Modeling Hawaii's ALICE Households in the COVID Epoch

Validation Design, Process, and Results for Hawaii Data Collaborative's Synthetic Population Model

Table of Contents

Overview

This paper describes the assumptions that are critical to the development, utilization, and interpretation of the synthetic population model for the state of Hawaii covering the post-COVID time period (i.e., March 2020 to March 2021).

Validation of the synthetic population model looks at four areas: 1) representativeness of the baseline population, 2) model constructs, 3) data inputs, and 4) model theory. Within these four topics, 8 categories of assumptions are reviewed:

- Geographic representativeness
- Person and household representativeness
- Temporal shift
- ALICE framework
- Construct validity
- Accuracy of data inputs
- Key drivers of employment and wages
- Scenarios as estimates

Descriptions of each category are given at the beginning of each section with the following information:

- **Assumption** – explanation of the assumption the synthetic population model makes.
- **Test** – description of the assumption test.
- **Implications** – the supposed effect of violating the assumption.
- **Considerations** – additional information to consider when thinking about and testing the assumption.
- **Result** – conclusions made from the assumption test.

Following each description, a detailed review of the assumption test is given. Conclusions and model limitations are provided subsequently.

Representativeness of the Baseline Population

This section tests the assumption that the synthetic population used in the model is representative of the geographic, person, and household characteristics of the Hawaii population at the beginning of 2020. Representativeness is crucial to the interpretation of model outputs. A representative model (i.e., synthetic) population means that model outputs can be reflective of the current state of the population—assuming that data inputs, model constructs, and theory are valid. An unrepresentative model means that model outputs are not reflective of the current Hawaii population at the baseline point in time—and likely, not over the projected period.

Geographic representativeness, person and household representativeness, and the temporal shift from the data collection period (i.e., 2015–2019) to the baseline year (i.e., 2020) are examined below.

1. Geographic Representativeness

Assumption: The synthetic population is representative of geographic distributions for residence and place of work for the population to the County Subdivision level.

Test: Compare the distributions of the population for place of residence and place of work from model data (i.e., synthesized data) to the distributions of target data (i.e., ACS Summary Tables and PUMS) at the county subdivision level (CCD).

Implications: *If the assumption is broken,* modeled data cannot be used to analyze geographies lower than state-level.

Considerations: A built-in discrepancy exists in the target data (i.e., 2015–2019 ACS PUMS) where the sum of person weights does not equal the sum of housing weights multiplied by the number of people in households. This presents a problem of reconciliation between person and household totals when creating the synthetic population.

Results: The geographic distribution of people and households is reasonable for the majority of geographies. Spreckelsville and Kahului are obvious exceptions, where the population is overestimated and underestimated significantly, respectively. There are **a handful of smaller geographies where person and household totals differ** from target totals more than would seem reasonable to make conclusions about these geographies from modeled data.

The assignment of workers to workplaces in modeled data is unreasonable considering island geography and existing infrastructure for and data on travel time to work. This presents a problem in interpreting model results for smaller geographies when considering the differential impacts of COVID-19 by community (e.g., some communities are more reliant on tourism). If the distribution of industry is valid in the synthetic population, then **the current model assumes equal economic impact statewide due to COVID-19, regardless of geography**.

1.1 Distribution of the Population

Determine that the population is distributed proportionately with target data to the County Subdivision level (CCD).

Table 1. Distribution of the population[1]

CCD	Model Estimate	Validation Estimate	Model Percent	Validation Percent	Difference	Percent Difference
Spreckelsville	6401	273	0.45	0.02	0.43	2242.58
Honokaa-Kukuihaele	4152	3529	0.29	0.25	0.04	17.55
Hana	1644	1405	0.12	0.1	0.02	16.91
Kalawao	76	66	0.01	0	0	15.05
East Molokai	4380	3892	0.31	0.27	0.03	12.44
Koolauloa	20740	19681	1.46	1.38	0.07	5.29
Lahaina	23469	22309	1.65	1.57	0.08	5.11
Lanai	2826	2730	0.2	0.19	0.01	3.42
Waialua	13443	13090	0.94	0.92	0.02	2.6
Hanalei	6232	6073	0.44	0.43	0.01	2.53
Hilo	48773	47637	3.43	3.35	0.08	2.29
West Molokai	2368	2317	0.17	0.16	0	2.11
Lihue	7212	7058	0.51	0.5	0.01	2.09
Wahiawa	44654	43774	3.14	3.08	0.06	1.92
Kekaha-Waimea	5524	5424	0.39	0.38	0.01	1.75
Puhi-Hanamaulu	11060	10871	0.78	0.76	0.01	1.65
Pahoa-Kalapana	11215	11068	0.79	0.78	0.01	1.24
Kula	13479	13317	0.95	0.94	0.01	1.13
Koloa-Poipu	6574	6513	0.46	0.46	0	0.85
North Kohala	6045	5998	0.42	0.42	0	0.69
Honolulu	402951	400317	28.31	28.15	0.16	0.57
Waianae	50134	49971	3.52	3.51	0.01	0.24
Ewa	344822	345159	24.23	24.27	-0.05	-0.19
Keaau-Mountain View	35552	35700	2.5	2.51	-0.01	-0.5
Koolaupoko	112213	112829	7.88	7.93	-0.05	-0.64
Makawao-Paia	20438	20558	1.44	1.45	-0.01	-0.67
South Kohala	19856	20289	1.39	1.43	-0.03	-2.22
South Kona	10768	11027	0.76	0.78	-0.02	-2.44
Paauhau-Paauilo	2520	2584	0.18	0.18	0	-2.56
Wailua-Anahola	13554	13913	0.95	0.98	-0.03	-2.67
Papaikou-Wailea	4162	4280	0.29	0.3	-0.01	-2.84
Kapaa	7926	8164	0.56	0.57	-0.02	-3
Eleele-Kalaheo	9246	9552	0.65	0.67	-0.02	-3.29

[1] Model source: sample_individuals, experiement.covid_individuals
Validation source: 2019 5-year ACS summary file, table B01001

CCD	Model Estimate	Validation Estimate	Model Percent	Validation Percent	Difference	Percent Difference
Wailuku	22783	23635	1.6	1.66	-0.06	-3.69
Kaumakani-Hanapepe	4049	4201	0.28	0.3	-0.01	-3.7
Kau	9473	9885	0.67	0.7	-0.03	-4.25
North Hilo	1510	1580	0.11	0.11	-0.01	-4.52
North Kona	43631	45882	3.07	3.23	-0.16	-4.99
Kihei	23950	25294	1.68	1.78	-0.1	-5.4
Waihee-Waikapu	7427	7879	0.52	0.55	-0.03	-5.82
Haiku-Pauwela	9933	11034	0.7	0.78	-0.08	-10.06
Kahului	26207	31336	1.84	2.2	-0.36	-16.44

- Spreckelsville shows a large increase in population (from 273 to 6,401) between model and target data.
- Kahului shows a large decrease in population (from 31,336 to 26,207) between model and target data.
- Other geographies with a notable difference between model and target estimates (> 5 or < -5 percent difference) include: Honokaa-Kukuihaele, Hana, Kalawao, East Molokai, Koolauloa, Lahaina, Kihei, Waihee-Waikapu, and Haiku-Pauwela.

