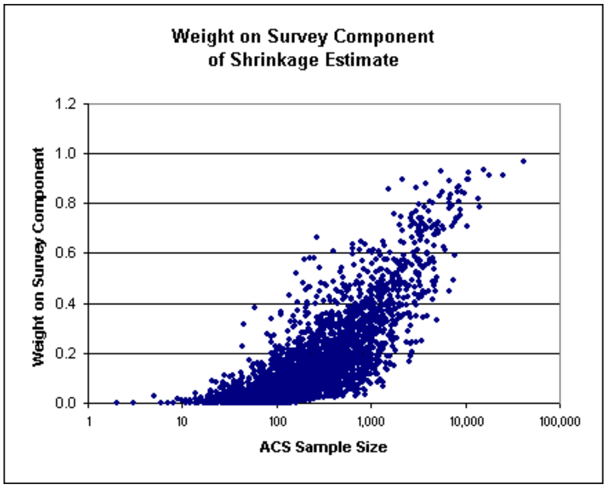


Figure 2.18: Weight on the survey component of the shrinkage estimates against 2005 ACS sample size.

Estimates of number in poverty, ages 5-17 related

Figure 2.19 shows the frequency distribution of the ratios of the shrinkage estimates to the ACS direct survey estimates. Much of the density appears roughly between ratios of 0.8 and 1.5.

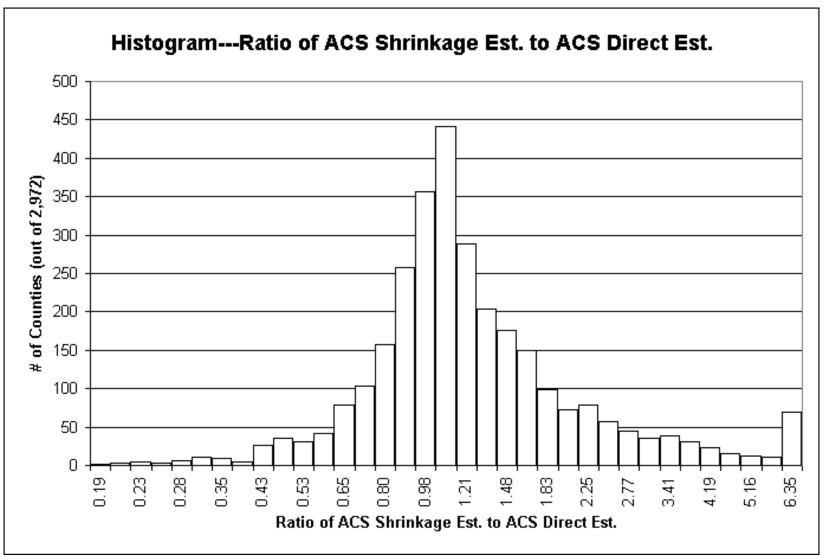
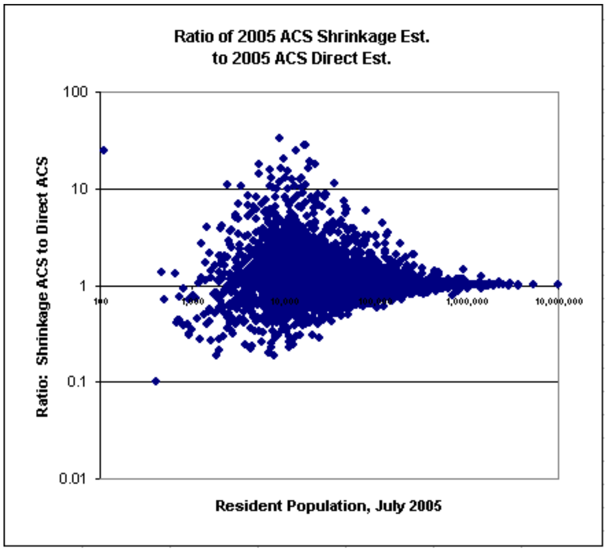


Figure 2.19: Frequency distribution of the ratios of 2005 ACS shrinkage estimates to the 2005 ACS direct survey estimates.
The numbers shown on the x-axis are the lower bound cutoff for each bar.
The rightmost bar contains all points higher than 6.35.

Figure 2.20 plots the ratios of the shrinkage estimates to the 2005 ACS direct survey estimates against 2005 PEP county total resident population. Most of the ratios range from about 0.2 to about 33, signifying that the model-based estimates are at times from one-fifth to about thirty-three times large as the ACS direct survey estimates. Figure 2.21 plots the ratios of the Census 2000 poverty estimates for ages 5 – 17 to the 2005 ACS direct survey estimates against the 2005 PEP county total resident population. The ratios vary similarly from about 0.1 to about 35. Figures 2.20 and 2.21 show the expected instability of the direct survey estimates for counties with low sample size. Similar plots of the survey estimates against Food Stamp Program participants and child tax-poor exemptions (not shown) visually have similar dispersion.



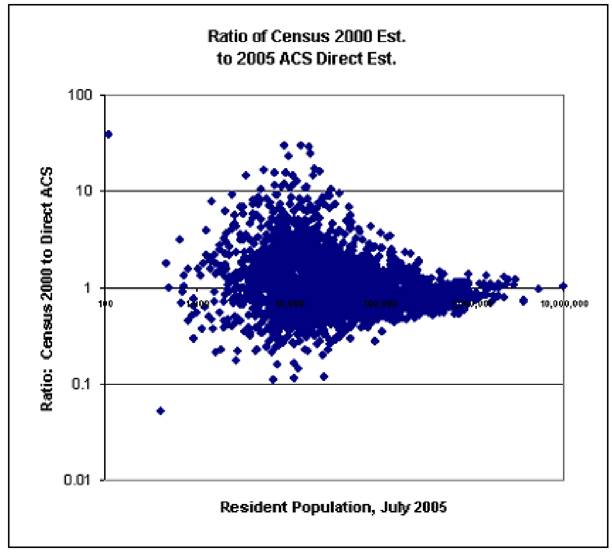
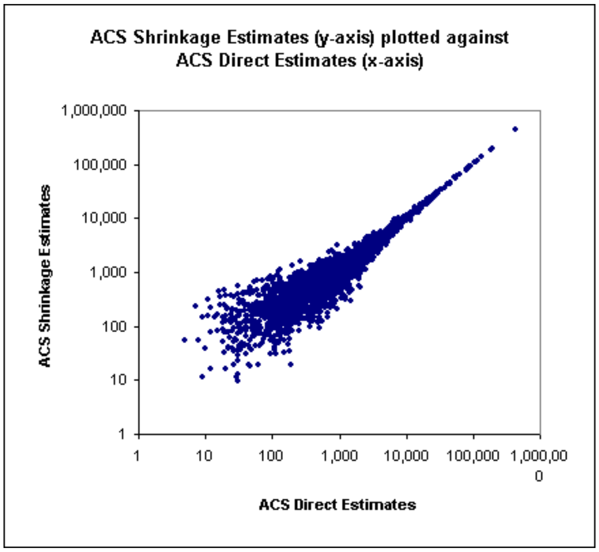


Figure 2.20 (left): Ratios of the 2005 ACS shrinkage estimates to the 2005 ACS survey estimates against 2005 PEP county total resident population.

Figure 2.21 (right): Ratios of Census 2000 poor, ages 5 - 17 related, to the 2005 ACS direct survey estimates against 2005 PEP county total resident population.

Figure 2.22 plots the shrinkage estimates against the 2005 ACS direct survey estimates. There is greater dispersion for counties with small shrinkage estimates or small direct survey estimates. Also, there appears to be some skewness at the bottom left tail suggesting that some counties with low ACS direct survey estimates have much higher ACS shrinkage estimates. This observation is consistent with the observation in Figure 2.20 of several counties with ratios above 10.0. Figure 2.23 plots the shrinkage estimates against the corresponding regression predictions. The shrinkage estimates generally appear fairly close to the regression predictions. This is the case even for counties with large shrinkage estimates or regression predictions, which presumably have high weights on the ACS direct survey estimates.



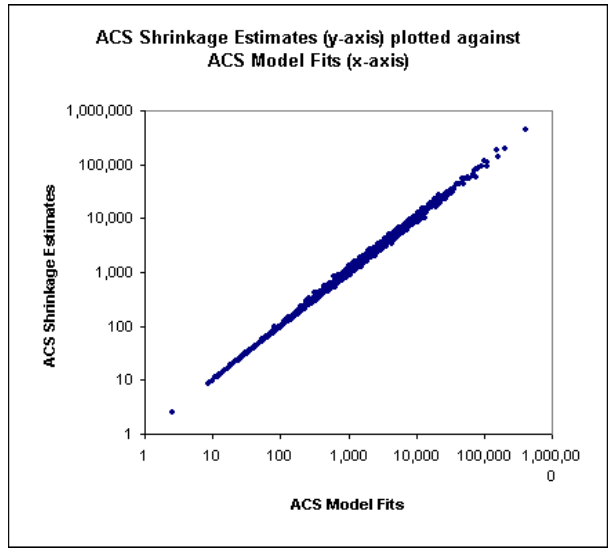


Figure 2.22 (left): 2005 ACS shrinkage estimates against 2005 ACS direct survey estimates.
Figure 2.23 (right): 2005 ACS shrinkage estimates against 2005 ACS regression predictions (or model fits).

2.2.5 Testing alternate restrictions and regressor sets

Two types of model comparisons are discussed in this report. This section discusses comparisons among different versions of log-level models. Because these involve comparisons among nested models, they can be analyzed in a classical hypothesis-testing framework. Section 2.4 examines non-nested comparisons between the log-level models and log-rate models.

We first discuss the comparison between the model with the 8-regressor set and the model with the 6-regressor set. Following that, we discuss tests of scale invariance restrictions on the regression coefficients.

The motivation for the unrestricted model with the eight regressor set, labeled LL8u, is to include as regressors denominator terms appropriate for transforming all log-level regressors into log-rate regressors. With this structure, the model's predicted values and residuals are invariant to whether the dependent variables are expressed as log-rate or log-level. To illustrate, consider terms involving the IRS tax-poor exemption variable and its appropriate denominator, IRS child tax exemptions, using the notation given in Table 2.6:

$$\begin{aligned}&\beta_1 \log(\text{Tax Poor}) + \beta_4 \log(\text{Ch Exemp}) \\ &= \beta_1 \{\log(\text{Tax Poor}) - \log(\text{Ch Exemp})\} + (\beta_4 + \beta_1)\log(\text{Ch Exemp})\end{aligned} \tag{9}$$

When denominators for all the regressors are included in the model, whether the dependent variable is expressed as log-rate or log-level has no impact on the residuals or predicted values. Only a re-definition of the coefficients occurs. A more general statement of this result is included in the Appendix B.2.

With the regressor set limited to six variables, two of the denominator variables are omitted, so this result will not hold exactly. Since all the denominator terms are highly correlated, however, similar regression predictions are expected. The hypothesis tests reported below comparing LL6 and LL8 provide a formal evaluation of this expectation.

The coefficient-restricted models, LL6r and LL8r, reported in Table 2.6 test the imposition of scale invariance restrictions. The motivation for these tests is the scale-invariance property that holds for a pure log-rate model, i.e., a model involving only log-rate variables. The definition of scale invariance can be stated as follows: if all scale variables increase by a scalar multiple, while all rates are unchanged, then the predicted number in poverty increases by the same scalar multiple. For the log-level model, the restriction implied by scale invariance is that the coefficients for all non-intercept terms sum to one. A proof of this condition can be found in the Appendix B.2.

Table 2.6 displays regression prediction results for all four versions of the log-level model considered. Restricted and unrestricted refer to the scaling restrictions discussed above, namely, restricting the slope coefficients to sum to one.

Table 2.6: Regression prediction results for four versions of the log-level model

These results use the 2005 ACS as the dependent variable with income year 2004 regressors and report: coefficient (standard error).

Variable		Model			
		LL6r	LL6u	LL8r	LL8u
Intercept	β_0	-0.394 (0.041)	-0.421 (0.057)	-0.218 (0.095)	-0.256 (0.104)
Log (IRS child tax-poor exemptions)	β_1	0.542 (0.044)	0.548 (0.045)	0.544 (0.047)	0.563 (0.052)
Log (Food Stamps)	β_2	0.173 (0.022)	0.173 (0.022)	0.181 (0.023)	0.180 (0.023)
Log (PEP population, ages 0-17)	β_3	1.082 (0.112)	1.050 (0.122)	1.118 (0.113)	1.069 (0.127)
Log (IRS child tax exemptions)	β_4	-1.064 (0.106)	-1.037 (0.114)	-1.062 (0.142)	-1.073 (0.143)
Log (Census 2000 poor, ages 5-17 related)	β_5	0.268 (0.030)	0.268 (0.030)	0.250 (0.035)	0.242 (0.037)
Log (PEP population, all ages)	β_6			-0.122 (0.061)	-0.117 (0.061)
Log (Census 2000 poverty universe, ages 5-17 related)	β_7			0.091 (0.097)	0.141 (0.114)
Degrees of freedom		2,966	2,966	2,964	2,964
Model error variance		0.0216	0.0217	0.0209	0.0211
AIC		4100.3	4101.8	4099.8	4101.0
Maximum log likelihood		-2044.14	-2043.90	-2041.91	-2041.52
Sum of slopes		1.0000	1.0036	1.0000	1.0055

Table 2.7 displays likelihood ratio test results for all the comparisons implied by the four models in Table 2.6. Asymptotically, this statistic should be distributed as chi-squared under the respective H0s. Wald and Lagrange multiplier tests for all these comparisons were also examined, but no different conclusions were obtained than those reported below.

The LL6r versus LL6u and LL8r versus LL8u tests examine the imposition of the scale invariance restriction. The LL6r versus LL8r and LL6u versus LL8u tests examine the omission of the two additional scale or denominator terms listed in Table 2.1. The final test in Table 2.7

tests the joint hypothesis of the additional denominator terms not being needed and the scale invariance restriction holding.

We fail to reject the null hypothesis for all these tests except one. For the test of the reduced variable set, with unrestricted coefficients (LL6u vs. LL8u), the null hypothesis is rejected at the 10% significance level but not the 5% level. The ordering of these results is nearly coincident with the AIC ordering in Table 2.6, since for these simple linear models there is little difference between the likelihood ratio and the AIC formula.

Overall, these test results do not reject the model with 6 explanatory variables and the scale invariance restriction on the coefficients (LL6r). Any variance advantage under LL6r will be relatively minor, as can be seen by comparing the standard errors of the regression coefficients between LL6r and LL6u. Thus, these tests generate confidence in use of either the LL6r or LL6u models for prediction.

Table 2.7: Likelihood ratio tests for the log-level models

Models	H0	Likelihood Ratio	Degrees of Freedom	Critical Value (α=0.10)	Critical Value (α=0.05)
LL6r vs. LL6u	sum of slopes = 1	0.48	1	2.71	3.84
LL8r vs. LL8u	sum of slopes = 1	0.78	1	2.71	3.84
LL6r vs. LL8r	$\beta_6 = \beta_7 = 0$	4.46	2	4.61	5.99
LL6u vs. LL8u	$\beta_6 = \beta_7 = 0$	4.76	2	4.61	5.99
LL6r vs. LL8u	joint both above	5.24	3	6.25	7.82

2.3 LOG-RATE MODELS

This section contains detailed results and tests for the log-rate model structure. The structure and numbering of the section is parallel with that from Section 2.2. The section starts by defining the variables in the model and reporting summary statistics and relations. Regression prediction results for the version of this model similar to the log-level model (LR6u) are given next, followed by residual analysis and shrinkage estimates. The section concludes by comparing several alternative versions of the log-rate model, with discussion of the structural meaning of the restrictions required for each. The result of these comparative tests reinforces the validity of focusing on the log-level version similar to that used in past SAIPE production.

2.3.1 Variable definitions

Throughout this section, the dependent variable is the log of the direct 2005 ACS county estimate of the poverty rate for related children ages 5 - 17. Table 2.8 lists variable definitions for the models evaluated in this section.

Table 2.8: Variable definitions for the log-rate models (LR6 and LR8)

Short Name	Description
Dependent Variable	
Log (ACS poverty rate, ages 5-17 related)	Log estimated county poverty rate of 5 – 17 related children from the 2005 ACS.
Regressors for LR6	
Log (IRS child tax-poor exemption rate)	Log number of county tax-poor child exemptions divided by total child exemptions, from IRS administrative records.
Log (Food Stamp rate)	Log number of county Food Stamp Program participants divided by total all-ages population from the PEP intercensal demographic estimates.
Log (PEP population, ages 0-17)	Log county population, ages 0-17, as of July 1, 2005, from the Census Bureau's Population Estimates Program (PEP) of intercensal demographic estimates.
Log (IRS child filing rate)	Log total number of county child exemptions from IRS administrative records divided by county ages 0 – 17 PEP intercensal demographic estimate.
Log (Census 2000 poverty rate, ages 5-17 related)	Log estimated county poverty rate for 5 - 17 related children from Census 2000.
Additional Regressors for LR8	
Log (PEP population, all ages)	Log county population, all ages, as of July 1, 2005, from the from the PEP intercensal demographic estimates.
Log (Census 2000 poverty universe, ages 5-17 related)	Log number of county 5 - 17 related children from Census 2000.
For further details and source information, see Table 2.1.	

Sampling error variances for the log rate values are estimated directly using a set of replicate weights included in the micro-data file. Summary statistics for these variance estimates are given in Table 2.9. The standard deviation corresponding to the median of the variance estimates is $(.160)^{.5} = .40$. Since standard deviations for the log estimates can approximate CVs on the original scale (though for standard deviations this large the approximation is not very good), this value reflects substantial uncertainty in the direct ACS poverty estimates for many counties. This median standard deviation for the log poverty rates is somewhat smaller than the median standard deviation (.43) for the log estimates of the number of 5 - 17 in poverty, due to the positive correlation between these estimates (which form the numerators of the rates) and the direct survey estimates of 5 - 17 population (actually, the poverty universe, which are the denominators of the rates). Thus, the log estimated 5 - 17 poverty universe variable is explaining part of the variation in the log estimated number of 5 - 17 in poverty.

Table 2.9: Estimated sampling error variances for log (2005 ACS poverty rates, ages 5-17 related)

Mean	0.325	Median	0.160
Standard Deviation	0.423	Variance	0.179
Skewness	2.271	Kurtosis	4.974
Percentiles			
100%	2.386	50%	0.160
99%	1.892	25%	0.068
95%	1.433	10%	0.031
90%	0.834	5%	0.019
75%	0.375	1%	0.006

Figures 2.24 through 2.28 plot the 2005 ACS poverty rate for related children ages 5 - 17 against the rate version of the independent variables. In the figures (except for Figure 2.27), both

the x- and y-axes are shown in log scale but labeled in linear scale. In Figure 2.27, the x-axis is in linear scale because the data do not vary over as wide a range.

The relations of the regressors to the ACS poverty estimates (for ages 5 - 17 related) are less strong with the variables expressed in rate form than in level form, i.e., there is more noise in the plots. This is as expected since some of the strength of the relations between the variables in level form is due to their all being strongly related to population size, and this relation is removed when the variables are expressed as rates. Figures 2.24, 2.25 and 2.26 show positive individual correlations between the ACS poverty rate and the child tax-poor rate, the Food Stamp Program participation rate, and the Census 2000 poverty rate. Figure 2.27 depicts a negative correlation between the IRS child filing rate and the ACS poverty rate. The IRS child filing rate on the x-axis is often greater than 100 percent because individuals can be claimed as child exemptions on tax returns even when they are ages 18 and over, though the computation of the IRS child filing rate uses ages 0 - 17 population in the denominator. Figure 2.28 plots the 2005 ACS county poverty rate estimates, ages 5-17 related against the 2005 PEP county population, ages 0 - 17. This variable is also one of the variables included in the log-level model. It is included in the unrestricted log-rate models (LR6u and LR8u) to account for a potential relation between county size and poverty rate.

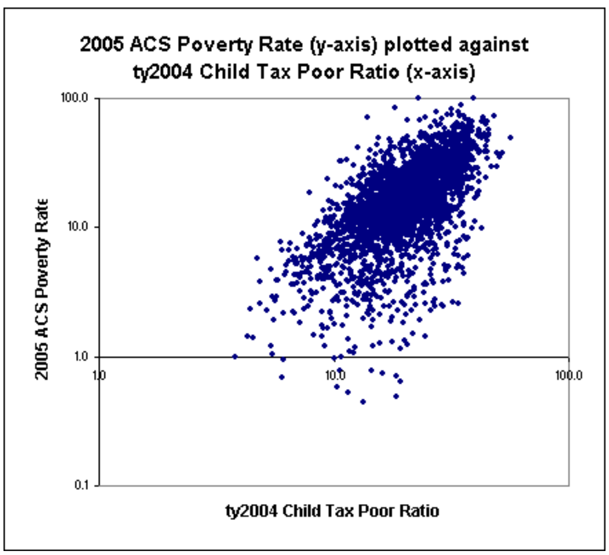
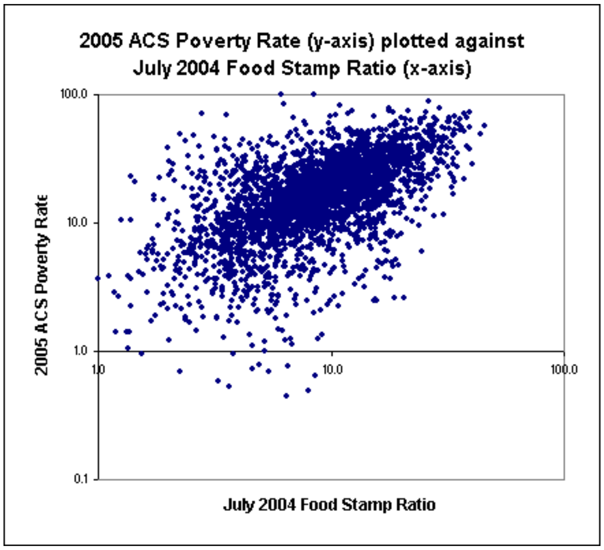


Figure 2.24: 2005 ACS rate survey estimates, ages 5-17 related against 2004 IRS child tax-poor rate.



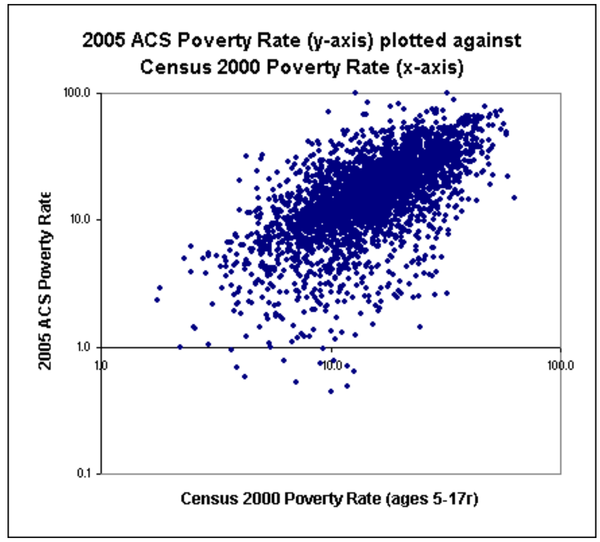
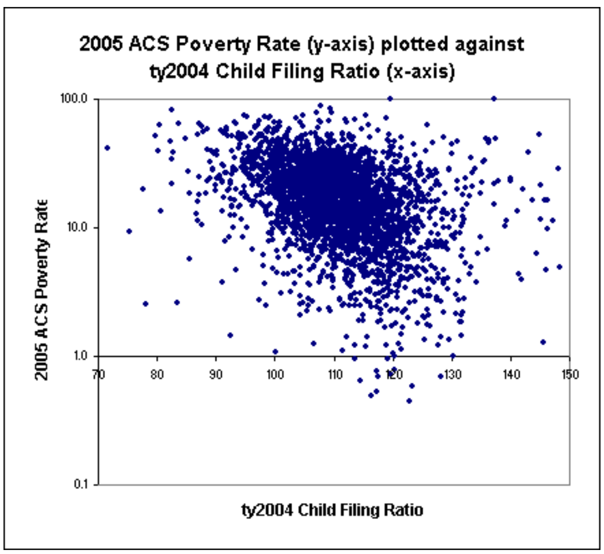


Figure 2.25 (left): 2005 ACS rate survey estimates, ages 5-17 related against July 2004 Food Stamp Program enrollment rate.
Figure 2.26 (right): 2005 ACS rate survey estimates, ages 5-17 related against Census 2000 poverty rate estimates, ages 5 - 17 related.



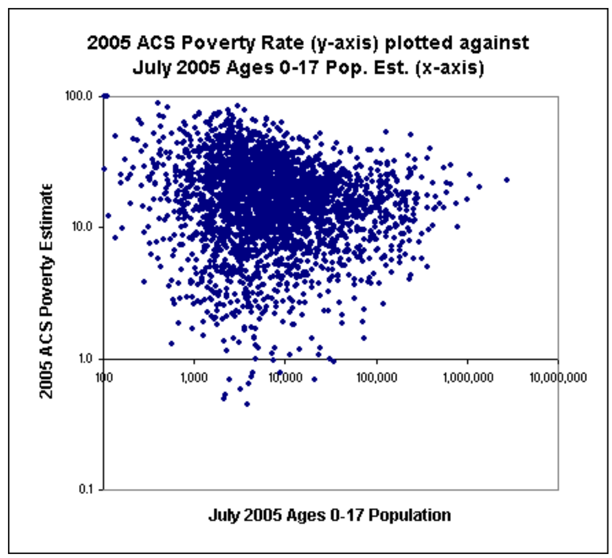


Figure 2.27 (left): 2005 ACS rate survey estimates, ages 5-17 related against 2004 IRS child filing rate.
Figure 2.28 (right): 2005 ACS rate survey estimates, ages 5-17 related against 2005 PEP population, ages 0 - 17.

2.3.2 Regression prediction results

Table 2.10 shows regression prediction results for the 2005 ACS unrestricted log-rate model with six 2004 regressor inputs (LR6u). The correlation matrix for the regression coefficients is given in Table 2.11. Discussion of models with the expanded regressor set, and/or the scale restriction on that sets the coefficient on the log (PEP population, ages 0-17) variable to zero are given in Section 2.3.5. Model estimation was again by maximum likelihood via an iterated weighted least squares algorithm.

Most of the results in Table 2.10 are analogous to those in Table 2.3 for the log-level model. First, the regression coefficient estimates are all statistically significant and all bear the expected sign, with a negative coefficient for the log (IRS child filing rate) variable. Second, the most significant coefficient is that for the log (child tax-poor rate). Since the fundamental difference between the log-rate models and the log-level models is the denominator (PEP population, ages 0-17) in the dependent variable, the significance of this coefficient provides some evidence against a pure rate model (such as LR6r and LR8r discussed later). Finally, the R-squared value (.620) is substantially lower than the R-squared value for LL6u (.936) since the latter benefits from the model explaining variation across counties in population size, something that is mostly explained by the denominator of the poverty rate variable in the log-rate model.

Table 2.10: Regression prediction results for unrestricted log-rate model with six regressors (LR6u)

These results are based on direct ACS sampling error variances of dependent variable log (2005 ACS poverty rate, ages 5-17 related).

Variable	Regression Coefficient	Standard Error	t	Pr > |t|
Intercept	0.372	0.048	7.73	<0.0001
Log (IRS child tax-poor exemption rate)	0.556	0.049	11.25	<0.0001
Log (Food Stamp rate)	0.167	0.022	7.64	<0.0001
Log (PEP population, ages 0-17)	-0.020	0.005	-3.70	0.0002
Log (IRS child filing rate)	-0.414	0.117	-3.55	0.0004
Log (Census 2000 poverty rate, ages 5-17 related)	0.289	0.035	8.33	<0.0001
Degrees of freedom = 2966				
Model error variance = 0.030				
R-squared = 0.620				

Table 2.11 reports the correlation matrix for the regression coefficient estimates. There is a pattern of negative correlations somewhat similar to that of Table 2.4, but without any of the correlations approaching negative one.

Table 2.11: Correlation matrix of coefficient estimates (LR6u)

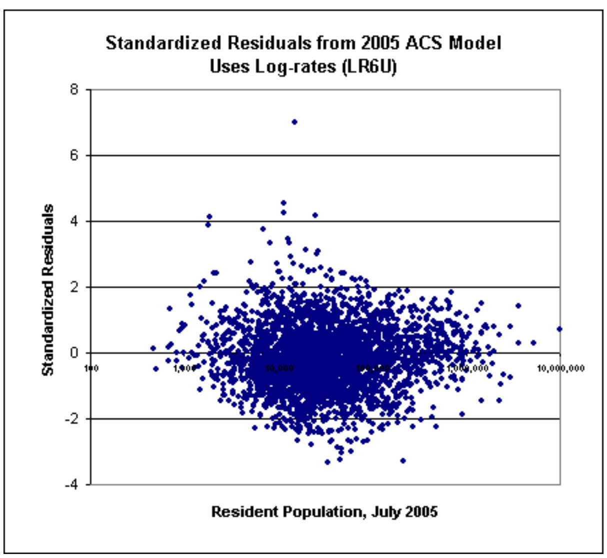
	Intercept	Log (IRS child tax-poor rate)	Log (Food Stamp rate)	Log (PEP population, ages 0-17)	Log (IRS child filing rate)	Log (Census 2000 poverty rate, ages 5-17 related)
Intercept	1.000					
Log (IRS child tax-poor exemption rate)	-0.043	1.000				
Log (Food Stamp rate)	0.134	-0.334	1.000			
Log (PEP population, ages 0-17)	-0.765	0.359	-0.003	1.000		
Log (IRS child filing rate)	-0.131	0.155	0.001	0.371	1.000	
Log (Census 2000 poverty rate, ages 5-17 related)	0.024	-0.671	-0.296	-0.126	0.163	1.000

2.3.3 Analysis of residuals

Figures 2.29 through 2.33 plot the standardized residuals from the LR6u model against the values of the following variables: 2005 PEP county total resident population, 2005 ACS county sample sizes, LR6u regression predictions, Census 2000 county poverty rates for all ages, and the Census 2000 county percent rural.

These figures contain a small number of standardized residuals larger than three. The large, positive residuals all represent counties with high ACS sample poverty levels compared with the Census poverty levels or tax poor numbers. Seven of the twelve counties with large, positive residuals in this model also had standardized residuals larger than three in the log-level regression. The explanation for these is the same as discussed in Section 2.2.3. Namely, from Figure 2.29 it appears that most of these come from counties with fairly small ACS sample sizes. One possible explanation for these outliers is that the sampling distribution of the estimates from counties with small ACS samples has wider tails than the normal distribution. Another possible explanation is that the occurrence of these outliers has to do with the direct sampling error variance estimates used here being not very reliable for counties with small ACS samples. This may be addressed by replacing the direct variance estimates with smoothed estimates from fitted GVFs, something that we are currently investigating.

From Figure 2.33 it appears that many of the large positive residuals occur for counties that are either all or mostly rural. For other counties, each case is somewhat different. In particular, the one county with a standardized residual near seven appears to be a county with large ACS sampling errors of opposite signs for the numerator and denominator. For the denominator estimate, the 2005 ACS poverty universe is about 20% below the decennial estimate, whereas the PEP intercensal demographic estimates imply strong growth in the county. For the numerator, the 2005 ACS number in poverty is more than 2.5 times the decennial number.



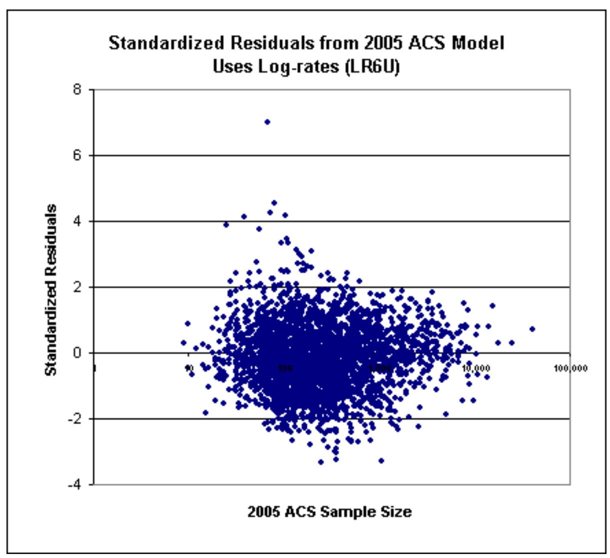


Figure 2.29 (left): Standardized residuals from 2005 ACS log-rate regression against 2005 PEP county total resident population.
Figure 2.30 (right): Standardized residuals from 2005 ACS log-rate regression against 2005 ACS sample size.

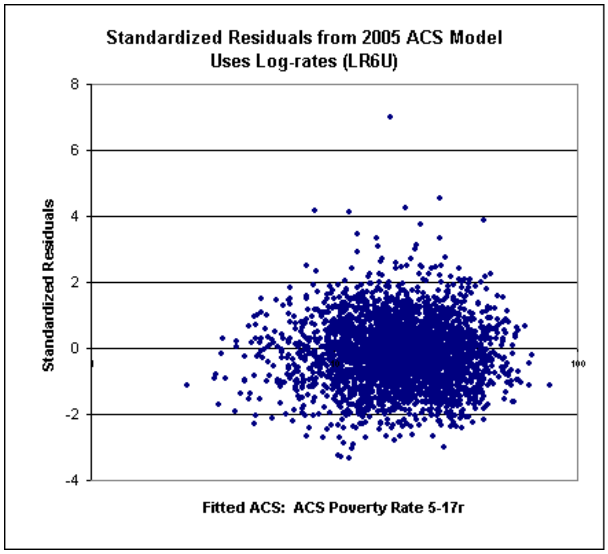
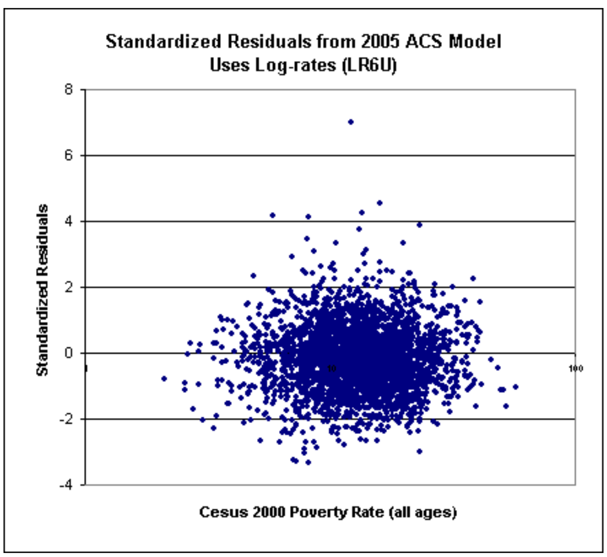


Figure 2.31: Standardized residuals from 2005 ACS log-rate regression against 2005 ACS regression prediction.



