RESEARCH REPORT SERIES
(Statistics #2007-18)

An Empirical Study on Using CPS and ACS Survey Data in Bivariate State Poverty Models

Elizabeth T. Huang
William R. Bell*

Statistical Research Division
Office of the Director*

U.S. Census Bureau
Washington, DC 20233

Report Issued: November 16, 2007

An Empirical Study on Using CPS and ACS Survey Data in Bivariate State Poverty Models[1]

Elizabeth T. Huang and William R. Bell

U.S. Census Bureau, Washington D.C. 20233

November 14, 2007

[1] Disclaimer: This report is released to inform interested parties of ongoing research and to encourage discussion of work in progress. The views expressed on statistical, methodological, technical, or operational issues are those of the authors and not necessarily those of the U. S. Census Bureau.

1. Introduction

The Census Bureau's Small Area Income and Poverty Estimates (SAIPE) program produces poverty estimates for various age groups for states, counties, and school districts. For states the age groups are 0-4, 5-17, 18-64, and 65+. Through 2006 the state estimates have come from a regression model with state random effects (Fay and Herriot 1979) applied to direct state estimates from the Current Population Survey (CPS) Annual Social and Economic Supplement (ASEC, formerly known as the CPS March income supplement).[1] The models borrow information from regression variables related to poverty that are constructed from administrative records data and from poverty estimates from the previous decennial census. Estimates are identified by the "income year" (IY), which refers to the year for which income is reported in the ASEC. From IY 2000 on, the CPS ASEC sample size has been about 100,000 addresses.[2] Further information is available on the SAIPE web site at http://www.census.gov/hhes/www/saipe/index.html. For simplicity, in what follows we shorten references to "CPS ASEC" to just "CPS."

From 2000-2004 demonstration surveys fielded to test data collection procedures for the American Community Survey (ACS) have also provided state poverty estimates. The ACS asks essentially the same questions as previous decennial census long form surveys, and is replacing the long form, but with the data collection spread continuously throughout the decade, rather than at a single point in time. The demonstration surveys of ACS had sample sizes on the order of 700,000 to 800,000 addresses, significantly larger than the CPS.

The full implementation of the ACS sample started in January 2005, with a national annual sample size of approximately 3 million addresses. Poverty estimates for states (as well as for counties and other places with populations of 65,000 or more) are now available from the 2005 ACS and, more recently, the 2006 ACS. Further information on the ACS may be found at http://www.census.gov/acs/www/.

ACS poverty estimates differ from CPS poverty estimates as discussed in Bishaw and Stern (2006) and Nelson (2006). ACS has tended to produce higher poverty rates, with those at the national level exceeding those from CPS by statistically significant amounts. Also, Nelson (2006) reports significant results of chi-squared tests of the equality of the distribution of poverty across states from the ACS and CPS estimates. That is, there is evidence that the distribution of poverty differs between the ACS and CPS estimates. The reasons for these differences are not fully understood, but some known reasons to expect differences arise from ways that the ACS procedures for collecting income data differ from those of the CPS. One relevant point is that ACS collects income data continuously with a reference period of the previous 12 months (at the time income is reported) whereas the CPS collects income data in February–April with a reference period of the previous calendar year. Annual ACS state estimates use data collected over a full year, and thus involve income reports that cover different 12 month time frames extending over a period of nearly

[1] As of this writing, the 2006 estimates are the latest production state estimates that SAIPE has released. The SAIPE program is currently investigating use of ACS data as the basis for the production estimates.

[2] Prior to IY 2000 the CPS ASEC sample size was about 60,000 households. Starting in IY 2000 the sample was expanded to produce estimates of health insurance coverage for the State Children's Health Insurance Program.

two years (23 months). These timing differences suggest some alternatives for joint modeling of the ACS and CPS data that will be examined later in this paper.

In this paper we report the results of an empirical study investigating the potential benefits to state poverty models of using data from both the CPS and ACS via a bivariate model. Our initial efforts (Huang and Bell 2004) combined CPS data with data from ACS demonstration surveys for 2000 and 2001. Here we add similar results for IY 2002 and also report the results of combining CPS data (for IY 2004) with full production 2005 ACS estimates. We assess the potential benefits of using both data sources by comparing prediction error variances from the bivariate model with those from corresponding univariate models applied to the CPS or ACS data separately. We focus mostly on the potential of the bivariate models to reduce prediction error variances for the CPS equation, but present results for the ACS equation as well. In the first case the targets are the "true" state poverty ratios (underlying population quantities) being estimated by the CPS data, in the second case the targets are the "true" state poverty ratios being estimated by the ACS data. As noted above, these two targets are different.

Our empirical results suggest that use of the ACS data has some potential to reduce prediction error variances below those of the CPS univariate state poverty ratio models, but there are two qualifications. First, the results vary over states, with some states actually showing increased variances. Second, we tried alternative bivariate models for using the ACS survey data, and results varied across the alternative models. Models that made more restrictive assumptions yielded apparently greater improvements in prediction error variances. The validity of these results depends, though, on the more restrictive model assumptions holding. Therefore, we also examined statistical tests (chi-squared tests) of these restrictions. The most restrictive assumptions (such as assuming no difference between what CPS and ACS are estimating) were rejected. Results from the ACS demonstration survey data and the full production ACS 2005 data lead to similar conclusions about the potential for reducing CPS equation prediction error variances by borrowing information from ACS data.

Other results show that there is essentially no potential for using CPS data to improve prediction error variances in the ACS equation (below those from the ACS univariate model). The sampling error in the CPS estimates is simply too large relative to that in the ACS data for the CPS data to provide useful additional information about state poverty as estimated by ACS.

Section 2 presents the alternative models we examined for the CPS and ACS state poverty ratios. The prediction error variances for our models are posterior variances computed via a Bayesian approach, which is also discussed in Section 2. Section 3 contains the empirical results from using ACS demonstration survey data (for IY2000–2002), including the prediction error variance comparisons, results of the chi-squared tests, and comparisons of point estimates from alternative models. Section 4 contains the empirical results from using the full production ACS 2005 data and CPS data for IY 2004 in our models. Finally, Section 5 summarizes our results and draws conclusions.

2. Alternative Models for State Poverty Ratios

To incorporate information from both CPS and ACS data we use a general bivariate regression model with random effects. Bell (2000) discussed this model in the context of county poverty models. Section 2.1 discusses the general bivariate model, and Section 2.2 some alternative (restricted) bivariate models, as well as the univariate model currently used in SAIPE production. Section 2.3 then discusses Bayesian treatment of the models, i.e., how we obtain posterior means and variances for the state poverty ratios.

2.1 General Bivariate Model

For any given year and age group, let Y_{1i} and Y_{2i} be the "true poverty ratios" (number poor / population) for state i that are being estimated by the CPS and ACS, respectively, for $i = 1,\dots,51$ (including the 50 states and the District of Columbia). As noted earlier, due to data collection and possibly other differences between the CPS and ACS, we assume that $Y_{1i} \neq Y_{2i}$, in general. Let y_{1i} and y_{2i} be the direct sample estimated poverty ratios for state i from the CPS and ACS, respectively. Then we have

$$y_{1i} = Y_{1i} + e_{1i}$$

$$y_{2i} = Y_{2i} + e_{2i},$$

where the sampling errors e_{1i} and e_{2i} are assumed to be independently distributed as $N(0, \mathrm{v}_{ji})$, $j = 1,2$. The v_{ji} are assumed known, though they are actually estimates of the true sampling variances. In the case of CPS, the direct variance estimates are smoothed using a sampling error model (Otto and Bell 1995) to get the v_{1i}. In the case of ACS, we use the direct sampling variance estimates as the v_{2i}. Finally, we assume $\mathrm{Cov}(e_{1i}, e_{2i}) = 0$, because the CPS and ACS are independent samples.

Our model for the true poverty ratios is:

$$Y_{1i} = \alpha_1 + x_i' \beta_1 + u_{1i}$$

$$Y_{2i} = \alpha_2 + x_i' \beta_2 + u_{2i}$$

where the α's and β's are regression parameters, x_i' is a row vector of regression variables, and $(u_{1i}, u_{2i})'$ are independently and identically normally distributed with zero means. Note the same vector of regression variables x_i' is used in both the CPS and ACS equations. We let

$$\mathrm{Var}(u_{1i}) = \mathrm{s}_{11},\ \ \mathrm{Var}(u_{2i}) = \mathrm{s}_{22},\ \text{ and }\ \mathrm{Corr}(u_{1i}, u_{2i}) = \rho.$$

The regression variables in x_i' include pseudo state poverty rates constructed from Internal Revenue Service (IRS) tax data, tax non-filer ratios constructed from IRS data and state population estimates, Supplementary Security Income (SSI) state participation rates (for age 65+ only) constructed from

Social Security Administration data and state population estimates, and Census 2000 state poverty ratios.[3] For more information see the SAIPE web site mentioned earlier.

Noninformative prior distributions for the model parameters are assumed as follows:

$\boldsymbol{\beta} = (\alpha_1, \beta_1', \alpha_2, \beta_2')'$ is distributed as multivariate N($\mathbf{0}$, cI), with c large,

s_{11} and s_{22} are independently distributed as Uniform (0, m_1) and Uniform (0, m_2), with m_1 and m_2 large, and

ρ is distributed as Uniform (-1, 1).

The values of c, m_1, and m_2 were chosen to be sufficiently large so that the priors could effectively be regarded as flat on $(-\infty, +\infty)$ and $(0, +\infty)$ as appropriate. We used c = 1,000 for all age groups and chose values for m_1 and m_2 separately for each age group so that the likelihood (for the univariate models discussed below) was effectively zero beyond m_1 and m_2. (E.g., for age 5-17, this led to choosing $m_1 = m_2 = 20$.)

2.2 Alternative Models

Bivariate Model A is the general bivariate model discussed above with no restrictions on the model parameters (except $0 \leq s_{11} \leq m_1$, $0 \leq s_{22} \leq m_2$, and $|\rho| \leq 1$.)

Bivariate Model B1 assumes that the CPS and the ACS estimate the same state poverty ratio, that is, $Y_{1i} = Y_{2i}$. (For Model A this implies the constraints $\alpha_1 = \alpha_2$, $\beta_1 = \beta_2$, and $u_{1i} = u_{2i}$, which in turn imply that $s_{11} = s_{22}$ and $\rho = 1$.)

Bivariate Model B2 assumes that the CPS and ACS models have the same regression parameters ($\alpha_1 = \alpha_2$ and $\beta_1 = \beta_2$), but with different model errors ($u_{1i} \neq u_{2i}$, so that $s_{11} \neq s_{22}$ and $\rho \neq 1$, in general).

Bivariate Model B3 assumes that, excluding the intercepts, the regression coefficients in the CPS and ACS regression equations are the same ($\beta_1 = \beta_2$).

Univariate Models U: If $\rho = 0$, then Model A reduces to separate univariate regression models and we fit the CPS and ACS equations separately.

Through IY 2004 (results released in November 2006), the univariate model using the CPS equation

[3] Starting with IY 2004 two changes were made to the regression variables in the state poverty ratio models for ages 0-4, 5-17, and 18-64. First, we added state participation rates in the food stamp program. Second, we replaced the Census 2000 state poverty ratios by corresponding "census residuals" obtained by regressing the Census 2000 estimates on the other regression variables defined for IY 1989. These changes are discussed in connection with the IY 2004 estimates on the SAIPE web site. The model for the 65+ poverty ratios remained unchanged.

has been the SAIPE production state model.[4] If $\rho \neq 0$ then a bivariate model has potential benefits compared to the univariate model.

2.3 Bayesian Inference for the Models

For each model above,[5] we used Gibbs sampling via WinBUGs Version 1.4 (Spiegelhalter, et al. 2003) to simulate 10,000 sets of model parameters (ρ, $s_{11,}$ s_{22}, α_1, β_1, α_2, β_2) from their joint posterior distribution given the data.(Note for Univariate Model U, $\rho = 0$ so there is no need to simulate ρ.) Many empirical runs were carried out to check the convergence of the MCMC runs using WinBUGs, and it was decided that single chain runs of 10,500 simulations with the first 500 simulations discarded was sufficient to achieve convergence. The posterior means and variances of Y_{1i} from the CPS equation were approximated by averaging results over the simulations of (ρ, $s_{11,}$ s_{22}) to approximate the following formulas under each model respectively:[6]

$$E(Y_{1i} \mid \mathbf{y}) = E_{\rho,\, s11,\, s22 \mid \mathbf{y}} [E(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})] \quad (1)$$

$$Var(Y_{1i} \mid \mathbf{y}) = E_{\rho,\, s11,\, s22 \mid \mathbf{y}} [Var(Y_{1i} \mid \mathbf{y}, \rho, s_{11,} s_{22})] + Var_{\rho,\, s11,\, s22 \mid \mathbf{y}} [E(Y_{1i} \mid \mathbf{y}, \rho, s_{11,} s_{22,})] \quad (2)$$

where $\mathbf{y} = \{(y_{1i}, y_{2i}), i = 1,...,51\}$ is the observed data. In (1) and (2), $E(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$ and $Var(Y_{1i} \mid \mathbf{y}, \rho, s_{11,} s_{22})$ can be readily calculated from standard formulas that account for the effects of uncertainty about $\boldsymbol{\beta}$. (See, e.g., Bell 1999.) $E_{\rho,\, s11,\, s22 \mid \mathbf{y}}[\bullet]$ and $Var_{\rho,\, s11,\, s22 \mid \mathbf{y}}[\bullet]$ were approximated by taking the sample mean and variance across the simulations of the terms as indicated. The analogous calculations were also made to obtain the posterior means and variances of the Y_{2i}, the true poverty ratios in the ACS equation.

Tables 2.1 and 2.2 show posterior means and standard deviations of the parameters (ρ, $s_{11,}$ s_{22}) from the Gibbs sampling via WinBUGs of 10,000 simulations from bivariate Models A and B3 for IYs 2000-2002. Table 2.3 shows corresponding results for the model parameters ($s_{11,}$ s_{22}) from univariate Model U (for which ρ is set to 0). Notice that in most cases the posterior means and standard deviations for ($s_{11,}$ s_{22}) for a given year are fairly close across all three models (though some larger differences show up for age 0–4). Notice also that the standard deviations for s_{11} and s_{22} are not small relative to their posterior means (so their posterior coefficients of variation are large), and that the posterior standard deviations for ρ are not small relative to the width of the interval $(-1,1)$. This reflects considerable uncertainty about these model parameters. This uncertainty can also be seen from estimates of the posterior densities for (ρ, $s_{11,}$ s_{22}, $\boldsymbol{\beta}$) from Model A, which are plotted in Figure 1 for age 5-17 in IY 2000. In fact, in regard to ρ, in no case does the posterior for ρ give

[4] As of this writing, current plans call for switching the production SAIPE state estimates from using CPS data to using ACS data.

[5] Actually, Model B1 was not implemented, because, as will be seen in Section 3.2, even the less restrictive Model B2 is rejected by chi-squared tests.

[6] Note that we actually discarded the simulations of (α_1, β_1, α_2, β_2) and used only those for (ρ, $s_{11,}$ s_{22}); thus, we were using simulations of (ρ, $s_{11,}$ s_{22}) from their marginal posterior distribution. This was done because equations (1) and (2) use exact analytical results in dealing with the regression parameters (α_1, β_1, α_2, β_2), which is more numerically accurate than are calculations that use the simulations of the regression parameters.

conclusive evidence that ρ > 0 (i.e., the 95 percent highest posterior density intervals for ρ always include negative values). Plots of the posterior densities of the regression coefficients from Model A shown in Figure 1 also show considerable uncertainty, more so for the first four coefficients shown (which refer to the CPS equation) than for the last four coefficients (which refer to the ACS equation). This is because the lower level of sampling error in the ACS estimates leads to more precise estimates of the ACS equation regression coefficients. Also note that the posterior densities of the regression coefficients appear reasonably normal. (The notation in Figure 1 of beta[1],...,beta[8] corresponds to the regression parameters $\boldsymbol{\beta} = (\alpha_1, \beta_1', \alpha_2, \beta_2')'$.)

