METHODS AND ASSUMPTIONS FOR POPULATION PROJECTIONS

IDSER is responsible for preparing the 2020 population and land area estimates for urbanized and non-urbanized areas in Texas and in the United States. The projected population groups include total population, population aged 65 and over (senior population afterwards), persons with at least one disability (disabled population afterwards), and population living under 150% of the federal poverty line (low-income population afterwards).

We employed a mixture of two extrapolation approaches: linear extrapolation and exponential extrapolation. Both extrapolation approaches account for rates of population change in different time periods. The input data include the 2010 Decennial Census and the 2013 and 2018 5-Year American Community Survey (ACS) Summary File data. These data provide counts of the four population groups in each urbanized area in the United States over time, allowing us to calculate the average rate of population change in the linear and the exponential format for multiple time periods.

The linear extrapolation approach utilizes the average annual numeric change (AANC), which can be calculated by dividing the numeric change in total population between time 1 and time 0 by the total elapsed time between time 1 and time 0 (see Equation 1).

- $AANC = \frac{P_1 - P_0}{T}$ (1)

  where P_1 is the population at time 1, P_0 is the population at time 0, and T is the total elapsed time between time 1 and time 0.

The exponential extrapolation approach utilizes the average population growth rate (denoted by r), which can be calculated by taking the logarithm of the ratio of population at time 1 over the population at time 0, and then dividing it by the total elapsed time between time 1 and time 0 (see Equation 2).

- $r = \frac{\ln\left(\frac{P_1}{P_0}\right)}{T}$ (2)

  where P_1 is the population at time 1, P_0 is the population at time 0, and T is the total elapsed time between time 1 and time 0.

Counts of total population and senior population in urbanized areas are available in the 2010 Census and in the 2013 and 2018 5-Year ACS data. Therefore, we were able to calculate the average annual numeric changes (AANC) and the average population growth rate (r) for three time periods, which are: 2010-2011 (the mid-point year of the 2013 ACS), 2011-2016 (the mid-point year of the 2018 ACS), and 2010-2016. However, counts of the disabled population and low-income population in urbanized areas are only available in the 2013 and 2018 ACS data, allowing us to calculate the population change rates for only one time period: 2011-2016.

Population Groups in Urbanized Areas

The population projection process started with estimating the total population in 2020 for each urbanized area in the United States. We applied both the linear extrapolation and exponential extrapolation approach for the three time periods, yielding 6 forecasted values of the 2020 total population. We then dropped the highest and the lowest forecasted values and averaged the remaining four values for the forecasted 2020 total population. Equation 3 and 4 illustrate the extrapolation procedures in more detail.

- $P_t = P_b + T \times AANC$ (3)

where P_t is the target population, 2020 total population here, P_b is the base year population used to project the target population, AANC is the average annual numeric change rate, and T is the time span between year 2020 and the base year. For example, in the case of the linear extrapolation by using the AANC between 2010 and 2011 (the mid-point year of the 2013 ACS data), P_b is the total population in the 2013 ACS data, and T is 9 years (from 2011 to 2020).

- $P_t = P_b \times e^{r \cdot T}$ (4)

where P_t is the target population, 2020 total population here, P_b is the base year population used to project the target population, r is the average population growth rate, and T is the time span between year 2020 and the base year. For example, in the case of the exponential extrapolation by using the r between 2011 (the mid-point year of the 2013 ACS data) and 2016 (the mid-point year of the 2018 ACS data), P_b is the total population in the 2018 ACS data, and T is 4 years (from 2016 to 2020).

After the initial projection procedure, the projected 2020 total population in each urbanized area were treated as the control for the projections for the three other population groups (low-income population, senior population, and disabled population). For the senior population in each urbanized area, we applied the linear and exponential extrapolation methods to its complimentary population, that is, population aged 0 to 64. Similar to the procedure for estimating the total population, we dropped the highest and the lowest forecasted values and averaged the remaining four values. We then subtracted the forecasted population aged 0 to 64 from the projected 2020 total population to obtain the "controlled" projected 2020 senior population.

To validate the projected 2020 senior population, we cross-checked the proportion of senior population in 2020 against the average proportion of senior population from the 2010 Census, 2013 ACS, and 2018 ACS data. If the projected proportion of senior population in 2020 was 10% higher or lower than the average proportion of senior population over time, we first replaced it with the forecasted 2020 senior population that directly extrapolated from the base-year senior population (instead of extrapolating from its complimentary population), and checked again against the average proportion of senior population. If the substituted value was still 10% higher or lower than the average proportion of senior population, we further replaced the forecasted value with the product of the projected 2020 total population multiplied by the average proportion of senior population.

For low-income population and disabled population, we can only obtain the AANC and population growth rate (r) in each urbanized area for one time period: 2011-2016 (the mid-point year for 2013 and 2018 ACS, respectively). Similar to the process of forecasting the 2020 senior population, we linearly and exponentially extrapolated the complimentary population of low-income and disabled population, which are population living above the 150% federal poverty line and population with no disability, respectively. We retained both the linear and exponential results and averaged them to obtain the forecasted 2020 non-low-income population and 2020 non-disabled population. We then subtracted both of the forecasted complimentary populations in 2020 from the projected 2020 total population, yielding the "controlled" projected 2020 low-income population and 2020 disabled population. The same validation process as the one used for the projected 2020 senior population was conducted for these two population groups. We compared the proportions of these two population groups in 2020 with their associated average proportions over time. If their projected proportions in 2020 was 10% higher or lower than their average proportions over time, we first replaced them with their forecasted 2020 population that was directly extrapolated from the base-year low-income and disabled population (instead of extrapolating from their complimentary populations). We then cross-checked against the average proportions of these two population groups again. If the substituted values were still 10% higher or lower than the average proportions of these two population groups, we then replaced their forecasted values with the products of the projected 2020 total population multiplied by each of their associated average proportions.