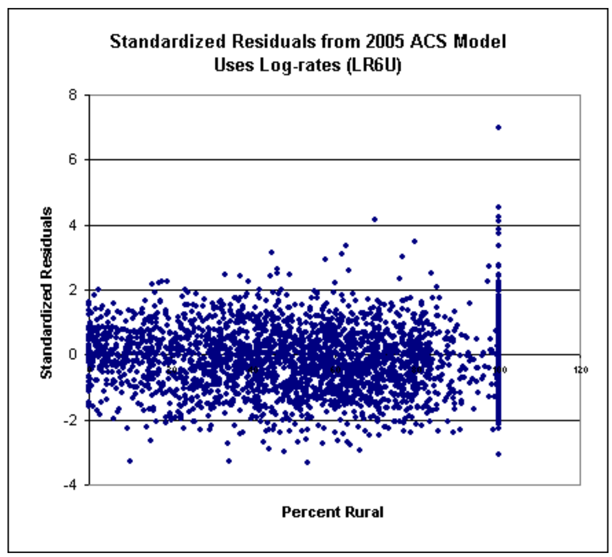


Figure 2.32 (left): Standardized residuals from 2005 ACS log-rate regression against Census 2000 poverty rate, all ages.
Figure 2.33 (right): Standardized residuals from 2005 ACS log-rate regression against Census 2000 percent rural.

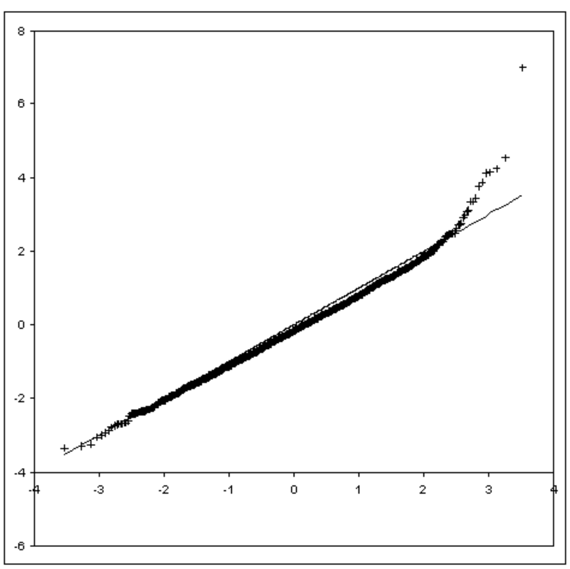
Box-whisker plots and additional scatter plots of the standardized residuals are included in Appendix B.1. These box-whisker plots group the categorization variables into quintiles. Each quintile contains 594 counties of the 2,972 counties from regression prediction. The categorization variables are mainly from Census 2000 and include the following concepts: B.1 Total population, B.2 Percent in poverty, B.3 Population growth 1990-2000, B.4 Population growth 2000-2005, B.5 Percent Black, B.6 Percent Hispanic, B.7 Percent Asian, B.8 Percent group quarters. The Appendix figures show a slight visual upward trend with respect to percent Asian. These indicate very mild model underprediction in counties with a larger share of population Asian. The figures also show some small model overprediction for counties within the lowest quintile of percent population Black. Other classification variables showed no discernible patterns.

Spearman's rank correlation coefficients between the squared standardized residuals and four covariates used in the plots are reported in Table 2.12, along with asymptotic t-statistics. As with Table 2.5, significant correlations would be taken as evidence of possible heteroskedasticity related to the relevant variable. The Spearman's statistics reported in Table 2.12 confirm the general pattern seen in the plots; however, some small, but significant, heterogeneity is detected. It is not an induced problem caused by the reduced list of regressors; the same significant heterogeneity is found for the model with 8 regressors (LR8u).

Table 2.12: Spearman's rank correlation coefficients for squared standardized residuals of model LR6u

Covariates	Spearman's ρ	Asymptotic t-statistic
2005 PEP total county resident population, all ages (log)	-0.043	-2.25
2005 ACS sample size, household count (log)	-0.038	-2.02
Census 2000 poverty rate, all ages	-0.015	-0.80
Census 2000 county percent rural	0.026	1.46

Figures 2.34 and 2.35 describe the distribution of the standardized residuals for the log-level model. Figure 2.34 clearly displays a substantially higher number of large, positive residuals than expected. The positive skewness is even more pronounced than for the log-level model, primarily because of the one standardized residual of about seven. The standard normal distribution predicts approximately four standardized residuals should be higher than three. In this regression, there are twelve such large, positive residuals. Large negative residuals total four, which meets expectations.



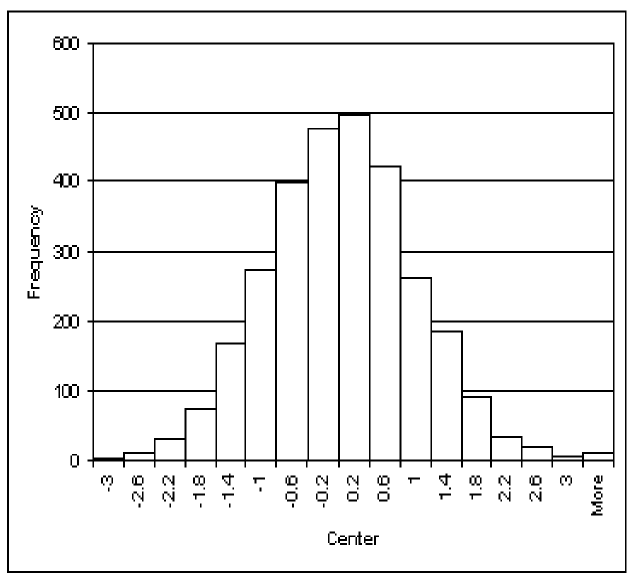


Figure 2.34 (left): Quartiles of standardized residuals for LR6u vs. standard normal quartiles.
Figure 2.35 (right): Frequency distribution of standardized residuals for LR6u.

2.3.4 Analysis of shrinkage estimates

This section compares the shrinkage estimates and their standard errors obtained from the LR6u model with results for the single-year 2005 ACS direct survey estimates.

Precision – Standard errors and CV reduction

The computation of the standard errors of the ACS shrinkage estimates for poverty rates on the original, not logged, scale (the square roots of the prediction error variances) is via equations (6) and (8) of Section 2.1. The standard errors for the direct ACS survey estimates are the square roots of the direct sampling error variance estimates (with limitations for counties with small samples as noted earlier) obtained as discussed in Section 2.6. The CV for the survey estimate is computed as the standard error of the survey estimate divided by the direct survey estimate, and the CV of the shrinkage estimate is computed as the standard error of the shrinkage estimate divided by the shrinkage estimate.

Figure 2.36 shows the frequency distribution of the ratios of the standard errors of the ACS shrinkage estimates to the standard errors of the ACS direct survey estimates. Much of the density appears roughly between ratios of 0.2 and 0.8.

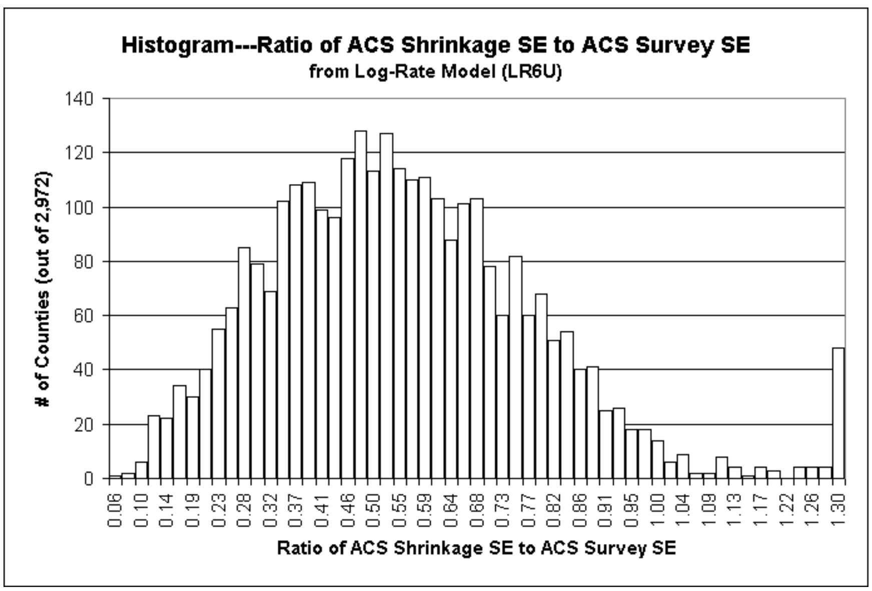
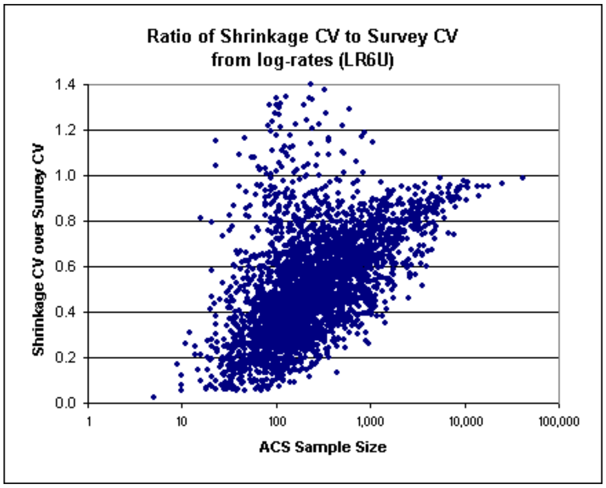


Figure 2.36: Frequency distribution of the ratios of the standard errors of the 2005 ACS shrinkage estimates to the standard errors of the 2005 ACS direct survey estimates.
The numbers shown on the x-axis are the lower bound cutoff for each bar.

Figure 2.37 shows the ratios of the standard errors of the ACS shrinkage estimates to the standard errors of the ACS direct survey estimates against population. The median ratio is 0.52, i.e., 52 percent, which corresponds to the model producing a 48 percent decline in the CV of the ACS county poverty survey estimates. Figure 2.38 shows the ratios of the CVs of the ACS shrinkage estimates to the CVs of the ACS direct survey estimates against 2005 ACS sample size. The median ratio is 0.43, i.e., 43 percent, which corresponds to the model producing a 57 percent decline in the CV of the ACS county poverty survey estimates. The direct estimate standard errors and CVs are more reduced by modeling for counties with smaller ACS sample sizes than counties with larger ACS sample sizes. Counties with large sample size have relatively precise direct survey estimates that receive a high weight in the shrinkage estimate. In Figure 2.37, the ratios of standard error greater than one occur for counties in which the shrinkage estimate is much larger than the direct survey estimate. The standard errors for the shrinkage estimates are accordingly higher. For these counties, despite reduction in the CV from modeling, the standard errors are higher from modeling due to the higher shrinkage estimates.



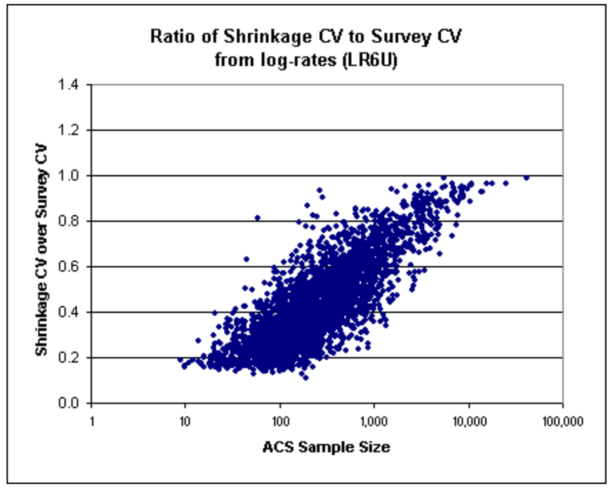


Figure 2.37 (left): Ratios of standard errors of the 2005 ACS shrinkage estimates to the standard errors of the 2005 ACS direct survey estimates against 2005 ACS sample size.
Figure 2.38 (right): Ratios of CV of the 2005 ACS shrinkage estimates to the CV of the 2005 ACS direct survey estimates against 2005 ACS sample size.

Figure 2.39 is a scatter plot of the CVs of the ACS shrinkage estimate against the CVs of the ACS direct survey estimates. The CVs of the ACS shrinkage estimates are uniformly below 19 percent, while the CVs of the ACS direct survey estimates range up to 160 percent. We see a clear improvement from modeling for the majority of counties (which typically have small to medium size ACS samples).

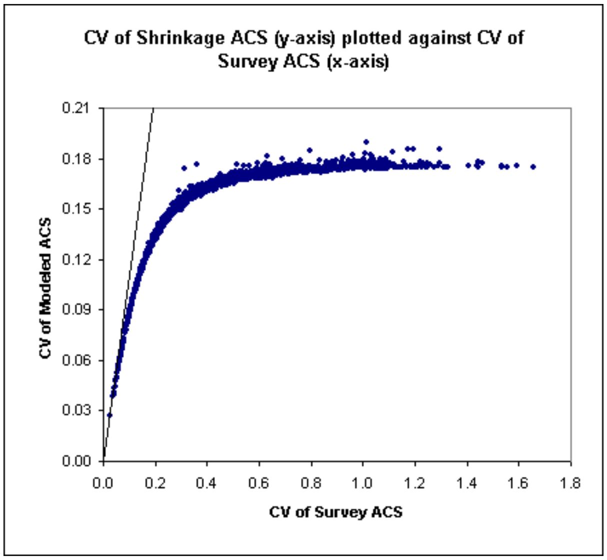


Figure 2.39: CVs of the 2005 ACS shrinkage estimates against the CVs of the 2005 ACS direct survey estimates.

Weights on direct survey estimates

Figure 2.40 shows the frequency distribution of the weights given to the direct survey estimates by the shrinkage estimates. These are the quantities $1 - w_i$ where w_i is given in conjunction with equation (5) of Section 2.1. Most counties have weights on their direct survey estimates of less than 16 percent, the median weight being 0.158. Counties with direct estimates of zero for number 5 - 17 in poverty are not included in the histogram.

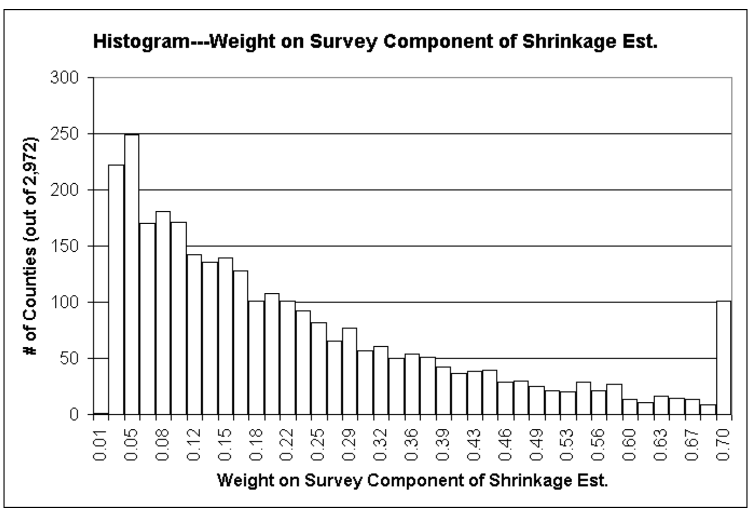


Figure 2.40: Frequency distribution of the weight on the survey component of the shrinkage estimate.
The numbers shown on the x-axis are the lower bound cutoff for each bar. The rightmost bar contains all point greater than 0.70.

Figure 2.41 plots the weights given to the direct survey estimates by the shrinkage estimates against the ACS sample sizes. Many counties with sample sizes larger than 1,000 have weights on the direct survey estimates higher than 60 percent.

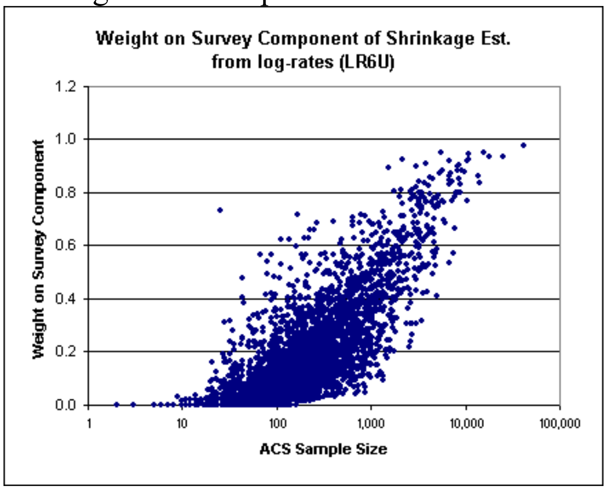


Figure 2.41: Frequency distribution of the weight on the survey component of the shrinkage estimates against 2005 ACS sample size.

Poverty ratio estimates

Figure 2.42 shows the frequency distribution of the ratios of the shrinkage estimates to the ACS direct survey estimates. Much of the density appears roughly between ratios of 0.8 and 1.5.

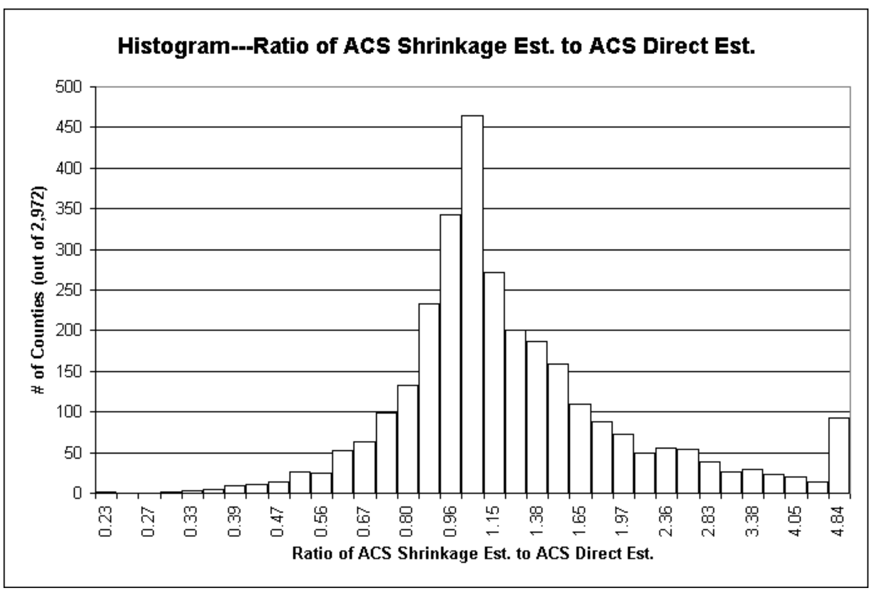
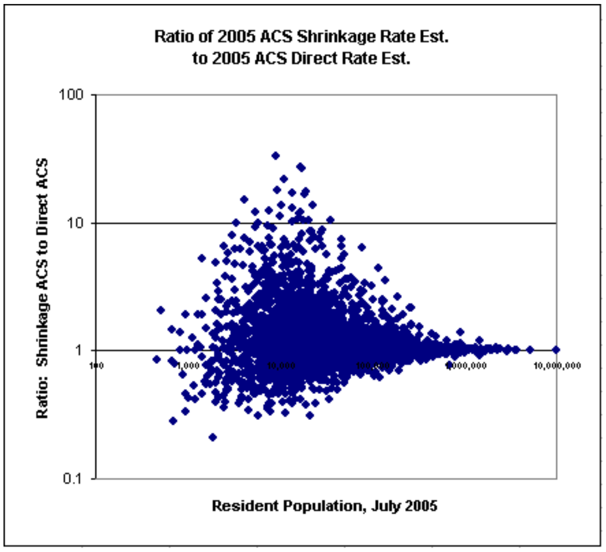


Figure 2.42: Frequency distribution of the ratios of 2005 ACS shrinkage estimates to 2005 ACS survey estimates.
The numbers shown on the x-axis are the lower bound cutoff for each bar. The rightmost bar contains all points greater than 4.84.

Figure 2.43 plots the ratios of the shrinkage estimates to the 2005 ACS direct survey estimates of poverty rate against 2005 PEP county total resident population. The ratios range from about 0.2 to about 33, signifying that the model-based estimates are at times from one-fifth to thirty-three times as large as the ACS direct survey estimates. Figure 2.44 plots the ratios of the Census 2000 poverty rates, ages 5 - 17 related to the 2005 ACS direct survey estimates of poverty rate, ages 5-17 related against the 2005 PEP county total resident population. The ratios vary similarly from about 0.1 to about 35. Figures 2.43 and 2.44 show the expected variation of the direct survey estimates for counties with low sample size. Similar plots of the survey estimates against Food Stamp rate and child tax-poor rate (not shown) visually have similar dispersion.



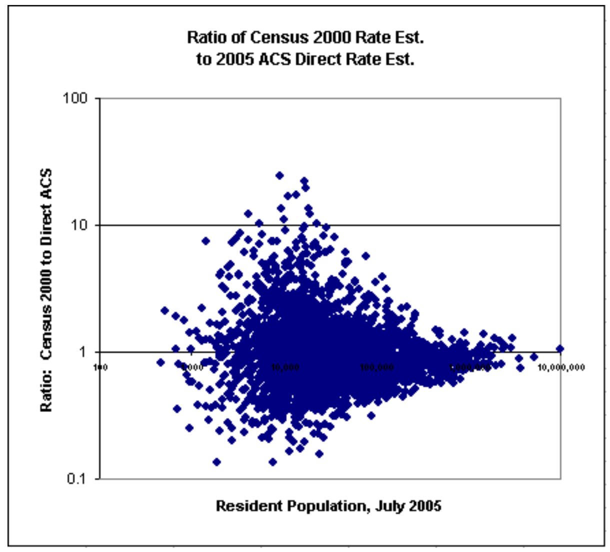
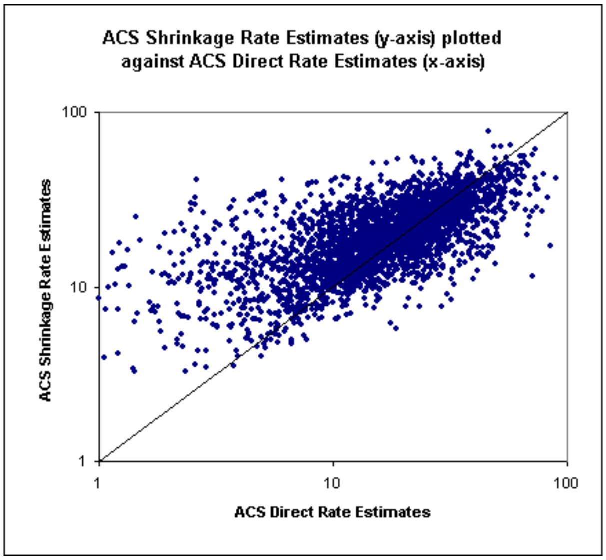


Figure 2.43 (left): Ratios of the 2005 ACS shrinkage estimates, ages 5-17 related to the 2005 ACS direct survey estimates, ages 5-17 related.

Figure 2.44 (right): Ratios of the Census 2000 poverty rates, ages 5-17 related to the 2005 ACS direct survey estimates of poverty rate, ages 5 - 17 related.

Figure 2.45 plots the shrinkage estimates against the 2005 ACS direct survey estimates. There is greater dispersion for counties with small shrinkage estimates or small direct survey estimates. Also, there appears to be some skewness at the bottom left tail, suggesting that some counties with low ACS direct survey estimates have much higher ACS shrinkage estimates. This observation is consistent with the observation of several counties in Figure 2.43 with ratios above 10.0. Figure 2.46 plots the shrinkage estimates against the corresponding regression predictions. The shrinkage estimates generally appear fairly close to the regression predictions. This is the case even for counties with large shrinkage estimates or regression predictions, which presumably have high weights on the ACS direct survey estimates.



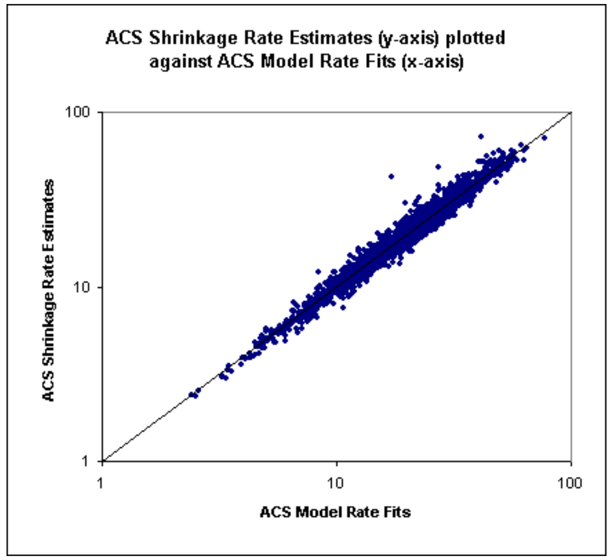


Figure 2.45: 2005 ACS rate shrinkage estimates against 2005 ACS rate survey estimates.
Figure 2.46: 2005 ACS rate shrinkage estimates against 2005 ACS rate regression predictions (or model fits).

2.3.5 Testing alternate restrictions and regressor sets

Tables 2.13 and 2.14 repeat the tests discussed in Section 2.2.5, but done here for the log-rate model. The same derivation of the equivalence of log-level versus log-rate as regressors that motivated the 8-regressor models for log-level can be applied to models with the log-rate dependent variable. The 8-variable model with the log-rate dependent variable (LR8 in Table 2.13) is similar in concept to the "hybrid models" examined in the second NAS panel interim report (NAS 2000, page 101).

The scale invariance restriction for the log-rate model amounts to having the coefficients on all the log denominator terms sum to zero. For the LR6 model this reduces to simply having the coefficient on log(PEP population, ages 0-17) be zero (as shown in the table for LR6r).

The LR6r versusLR6u and LR8r versus LR8u statistics test the imposition of the scale invariance restriction, namely, the sum of the denominator term coefficients equals zero. (See discussion in Section 2.2.5.) The LR6r versus LR8r and LR6u versus LR8u tests examine the omission of the two additional scale or denominator terms listed in Table 2.13. The motivation for these additional variables was discussed in Section 2.2.5. The final test in Table 2.14 tests the joint hypothesis.

Every null hypothesis in this table is rejected, except for a borderline result when comparing the unrestricted, six-variable model to the unrestricted, eight-variable model. The ordering of these results is nearly coincident with the AIC ordering, which prefers the two unrestricted models (LR6u and LR8u) to the two restricted models (LR6r and LR8r).

Table 2.13: Regression prediction results for four versions of the log-rate model
These results use the 2005 ACS as the dependent variable with income year 2004 regressors and report: coefficient (standard error).

Variable		Model			
		LR6r	LR6u	LR8r	LR8u
Intercept	β_0	0.235 (0.031)	0.372 (0.048)	0.154 (0.094)	0.242 (0.102)
Log (IRS child tax-poor exemption rate)	β_1	0.624 (0.046)	0.556 (0.049)	0.623 (0.047)	0.576 (0.052)
Log (Food Stamp rate)	β_2	0.166 (0.022)	0.167 (0.022)	0.163 (0.022)	0.167 (0.022)
Log (PEP population, ages 0-17)	β_3	0	-0.020 (0.005)	-0.321 (0.086)	-0.222 (0.098)
Log (IRS child filing rate)	β_4	-0.255 (0.109)	-0.414 (0.117)	-0.507 (0.128)	-0.521 (0.128)
Log (Census 2000 poverty rate, ages 5-17 related)	β_5	0.272 (0.035)	0.289 (0.035)	0.253 (0.035)	0.273 (0.036)
Log (PEP population, all ages)	β_6			0.090 (0.061)	0.079 (0.061)
Log (Census 2000 poverty universe, ages 5-17 related)	β_7			0.232 (0.083)	0.130 (0.097)
Degrees of freedom		2,966	2,966	2,964	2,964
Model error variance		0.0319	0.0303	0.0306	0.0300
AIC		3866.4	3854.8	3856.5	3854.2
Maximum log likelihood		-1927.2	-1920.4	-1920.3	-1918.1
Sum of denominators		0.0000	-0.0201	0.0000	-0.0135

Table 2.14: Likelihood ratio tests for the log-rate models

Models	H0	Likelihood Ratio	Degrees of Freedom	Critical Value (α=0.10)	Critical Value (α=0.05)
LR6r vs. LR6u	sum of denominators = 0	13.62	1	2.71	3.84
LR8r vs. LR8u	sum of denominators = 0	4.32	1	2.71	3.84
LR6r vs. LR8r	$\beta_6 = \beta_7 = 0$	13.88	2	4.61	5.99
LR6u vs. LR8u	$\beta_6 = \beta_7 = 0$	4.58	2	4.61	5.99
LR6r vs. LR8u	joint both above	18.20	3	6.25	7.82

2.4 COMPARISONS OF REGRESSION PREDICTIONS FROM LOG-LEVEL AND LOG-RATE MODELS

The hypothesis test results and AIC comparisons of Section 2.3.5 showed a preference for the unrestricted log-rate models over log-rate models with scale invariance restrictions. Since unrestricted log-rate models have a similar structure to log-level models (note related discussion in Section 2.1), these results show some preference for log-level over log-rate models for the 2005 ACS county 5 - 17 poverty estimates. More direct statistical inferences about comparisons of log-level and log-rate models are not immediate since the two types of models involve different dependent variables: for log-level models the dependent variable is log (y_i),[22] whereas for log-rate models it is log (y_i) − log (z_i), where z_i is the 2005 ACS direct survey estimate of the 5 - 17 poverty universe for county *i*.

[22] Recall that y_i is the ACS direct survey estimate of the number of related 5 - 17 in poverty in county *i*.

One approach to comparing log-level and log-rate models would be to construct a simultaneous (bivariate) model of log (y_i) and log (z_i) under which the LL6u and LR6u models (or LL8u and LR8u models) would be nested. As developing a model for the log poverty universe estimates (log (z_i)) would be somewhat extraneous to the present focus, we compare model predictions from log-level and log-rate models in two ways: by converting log-rate regression predictions to corresponding predictions of number in poverty and by converting log-level regression predictions to corresponding predictions of log-rate. Section 2.4.1 makes some "goodness of fit" comparisons in terms of root mean squared error (RMSE) statistics for prediction of the dependent variables (ACS direct estimates) from the models. Section 2.4.2 simply compares the prediction results from the two types of models to illustrate differences rather than to assess goodness of fit.

2.4.1 Goodness of fit comparisons

From the test results in Sections 2.2.5 and 2.3.5 above, the log-level and log-rate models should be comparable for either the six- or eight-variable, unrestricted versions. The six-variable version is presented below, as the log-level version corresponds to the current SAIPE production model.

The statistic used here to compare goodness of fit of the two models is root mean squared error (RMSE) in the log scale for all counties with nonzero 2005 ACS poverty estimates, ages 5 - 17 related (so the logs can be taken of the direct ACS estimates). For the log-level models, the errors are just the regression residuals from the fitted models. For the log-rate model, a transformation must first be calculated. Analogous to the production methodology, regression predictions are transformed from the log rate model by adding a demographic population estimate of the log (poverty universe, ages 5-17 related), labeled log (z demog) in Table 2.15.

In general, the same results will not be obtained from comparing log-level model predictions with transformed log-rate model predictions as would be obtained by performing the opposite transformation and comparing on the log-rate scale. Both comparisons are reported in Table 2.15, but since estimates of poverty level is the eventual goal of SAIPE production, the discussion focuses more on the comparisons in the log-level scale.

Table 2.15 displays RMSEs for the two models. This is equivalent to the square root of the unweighted residual mean square for each model. RMSEs are reported for the entire set of observations, as well as for two partitions of counties – one by population size and one by Census 2000 poverty rates. The bounds for the population partition were chosen based on the three levels of ACS reporting. As full implementation of the ACS proceeds, counties with over a 65,000 population will have single-year survey estimates published, those with a population of 20,000 - 65,000 will have 3-year average estimates published, and those under 20,000 will have 5-year estimates. The bounds for the poverty rate partition were chosen to approximate equal-sized partitions.

Comparisons should primarily be made horizontally for this table, comparing results for log-level versus log-rate for a given partition of counties. Different subsets of counties will have different mean values, and thus the expected RMSE is different. Even if the difference in means is adjusted, sampling error variances will be much larger for smaller counties in general, and thus larger RMSEs are expected.

The comparison between log-level model predictions and log-rate model predictions results in a lower RMSE for the log-level predictions in every case of Table 2.15. On either scale, for any partition of the data, the log-level predictors result in lower RMSEs. It is not known if any of these differences are significant, since constructing statistics with known distributions between two dependent series with different variances is not straightforward. Nonetheless, the uniformity of the results does appear notable.

Table 2.15: RMSE comparisons for the LL6u andLR6u models

	Comparison on the log-level scale		Comparison on the log-rate scale	
	LL6u	LR6u + log (z demog)	LL6u − log (z demog)	LR6u
Full-sample: 2,972 counties	0.638	0.650	0.597	0.609
Partition by 2005 PEP county total resident population				
> 65k: 775 counties	0.321	0.326	0.317	0.323
20k – 65k: 1,032 counties	0.547	0.556	0.531	0.540
< 20k: 1,165 counties	0.839	0.857	0.770	0.786
Partition by Census 2000 poverty rate, all ages				
> 20%: 936 counties	0.617	0.634	0.547	0.562
12.5 – 20%: 1,000 counties	0.605	0.618	0.561	0.573
< 12.5: 1,036 counties	0.685	0.695	0.670	0.680

2.4.2 Comparisons of regression prediction results

This section compares the county-level regression predictions, shrinkage weights, shrinkage estimates, standard errors, and CVs between the log-level and log-rate models.

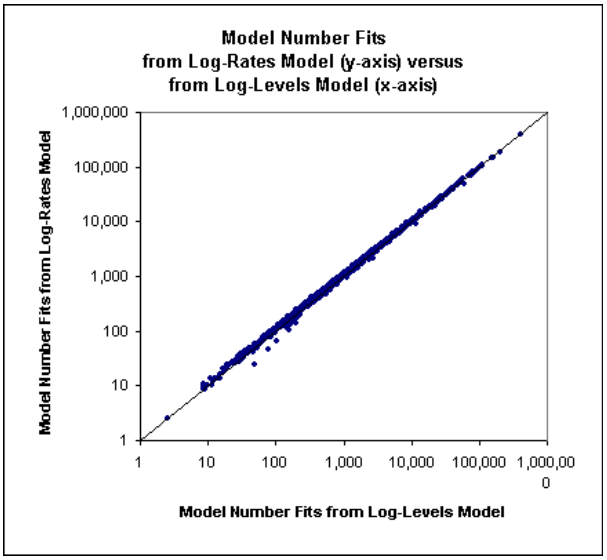
For comparisons, the log-level and log-rate models need to be transformed into a common scale. The transformations apply to the regression predictions and shrinkage estimates, but are slightly different for the log-rate and log-level models. The following terms refer to the transformed variables: “number regression prediction,” “rate regression prediction,” “number shrinkage estimates,” and “rate shrinkage estimates.”