Table 2.1 Posterior means and standard deviations of the parameters of Model A

IY 2000

age	0-4	5-17	18-64	65+
ρ	0.53 (0.38)	0.29 (0.46)	0.34 (0.41)	− 0.20 (0.47)
s_{11}	2.92 (2.23)	0.81 (0.70)	0.23 (0.19)	0.76 (0.71)
s_{22}	2.24 (1.21)	1.26 (0.58)	0.48 (0.17)	0.66 (0.30)

IY 2001

age	0-4	5-17	18-64	65+
ρ	0.17 (0.53)	0.54 (0.33)	0.49 (0.40)	−0.07 (0.53)
s_{11}	1.80 (1.70)	1.85 (1.21)	0.18 (0.16)	0.39 (0.40)
s_{22}	1.53 (0.97)	0.92 (0.45)	0.37 (0.13)	0.38 (0.21)

IY 2002

age	0-4	5-17	18-64	65+
ρ	0.44 (0.40)	0.25 (0.50)	0.56 (0.29)	0.32 (0.44)
s_{11}	3.40 (2.40)	0.84 (0.80)	0.36 (0.24)	1.40 (0.98)
s_{22}	1.47 (0.91)	0.77 (0.43)	0.58 (0.19)	0.25 (0.17)

Table 2.2 Posterior means and standard deviations of the parameters of Model B3

IY 2000

age	0-4	5-17	18-64	65+
ρ	0.56 (0.37)	0.34 (0.46)	0.40 (0.42)	−0.15 (0.48)
s_{11}	2.89 (2.23)	0.74 (0.63)	0.20 (0.17)	0.73 (0.67)
s_{22}	2.44 (1.26)	1.20 (0.58)	0.47 (0.17)	0.63 (0.29)

IY 2001

age	0-4	5-17	18-64	65+
ρ	0.07 (0.53)	0.59 (0.32)	0.45 (0.41)	0.02 (0.55)
s_{11}	2.04 (1.92)	1.73 (1.12)	0.17 (0.16)	0.32 (0.34)
s_{22}	1.63 (1.03)	0.90 (0.45)	0.38 (0.14)	0.35 (0.19)

IY 2002

age	0-4	5-17	18-64	65+
ρ	0.39 (0.41)	0.25 (0.49)	0.55 (0.28)	0.28 (0.44)
s_{11}	3.70 (2.55)	0.84 (0.76)	0.37 (0.24)	1.30 (0.94)
s_{22}	1.50 (0.94)	0.79 (0.43)	0.59 (0.19)	0.27 (0.19)

Table 2.3 Posterior means and standard deviations of the parameters of Model U

IY 2000

age	0-4	5-17	18-64	65+
s_{11}	2.50 (2.11)	0.81 (0.69)	0.24 (0.20)	0.80 (0.69)
s_{22}	2.09 (1.15)	1.24 (0.57)	0.46 (0.17)	0.63 (0.28)

IY 2001

age	0-4	5-17	18-64	65+
s_{11}	1.84 (1.73)	1.79 (1.23)	0.16 (0.15)	0.41 (0.41)
s_{22}	1.52 (0.97)	0.90 (0.45)	0.36 (0.13)	0.38 (0.21)

IY 2002

age	0-4	5-17	18-64	65+
s_{11}	3.38 (2.46)	0.86 (0.78)	0.36 (0.25)	1.45 (1.00)
s_{22}	1.45 (0.92)	0.76 (0.41)	0.58 (0.18)	0.26 (0.18)

3. Empirical Model Comparisons Using ACS Demonstration Survey Data (2000-2002)

Our primary focus here is on whether using ACS demonstration survey data in conjunction with the CPS data can reduce prediction error (posterior) variances of Y_{1i}, the "true" poverty ratio as estimated by CPS? The first four subsections here present empirical results directly relevant to this question. First, in Section 3.1 we compare posterior variances of Y_{1i} from bivariate Model A with those from the CPS univariate model. We find Model A generally provides small improvements on average, though we also find a few instances where it produces large *increases* in posterior variance. Section 3.2 then presents results of chi-squared tests of the restrictions imposed on the regression coefficients by Models B2 and B3, testing if these more restrictive models are consistent with the data. The tests reject Model B2 (and by implication, also the more restrictive Model B1), while we fail to reject Model B3. Section 3.3 thus compares posterior variances of Y_{1i} from Model B3 with those from the univariate model U to see what improvements may result from use of the more restrictive bivariate Model B3. We find larger average improvements in posterior variance than with Model A, though we still find a few instances of large posterior variance increases. Section 3.4 examines the mathematical reasons for the occasional large posterior variance increases, and notes the instances from our models in which these occur. Section 3.5 reports summary results on estimates of Y_{2i} and on corresponding posterior variances, where Y_{2i} is the "true" poverty ratio estimated by ACS. Section 3.6 examines the effects on point estimates from the bivariate models of shifting the ACS data one year ahead. Finally, Section 3.7 compares point estimates of poverty ratios for individual states from the alternative models.

To compare posterior variances we examine their relative percentage differences under an alternative model with those from the univariate model U. For example, we compare posterior variances from Model A with those from the univariate model U by computing ($\mathbf{y}_1 = (y_{11},\ldots,y_{1,51})'$)

$$100 \times \frac{\mathrm{Var}(Y_{1i}|\mathbf{y},\ \text{Model A}) - \mathrm{Var}(Y_{1i}|\mathbf{y}_1,\ \text{Model U})}{\mathrm{Var}(Y_{1i}|\mathbf{y}_1,\ \text{Model U})}.$$

3.1 Posterior Variance Comparisons for Model A – CPS Equation

Table 3.1 summarizes the comparisons of the posterior variances of the state poverty ratios Y_{1i} from Model A with those from the univariate model using the relative percentage differences as defined earlier. We see, for all three years and across all age groups at most only small improvements in posterior variances on average from use of the bivariate model A. For age 0-4 in IY 2000, and age 18-64 in IY 2001, using Model A actually produces a small average increase in posterior variance.[7] The min values in Table 3.1 show that some states show more substantial variance reductions than others, while the max values show that some states show substantial variance increases. Appendix A shows more detail from these results, presenting the frequency distribution of the percentage differences in posterior variances between Model A and Model U. These tables show that substantial variance increases for states are relatively rare, while small to moderate variance reductions predominate (except for age 0-4 in IY 2000). The instances of substantial variance increases with Model A are noted in Section 3.4, and the reasons for this are discussed.

The best cases for variance reductions from the bivariate Model A are for age 18-64 in IY 2000, age 5-17 in IY 2001, and age 18-64 in IY 2002. Even for these cases the average variance reductions are small, and for most individual states the percentage differences in posterior variances are small or moderate at best. For the other cases, it is difficult to claim any overall advantage from using bivariate Model A. Also, the fact that the age group showing the most improvement varies over the three years is not encouraging.

[7] This increase may be due to the posterior mean of s_{11} being larger in Model A than in Model U.

Table 3.1 Relative percent differences of posterior variances: Model A versus univariate model (means are unweighted averages of percent differences across the 50 states and DC)

IY 2000

Age	Mean	Min	Max
0-4	2.1	−10.4	29.8
5-17	−2.8	−12.1	17.1
18-64	−5.6	−14.7	18.4
65+	−4.4	−12.9	18.9

IY 2001

Age	Mean	Min	Max
0-4	−2.7	−12.2	32.0
5-17	−5.3	−13.9	8.3
18-64	1.3	−10.2	40.8
65+	−3.3	−11.9	15.8

IY 2002

Age	Mean	Min	Max
0-4	−2.9	−10.4	4.0
5-17	−3.4	−12.7	52.4
18-64	−11.2	−21.2	7.6
65+	−3.7	−9.8	10.1

3.2 Chi-Squared Tests of Model Restrictions

Section 2.2 presented three alternative bivariate models (B1, B2, and B3), all of which impose restrictions on the general bivariate Model A. The restrictions implied by Models B2 and B3 can be tested by testing the following null hypotheses:

H2: $\alpha_1 = \alpha_2, \quad \beta_1 = \beta_2$ (Model B2)
H3: $\beta_1 = \beta_2$ (Model B3)

Hypothesis H2 postulates equality of all the regression coefficients in the CPS and ACS equations. Hypothesis H3 postulates this equality apart from the intercept terms. We test these hypotheses against the alternative hypothesis of Model A holding with no restrictions. While we will not explicitly test Model B1 (CPS and ACS estimate the same poverty ratios), note that if H2 is rejected so, by implication, is the more restrictive Model B1.

To test the hypotheses H2 and H3 we formulate chi-squared statistics using the posterior means and covariance matrices of the regression coefficients under Model A. From a Bayesian perspective, this is equivalent to checking if the null hypothesis values (here zero) of the differences of the regression coefficients from the two equations lie within a given highest posterior density (HPD) region of the parameter space. Thus, the chi-squared statistics check if the restrictions under H2 and H3 are reasonably consistent with the posterior distribution of the regression parameters under the general Model A.

More specifically, the chi-squared statistic for testing H3 is

$$\chi^2 = (b_1 - b_2)'[\mathrm{Var}\,(b_1 - b_2)]^{-1}(b_1 - b_2)$$

where b_1 and b_2 are the posterior means of β_1 and β_2, while $\mathrm{Var}(b_1 - b_2)$ is the posterior covariance matrix of $\beta_1 - \beta_2$. We assume that, under H3, the posterior distribution of $\beta_1 - \beta_2$ is approximately normal with mean vector 0 and covariance matrix $\mathrm{Var}\,(b_1 - b_2)$. The chi-squared statistic has three degrees of freedom for ages 0-4, 5-17, and 18-64, and four degrees of freedom for age 65+ (the difference being due to the additional inclusion of the SSI participation rate in x_i' for age 65+). For testing H2 an analogous statistic is used that also involves the intercepts, α_1 and α_2, and which has four degrees of freedom for ages 0-4, 5-17, and 18-64, and five degrees of freedom for age 65+. We compare χ^2 to five percent critical values from the chi-squared distribution; these are 7.8, 9.5, and 11.1 for three, four, and five degrees of freedom, respectively. For H3, this corresponds to checking if the zero vector lies in the 95 percent HPD region of the parameter space of $\beta_1 - \beta_2$.

The results of the Chi-squared tests for IYs 2000 - 2002 are given in Table 3.2 below. Values that are significant at the five percent level are shown in bold. First note that we reject the hypothesis H2 for ages 5-17 and 18-64 for IYs 2000 and 2001, and for ages 0-4 and 5-17 for IY 2002. Moreover, in most cases where H2 is not rejected, the chi-squared statistic is close to being significant at the five percent level. These results suggest that assuming all the regression parameters, including the intercepts, are the same between the CPS and ACS equations is not tenable, so we should reject Model B2 and, by implication, the more restrictive Model B1. These results are expected given the overall level difference between the CPS and ACS poverty estimates that was discussed earlier and in the reports of Bishaw and Stern (2006) and Nelson (2006). On the other hand, we fail to reject H3 for all age groups. This suggests that perhaps the regression parameters other than the intercepts can be assumed to be the same in the CPS and ACS equations, so we might consider using Model B3 instead of Model A. The consequences of this for posterior variances are examined in the next section.

Table 3.2 Chi-Squared statistics for testing hypotheses H2 and H3 for IYs 2000 - 2002

IY 2000

Age	0-4	5-17	18-64	65+
H2	8.4	**35.3**	**44.1**	8.6
H3	5.5	2.5	2.5	5.2

IY 2001

Age	0-4	5-17	18-64	65+
H2	7.7	**18.6**	**12.2**	0.7
H3	7.0	0.4	2.5	0.6

IY 2002

Age	0-4	5-17	18-64	65+
H2	**11.2**	**16.9**	7.9	10.1
H3	6.4	3.7	3.8	3.2

3.3 Posterior Variance Comparisons for Model B3 – CPS Equation

Posterior variances for Model B3 were computed as discussed in Section 2.3, i.e., using equations (1) and (2) with the simulations of (ρ, $s_{11,}$ s_{22}) obtained under Model B3, and with $E(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$ and $Var(Y_{1i} \mid \mathbf{y}, \rho, s_{11,} s_{22})$ computed to account for the Model B3 restriction, $\beta_1 = \beta_2$. Table 3.3 presents summaries of the relative percent differences of the resulting posterior variances of Model B3 from those for the univariate model ; these results can be compared to those of Table 3.1. Doing so we see that the average variance reductions from Model B3 are substantially larger than those from Model A. Also, the largest variance reductions are about 50 percent or more, and while there are some variance increases, the maximum increases from use of Model B3 are not as severe as those from Model A (except for age 0-4 in IY 2001-2002). Note also that for age 5-17 all states show variance reductions with Model B3 for IYs 2000-2001. However, for IY 2002, there is one state for which the posterior variance for age 5-17 increases 44.3 percent with use of Model B3.

Table 3.3 Relative percent differences of posterior variances: Model B3 versus univariate model (means are unweighted averages of percent differences across the 50 states and DC)

IY 2000

Age	Mean	Min	Max
0-4	−13.5	−48.9	24.3
5-17	−23.6	−52.2	−6.3
18-64	−26.8	−51.0	2.4
65+	−22.0	−50.7	11.9

IY 2001

Age	Mean	Min	Max
0-4	−10.2	−54.5	34.5
5-17	−21.2	−45.8	−4.5
18-64	−18.9	−54.1	34.9
65+	−38.5	−67.8	0.4

IY 2002

Age	Mean	Min	Max
0-4	−9.0	−52.3	9.1
5-17	−19.5	−51.4	44.3
18-64	−19.4	−46.3	2.5
65+	−21.3	−53.2	-5.4

The larger variance reductions under Model B3 than under Model A are presumably due to increased precision in the estimation of the regression coefficients under Model B3's assumption that, apart from the intercepts, the regression coefficients are common to both equations. Given the substantially lower sampling variances from ACS, under this assumption the ACS data should provide relatively more information for estimation of the common regression coefficients than does the CPS data. So there appears to be more potential for improvement from using the ACS data to improve estimation of the regression coefficients (if the assumption that they are common to both equations holds) than from using the ACS data to improve prediction of the state random effects (which is done by all the bivariate models.)

Note one important qualification to these results. The posterior variances quoted from Model B3 assume that the Model B3 restriction of $\beta_1 = \beta_2$ holds exactly (and that the other model assumptions are true, these being the assumptions for Model A). We determined that the assumption $\beta_1 = \beta_2$ was plausible by performing the chi-squared tests of Section 3.2. So to decide to use Model B3, we had to actually estimate Model A and perform the tests of $\beta_1 = \beta_2$. If we then quote posterior variances calculated as if Model B3 were known to be true we would not be accounting for this estimation and testing and we would thus understate our uncertainty about the true poverty ratios. One way around this dilemma is to use an informative prior distribution with mean zero for $\beta_1 - \beta_2$. Such a prior might be developed from estimation results for previous years.

3.4 Investigating Large Posterior Variance Increases for the Bivariate Models

Tables 3.1 and 3.3 above showed there were some instances of moderate to large posterior variance increases from use of either bivariate Model A or the restricted bivariate Model B3 instead of the univariate model. Table 3.4 shows all the instances of posterior variance increases for the bivariate models that were 15 percent or greater. Increases just above 15 percent are not of that much concern; increases exceeding 25 percent (shown in bold in the table) are of more concern. Of particular concern is the approximate 50 percent increase for age 5-17 for Alaska in IY 2002 for model A and 44% for model B3. In this section we examine what can lead to such large posterior variance increases. For concreteness, most of the discussion here focuses on the general bivariate Model A.

Table 3.4 Large ($\geq$ 15 percent) posterior variance increases from the bivariate models compared to the univariate model (increases of 25 percent or more are in bold)

Age	Model	IY 2000	IY 2001	IY 2002
0-4	A	OR(24%),**HI (29%)**, **WA (30%)**	**AK (32%)**, NC (15%), TN (23%)	
	B3	OR(19%),WA(24%), HI(16%)	**AK(35%)**,TN(18%), NC(16%)	
5-17	A	SD (17%)		**AK (52%)**, NC (18%)
	B3			**AK (44%)**
18-64	A	AZ (17%), HI(18%)	**AK (40%)**, **KS (40%), NC (41%)**, **OR (32%)**	
	B3		**AK (28%), KS (35%**)	
65+	A	OK (19%)	IL (16%)	
	B3			

The posterior variance of Y_{1i} for a given model is obtained from equation (2) above, which we repeat here for convenience:

$$\mathrm{Var}(Y_{1i} \mid \mathbf{y}) = \mathrm{E}_{\rho,\, s11,\, s22 \mid \mathbf{y}} [\mathrm{Var}(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})] + \mathrm{Var}_{\rho,\, s11,\, s22 \mid \mathbf{y}} [\mathrm{E}(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})] \qquad (2)$$

where $\mathrm{E}_{\rho,\, s11,\, s22 \mid y}$ [•] and $\mathrm{Var}_{\rho,\, s11,\, s22 \mid y}$ [•] are taken with respect to $p(\rho, s_{11}, s_{22} \mid \mathbf{y})$, the marginal posterior distribution of the model error correlation and variance parameters, (ρ, s_{11}, s_{22}), for the respective model. The first term in equation (2) is smaller for the bivariate model than is the corresponding term for the univariate model because, conditional on any values of the parameters (ρ, s_{11}, s_{22}), the variance conditional on $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$ is lower than the variance conditional on just y_1, that is, $\mathrm{Var}(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22}) \le \mathrm{Var}(Y_{1i} \mid \mathbf{y}_1, \rho, s_{11}, s_{22})$.[8] Therefore, increases in posterior variance from using the bivariate model must come from an increase in the second term. This term reflects uncertainty about the parameters (ρ, s_{11}, s_{22}). For many states the first term in (2) is much larger than the second, but occasionally the second term makes a substantial contribution. Sometimes this substantial contribution from the second term in (2) is much larger for the bivariate than for the univariate model, leading to a substantial increase in posterior variance from using the bivariate model. We now examine why this occurs.

Let $\hat{Y}_{1i} = \mathrm{E}(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$ and $\hat{Y}_{2i} = \mathrm{E}(Y_{2i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$. (We use this notation generically for both bivariate Model A and the univariate models. For the latter we should more properly fix ρ at zero and condition on only $\mathbf{y}_1$ or $\mathbf{y}_2$ rather than $\mathbf{y}$.) We obtain expressions for $\hat{Y}_{1i}$ and $\hat{Y}_{2i}$ from expressions for the expectations of Y_{1i} and Y_{2i} conditional on all model parameters. For Y_{1i} this is

$$\mathrm{E}(Y_{1i} \mid \mathbf{y}, \boldsymbol{\beta}, \rho, s_{11}, s_{22}) = (\alpha_1 + x_i' \beta_1) + c_{1i} \times [y_{1i} - \alpha_1 - x_i' \beta_1] + c_{2i} \times [y_{2i} - \alpha_2 - x_i' \beta_2] \qquad (3)$$

and an analogous expression holds for $\mathrm{E}(Y_{2i} \mid \mathbf{y}, \boldsymbol{\beta}, \rho, s_{11}, s_{22})$. In equation (3), c_{1i} and c_{2i} depend on the model error correlation and variance parameters (ρ, s_{11}, s_{22}) as well as on the sampling error variances of the CPS and ACS estimates (the latter depend on i). (For the univariate CPS model $c_{2i} = 0$ so the last term in (3) disappears.) To obtain an expression for $\hat{Y}_{1i}$ we take the expectation of (3) over the conditional posterior distribution of $\boldsymbol{\beta}$ given $(\mathbf{y}, \rho, s_{11}, s_{22})$. That is, we compute $\hat{Y}_{1i} = \mathrm{E}_{\boldsymbol{\beta} \mid (\mathbf{y}, \rho, s11, s22)} [\mathrm{E}(Y_{1i} \mid \mathbf{y}, \boldsymbol{\beta}, \rho, s_{11}, s_{22})]$, which yields

$$\hat{Y}_{1i} = (\hat{\alpha}_1 + x_i' \hat{\beta}_1) + c_{1i} \times [y_{1i} - \hat{\alpha}_1 - x_i' \hat{\beta}_1] + c_{2i} \times [y_{2i} - \hat{\alpha}_2 - x_i' \hat{\beta}_2]. \qquad (4)$$

There is an analogous expression for $\hat{Y}_{2i}$. In equation (4) $[\hat{\alpha}_1, \hat{\beta}_1] = \mathrm{E}([\alpha_1, \beta_1] \mid \mathbf{y}, \rho, s_{11}, s_{22})$ and $[\hat{\alpha}_2, \hat{\beta}_2] = \mathrm{E}([\alpha_2, \beta_2] \mid \mathbf{y}, \rho, s_{11}, s_{22})$ are generalized least squares (GLS) estimators of the regression parameters given the covariance matrix of the data determined by (ρ, s_{11}, s_{22}) and the sampling error variances. Although these GLS estimators are thus functions of the unknown parameters (ρ, s_{11}, s_{22}),

[8] This is obvious if the regression parameters are known, but it can also be shown that the contribution to $\mathrm{Var}(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$ from uncertainty about the regression parameters is less for bivariate Model A than for the univariate Model U.

their dependence on (ρ, s_{11}, s_{22}) may not be very great over the range of the appreciable posterior density, since the sampling error variances for most states are much larger than the model error variances (at least for the CPS equation). We therefore assume that uncertainty about (ρ, s_{11}, s_{22}) produces little uncertainty about the regression parameters, so that $\hat{Y}_{1i}$ can be approximately regarded as given by equation (4) with the GLS estimators of the regression parameters taken as fixed. A similar analysis applies to $\hat{Y}_{2i}$.