Table 2. Distribution of households[2]

CCD	Model Estimate	Validation Estimate	Model Percent	Validation Percent	Difference	Percent Difference
Spreckelsville	1710	102	0.37	0.02	0.35	1583.06
Kalawao	47	39	0.01	0.01	0	20.99
East Molokai	1513	1379	0.33	0.3	0.03	10.15
Hana	498	467	0.11	0.1	0.01	7.06
Lahaina	7757	7390	1.7	1.61	0.09	5.38
Lihue	2266	2175	0.5	0.47	0.02	4.59
Honokaa-Kukuihaele	1186	1146	0.26	0.25	0.01	3.9
Waialua	4122	4026	0.9	0.88	0.02	2.79
Kula	5282	5172	1.15	1.13	0.03	2.53
Koolauloa	4960	4873	1.08	1.06	0.02	2.19
Hilo	16900	16615	3.69	3.62	0.08	2.12
Pahoa-Kalapana	4495	4429	0.98	0.96	0.02	1.89
Koloa-Poipu	2267	2245	0.5	0.49	0.01	1.38
Waihee-Waikapu	2252	2235	0.49	0.49	0.01	1.16
Papaikou-Wailea	1631	1620	0.36	0.35	0	1.08
Hanalei	2199	2191	0.48	0.48	0	0.76
Honolulu	147049	146529	32.13	31.89	0.24	0.75
Koolaupoko	34364	34269	7.51	7.46	0.05	0.67

[2] Model source: sample_households, experiements.covid_households
Validation source: 2019 5-year ACS summary file, table B11001

CCD	Model Estimate	Validation Estimate	Model Percent	Validation Percent	Difference	Percent Difference
Waianae	12057	12098	2.63	2.63	0	0.05
West Molokai	807	811	0.18	0.18	0	-0.1
Wahiawa	12070	12137	2.64	2.64	0	-0.16
Eleele-Kalaheo	2877	2895	0.63	0.63	0	-0.23
Kaumakani-Hanapepe	1186	1195	0.26	0.26	0	-0.36
Wailua-Anahola	4658	4698	1.02	1.02	0	-0.46
Ewa	97756	98863	21.36	21.52	-0.16	-0.73
Makawao-Paia	7065	7172	1.54	1.56	-0.02	-1.1
Keaau-Mountain View	12532	12739	2.74	2.77	-0.03	-1.24
South Kohala	6495	6605	1.42	1.44	-0.02	-1.28
Puhi-Hanamaulu	2666	2714	0.58	0.59	-0.01	-1.38
Kau	3060	3131	0.67	0.68	-0.01	-1.88
Kekaha-Waimea	1789	1842	0.39	0.4	-0.01	-2.5
Wailuku	7447	7682	1.63	1.67	-0.04	-2.68
Lanai	1143	1181	0.25	0.26	-0.01	-2.84
Kapaa	2615	2703	0.57	0.59	-0.02	-2.88
South Kona	3609	3748	0.79	0.82	-0.03	-3.33
Kihei	8891	9248	1.94	2.01	-0.07	-3.48
Paauhau-Paauilo	834	868	0.18	0.19	-0.01	-3.54
North Kohala	2013	2112	0.44	0.46	-0.02	-4.31
North Kona	15073	15821	3.29	3.44	-0.15	-4.35
Haiku-Pauwela	3587	3794	0.78	0.83	-0.04	-5.08
North Hilo	585	619	0.13	0.13	-0.01	-5.12
Kahului	6311	7846	1.38	1.71	-0.33	-19.25

- Similar to Table 1 above, Spreckelsville has a large increase in the number of households and Kahului has a large decrease in the number of households between model and target data.
- Other geographies with a notable difference include: East Molokai, Lahaina, and Haiku-Pauwela.

1.2 Geographic Distribution of Workers to Places of Work[3]

Determine that place of work is reasonable for the population at each CCD.

Figure 1. Place of work among Koolauloa residents

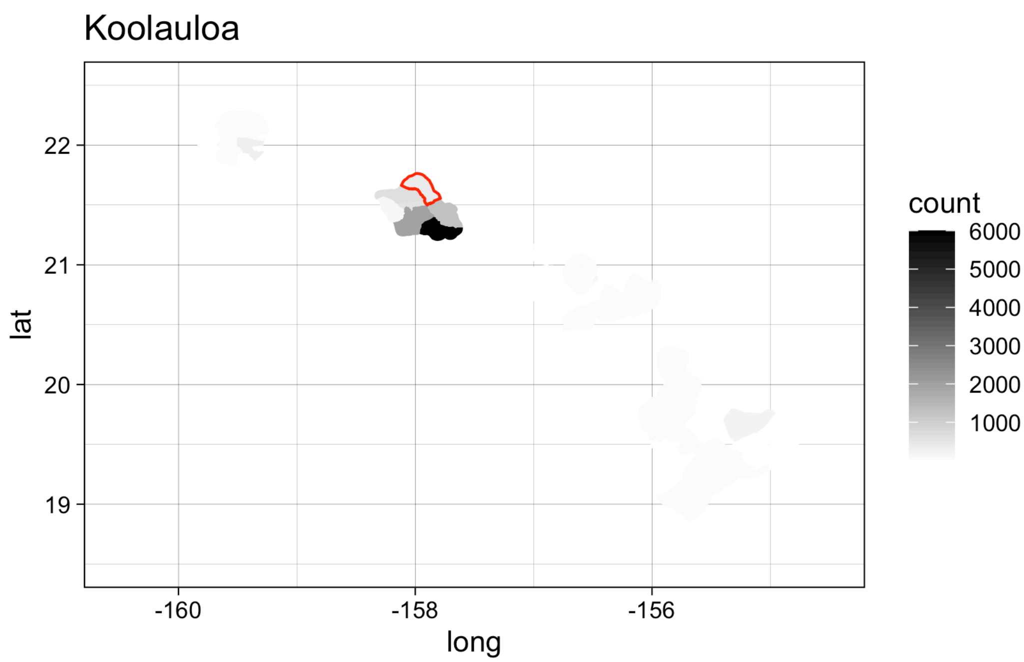
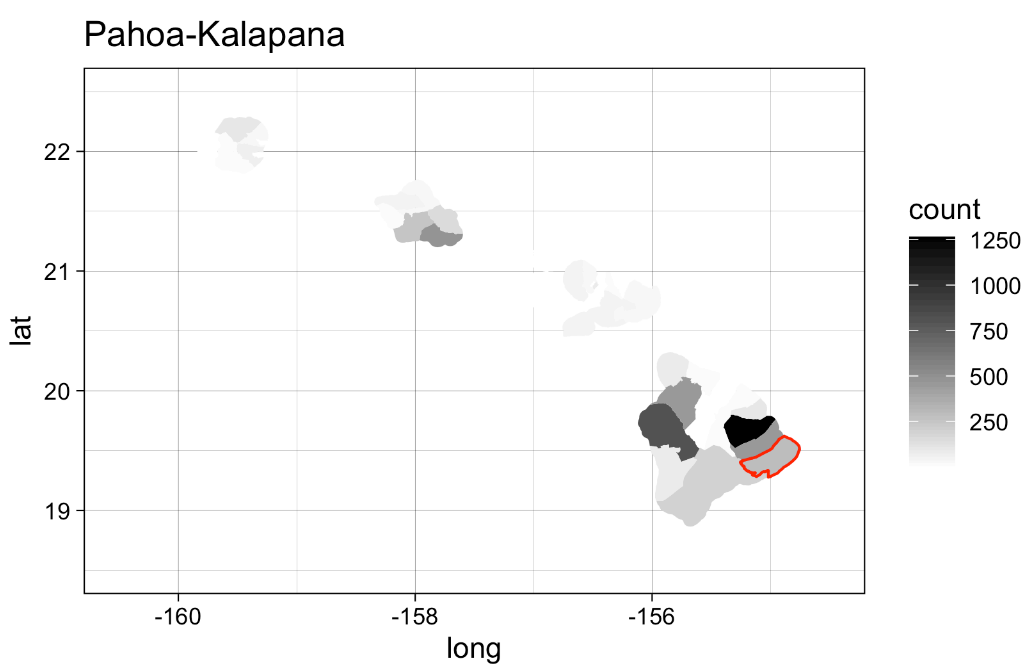


Figure 2. Place of work among Pahoa–Kalapana residents



[3] Model source: experiement.covid_households, inputs.households_workplaces_links, inputs.firms

- Workers tend to travel extreme distances to their place of work at high rates.
    - Koolauloa: nearly 10 percent of Koolauloa workers work on neighbor islands.
    - Pahoa-Kalapana: twenty-eight percent of workers work on neighbor islands.
- The number of Koolauloa residents who work in Koolauloa (384) appears low considering roughly 1,800 available jobs in the tourism industry for the region and roughly 4,500 Koolauloa residents who travel less than 25 minutes to work (i.e., work within the Koolauloa area).
    - The assignment of workers to workplaces is not informed by the "travel time to work" variable in the ACS (see also 2.3 Distribution of Relevant Non-matching Variables, *Travel Time to Work*).

2. Person and Household Representativeness

Assumption: Characteristics of people and households in the synthesized data resemble characteristics of people and households in the survey data *for matching variables and relevant non-matching variables* to the county subdivision geographic level.

Test: Compare the distributions of matching and *non-matching* variables of interest from model data (i.e., synthetic population) to the distributions of target data (i.e., ACS summary files) on those same variables.

Implications: *If the assumption is broken,* the modeled data are not representative of people or households in Hawaii.