From the log-level model, exponentiating and then adjusting for the log-bias in the regression model (using the model error variance in the adjustment) transforms the log-level regression predictions to predictions of numbers in poverty. The predictions for poverty rates from the log-level model use these results divided by the demographic poverty universe estimates.

For the log-rate model, exponentiating and then adjusting for the log-bias in the regression model (using the model error variance in the adjustment) transforms the log-rate regression predictions to predictions of poverty rates. The predictions of numbers in poverty from the log-rate model use these same results multiplied by the demographic poverty universe estimates.

Comparisons of regression predictions from fitted models

Figures 2.47 and 2.48 are scatter plots of log-rate model regression predictions versus log-level model regression predictions. These figures show a small number of points noticeably below the diagonal (with lower number regression prediction from the log-rate model than from the log-level model). It appears these counties are not large counties (as per Figure 2.47) and also that these counties are not low-poverty-rate counties (as per Figure 2.48).



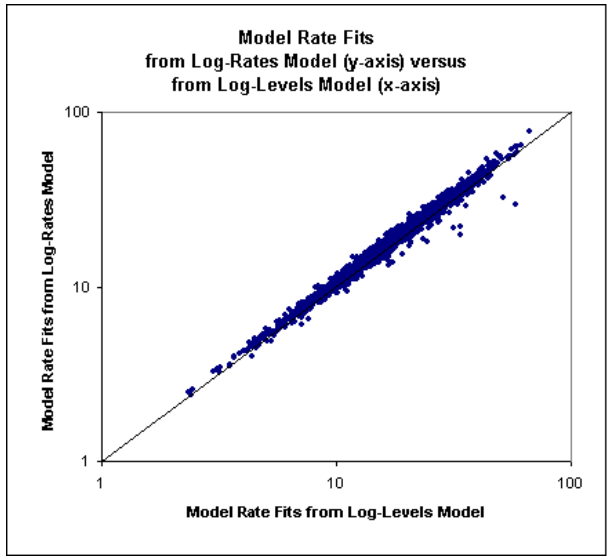
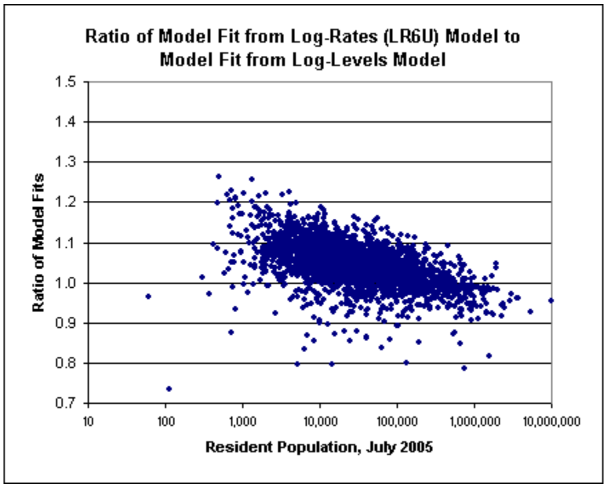


Figure 2.47 (left): LR6u against LL6u number regression predictions.
Figure 2.48 (right): LR6u against LL6u rate regression predictions.

Figures 2.49 and 2.50 show ratios of predictions from the log-rate model to predictions from the log-level model, plotted against resident population. In these figures there appears to be a trend in which the regression predictions from the log-rate model exceed those from the log-level model for small counties, with the reverse holding for large counties. For example, in Figure 2.49 there are many counties with population size near 1,000 whose log-level number regression predictions exceed the log-rate number regression predictions by a multiplicative 10 percent. Similarly, in Figure 2.50 there are many counties of population size near 1,000 whose log-level rate regression predictions exceed the log-rate rate regression predictions by 2.0 percentage points.



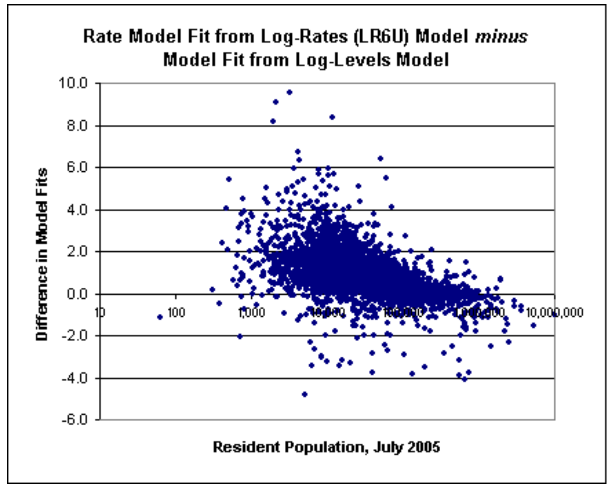


Figure 2.49 (left): Ratios of number regression predictions from theLR6u model to number regression predictions from the LL6u model against 2005 PEP county total resident population.

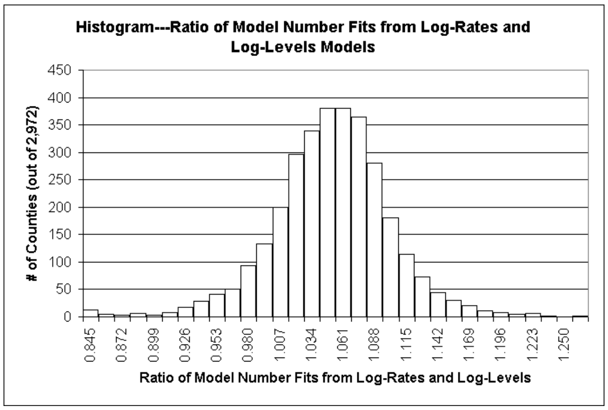
Figure 2.50 (right): Differences between rate regression predictions from theLR6u model and rate regression predictions from the LL6u model against 2005 PEP county total resident population.

The trends in the log-rate and log-level model predictions against population size are further illustrated in Table 2.16. The left-hand column of Table 2.16 refers to the ratio of number regression prediction from the log-rate models to the number regression prediction of the log-level models. The median (mean) ratio for the smallest 628 counties is 1.078 (1.075), indicating a number regression prediction that is 7.8 percent (7.5 percent) higher from the log-rate model than the log-level model. The median ratio for the largest 628 counties is 1.005 (1.003). The right-hand column of Table 2.16 refers to the difference between rates regression prediction from the log-rate models to the rates regression prediction of the log-level models. The median (mean) difference for the smallest 628 counties is 1.39 percentage points (1.48 percentage points), and the median difference for the largest 628 counties is 0.06 (−0.02).

Table 2.16: Comparing log-rate and log-level regression predictions by 2005 PEP county total resident population quintiles

	Ratio of number regression predictions: Log-rate over log-level	Difference in number regression predictions: Log-rate minus log-level
	Median, by Population Quintiles	
1st (lowest) population quintile	1.078	1.39
2nd population quintile	1.064	1.30
3rd population quintile	1.050	0.95
4th population quintile	1.034	0.54
5th (highest) population quintile	1.005	0.06
	Mean, by Population Quintiles	
1st (lowest) population quintile	1.075	1.48
2nd population quintile	1.060	1.39
3rd population quintile	1.047	0.98
4th population quintile	1.033	0.61
5th (highest) population quintile	1.003	-0.02

Histograms corresponding to Figures 2.49 and 2.50 above follow below in Figures 2.51 and 2.52, which show the frequency distribution of the ratio of number regression predictions and the difference between rate regression predictions from the log-rate and log-level models. Most of the density in Figure 2.51 is between 1.007 and 1.102, and most of the density in Figure 2.52 is between 0.5 and 1.9. These figures indicate a majority of counties have larger number regression predictions under the log-rate model than under the log-level model. This aligns with the observation that larger counties tend to have smaller number regression predictions under the log-rate model. While there are comparatively few counties with smaller number regression predictions under the log-rate model, since they tend to be larger counties, they contribute disproportionately to the estimate of number of children in poverty, ages 5-17 related.[23]



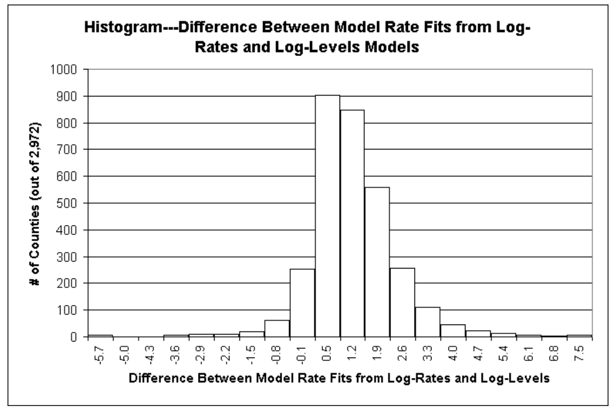


Figure 2.51 (left): Frequency distribution of the ratios of number regression predictions from the log-rate and log-level models.
Figure 2.52 (right): Frequency distribution of the differences between rate regression predictions from the log-rate and log-level models.
In both figures, the numbers shown on the x-axis are the lower bound cutoff for each bar.

Comparisons of shrinkage estimates

Table 2.17 presents summary statistics on the weights given to the direct ACS estimates in the shrinkage estimation for the log-level and log-rate models. As these are determined by the model error variance, which is constant for all counties, and the sampling error variances, summary statistics on the latter are also presented. Statistics do not include counties with 5 - 17 poverty of zero in the ACS sample.

We see from the table that sampling error variances tend to be lower for the log-rate models than for the log-level models. The reason for this can be seen as follows. The sampling error variance for the log-rate can be written as:

$$\mathrm{Var}(\log(y_i) - \log(z_i)) = \mathrm{Var}(\log(y_i)) + \mathrm{Var}(\log(z_i)) - 2\times\mathrm{Corr}(\log(y_i),\log(z_i))[\mathrm{Var}(\log(y_i))\ \mathrm{Var}(\log(z_i)]^{.5}.$$

[23] The number regression predictions examined are unraked, and the sum of the 2,972 counties considered is 0.5 percent lower for log-rate than for log-level. Despite the higher number of counties with larger rate regression predictions than number regression predictions, the total number in poverty across counties is slightly lower for the rate regression predictions.

Given that both y_i and z_i are strongly related to population size, the correlation is expected to be positive, and it can easily be large enough for the third term to offset the second, leading to a lower sampling error variance for the log-rate. Coupled with the higher model error variance for the log-rate model, the weights given to the direct survey estimates (note equation (5) in Section 2.1) tend to be higher than those for the log-level model.

Table 2.17: Summary statistics for sampling error variances and resulting shrinkage weights on the ACS direct survey estimates

	Median		Mean		Standard Deviation	
	log-level	log-rate	log-level	log-rate	log-level	log-rate
Sampling error variance	0.186	0.159	0.359	0.325	0.448	0.423
Model error variance	0.022	0.030	0.022	0.030	0.000	0.000
Weight on county ACS direct survey estimates	0.106	0.158	0.167	0.216	0.171	0.189

Figure 2.53 plots the weights on the direct survey estimates from the two models against each other, showing in detail that these weights are higher for the log-rate model with only a few exceptions of any substance. There is more dispersion for counties with smaller weights on the survey component, but the line is most bowed out at around 50 percent weight.

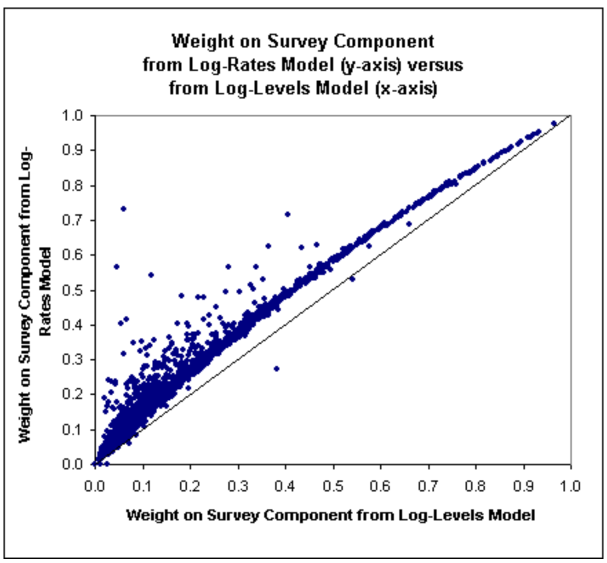
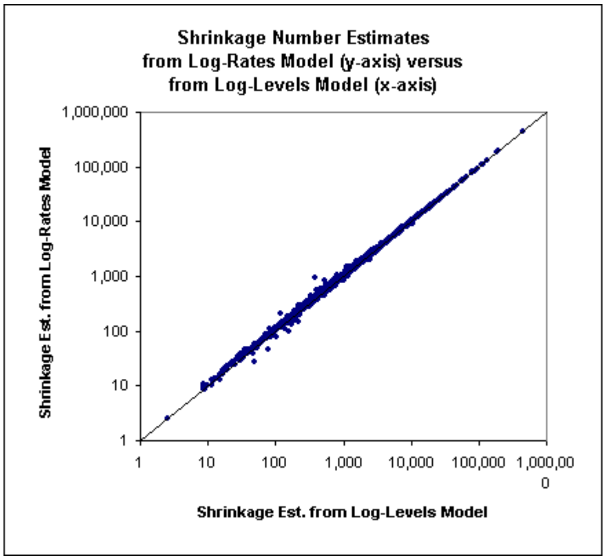


Figure 2.53: Weights on survey component from log-rate model against weights on survey component from log-level model.

Figures 2.54 and 2.55 plot shrinkage estimates from the log-level and log-rate models against each other; Figure 2.54 showing results for estimates of number in poverty, and Figure 2.55 showing results for estimates of poverty rates. The points in Figures 2.54 and 2.55 appear slightly less tight along the diagonal cluster than we saw for the regression predictions in Figures 2.47 and 2.48, which is due to the effect of the direct survey estimates on the shrinkage estimates. There are now two counties above the diagonal, and these are neither large counties nor low-poverty-rate counties. In both figures, counties with zero ACS poverty are also included. For these counties, the shrinkage estimates are simply the regression predictions from the fitted models.



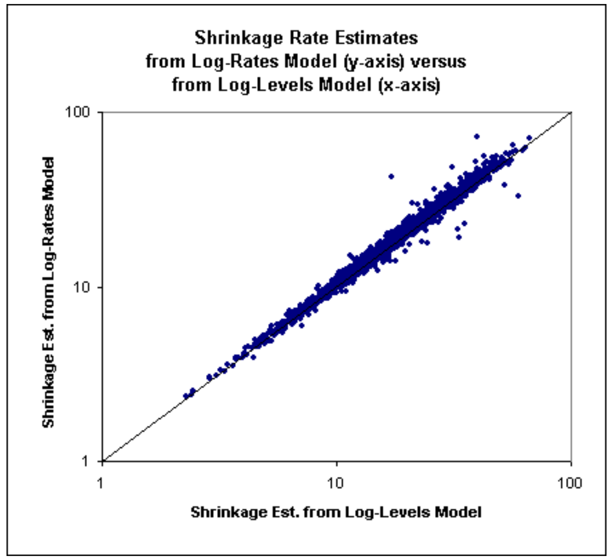
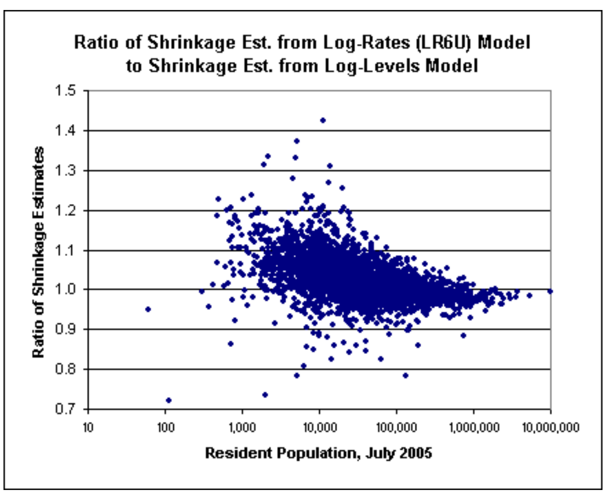


Figure 2.54 (left): Number shrinkage estimates from the LR6u model against shrinkage number fits from the LL6u model.

Figure 2.55 (right): Number shrinkage estimates from the LR6u model against shrinkage number fits from the LL6u model.

To investigate these relations further, Figures 2.56 and 2.57 plot the ratios of shrinkage estimates from the log-rate to the log-level model for estimates of number in poverty (Figure 2.56) and of poverty rates (Figure 2.57). From these figures it appears that much of the downward-sloping trends from Figures 2.49 and 2.50 remain after combining the regression predictions with the ACS direct survey estimates.



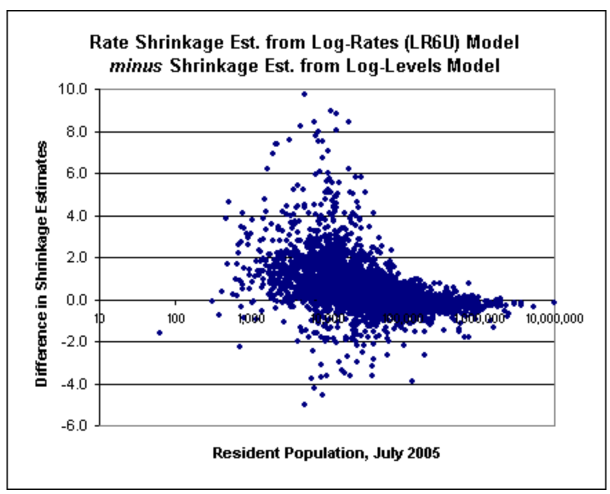
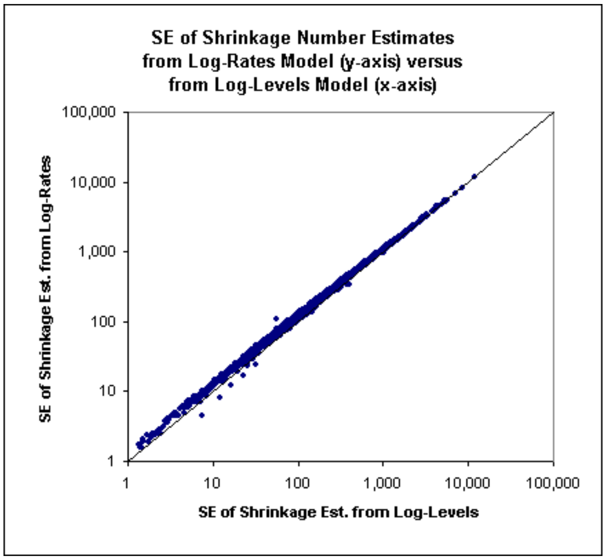


Figure 2.56 (left): Ratios of number shrinkage estimates from the LR6u model to number shrinkage estimates from theLL6u model against 2005 PEP county total resident population.

Figure 2.57 (right): Differences between rate shrinkage estimates from the LR6u model and rate shrinkage estimates from the LL6u model against 2005 PEP county total resident population.

Figure 2.58 plots the standard errors of the number shrinkage estimates from the LR6u log-rate model against the standard errors of the number shrinkage estimates from the LL6u log-level model. Figure 2.59 plots the standard errors of the rate shrinkage estimates from the LR6u log-rate model against the standard errors of the rate shrinkage estimates from the LL6u log-level model. It appears that the standard errors of the shrinkage estimates from the log-rate model are systematically higher than the standard errors of the shrinkage estimates from the log-level model, both for numbers and rates. For rates, this is especially true for high-standard-error counties.



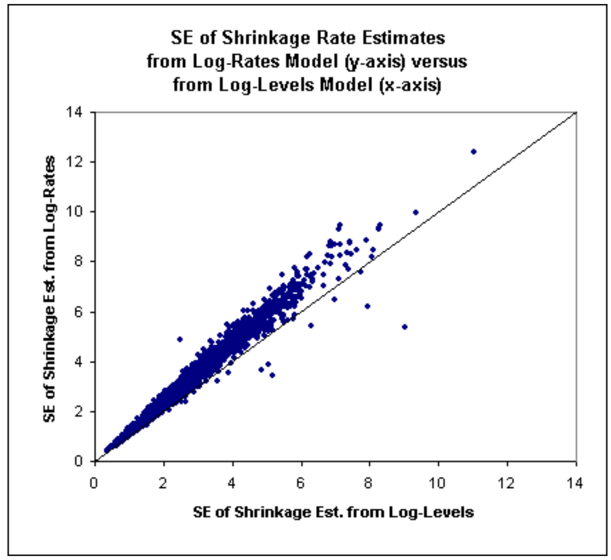


Figure 2.58 (left): Standard errors of number shrinkage estimates from the LR6U model against standard errors of number shrinkage estimates from the LL6u model.
Figure 2.59 (right): Standard errors of rate shrinkage estimates from the LR6U model against standard errors of rate shrinkage estimates from the LL6u model.

Figure 2.60 plots the CVs of the number shrinkage estimates from the LR6u log-rate model against the CVs of the number shrinkage estimates from the LL6u log-level model. (The CVs of the rate shrinkage estimates are identical to those of the number shrinkage estimates since the PEP intercensal demographic estimates in the denominators of each rate cancel out.) It appears that the CVs of the number shrinkage estimates from the log-rate model are systematically higher than those from the log-level model, especially for high-CV counties (likely the smaller counties).

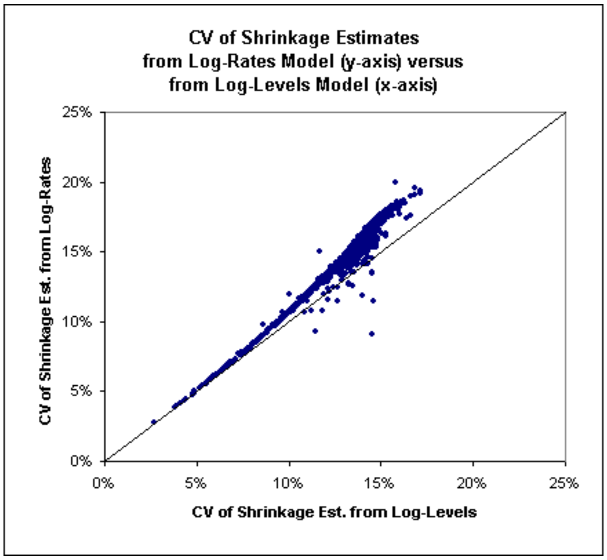
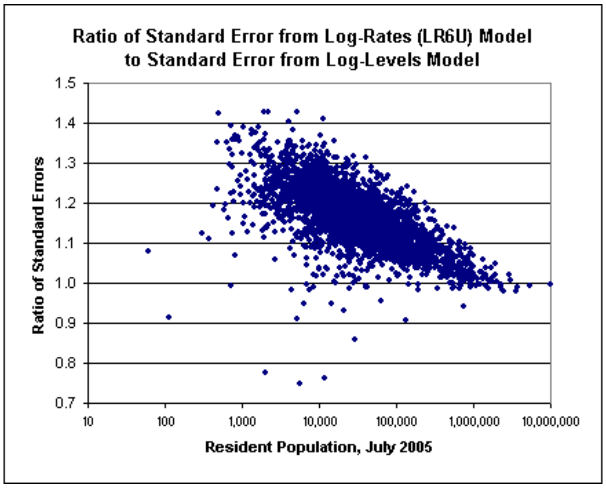


Figure 2.60: CVs of number shrinkage estimates from the LR6U model against CVs of number shrinkage estimates from the LL6U model.

Figure 2.61 plots the ratios of the standard errors of the number shrinkage estimates from the LR6U model to the standard errors of the number shrinkage estimates from the LL6U model against 2005 PEP county total resident population. For nearly all counties, the ratio is larger than 1.0, indicating higher standard errors from the log-rate model. Also, it appears that larger counties have lower ratios, indicating closer standard errors from the two models than the standard errors of smaller counties.

Figure 2.62 plots the ratios of the CVs of the shrinkage estimates from the LR6u model to the CVs of the shrinkage estimates from the LL6u model against resident population. The ratio of the CVs is greater than 1.0 for nearly all counties, and the ratio tends to be larger for the smaller counties, with the ratio nearest to 1.0 for the largest counties. A small set of counties with population between 5,000 and 20,000 have ratios of CVs less than 1.0.



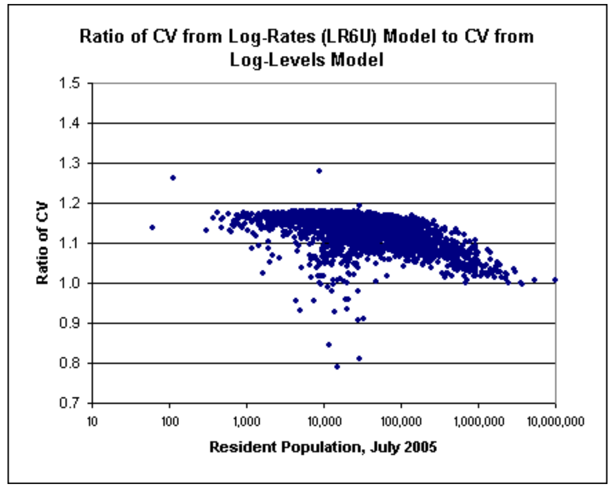


Figure 2.61 (left): Ratios of standard errors of shrinkage estimates from the LR6U) model to standard errors of shrinkage estimates from the LL6U model against 2005 PEP county total resident population.
Figure 2.62 (right): Ratios of CVs of shrinkage estimates from theLR6U model to CVs of shrinkage estimates from theLL6U model against 2005 PEP county total resident population.

2.5 TIME REFERENCE OF REGRESSORS

As discussed in Section 1.4.3, the 2005 ACS uses income reports spanning twenty-three months, from January 2004 through November 2005. The income reference is centered on December 15, 2004, and the population control is to July 2005. The tax data refer to calendar year 2004 and the majority were filed between January and April of 2005. The food stamps data reflect program participation as of July 2004, and the Census 2000 data refer to income in 1999 and were collected in April of 2000. Given these various timeframes, the appropriate reference year (2004 or 2005) for the regressor variables is not immediately apparent.

This section compares results from the 2005 ACS LL6u log-level model using two alternatives for the regressor variables – 2004 regressors (the base case used throughout most of this report) and 2005 regressors. Regression prediction results with these two sets of regressors are quite similar. While there are some differences of substance in the regression predictions for particular counties, much of this is attributable to substantial changes in the values of the regressors for particular counties rather than differences in the estimated model coefficients.

Regression prediction results with 2004 and 2005 regressors

Tables 2.18 and 2.19 show regression prediction results for the 2005 ACS county log-level poverty model using 2004 regressors and using 2005 regressors, respectively.[24] The results of the two regression predictions are very similar. Tables 2.18 and 2.19 show that most of the coefficient estimates are very close, and the R-squared measure is nearly the same, at 0.936 with 2004 regressors and 0.935 with 2005 regressors. Finally, the estimated model error variances are

[24] For the 2005 regressors, imputations were made for the two counties most affected by Hurricane Katrina (Orleans Parish, LA and St. Bernard Parish, LA). This required imputation for three of the 2005 regressors that measure post-Katrina conditions: Tax-year 2005 poor child exemptions, Tax-year 2005 total child exemptions, and July 2006 population estimates.

nearly identical at 0.022.

Table 2.18: Regression output for 2005 ACS model using 2004 regressors
These results are based on direct ACS sampling error variances of dependent variable Log (2005 ACS poor, ages 5-17 related).

Variable	Regression Coefficient	Standard Error	t	Pr>|t|
Intercept	-0.42	0.06	-7.39	<0.0001
Log (IRS child tax-poor exemptions)	0.55	0.04	12.26	<0.0001
Log (Food Stamps)	0.17	0.02	7.90	<0.0001
Log (PEP population, ages 0-17)	1.05	0.12	8.62	<0.0001
Log (IRS child tax exemptions)	-1.04	0.11	-9.13	<0.0001
Log (Census 2000 poor, ages 5-17 related)	0.27	0.03	8.90	<0.0001
Degrees of freedom = 2966				
Model error variance = 0.022				
R-squared = 0.936				

Table 2.19: Regression output for 2005 ACS model using 2005 regressors
These results are based on direct ACS sampling error variances of dependent variable Log (2005 ACS poor, ages 5-17 related).

Variable	Regression Coefficient	Standard Error	t	Pr>|t|
Intercept	-0.40	0.06	-6.90	<0.0001
Log (IRS child tax-poor exemptions)	0.54	0.04	12.14	<0.0001
Log (Food Stamps)	0.16	0.02	7.02	<0.0001
Log (PEP population, ages 0-17)	0.98	0.12	7.88	<0.0001
Log (IRS child tax exemptions)	-0.97	0.11	-8.47	<0.0001
Log (Census 2000 poor, ages 5-17 related)	0.30	0.03	10.31	<0.0001
Degrees of freedom = 2966				
Model error variance = 0.022				
R-squared = 0.935				

Comparing residuals and regression predictions for the two models

Figure 2.63 plots the 2005 ACS standardized residuals using 2005 regressors on the y-axis against the 2005 ACS standardized residuals using 2004 regressors on the x-axis. It appears that nearly all large (in magnitude) residuals remain large and nearly all small (in magnitude) residuals remain small upon moving from the 2004 regressors to the 2005 regressors. The residuals that are large in magnitude may be primarily the result of extreme ACS direct survey estimates (which do not change upon switching from 2004 to 2005 regressors) as opposed to extreme regressor input data.

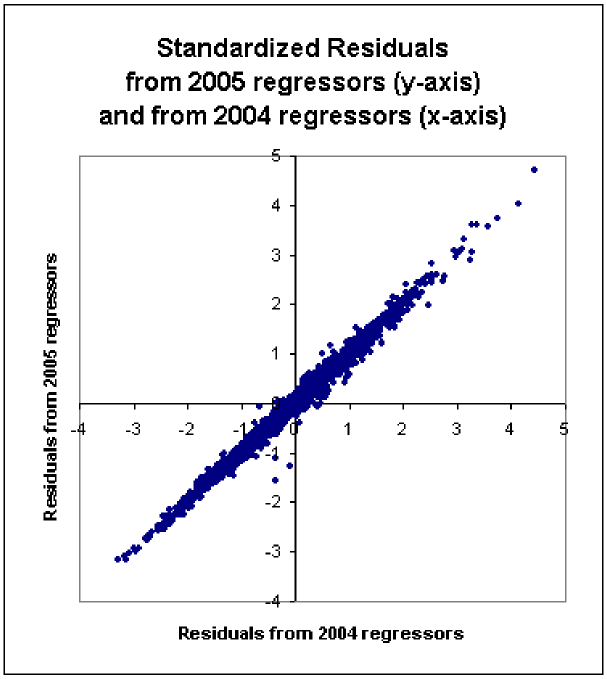
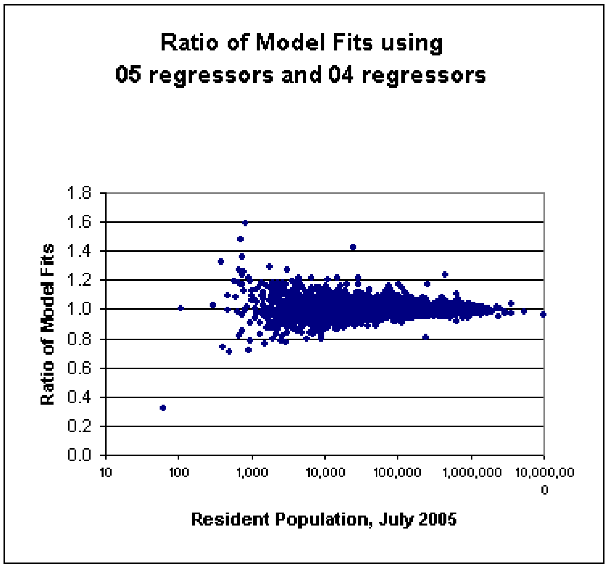


Figure 2.63: 2005 ACS standardized residuals from 2005 regressors versus 2005 ACS standardized residuals using 2004 regressors.

It is instructive to compare regression predictions using the two different years of regressors. In this case, differences can come from two sources: 1) changes in the coefficient estimates, and 2) changes in the regressor data themselves. Figure 2.64 plots the ratios of the 2005 ACS regression fitted values using 2005 regressors over those using 2004 regressors. In this figure, the observed differences can come from either sources (1) or (2). Since the coefficient estimates are almost identical between the two years, the observed differences are mostly due to changes in the regressor data themselves. Figure 2.65 plots ratios with the same denominators (regression fitted values from the model using 2004 regressors) but with different numerators. In this plot the numerators of the ratios come from taking regression parameters estimated using the 2005 regressors but with these parameters applied to the 2004 regressors. In this figure, the observed differences come from only source (1) since, in forming the two sets of regression predictions, the regressor variables in both cases refer to the 2004 values.

In Figure 2.64, most of the differences in the regression predictions are within 20 percent, and the percent differences are larger in absolute value for less populous counties. In Figure 2.65, most of the differences are within 5 percent, and there is little observable trend against county size. Given the much tighter cluster of points in Figure 2.65, it appears that much of the dispersion in regression predictions from using the 2005 rather than the 2004 regressors is due to differences in the 2005 and 2004 regressor data for particular counties (especially small counties), rather than differences in the estimated regression parameters. This observation supports the earlier observation from the regression output in Tables 2.18 and 2.19 that the summary statistics from models using either 2005 or 2004 regressors are very similar.



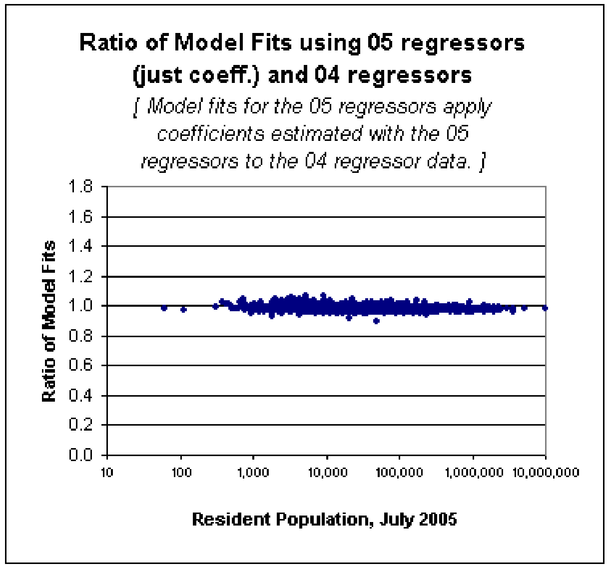


Figure 2.64 (left): Ratio of 2005 ACS regression prediction using 2005 regressors over that from using 2004 regressors.
Figure 2.65 (right): Ratio of 2005 ACS regression prediction using 2005 regressors (just coefficients) over that from using 2004 regressors.
The regression predictions for the 2005 regressors (just coefficients) apply coefficients estimated with the 2005 regressors to the 2004 regressor data.

