Supplementary Information for "Dynamic Ecological Inference for Time-Varying Population Distributions Based on Sparse, Irregular, and Noisy Marginal Data" (Caughey and Wang, *Political Analysis*)

A Online Appendix

A.1 Example Application (Phone by Race by Region)

A.1.1 Stan Code for Example Application

```
data {
  int<lower=1> N; // number of cells
  int<lower=1> Y; // number of time periods
  int<lower=1> M; // maximum number of margins in any period
  matrix<lower=0>[Y, M] n_sample; // sample size of observed margins
  real<lower=0> n_prior;  // governs precision of priors
  real<lower=0> n_evolve; // governs precision of transition model
  real Ygaps[Y];  // gap between periods -- all 1 in this application
  simplex[N] pi0; // prior means for pi in first period
  simplex[2] props_y1m1; // marginal count data for t=1, m=1
  matrix<lower=0,upper=1>[2, N] A_y1m1; // 'A' matrix for t=1, m=1
  simplex[4] props_y1m2; // etc.
  matrix<lower=0,upper=1>[4, N] A_y1m2;
  simplex[2] props_y6m1;
  matrix<lower=0,upper=1>[2, N] A_y6m1;
  simplex[4] props_y8m1;
  matrix<lower=0,upper=1>[4, N] A_y8m1;
  simplex[4] props_y11m1;
  matrix<lower=0,upper=1>[4, N] A_y11m1;
  simplex[4] props_y11m2;
  matrix<lower=0,upper=1>[4, N] A_y11m2;
  simplex[4] props_y16m1;
  matrix<lower=0,upper=1>[4, N] A_y16m1;
  simplex[4] props_y21m1;
  matrix<lower=0,upper=1>[4, N] A_y21m1;
  simplex[8] props_y31m1;
  matrix<lower=0,upper=1>[8, N] A_y31m1;
}
parameters {
  // period-specific cell proportions
  simplex[N] pi[Y];
}
```

```
model {
  // Priors for first period
  pi[1] ~ dirichlet(pi0 * n_prior);
  // Transition model
  for (y in 2:Y) {
    pi[y] ~ dirichlet(pi[y - 1] * n_evolve / Ygaps[y]);
  }
  // Observation model
  props_y1m1 ~ dirichlet(n_sample[1, 1] * A_y1m1 * pi[1]);
  props_y1m2 ~ dirichlet(n_sample[1, 2] * A_y1m2 * pi[1]);
  props_y6m1 ~ dirichlet(n_sample[6, 1] * A_y6m1 * pi[6]);
  props_y8m1 ~ dirichlet(n_sample[8, 1] * A_y8m1 * pi[8]);
  props_y11m1 ~ dirichlet(n_sample[11, 1] * A_y11m1 * pi[11]);
  props_y11m2 ~ dirichlet(n_sample[11, 2] * A_y11m2 * pi[11]);
  props_y16m1 ~ dirichlet(n_sample[16, 1] * A_y16m1 * pi[16]);
  props_y21m1 ~ dirichlet(n_sample[21, 1] * A_y21m1 * pi[21]);
  props_y31m1 ~ dirichlet(n_sample[31, 1] * A_y31m1 * pi[31]);
}
```

A.1.2 Data for Example Application

Table A1: Year = 1930

	PHONE	Prop	Freq
1	No Phone	0.62	620000.00
2	Phone	0.38	380000.00

Table A2: Year = 1930

	BLACK	SOUTH	Prop	Freq
1	Non-Black	Non-South	0.70	4238482.24
2	Black	Non-South	0.02	147023.01
3	Non-Black	South	0.20	1236172.22
4	Black	South	0.07	437058.41

Table A3: Year = 1935

	PHONE	Prop	Freq
1	No Phone	0.70	700000.00
2	Phone	0.30	300000.00

Table A4: Year = 1937

	PHONE	SOUTH	Prop	Freq
1	No Phone	Non-South	0.47	471935.97
2	Phone	Non-South	0.28	275951.33
3	No Phone	South	0.21	208389.10
4	Phone	South	0.04	43723.60