Considerations: The relationship between variables at lower level geographies (i.e., county subdivision and census tract) may not remain constant at an aggregate level after synthesis (particularly for those variables that were not used for matching). Extreme care should be taken when using synthesized data to analyze variables that have not been used for matching in the synthesis process.

The testing done in sections 2.2 Distribution of Matching Variables and 2.3 Distribution of Relevant Non-matching Variables were conducted as the greatest level of detail possible for each relevant summary file table. Model accuracy at this level is not always necessary (e.g., age by sex), but in other cases is important (e.g., number of earners in family).

Results: The reliability of underlying (target) data for small geographies is too low to be used as a target for generating a synthetic population. In such cases, small geographies should be combined to improve the reliability of estimates of target data.

There are roughly 28,000 fewer employed individuals in the model data compared with target data. The majority of these missing cases are employed in industry categories that include *educational services* and *accommodation and food services*.

Household type by size shows large differences in distribution between model and target data. The source data used to develop the synthetic population (i.e., IPUMS) contains a code '9' (household type could not be determined) that does not exist in the original PUMS data. This could explain (at least in part) the **underestimation of households by type and size**.

At detailed levels of household income, there is considerable error between model and target data. An additional issue here is the reliability of target data (as mentioned above). Income data are critical to the utility of the model as need is determined by the ratio of income to housing costs.

Modeled data have a tendency toward the overall distribution for variables that were not used for matching or variables not sufficiently correlated with matching variables.

There is an **underestimation of no-earner and 1-earner households**, which may result in an understatement of the negative effects due to COVID-19 in modeled scenarios.

Modeled data for specific race groups require race matching as demonstrated in the Poverty Status among NHOPI section below.

2.1 Person and Household Totals

Determine that total population size and total number of households match in model and target data.

Table 3. Comparison of population and household totals[4]

Subject	Model Value	Validation Value
Total population size	1,423,372	1,422,094
Number of households	457,582 (where GQTYPE = 0)	459,424 (occupied housing units)

- Population totals at the state level match.
- Total number of occupied households (non-GQ) at the state level match.

[4] Model source: experiements.covid_individuals, sample_individuals, experiments.covid_households, sample_households
Validation source: 2019 5-year ACS summary file, tables B01001 and B25009

2.2 Distribution of Matching Variables

Determine the level of congruence between distributions of target (original summary file) data and model (synthetic population) data.

Table 4. Level of congruence between model and validation data by matching variable [0 = perfect match]

Name	Age	Employment	OccXInd	Tenure	HH_Type	HH_Income	Total	HH(n)
Spreckelsville	23.7037	37.4279	64.7073	16.0156	21.7895	18.72	182.364	102
Kalawao	0.6897	0.7126	0.9644	0.2293	0.4522	0.8027	3.8509	39
Hana	0.6513	1.1955	1.1756	0.117	0.2779	0.4048	3.8221	467
North Hilo	0.4476	0.5689	0.8975	0.0733	0.2149	0.5126	2.7148	619
Honokaa-Kukuihaele	0.4526	0.7842	0.7779	0.0891	0.2328	0.362	2.6986	1146
Lanai	0.3008	0.5013	1.0264	0.0337	0.2584	0.5093	2.63	1181
East Molokai	0.3786	0.5736	0.9049	0.1171	0.227	0.3491	2.5502	1379
Kau	0.2196	0.5244	1.0698	0.0279	0.3553	0.2442	2.4412	3131
KaumakaniHanapepe	0.3771	0.667	0.7153	0.0865	0.2017	0.389	2.4366	1195
Paauhau-Paauilo	0.3622	0.4551	0.6896	0.0641	0.2848	0.4176	2.2733	868
Lihue	0.3069	0.5806	0.6338	0.048	0.2454	0.4088	2.2235	2175
North Kohala	0.2488	0.518	0.6836	0.1618	0.2671	0.2806	2.1599	2112
Papaikou-Wailea	0.2911	0.4517	0.6933	0.0836	0.3261	0.3027	2.1485	1620
Kapaa	0.313	0.5121	0.5732	0.0593	0.2622	0.4144	2.1344	2703
Haiku-Pauwela	0.2802	0.4399	0.6445	0.0548	0.419	0.25	2.0885	3794
Kahului	0.2403	0.3509	0.5427	0.1957	0.414	0.3145	2.058	7846
South Kona	0.2857	0.4863	0.7334	0.0963	0.1445	0.2187	1.9648	3748
Hanalei	0.3196	0.3956	0.8014	0.0222	0.228	0.1826	1.9494	2191
West Molokai	0.2882	0.4623	0.6029	0.0324	0.2177	0.3448	1.9483	811
Wahiawa	0.0996	1.2403	0.3165	0.0156	0.1063	0.1614	1.9396	12137
Kekaha-Waimea	0.2705	0.49	0.5129	0.0294	0.2288	0.3494	1.881	1842
South Kohala	0.2275	0.4187	0.6436	0.0356	0.1558	0.2554	1.7365	6605
Pahoa-Kalapana	0.1911	0.4561	0.5198	0.0514	0.1939	0.3107	1.7231	4429
Waihee-Waikapu	0.1968	0.4543	0.4994	0.0101	0.201	0.2857	1.6473	2235
Eleele-Kalaheo	0.2396	0.4492	0.5555	0.046	0.1618	0.1931	1.6452	2895
Waialua	0.1682	0.6118	0.5533	0.0378	0.1164	0.1283	1.6158	4026
Puhi-Hanamaulu	0.229	0.3394	0.5048	0.0504	0.1881	0.2355	1.5471	2714
Koolauloa	0.1712	0.3221	0.5044	0.0499	0.3113	0.1759	1.5347	4873
Koloa-Poipu	0.2175	0.3858	0.465	0.0196	0.1492	0.2189	1.456	2245
Makawao-Paia	0.1787	0.3373	0.4328	0.0422	0.1993	0.2431	1.4334	7172
Kula	0.1985	0.3543	0.4321	0.0353	0.1572	0.1981	1.3755	5172
North Kona	0.1444	0.3034	0.4416	0.0653	0.2194	0.1783	1.3523	15821
Lahaina	0.1926	0.3253	0.4289	0.057	0.1624	0.1683	1.3346	7390
Wailua-Anahola	0.192	0.3085	0.3791	0.0112	0.268	0.1476	1.3062	4698
Kihei	0.1468	0.2642	0.3948	0.0929	0.2483	0.1529	1.3	9248
Wailuku	0.1636	0.2768	0.3399	0.0842	0.1806	0.1537	1.1989	7682
Waianae	0.1596	0.2949	0.4035	0.0126	0.1594	0.1303	1.1604	12098

Name	Age	Employment	OccXInd	Tenure	HH_Type	HH_Income	Total	HH(n)
Keaau-Mountain View	0.1633	0.2687	0.374	0.0542	0.1246	0.1088	1.0936	12739
Hilo	0.1139	0.1932	0.3698	0.0172	0.117	0.0974	0.9085	16615
Koolaupoko	0.0691	0.3697	0.2825	0.0175	0.0588	0.0876	0.8852	34269
Ewa	0.0482	0.2044	0.1882	0.0241	0.108	0.0647	0.6377	98863
Honolulu	0.0461	0.1023	0.1723	0.0039	0.1125	0.055	0.492	146529

Age by sex[5]

Figure 3. Distribution of age—comparison of model and validation data [sufficient match and shape]