An alternative choice for the 2005 regressors

The 2004 regressors include July 1, 2005 PEP population, ages 0 – 17 estimates and the tax year 2004 IRS child tax exemptions. These timeframes are used since most tax-year 2004 returns are filed in January to mid-April of 2005, and the filing address used in the winter and spring of 2005 in most cases is the same household address as that of the July 1, 2005 population. Thus, as discussed in Section 2.2, when the estimated coefficients on these two variables are of similar magnitude but opposite signs (as is the case in our results), the effect on the regression predictions of these two variables reflects the effects of a "log filer rate" variable. The July 2005 population estimates are important because they provide the population controls for the 2005 ACS county estimates, and thus they directly affect the dependent variable used in the models.

Analogously, the standard set of the 2005 regressors uses the July 1, 2006 PEP population, ages 0-17 and the tax-year 2005 IRS child tax exemptions. With this choice, however, the July 1, 2006 population estimates do not match the population controls used for the 2005 ACS. Therefore, an alternative choice of 2005 regressors would use the 2005 PEP population, ages 0 – 17. This would maintain consistency with the population controls that affect the dependent variable but lead to inconsistent timing with the 2005 IRS child tax exemptions, lessening the connection with the log filer rate variable. To see if this choice matters to the modeling, we also fit the model with this alternative version of the 2005 regressors. We found that results from using the alternative version of the 2005 regressors was very similar to those obtained using the standard version of the 2005 regressors. Thus, we have shown detailed results only for the model with the standard version of the 2005 regressors.

2.6 ESTIMATION OF ACS SAMPLING ERROR VARIANCES

Application of the models presented in Sections 2.2 and 2.3 requires estimates of the sampling error variances (denoted v_i) of the ACS direct county estimates being modeled. These estimates are either (*i*) the logs of the direct ACS county estimates of numbers of related 5 - 17 poverty, or (*ii*) the logs of the county estimates of related 5 - 17 poverty rates. We examined several approaches to obtaining these sampling error variance estimates.

A standard approach to such problems is to start with direct sampling error variance estimates of estimated totals or rates (in this case, of the direct county estimates of number in poverty or poverty rates). From these, obtain variances of any nonlinear functions (here logarithms) of the estimated totals or rates by Taylor series linearization (Wolter 1995). The linearization approximation to the variance of the log of a survey estimate is its relative variance, which is the square of the CV. If we view the survey estimate y_i to have mean $E(y_i) = \mu_i$ and variance $\text{Var}(y_i) = \sigma_i^2$, then this approximation of Var(log(y_i)) by the relative variance of y_i is

RelVar approximation:

$$\text{Var}(\log(y_i)) \equiv \text{v}_i \approx \text{RelVar}(y_i) = CV_i^2 = \sigma_i^2 / \mu_i^2 \tag{10}$$

and this is usually estimated by plugging in the direct sampling error variance estimate for σ_i^2 and plugging in y_i for μ_i.[25] (Note the possible alternative estimate of μ_i given in the footnote below.)

Phillip Kott, in discussing a paper by Jiang, et. al. (2001), offered a cautionary note on this approach, pointing out that Taylor series linearization is a large sample result, and thus questioning its applicability in the context of small area estimation modeling.[26]

Another approach makes use of properties of the lognormal distribution. Thus, if y_i is assumed lognormal with mean and variance $E(y_i) = \mu_i$ and $\text{Var}(y_i) = \sigma_i^2$ as above, then an exact result for the variance of log(y_i) is

$$\text{Log normal result: } \text{Var}(\log(y_i)) \equiv \text{v}_i = \log(1 + \sigma_i^2 / \mu_i^2) \tag{11}$$

Comparing (10) and (11) we note that if $CV_i^2 = \sigma_i^2 / \mu_i^2$ is not large, then the Taylor series approximation of (11) gives $\log(1 + CV_i^2) \approx CV_i^2$, and (10) and (11) will be approximately equal. This approximation rapidly diverges for CV_i greater than one-half, however, so if a log-

---

[25] Since here we are reflecting sampling error variation only, we can view the distribution of y_i as conditional on the true population quantity Y_i. The mean, μ_i, can thus be thought of as Y_i, though, as noted above, in practice one usually substitutes in the direct estimate y_i as an estimate of μ_i. In the modeling context here an alternative, and more stable, estimate of Y_i could be obtained from the regression predictions of the model (e.g., $\exp(x_i'\hat{\beta})$).

[26] Jiang, et al. (2001) took direct survey estimates of proportions, applied the arc-sin transformation, and used linearization to obtain approximate sampling error variances of the transformed estimates.

normal distribution is truly appropriate, one should dispense with the approximation (10) and use the exact formula (11).

Initial regression predictions to the ACS data used either (10) or (11) to obtain sampling error variances of the log estimates. Direct sampling error variance estimates for y_i were obtained by replication methods using replicate weight files available for the ACS micro data. Strong evidence of residual variance heterogeneity was found, however, leading us to seek another method.

The third approach we examined was to directly estimate the sampling error variance of the county estimates of log number of 5 - 17 poverty (and of log poverty rates). The same sets of replicate weights to estimate the sampling error variances of the un-logged direct survey estimates can be used to directly estimate the sampling error variances for any (suitably smooth) nonlinear transformations of these estimates.

Figure 2.66 plots the outcomes of these three methods for estimating the sampling error variances of the logs of the direct 2005 ACS county estimates of related 5 - 17 in poverty. It can be seen that both (10) and (11) provide reasonable smoothed approximations to these direct variances for low-levels of CV_i. As CV_i increases, however, it becomes apparent that neither of the results from (10) or (11) are tenable.

Using these direct estimates for the variance of log(y_i) in the models resulted in a reduction of the residual heterogeneity, as well. As seen in the Spearman's statistic results reported in Table 2.5, no significant heterogeneity was detected for the log-level model. Using the methods of (10) or (11), t-statistics near 8 were obtained for the same estimator.

The wide scatter seen for direct variance estimates in Figure 2.66 for counties with high CVs presumably reflects instability in the direct variance estimates for counties with relatively small ACS samples. As discussed in Sections 2.2.3 and 2.3.3, this instability impacts outlying residuals. To address this issue, we are investigating fitting GVFs to smooth the direct replicate weight variance estimates.

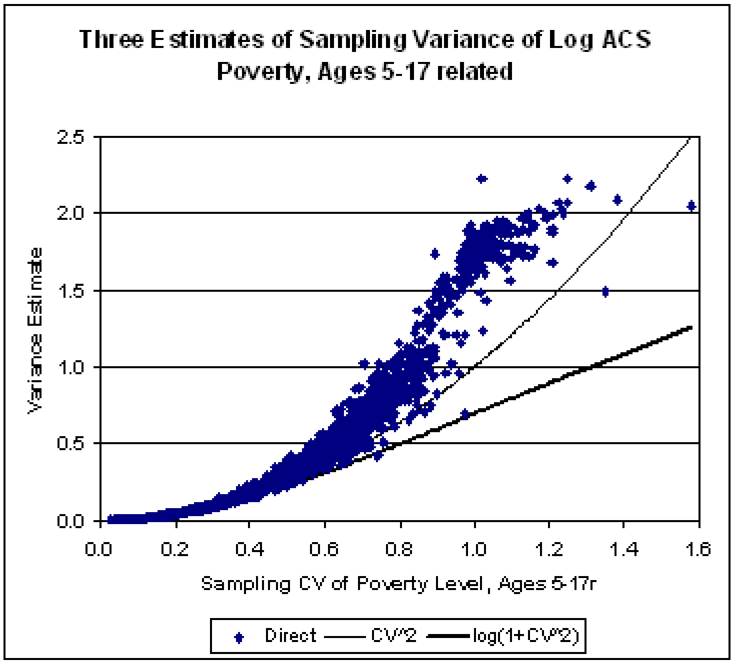


Figure 2.66: Estimates of sampling error variances of log (ACS survey estimates of number in poverty, ages 5-17 related) against sampling CVs of poverty-level.

2.7 ACS STATE RAKING FACTORS

2.7.1 State poverty models

SAIPE state poverty estimates have been produced to date from Fay-Herriot type models applied to CPS ASEC direct survey estimates of state poverty ratios.[27] Large sampling error variances in the CPS ASEC survey estimates for many states necessitated the use of the models to improve the estimates.[28] Given the much lower sampling error variances of the ACS direct survey estimates, modeling the ACS state data may not be necessary. The SAIPE program thus faces the question of whether or not to model the ACS data at the state level. The decision on this question will affect the ACS county shrinkage estimates, as county results must be raked to agree with the ACS state shrinkage estimates.

The SAIPE state poverty ratio models are in the form of equations (1) and (2), except that the models are applied to untransformed poverty ratios, that is, with the log (•) notation dropped in equations (1) and (2). The regression variables in x_i, in addition to the intercept term, are ratios

[27] As noted earlier, poverty ratios are defined as the number in poverty in a given age group divided by the population of the age group, while for poverty rates, the denominator is the "poverty universe" for the age group, not the total population for the age group.
[28] In the annual CPS ASEC direct survey estimates of state poverty, the Census Bureau used three-year averages of the CPS ASEC data to reduce the sampling variability.

related to poverty. In the most recent SAIPE state 5 - 17 poverty ratio model, they include the following features (definitions, stipulations, etc.): [29]

- IRS tax-poor child exemption rate = (number of child exemptions on tax returns with incomes below the poverty threshold in the state) / (total child exemptions in the state),
- IRS nonfiler ratio = 1 – (number of exemptions in the state) / (state population),
- state rate of participation in the food stamp program = (number of food stamp participants) / (state population), and
- "census residuals" obtained by regression of the previous census state 5 - 17 poverty ratio estimates on the other regression variables defined in the census income year (1999 for Census 2000).

This model was applied to 2005 ACS data (rates not ratios) using the regression variables as defined for the model for the CPS ASEC estimates of 2004 poverty ratios. [30] With the differences in timing between ACS and CPS ASEC poverty estimates (discussed in Section 1.4.3), one could question whether 2005 ACS data should instead be modeled with regression variables defined for 2005. Huang and Bell (2004) examined this timing issue using data from the ACS demonstration surveys for several years and found that shifting the timing of the ACS model regression variables this way by one year made little difference to the results.

As is the case with the CPS ASEC state models, the model with ACS data was given a Bayesian treatment with flat priors on σ_u^2 and on the regression parameters β (Bell 1999). The results obtained were posterior means and variances of the true poverty rates Y_i, which were obtained from the Bayesian analogues of equations (5) and (6).[31] These results were then used to construct 90 percent prediction intervals assuming normality.

The results from modeling the 2005 ACS data showed little effect on the point estimates or their model error variances for most states – that is, for most states the shrinkage estimates and their variances from equations (5) and (6) were about the same as the direct ACS results. This is due to the small sampling error variances of the ACS survey estimates for most states. When the sampling error variance, v_i, is small, then the weight, $1 - w_i$, given to the ACS survey estimate in equation (5) is close to 1. Also, the first term on the right hand side in equation (6) for the variance of the shrinkage estimate can be shown to approach v_i as the ratio $v_i / \hat{\sigma}_u^2 \to 0$, and the second term on the right hand side of (6) approaches zero since w_i approaches zero. For most states, the v_i was small enough that these results seemed to hold approximately. However, for the ten or so states with the largest sampling error variances, the modeling and shrinkage

[29] For further details see the SAIPE web site at http://www.census.gov/hhes/www/saipe/techdoc/methods/state/04statemod.html

[30] When this work was done, direct ACS poverty rates were available whereas direct ACS poverty ratios were not. As the difference between poverty rates and ratios is not that large and should be fairly consistent across states (poverty ratios are lower than poverty rates), this difference matters little for the purposes here, which are to examine the potential benefits of modeling the ACS data.

[31] The Bayesian version extends equation (6) to account for uncertainty about σ_u^2.

estimation did have an appreciable effect on the results. This is illustrated in Figures 2.67 and 2.68.

Figure 2.67 plots differences between the 2005 ACS direct survey estimates of state poverty rates, ages 5-17 related and the 2005 ACS shrinkage estimates of state poverty rates, ages 5-17 related against 2005 PEP state total resident population (the latter plotted on a log scale). Most of the differences are less than one percent in magnitude, except for the smallest states (which tend to have the smallest sample sizes and largest sampling error variances).

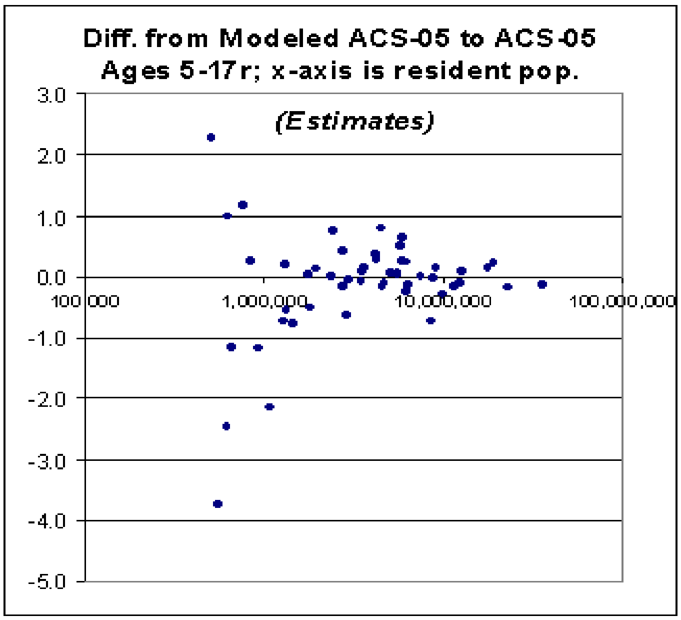


Figure 2.67: The 2005 SAIPE shrinkage ratio minus 2005 ACS direct survey ratio against 2005 PEP state total resident population.

Figure 2.68 shows plots of the half-widths of 90 percent confidence intervals for 5 - 17 state poverty rates from the ACS direct survey estimates and the fitted model. The half-width is 1.645 times the appropriate prediction standard error, thus a plot of the prediction standard errors would have exactly the same appearance. Multiplying by 1.645 changes the scales so they may be interpreted as the "plus or minus" margins of error in the corresponding model-based prediction and survey estimate confidence intervals, aiding interpretability of the results.

All of the points in Figure 2.68 lie below the 45-degree line, reflecting predicted improvements from the model-based estimator. (Keep in mind, though, that these model-based results are optimistic in that they assume that the model being used is correct.) However, for most of the states the points are not that far below the 45-degree line, so that modeling does not provide much improvement and, from Figure 2.67, also does not change the estimates much for these states. For the 10-20 states with the largest standard errors on the direct ACS estimates, though, more substantial improvements are realized. Opinions can differ about the point at which the statistical uncertainty in the direct ACS estimates is large enough that improvements from modeling appear to be worth pursuing, as well as about how worthwhile the apparent improvements from modeling actually are (weighing these against risks of model failure). The "state" (state-equivalent) with the largest standard error on the direct ACS estimate and the largest predicted improvement from modeling is Washington, DC. This is interesting given that Washington also appears in the county model, though as long as the county poverty estimates are raked to the state poverty estimates (direct or model-based), its estimate will be determined by the state results.

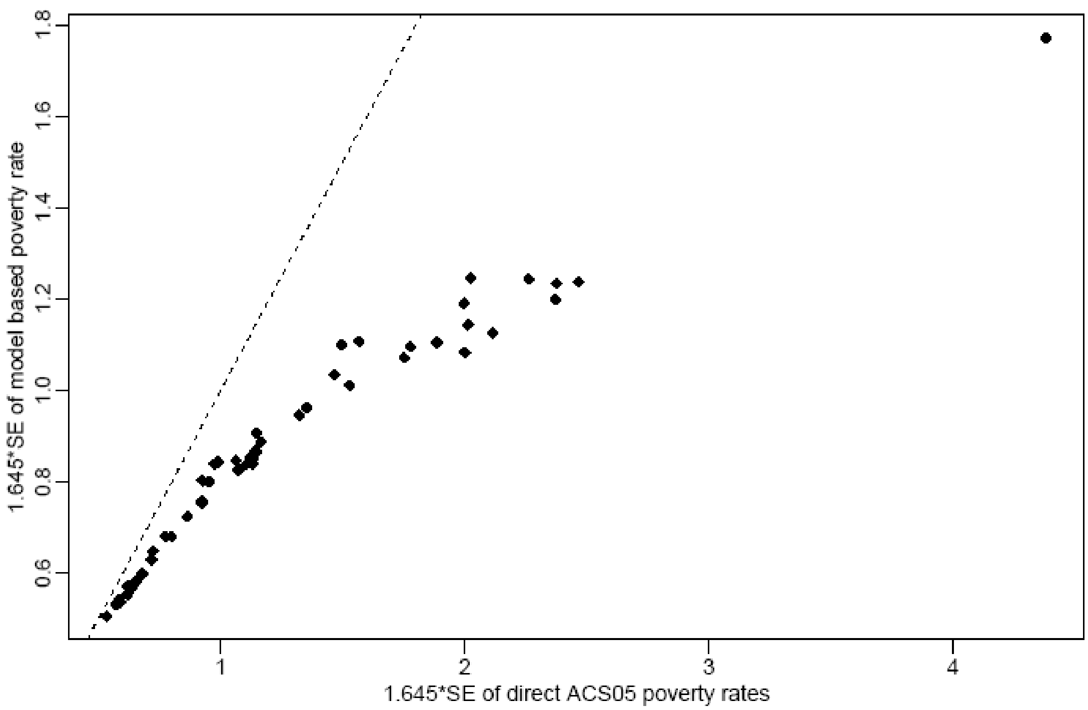


Figure 2.68: Comparing 90 percent confidence interval half-widths (1.645 times the prediction standard errors) for 2005 ACS shrinkage estimates of state poverty rate estimates, ages 5-17 related versus the 2005 ACS direct survey estimates of state poverty rate, ages 5-17 related.
The dotted line is a 45-degree (y = x) line.

2.7.2 Raking of estimates

To date SAIPE state and county model-based estimates of number in poverty have been raked (proportionally adjusted) to maintain consistency from the county to the state to the national level. That is, state estimates of number from the state CPS ASEC model have been raked to direct CPS ASEC national estimates of number in poverty, and county estimates of number in poverty from the county CPS ASEC model have been raked to these raked state model estimates of number in poverty.[32] In moving to use of ACS data for the state estimates (be they direct or model-based) and county estimates, we can reconsider the raking for both the state and county-level estimates.

With regard to raking the state estimates, if direct ACS state estimates were to replace the CPS ASEC model-based state estimates, then it would not be necessary to rake them to the direct ACS national estimate, because doing so would not change the results. If state estimates came instead from modeling direct ACS state estimates, then we could rake these to the ACS national total, but this does not appear to be advisable since doing so would modify all the state estimates, including those for states with very low sampling error variances (for which the shrinkage estimate is very nearly equal to the direct estimate). Another possibility would be to rake model-based state estimates from ACS data to the CPS ASEC national total, since the CPS ASEC defines the official direct poverty estimates at the national level. This would bring up other considerations, however, that are outside the scope of this paper.

The remainder of this section considers the effects of raking county model-based estimates (using 2005 ACS data) to state estimates. We consider these effects for both the log-level and log-rate county ACS models, and for raking to either ACS direct survey estimates or model-based shrinkage estimates.

Effects of raking county estimates to state estimates

Table 2.20 presents some summary statistics on the raking factors for the four combinations defined by the two model choices (log-level or log-rate) and two choices of state estimates of which to be raked (direct ACS or model-based). For any of these combinations the "raking factor" for a given state is the multiplier applied to the county model-based estimate of related 5 - 17 poverty to force the raked county estimates to sum to the appropriate state estimate being used as a control.[33] Since each combination has one raking factor for each state, these summary statistics are calculated across 51 state raking factors.

The mean raking factors are not very different when using direct ACS or ACS shrinkage estimates as controls, and similarly for the median raking factors. The means and medians are all slightly less than one. However, the spread of the raking factors using the ACS direct state estimates is considerably larger than the spread of the raking factors using the ACS state shrinkage estimates. These results hold for both the log-level and log-rate models. In fact, for a given choice of controls (direct or shrinkage ACS), the differences between the summary statistics for the log-level and log-rate models are rather small.

[32] State number poor estimates are obtained by multiplying the (Bayesian) shrinkage estimates of 5 - 17 related poverty ratios obtained from the state model by state 5 - 17 population estimates.
[33] For the log-rate model we convert model-based estimates of the log-rate to estimates of number poor as discussed in Section 2.4.

Table 2.20: Summary statistics for state raking factors, ages 5 - 17 related

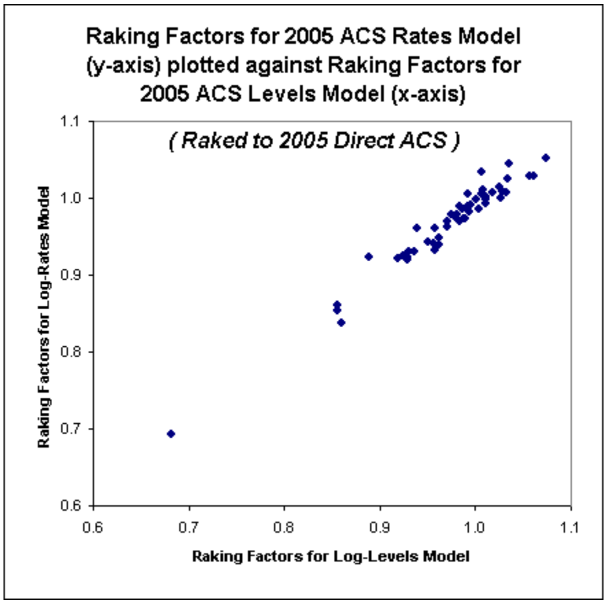
State Control	County Model	Raking Factors				
		Mean	Median	Standard Deviation	Minimum	Maximum
Direct ACS						
	Log-level (LL6u)	0.972	0.984	0.064	0.681	1.075
	Log-rate (LR6u)	0.966	0.978	0.060	0.692	1.051
Shrinkage ACS						
	Log-level (LL6u)	0.967	0.967	0.043	0.835	1.030
	Log-rate (LR6u)	0.960	0.965	0.040	0.859	1.032

Figures 2.69 and 2.70 plot raking factors for the log-rate model against raking factors for the log-level model. Figure 2.69 does this for the case where the ACS direct state estimates are used as controls, while Figure 2.70 does this for the case where the ACS state shrinkage estimates are used as controls. Within each plot, we see the fairly close agreement of the raking factors for any given state from the two models. Comparisons of the two plots show the larger variation of the raking factors across states when the direct ACS state estimates serve as controls instead of the ACS state shrinkage estimates.

When the direct ACS state estimates are used as controls (Figure 2.69), the largest downward raking factors for the log-level model (and for the log-rate model in parentheses) are: 0.68 (0.69) for Wyoming, 0.86 (0.86) for Utah, 0.86 (0.84) for South Dakota, and 0.86 (0.85) for North Dakota. The largest upward raking factors for the log-level model (and log-rate model in parentheses) are: 1.07 (1.05) for Vermont, 1.06 (1.03) for North Carolina, 1.06 (1.03) for Mississippi, and 1.04 (1.04) for Massachusetts.[34]

When the ACS state shrinkage estimates are used as controls (Figure 2.70), the largest downward raking factors for log-level (and log-rate in parentheses) are: 0.84 (0.87) for Hawaii, 0.86 (0.87) for Wyoming, 0.88 (0.91) for Washington, DC, and 0.88 (0.86) for Vermont. The largest upward raking factors for log-level (and log-rate in parentheses) are: 1.03 (1.00) for Mississippi, 1.02 (1.02) for Ohio and 1.01 (1.00) for Louisiana, and 1.02 (1.03) for Massachusetts.

[34] A raking factor of 0.67 means that the county estimates are scaled down by 33 percent in order to reach the state control, or that the sum of the pre-raked county estimates is 1 / 0.67 = 1.49 times the state control.



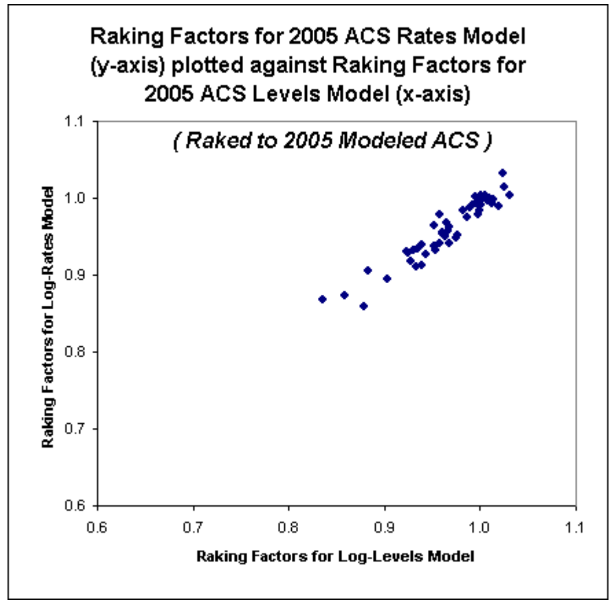
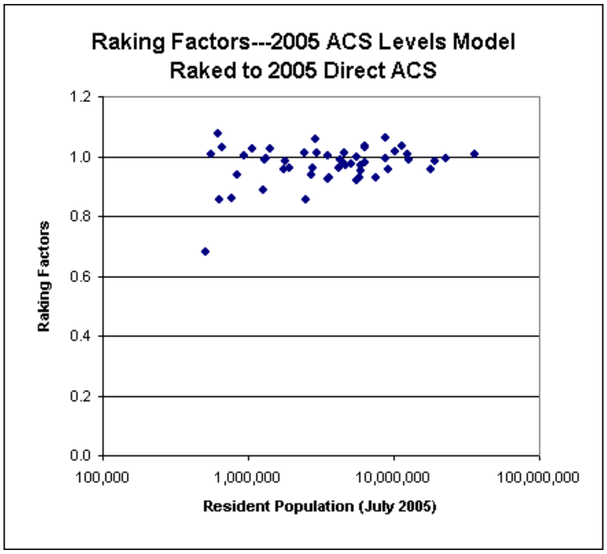


Figure 2.69 (left): State raking factors of the ACS state direct survey estimates for the log-rate county shrinkage estimates against those for the log-level county shrinkage estimates.
Figure 2.70 (right): State raking factors of the ACS state shrinkage estimators for the log-rate county shrinkage estimates against those for the log-level county shrinkage estimates.

Figures 2.71 and 2.72 plot the raking factors from the log-level model against state population (on a log scale), the former when the direct ACS state estimates are used as controls and the latter when the ACS state shrinkage estimates are used as controls. The corresponding plots for log-rate models appear very similar and so are not shown.

These plots again show that the raking factors for the ACS shrinkage estimates used as controls are less diffuse than those for the ACS direct state estimates used as controls. The plots also show more variation in the raking factors for the smaller states, particularly when the direct ACS state estimates are used as controls (Figure 2.71). This is expected given the higher sampling error in the ACS direct estimates for small states. Comparing the two plots we can see that for the largest states the raking factors appear to be very similar, which should be the case since for the largest states the model-based state estimates are very close to the direct ACS state estimates. Finally, the raking factors to the ACS shrinkage estimates appear to show a positive correlation with population size.



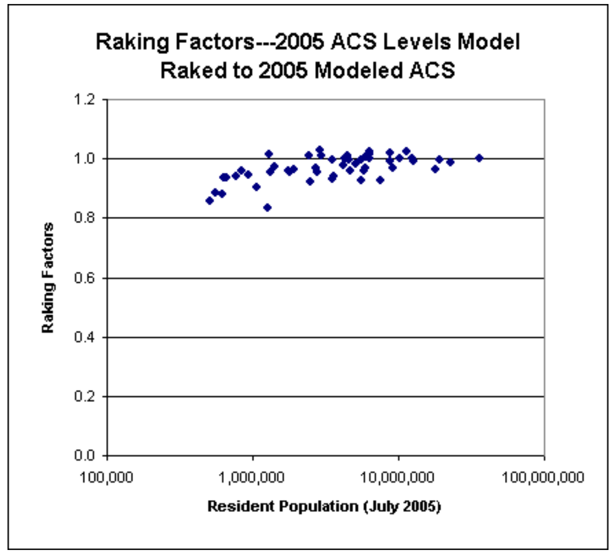


Figure 2.71 (left): Raking factors for ACS state direct survey estimates against 2005 PEP state total resident population

Figure 2.72 (right): Raking factors for ACS state shrinkage estimates against 2005 PEP state total resident population.

Chapter 3: Modeling Additional Years of ACS Data

The first full implementation of ACS was the 2005 ACS. Prior to this, there was an ACS demonstration period from 2000 through 2004. We will sometimes use the term "ACS demonstration surveys" to refer to the ACS surveys from this period, as well as referring specifically to the "2004 ACS," "2003 ACS," etc. The 2000 ACS survey is also known as the "Census 2000 Supplementary Survey (C2SS)," because it was done as a supplemental effort to Census 2000. Our interest in these additional years of "ACS data" is in using them to test the fit of our proposed county model for the 2005 ACS estimates. In doing this, however, we must be aware of some limitations that could arise from lack of comparability of the data from the full production ACS and the ACS demonstration surveys.

Two major differences between the ACS demonstration surveys and the full-production ACS are the following. 1) The sample for the ACS demonstration surveys was from roughly 800,000 addresses, while the sample for the full-production 2005 ACS was from roughly 3,000,000 addresses. This means, of course, that estimates obtained using ACS demonstration survey data, including direct survey estimates, estimates of model parameters, and model predictions, will have larger standard errors than corresponding results obtained from full-production ACS data. 2) Although the demonstration surveys used ACS methods, the sample design did not reflect the ACS sample design for full production because the demonstration surveys were designed to provide characteristic data for states and large entities of 250,000 or more, not for counties and other smaller entities.

The much larger sample size, county-specific design, and methodology improvements in the 2005 ACS relative to the ACS demonstration surveys may influence the distribution of county level ACS direct poverty estimates, which in turn may influence the corresponding ACS predictions of poverty and ACS shrinkage estimates. We need to keep this in mind in Section 3.1, which makes some comparisons of direct ACS poverty estimates over the period 2000-2005, and in Section 3.2, which examines results of log-level predictions of poverty for ACS data for 2000-2005. Fortunately, we find a fair degree of comparability of the results over the years of the ACS demonstration surveys, and these results also seem reasonably comparable to the results obtained using the 2005 ACS data that were discussed in Chapter 2. We do need to make obvious allowances for the higher sampling variability in the demonstration period. Thus, the changes over time in ACS do not appear to have large enough effects to invalidate this approach to assessing the model. Section 3.3 discusses some modeling results using the 2006 ACS data, the second year of full-production ACS, and compares these results with the modeling results from the 2005 ACS.

All modeling results presented here pertain to the LL6u log-level model not the log-rate model.

3.1 ACS DIRECT SURVEY ESTIMATES: 2000-2005

This section discusses ACS direct poverty estimates from 2000 through 2004, comparing these results over time and with the 2005 ACS direct poverty estimates. Various graphs compare results involving estimates from the 2005, 2004, and 2003 ACS surveys.

While the 2005 ACS surveyed from all 3,141 U.S. counties and published single-year estimates for about 775 counties, the ACS demonstration surveys typically sampled from about 1,200 counties and published single-year estimates for about 240 counties. Table 3.1 contains published ACS poverty rate estimates (for ages 5 - 17 related) at the national level and for the ten U.S. counties with the largest total populations and among the largest ACS sample sizes. The larger sample size of the 2005 ACS produces the much lower standard errors of the estimates. For example, the standard error of the age 5 - 17 related poverty rate for Los Angeles County is 1.1 percent for the 2004 ACS (on a sample size of 10,811) and is 0.6 percent for the 2005 ACS (on a sample size of 41,020).[35] From year to year, these ten large-sample county poverty rate estimates are fairly stable, and nearly all year-to-year changes are within the 90 percent confidence interval for the difference. Some larger changes occur in Maricopa County, AZ, with an estimated poverty rate (and standard error) of 18.0 percent (1.9 percent) in C2SS, 13.1 percent (1.5 percent) in the 2001 ACS, and 17.8 percent (1.5 percent) in the 2002 ACS; and Harris County, TX, with an estimated rate (and standard error) of 18.8 percent (1.1 percent) in the 2004 ACS and 25.1 percent (1.0 percent) in the 2005 ACS. The statistically significant changes across years could be due, of course, to economic changes, not just to differences in ACS samples or in ACS procedures.

[35] Reductions in the sampling standard errors from 2004 to 2005 ACS are not necessarily proportional to the square roots of the ratios of sample size due to, among other things, the effects of population controls on the variances.

Table 3.1: ACS direct survey estimates of poverty rate, ages 5 - 17 related from the ten largest counties for C2SS, 2001-2005 ACS[36]

County		ACS Poverty Rate, Ages 5-17 (Standard Error) Sample Size					
Name	ACS Total Population	C2SS	2001 ACS	2002 ACS	2003 ACS	2004 ACS	2005 ACS
United States	288,378,137	15.9 (0.2) 555,630	15.5 (0.2) 564,922	**16.2** **(0.2)** **482,470**	16.1 (0.2) 535,713	**16.9** **(0.1)** **534,140**	17.0 (0.1) 1,767,827
Los Angeles County, CA	9,743,498	23.8 (1.2) 10,637	**21.0** **(1.1)** **10,955**	21.6 (1.3) 9,904	23.1 (1.0) 10,892	22.5 (1.1) 10,811	22.8 (0.6) 41,020
Cook County, IL	5,195,724	17.0 (1.6) 6,892	19.6 (1.6) 7,009	19.0 (1.6) 6,555	18.4 (1.4) 7,012	20.1 (1.3) 6,855	20.2 (0.9) 25,128
Harris County, TX	3,839,608	16.4 (1.0) 8,057	17.6 (1.0) 6,703	**20.6** **(1.3)** **6,076**	20.6 (1.4) 6,885	18.8 (1.1) 6,882	**25.1** **(1.0)** **15,811**
Maricopa County, AZ	3,728,877	18.0 (1.9) 4,278	**13.1** **(1.5)** **4,454**	**17.8** **(1.5)** **4,143**	16.1 (1.4) 4,648	14.0 (1.7) 4,768	16.2 (0.7) 17,873
Orange County, CA	2,966,983	12.5 (1.4) 3,524	9.6 (1.4) 3,217	11.5 (1.7) 2,788	12.6 (1.3) 3,598	12.2 (1.4) 3,654	10.2 (0.7) 13,663
San Diego County, CA	2,829,144	18.1 (2.3) 3,814	14.3 (1.5) 3,924	15.9 (1.9) 3,562	15.5 (1.8) 3,865	16.6 (1.6) 3,882	15.2 (1.2) 14,219
Kings County, NY	2,470,756	25.6 (2.6) 2,457	27.1 (2.4) 2,623	31.7 (2.8) 2,348	28.5 (2.3) 2,644	32.6 (2.3) 2,564	29.6 (1.3) 10,834
Miami-Dade County, FL	2,340,671	21.7 (2.1) 2,626	25.3 (2.5) 2,591	23.8 (2.6) 2,384	21.7 (2.0) 2,710	22.4 (2.2) 2,605	21.6 (1.4) 9,019
Dallas County, TX	2,312,043	15.4 (2.3) 2,809	15.9 (1.7) 2,947	19.6 (2.0) 2,558	20.6 (1.9) 2,771	22.3 (2.0) 2,793	21.4 (1.1) 10,387
Queens County, NY	2,230,829	15.8 (2.4) 2,326	18.7 (2.3) 2,495	15.7 (2.2) 2,219	18.4 (2.5) 2,523	19.8 (2.3) 2,497	**15.4** **(1.0)** **9,787**

Since the county-level single-year ACS estimates are inputs to county level models, it is instructive to consider the properties of the ensemble of ACS county estimates across the years. However, two issues arise. First, some counties are in-sample one year and out-of-sample the next year (or have zero estimated poverty one of the years). Second, not all sampled counties in the C2SS are self-representing; that is, not all sampled counties are sampled with probability one.