Table A5: Year = 1940

	PHONE	SOUTH	Prop	Freq
1	No Phone	Non-South	0.45	449717.42
2	Phone	Non-South	0.30	295967.38
3	No Phone	South	0.21	205127.97
4	Phone	South	0.05	49187.23

Table A6: Year = 1940

	BLACK	SOUTH	Prop	Freq
1	Non-Black	Non-South	0.69	930615.80
2	Black	Non-South	0.03	34851.81
3	Non-Black	South	0.21	284635.57
4	Black	South	0.07	94764.83

Table A7: Year = 1945

	PHONE	SOUTH	Prop	Freq
1	No Phone	Non-South	0.36	360332.20
2	Phone	Non-South	0.38	383832.20
3	No Phone	South	0.19	189554.94
4	Phone	South	0.07	66280.66

Table A8: Year = 1950

	BLACK	SOUTH	Prop	Freq
1	Non-Black	Non-South	0.69	1311123.87
2	Black	Non-South	0.04	68376.51
3	Non-Black	South	0.22	413181.21
4	Black	South	0.06	120031.41

Table A9: Year = 1960

	PHONE	BLACK	SOUTH	Prop	Freq
1	Non-Phone	Non-Black	Non-South	0.10	16845128.00
2	Phone	Non-Black	Non-South	0.58	100600323.00
3	Non-Phone	Black	Non-South	0.02	2620611.00
4	Phone	Black	Non-South	0.03	5264539.00
5	Non-Phone	Non-Black	South	0.07	11679308.00
6	Phone	Non-Black	South	0.15	25821087.00
7	Non-Phone	Black	South	0.04	6541732.00
8	Phone	Black	South	0.02	3501860.00

A.1.3 Sensitivity Analysis for Example Application

There are two main tuning parameters in our model: the implied sample size of the Dirichlet prior for the cell proportions in $t = 1$ (n^{prior}) and and the implied sample size of the Dirichlet transition model (n^{evolve}).[15] We generally recommend setting n^{prior} high enough that the prior is proper but low enough to be almost completely uninformative. We use $n^{\text{prior}} = 8$ in the main text and $n^{\text{prior}} = 100$ in the simulations below.

Of the two tuning parameters, n^{evolve} requires more substantive judgement. The results we report in the main text use $n^{\text{evolve}} = \exp(10) \approx 22{,}026$, which corresponds to a belief that a cell comprising half the population would be expected to change by a third of a percentage point per year. We view this as a reasonable choice for both empirical and theoretical reasons. Empirically, this prior is on the same order of magnitude as the typical *observed* change in the data on marginal proportions. Second, this value of n^{evolve} is large enough to yield informative priors even across 31 years, while still being an order of magnitude smaller than the typical n_{tm}^{samp} of around 1,000,000, thus allowing the data to dominate when available.

Even if this value of n^{evolve} is reasonable, however, it is nevertheless valuable to examine the results' sensitivity to this choice. Figure A1 reports the results of such a sensitivity analysis, which varies the value of n^{evolve} from 100 to 10,000,000 by powers of 10. For values of n^{evolve} between 1,000 and 1,000,000, the point estimates are quite similar, but the credible intervals differ greatly. For $n^{\text{evolve}} = 100$, the point estimates are very different from the other panels and the confidence intervals are almost completely uninformative. The main reason for this is that the 1960 IPUMS data have very little impact on estimates more than a few years previous. In 1940, for example, the 1960 IPUMS data have about as much informative value for estimates as a sample size of 5 (100/20). At the other extreme, a value of $n^{\text{evolve}} = 10{,}000{,}000$ (lower-right) implies that estimates from periods adjacent to t receive an order of magnitude more weight than the data in t itself. This has the effect of yielding posterior distributions that may not even cover the *observed* group proportions. This occurs most clearly in 1960, when all the cell estimates are substantially below their proportions in the 1960 IPUMS. In sum, this analysis suggests that although credible intervals differ, point estimates are relatively insensitive to the choice of n^{evolve} as long as it is moderately informative but not so large as to rival the influence of the data themselves.