With this assumption, and ignoring posterior covariance between c_{1i} and c_{2i}, we have the approximation (with an analogous expression for $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{2i}\,]$)

$$\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{1i}\,] \approx \mathrm{Var}(c_{1i}) \times [r_{1i}]^2 \;+\; \mathrm{Var}(c_{2i}) \times [r_{2i}]^2 \qquad (5)$$

where $r_{1i} = y_{1i} - \hat{\alpha}_1 - x_i'\hat{\beta}_1$ and $r_{2i} = y_{2i} - \hat{\alpha}_2 - x_i'\hat{\beta}_2$ are the GLS regression residuals for the CPS and ACS equations. While this is at best a crude approximation, equation (5) nonetheless suggests that $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{1i}\,]$ could be (*i*) large for both Model A and univariate Model U if the CPS equation regression residual, r_{1i}, is large, and (*ii*) large for Model A if the ACS equation residual, r_{2i}, is large. (Recall that $c_{2i} = 0$ in the CPS univariate model, so for this model the second term in (5) is zero.) Empirical results bear this out. We focus on the results for age 5-17 poverty ratios in IY 2002. First, we note that, for both the CPS and ACS equations, the residuals from Model A are approximately the same as the residuals from Model U. We now examine the results for three states: one for which r_{1i} is small but r_{2i} is large; one for which r_{1i} is large but r_{2i} is small; and one for which r_{1i} and r_{2i} are both somewhat large.[9]

- For Alaska, r_{1i} is small (std. res. = −.3) but r_{2i} is very large (std. res. = −3.1). This leads to very large values of $\mathrm{Var}_{\rho,s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{1i}\,]$ and $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{2i}\,]$ in the bivariate Model A, and to a large value for $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{2i}\,]$ in the univariate model (ACS equation). In the CPS univariate model for y_{1i}, $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{1i}\,]$ is small because r_{1i} is small and $c_{2i} = 0$.

- For Maine, r_{1i} is large (std. res. = 2.2) but r_{2i} is small (std. res. = .5). The large value of r_{1i} leads to large values of $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{1i}\,]$ in both the bivariate Model A and in the univariate model, and also to a moderately large value of $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{2i}\,]$ in the bivariate model. The small value of r_{2i} leads to a small value of $\mathrm{Var}_{\rho,s11,s22\,|\,\mathbf{y}}[\,\hat{Y}_{2i}\,]$ in the ACS univariate model (for which $c_{1i} = 0$.) The value of $\mathrm{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\,\hat{Y}_{2i}\,]$ in the bivariate model is not nearly as large

[9] We quote standardized residuals here, and use them subsequently in the plots, rather than the ordinary regression residuals, to clarify which of the residuals might be classified as somewhat extreme. The standardized residuals are defined as $r_{1i} / (\hat{s}_{11} + v_{1i})^{.5}$ and $r_{2i} / (\hat{s}_{22} + v_{2i})^{.5}$ where $\hat{s}_{11}$ and $\hat{s}_{22}$ are the posterior means of s_{11} and s_{22}. The use of standardized residuals does not change the results of the comparisons made here: plots using the ordinary residuals have similar appearances to those using the standardized residuals.

as are the values for Alaska and North Carolina, and it is not much larger than the values for some other states. This is because the estimate of $\hat{Y}_{2i}$ does not depend much on the CPS data y_{1i}, because y_{1i} has such high sampling error variance compared to the ACS data y_{2i}.

- For North Carolina, r_{1i} (std. res. = 1.4) and r_{2i} (std. res. = 2.1) are both somewhat large, though not as large in magnitude as are r_{1i} for Maine and r_{2i} for Alaska. The result for North Carolina is large values for $\text{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\hat{Y}_{1i}]$ and $\text{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\hat{Y}_{2i}]$ in both bivariate Model A and the univariate models. Actually, North Carolina's value for $\text{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\hat{Y}_{1i}]$ in the univariate model is smaller than those for a few other states (besides Maine), but these other states also have large values of r_{1i}.

To sum up the results for the posterior variances for Y_{1i} (CPS equation): For Alaska the posterior variance from the bivariate Model A ($\text{Var}(Y_{1i}\,|\,\mathbf{y}) = 1.24$) is much higher than the posterior variance from the univariate model ($\text{Var}(Y_{1i}\,|\,\mathbf{y}_1) = 0.81$) because the former is inflated due to the large value of r_{2i} but the latter is not. In contrast, for Maine the posterior variance for Y_{1i} from the bivariate Model A (1.04) is slightly lower than that from the univariate model (1.09), since r_{2i} is small but r_{1i} is large and inflates both posterior variances. Meanwhile, for North Carolina the somewhat large value of r_{1i} inflates both the bivariate and univariate model posterior variances for Y_{1i}, but the former is further inflated by the large value of r_{2i} (to $\text{Var}(Y_{1i}\,|\,\mathbf{y}) = 1.13$) while the latter is not ($\text{Var}(Y_{1i}\,|\,\mathbf{y}_1) = .96$).

We see that if r_{2i} is large then $\text{Var}_{\rho,\, s11,\, s22\,|\,\mathbf{y}}[\hat{Y}_{1i}]$ can be large for the bivariate Model A but not the univariate model, and it appears to be in these cases that $\text{Var}(Y_{1i}\,|\,\mathbf{y})$ from the bivariate Model A can be substantially larger than $\text{Var}(Y_{1i}\,|\,\mathbf{y}_1)$ from the univariate model. If r_{1i} is small this will accentuate the difference in posterior variances between the bivariate and univariate models, as was the case for the 5-17 poverty ratio estimate for Alaska in IY 2002.

To further illustrate how the potential for improvement of the posterior variance with a bivariate model depends on the regression residual from the ACS equation, Figures 2-4 plot, for IYs 2000-2002, percent differences of the bivariate Model A and univariate model posterior variances against ACS equation standardized residuals (Figures 2.a, 3.a, and 4.a) and squared standardized residuals (Figures 2.b, 3.b, and 4.b). In the plots, points below the horizontal dotted line are instances where the bivariate model A had a lower posterior variance, while points above the horizontal dotted line are instances where the bivariate model A had a higher posterior variance. In Figures 2.a, 3.a, and 4.a we see a quadratic pattern that might be expected from equation (5), if we also take into account that (*i*) differences in the first term of equation (2) involve expectations of $\text{Var}(Y_{1i}\,|\,\mathbf{y}, \rho, s_{11}, s_{22}) - \text{Var}(Y_{1i}\,|\,\mathbf{y}_1, \rho, s_{11}, s_{22})$, which does not depend on the ACS equation residuals, and (*ii*) both bivariate Model A and univariate model posterior variances may be similarly inflated by the first term in equation (5), which depends on the CPS equation residuals. We can also note in Figures 2.a, 3.a, and 4.a how the largest increases in posterior variances from use of bivariate Model A generally correspond to ACS equation standardized residuals exceeding 2 in absolute value. The plots in Figures 2.b, 3.b, and 4.b convey much the same information, but help clarify the relation by suggesting an approximate linear dependence of the posterior variance difference on the squared standardized residuals from the ACS equation.

3.5 Posterior Variance Comparisons for the ACS Equation of the Bivariate Models

To put the results of Sections 3.1 and 3.3 in context, Appendix B presents tables showing the average posterior means and variances from Models A and B3, as well as those from the univariate models, for the CPS and ACS equations. The tables also show the average direct survey estimates and their average sampling error variances. The latter are the average variances one would have from each data source if modeling was not used to improve the estimates. The tables show the substantially lower variances of the ACS direct estimates compared to those for CPS, due to the larger sample size of the ACS demonstration surveys. (This difference, of course, increases with the full production ACS data examined in Section 4.) They also show the results noted earlier that the univariate models achieve substantial variance reductions in the CPS equations compared to the direct estimates, and that further reductions under Models A and B3 are not as large.

We now focus on the posterior variances for the ACS equations. The tables in Appendix B show that the average posterior variance for the ACS equation from a given model is substantially lower than the corresponding average posterior variance for the CPS equation. This makes sense since the ACS model-based estimates start with the ACS direct estimates, which have lower variance than the CPS estimates, and we then use the model to improve them. In fact, in some cases (e.g., 5-17 in IY2001, and 18-64 and 65+ in IY2002) the direct ACS estimates have an average sampling variance comparable to or lower than that of the average posterior variance for the CPS univariate model.

Comparing now average posterior variances from the ACS univariate models to the average ACS sampling error variances, we see substantial average improvement from the univariate models. The percentage reductions in variance from modeling the ACS data (not shown) are less than those for modeling the CPS data, but are still substantial. This makes sense given the lower variances of the ACS direct estimates – there is less opportunity to improve them. Using bivariate Model A achieves generally negligible improvements in average posterior variances for the ACS equations (compared to the somewhat larger though still small average variance improvements from using bivariate Model A for the CPS equation that were discussed in Section 3.1). Using bivariate Model B3 also achieves little, if any, improvement in posterior variances for the ACS equation (in contrast with the results for the CPS equation, for which Model B3 seems more promising.) A general conclusion from these results is that, because of the higher variance of the CPS direct estimates, the bivariate models cannot usefully borrow information from them to improve estimates in the ACS equation.

3.6 Effects on Bivariate Model Results from Shifting the ACS Estimates by One Year

Appendix C provides tables comparing results (point estimates and posterior variances) for the CPS equation from the univariate model with two sets of results from the bivariate models: one set uses ACS data collected in the IY (these results are the same as in Appendix B); the other uses ACS data collected one year later. It makes sense to explore the latter alternative since the ACS data collected in IY+1 involve income reports for a period of almost two years from the start of IY to nearly the end of IY+1. Thus, for example, for IY 2000 we have previously been modeling ACS 2000[10] estimates with the CPS estimates for IY 2000, but we could alternatively use the ACS 2001 estimates, instead of the ACS 2000 estimates, in the model. Similarly, we have so far modeled ACS2001 estimates with CPS estimates for IY 2001, but we could use the ACS2002 estimates instead. The tables in Appendix C provide results from both these options for CPS equation estimates for IYs 2000 and 2001.

Comparisons of the CPS equation average point estimates for a given model in Appendix C show little difference depending on which year of ACS data is used. (Comparisons of individual state point estimates are made in the next section.). Comparisons of the average posterior variances for a given model show mostly small differences, and some of these for ages 0-4 and 5-17 are partly due to differences in the posterior distribution of the model parameters, particularly s_{11}.[11] In any case, the tables in Appendix C do not show a consistent pattern suggesting that shifting the ACS data used in the models by one year would yield consistently lower or higher posterior variances.

3.7 Comparing Individual State Poverty Ratio Point Estimates from Alternative Models

Results given in section 3.5 show that ACS poverty ratio estimates tend to be higher than CPS poverty ratio estimates, on average, and this difference persists when comparing ACS with CPS model-based estimates for a given model. On the other hand, the average estimates from the different models for a given equation, ACS or CPS, generally do not differ greatly, and are not too different from the average of the direct estimates. While these results on the average behavior are interesting, the question arises as to whether estimates for individual states may differ appreciably depending on the model chosen? We now turn to this question.

Figures 5, 6, 7, and 8 plot individual state poverty ratio estimates (for given CPS equations) from bivariate Models A and B3 against corresponding univariate model estimates for ages 0-4, 5-17, 18-64, and 65+, respectively, for IYs 2000, 2001, and 2002. Two sets of estimates are shown for the bivariate models for IYs 2000 and 2001 – one set from the models that use ACS data collected in the IY, and one set from the models that use ACS data collected in IY+1. On each page of plots those on the left are for bivariate Model A while those on the right are for bivariate Model B3, and those on top use ACS IY data while those on the bottom use ACS IY+1 data. (For IY 2002 only one set of estimates are plotted since, at the time of this work, for IY 2002 we had only results using models with ACS SS02 data.) Each plot also contains a y = x line to help judge how close the bivariate model estimates are to the univariate model estimates.

[10] The ACS demonstration survey conducted in 2000 was known as the Census 2000 Supplementary Survey, or C2SS.

[11] These effects observed for ages 0-4 and 5-17 can be seen by comparing posterior variances for the two versions of the univariate model for each age group, since these differences are due entirely to differences in the posterior distribution of the model parameters. For age groups 18-64 and 65+ the posterior distribution of s_{11} differs little depending on which two years of ACS data are used for all models considered.

Focusing first on the plots for bivariate Model A, we notice that the plotted points almost always cluster tightly around the y = x line, showing that use of Model A has little effect on the point estimates relative to those from the univariate model. (There is somewhat more variation for age 0-4 in IY 2000, but still not that much.) That this holds both for the plots of estimates from models using ACS IY data and those using ACS IY+1 data shows that also the choice of whether to use the ACS IY or ACS IY+1 estimates in the model makes little difference to the state point estimates.

With two qualifications and one major exception, the same conclusions can be drawn from the plots of point estimates from bivariate Model B3 against the univariate model estimates. One qualification is that in some cases the points appear to cluster about a line close to but different from y = x, e.g., with the Model B3 estimates higher than the univariate model estimates at the low end of poverty ratios, and with the reverse true at the high end. This is due to different regression coefficients being used in the CPS equation under Model B3. Since Model B3 assumes the same regression coefficients (apart from the intercepts) in both the CPS and ACS equations, and the ACS direct estimates have much lower sampling error than the CPS direct estimates, the estimated regression coefficients are weighted more strongly towards those from the ACS equation (e.g., those that would be obtained from the ACS equation under Model A). While this effect is not "statistically significant," as shown by the results of Section 3.2, it is enough to make the Model B3 point estimates appear to visibly shift relative to those from the univariate model. The second qualification is that in the plots of Model B3 versus univariate model estimates the points do not cluster as tightly about a line – there is more variation over states around whatever line would best fit the points plotted. Still, this variation is not very large.

The one major exception mentioned above is the point way above the y = x line in the Model B3 plots for age 0-4 for IY 2000, and the corresponding point above the y = x line among the high poverty ratio estimates for IY 2001. These points are for Washington, DC. The exact reason why these points stick out as outliers in the plots is unclear because similar results are not obtained for the other age groups or for Model A. However, we have observed that DC is unusual in comparison to the 50 states in regard to its regression variables – something perhaps not surprising given that, unlike any of the 50 states, DC is entirely an urban area. This distinction may have something to do with these results, but the results still deserve further scrutiny.

4. Empirical Model Comparisons Using Full Production ACS 2005 Data

In the previous sections, we used bivariate models incorporating both CPS data for IYs 2000-2002 and ACS demonstration surveys data for suitably comparable years. In this section, we present some empirical results from bivariate models using CPS data for IY 2004 and data from the full production 2005 ACS,[12] again making comparisons with corresponding results from univariate models. For ages 0-4, 5-17, and 18-64 the models involve two changes to the regression variables

[12] Notice that we have now shifted the ACS data ahead one year relative to the CPS IY, in contrast to most of the analyses of Section 3. From Section 3.6 this should have little effect on the results, and the ACS 2005 data are about as comparable to the CPS IY 2004 data as would be the ACS 2004 data. One reason for this shift is that the CPS estimates for IY and the ACS IY+1 data are released at about the same time, namely August of IY+1.

from those used before: addition of the state food stamp participation rate, and replacement of the Census 2000 state poverty ratio estimates for the age group with corresponding "census residuals." (These changes were mentioned in footnote 2 of Section 2.1.) For age 65+ the models remained the same as before. Our goal in the analyses of this section is to assess the benefits to using a bivariate model combining CPS and full production ACS data over the corresponding univariate models. Recall that the full production ACS survey has a much larger sample size (3 million addresses) than the pre-2005 ACS demonstration surveys (700,000 – 800,000 addresses). The direct variance estimates of the state age-group poverty rates of the full production 2005 ACS are about half of those from the pre-2005 ACS demonstration surveys. For the large states these full production ACS variances are quite small, implying very reliable direct estimates (of Y_{2i}).

4.1 Posterior Variance Comparisons for Model A – CPS equation

The posterior means and standard deviations of the model parameters (ρ, $s_{11,}$ s_{22}) from the Gibbs sampling via WinBUGs based on 10,000 simulations from bivariate Models A and B3, and of the model parameters ($s_{11,}$ s_{22}) from Model U (where ρ is set to 0), for IY 2004, are shown in Tables 4.1.a – 4.1.c, respectively. As with the earlier results using ACS demonstation survey data, the standard deviations for s_{11} are not small relative to their posterior means, reflecting considerable uncertainty in this parameter. Note that the posterior standard deviations for s_{22} from these models using ACS 2005 data are smaller than those from the models using ACS demonstration survey data (in Tables 2.1 – 2.3). The posterior standard deviations for ρ, however, are still substantial, and are not necessarily smaller than those for IYs 2000 – 2002. Notice also that the posterior means of ρ are fairly close to zero except for age 65+, which would suggest that we will get little improvement from use of the bivariate model in IY 2004 for the first three age groups.