[36] Boldface indicates statistically significant change from the prior year at the 10 percent level of significance.

For the sake of comparability of county-level ACS data across years, only self-representing counties that have non-zero estimated poverty in every year of the ACS demonstration surveys are included in most of the tables and figures that follow below. This constitutes a 621-county subset.

Figure 3.1 plots the 2005 ACS direct estimated county poverty rates for age 5 - 17 related against population size (all ages) for both the 621-county subset and the full 3,141 counties. As expected, the subset counties are uniformly among the larger counties, but it appears that the average poverty rate for the subset counties does not diverge much from the average poverty rate for counties with population greater than 100,000. However, the average poverty rate for counties with population below 100,000 (most of which are not included in the county subset) is slightly higher.

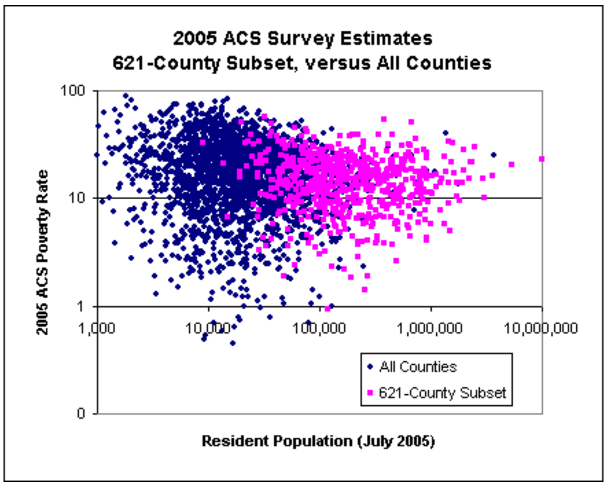
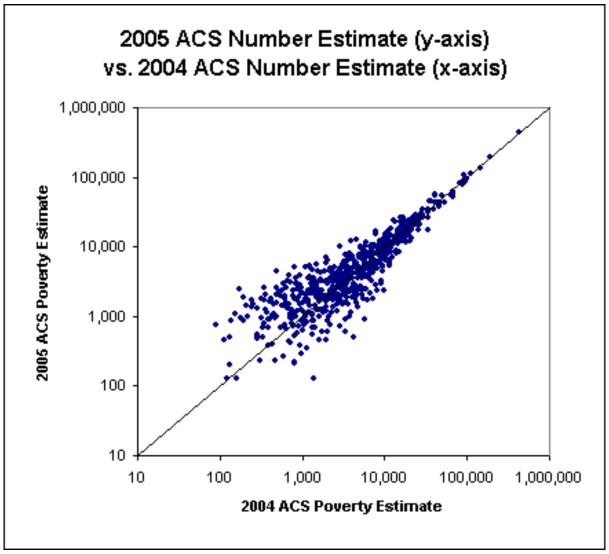


Figure 3.1: 2005 ACS rate survey estimates for the 621-county subset and for all counties, against 2005 PEP county total resident population.
Both the x- and y-axes are shown in log scale, but labeled in linear scale.

Figures 3.2 through 3.5 compare county-level ACS poverty estimates (for numbers and rates) across years for the county subset. Figure 3.2 plots the 2005 ACS survey estimates of number of 5 - 17 poverty against the 2004 estimate. Figure 3.3 similarly compares estimates from the 2004 ACS to the 2003 ACS. We see that the points in Figure 3.2 are a bit less diffuse than those in Figure 3.3, particularly at the low end, this being due to the effects of the higher sample size of the 2005 ACS. Also, while the points appear to be centered at the diagonal line in Figure 3.3, the points appear to be concentrated some to the left of the diagonal for the smaller counties in Figure 3.2.

Figures 3.4 and 3.5 are similar to Figures 3.2 and 3.3, except that they display estimated poverty rates instead of estimated numbers in poverty. Figure 3.4 appears less diffuse than Figure 3.5, again due to the higher 2005 ACS sample size, and the points in both figures appear to be generally centered at the diagonal line. The concentration above the diagonal in Figure 3.2 for the levels data is not as apparent in Figure 3.4 for the poverty rates. The same plots for C2SS through the 2003 ACS appear very similar to those below.



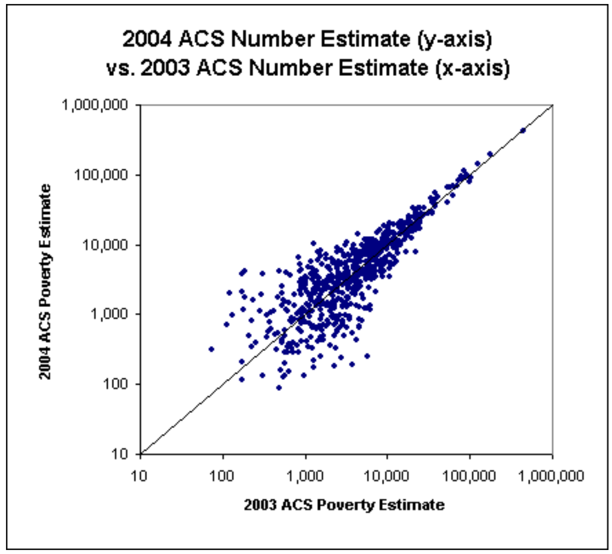
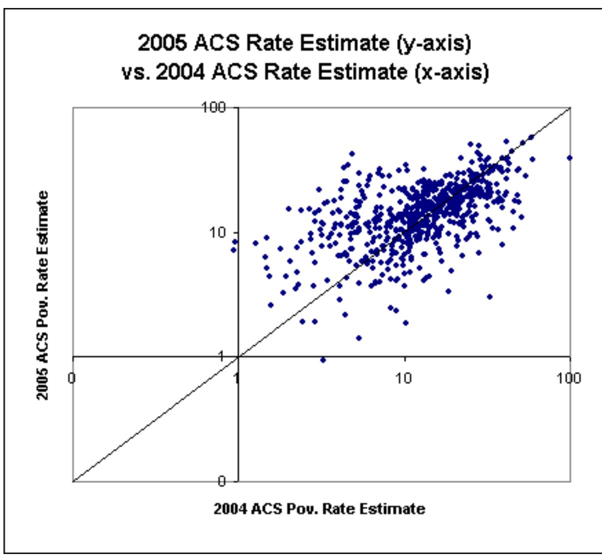


Figure 3.2 (left): 2005 ACS number survey estimates against 2004 ACS number survey estimates for the county subset.

Figure 3.3 (right): 2004 ACS number survey estimates against 2003 ACS number survey estimates for the county subset.

In both figures both the x- and y-axes are shown in log scale but labeled in linear scale.



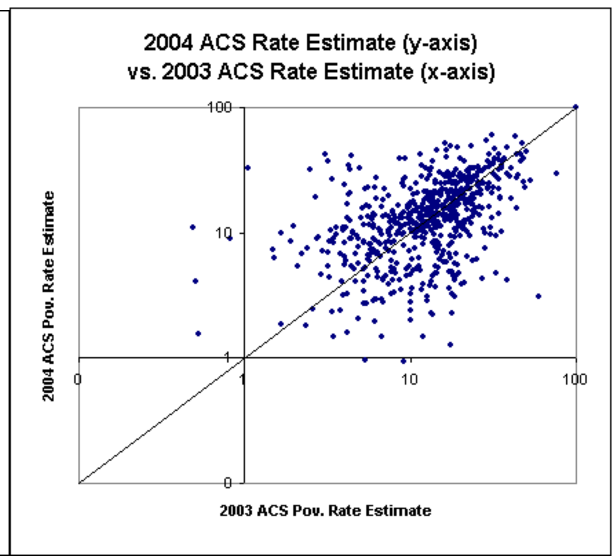


Figure 3.4 (left): 2005 ACS rate survey estimates against 2004 ACS rate survey estimates for the county subset.

Figure 3.5 (right): 2004 ACS rate survey estimates against 2003 ACS rate survey estimates for the county subset.

Regarding the relative precision of the ACS survey estimates, Figure 3.6 shows that the 2005 ACS survey estimates overall had much lower CVs than the ACS demonstration survey estimates because of the much greater sample size in the 2005 ACS. Between prior years of ACS demonstration surveys, however, the average CVs were fairly constant.

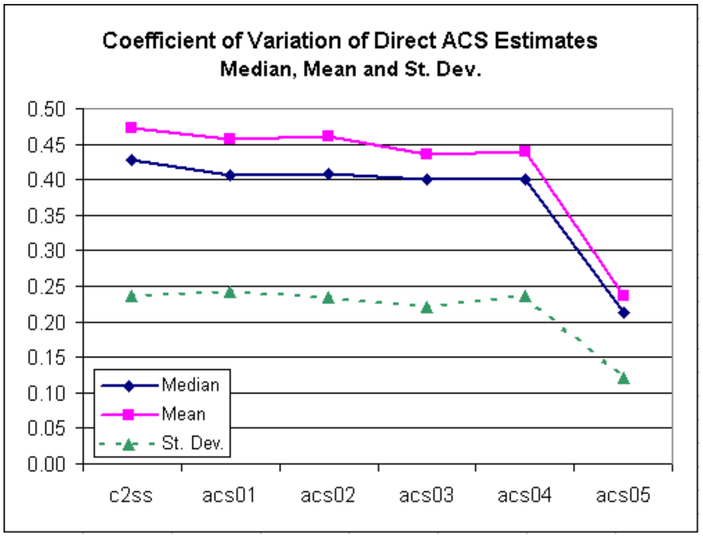
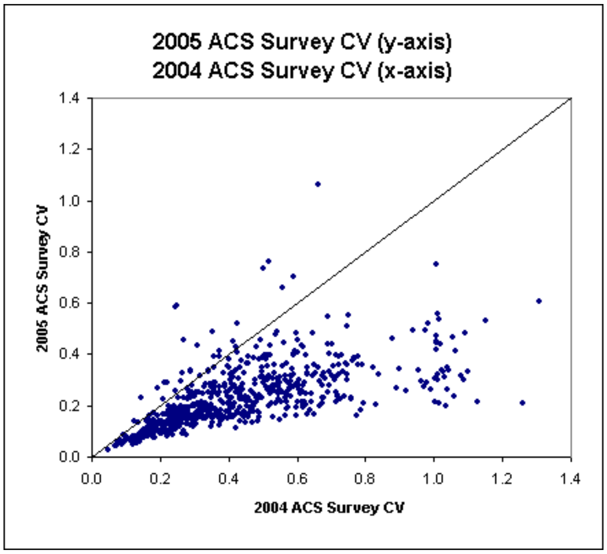


Figure 3.6: Mean, median and standard deviation of the CVs of the ACS survey estimates from C2SS through 2005 ACS for the county subset.

Figures 3.7 and 3.8 compare CVs of ACS poverty estimates over time for individual counties. Figure 3.7 show that the CVs of the 2005 ACS survey estimates are systematically lower than the CVs of the 2004 ACS survey estimates. In contrast, the CVs of the 2003 ACS and 2004 ACS demonstration surveys, shown in Figure 3.8, appear more comparable, with year-to-year differences probably due mostly to random estimation error.



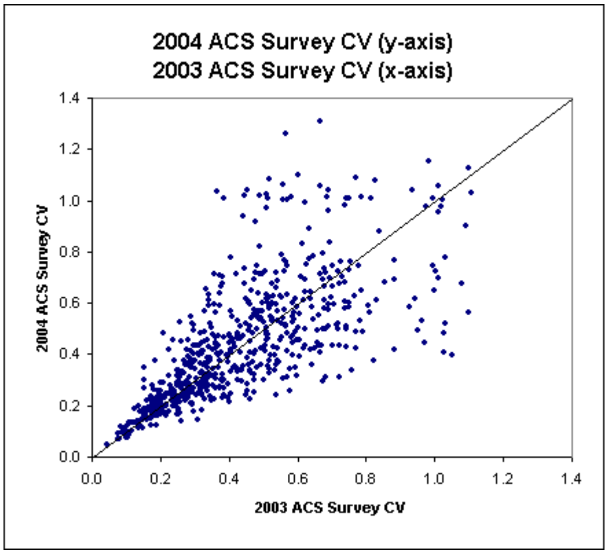


Figure 3.7 (left): CVs of the 2005 ACS survey estimates against CVs of the 2004 ACS survey estimates for the county subset.
Figure 3.8 (right): CVs of the 2004 ACS survey estimates against CVs of the 2003 ACS survey estimates for the county subset.

3.2 ACS REGRESSION PREDICTIONS: 2000-2005

This section discusses regression predictions from fitting models to data from the ACS demonstration surveys, 2000-2004. Results from the full-production 2005 ACS are included for comparison.

Model-fitting results

Since the 2000-2005 ACS models have the same yearly input data as the income years (IY)1999–2004 CPS ASEC models from the SAIPE program (though with a different set of county observations), CPS ASEC modeling results are also compared. These SAIPE production results with CPS ASEC data may serve to inform whether or not the differences in coefficient estimates across years in the ACS models can be attributed to changes in the distribution of input data rather than changes in the ACS survey itself.

The modeling results presented here for the ACS demonstration surveys are obtained from the same group of 621 counties for which results of direct ACS estimates were presented in Section 3.1, for the reasons discussed there. In contrast, the SAIPE 1999 through SAIPE 2004 results with CPS ASEC data presented below include results for both self-representing and non-self-representing counties in the CPS sample, and thus involve a different number of observations (counties) each year.

The regressors in the prior-year models are conceptually the same as the regressors in the 2005 ACS model discussed at length in Chapter 2. For example, the regressors used in the 2004 ACS model are the same as the regressors in the 2005 ACS model, except with a reference date one year earlier. Similarly, the regressors for the C2SS model are the same as the regressors for the 2005 ACS model, except with a reference date five years earlier. The CPS ASEC-based SAIPE production models also use the same regressors as the ACS-based models (with appropriate timing considerations). Thus, the SAIPE-CPS ASEC model for IY 2004 used the same regressors as the 2005 ACS model, and the SAIPE-CPS ASEC model for IY 2003 used the same regressors as the 2004 ACS model, etc.

Model estimation using the ACS poverty estimates for the 621-county subset is, as before, by maximum likelihood via an iterated weighted least squares procedure. Sampling error variances of the ACS estimates are set to the corresponding direct variance estimates (as discussed in Section 2.6 of Chapter 2) and then treated as known. Recall that this estimation approach differs from the approach used when the models were applied to the CPS ASEC estimates. In the latter case for each year the model error variance was set to the value obtained by fitting the Census 2000 auxiliary equation, and the CPS ASEC sampling error variances were obtained as part of the regression prediction using a simple parametric function dependent on sample size.

Table 3.2 shows regression prediction results from the years of ACS demonstration surveys alongside results from the 2005 ACS model. Results for the 2005 ACS model are shown twice: once based on the full-county estimation and once based on the 621-county estimation. Also included are results from corresponding years of the CPS ASEC models.

The results from the 621-county subset 2005 ACS model are similar to those from the all-county 2005 ACS model. The estimated coefficients of log(PEP population, ages 0-17) and log(IRS child tax exemptions) are both larger in magnitude for the all-county model. However, as discussed in Section 2.2 of Chapter 2, the more relevant consideration with these regressors appears to be the sum of their two coefficient estimates, which is close to zero in both cases here. So, overall, the county subset 2005 ACS model results can represent a base case for comparison with model results based on prior ACS years.

In the ACS models (left side of table), the estimated model intercept becomes progressively more negative between the first and last years of estimation. The estimated coefficient on IRS child tax-poor exemptions is greater in the years further from the Census year (and is significant in all years except for the C2SS). The Food Stamps coefficient estimates show a similar trend. Conversely, the estimated coefficient on the Census 2000 variable is largest near the Census year. While this coefficient is significant in all years, its larger values near the census year (and the corresponding smaller values of the coefficients on the IRS child tax-poor exemptions and Food Stamp variables,) may be due to the presumably greater relevance of the Census 2000 variable for years close to the census year. The estimated coefficients on log (PEP population, ages 0-17) and log (IRS child tax exemptions) vary over the years, but their sum is always close to zero.

Somewhat similar patterns can also be observed in the CPS ASEC regression predictions (right side of table). In particular: 1) the estimated coefficients on the IRS child tax-poor exemptions and Food Stamp variables tend to be larger in years further from Census 2000; this behavior is not as well defined in the CPS ASEC model, 2) the estimated coefficient on the Census 2000 poverty variable is higher near the census year and lower in recent years, 3) the estimated coefficients on the log (PEP population, ages 0-17) and log (IRS child tax exemptions) variables vary over time, but their sum is more stable and often close to zero. The larger amount of variation over time seen in the results from the CPS ASEC models may be due to the higher sampling error variances of the CPS ASEC county estimates, which lead to higher variances on the parameter estimates for the CPS ASEC models.

The improvement in R-squared in the 2005 ACS model relative to the 2000-2004 values is expected due to the lower sampling error variances of the 2005 ACS estimates. The improvement in R-squared in the CPS ASEC models for income years 2000-2004 relative to income year 1999 is also expected due to an increase in the CPS ASEC sample size for income year 2000 onward.[37]

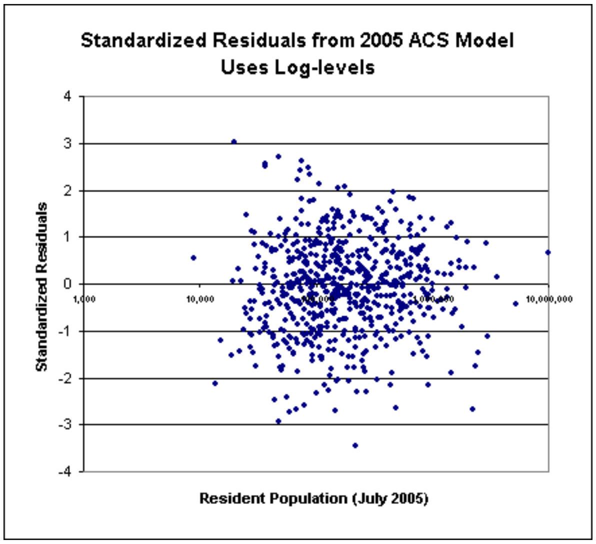
[37] Starting with the 2001 CPS ASEC, which collected income for 2000, the sample was significantly expanded to improve the reliability of certain estimates used in the funding formula for the State Children's Health Insurance Program (SCHIP).

Table 3.2: Regression prediction results from the 621-county subset C2SS - 2005 ACS log-level models (and full 2005 ACS) and from the corresponding CPS ASEC models

	ACS models by survey year: Log (ACS poverty) coefficient (t-statistic)							CPS ASEC models by income year: Log (CPS ASEC poverty) coefficients (t-statistic)					
	C2SS	2001	2002	2003	2004	(subset) 2005	(all) 2005	1999	2000	2001	2002	2003	2004
Intercept	0.13	-0.01	-0.24	-0.19	-0.65	-0.51	-0.42	-0.29	-0.24	-0.38	-0.61	-0.85	-0.99
	(0.55)	(-0.06)	(-1.14)	(-0.97)	(-3.01)	(-3.93)	(-7.39)	(-1.31)	(-1.28)	(-2.20)	(-3.57)	(-5.67)	(-6.77)
Log (IRS child tax-poor exemptions)	0.27	0.22	0.44	0.39	0.63	0.50	0.55	0.28	0.69	0.84	0.83	0.51	0.40
	(1.98)	(2.07)	(4.19)	(4.21)	(6.59)	(8.23)	(12.26)	(1.49)	(4.88)	(6.30)	(6.42)	(4.50)	(3.66)
Log (Food Stamps)	0.07	0.10	0.04	0.11	0.20	0.21	0.17	0.10	-0.02	-0.02	0.14	0.19	0.19
	(1.18)	(1.97)	(0.88)	(2.45)	(4.04)	(6.25)	(7.90)	(1.45)	(-0.34)	(-0.27)	(2.43)	(3.75)	(3.66)
Log (PEP population, ages 0-17)	0.41	0.56	0.64	0.65	0.14	0.85	1.05	0.29	0.92	0.85	0.93	0.79	1.10
	(1.27)	(1.58)	(1.75)	(1.93)	(0.38)	(3.90)	(8.62)	(0.75)	(2.14)	(1.98)	(2.31)	(2.26)	(3.32)
Log (IRS child tax exemptions)	-0.43	-0.56	-0.60	-0.62	-0.16	-0.83	-1.04	-0.15	-0.91	-0.87	-0.97	-0.68	-0.89
	(-1.37)	(-1.66)	(-1.73)	(-1.97)	(-0.46)	(-4.07)	(-9.13)	(-0.38)	(-2.19)	(-2.14)	(-2.54)	(-2.06)	(-2.87)
Log (Census 2000 poor, ages 5-17 reltd)	0.66	0.67	0.47	0.47	0.21	0.27	0.27	0.45	0.30	0.20	0.08	0.19	0.21
	(4.40)	(5.86)	(4.55)	(5.83)	(2.81)	(6.01)	(8.90)	(2.56)	(2.31)	(1.77)	(0.82)	(2.25)	(2.85)
Degrees of freedom	615	615	615	615	615	615	2966	894	995	1031	1043	1227	1258
Model error variance	0.022	0.013	0.022	0.022	0.021	0.012	0.022	0.017	0.017	0.017	0.017	0.017	0.017
R-squared	0.909	0.937	0.925	0.931	0.917	0.964	0.936	0.796	0.831	0.840	0.843	0.852	0.862

Analysis of residuals

Standardized residuals from the 2005 ACS and 2004 ACS LL6u log-level models are plotted below. All models were fit with the 621-county subset discussed in Section 3.1. Figures 3.9 and 3.10 plot the standardized residuals of the 2005 ACS model and 2004 ACS model against the 2005 PEP county total resident population. Figures 3.11 and 3.12 have the same y-axis as Figures 3.9 and 3.10, but the x-axis is the Census 2000 poverty rate, all ages. As these four figures show, the dispersions of the standardized residuals for the county subset 2005 ACS and 2004 ACS models are similar to one another, and do not give evidence of any dependence on population size or on Census 2000 poverty rates. The same plots for C2SS through the 2003 ACS appear very similar to those below.



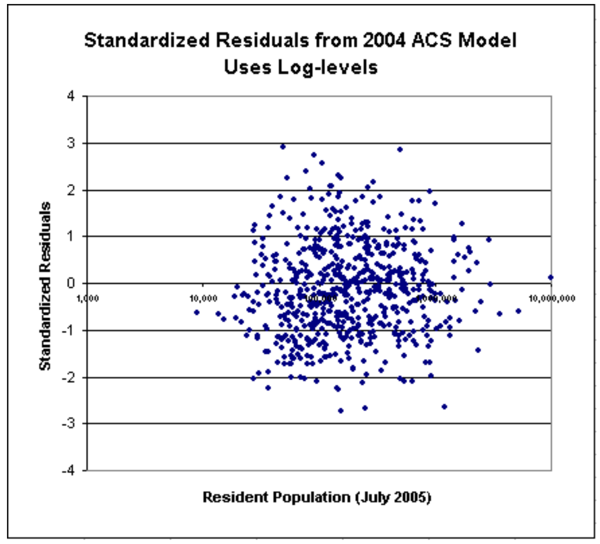
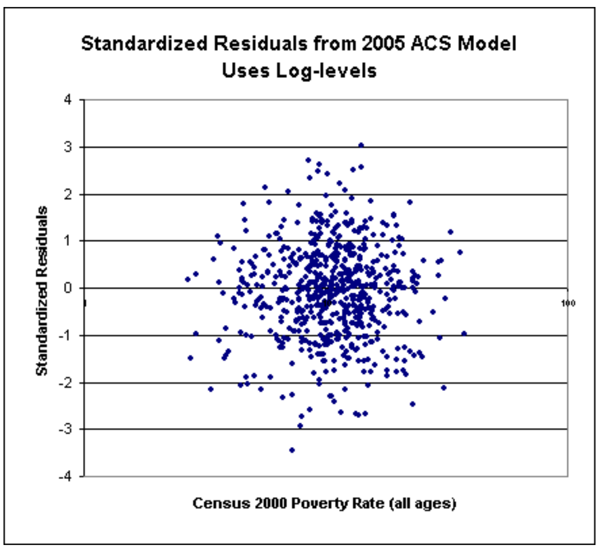


Figure 3.9 (left): Standardized residuals from the 2005 ACS model against 2005 PEP county total resident population for the 621-county subset.
Figure 3.10 (right): Standardized residuals from the 2004 ACS model against 2005 PEP county total resident population for the 621-county subset.



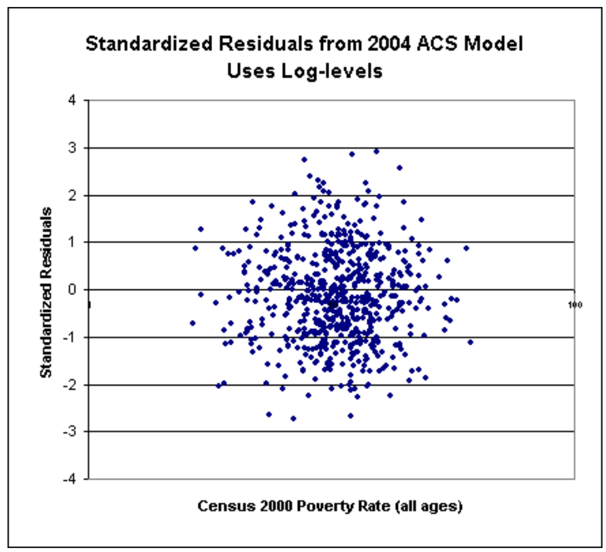
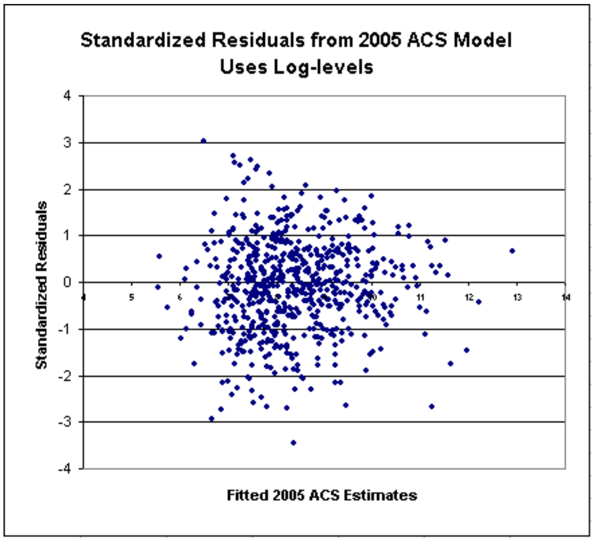


Figure 3.11 (left): Standardized residuals from the 2005 ACS model against the Census 2000 poverty rate for all ages for the county subset.
Figure 3.12 (right): Standardized residuals from the 2004 ACS model against the Census 2000 poverty rate for all ages for the county subset.

Figures 3.13 and 3.14 again show standardized residuals on the y-axis (as in Figures 3.9 to 3.12), but now have the ACS predictions of poverty on the x-axis. Figure 3.13 shows a small number of large negative residuals for large 2005 ACS regression predictions. This is not so evident in Figure 3.14. The same plots for C2SS through the 2003 ACS appear very similar to those below.



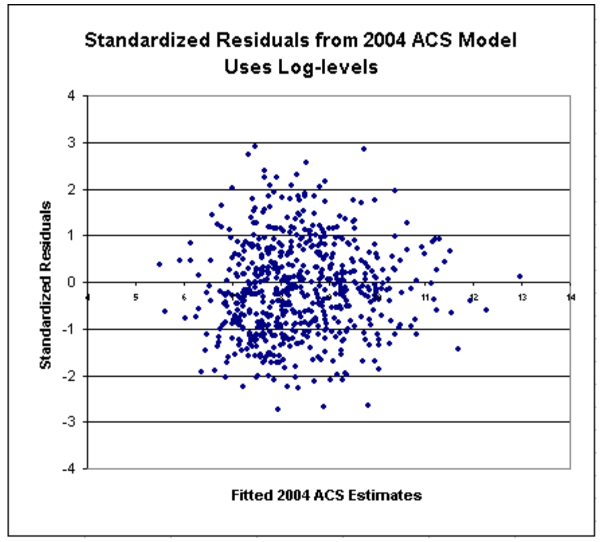
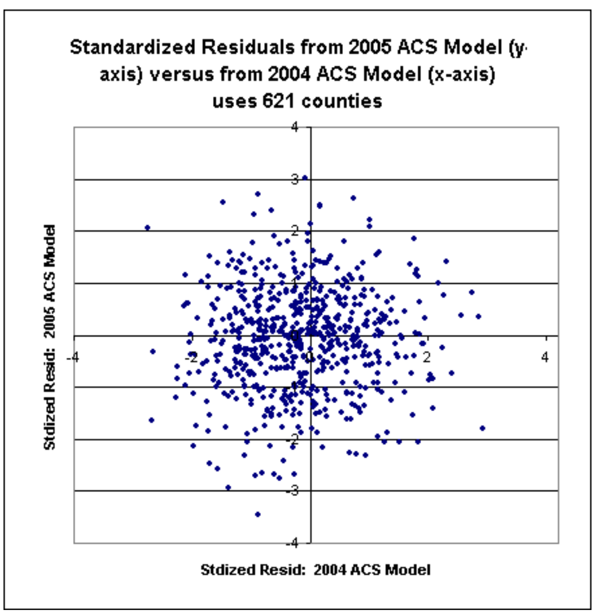


Figure 3.13 (left): Standardized residuals from the county subset 2005 ACS model against the 2005 ACS regression predictions.
Figure 3.14 (right): Standardized residuals from the 2004 ACS model against the 2004 ACS regression predictions.

Figures 3.15 and 3.16 plot the standardized residuals from one year of ACS model against the prior year of ACS model (for the 621-county subset), providing some assessment of the stability of individual county residuals from year to year. Both figures show neutral cross-year patterns in residuals. This indicates that model underpredictions in one year are just as likely model over- or underpredictions in the next year. High survey observations or model predictions from one year are not systematically high or in the same direction the next year.



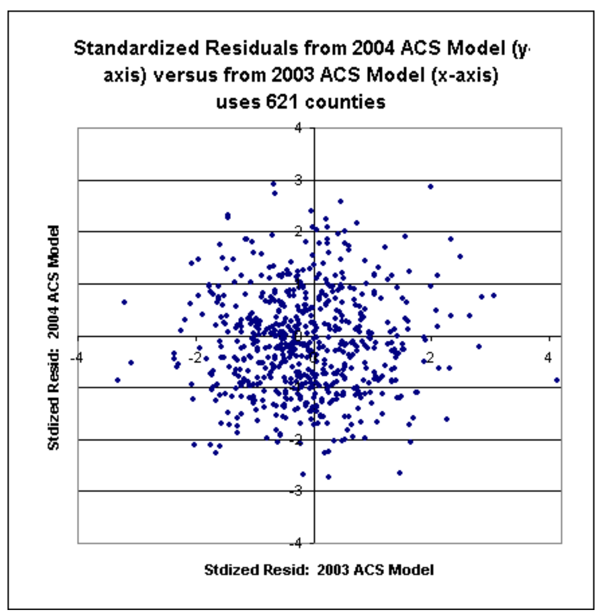


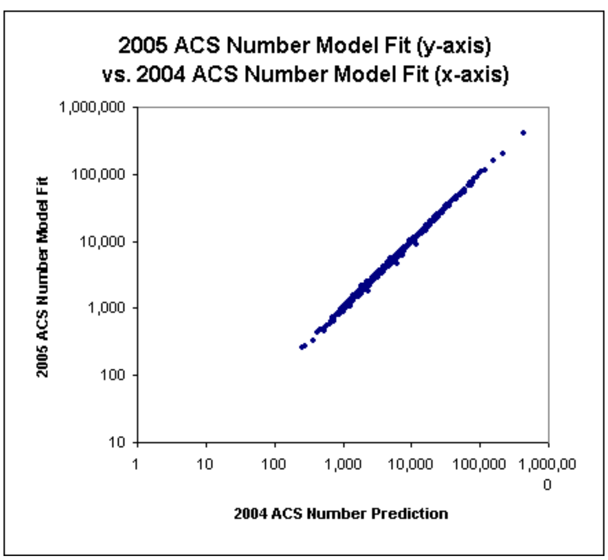
Figure 3.15 (left): Standardized residuals from the 2005 ACS model against standardized residuals from the 2004 ACS model for the county subset.
Figure 3.16 (right): Standardized residuals from the 2004 ACS model against standardized residuals from the 2004 ACS model for the county subset.

Box-whisker plots of the standardized residuals from C2SS through 2005 are included in Appendix C. These box-whisker plots group the categorization variables into quintiles. Each quintile contains 124 counties of the 621 counties from regression predictions. The categorization variables are mostly from Census 2000 and include the following concepts: C.8 Total population, C.9 Percent poor, C.10 Population growth 1990-2000, C.11 Population growth 2000-2005, C.12 Percent Black, C.13 Percent Hispanic, C.14 Percent Asian, C.15 Percent group quarters. The Appendix C figures show median residuals slightly less than those for the smallest population quintile of the 621 counties for the 2004 ACS. There were no discernible patterns with respect to the other classification variables.

Comparisons of regression prediction results

This section discusses the 621-county subset for the C2SS through the 2005 ACS. Results from log-level number regression predictions are examined both in numbers form and in rates form. To transform the ACS log-level regression predictions from log numbers into numbers, the predictions are first exponentiated and adjusted for log-bias. As in Section 2.4.2, the following terms refer to the transformed variables: "number regression prediction," and "rate regression prediction."

As shown in Figures 3.17 and 3.18, the ACS number regression predictions are fairly consistent across years. The points in Figure 3.17 are a little tighter along the diagonal than the points in Figure 3.18, which is expected due to the lower sampling error variances of the 2005 ACS estimates, and the resulting lower variances of the 2005 ACS model regression predictions. Comparing Figures 3.17 and 3.18 to the corresponding plots of ACS direct survey estimates of number in poverty (Figures 3.2 and 3.3), the data also characteristically show much less year-to-year variation in the ACS number regression predictions than in the ACS direct survey estimates. Also, as with Figures 3.2 and 3.3, when Figures 3.17 and 3.18 are plotted with C2SS through the 2003 ACS data, the results appear very similar to those below.