On a side note, observe how much the uncertainty of estimated phone-ownership rates differ across groups. This is mainly a result of the fact that non-blacks constituted a larger proportion of the population, especially in the non-South, which was less than 5% black for much of this period. As a consequence, the phone ownership rate for the larger group (especially non-black non-Southerners) is much more tightly constrained by the known regional phone-ownership rate than the smaller group.

15. The sample size of each observation model, n_{tm}^{samp}, will typically be determined by the actual sample size of the data, but in some cases (as with our AT&T data) the sample size can be stipulated by the analyst to reflect their beliefs about the measurement error in the estimates.

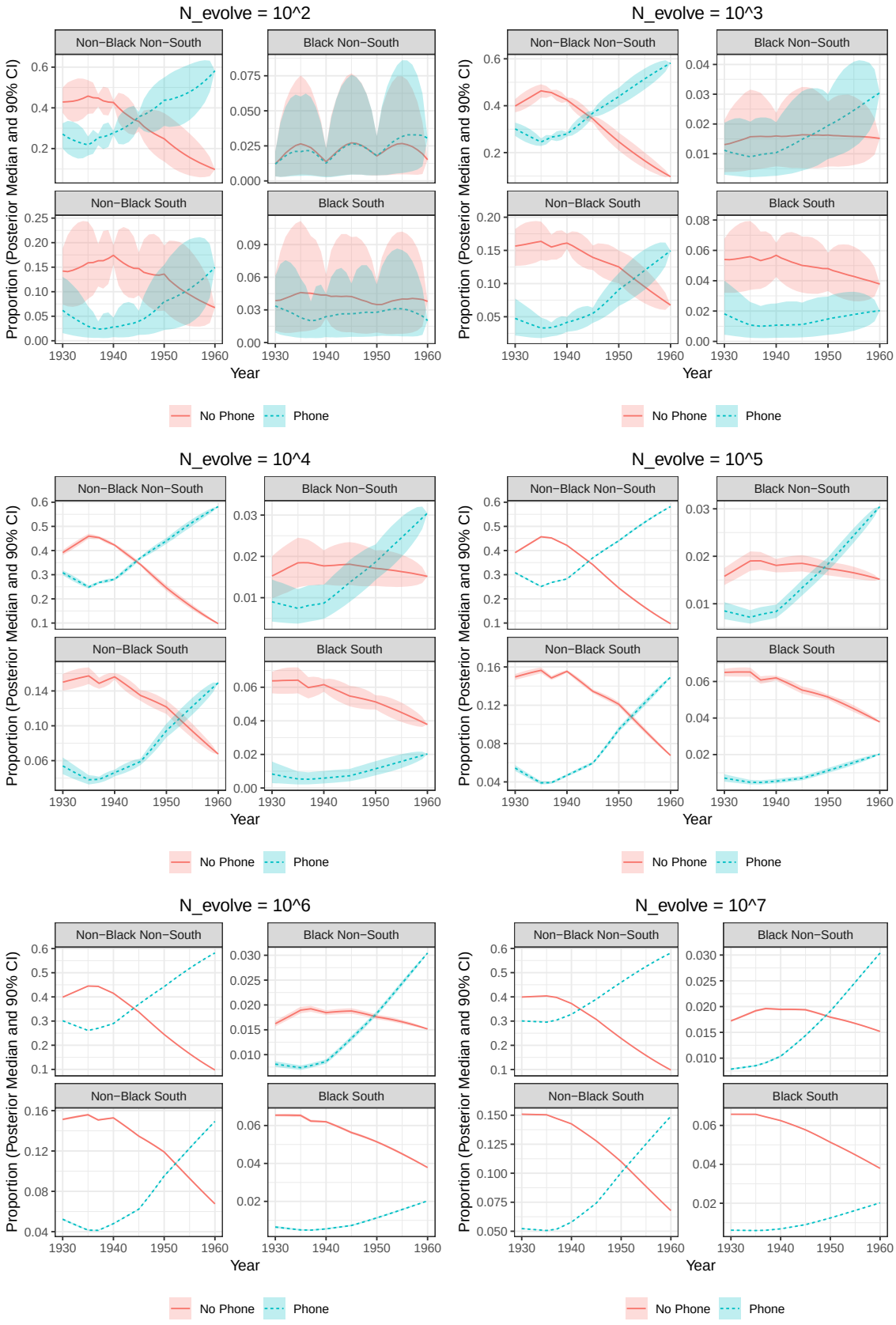

Figure A1: Sensitivity to different choices of "sample size" (n^{evolve}) for the transition model.