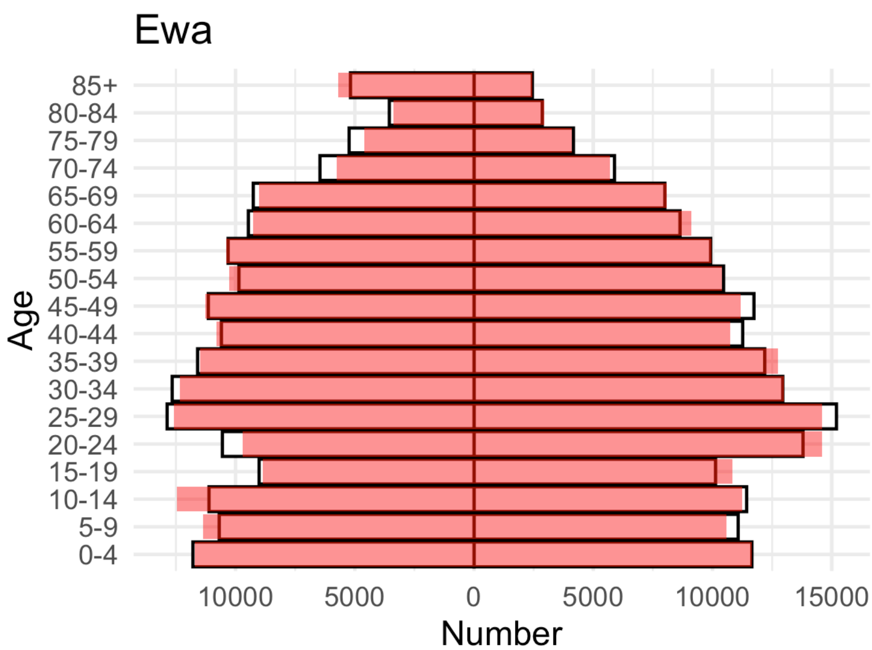
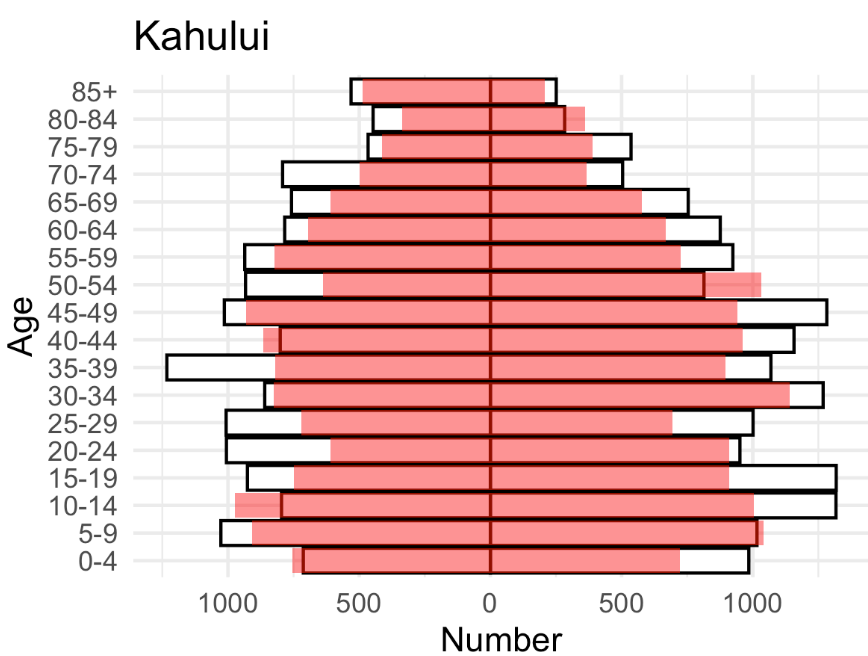


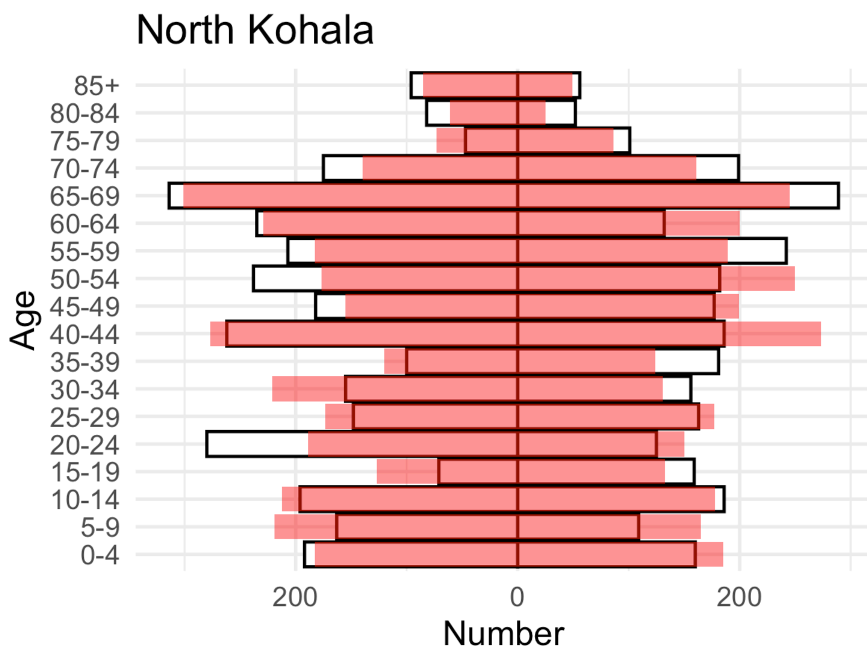
Figure 4. Distribution of age—comparison of model and validation data [insufficient match, adequate shape]



---

[5] Model source: experiments.covid_individuals
Validation source: 2019 5-year ACS summary file, table B01001

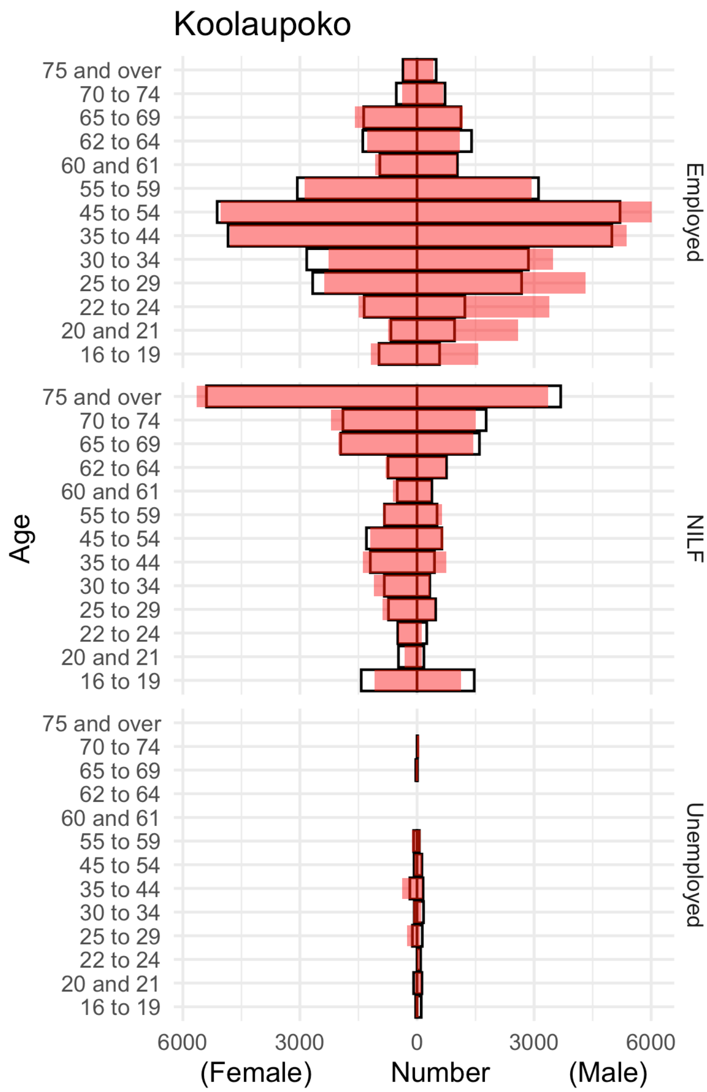
Figure 5. Distribution of age—comparison of model and validation data [adequate match, insufficient shape]



- For larger geographies, both the sample data (i.e., original PUMS data) and the synthesized data are sufficient. Specifically, the underlying population structure of target data is what one would expect of a "normal" pyramid *and* the level of congruence between target data and model data is adequate (see Figure 3 for example).
- Certain geographies have an adequate structure for target data, but the match between target and model data is inadequate. In the example shown in Figure 4, the model data underestimates the number of individuals for nearly all age groups, resulting in a systematic underestimation of the population in Kahului.
- In other instances, a geography may have an inadequate population structure from the source (i.e., target) data, though the match between target data and model data may be adequate (see Figure 5).