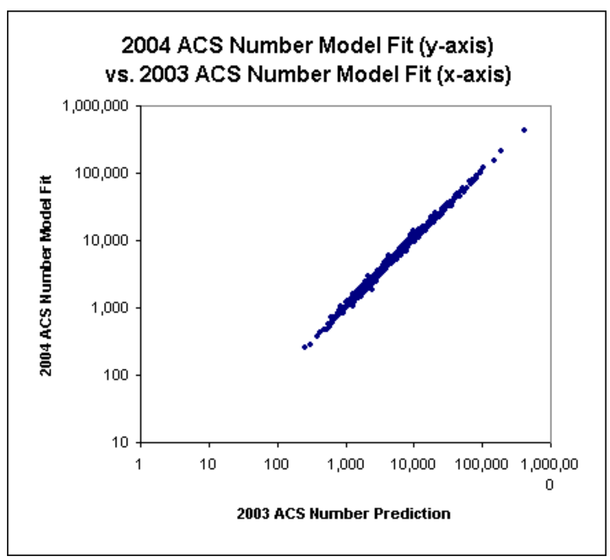
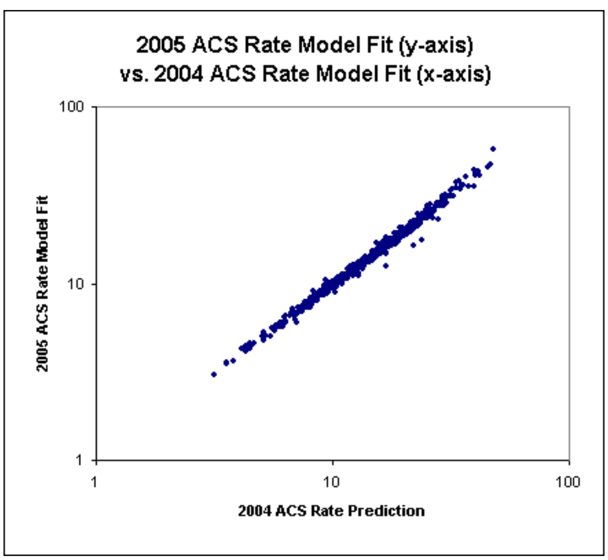


Figure 3.17 (left): 2005 ACS number regression predictions against 2004 ACS number regression predictions.
Figure 3.18 (right): 2004 ACS number regression predictions against 2003 ACS number regression predictions.

Special denominators are computed to transform the ACS log-level regression predictions from predicted poverty numbers to predicted poverty rates. The denominators used are the household population estimates from the Census Bureau's Population Estimates Program (PEP). The reason household population is used instead of total resident population is that the universe for the 2005 ACS is the household population.[38] These household population estimates are then adjusted to reflect the poverty universe of the household population; i.e., multiplied by the ratio of the number of children ages 5 - 17 in the Census 2000 poverty universe to the number of all children ages 5 - 17 in Census 2000. These county-level estimated poverty universes are then raked to the national ACS estimated poverty universe (though the poverty estimates are not raked to the national ACS poverty estimate). The predicted rates are computed by dividing the predicted numbers (unraked) by these raked poverty universe estimates. These predicted rates are denoted "rate regression predictions." Note that, while the rate regression predictions use PEP intercensal population estimates in the denominators, the rate survey estimates use survey estimates in both the numerator and denominator.

Figures 3.19 and 3.20 are analogous to Figures 3.17 and 3.18, except that they display rate regression predictions instead of number regression predictions. The dispersion in Figure 3.19 is, as expected, lower than that in Figure 3.20 due to the lower sampling error variances of the 2005 ACS estimates. The dispersion of both the ACS number and rate regression predictions is greatly reduced from the dispersion observed in the ACS number survey estimates and rate survey estimates in Figures 3.4 to 3.5. Still, the year-to-year dispersion in rate regression predictions in Figures 3.19 and 3.20 is somewhat larger than the year-to-year dispersion in number regression predictions in Figures 3.17 and 3.18, as expected due to the rate transformation. Also, as with Figures 3.4 and 3.5, when Figures 3.19 and 3.20 are plotted with C2SS through the 2003 ACS data, the results appear very similar to those below.



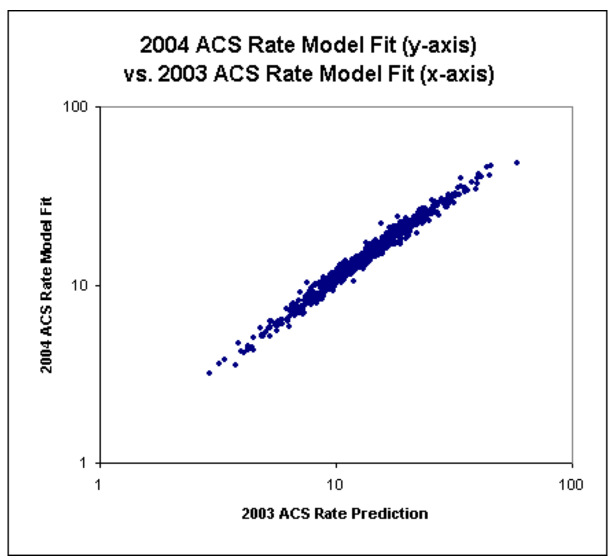
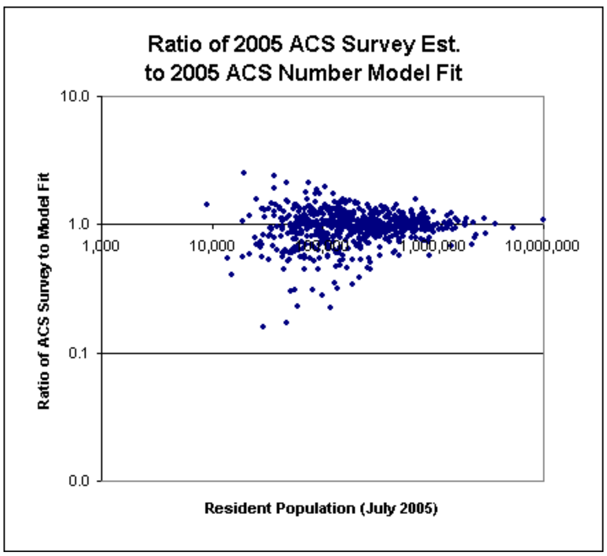


Figure 3.19 (left): 2005 ACS rate regression predictions against 2004 ACS rate regression predictions.
Figure 3.20 (right): 2004 ACS rate regression predictions against 2003 ACS rate regression predictions.

[38] Group quarters (GQ) were included in the ACS starting with the 2006 survey.

These regression predictions do not involve raking to the state level, so the comparisons of the ACS demonstration survey estimates and the corresponding ACS predictions of poverty do not provide perfect indications of the nature of the results with the full production ACS. Raking[39] was not done for the regression predictions from the ACS demonstration surveys since their direct ACS state estimates are less precise than the full-production direct ACS state estimates.

The ACS number survey estimates and number regression predictions can also be compared by computing their ratios, as shown in Figures 3.21 and 3.22. Figure 3.21 shows the ratios of 2005 ACS number survey estimates to 2005 ACS number regression predictions against 2005 PEP county total resident population. Figure 3.22 shows the same for the prior year, 2004. As expected, there is much less dispersion in Figure 3.21 for the 2005 ACS than in Figure 3.22 for the 2004 ACS due to the lower sampling error variances of the 2005 ACS survey estimates. There is a group of counties in Figure 3.21 with low survey estimates relative to the regression predictions in the ratio of the 2004 ACS data. This feature was examined in Chapter 2 though the numerator and denominator were flipped in that analysis. It appears that some mid-sized ACS counties have ACS estimates that are low relative to other data sources such as Census 2000, tallies of IRS tax returns, and the Food Stamp Program.



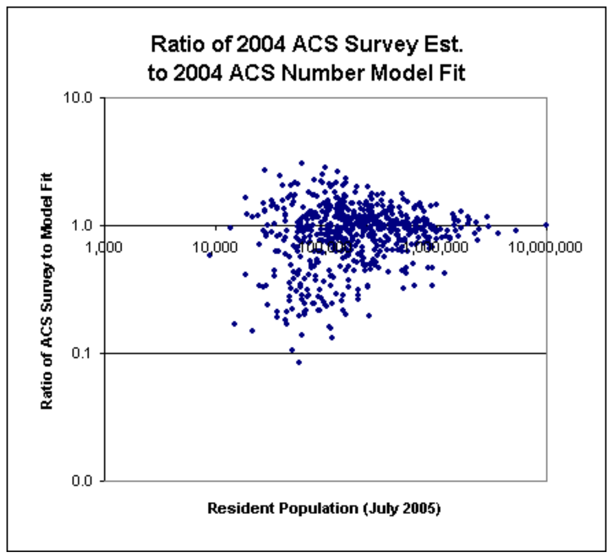
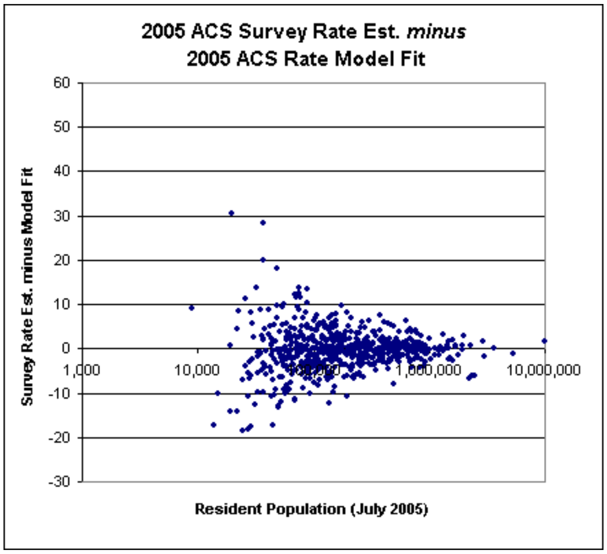


Figure 3.21 (left): Ratios of 2005 ACS number survey estimates to 2005 ACS number regression predictions against 2005 PEP county total resident population.
Figure 3.22 (right): Ratios of 2004 ACS number survey estimates to 2004 ACS number regression predictions against 2005 PEP county total resident population.

The ACS survey estimates and number predictions can also be compared in rates form, as in Figures 3.23 and 3.24. Figure 3.23 shows the difference between 2005 ACS rate survey estimates and 2005 ACS rate regression predictions against 2005 PEP county total resident population, and Figure 3.24 shows the same for the prior year, 2004. Although the dispersion for the 2004 figure is again greater than that in the 2005 figure, there are fewer downside outliers in

[39] In this case, raking would mean summing, by state, the number regression predictions for the 621 counties and the 2,520 out-of-sample counties, and scaling these state-level sums to match the state-level ACS survey estimates or the state-level model-based estimates for the given year.

Figures 3.23 and 3.24 than in Figures 3.21 and 3.22. This can happen due to correlation in the numerator and denominator of the rate survey estimates. Even if a given number survey estimate is low, the survey poverty universe denominator used in computing the rate survey estimate may also be low, so that taking the ratio of the two still produces a moderate rate survey estimate.



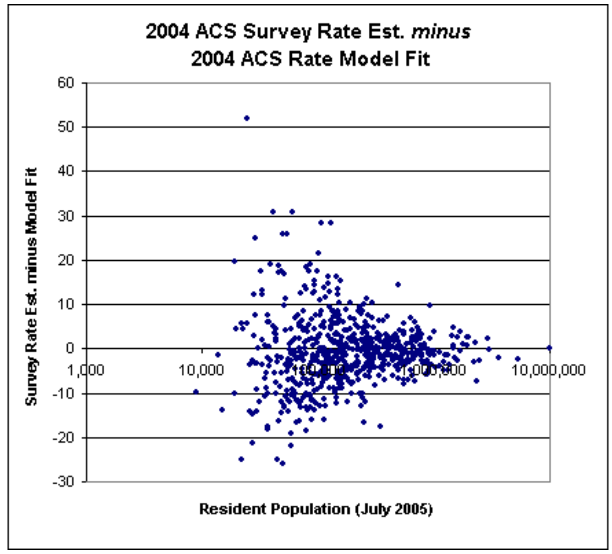


Figure 3.23 (left): 2005 ACS rate survey estimates minus 2005 ACS rate regression predictions against 2005 PEP county total resident population.
Figure 3.24 (right): 2004 ACS rate survey estimates minus 2004 ACS rate regression predictions against 2005 PEP county total resident population.

3.3 SOME RESULTS FROM MODELING 2006 ACS DATA

The 2006 ACS was the second year of full-production ACS. Current availability of the 2006 ACS direct survey estimates allows an initial look at the year-to-year stability of the full-production ACS modeling results by comparison to modeling results reported in Chapter 2 for the 2005 ACS.

Tables 3.3 and 3.4 display results of the model estimation for 2005 and 2006 ACS side-by-side for the six-variable log-level and log-rate models, respectively. There are slightly fewer counties with zero estimated poverty in 2006 than in the 2005 ACS. The timing of the inputs is the same as described in Sections 2.2 and 2.3, with food stamps from the year prior to the ACS survey, data from tax filings coincident with the ACS survey year, and the population estimates matching those used for the survey population controls.

Overall, for both the log-level and log-rate models, the estimation results are relatively comparable between the two years when one makes allowance for the standard errors of the estimated coefficients.[40] Estimates of the model error variance for the two years are also not

[40] Confidence intervals for the difference of two coefficients across years are readily constructed assuming independence of the results for the two years (and we mention a couple such results here). The independence assumption most likely does not hold for the model errors, but we do not have an estimate of their year-to-year correlation. The independence assumption is more palatable for the sampling errors in the ACS survey estimates,

dramatically different. One notable difference between the 2005 and 2006 ACS models is that for the log-level model, the coefficients on the filing-rate variables, log (PEP population, ages 0 - 17) and log (IRS child tax exemptions) are much smaller in 2006 than in 2005. A 90 percent confidence interval for the difference between these coefficient estimates for 2006 and 2005 is (assuming independence across the two years) −.57 to −.02, so the difference between these two years of coefficient estimates is marginally statistically significant.

Other model evaluations and comparisons show some differences between the 2005 and 2006 ACS data. In 2006, the Spearman's rank correlation coefficients of the squared standardized residuals indicate significant variance heterogeneity remaining in the standardized residuals. Also, hypothesis tests rejected the models with six regressors in favor of the models with eight regressors. Both these results may be related to increased dispersion in population size in 2006, and further research on the 2006 regressors will be required.

Table 3.3: Log-level model (LL6u) – 2005 ACS and 2006 ACS estimation

Variable	Log (2005 ACS poverty)			Log (2006 ACS poverty)		
	Regression Coefficient	Standard Error	t	Regression Coefficient	Standard Error	t
Intercept	-0.421	0.057	-7.39	-0.381	0.057	-6.73
Log (IRS child tax-poor exemptions)	0.548	0.045	12.26	0.614	0.044	13.86
Log (Food Stamps)	0.173	0.022	7.90	0.138	0.022	6.34
Log (PEP population, ages 0-17)	1.050	0.122	8.62	0.678	0.126	5.40
Log (IRS child tax exemptions)	-1.037	0.114	-9.13	-0.698	0.116	-6.00
Log (Census 2000 poor, ages 5-17 related)	0.268	0.030	8.90	0.270	0.028	9.55
Degrees of freedom	2,966			2,989		
Model error variance	0.022			0.020		
R-squared	0.936			0.938		
Sum of slopes	1.0036			1.0023		

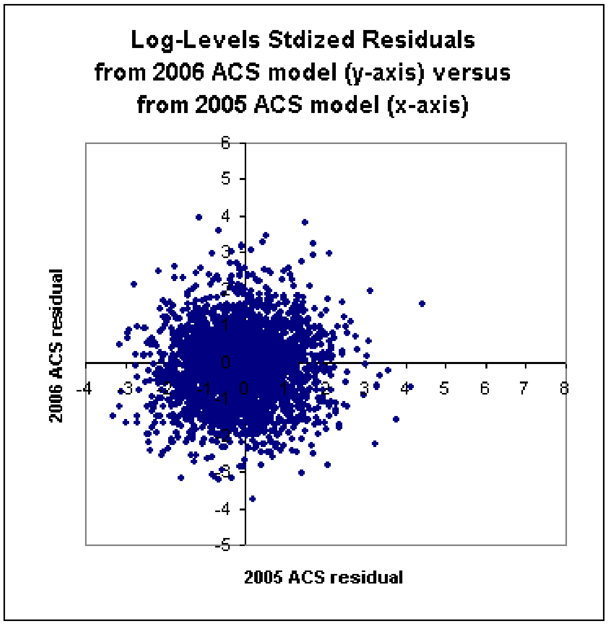
Table 3.4: Log-rate model (LR6u) – 2005 ACS and 2006 ACS estimation

Variable	Log (2005 ACS poverty)			Log (2006 ACS poverty)		
	Regression Coefficient	Standard Error	t	Regression Coefficient	Standard Error	t
Intercept	0.372	0.048	7.73	0.343	0.047	7.37
Log (IRS child tax-poor rate)	0.556	0.049	11.25	0.579	0.050	11.69
Log (Food Stamp rate)	0.167	0.022	7.64	0.133	0.021	6.23
Log (PEP population, ages 0-17)	-0.020	0.005	-3.70	-0.023	0.005	-4.45
Log (IRS child filing rate)	-0.414	0.117	-3.55	-0.121	0.119	-1.01
Log (Census 2000 poverty rate, ages 5-17 related)	0.289	0.035	8.33	0.308	0.033	9.34
Degrees of freedom	2,966			2,989		
Model error variance	0.030			0.025		
R-squared	0.620			0.614		
Sum of denominators	-0.020			-0.023		

---

and since these contribute most of the statistical variability, the confidence intervals constructed assuming independence across years should have some approximate validity.

Residual plots, comparing the standardized residuals from the two years of estimation, as in Figures 3.25 and 3.26, display no persistence in over-estimation or under-estimation over the two years. There appears to be little correlation in the standardized residuals from the 2006 ACS model with those from the 2005 ACS model, for both the log-level and log-rate models.



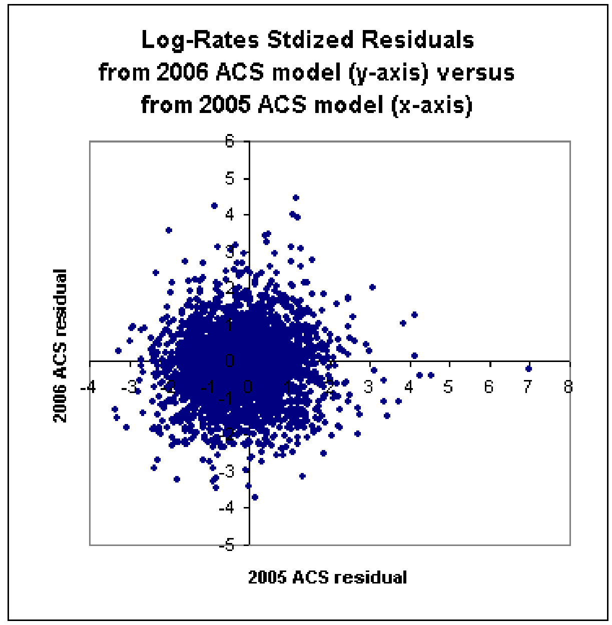


Figure 3.25 (left): Log-level standardized residuals from the 2006 ACS model versus log-level standardized residuals from the 2005 model.

Figure 3.26 (right): Log-rate standardized residuals from the 2006 ACS model versus log-rate standardized residuals from the 2005 model.

Table 3.5 displays the same RMSE results for the 2006 ACS results as reported in Section 2.2 for the 2005 ACS model comparisons. These results are very similar, with the log-level model displaying a lower RMSE for every partition. For example, for the 20k – 65k group, 0.585 is less than 0.597 for comparisons on the log-level scale; and 0.567 is less than 0.579 for comparisons on the log-rate scale.

Table 3.5: 2006 ACS RMSE comparison for 6-regressor log-level (LL6u) and log-rate models (LR6u)

	Comparison on the log-level scale		Comparison on the log-rate scale	
	LL6u Regressors	LR6u Regressors +log(z demog)	LL6u Regressors -log(z demog)	LR6u Regressors
Full-sample: 2,995 counties	0.652	0.665	0.619	0.631
Partition by 2005 PEP county total resident population				
> 65k: 783 counties	0.304	0.307	0.302	0.305
20k – 65k: 1,030 counties	0.585	0.597	0.567	0.579
< 20k: 1,182 counties	0.848	0.865	0.794	0.809
Partition by Census 2000 poverty rate, all ages				
> 20%: 934 counties	0.627	0.645	0.565	0.583
12.5 – 20%: 1,007 counties	0.667	0.677	0.629	0.638
< 12.5: 1,054 counties	0.660	0.671	0.654	0.664

Finally, the shrinkage estimates obtained from the two models yield stable predicted values when compared to the survey estimates. Table 3.6 displays summary statistics for the poverty rate predictions produced by the log-level model for each year[41]. Nearly 90% of the counties (2,793 out of 3,141) had predicted values for the poverty rate that changed less than 2 percentage points on a rate scale. Table 3.7 reports the same distribution for the ACS direct survey estimates.

Table 3.6: Change from 2005 to 2006 ACS-based shrinkage estimates of poverty rate from log-level model

Mean difference from 2005 to 2006 ACS	-0.1%
Mean absolute difference	1.5%
Median absolute difference	1.0%
Number of counties with:	
Absolute difference < 1 percentage point	1,584
1% ≤ Absolute difference < 2%	813
2% ≤ Absolute difference < 3%	395
3% ≤ Absolute difference < 4%	160
4% ≤ Absolute difference < 5%	81
Absolute difference ≥ 5%	108

Table 3.7: Change from 2005 to 2006 ACS survey estimates of poverty rate for counties with non-zero estimates in both years

Mean absolute difference from 2005 to 2006 ACS	9.3%
Median absolute difference	6.1%
Number of counties with:	
Absolute difference < 1 percentage point	308
1% ≤ Absolute difference < 2%	284
2% ≤ Absolute difference < 3%	246
3% ≤ Absolute difference < 4%	225
4% ≤ Absolute difference < 5%	194
Absolute difference ≥ 5%	1,629

Overall, the 2006 ACS modeling results are comparable with those from modeling the 2005 ACS, and the corresponding number and rate regression predictions are fairly stable across the two years. The log-level model again has an RMSE advantage over the log-rate model.

[41] Poverty rates were produced from the log-level shrinkage estimates by first exponentiating the prediction, including a log-bias adjustment, and then converting to a rate using a demographically-estimated poverty universe as a denominator. This process is discussed in detail in Section 2.4.

Appendix A: Reference to Chapter 1

This appendix contains additional quotations of recommendations from the 2000 NAS Panel review of SAIPE. It refers to Section 1.2.

> In its third interim report (National Research Council, 1999), the panel concluded that although the Census Bureau's 1995 estimates of poverty school-age children had potentially large errors for many school districts, the estimates were nonetheless not inappropriate or unreliable to use for direct Title I allocations to districts as intended by the 1994 legislation. In reaching this conclusion, the panel interpreted "inappropriate and unreliable" in a relative sense. Some set of estimates must be used to distribute Title I funds to school districts. The panel concluded that the Census Bureau's estimates were generally as good as—and, in some instances, better than—estimates that were previously used. On the basis of the panel's study, the Department of Education made direct allocations to school districts for the 1999-2000 and 2000-2001 school years by using the Census Bureau's 1995 school district estimates and other elements of the allocation formula. The department also notified the states of a recommendation by the panel that states electing to reallocate amounts for school districts with fewer than 20,000 people on the basis of some other data source (e.g., school lunch data) should do so on a county-by-county basis so as to reflect (approximately) the Census Bureau's updated estimates of in poverty school-age children from the county model (National Research Council, 2000a, p. 7).
>
> Research is needed to take account of likely future developments in the availability and characteristics of data sources that have implications for the modeling effort and to work on longer term modeling issues. Continued work to improve the county model is important not only for county estimates, but also to improve school district estimates that are developed by using the within-county shares estimation procedure (National Research Council, 2000a, p. 157).
>
> Research and development by the Census Bureau should begin now to explore two possible uses for the ACS in SAIPE models for counties. One use is for ACS estimates to form one of the predictor variables in regression models for which the official source of income and poverty estimates, the March CPS, continues to provide the dependent variable. Another use for ACS estimates to serve as the dependent variable in county models, which could thereby include all, or nearly all, counties in the estimation (National Research Council, 2000b, p. 7). For the latter use the ACS estimates might possibly be calibrated in some way to selected estimates from the March CPS (National Research Council, 2000b, p. 123). The Census Bureau should also conduct research on using ACS estimates for school districts and other subcounty areas to form within-county shares or proportions to apply to updated county model poverty estimates (National Research Council, 2000b, p. 7).

Even when it is possible to produce direct survey estimates by averaging over several years or months, as is sometimes done for states from the Current Population Survey and as is planned for states and smaller areas from the ACS, model-based estimates should be considered by users and may be preferred (National Research Council, 2000b, p. 162).

Appendix B: Reference to Chapter 2

B.1 LOG-LEVEL AND LOG-RATE RESIDUAL ANALYSIS

This section refers to Sections 2.2.3 and 2.3.3.

The following figures compare the standardized residuals of log-level and log-rate from the 2005 ACS log-level model on income year 2004 log model inputs. In the plots, the y-axis is the standardized residual and the x-axis is one of several classification variables. The 2,972 counties used to fit the regression models are depicted. Each page has four plots with the four variations of the same classification variable. These variations are designated as:

LS=Levels (from LL6u), scatter-plot	LB=Levels (from LL6u), box-plot
RS=Rates (from LR6u), scatter-plot	RB=Rates (from LR6u), box-plot

The upper plots on each page are from the log-level model, and the lower plots are from the log-rate model. The left-hand plots are scatter plots, and the right-hand plots are box-and-whisker plots.

For each box-lot, the bottom and top edges of the box are located at the sample 25th and 75th percentiles. The center horizontal line is drawn at the 50th percentile (median). The vertical lines (or whiskers) are drawn from the box to the most extreme point within 1.5 interquartile ranges. (An interquartile range is the distance between the 25th and the 75th sample percentiles.) Any value more extreme than this is marked with a square. The x-axis in the box-plots shows county quintiles (each with about 594 of the 2,972 counties) from a sort of the counties by the particular classification variable.

<u>Note:</u> In this section of Appendix B, the x-axis is shown in log scale but labeled in linear scale.

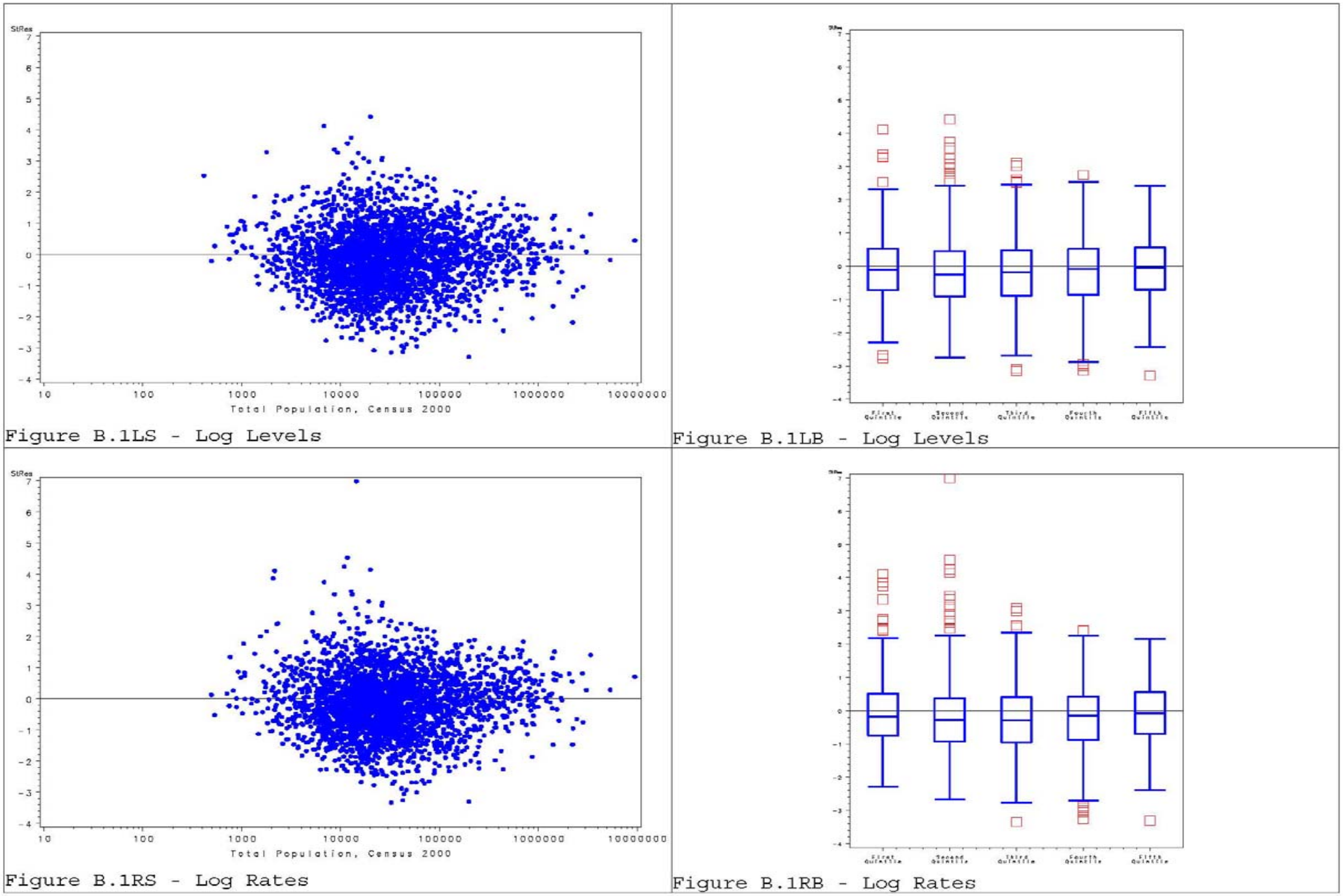


Figure B.1: Standardized residuals against Census 2000 county total resident population.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

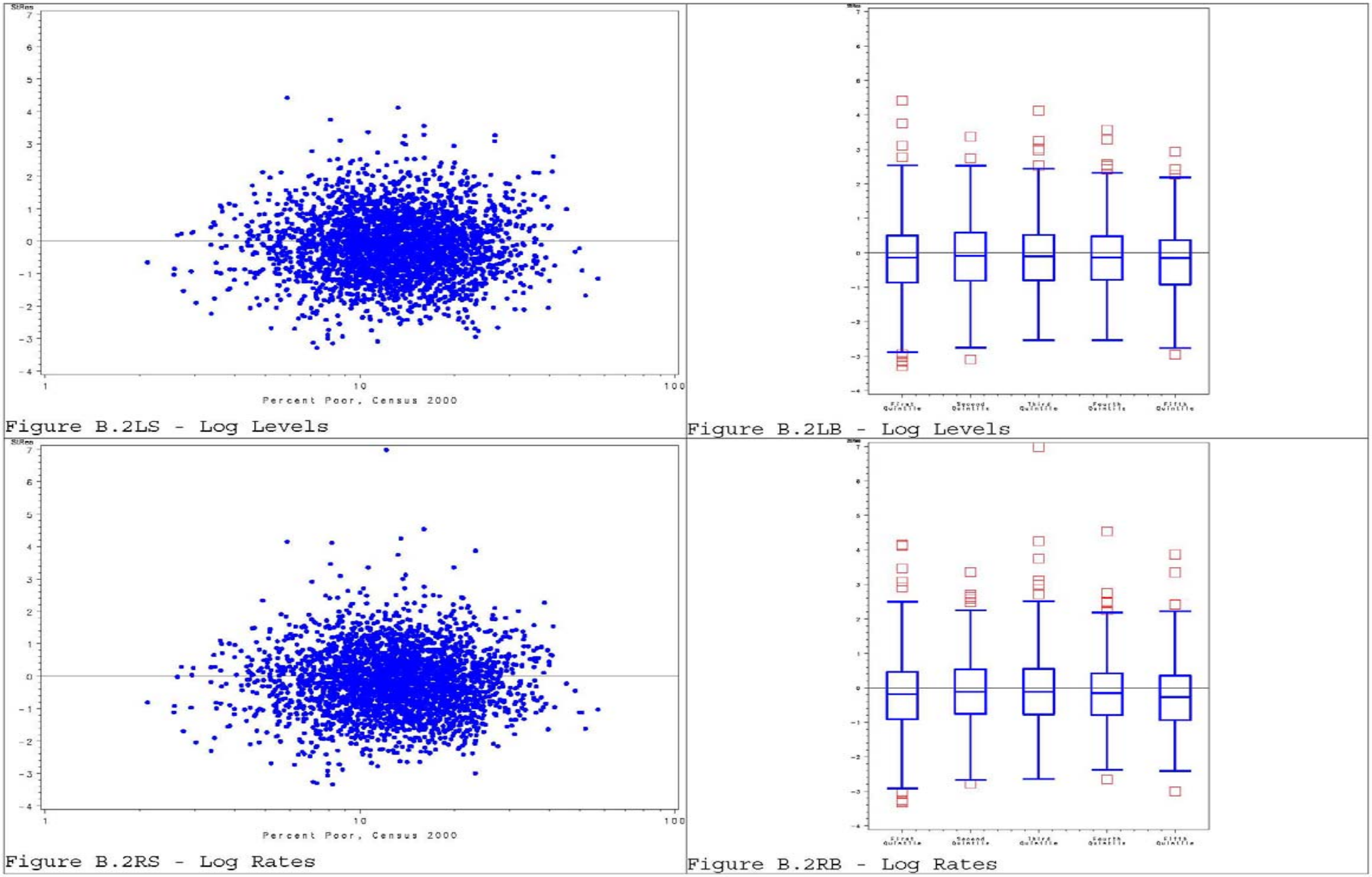


Figure B.2: Standardized residuals against Census 2000 percent poor, all ages.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