A.2 Estimates of Known Proportions using Different Data

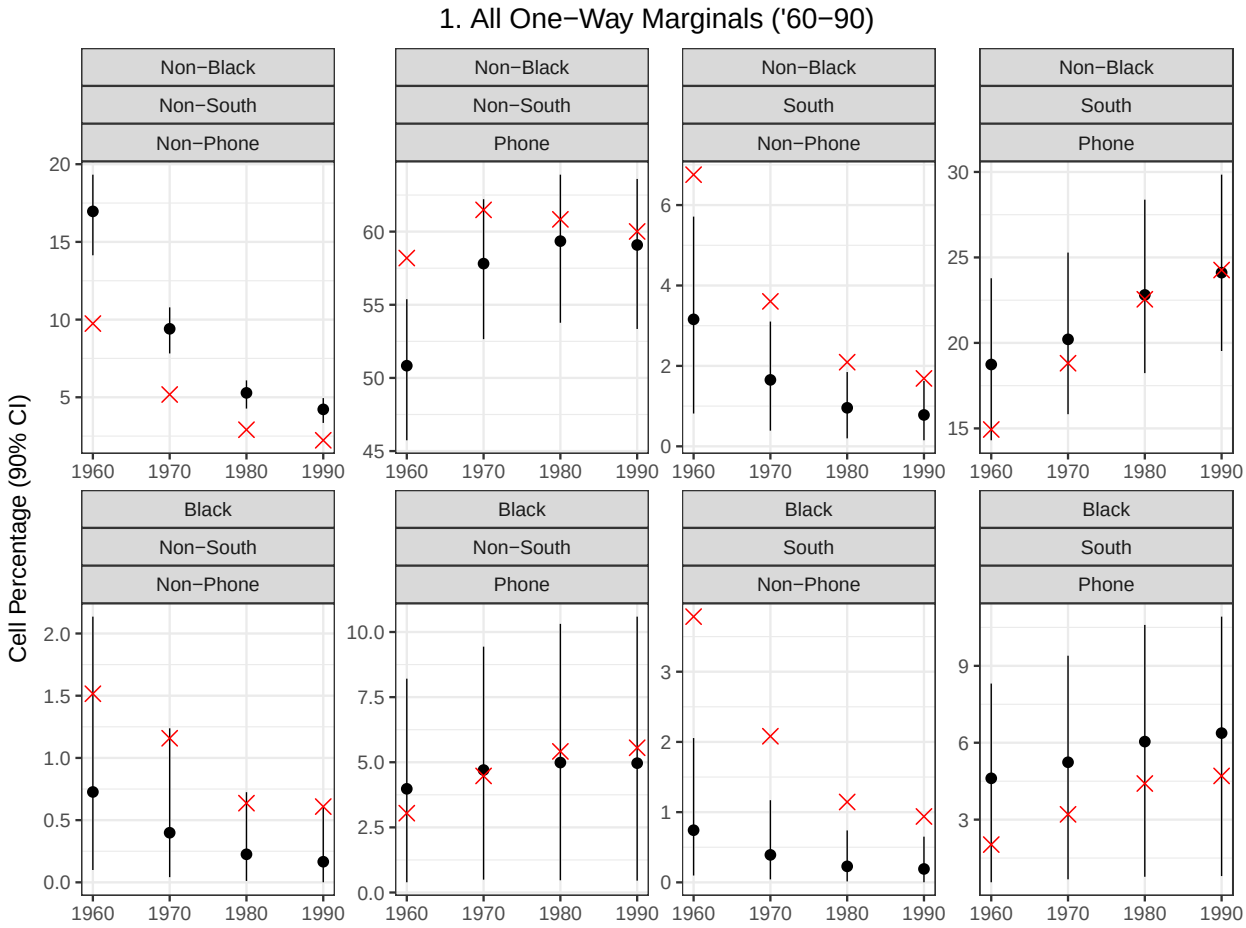

Figure A2: Dynamic cell estimates based on one-way marginal data. IPUMS targets are indicated with X.