Employment status by sex by age[6]

Figure 6. Employment status by sex and age—comparison of model and validation data



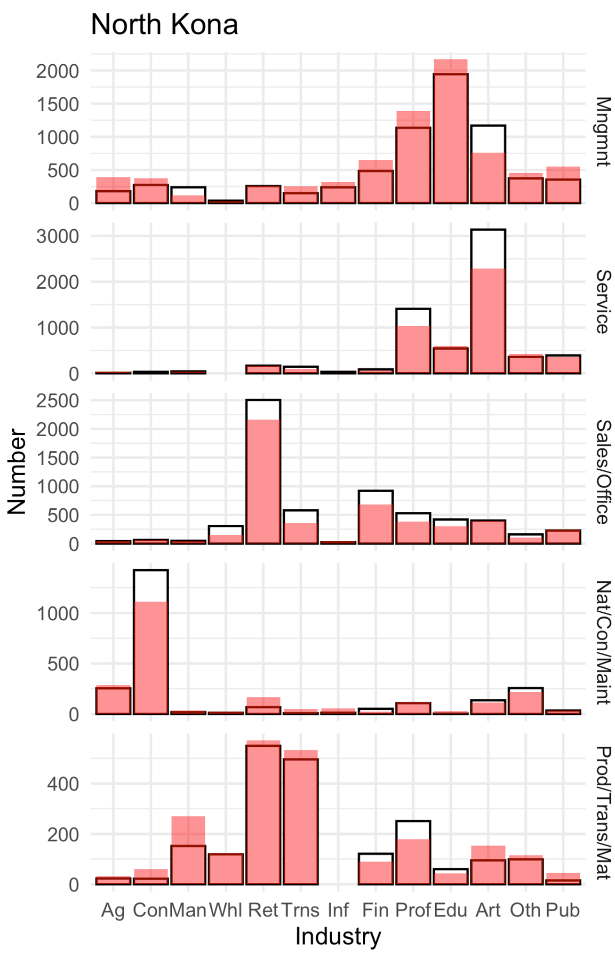
- There are 28,000 fewer employed civilians in the model data compared with target data (652,036 and 680,253, respectively). This results in an underestimation of the employed population for 2/3 of all geographies.
    - The model simulation corrects this number for the time period following the baseline estimate (see Figure 27).
    - The model-target discrepancy for employment is considerably lower for the total population (i.e., including military).
- Many small geographies are too incongruent for practical use.
- Five geographies show a bias toward young employed males: Koolauloa, Koolaupoko, Wahiawa, Waialua, and Waianae.

---

[6] Model source: experiements.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table B23001

Occupation by industry[7]

Figure 7. Occupation by industry—comparison of model and validation data

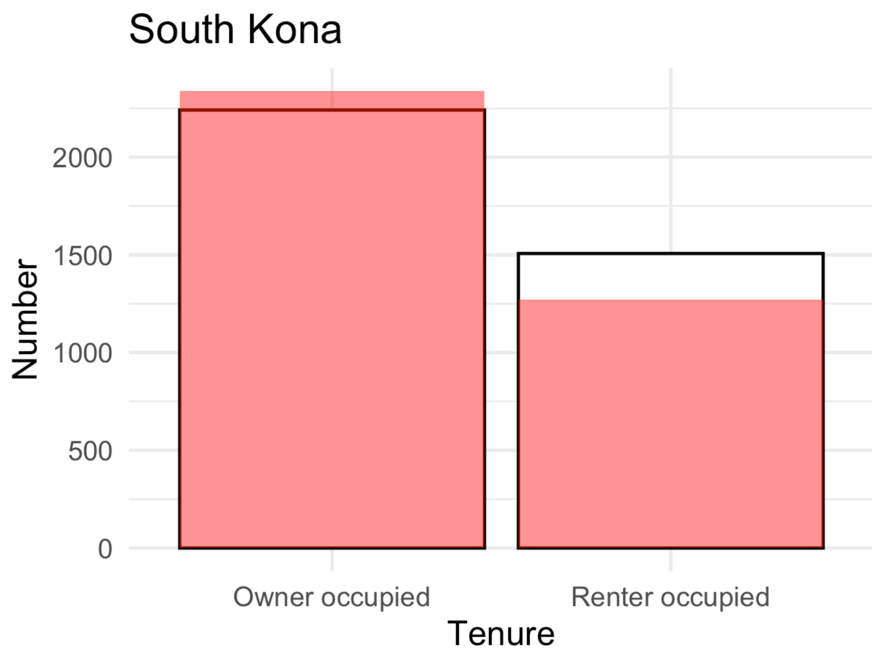


- The underestimation of 28,000 employed individuals in model data statewide (see *Employment status by sex by age* above) results in an underrepresentation of individuals in industry. Matching error (in terms of under- and overestimation) is not systematic (i.e., nearly equivalent in the error rates across groups); however, underestimates are underestimated more than overestimates are overestimated.
- Two industry categories that include *education services* and *accommodation and food services*, show the largest underestimation of workers (10,652 and 8,699, respectively) when comparing model data with target data.
  - The underrepresentation of individuals in education services and accommodation and food services means that model scenarios will likely understate the economic impact of COVID-19 on the state in general and on these industries in particular (see also Figure 28).

[7] Model source: experiments.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table C24050

Household Tenure[8]

Figure 8. Household tenure—comparison of model and validation data



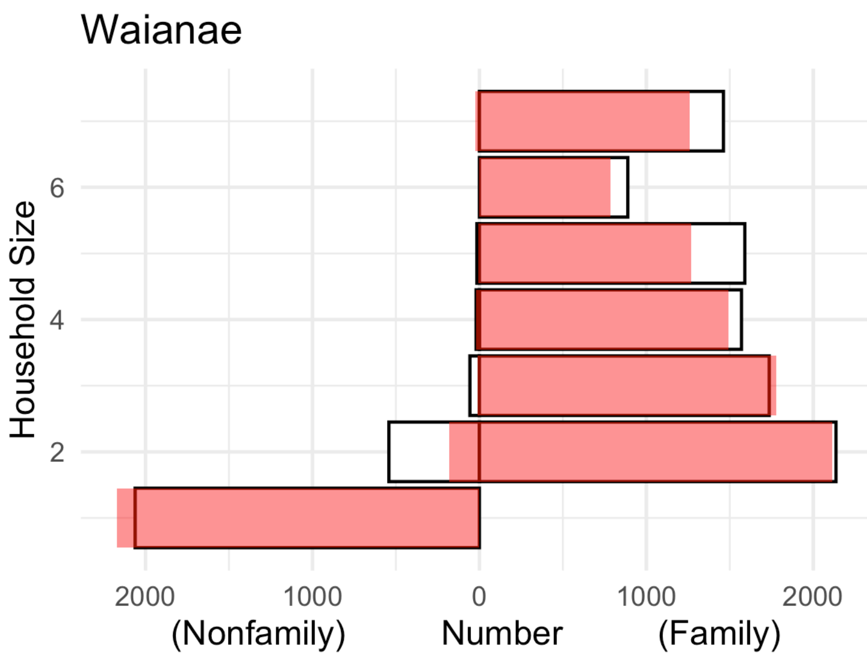
- Modeled estimates of tenure for several geographies are incongruent beyond many practical use cases, namely: Spreckelsville, Kalawao, Kahului, North Kohala, East Molokai, Hana, South Kona, Kihei, Honokaa-Kukuihaele, Kaumakani-Hanapepe, Wailuku.

---

[8] Model source: experiments.covid_households, sample_households
Validation source: 2019 5-year ACS summary file, table B25003

Household Type by Size[9]

Figure 9. Household type by size—comparison of model and validation data

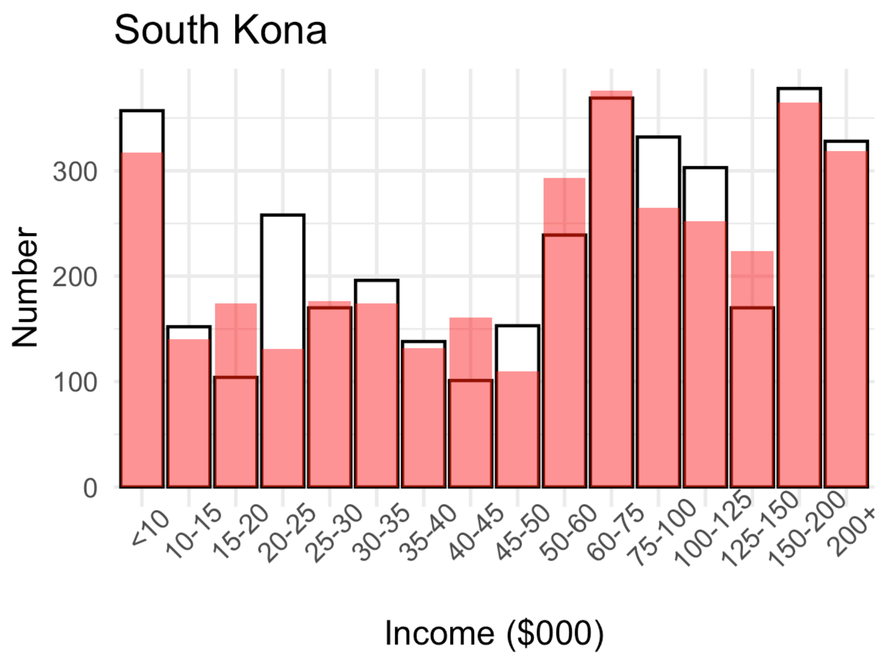


- Two- or more person households have more matching error than 1-person households, which are always nonfamily households. Nonfamily households with two or three people show the most frequently discrepant. The direction of the error is more likely to be an underestimation.
- "Code '9' – HHTYPE could not be determined" exists in the model dataset. This classification does not affect the total number of households in the model data. However, analyses utilizing family type (e.g., single-parent households) will likely have discrepancies with target data. Code "9" is unique to iPUMS data, but does not exist in the original Census data.