Table 4.1.a Posterior means and standard deviations of the parameters of Model A (IY 2004)

age	0 - 4	5 - 17	18-64	65+
ρ	-0.01 (0.46)	0.17 (0.37)	0.03 (0.28)	0.50 (0.37)
s_{11}	2.50 (2.26)	1.64 (1.17)	0.74 (0.41)	0.64 (0.55)
s_{22}	1.51 (0.60)	0.60 (0.25)	0.26 (0.08)	0.38 (0.13)

Table 4.1.b Posterior means and standard deviations of the parameters of Model B3 (IY 2004)

age	0 - 4	5 - 17	18-64	65+
ρ	0.05 (0.48)	0.15 (0.36)	-0.00 (0.28)	0.56 (0.34)
s_{11}	1.93 (1.85)	1.71 (1.22)	0.75 (0.42)	0.66 (0.51)
s_{22}	1.46 (0.57)	0.61 (0.25)	0.26 (0.08)	0.38 (0.12)

Table 4.1.c Posterior means and standard deviations of the parameters of Model U (IY 2004)

age	0 - 4	5 - 17	18-64	65+
s_{11}	2.91 (2.36)	1.76 (1.17)	0.78 (0.42)	0.60 (0.54)
s_{22}	1.49 (0.60)	0.59 (0.24)	0.26 (0.08)	0.38 (0.12)

Table 4.2 summarizes the comparisons of the posterior variances of the state poverty ratios Y_{1i} (CPS equation) from Model A with those from the univariate model using the relative percentage differences as defined in Section 3. We see, for all age groups for IY 2004, only small improvements in posterior variances on average from use of the bivariate Model A. The min values show that some states show somewhat larger variance reductions than others, though still not very much. The max values show that some states show variance increases, though we do not see any extremely large increases of the sort observed for a few cases in Section 3.4. Appendix D shows more detail from these results, presenting the frequency distribution of the percentage differences in posterior variances of between Model A and Model U. The table in Appendix D shows that substantial variance increases for states are relatively rare, while small to moderate variance reductions predominate for IY 2004.

Table 4.2 Relative percent differences of posterior variances : Model A versus univariate model (means are unweighted averages of percent differences across the 50 states and DC)

IY 2004

Age	Mean	Min	Max
0-4	−8.1	−15.8	10.0
5-17	−4.8	−12.0	9.8
18-64	−2.3	−6.5	17.6
65+	−5.2	−17.1	24.6

The max relative percent difference of posterior variance in Table 4.2 occurs for age 65+ for Rhode Island. It shows a 24.6 percent increase in posterior variance, with standardized residuals of -2.29 in the ACS equation and -0.93 in the CPS equation. For age 18-64, the max relative percent variance increase of Model A versus the univariate model is 17.6 percent for NY, with a standardized residual of -2.61 in the ACS equation.

4.2 Chi-Squared Tests of Model Restrictions

Section 3.2 presented results for IYs 2000 – 2002 on chi-squared tests of model restrictions corresponding to the alternative bivariate models B2 and B3. We applied the same procedures to the CPS IY 2004 and ACS 2005 data. Recall that we are testing the following null hypotheses:

H2: $\alpha_1 = \alpha_2, \quad \beta_1 = \beta_2$ (Model B2)
H3: $\beta_1 = \beta_2$ (Model B3)

Because the food stamp participation rate has been added as a regression variable to the models for ages 0-4, 5-17, and 18-64, the tests of H3 now have four degrees of freedom, and the tests of H2 have five degrees of freedom, for all four age groups. The five percent critical values from the chi-squared distribution are 9.5 and 11.1 for four and five degrees of freedom, respectively.

The results of the Chi-squared tests for IY 2004 are given in Table 4.3. Values that are significant at the five percent level are shown in bold.

Table 4.3 Chi-Squared statistics for testing hypotheses H2 and H3 for IY 2004

Age	0-4	5-17	18-64	65+
H2	5.5	**24.6**	**19.3**	2.5
H3	0.2	6.7	5.3	2.2

For IY 2004, we reject the hypothesis H2 for ages 5-17 and 18-64 but not for 0-4 and 65+. The latter results do differ some from those obtained earlier, which showed most tests of H2 at least close to significant at the 5 percent level. Still, the significant results for ages 5-17 and 18-64 suggest rejecting model B2. The tests of H3, however, are all insignificant, consistent with the earlier results, again suggesting that we might assume regression coefficients other than the intercepts to be the same in both equations.

4.3 Posterior Variance Comparisons for Model B3 – CPS Equation

Posterior variances for Model B3 were computed as discussed in Section 2.3, i.e., using equations (1) and (2) with the simulations of (ρ, s_{11}, s_{22}) obtained under Model B3, and with $E(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$ and $Var(Y_{1i} \mid \mathbf{y}, \rho, s_{11}, s_{22})$ computed to account for the Model B3 restriction, $\beta_1 = \beta_2$. Table 4.4 presents summaries of the relative percent differences of the resulting posterior variances of Model B3 from those for the univariate model; these results can be compared to those of Tables 3.4 and 4.2. Doing so we see that the average variance reductions from Model B3 are again substantially larger than those from Model A. Also, Model B3 avoids the large variance increase for Rhode Island for age 65+ that occurred with Model A. Compared to the results for IYs 2000 – 2002, the average posterior variance reductions are greater for ages 0-4 and 65+, and not as large for ages 5-17 and 18-64. Overall, though, the results seem reasonably consistent with those for the earlier years, noting that no large posterior variance increases happened to occur in IY 2004.

Table 4.4 Relative percent differences of posterior variances: Model B3 versus univariate model (means are unweighted averages of percent differences across the 50 states and DC)

IY 2004

Age	Mean	Min	Max
0-4	−38.9	−66.0	−12.9
5-17	−16.2	−50.8	10.4
18-64	−10.2	−40.8	14.4
65+	−32.5	−63.6	1.4

4.4 Posterior Variances for the ACS Equation of the Bivariate Models

To put the results of Sections 4.1 and 4.3 in context, Appendix E presents a table showing the average posterior means and variances from Models A and B3, as well as those from the univariate model, for the CPS IY 2004 and ACS 2005 equations. The table also shows the average variances of the direct survey estimates; these are the average variances one has from each data source if modeling is not used to improve the estimates. The table shows the substantially lower variances of the full production ACS 2005 direct estimates compared to those for the CPS IY 2004 direct estimates, due to the much larger sample size of the 2005 ACS. Also, if we compare the average of the ACS 2005 direct variance estimates in Appendix E with corresponding entries for 2000-2002 from the tables in Appendix B, we see that the ACS 2005 direct variance estimates for poverty rates are about half or less than those from the ACS demonstration surveys because of the much larger full production sample size (3 million compared to 700,000 addresses).

We now focus on the posterior variances for the ACS equations. The table in Appendix E shows that the average posterior variance for the ACS equation from a given model is substantially lower than the corresponding average posterior variance for the CPS equation for each age group. This is because the ACS model-based estimates start with the ACS direct estimates, which have much lower variance than the CPS direct estimates, and then use the model to improve them. In fact, for all four age groups, the average sampling variance of the direct ACS 2005 estimates is lower than the average posterior variance of the univariate modeled estimates from the CPS equation for IY 2004.

Comparing average posterior variances from the 2005 ACS univariate models to the average 2005 ACS sampling error variances, we still see substantial improvement from modeling in terms of percentage variance reduction. (Note, though, that in some cases, particularly for age 18-64, the average variance of the direct ACS estimates might be regarded as sufficiently small as to not be in need of reduction.) Using either bivariate Model A or B3 achieves negligible (or no) improvements in average posterior variances for the ACS equations, as was the case with the ACS demonstration survey data.

Results in Appendix E also show that the ACS 2005 poverty rate estimates[13] tend to be higher than the CPS IY 2004 poverty ratio estimates, on average, and this difference persists when comparing ACS with CPS model-based estimates for a given model (except for age 65+). On the other hand, the average estimates from the different models for a given equation, ACS or CPS, generally do not differ greatly, and are not too different from the average of the direct estimates.

5. Summary and Conclusions

Our goal in this paper was to examine the potential benefits of jointly modeling state poverty ratio estimates from the CPS ASEC and ACS (demonstration surveys or full production) to improve model-based estimates of poverty ratios. We focused primarily on results for the CPS equation, but also examined results for the ACS equation. We examined alternative bivariate models for doing this using CPS estimates for IYs 2000 – 2002 with ACS data from the ACS demonstration surveys of these years, and also using CPS estimates for IY 2004 with full production ACS 2005 data. The models included regression variables constructed from administrative records data and Census 2000 poverty ratios (or Census 2000 residuals for IY 2004), along with state random effects and sampling error components. We can summarize the results from our empirical study as follows:

1. In comparing the general bivariate Model A with the CPS univariate model, there is, at best, a small average improvement in state posterior variances (of Y_{1i}) for most age groups and income years. Results are variable for individual states, with some states showing larger improvements, but a few showing substantial variance increases. Overall, these results do not suggest worthwhile benefits from using a bivariate model to bring in the ACS data.

2. For many of the combinations of age-group and IY, chi-squared tests rejected the restricted bivariate Models B2 and B1, giving evidence that these models should not be used. The chi-squared tests failed to reject Model B3[14], however, for all four age groups and all the IYs considered. This suggests that one might try using Model B3 to obtain reductions in posterior variance relative to the univariate model.

[13] When this analysis was done we had available ACS state poverty rate estimates but not ACS state poverty ratio estimates. The former are uniformly larger than the latter: the numerators of the two are identical but the denominator of the poverty rate is the poverty universe, which excludes some persons and thus is smaller than the total age-group population, thus making the poverty rate larger than the poverty ratio. This difference is small, however, compared to the inherent difference between the CPS and ACS estimates.

[14] Recall from Section 2.2 that Model B1 assumes CPS and ACS estimate the same thing ($Y_{1i} = Y_{2i}$), Model B2 assumes all regression parameters are the same between the CPS and ACS equations, and Model B3 assumes that regression parameters apart from the intercept are the same between the CPS and ACS equations ($\beta_1 = \beta_2$).

3. In fact, Model B3 had substantially lower posterior variances (for the CPS equation), on average, than did the general bivariate Model A. Model B3 did not, however, eliminate the occasional large posterior variance increases found for Model A. Also, in order to decide to use Model B3, we had to estimate Model A and test whether $\beta_1 = \beta_2$. Quoting posterior variances as if Model B3 is known to be true thus understates statistical uncertainty.

4. The occasional moderate to large posterior variance increases from use of bivariate Models A or B3 correspond to large regression residuals from the ACS equation.

5. Use of either of the bivariate models makes no improvements of substance in estimates for the ACS equation. There is too much sampling error in the CPS direct estimates for them to convey much useful information for estimation in the ACS equation.

6. Use of bivariate Model A seems to produce point estimates very similar to those from the univariate model (both CPS and ACS equations). Use of bivariate Model B3 produces slightly different point estimates for the CPS equation, due to obtaining somewhat different estimates of the regression parameters, but these differences are not very large (and are not statistically significant according to the chi-squared tests of Section 3.2 and 4.2).

7. Full production 2005 ACS state estimates have substantially lower sampling error variances than the estimates from the ACS demonstration surveys. However, the empirical results for the CPS equation using full production ACS 2005 data lead to similar conclusions as with using ACS demonstration survey data in regard to potential benefits from using these data in a bivariate model.

8. The average direct sampling variance of the full production ACS 2005 estimates is about half of the average posterior variance of the CPS univariate model-based estimates for all four age groups.

9. The average posterior variance of the univariate model-based estimate from the ACS 2005 equation is about half or less of the average sampling variance of the ACS 2005 direct estimates for all four age groups.

These conclusions bring out two trade-offs involved in a decision of whether or not to use the ACS data to attempt to improve estimates from the CPS equation via a bivariate model:

- Use of bivariate Model A can be expected to yield, at best, only small average improvements in posterior variance, and this is at the expense of a potential for occasional large posterior variance increases (when the regression residual in the ACS equation is large.)

- Use of bivariate Model B3 can be expected to yield larger apparent improvements in posterior variance, but use of Model B3 assuming it to be true may understate the statistical uncertainty. Also, use of Model B3 does not eliminate the potential for large posterior variance increases.

Potential resolutions of these two trade-offs can be considered:

- Results from the ACS equation could be examined for large regression residuals. When these occur the estimates for the corresponding states could be obtained from the CPS equation univariate model rather than from a bivariate model. There are some concerns as to whether this is really a "statistically principled" approach. It is perhaps easiest to justify if the instances of use of the univariate model are rare, corresponding only to distinct outliers in the ACS equation. The logical extension of this idea – compute posterior variances under both the univariate and bivariate model and for each state choose the model that produces the lowest posterior variance – seems definitely *not* statistically principled.

  As a more sophisticated generalization of the above approach, one could consider using a model assuming a long-tailed distribution for one of the random errors (model error or sampling error) in the ACS or CPS equation in an attempt to deal with potential outliers. Huang and Bell (2006) pursued this approach using a t-distribution and had some limited success with one of the examples considered here – the large standardized residual for age 5-17 for Alaska in IY 2002.

- Rather than assume bivariate Model B3 to be exactly true, we might put an informative prior distribution centered around 0 on the differences of the regression coefficients between the CPS and ACS equations (excepting the intercepts for the two equations, which would still be given flat priors). This might reap some of the benefits of Model B3 without making the assumption that the regression coefficients are exactly the same in the two equations.

References

Bell, William R (2000), "Models for County and State Poverty Estimates," appendix to *Small-Area Estimates of School-Age Children in Poverty – Evaluation of Current Methodology*, Panel on Estimates of Poverty for Small Geographic Areas, eds. Constance F. Citro and Graham Kalton, Committee on National Statistics, Washington D.C.:National Academy Press.

Bell, William R (1999), "Accounting for Uncertainty About Variances in Small Area Estimation," *Bulletin of the International Statistical Institute*, 52nd Session, Helsinki, available at www.census.gov/hhes/www/saipe under "Publications."

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

Fay, Robert E. and Herriot, Roger A. (1979), "Estimates of Income for Small Places: An Application of James-Stein Procedures to Census Data," *Journal of the American Statistical Association*, **74**, 269-277.

Fay, Robert E. and Train, George F. (1997). "Small Domain Methodology for Estimating Income and Poverty Characteristics for States in 1993," *American Statistical Association, Proceedings of the Social Statistical Section*, 183-188.

Huang, Elizabeth T. and Bell, William R. (2004), "An Empirical Study on Using ACS Supplementary Survey Data in SAIPE State Poverty Models," *2004 Proceedings of the American Statistical Association, Survey Research Methods Section* [CD-ROM], Alexandria, VA: American Statistical Association, 3677-3684.

Huang, Elizabeth T. and Bell, William R. (2006), "Using the t-Distribution in Small Area Estimation: An Application to SAIPE State Poverty Models," *2006 Proceedings of the American Statistical Association, Survey Research Methods Section* [CD-ROM], Alexandria, VA: American Statistical Association, 3142-3149.

Nelson, Charles (2006), "What Do We Know About Differences Between CPS and ACS Income and Poverty Estimates?" U.S. Census Bureau working paper, available at http://www.census.gov/hhes/www/poverty/nelson_082906.pdf.

Otto, Mark C. and Bell, William R. (1995), "Sampling Error Modelling of Poverty and Income Statistics for States," *American Statistical Association, Proceedings of the Section on Government Statistics*, 160-165.

Spiegelhalter, D., Thomas, A., Best, N. and Lunn, D. (2003), *WinBUGS Users Manual, Version 1.4*, MRC Biostatistics Unit, Institute of Public Health, Robinson Way, Cambridge CB2 2SR, UK.

Appendix A: Frequency distributions over states of the relative percentage differences of the state posterior variances (CPS eq.) from bivariate Model A and the univariate model

Table A.1: CPS IY 2000 (with ACS 2000)

Percentage Difference	age 0-4	age 5-17	age 18-64	age 65+
-20 to -15	0	0	0	0
-15 to -10	1	2	14	10
-10 to -5	4	15	18	18
-5 to 0	13	23	14	15
0 to 5	23	7	0	3
5 to 10	6	3	2	3
10 to 15	1	0	1	1
15 to 20	0	1	2	1
20 to 25	1	0	0	0
25 to 30	2	0	0	0

Table A.2: CPS IY 2001 (with ACS 2001)

Percentage Difference	age 0-4	age 5-17	age 18-64	age 65+
-20 to -15	0	0	0	0
-15 to -10	3	9	1	2
-10 to -5	19	21	15	18
-5 to 0	17	16	18	24
0 to 5	9	3	4	4
5 to 10	0	2	7	1
10 to 15	1	0	2	1
15 to 20	0	0	0	1
20 to 25	1	0	0	0
25 to 30	1	0	0	0
30 to 35	0	0	1	0
35 to 40	0	0	2	0
40 to 45	0	0	1	0

Table A.3: CPS IY 2002 (with ACS 2002)

Percentage Difference	age 0-4	age 5-17	age 18-64	age 65+
-25 to -20	0	0	2	0
-20 to -15	0	0	12	0
-15 to -10	1	3	18	0
-10 to -5	11	24	12	20
-5 to 0	29	20	5	27
0 to 5	10	2	1	3
5 to 10	0	0	1	0
10 to 15	0	0	0	1
15 to 20	0	1	0	0
20 to 25	0	1	0	0
25 to 30	0	0	0	0
30 to 35	0	0	0	0
35 to 40	0	0	0	0
40 to 45	0	0	0	0
45 to 50	0	0	0	0
50 to 55	0	1	0	0
55 to 60	0	0	0	0

Appendix B: Comparing point estimates and corresponding (error) variances of state poverty ratios from direct and alternative model-based estimates using CPS (IYs 2000-2002) and ACS (2000-2002) demonstration survey data (entries are unweighted averages across the 50 states and DC)

Age 0-4

	2000		2001		2002	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
CPS equation						
Direct estimate	17.70	11.79	17.89	11.63	18.43	12.00
Univariate Model	17.84	2.73	17.80	2.29	18.31	3.25
Bivariate Model A	17.83	2.80	17.79	2.24	18.33	3.16
Bivariate Model B3	17.80	2.29	17.64	1.96	18.08	2.85
ACS equation						
Direct estimate	18.72	4.08	18.17	3.48	19.34	3.99
Univariate Model	18.77	1.49	18.09	1.16	19.27	1.20
Bivariate Model A	18.75	1.49	18.09	1.15	19.27	1.18
Bivariate Model B3	18.73	1.49	18.07	1.15	19.29	1.14

Age 5-17

	2000		2001		2002	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
CPS equation						
Direct estimate	13.22	4.91	13.52	4.96	14.29	5.15
Univariate Model	13.29	1.00	13.53	1.54	14.27	1.05
Bivariate Model A	13.29	0.97	13.54	1.47	14.27	1.01
Bivariate Model B3	13.35	0.73	13.51	1.19	14.30	0.81
ACS equation						
Direct estimate	15.54	1.83	15.26	1.52	15.61	1.59
Univariate Model	15.45	0.76	15.04	0.57	15.60	0.56
Bivariate Model A	15.46	0.76	15.05	0.56	15.60	0.56
Bivariate Model B3	15.45	0.71	15.05	0.54	15.61	0.54