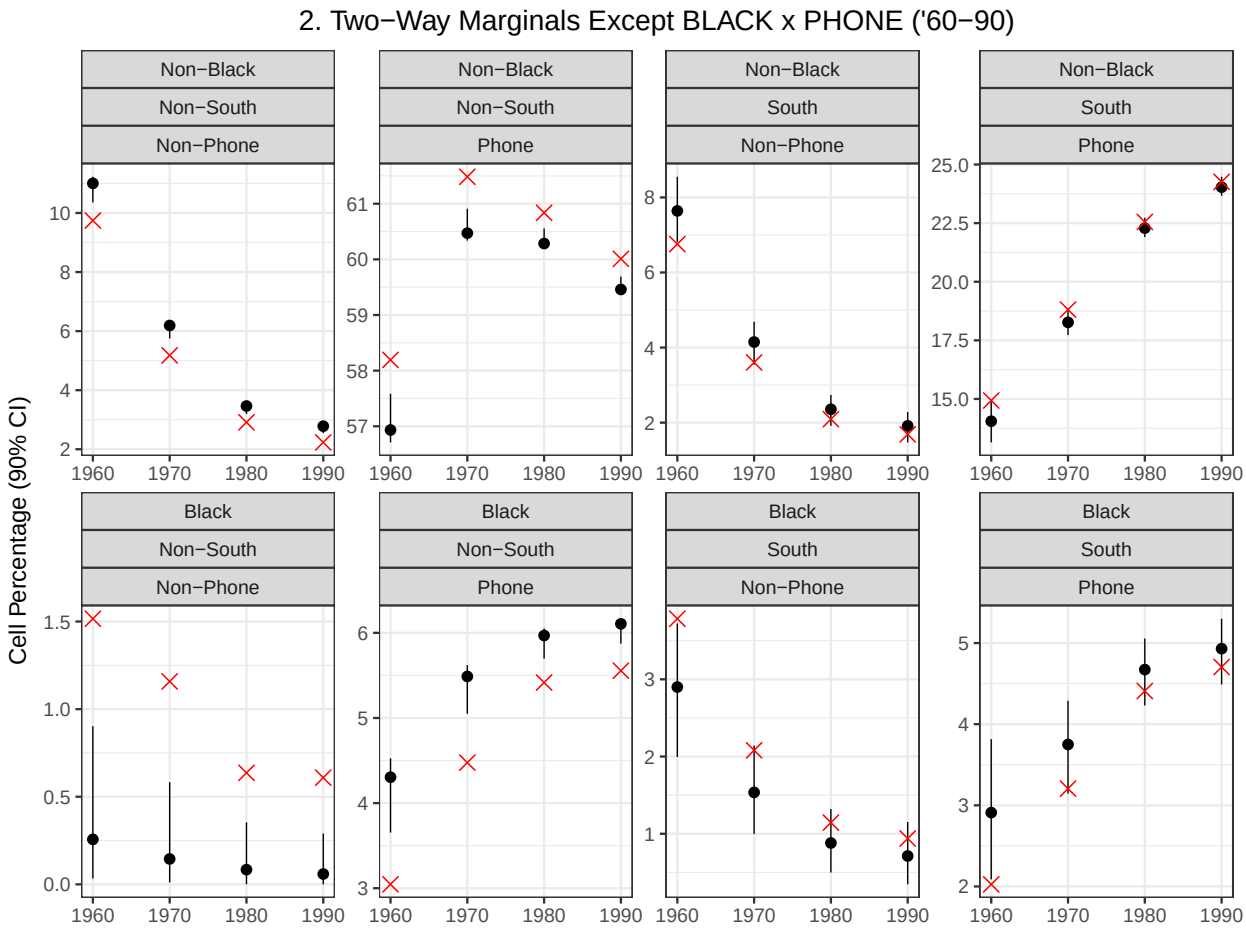

Figure A3: Dynamic cell estimates based on two-way marginal data except for BLACK x PHONE. IPUMS targets are indicated with X.