[9] Model source: experiments.covid_households, sample_households
Validation source: 2019 5-year ACS summary file, table B11016

Household Income[10]

Figure 10. Household income—comparison of model and validation data



- Similar to issues with *Age by Sex* above, the distribution of income in the target data appears inadequate for many small geographies.
    - Aggregated into larger income categories, model household income data match more closely with target data. However, it cannot be overstated the importance of income data to the model results, which rely on the ratio of income to household expenses (as represented by the ALICE framework) to determine need.

[10] Model source: experiments.covid_households
Validation source: 2019 5-year ACS summary file, table B19001

2.3 Distribution of Relevant Non-matching Variables

Determine the level of congruence between distributions of target (original summary file) data and current (synthetic population) data for non-matching variables relevant to the model's analytical purpose (i.e., projection of employment, income, and financial need)

Table 5. Level of congruence between model and validation data by non-matching variable [0 = perfect match]

Name	Travel_Time	N_Earners	Poverty_NHPI	Education	Benefits	Work_Time	Total
Spreckelsville	45.3803	16.4595		29.6908	20.0038	105.9502	
Hanalei	0.5084	0.0903	3.9501	0.1988	0.0358	0.4881	5.2715
North Hilo	1.1277	0.3405	1.6844	0.3489	0.0754	1.5517	5.1287
Waialua	0.6035	0.0936	2.6717	0.0864	0.0532	0.8325	4.3409
Honokaa-Kukuihaele	1.1718	0.4336	1.6008	0.2842	0.0875	0.745	4.3231
Lihue	0.7209	0.2457	2.6578	0.1687	0.1193	0.3915	4.3038
East Molokai	1.3556	0.4909	0.2981	0.2497	0.31	1.2551	3.9595
Kalawao	1.4031	0.4	0.8433	0.5276	0.2901	0.4125	3.8766
Hana	1.3836	0.369	1.1568	0.2703	0.097	0.5591	3.8357
Kaumakani-Hanapepe	0.9067	0.2516	1.3141	0.3118	0.045	0.4664	3.2957
Kau	1.0422	0.2188	0.7852	0.3202	0.1622	0.7423	3.2709
Paauhau-Paauilo	1.0762	0.2441	1.0502	0.2611	0.057	0.3424	3.031
Haiku-Pauwela	0.856	0.282	1.1032	0.2221	0.0666	0.496	3.026
Puhi-Hanamaulu	0.5731	0.1751	1.5091	0.2106	0.0644	0.366	2.8984
Lanai	0.9577	0.2033	0.8543	0.2533	0.075	0.5162	2.8597
Kapaa	0.548	0.1812	1.3974	0.2444	0.0326	0.3819	2.7856
Koolauloa	0.3739	0.2586	0.8314	0.3204	0.0841	0.7381	2.6065
West Molokai	0.5692	0.2368	0.181	0.4302	0.1382	1.0322	2.5877
Papaikou-Wailea	0.6992	0.085	0.9362	0.2711	0.0419	0.529	2.5625
Kahului	0.479	0.3258	0.5936	0.2806	0.209	0.668	2.556
Lahaina	0.5307	0.3099	0.8169	0.1645	0.0536	0.4845	2.3601
North Kohala	0.534	0.3824	0.8463	0.142	0.1411	0.3095	2.3552
Eleele-Kalaheo	0.6047	0.1704	0.6928	0.3816	0.0618	0.4151	2.3265
South Kona	0.7296	0.1932	0.8331	0.1099	0.0652	0.3266	2.2576
Kekaha-Waimea	0.7994	0.1686	0.6224	0.3315	0.0399	0.2655	2.2274
Koloa-Poipu	0.4508	0.126	0.7043	0.279	0.0196	0.5405	2.1201
Pahoa-Kalapana	0.6527	0.1647	0.5039	0.1552	0.2218	0.3621	2.0603
Hilo	0.5371	0.1132	0.8666	0.1788	0.0398	0.2746	2.0101
Makawao-Paia	0.6647	0.267	0.7407	0.0617	0.0482	0.2153	1.9977
South Kohala	0.5239	0.1604	0.6547	0.1721	0.1013	0.2834	1.8958
Kula	0.5882	0.1036	0.5643	0.1445	0.0329	0.4261	1.8596
Keaau-Mountain View	0.6598	0.1648	0.3857	0.1483	0.1517	0.3101	1.8204
Kihei	0.3927	0.124	0.9009	0.0811	0.0643	0.2482	1.8111
North Kona	0.4596	0.2015	0.4356	0.1162	0.1459	0.2281	1.5869
Wailua-Anahola	0.5516	0.0818	0.6131	0.095	0.0233	0.1929	1.5577
Wailuku	0.426	0.0884	0.4872	0.123	0.0453	0.3313	1.5011

Name	Travel_Time	N_Earners	Poverty_NHPI	Education	Benefits	Work_Time	Total
Waihee-Waikapu	0.4092	0.1666	0.397	0.1379	0.0911	0.2769	1.4787
Waianae	0.5009	0.1348	0.2727	0.1216	0.051	0.3744	1.4552
Wahiawa	0.2909	0.135	0.3973	0.1208	0.0074	0.2766	1.2279
Ewa	0.1304	0.0757	0.5662	0.0411	0.0156	0.1666	0.9956
Honolulu	0.0833	0.0571	0.6111	0.0393	0.0094	0.159	0.9593
Koolaupoko	0.2047	0.0345	0.2893	0.0504	0.0048	0.3598	0.9436
Hawaii	0.0741	0.0847	0.4495	0.028	0.0086	0.1608	0.8057

Travel Time to Work[11]

Figure 11. Travel time to work—comparison of model and validation data [sufficient match]

