Handling 1990 No-Data Blocks in Crosswalks

Jonathon Schroeder and James Gaboardi

May 2020

1   Problem

- General goal: Generate a crosswalk from source zones to target zones by aggregating up from a block-to-block crosswalk.

  - Ideally, the crosswalk should include at least one record for every source zone and at least one for every target zone, so all source and target zones are represented.

- This framework requires an association between each block in the base crosswalk and its encompassing source/target zone.

- To associate blocks with block group parts (BGPs) requires more information than is available in the base crosswalk...

  - The base crosswalk includes block IDs, from which we can derive codes for the state, county, tract, and block group that contain each block.

  - Identifying BGPs *additionally* requires (in 1990) codes for the place, county subdivision, AIANHH area, CD, and urban area.

- The additional required codes are available in 1990 block data tables from NHGIS, ***but***...

- 1990 data tables include no records for blocks that have no population and no housing units (i.e., "no-data blocks").

- Therefore, we **can't**[1] **directly associate 1990 no-data blocks with BGPs**.

2   Solution

General concept: Use the associations between no-data 1990 blocks and *whole block groups* (BGs) to ensure that any intersection between a 1990 BG and a target zone is represented in the crosswalk from 1990 BGPs to target zones...

1. Construct initial BGP crosswalk using only the "inhabited" blocks–those in 1990 data table.

   (a) Follow general steps 1-8, using option 4a for target zones and 4b for source zones

2. Construct BG crosswalk using *only* no-data blocks (all counts are 0) Repeat general steps 3-8 but...

   (a) In step 3, drop all block atoms that *have* a match in the 1990 summary data, keeping only atoms that have no match.

---

[1] We could determine associations from 2000 TIGER/Line data or NHGIS shapefiles, but the former would be complicated (3,000+ county-extent data files in an old format) and the latter would be time-consuming (overlay of several big nationwide shapefiles, some of which contain invalid offshore slivers that should first be cleaned up) and not totally reliable (potential inconsistencies due to differences in how each file was transformed from base TIGER files.) Meanwhile, determining exact associations shouldn't be important because we **already know** that no population or housing units were in these areas, and it's OK (even if not ideal) if we produce a crosswalk that includes "dummy" atoms for nonexistent source-target intersections, as long as the weights for those atoms are all zero, and all source and target zones are represented.

(b) Use option 4a for both source & target zones.

(c) Make sure to observe 4b: drop sub-zone atoms where the 1990 block GISJOIN = "".

(d) Skip steps 5-7: no need to compute counts because all are (should be) zero.

(e) In step 8, set all $\boldsymbol{w_{cst}} = 0$.

3. Add intersections from the "no-data BG crosswalk" (from b above) to the BGP crosswalk (from a), associating each "no-data atom" with all of the BG's parts...

    (a) Obtain and load summary data table for 1990 BGPs

        i. Need this to identify all BGPs that should be included in final crosswalk.

    (b) For every BGP in the BGP data table, identify its containing BG.

        i. One approach: Store the BG id in a temporary new field to the data table.

    (c) For every BGP in the initial BGP crosswalk, identify its containing BG.

        i. One approach: Store the BG id in a temporary new field to the crosswalk table.

    (d) "Expand" the no-data BG crosswalk for all associated BGPs...

        i. One approach: Join the no-data BG crosswalk to the BGP data table.

        ii. Outcome: Now any target zone associated with a 1990 BG through "no-data blocks" is also associated with all BGPs in that 1990 BG.

    (e) Remove any records in the expanded no-data BGP crosswalk where the 1990 BGP & target zone match a combo that already exists in the initial BGP crosswalk.

        i. Many source-target atoms contain a mix of inhabited and no-data 1990 blocks. The key info about these atoms was produced in step a above. We don't need to add the no-data records for these.

    (f) Add the remaining no-data BGP crosswalk records to the initial BGP crosswalk initial BGP crosswalk (from Step 1).

4. Proceed to general step 9.