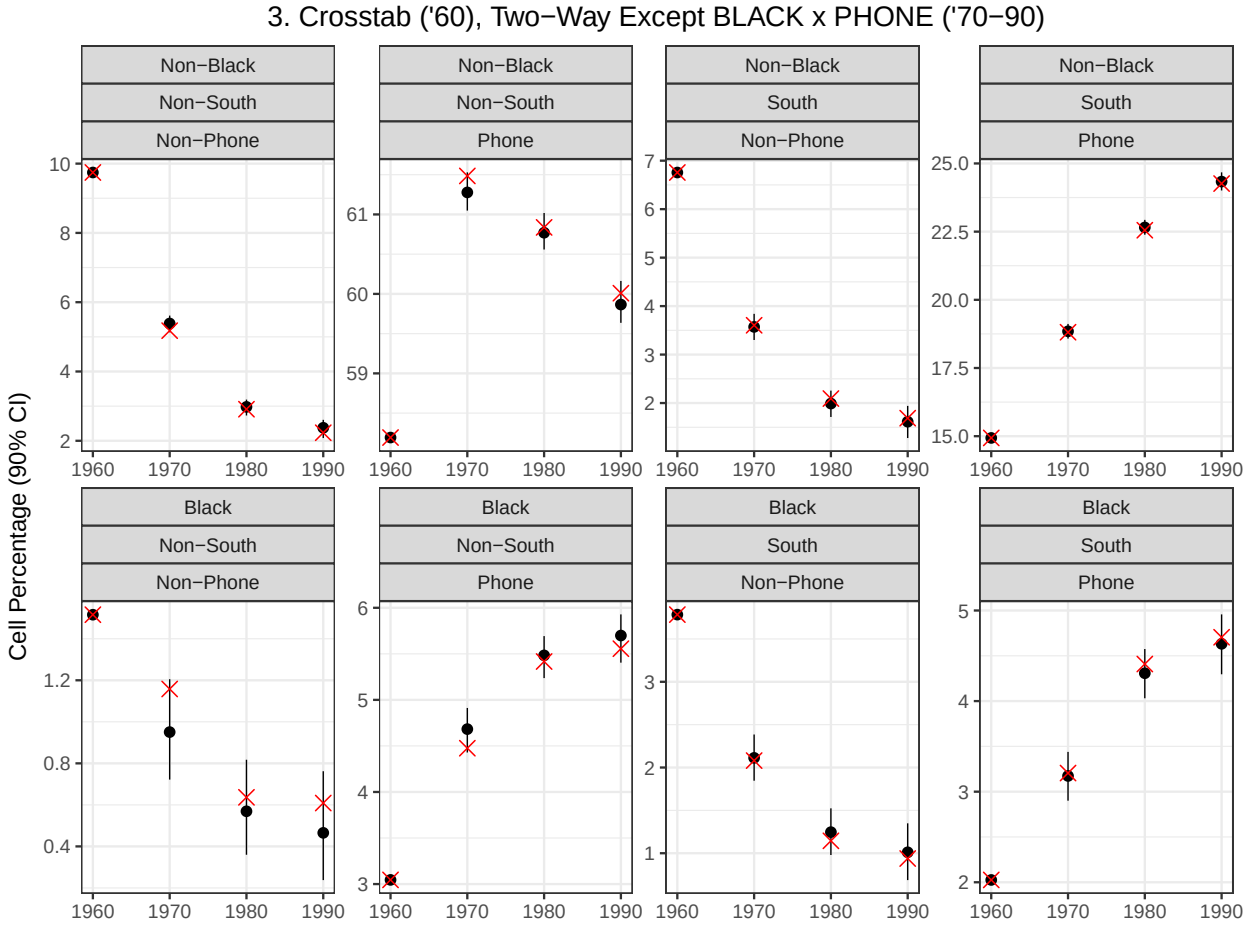

Figure A4: Dynamic cell estimates based on three-way crosstabs in 1960 and two-way marginal data except BLACK x PHONE in other years. IPUMS targets are indicated with X.