Age 18-64

	2000		2001		2002	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
CPS equation						
Direct estimate	9.29	1.14	9.95	1.19	10.36	1.08
Univariate Model	9.36	0.26	9.94	0.21	10.37	0.32
Bivariate Model A	9.37	0.25	9.95	0.22	10.39	0.29
Bivariate Model B3	9.41	0.19	9.95	0.16	10.43	0.25
ACS equation						
Direct estimate	10.67	0.36	10.54	0.31	10.85	0.34
Univariate Model	10.62	0.20	10.49	0.16	10.82	0.20
Bivariate Model A	10.61	0.20	10.49	0.16	10.82	0.20
Bivariate Model B3	10.60	0.19	10.49	0.16	10.80	0.20

Age 65+

	2000		2001		2002	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
CPS equation						
Direct estimate	10.03	3.99	10.24	4.05	10.43	4.16
Univariate Model	10.04	0.94	10.25	0.68	10.48	1.31
Bivariate Model A	10.04	0.90	10.25	0.66	10.48	1.26
Bivariate Model B3	10.00	0.70	10.23	0.39	10.54	1.00
ACS equation						
Direct estimate	10.73	1.16	10.30	0.75	9.82	0.83
Univariate Model	10.62	0.39	10.27	0.27	9.69	0.22
Bivariate Model A	10.62	0.40	10.27	0.27	9.68	0.21
Bivariate Model B3	10.62	0.38	10.27	0.25	9.68	0.22

Appendix C: Comparing point estimates and corresponding error variances of CPS equation state poverty ratios from alternative model-based estimates using ACS demonstration survey data for the IY and shifted one year (entries are unweighted averages across the 50 states and DC)

Age 0-4

Model (CPS and ACS eqs.)	CPS IY 2000 Poverty Ratio	CPS IY 2000 Variance Estimate	CPS IY 2001 Poverty Ratio	CPS IY 2001 Variance Estimate
with ACS data from:	**ACS 2000**		**ACS 2001**	
Univariate Models	17.84	2.73	17.80	2.29
Bivariate Model A	17.83	2.80	17.79	2.24
Bivariate Model B3	17.80	2.29	17.64	1.96
with ACS data from:	**ACS 2001**		**ACS 2002**	
Univariate Models	17.84	2.71	17.80	2.34
Bivariate Model A	17.83	2.76	17.79	2.20
Bivariate Model B3	17.82	2.16	17.64	1.79

Age 5-17

Model (CPS and ACS eqs.)	CPS IY 2000 Poverty Ratio	CPS IY 2000 Variance Estimate	CPS IY 2001 Poverty Ratio	CPS IY 2001 Variance Estimate
with ACS data from:	**ACS 2000**		**ACS 2001**	
Univariate Models	13.29	1.00	13.53	1.54
Bivariate Model A	13.29	0.97	13.54	1.47
Bivariate Model B3	13.35	0.73	13.51	1.19
with ACS data from:	**ACS 2001**		**ACS 2002**	
Univariate Models	13.29	1.02	13.53	1.55
Bivariate Model A	13.30	1.05	13.54	1.47
Bivariate Model B3	13.32	0.84	13.54	1.17

Age 18-64

Model (CPS and ACS eqs.)	CPS IY 2000		CPS IY 2001	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
with ACS data from:	**ACS 2000**		**ACS 2001**	
Univariate Models	**9.36**	**0.26**	**9.94**	**0.21**
Bivariate Model A	**9.37**	**0.25**	**9.95**	**0.22**
Bivariate Model B3	**9.41**	**0.19**	**9.95**	**0.16**
with ACS data from:	**ACS 2001**		**ACS 2002**	
Univariate Models	**9.36**	**0.26**	**9.94**	**0.22**
Bivariate Model A	**9.38**	**0.24**	**9.94**	**0.20**
Bivariate Model B3	**9.42**	**0.20**	**9.94**	**0.15**

Age 65+

Model (CPS and ACS eqs.)	CPS IY 2000		CPS IY 2001	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
with ACS data from:	**ACS 2000**		**ACS 2001**	
Univariate Models	**10.04**	**0.94**	**10.25**	**0.68**
Bivariate Model A	**10.04**	**0.90**	**10.25**	**0.66**
Bivariate Model B3	**10.00**	**0.70**	**10.23**	**0.39**
with ACS data from:	**ACS 2001**		**ACS 2002**	
Univariate Models	**10.04**	**0.95**	**10.25**	**0.68**
Bivariate Model A	**10.03**	**0.91**	**10.25**	**0.66**
Bivariate Model B3	**10.00**	**0.63**	**10.23**	**0.39**

Appendix D: Frequency distributions over states of the relative percentage differences of the state posterior variances (CPS eq.) from bivariate Model A and the univariate model

Table D: CPS IY 2004 (with ACS 2005)

Percentage Difference	age 0-4	age 5-17	age 18-64	age 65+
-25 to -20	0	0	0	0
-20 to -15	3	0	0	3
-15 to -10	21	1	0	14
-10 to -5	15	31	10	12
-5 to 0	6	14	33	16
0 to 5	3	3	6	3
5 to 10	2	2	1	1
10 to 15	1	0	0	0
15 to 20	0	0	1	1
20 to 25	0	0	0	1
25 to 30	0	0	0	0
30 to 35	0	0	0	0
35 to 40	0	0	0	0
40 to 45	0	0	0	0
45 to 50	0	0	0	0
50 to 55	0	0	0	0
55 to 60	0	0	0	0

Appendix E: Comparing point estimates and corresponding (error) variances of state poverty ratios from direct and alternative model-based estimates using CPS (IY 2004) and ACS 2005 survey data (entries are unweighted averages across the 50 states and DC)

	Age 0-4		Age5-17R		Age 18-64		Age 65+	
	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate	Poverty Ratio	Variance Estimate
CPS equation								
Direct estimate	19.09	13.27	14.23	4.95	10.84	1.27	9.81	4.37
Univariate Model	19.53	3.27	14.50	1.59	10.93	0.54	9.86	0.83
Bivariate Model A	19.55	3.02	14.51	1.51	10.94	0.52	9.84	0.79
Bivariate Model B3	19.62	1.91	14.65	1.29	10.99	0.47	9.72	0.53
ACS equation								
Direct estimate	20.90	1.80	16.93	0.80	11.77	0.12	9.94	0.32
Univariate Model	20.91	0.77	16.20	0.31	11.78	0.07	9.88	0.15
Bivariate Model A	20.91	0.77	16.20	0.31	11.78	0.07	9.87	0.15
Bivariate Model B3	20.90	0.75	16.19	0.31	11.77	0.07	9.88	0.15

Figure 1: Posterior Densities for Model A Parameters - Income Year 2000 - Age 5-17