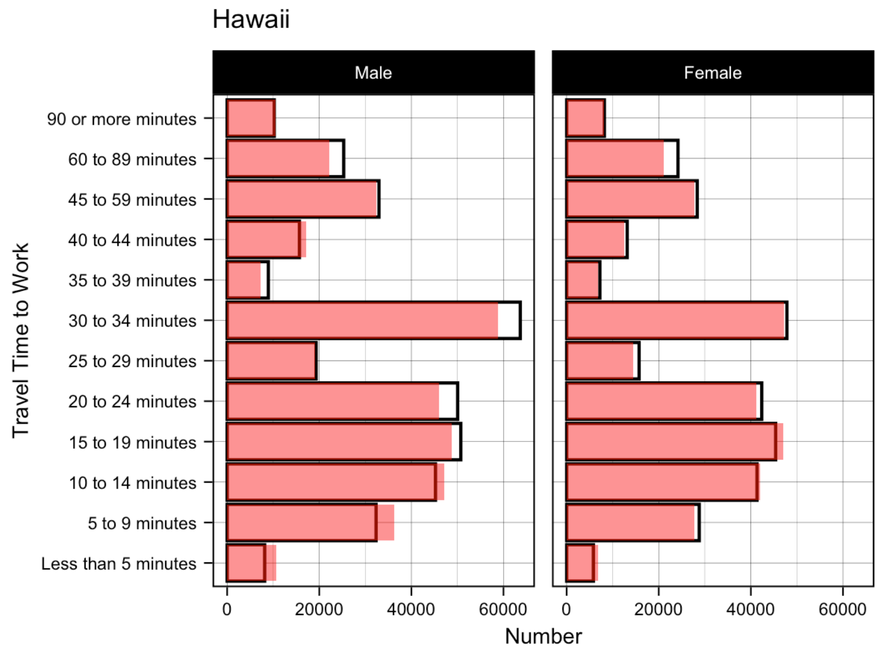
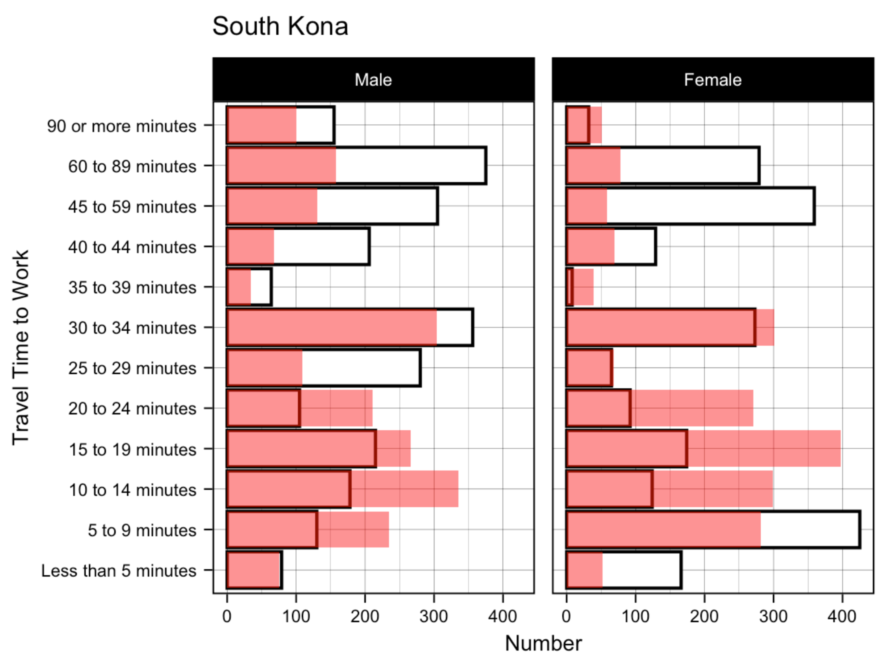
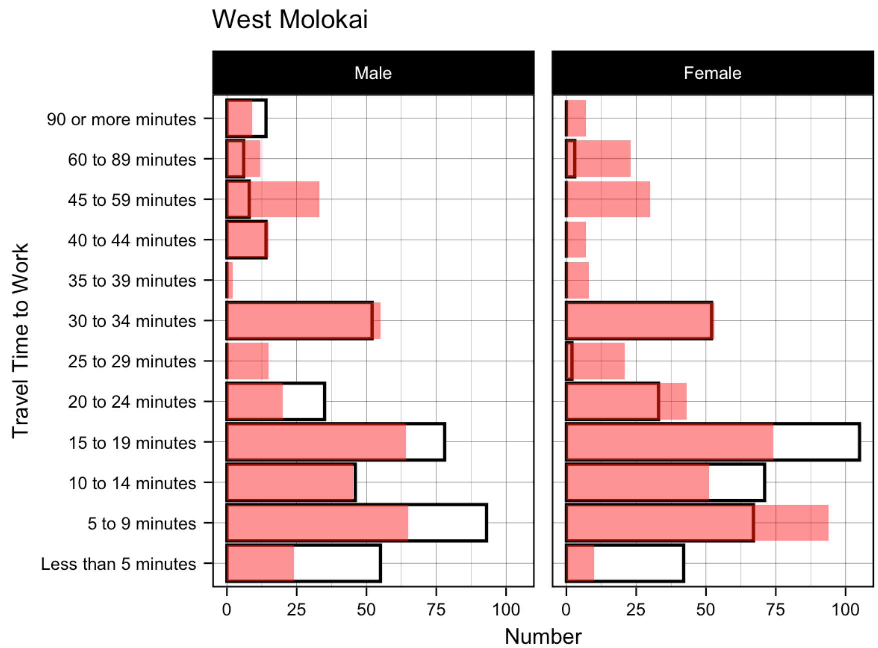


Figure 12. Travel time to work—comparison of model and validation data [bias toward overall I]



[11] Model source: experiments.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table B08012

Figure 13. Travel time to work—comparison of model and validation data [bias toward overall II]



- The modeled data have a tendency toward the overall distribution of travel time to work. This is most apparent in rural geographies where workers must travel far distances to work (e.g., South Kona) and in rural geographies where more workers work closer to home (e.g., West Molokai).
    - Variables not sufficiently correlated with matching variables will show less congruence than variables used for matching and variables highly correlated with matching variables.

Number of Earners in Family[12]

Figure 14. Number of earners in family—comparison of model and validation data [overall underestimation of no earners and 1 earner]

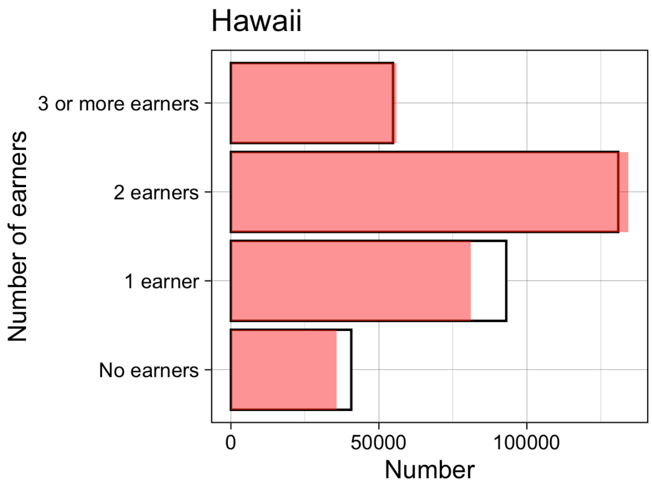
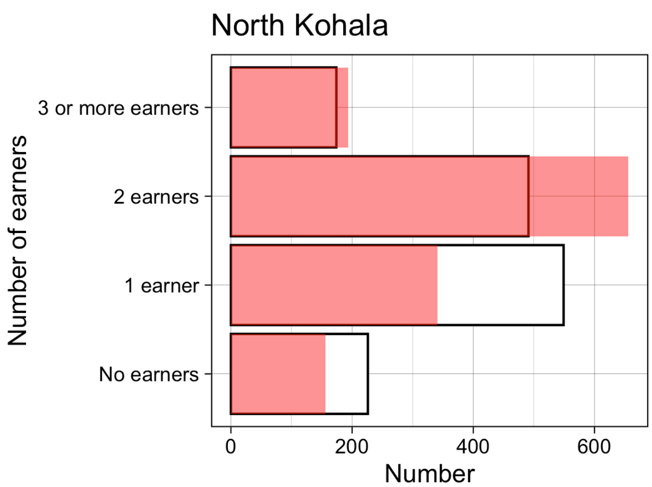


Figure 15. . Number of earners in family—comparison of model and validation data [small geography example]



- Overall, there is an underestimation of no-earner and 1-earner households and an overestimation of 2-earner and 3-or-more-earner households (Figure 14). The underestimation of no-earner and 1-earner households is apparent in numerous smaller level geographies (Figure 15).
    - Number of earners in the household is important when developing scenarios where individuals lose employment or wages. Households with multiple earners may be less impacted than households with one earner. The underestimation of 1-earner households suggests that the negative economic effects of COVID-19 will be underestimated.