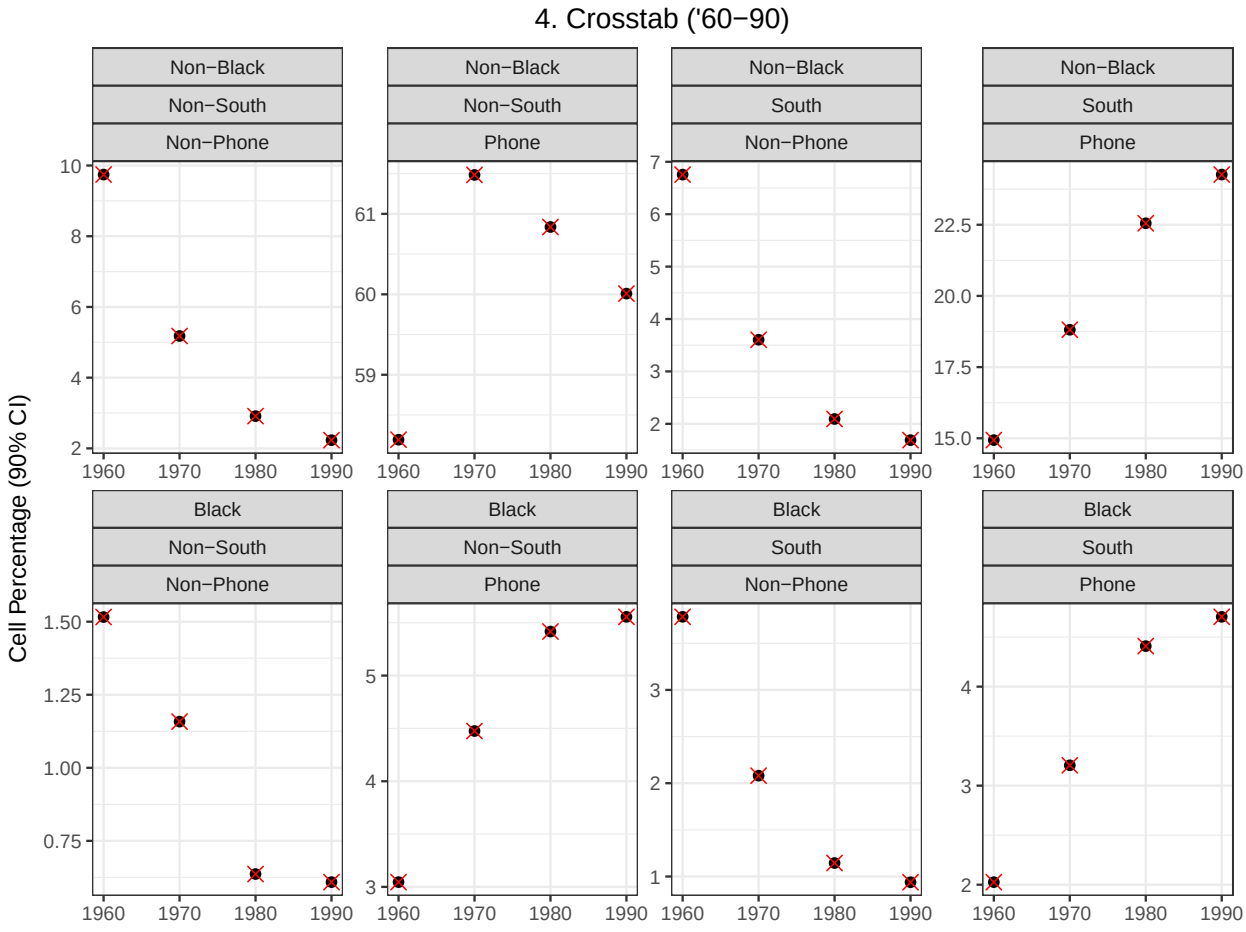

Figure A5: Dynamic cell estimates based on three-way crosstabs in all years. IPUMS targets are indicated with X.