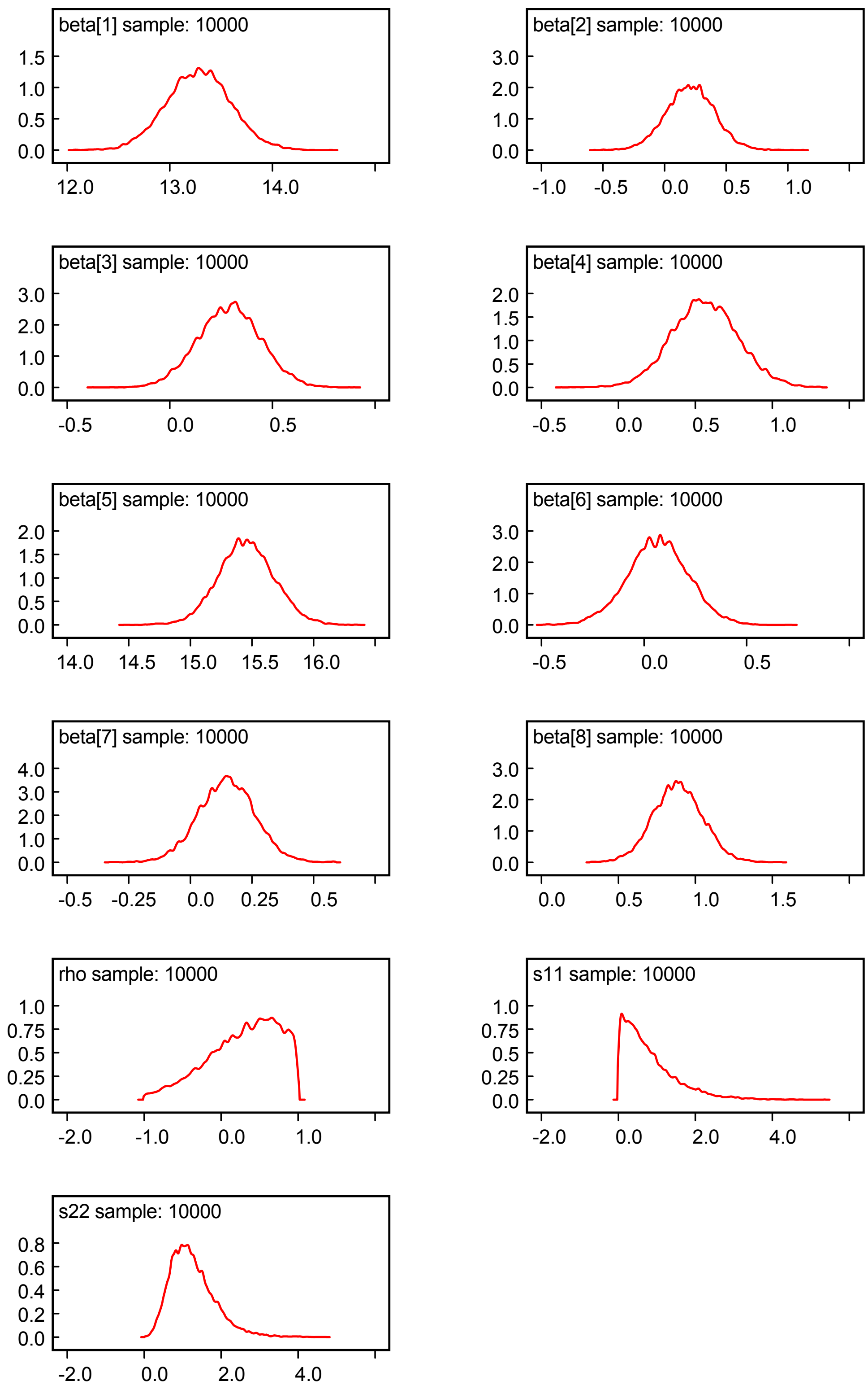


Figure 2. Posterior Variance Percent Differences, IY 2000

a. Var % diff for models A and U vs ACS std residuals

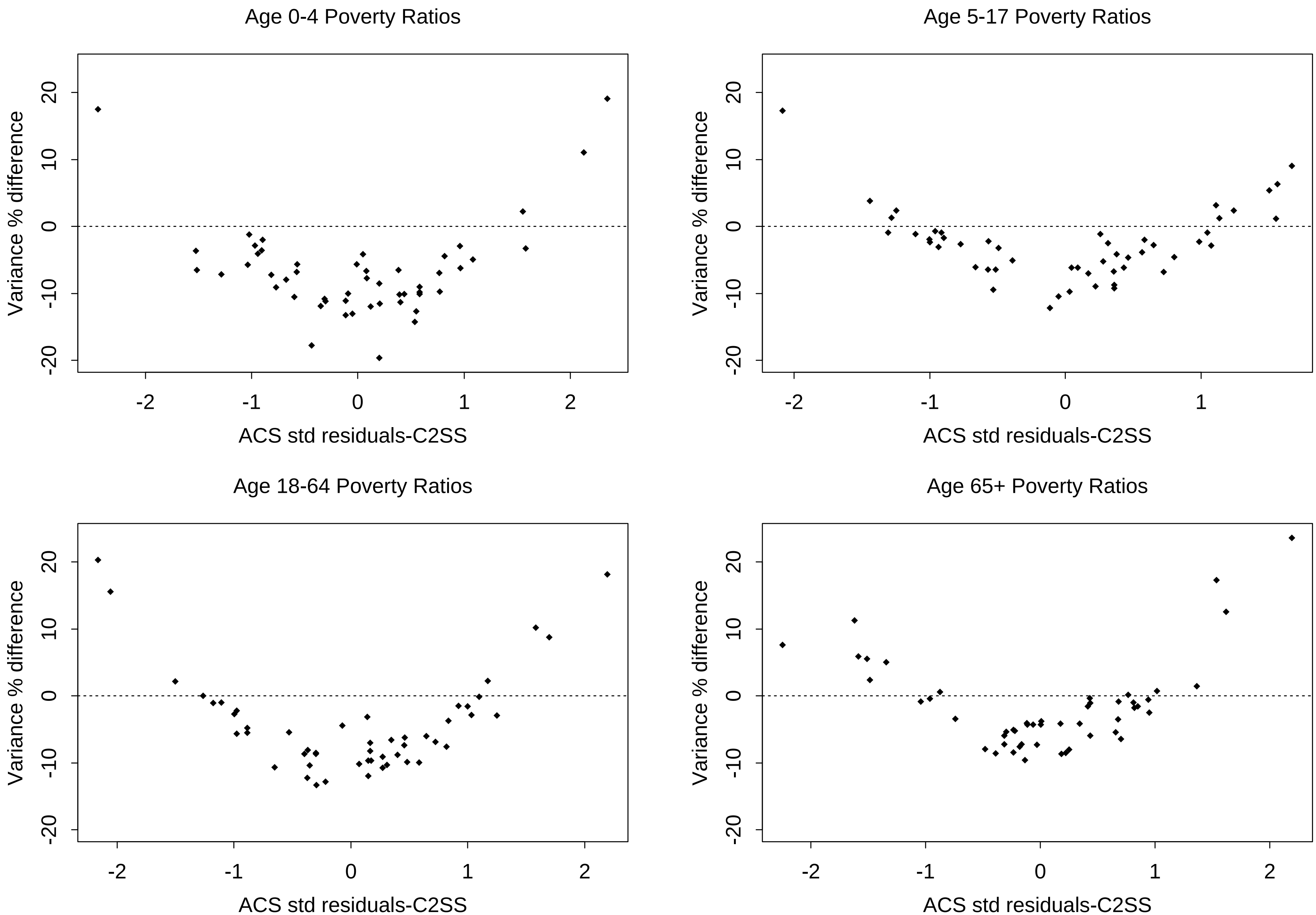


Figure 2. Posterior Variance Percent Differences, IY 2000

b. Var % diff for models A and U vs ACS squared std residuals

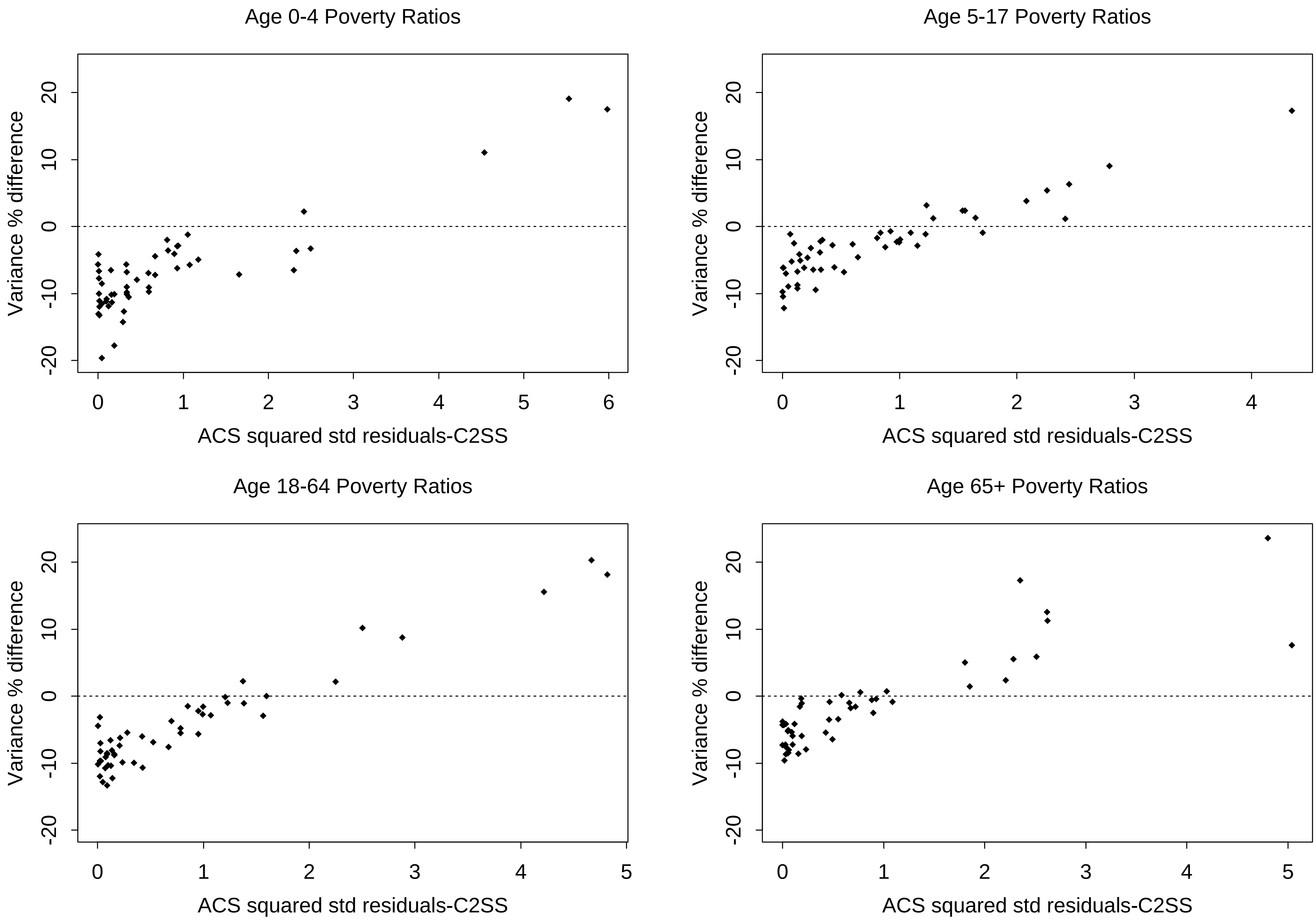


Figure 3. Posterior Variance Percent Differences, IY 2001

a. Var % diff for models A and U vs ACS std residuals

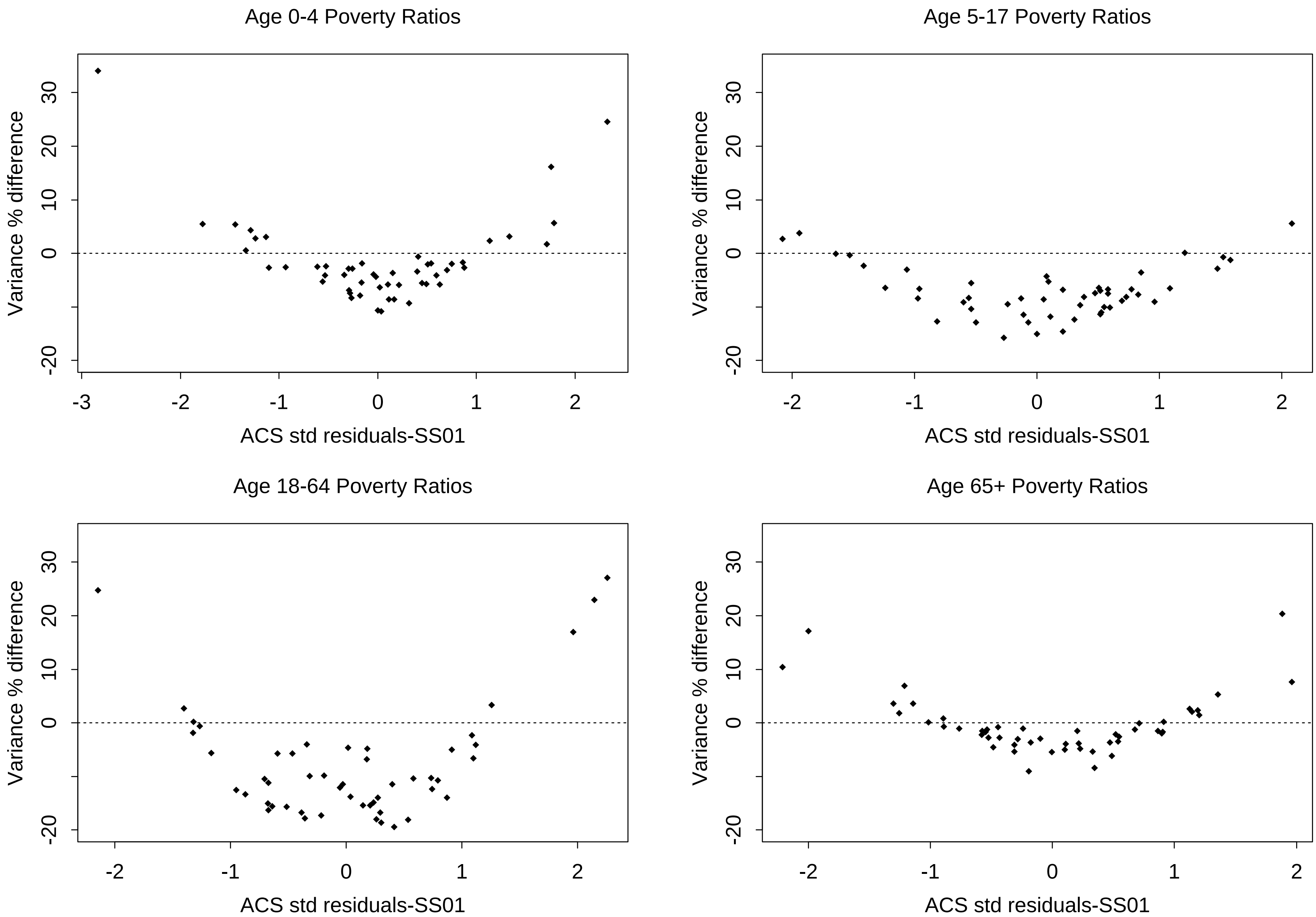


Figure 3. Posterior Variance Percent Differences, IY 2001

b. Var % diff for models A and U vs ACS squared std residuals

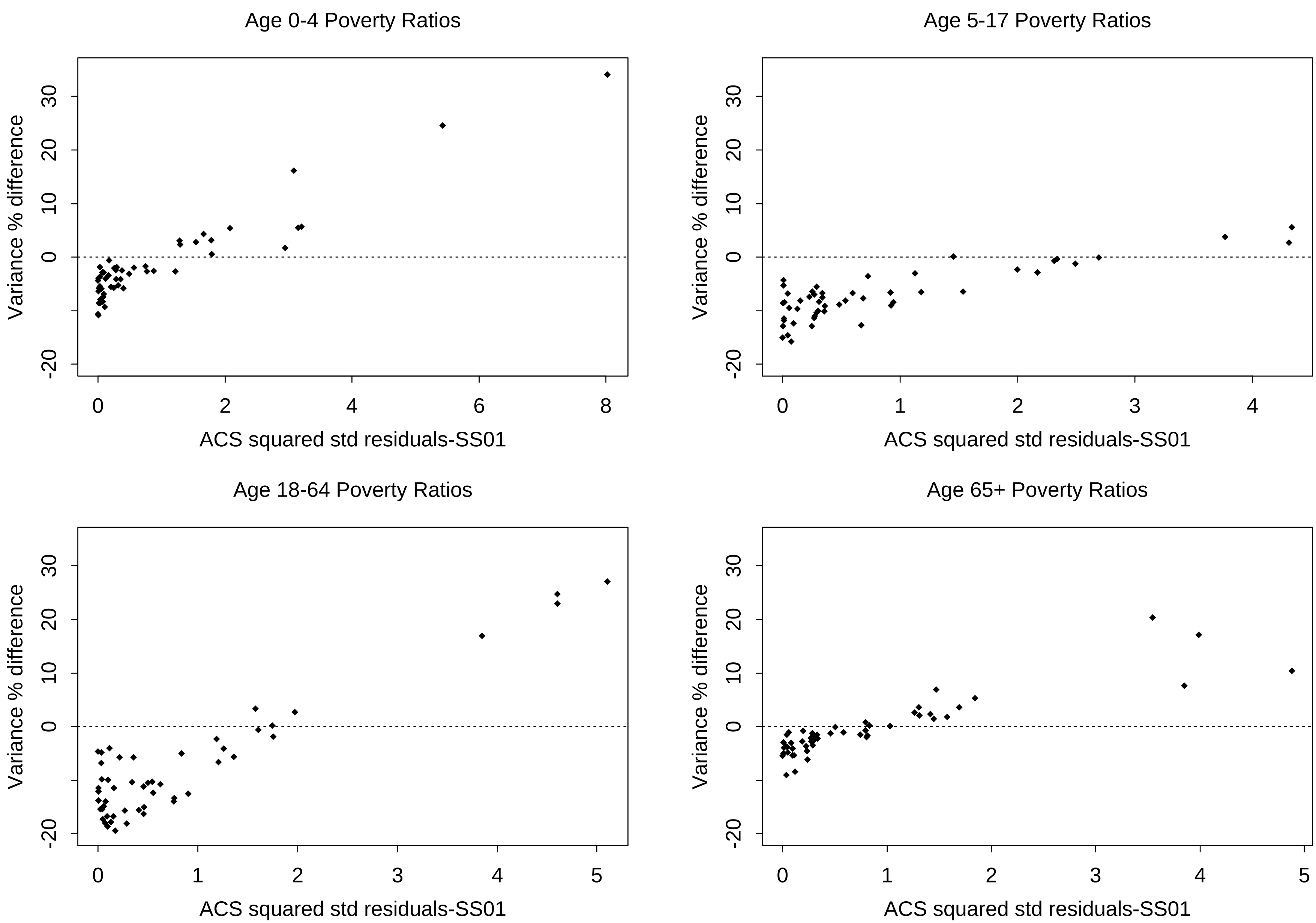