Population Groups in Non-Urbanized Areas at the State Level

To estimate the 2020 counts of the four population groups (total population, senior population, low-income population, and disabled population) in non-urbanized areas at the state level, we need to know the projected 2020 state total counts for these four population groups regardless of the population residency (i.e., in urbanized areas or in non-urbanized areas). Since the projected 2020 counts in urbanized areas for the four population groups had already been generated, we summarized associated urbanized areas in each state to get the state counts of total population, senior population, low-income population, and disabled population in urbanized areas. The projected 2020 counts in non-urbanized areas of each state for these groups was calculated by subtracting the 2020 state counts of the four population groups in urbanized areas from their associated 2020 state total counts.

The 2010 Census and the 2013 and 2018 5-Year ACS data contain the total counts of the four population groups in each county of all states over time. We employed the same extrapolation approaches as we did for projecting the four population groups in urbanized areas, allowing us to estimate the 2020 counts of total population, senior population, low-income population, and disabled population in all counties in the 50 States, D.C., and Puerto Rico. We then calculated the differences between the total counts and the counts in urbanized areas for the four population groups, and documented the calculated differences as the counts of the four population groups in non-urbanized areas in each county. Lastly, we aggregated the county counts to obtain the total counts of the four population groups in non-urbanized areas in the 50 states, D.C. and Puerto Rico.

For the state of Texas, the Geographic Correspondence Engine provided by the Missouri Census Data Center (MCDC) documented detailed proportions of the urbanized population, and urbanized senior population. We averaged the proportion of urbanized area population and the proportion of urbanized area senior population. This calculated average was then treated as the proportion for both the low-income population and disabled population in urbanized areas for each county. We assumed the proportions of counts in county urbanized areas for the four population groups remain unchanged between 2010 and 2020. Using the projected values for each urbanized area in Texas counties, combined with their associated proportions of urbanized area counts, we were able to obtain the total counts for these four population groups at the county level for Texas. For Texas counties with no population in urbanized areas, this proportion conversion approach is not available. Instead, we utilized the projected total counts of the four population groups in Texas counties from the non-urbanized area projection process.

Finally, we aggregated the forecasted 2020 counts of the four population groups at different geographic levels in Texas (urbanized area level and county level) and crosschecked the consistency of these forecasted numbers at the state-level. Some minor adjustments were made to account for rounding differences when aggregating counts for all Texas urbanized areas as well as for all Texas counties.

The Housing Unit Methods for Projected 2020 Total Population and Population Density

The logic underlying the housing unit projection method is that: we first calculated the average household size for each census block in 2010, and then we assumed the block-wise average household size remains stable between 2010 and 2020. By multiplying the average household size in each census block by its most current number of housing units, we estimated the current population in each census block. We then combined the population of all census blocks that would be merged into urbanized areas with the urbanized area population from the 2018 5-Year ACS data. This approach allowed us to calibrate the population for all urbanized areas in the 2018 5-Year ACS data, and the calibrated urbanized area population were then used to refine the population projections.

The 2010 Census data have counts of household population and population living in group quarters (GQ) for each census block. We first excluded the GQ population in each census block when calculating the average household size, and then merged the block level average household size data with the most recent Local Update of Census Addresses Operation (LUCA) housing unit data. Second, we multiplied the average household size in each census block by its most current number of housing units, allowing us to obtain the estimate for the current household population for each census block. Third, we added the 2010 GQ population in each census block to each block's current household population, and treated the result as the current total population for each census block. Finally, we calculated the current population density for each census block by dividing the current total population by its current land areas.

We further based upon the current population density of each census block to determine whether a census block would be merged into its nearby exiting urbanized area. The general idea of this

delineation was as follows: We used the GIS program to launch an iterative process to merge all census blocks with greater than 500 ppsm (persons per square mile) into their neighboring urbanized areas. The delineation process for the projected 2020 urbanized areas will be explained in more detail in the next section. After the process was completed, we extracted information of the total population and land areas for each newly delineated urbanized area from the GIS output data. Finally, we used the population for all newly delineated urbanized areas to substitute the urbanized area population from the 2018 ACS 5-Year data and re-calculate the population projections described in the previous sections. The results were incorporated to calibrate the projected 2020 urbanized area population. The projected 2020 senior population, low-income population, and disabled population were then updated accordingly.

Delineation of Projected 2020 Urbanized Areas

First, all the census blocks that are greater than 500 ppsm (persons per square mile) and adjacent to existing urbanized areas were selected by using a search radius of 0.1 miles (with the Spatial Join Tool in ArcGIS). After the initial merging step was completed, any qualified census blocks (greater than 500 ppsm) that were adjacent to the "updated" urbanized areas were then selected iteratively with each series of the combined urbanized area and newly added census blocks. The delineation process ended when all the qualified adjacent census blocks were added to the iteratively updated urban areas with no additional qualified census blocks can be further identified.

Twelve urbanized areas did not have any projected census blocks with greater than 500 ppsm within the search radius of 0.1 miles. None of these urbanized areas were in Texas. Specifically, eleven of them were in California, and one was in North Dakota. We further examined these urbanized areas with aerial imagery to verify the results and confirmed no projected blocks were located within the search distance.

There were 41 projected census blocks intersecting with two urbanized areas simultaneously. We further reviewed each case by comparing these census blocks with city and place boundaries. Finally, we made necessary adjustments to re-assign some census blocks to their neighboring urbanized areas and finalized the projected 2020 urbanized areas in Texas and in the United States.