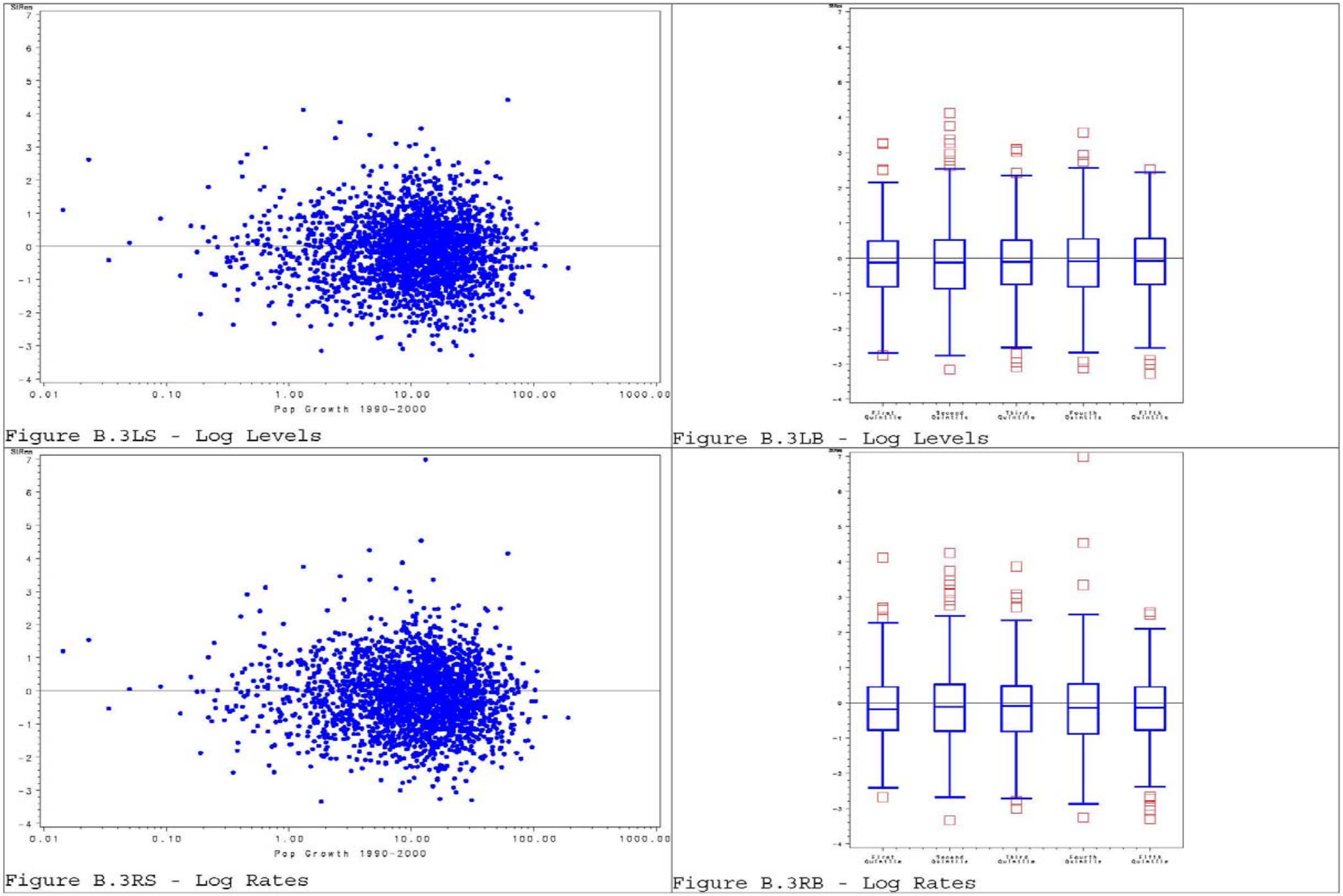


Figure B.3: Standardized residuals against population growth Census 1990 – Census 2000.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

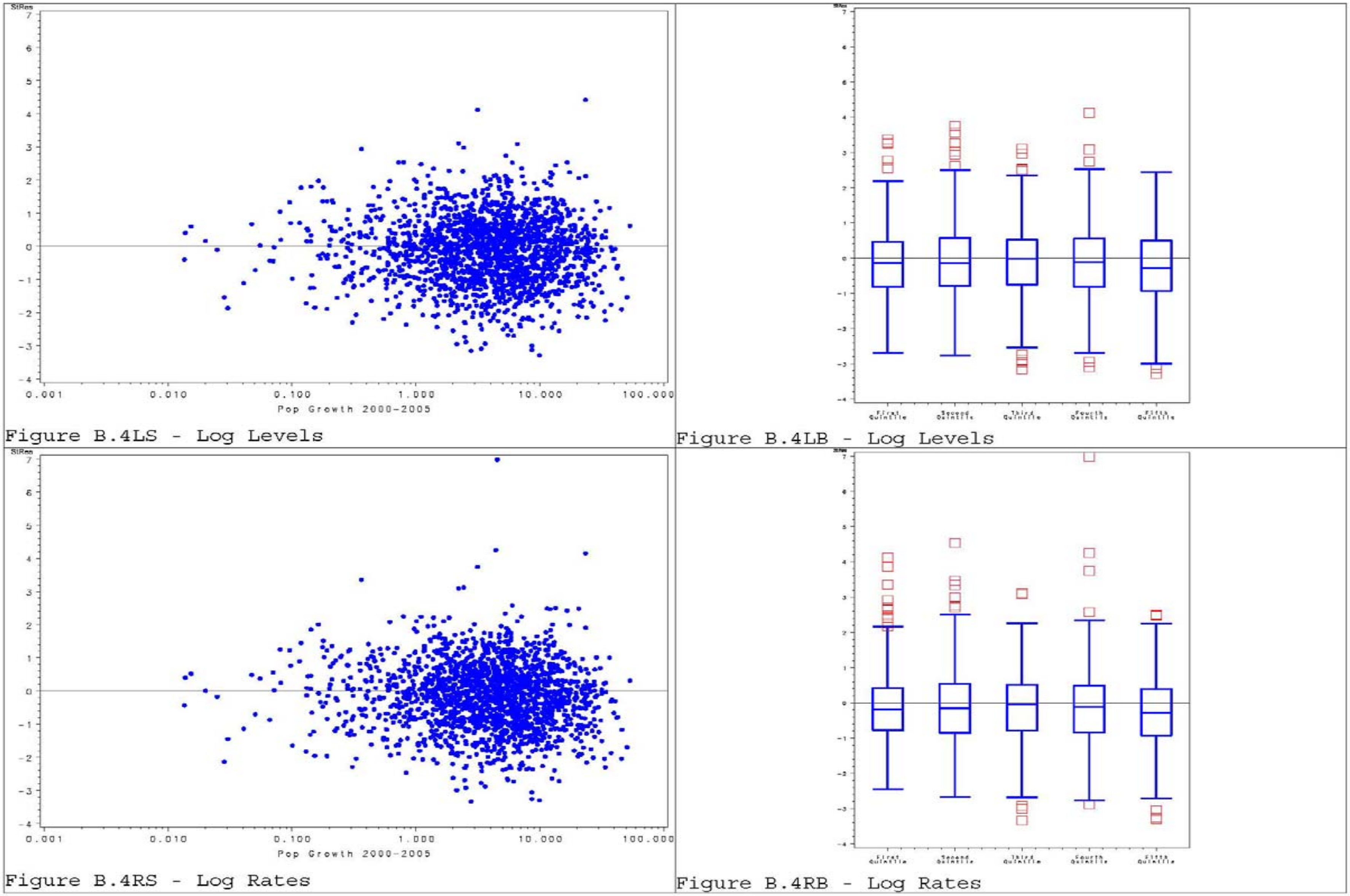


Figure B.4: Standardized residuals against population growth Census 2000 – 2005 PEP county total resident population.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

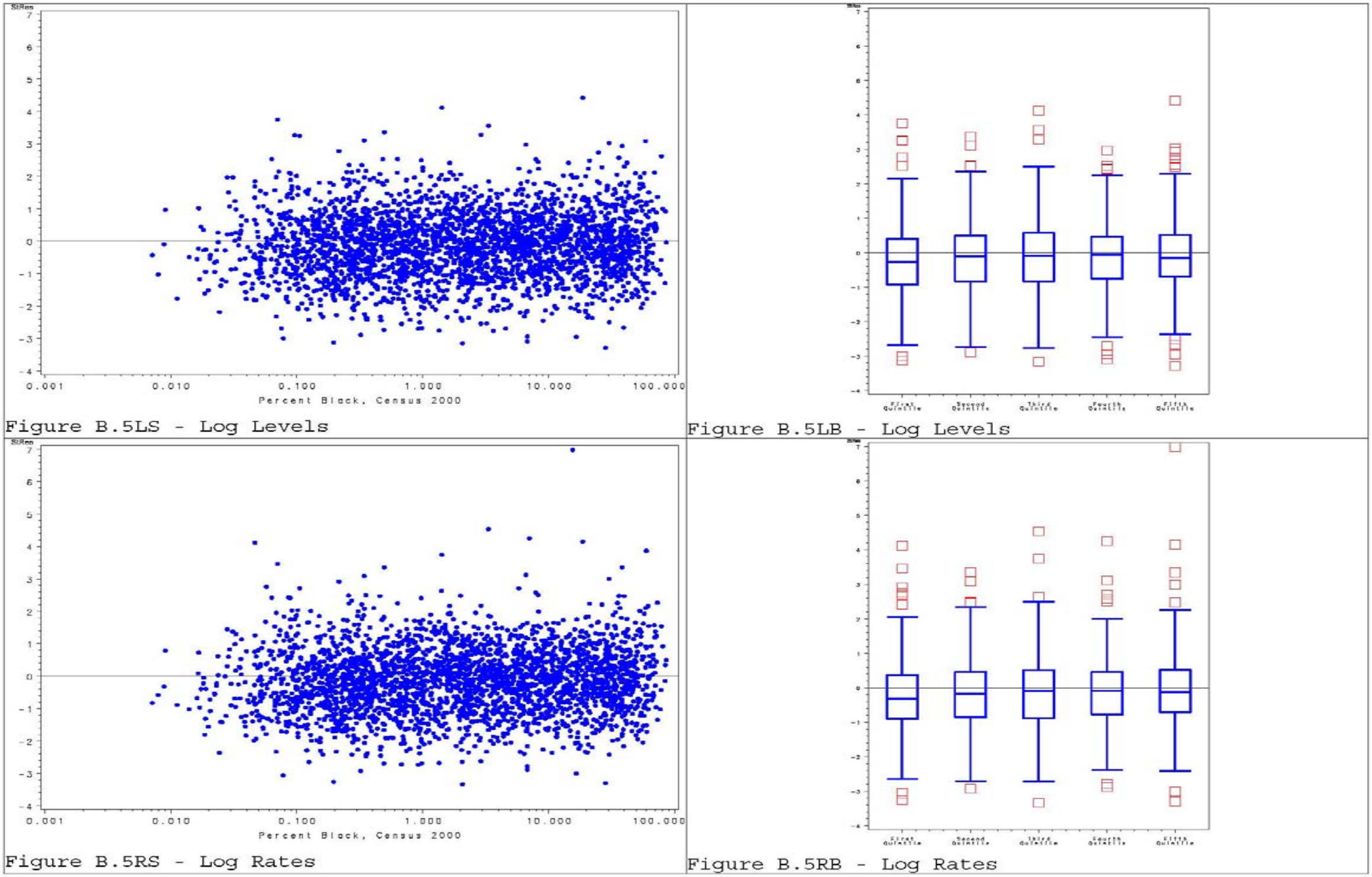


Figure B.5: Standardized residuals against Census 2000 percent Black.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

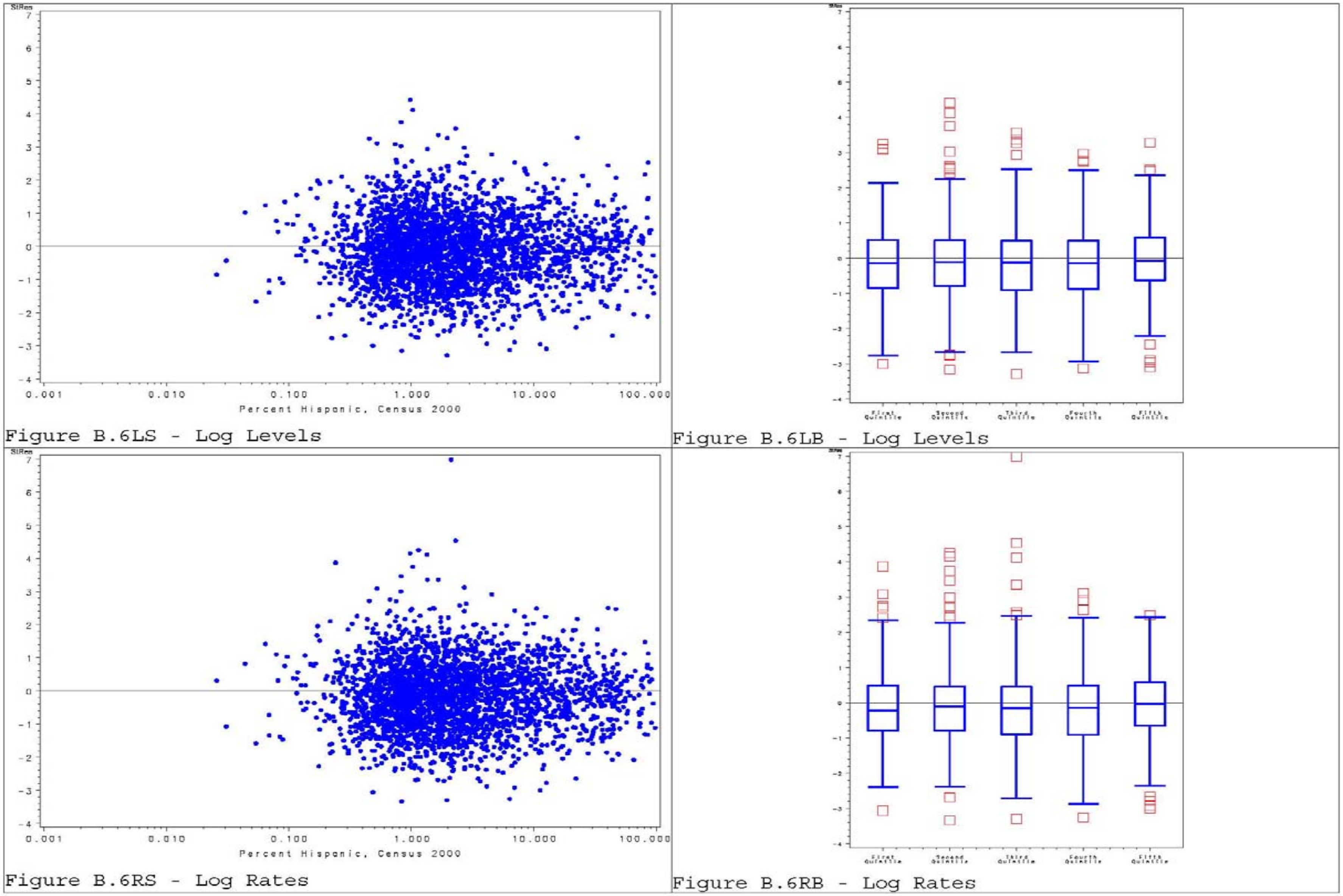


Figure B.6: Standardized residuals against Census 2000 percent Hispanic.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

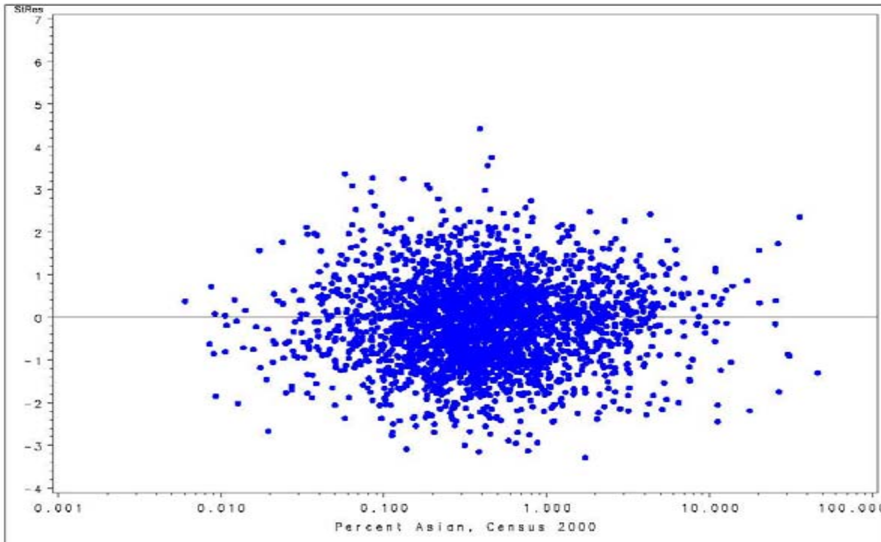


Figure B.7LS - Log Levels

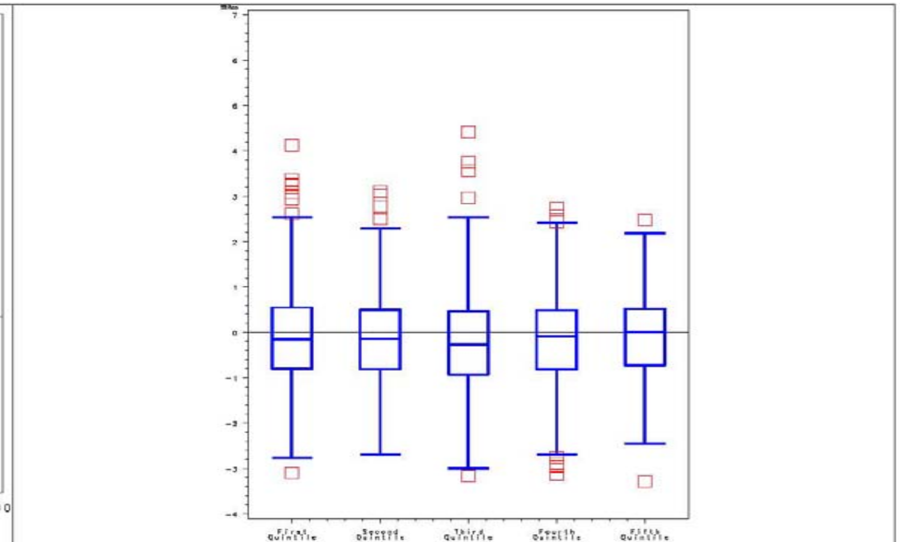


Figure B.7LB - Log Levels

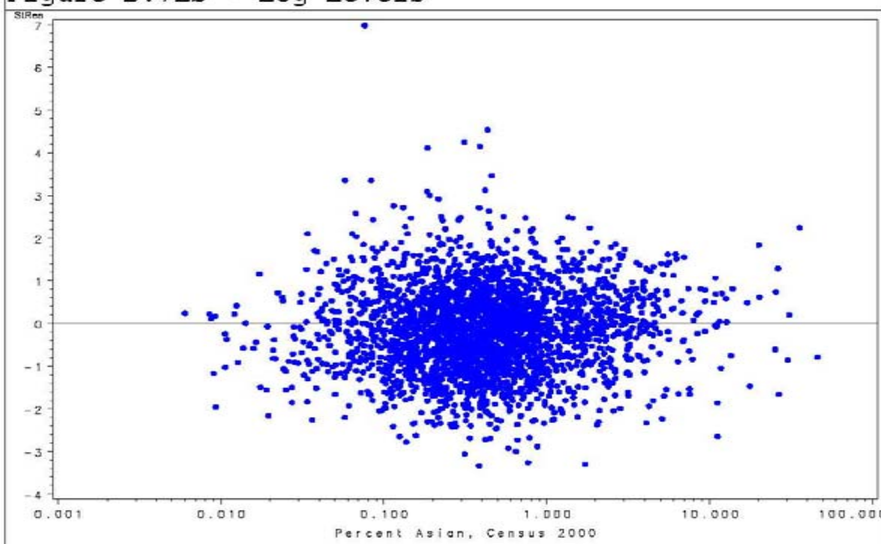


Figure B.7RS - Log Rates

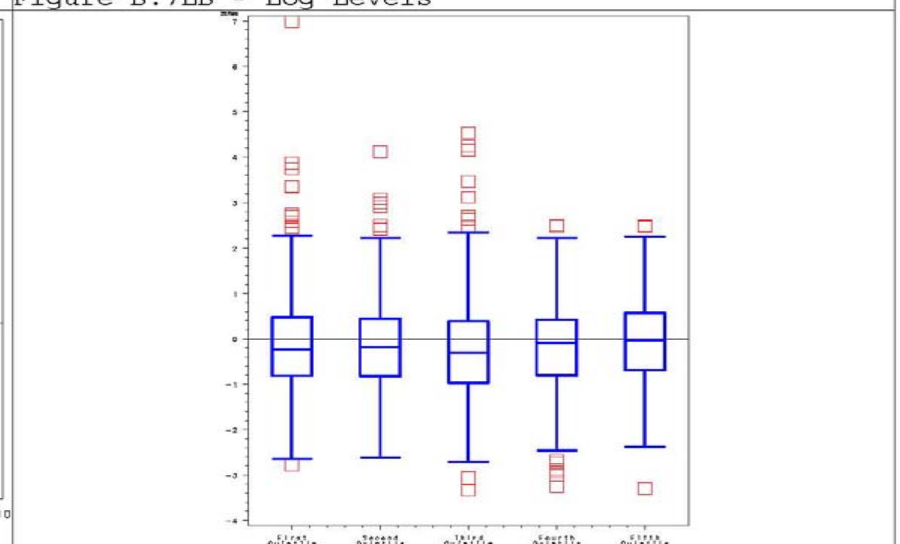


Figure B.7RB - Log Rates

Figure B.7: Standardized residuals against Census 2000 percent Asian.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

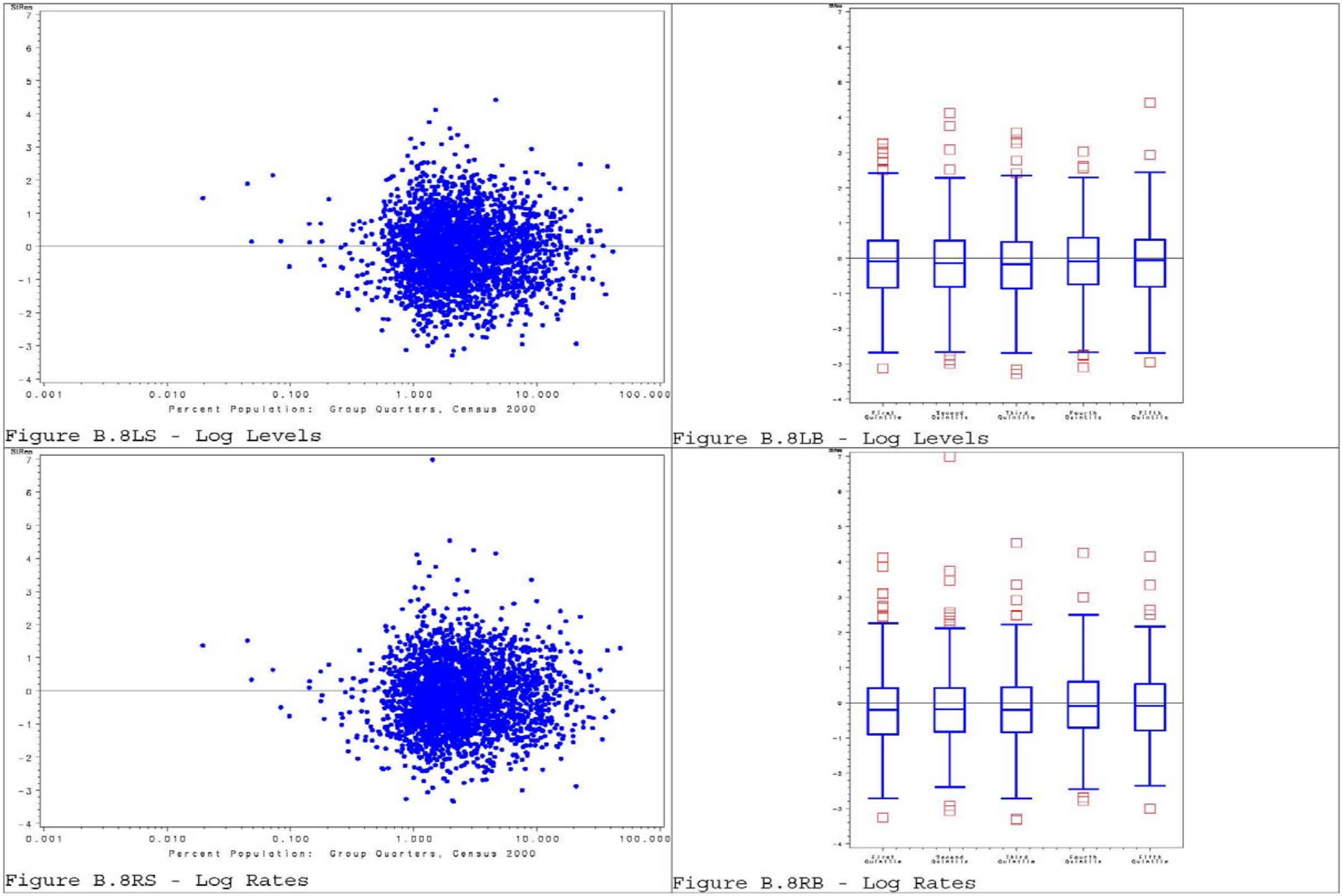


Figure B.8: Standardized residuals against percent Census 2000 group quarters' population.
The top two plots are from the log-level (LL6u) model, and the bottom two plots are from the log-rate (LR6u) model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 594-county quintile.

B.2 LOG-LEVEL VS. LOG-RATE MODELS

This section refers to Section 2.2.5.

B.2.1 Reparameterization

Given that log-rate is a simple linear combination of log(numerator) and log(denominator) terms, the change from log-level poverty indicators to log-rate indicators in the right-hand side matrix can be analyzed as a linear transformation. A general result is that any linear reshuffling of the columns of the right-hand side matrix will result in the same predicted values and residuals; providing the transformation is full rank (i.e. invertible).

Consider the general linear model:

$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u}$ where **X** is a n x k matrix including a constant column.

Define a linear transformation matrix, **A** k x k, with full rank such that the inverse exists.

Then consider the general model:

$$\mathbf{Y} = \mathbf{X}\beta + \mathbf{u} \qquad \text{where} \quad \mathbf{u} \mid \mathbf{X} \sim (\mathbf{0}, \Sigma) \tag{1}$$

Applying the transformation to the right-hand side matrix:

$$\mathbf{Y} = \mathbf{X}\mathbf{A}\mathbf{A}^{-1}\beta + \mathbf{u} = \mathbf{X}_A\beta_A + \mathbf{u}\,, \quad \mathbf{u} \mid \mathbf{X}_A \sim \mathbf{u} \mid \mathbf{X} \sim (\mathbf{0}, \Sigma)$$

The transformation has no effect on the conditional distribution or error term, but does represent a redefinition of the parameters.

Any estimators based on the cross-correlation matrices, such as ordinary least squares, weighted least squares, maximum likelihood estimation under normality, etc., will be similarly affected only through this linear transformation. Weighted least squares for example,

$$\begin{aligned}\hat{\hat{\beta}}_{WLS} &= \left(\mathbf{X}_A'\Sigma^{-1}\mathbf{X}_A\right)^{-1}\mathbf{X}_A'\Sigma^{-1}\mathbf{Y} = \left(\mathbf{A}'\mathbf{X}'\Sigma^{-1}\mathbf{X}\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \mathbf{A}^{-1}\left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{A}'^{-1}\mathbf{A}'\mathbf{X}'\Sigma^{-1}\mathbf{Y} \\ &= \mathbf{A}^{-1}\left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \mathbf{A}^{-1}\hat{\beta}_{WLS}\end{aligned}$$

And the predicted values and residuals will be identical,

$$\begin{aligned}\hat{\hat{\mathbf{Y}}}_{WLS} &= \mathbf{X}_A\left(\mathbf{X}_A'\Sigma^{-1}\mathbf{X}_A\right)^{-1}\mathbf{X}_A'\Sigma^{-1}\mathbf{Y} = \mathbf{X}\mathbf{A}\left(\mathbf{A}'\mathbf{X}'\Sigma^{-1}\mathbf{X}\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \mathbf{X}\mathbf{A}\mathbf{A}^{-1}\left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{A}'^{-1}\mathbf{A}'\mathbf{X}'\Sigma^{-1}\mathbf{Y} \\ &= \mathbf{X}\left(\mathbf{X}'\Sigma^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\Sigma^{-1}\mathbf{Y} = \hat{\mathbf{Y}}_{WLS}\end{aligned}$$

B.2.2 Scale invariance

Given P_B = m* P_A , all denominators scaled by the same multiple, and poverty rate indicators unchanged (so the numerators are scaled by the same multiple), what conditions ensure the predicted poverty level is also scaled by the same multiple.

The condition can be stated as:

Given $P_1 = mP_0$, $\mathbf{X}_{R,1} = \mathbf{X}_{R,0}$, $\mathbf{X}_{L,1} = m\mathbf{X}_{L,0}$, $\mathbf{D}_1 = m\mathbf{D}_0$, then
$\left(\overline{\overline{\log(Y/Z)}}\right)_1 = \left(\overline{\overline{\log(Y/Z)}}\right)_0$ for log-rate model[42],
or $\left(\overline{\overline{\log(Y)}}\right)_1 = \left(\overline{\overline{\log(Y)}}\right)_0 + \log(m)$ for log-level model, ignoring the transformation bias adjustments.

For log-rate model, parameterized with log-rate on the right-hand side,

$$\left(\overline{\overline{\log(Y/Z)}}\right)_1 = \left(\overline{\overline{\log(Y/Z)}}\right)_0 \Leftrightarrow \alpha_B + \log(\mathbf{X}_R)\delta_B + \log(m\mathbf{D})\gamma_B^* = \alpha_B + \log(\mathbf{X}_R)\delta_B + \log(\mathbf{D})\gamma_B^*$$

$$\Leftrightarrow \log(m)\gamma_B^* + \log(\mathbf{D})\gamma_B^* = \log(\mathbf{D})\gamma_B^* \Leftrightarrow \log(m)\gamma_B^* = 0 \ \forall m \Leftrightarrow \sum_{k=1}^{K_2} \gamma_{Bk}^* = 0$$

For log-level model,

$$\left(\overline{\overline{\log(Y)}}\right)_1 = \left(\overline{\overline{\log(Y)}}\right)_0 + \log(m)$$
$$\Leftrightarrow \alpha_A + \log(m\mathbf{X}_L)\delta_A + \log(m\mathbf{D})\gamma_A = \alpha_A + \log(\mathbf{X}_l)\delta_A + \log(\mathbf{D})\gamma_A + \log(m)$$
$$\Leftrightarrow \log(m)\delta_A + \log(\mathbf{X}_L)\delta_A + \log(m)\gamma_A + \log(\mathbf{D})\delta_A = \log(\mathbf{X}_L)\delta_A + \log(\mathbf{D})\gamma_A + \log(m)$$
$$\Leftrightarrow \log(m)\delta_A + \log(m)\gamma_A = \log(m) \ \forall m \Leftrightarrow \sum_{k=1}^{K_1}\delta_{Ak} + \sum_{k=1}^{K_2}\gamma_{Ak} = 1$$

[42] Double bars are used to indicate estimates in this section.

Appendix C: Reference to Chapter 3

This appendix refers to Section 3.2 and further compares the standardized residuals from the 2004 and 2005 ACS models based on the 621-county subset. In the plots, the y-axis is the standardized residual and the x-axis is one of several classification variables. Each page has four plots with the four variations of the same classification variable. These variations are designated as:

05S=2005 ACS, scatter-plot	05B=2005 ACS, box-plot
04S=2004 ACS, scatter-plot	04B=2004 ACS, box-plot

The upper plots on each page are from the log-level model, and the lower plots are from the log-rate model. The left-hand plots are scatter plots, and the right-hand plots are box-and-whisker plots.

For each box-plot, the bottom and top edges of the box are located at the sample 25th and 75th percentiles. The center horizontal line is drawn at the 50th percentile (median). The vertical lines, or whiskers, are drawn from the box to the most extreme point within 1.5 interquartile ranges. (An interquartile range is the distance between the 25th and the 75th sample percentiles.) Any value more extreme than this is marked with a square. The x-axis in the box-plots shows county quintiles (each with about 124 of the 621 counties) from a sort of the counties by the particular classification variable.

<u>Note:</u> For all figures in this Appendix, the x-axis is shown in log scale but labeled in linear scale.