Figure 4. Posterior Variance Percent Differences, IY 2002

a. Var % diff for models A and U vs ACS std residuals

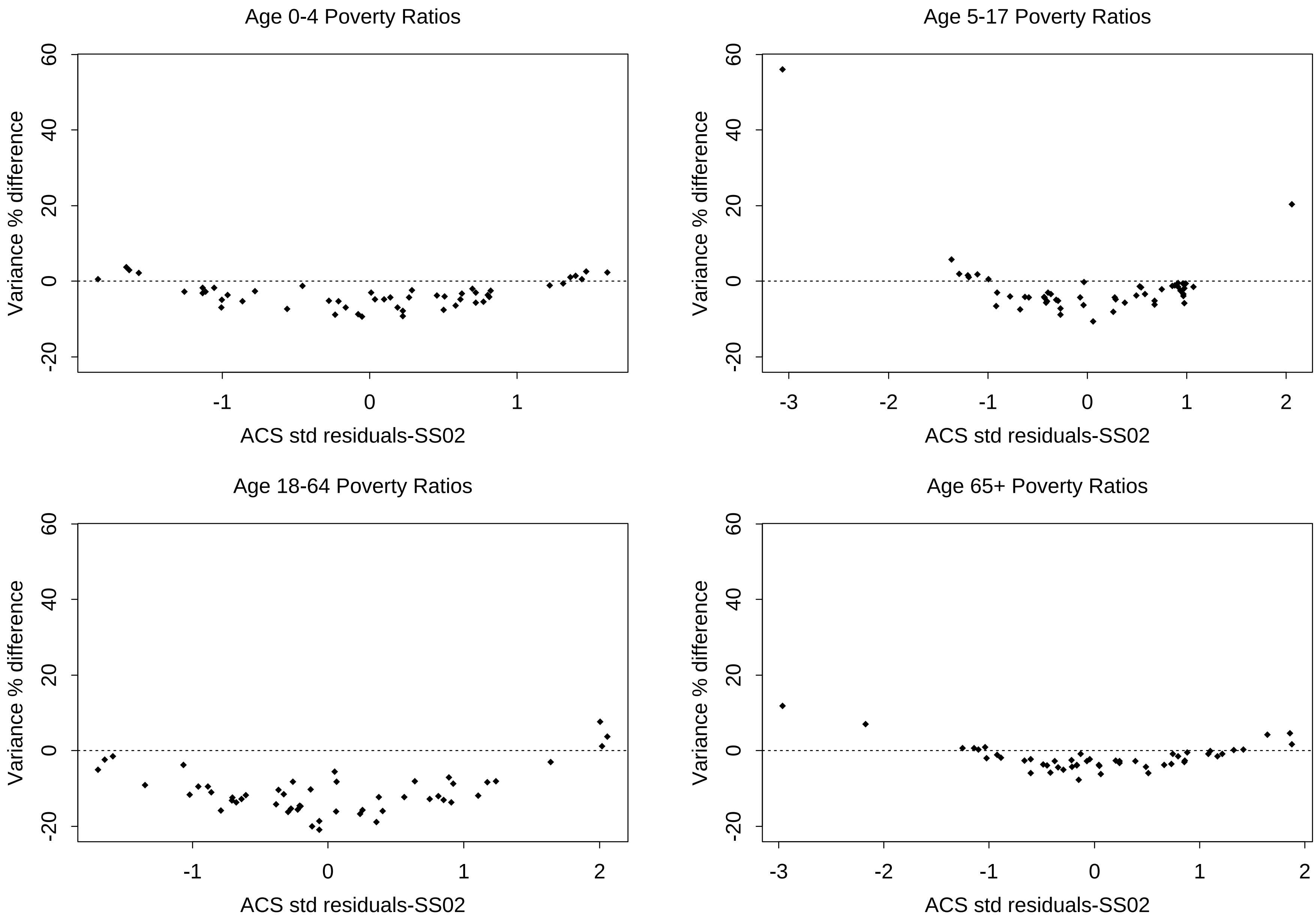


Figure 4. Posterior Variance Percent Differences, IY 2002

b. Var % diff for models A and U vs ACS squared std residuals

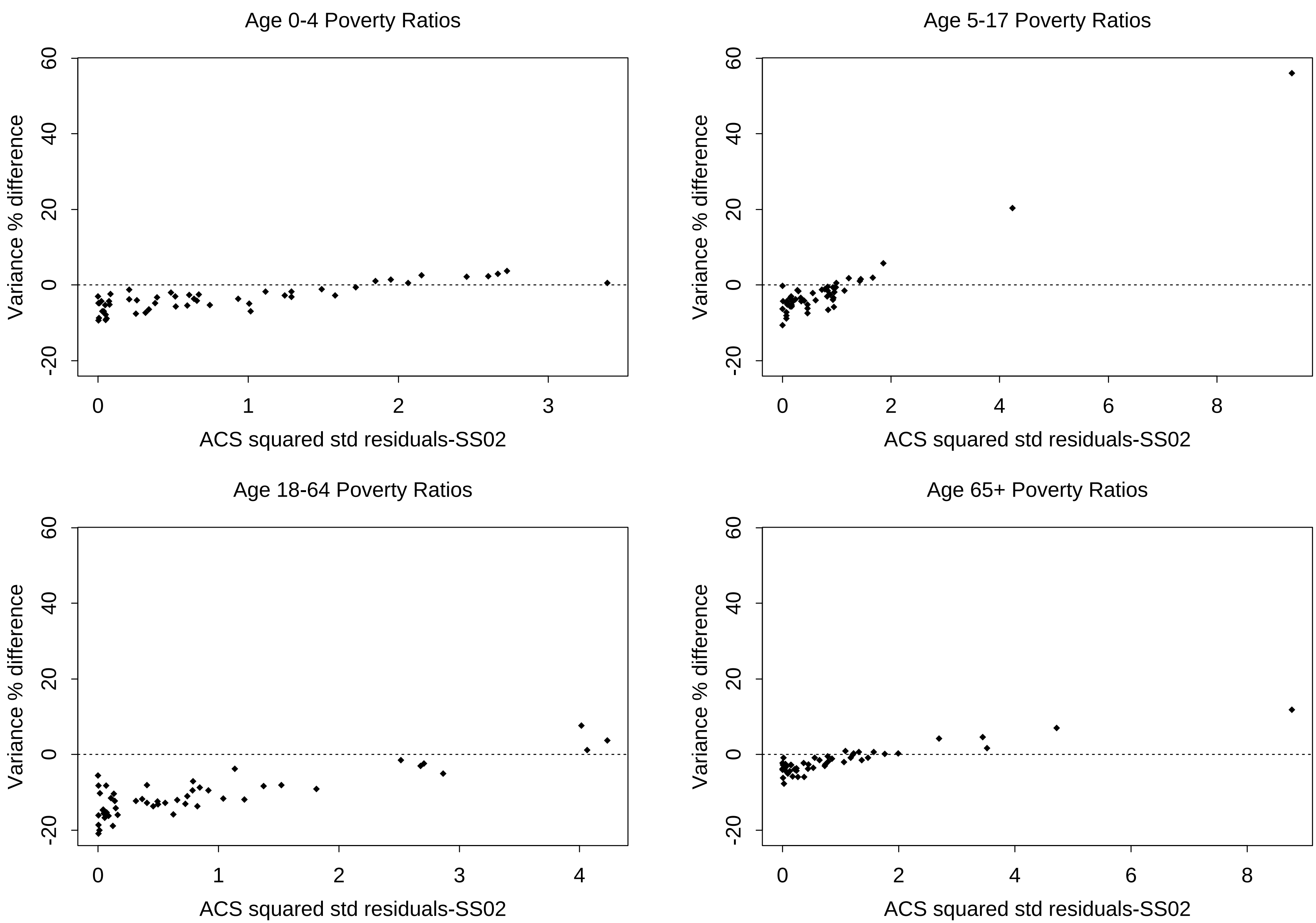


Figure 5. Poverty Ratio Estimates: Age 0-4

a. IY2000 CPS equation alternative vs univariate estimates

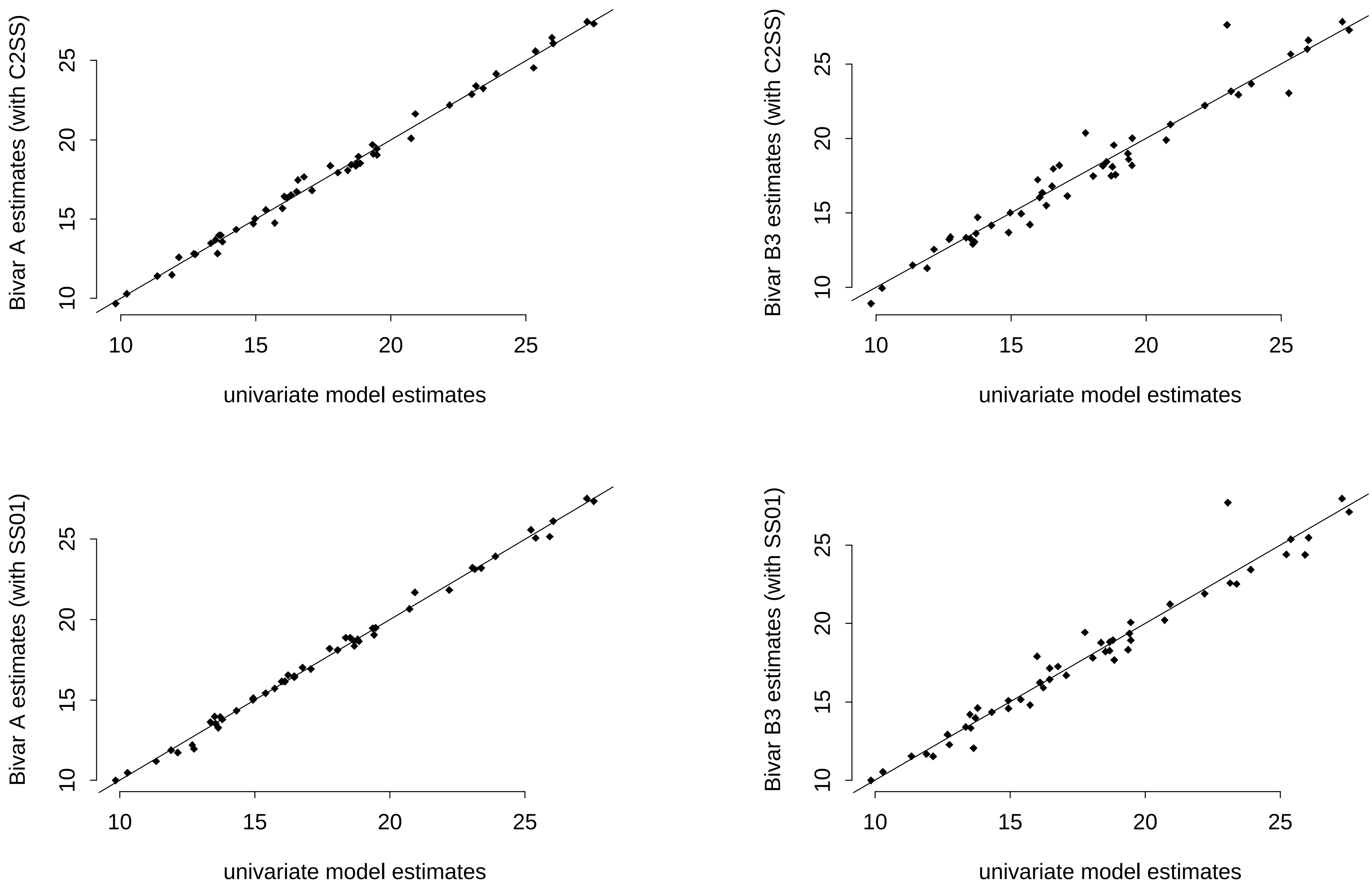


Figure 5. Poverty Ratio Estimates: Age 0-4

b. IY2001 CPS equation alternative vs univariate estimates

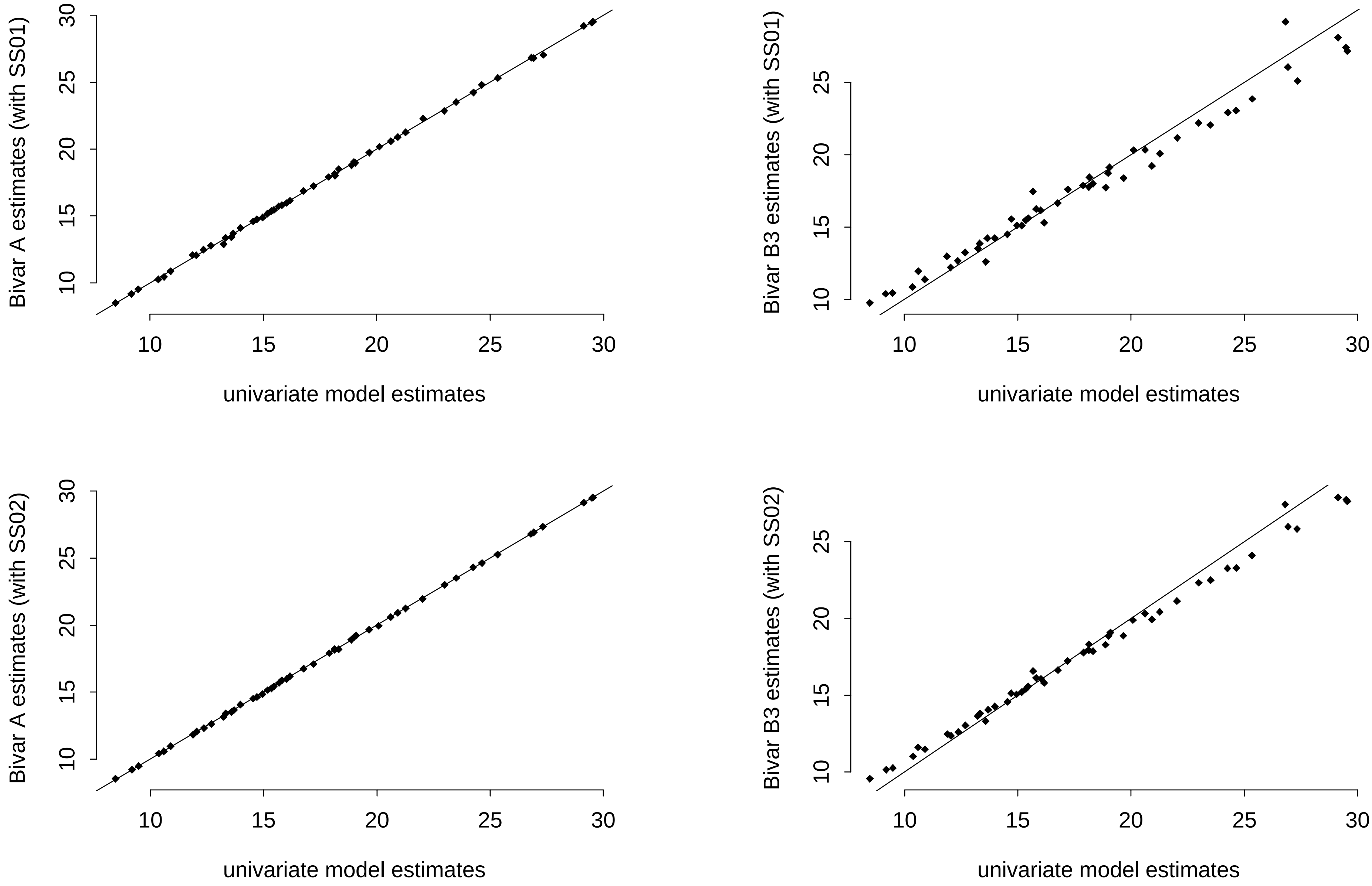
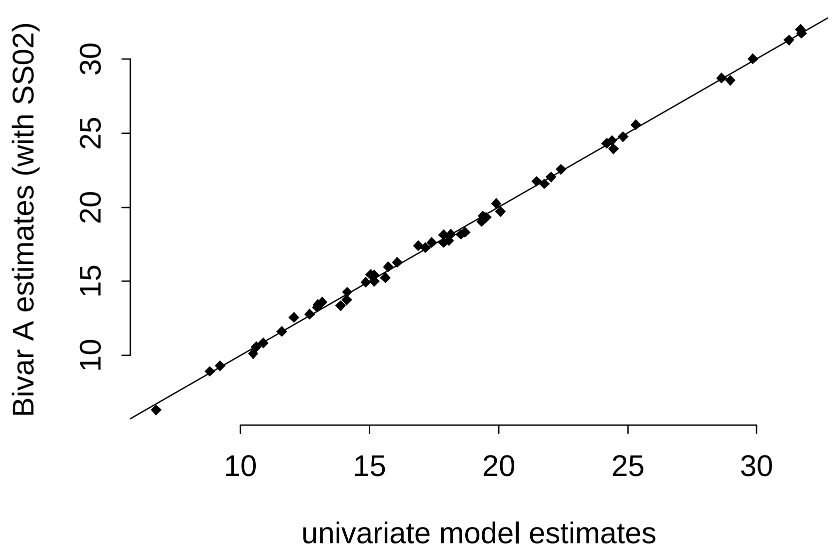