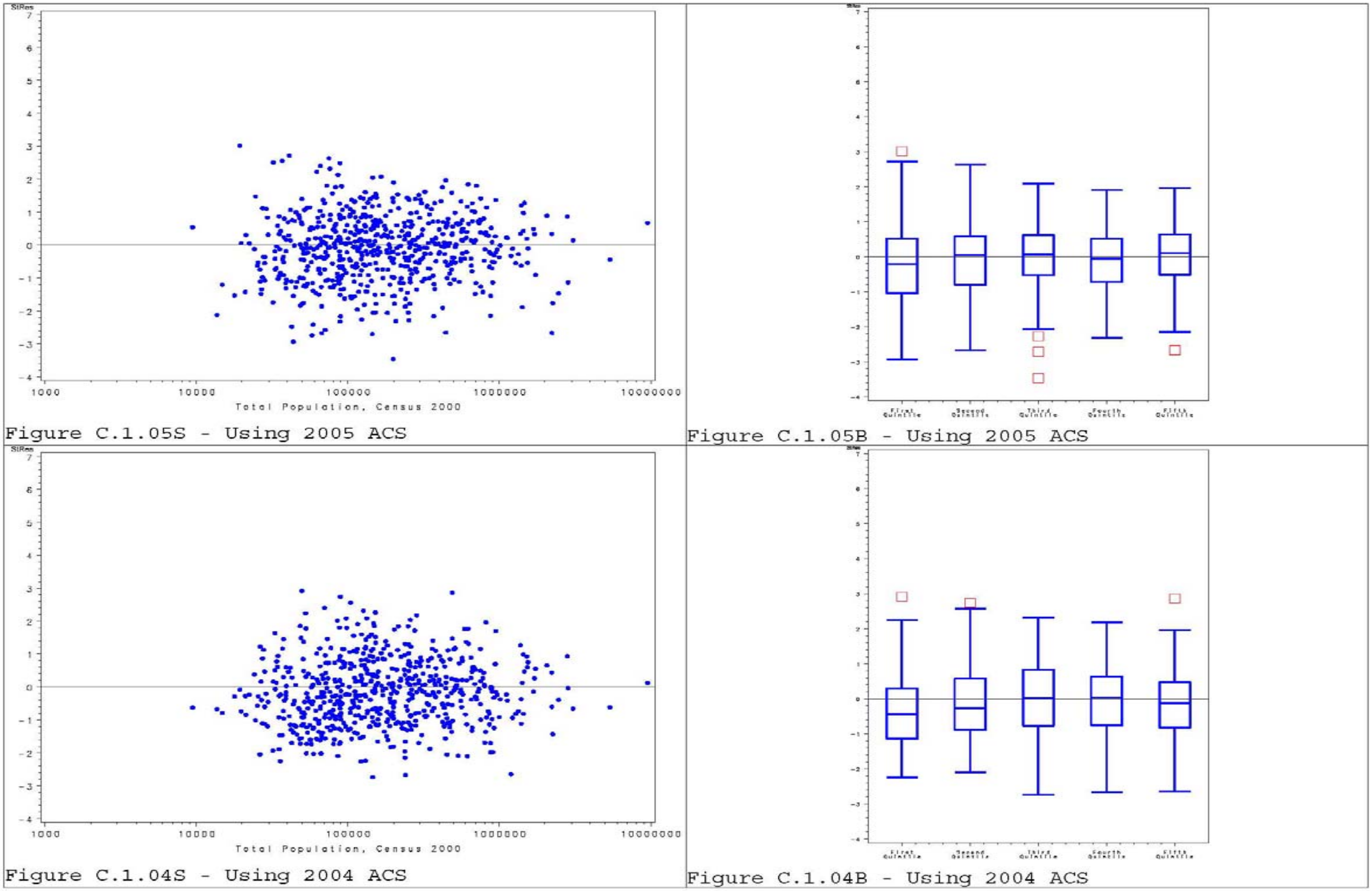


Figure C.1: Standardized residuals against Census 2000 county total resident population.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

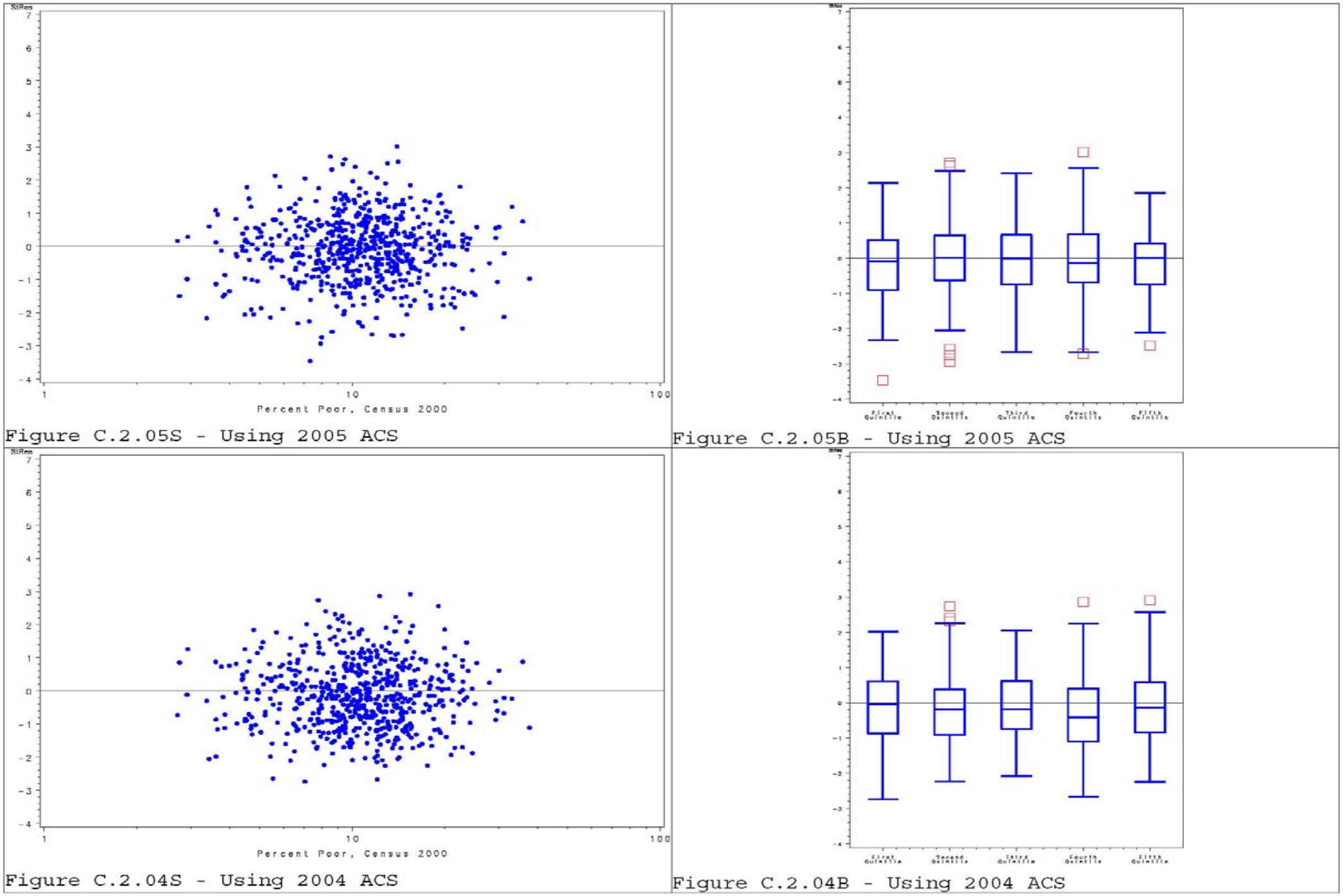


Figure C.2: Standardized residuals against Census 2000 percent poor, all ages.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

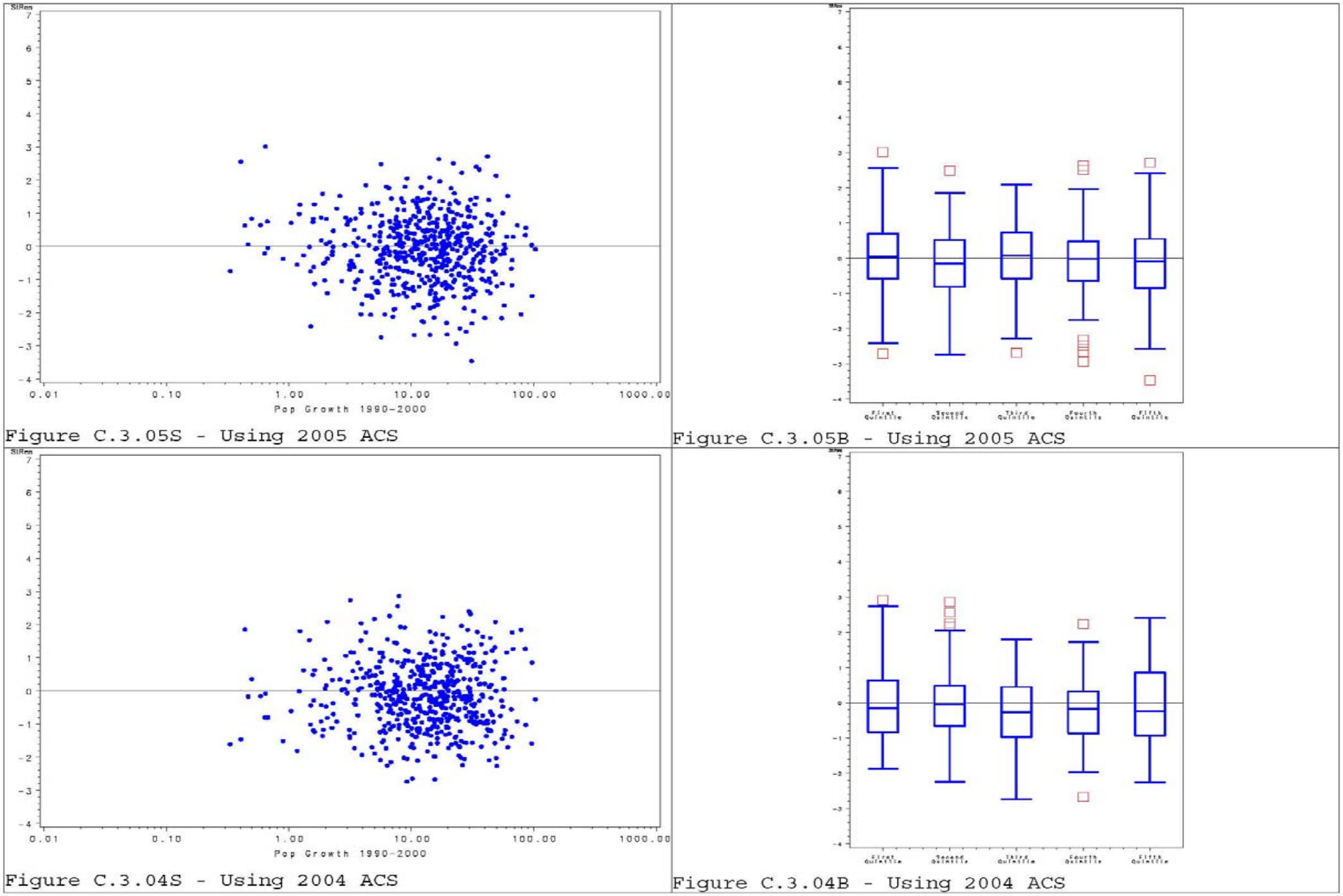


Figure C.3: Standardized residuals against population growth Census 1990 – Census 2000.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

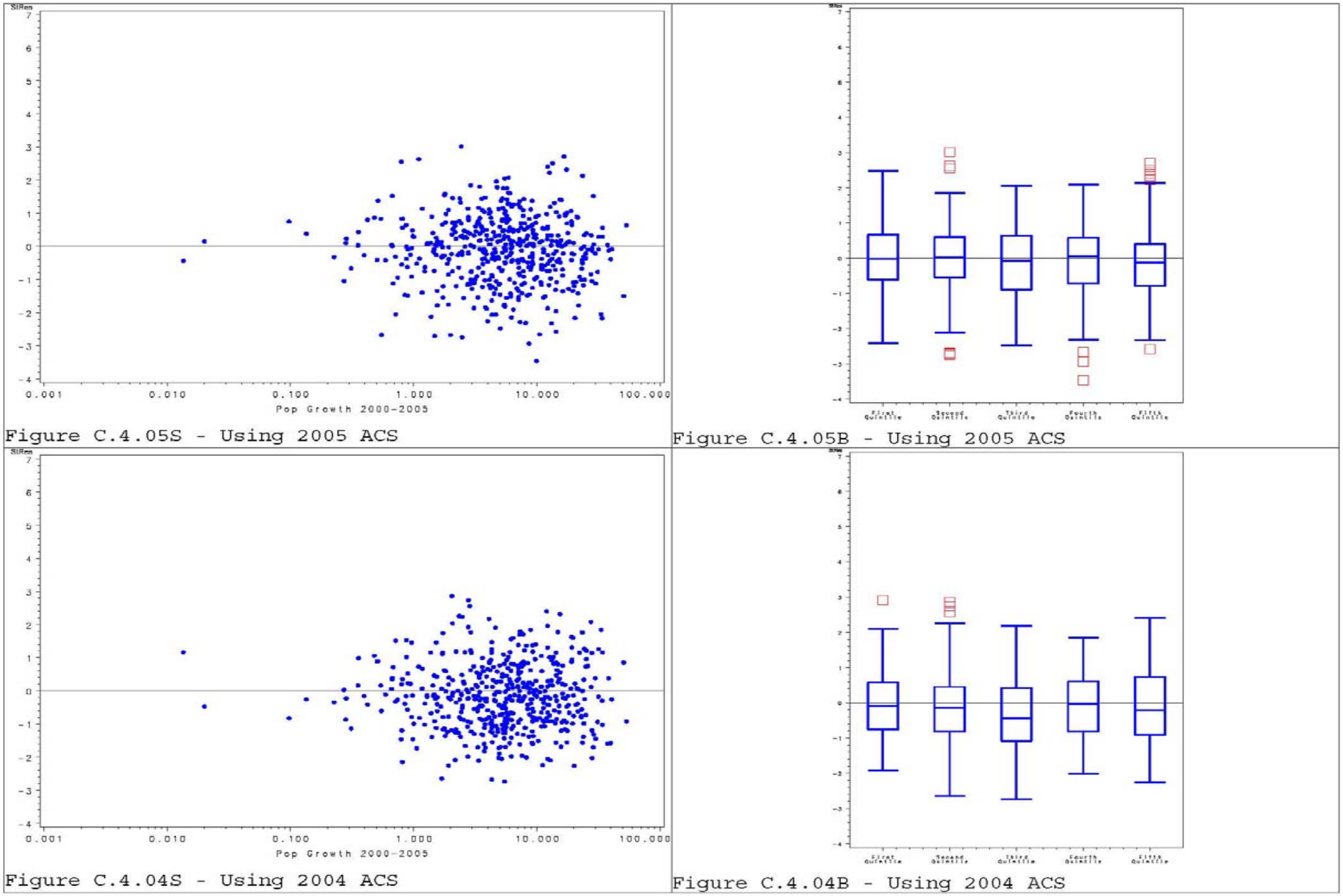


Figure C.4: Standardized residuals against population growth Census 2000 – 2005 PEP county total resident population.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

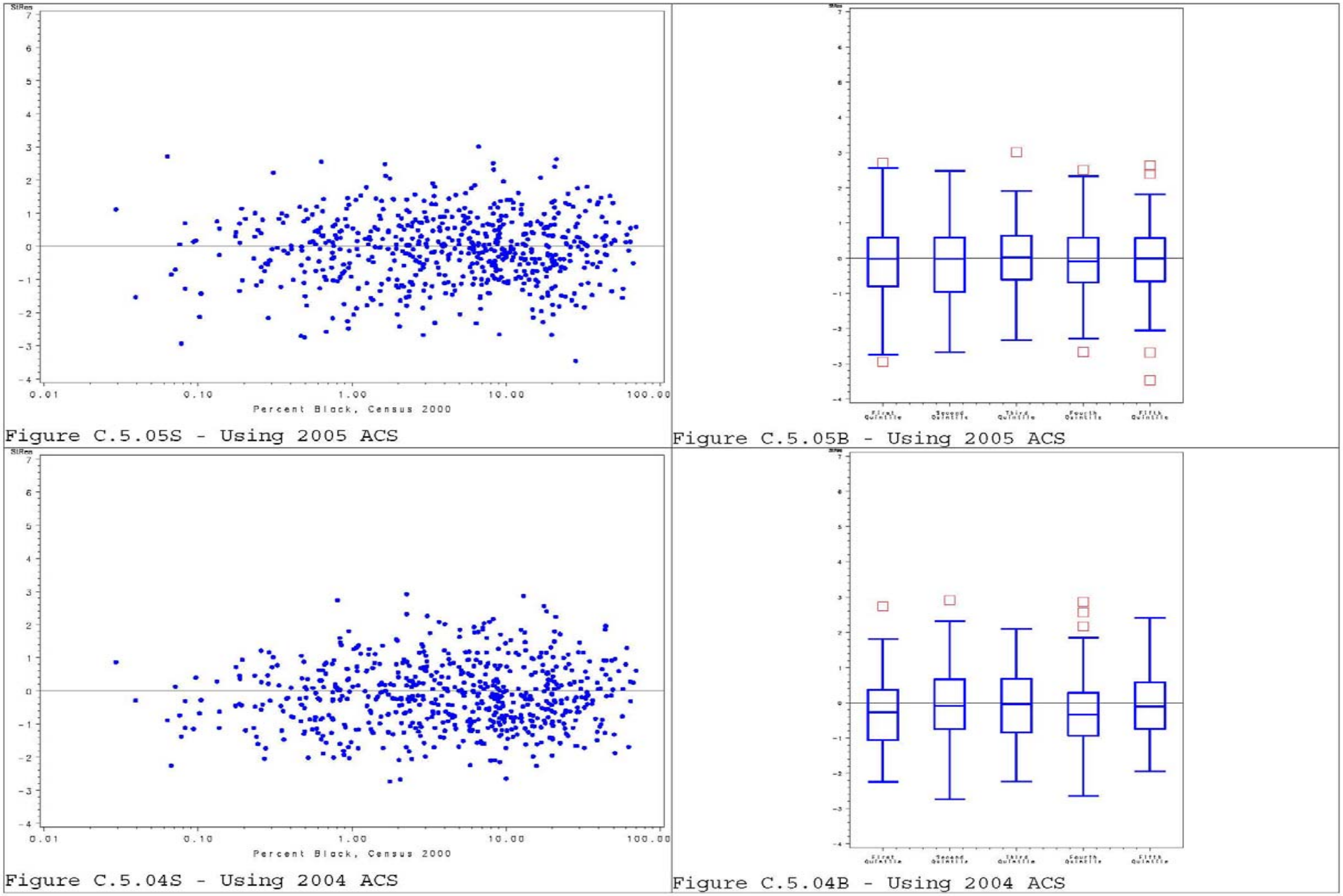


Figure C.5: Standardized residuals against Census 2000 percent Black.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

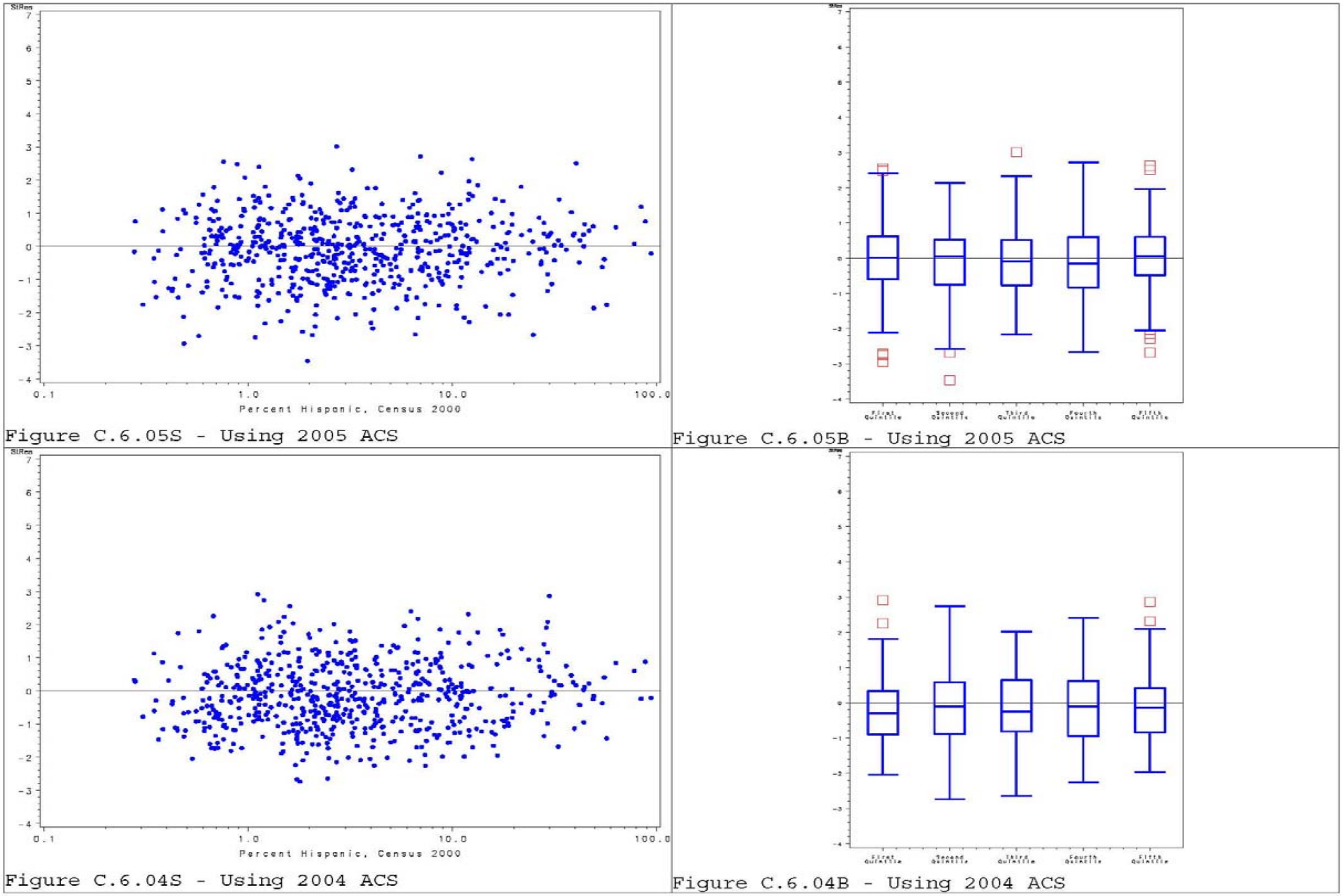


Figure C.6: Standardized residuals against Census 2000 percent Hispanic.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

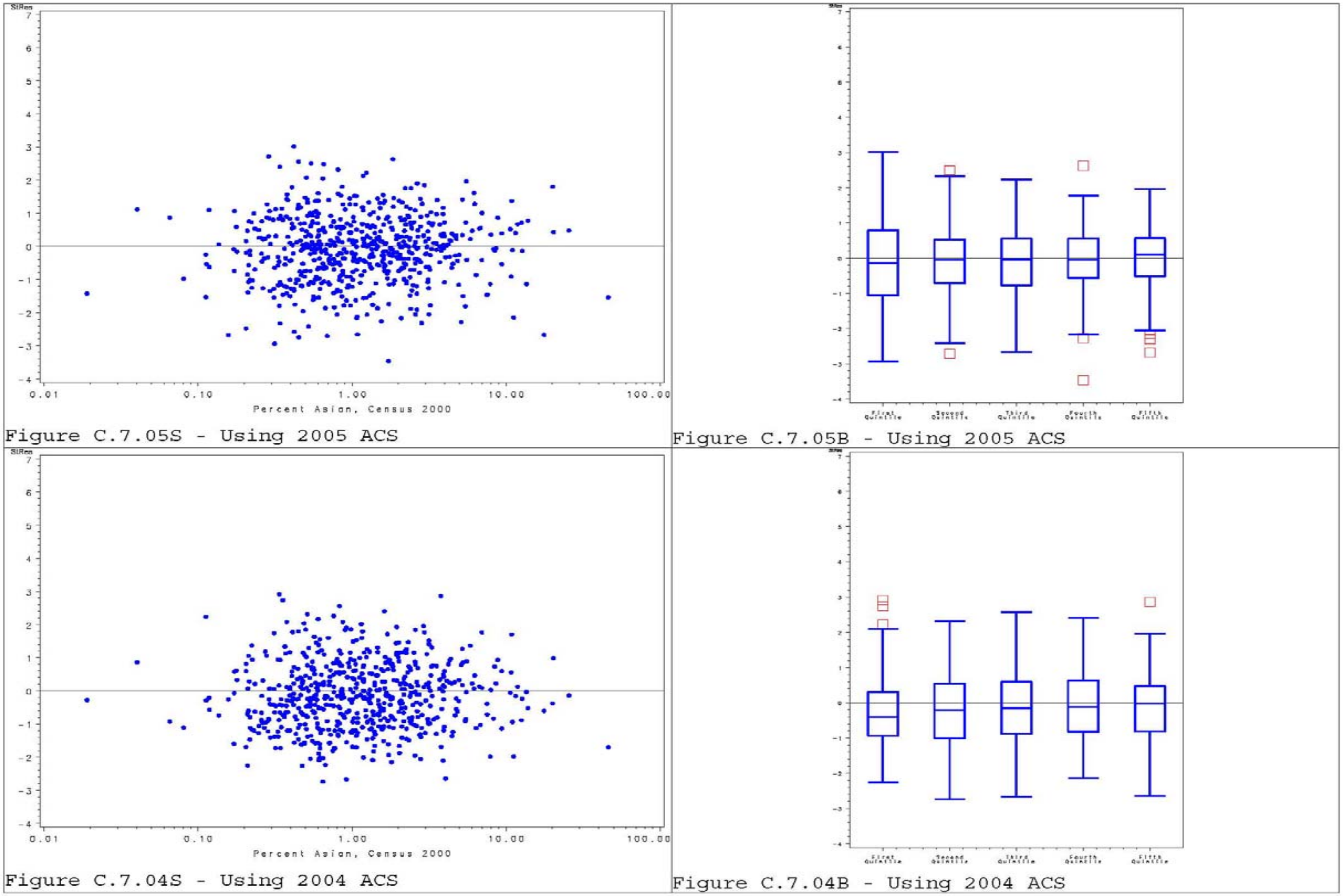


Figure C.7: Standardized residuals against Census 2000 percent Asian.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

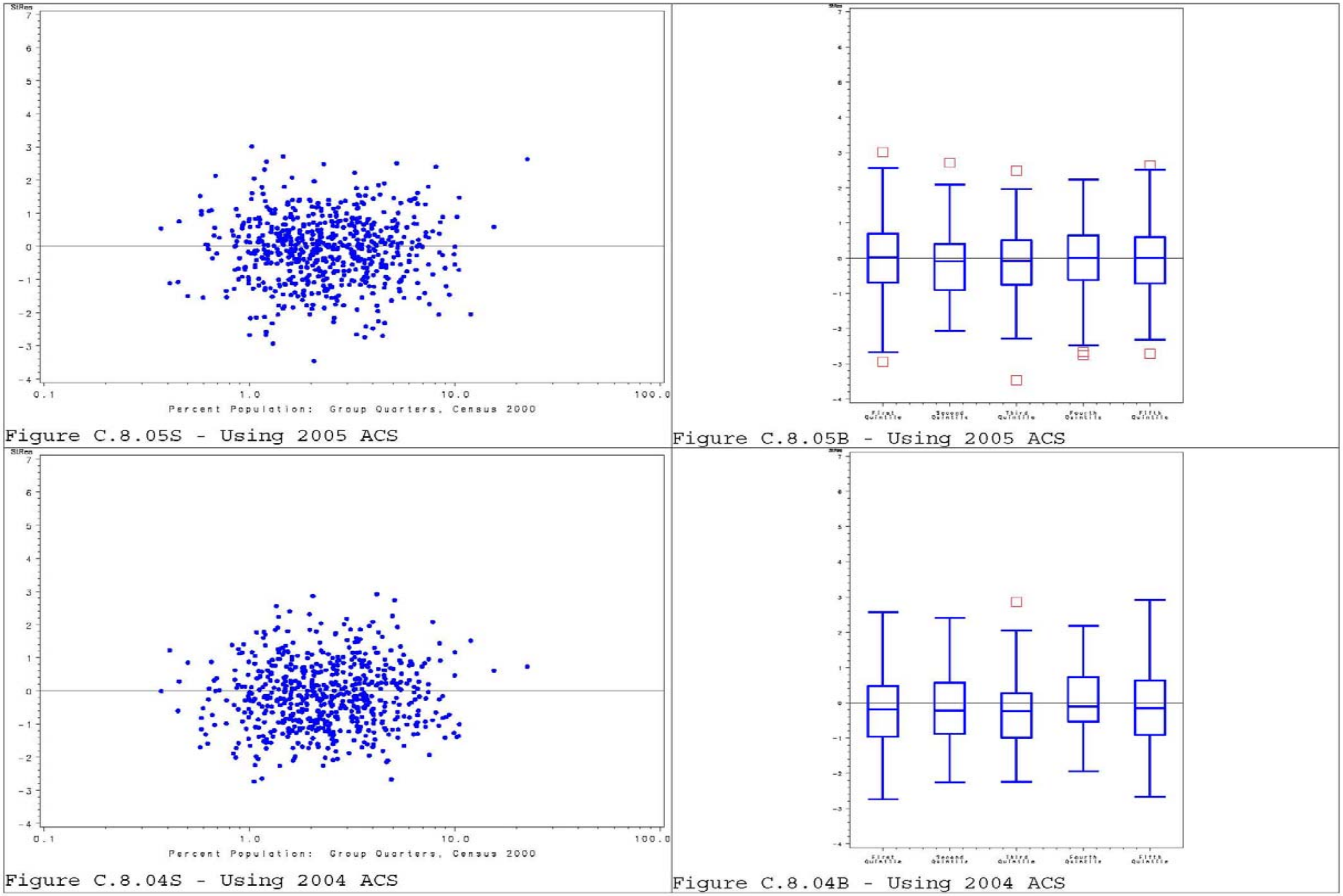


Figure C.8: Standardized residuals against Census 2000 percent group quarters' population.
The top two plots are from the 2005 ACS model, and the bottom two plots are from the 2004 ACS model. The left-hand plots are scatter plots, and the right-hand plots are box-whisker plots by 124-county quintile.

References

Asher, Jana and Robin Fisher, 2000. "Alternate CPS Sampling error variance Structures for Constrained and Unconstrained County Models: Additional Research". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/techrep/tech.report.4.pdf>

Basel, Wesley and Brett O'Hara, 2006. "Incorporating EITC Participation Data into the SAIPE Program's County Poverty Model". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/Basel2006asa.pdf>

Bell, William R., 1999. "Accounting for Uncertainty About Variances in Small Area Estimation". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/Bell99.pdf>

Bishaw, Alemayehu and Sharon Stern, 2006. "Evaluation of Poverty Estimates: A Comparison of the American Community Survey and the Current Population Survey". U.S. Census Bureau. <http://www.census.gov/hhes/www/poverty/acs_cpspovcompreport.pdf.>

Cruse, Craig and David Powers, 2006. "Estimating School District Poverty with Free and Reduced-Price Lunch Data". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/CrusePowers2006asa.pdf >

Fisher, Robin, 1997. "Methods Used for Small Area Poverty and Income Estimation". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/Fisher97.pdf>

Fisher, Robin and Jana Asher, 2000. "Alternate CPS Sampling error variance Structures for Constrained and Unconstrained County Models". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/techrep/tech.report.1.revised.pdf>

Fisher, Robin and Jana Asher, 1999. "Bayesian Hierarchical Modeling of U.S. County Poverty Rates". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/FisherAsher99.pdf>

Fisher, Robin and Jana Asher, 2000. "Alternate Scaling Parameter Functions in a Hierarchical Bayes Model of U.S. County Poverty Rates". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/FisherAsher00.pdf >

Fisher, Robin and Geof Gee, 2004. "Errors-In-Variables County Poverty and Income Models". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/FisherGee2004asa.pdf >

Fisher, Robin, 2003. "Errors-in-Variables Model for County-level Poverty Estimation". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/techrep/tech.report.5.pdf>

Huang, Elizabeth T. and William R. Bell, 2003. "Using the t-distribution in Small Area Estimation: An Application to SAIPE State Poverty Models". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/HuangBell2006asa.pdf >

Huang, Elizabeth T. and William R. Bell, 2004. "An Empirical Study on Using ACS Supplementary Survey Data in SAIPE State Poverty Models". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/HuangBell2004asa.pdf >

Jiang, Jiming, Partha Lahiri, Shu-Mei Wan, and Chien-Hua Wu, 2001. "Jackknifing in the Fay-Herriot Model with an Example". FCSM Seminar on the Funding Opportunity in Survey Research. <http://www.fcsm.gov/working-papers/spwp33_5.pdf.>

Maples, Jerry J. and William R. Bell, 2005. "Evaluation of School District Poverty Estimates: Predictive Models using IRS Income Tax Data". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/asa05finalmaples.pdf>

Maples, Jerry J., 2004. "Investigating the use of IRS Tax Data in the SAIPE School District Poverty Estimates". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/Maples2004asa.pdf>

National Research Council , 1997. *Small-Area Estimates of Children in Poverty, Interim Report 1, Evaluation of 1995 County and School District Estimates for Title I Allocations.* Panel on Estimates of Poverty for Small Geographic Areas. C.F. Citro, M.L. Cohen, G. Kalton, and K.K. West, eds., Committee on National Statistics. Washington, D.C.: The National Academies Press.

National Research Council, 1998. *Small-Area Estimates of Children in Poverty, Interim Report 2, Evaluation of 1995 County and School District Estimates for Title I Allocations.* Panel on Estimates of Poverty for Small Geographic Areas. C.F. Citro, M.L. Cohen, and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: The National Academies Press.

National Research Council, 1999. *Small-Area Estimates of Children in Poverty, Interim Report 3, Evaluation of 1995 County and School District Estimates for Title I Allocations.* Panel on Estimates of Poverty for Small Geographic Areas. C.F. Citro and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: The National Academies Press.

National Research Council, 2000a. *Small-Area Estimates of Children in Poverty: Evaluation of Current Methodology.* Panel on Estimates of Poverty for Small Geographic Areas. C.F. Citro and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: The National Academies Press.

National Research Council, 2000b. *Small-Area Income and Poverty Estimates: Priorities for 2000 and Beyond.* Panel on Estimates of Poverty for Small Geographic Areas. C.F. Citro and G. Kalton, eds., Committee on National Statistics. Washington, D.C.: The National Academies Press.

National Research Council, 2007. *Using the American Community Survey: Benefits and Challenges.* Panel on the Functionality and Usability of Data from the American Community Survey. Constance F. Citro and Graham Kalton, eds., Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Nelson, Charles, 2006. "What Do We Know About Differences Between CPS and ACS Income and Poverty Estimates?" U.S. Census Bureau. <http://www.census.gov/hhes/www/income/ nelson_082906.pdf>

Office of Management and Budget, 1978. "Statistical Policy Directive 14". U.S. Census Bureau. <http://www.census.gov/hhes/www/ povmeas/ombdir14.html>

Posey, Kirby G., Edward Welniak, and Charles Nelson, 2003. "Income in the ACS: Comparisons to Census 2000". U.S. Census Bureau. <http://www.census.gov/acs/www/Downloads/ACS/ASA_nelson.pdf>

Powers, David and William R. Bell, 2007. "Methodology for Testing for a Rise in Child Poverty Rate of Five Percent or Greater: 2003 to 2004". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/officialdocs/Tech-0304.pdf>

Powers, David, 2005. "Using Medicaid Participant Data in the Estimation of County Poverty Levels". U.S. Census Bureau. <http://www.census.gov/hhes/www/saipe/asapaper/asa2005dpowers.pdf>

Slud, Eric V, 2003. "Assessing Fit of SAIPE Models to Census and CPS County Child-Poverty Rates". <http://www.fcsm.gov/03papers/Slud.pdf >

Slud, Eric V, 2000. "Comparison Of Aggregate Versus Unit-Level Models For Small-Area Estimation". <http://www.math.umd.edu/~evs/myr.html/Census/asajsm00.pdf>

Slud, Eric V. and Tapabrata Maiti, 2005. "Small-Area Estimation based on Survey Data from a Left-Censored Fay-Herriot Model". <http://www.math.umd.edu/~evs/myr.html/LTruncSAE.pdf>

Slud, Eric V. and Tapabrata Maiti, 2005. "MSE Estimation in Transformed Fay-Herriot Models." University of Maryland-College Park. <http://www.math.umd.edu/~evs/myr.html/TransSAE5.pdf >

Stern, Sharon M, 2003. "Counting People with Disabilities: How Survey Methodology Influences Estimates in Census 2000 and the Census 2000 Supplementary Survey". U.S. Census Bureau. <http://www.census.gov/acs/www/Downloads/ACS/finalstern.pdf>

Stern, Sharon S. and Matthew Brault, 2005. “Disability Data From the American Community Survey: A Brief Examination of the Effects of a Question Redesign in 2003” U.S. Census Bureau. <http://www.census.gov/hhes/www/disability/ACS_disability.pdf>

Tapabrata, Maiti and Eric V. Slud, 2003. “Comparison of Small Area Models in SAIPE”. University of Maryland-College Park. <http://www.math.umd.edu/~evs/myr.html/SAErv5B.pdf>

U.S. Census Bureau, 1999. “Estimation Procedure Changes for the 1996 (revised) and 1997 Estimates”. <http://www.census.gov/hhes/www/saipe/techdoc/methods/97change.htm

Wolter, Kirk M., 1985. *Introduction to Variance Estimation.* New York: Springer-Verlag.