Figure 5. Poverty Ratio Estimates: Age 0-4

c. IY2002 CPS equation alternative vs univariate estimates



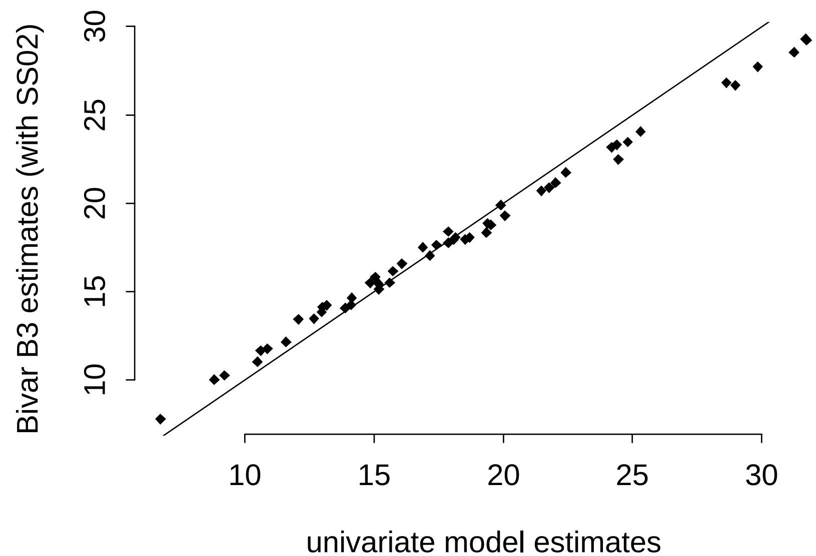


Figure 6. Poverty Ratio Estimates: Age 5-17

a. IY2000 CPS equation alternative vs univariate estimates

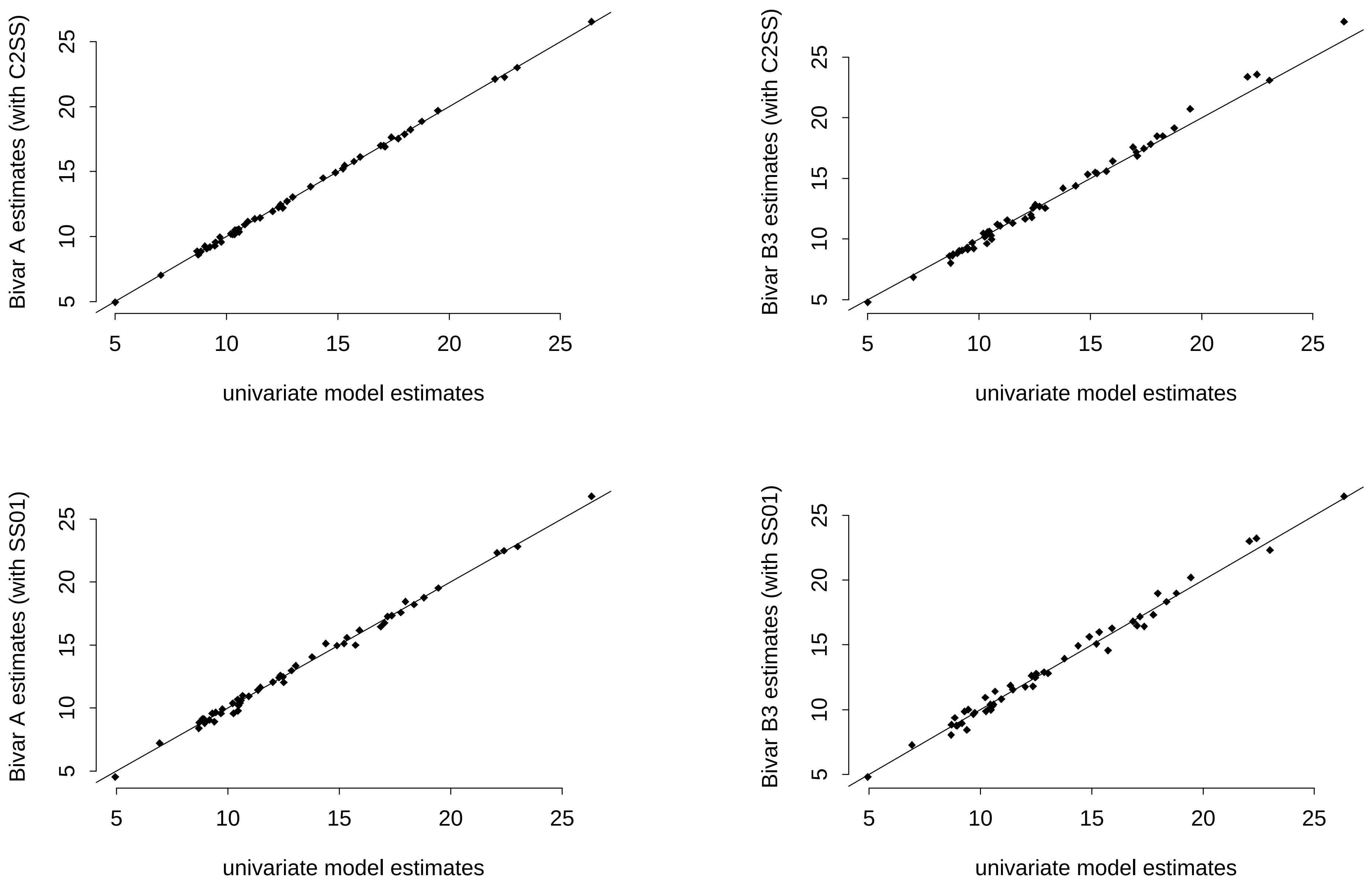


Figure 6. Poverty Ratio Estimates: Age 5-17

b. IY2001 CPS equation alternative vs univariate estimates

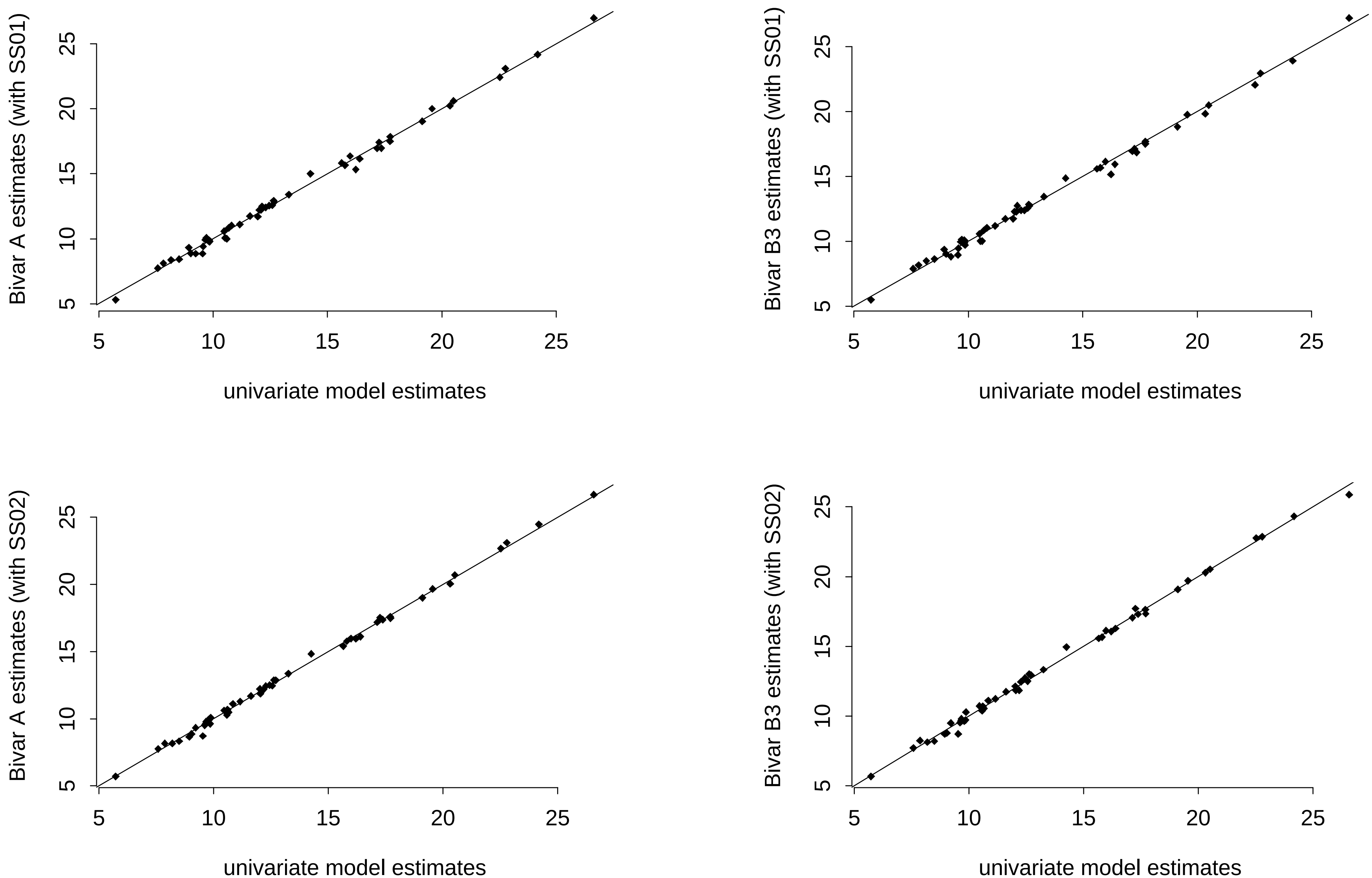


Figure 6. Poverty Ratio Estimates: Age 5-17

c. IY2002 CPS equation alternative vs univariate estimates

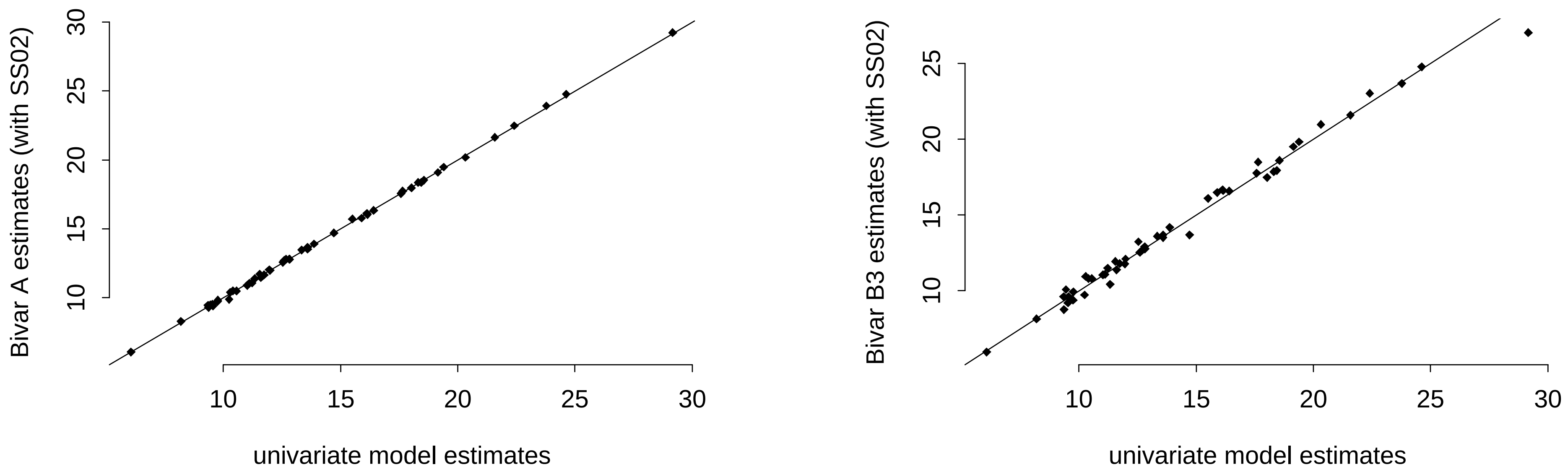


Figure 7. Poverty Ratio Estimates: Age 18-64

a. IY2000 CPS equation alternative vs univariate estimates

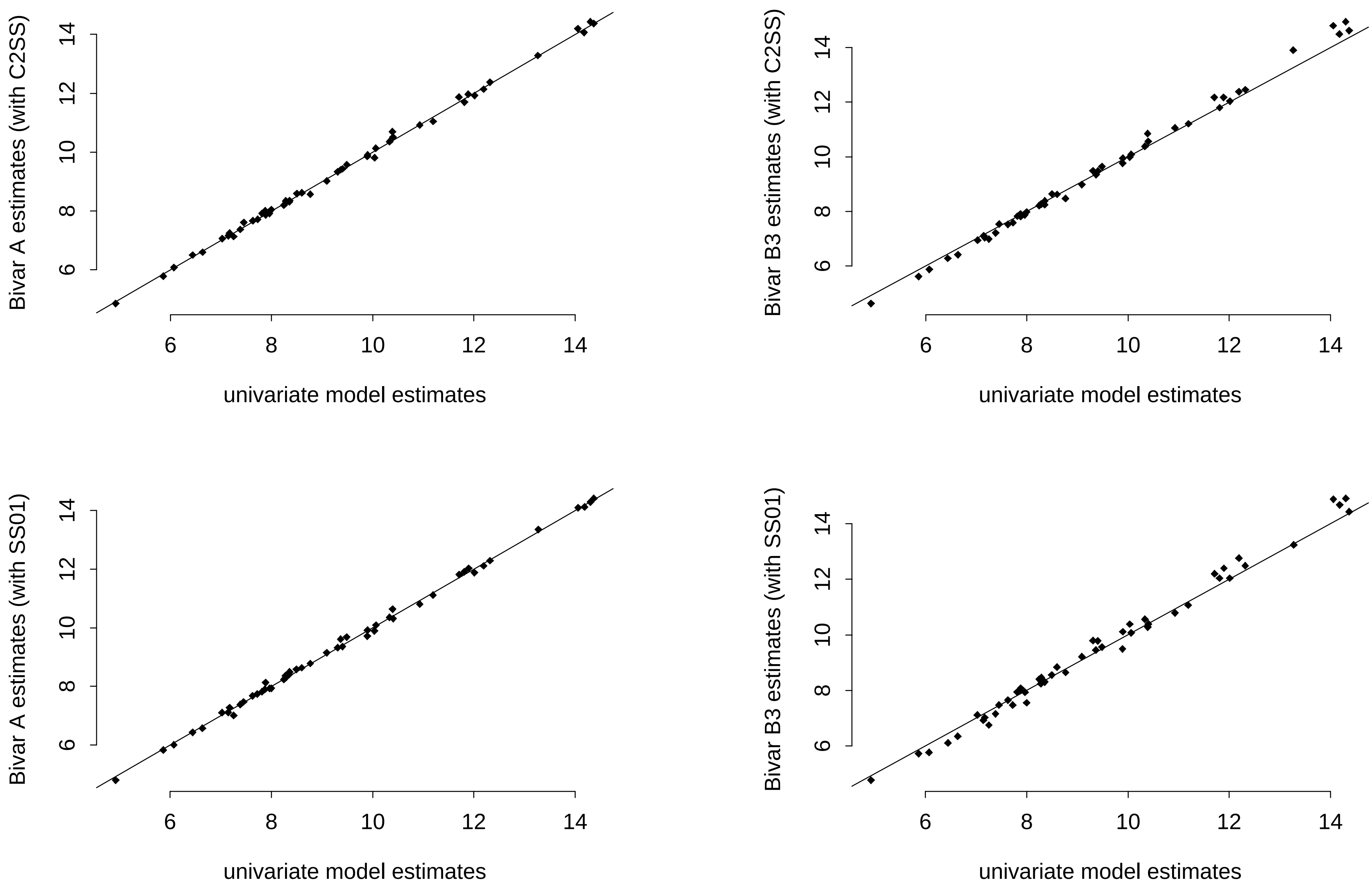


Figure 7. Poverty Ratio Estimates: Age 18-64

b. IY2001 CPS equation alternative vs univariate estimates

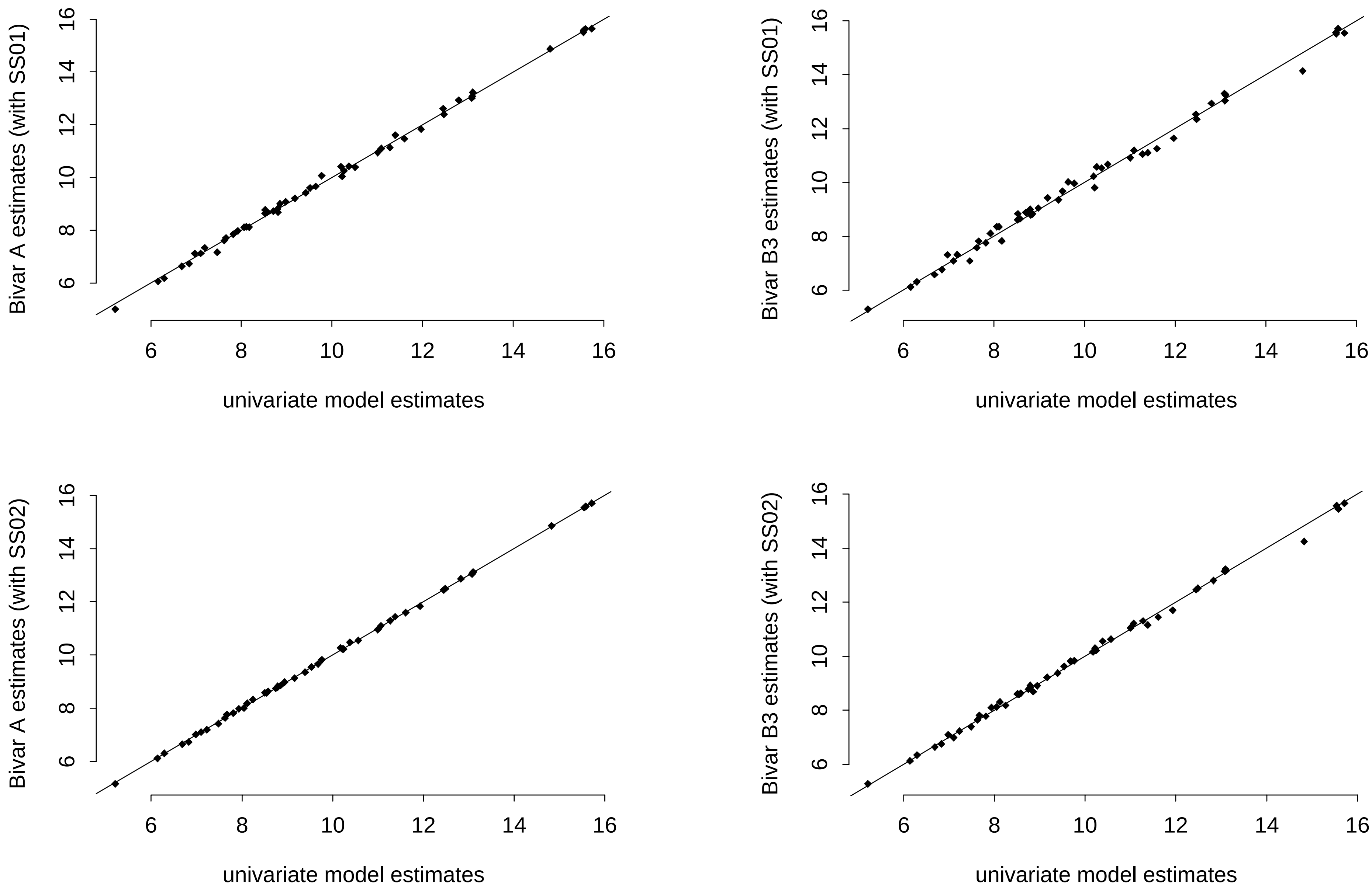


Figure 7. Poverty Ratio Estimates: Age 18-64

c. IY2002 CPS equation alternative vs univariate estimates

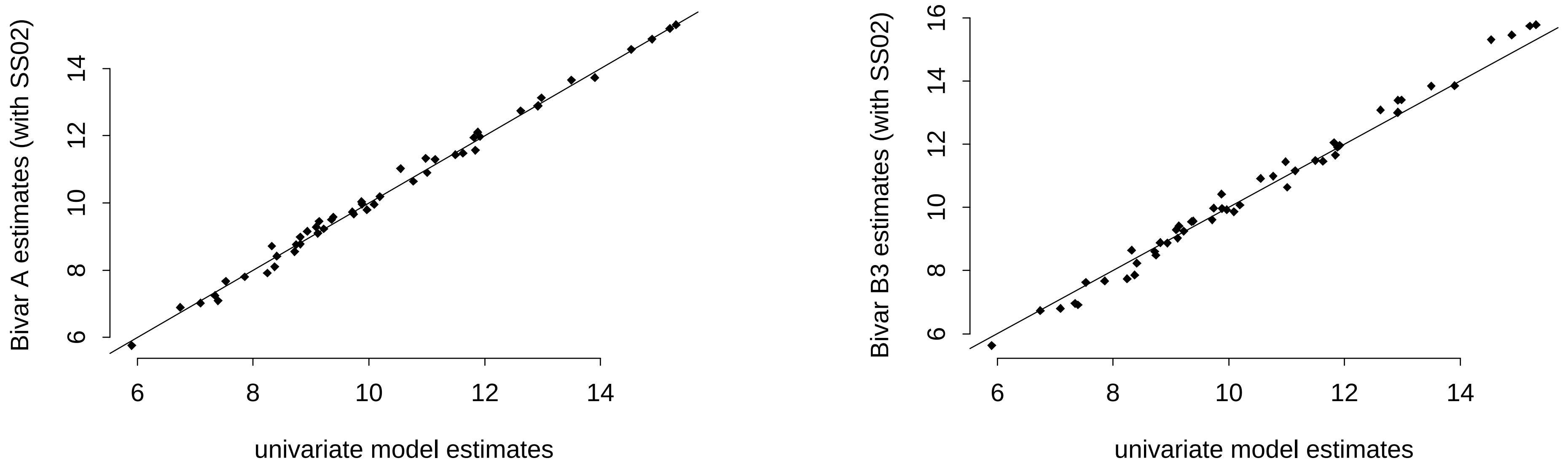


Figure 8. Poverty Ratio Estimates: Age 65+

a. IY2000 CPS equation alternative vs univariate estimates

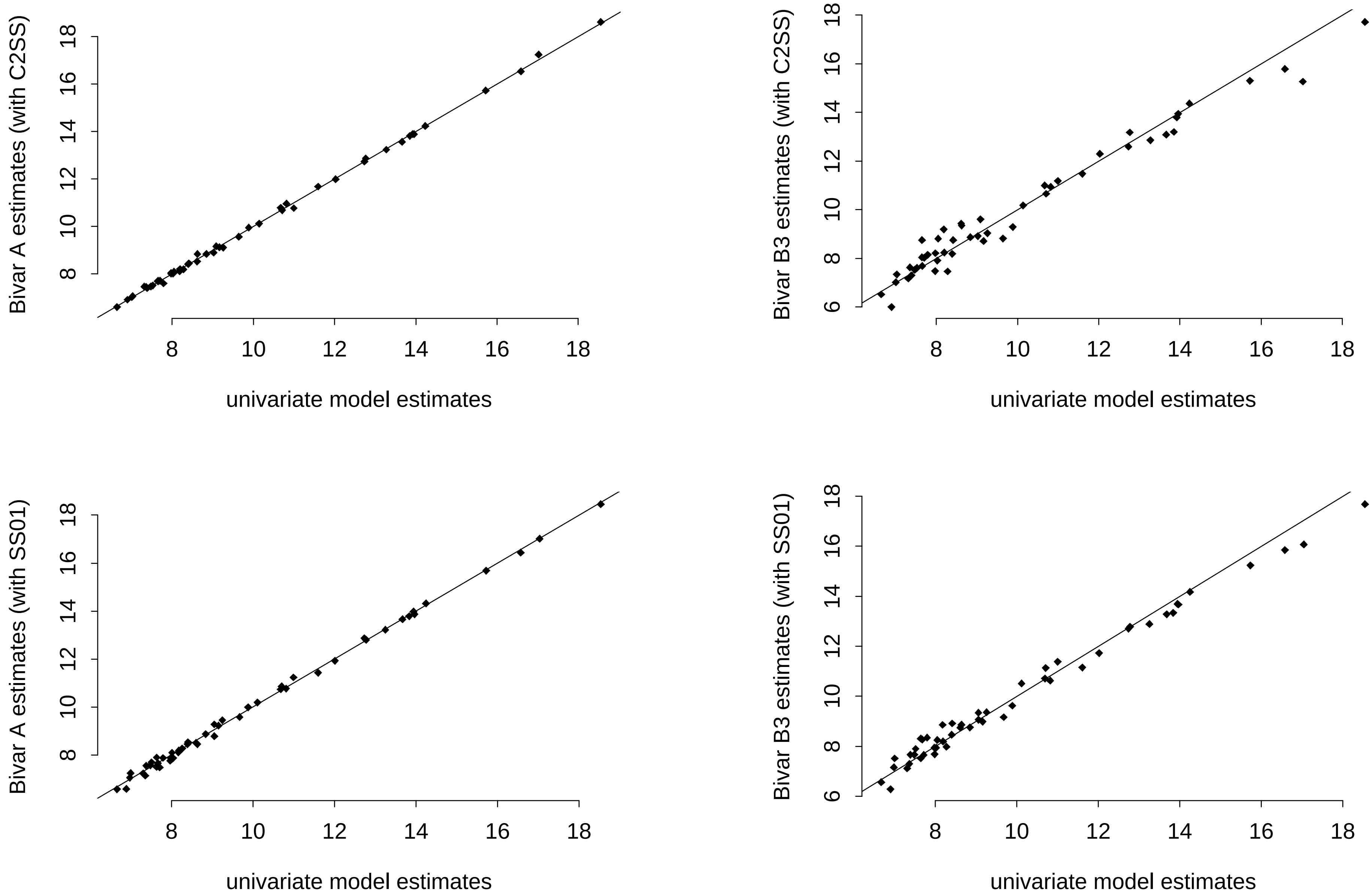


Figure 8. Poverty Ratio Estimates: Age 65+

b. IY2001 CPS equation alternative vs univariate estimates

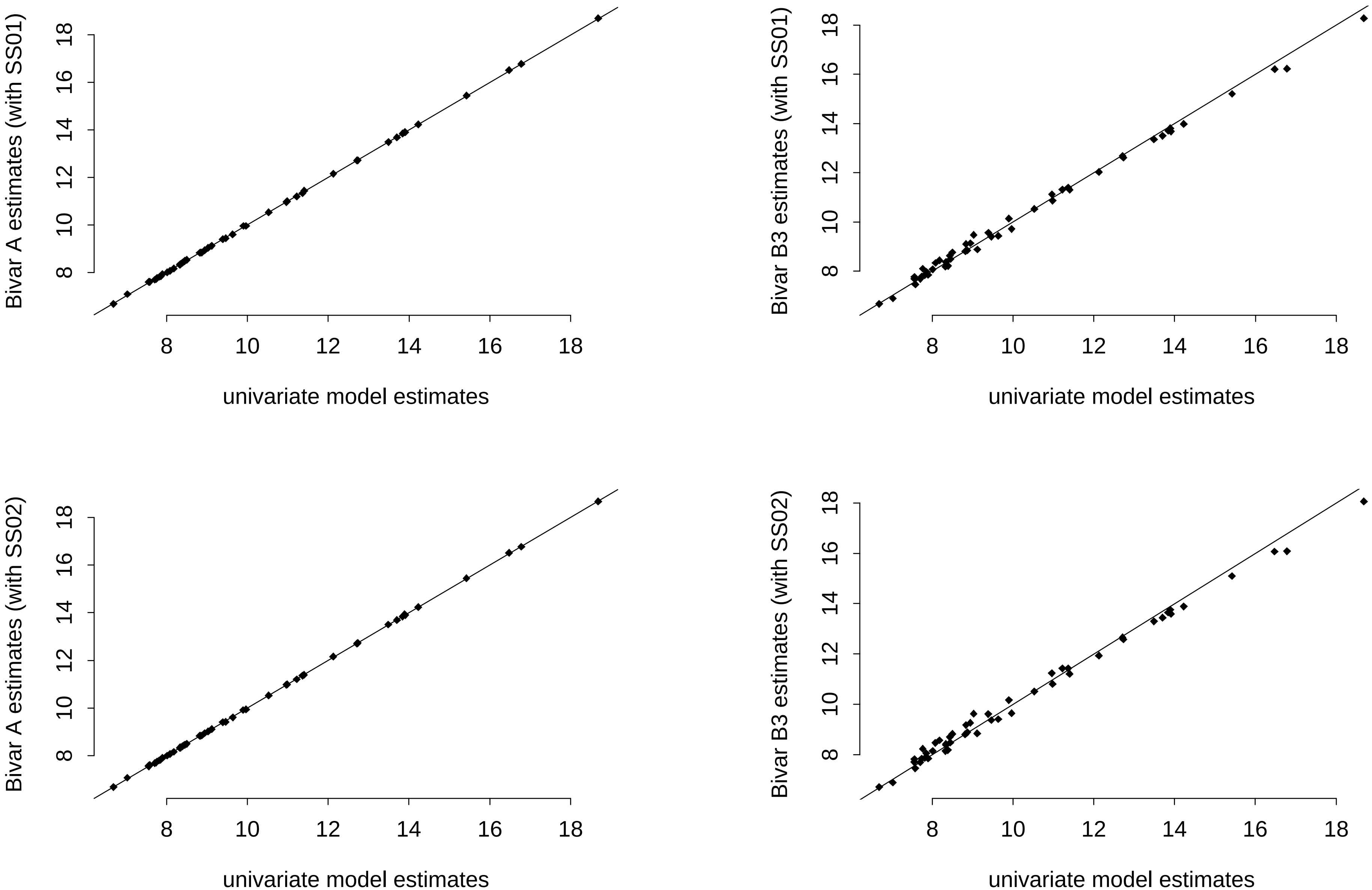


Figure 8. Poverty Ratio Estimates: Age 65+

c. IY2002 CPS equation alternative vs univariate estimates