---

[12] Model source: experiments.covid_households, experiments.covid_individuals, sample_household, sample_individuals
Validation source: 2019 5-year ACS summary file, table B19122

Poverty Status among NHOPI[13]

Figure 16. Poverty status among NHOPI—comparison of model and validation data [overestimation statewide]

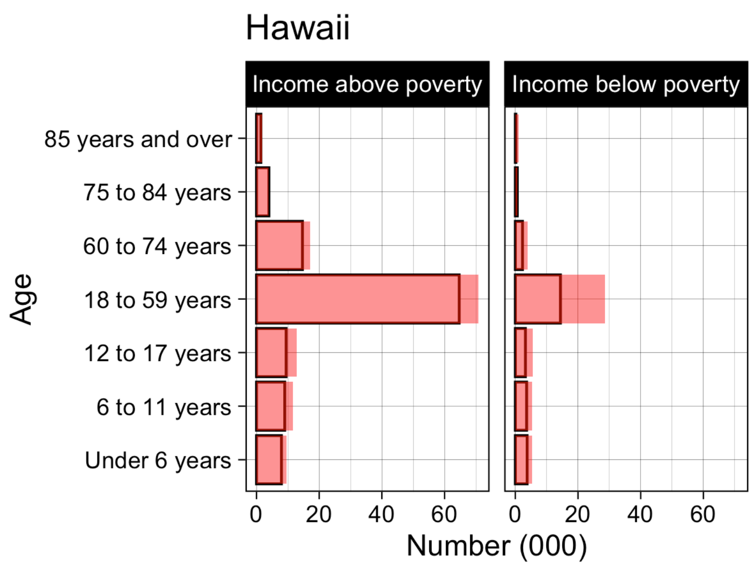
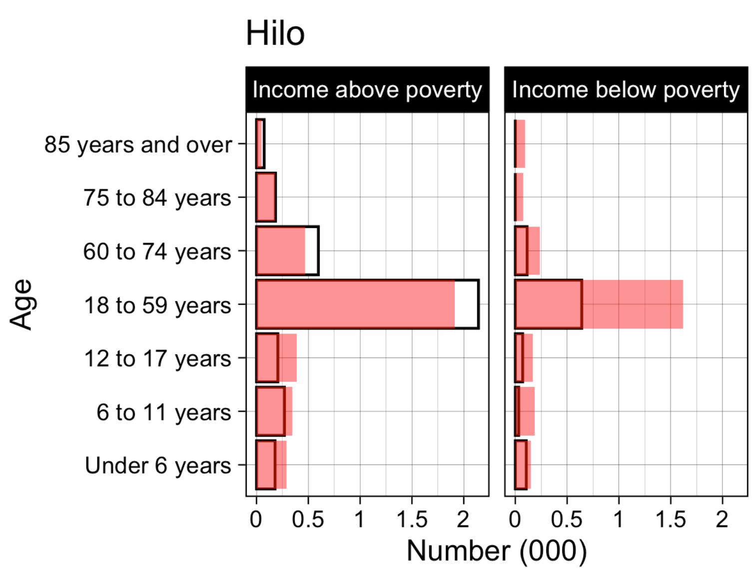


Figure 17. Poverty status among NHOPI—comparison of model and validation data [small geography example]



- There is an overestimation of roughly 40,000 NHOPI in the modeled data. Poverty status appears to be largely inaccurate for most geographies.

---

[13] Model source: experiments.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table B17020E

Educational Attainment[14]

Figure 18. Educational attainment—comparison of model and validation data [overall]

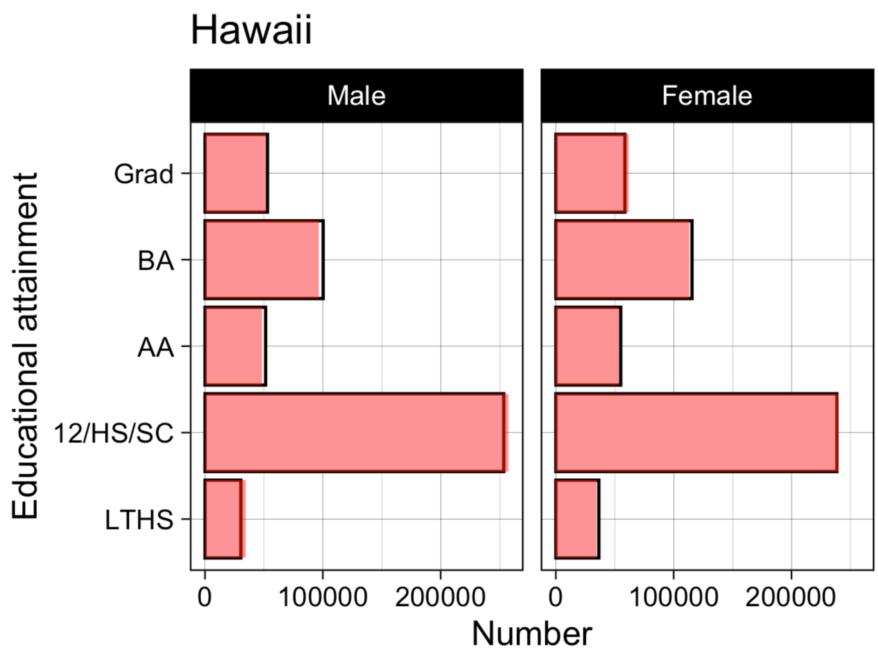
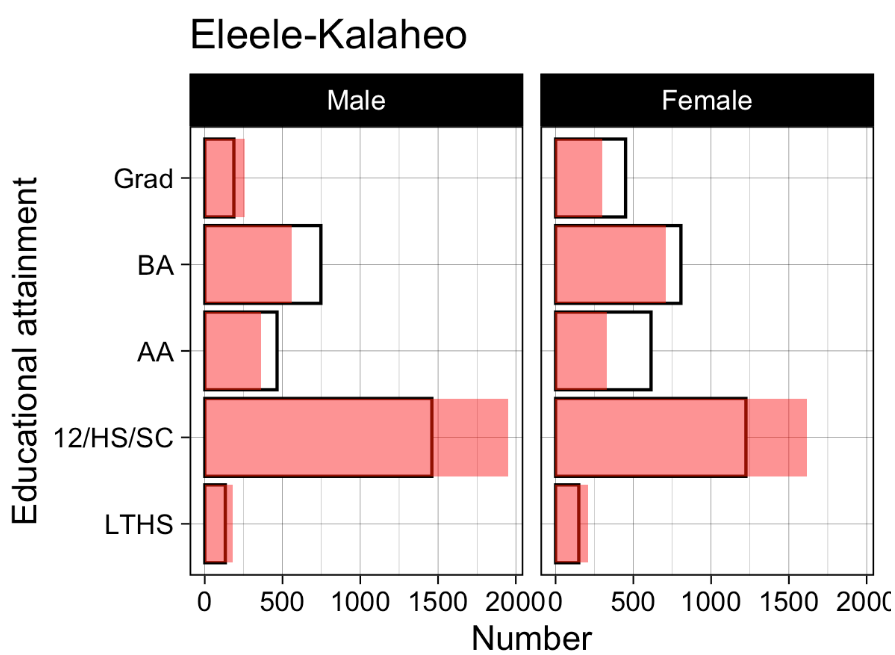


Figure 19. Educational attainment—comparison of model and validation data [small geography example]



[14] Model source: experiments.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table B15002

- The IPUMS variable for educational attainment 'EDUC' is recoded to be consistent with earlier versions of sample data. As a result, the categories offered are somewhat unintuitive and likely not ideal for matching (if used). It is recommended to use the Census version 'SCHL' for this variable in particular.
- Statewide, the distribution of educational attainment is sufficiently matched between model and target data. For some lower-level geographies, results show larger discrepancies between model and target data, which may affect both modeled scenarios and interpretation.

Public Assistance and Food Stamp Utilization[15]

Figure 20. Public assistance and food stamp utilization—comparison of model and validation data [overall]

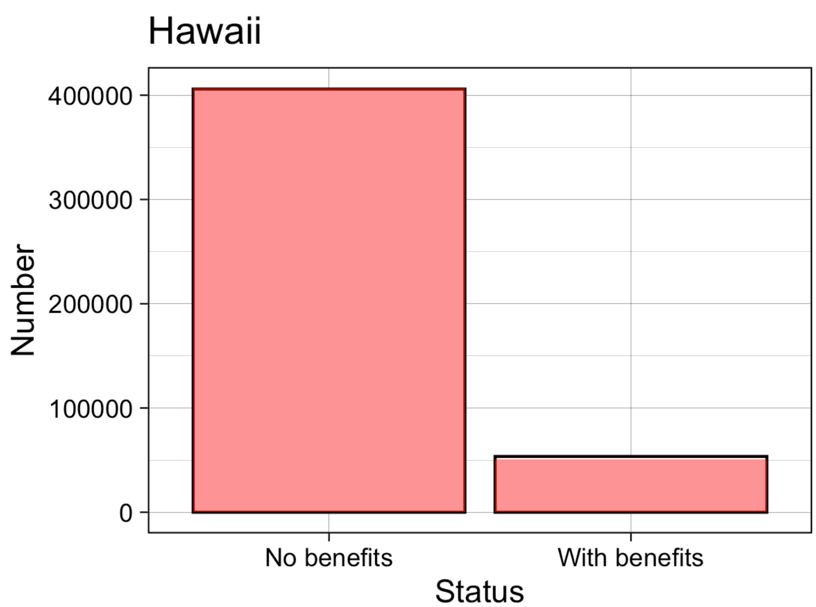
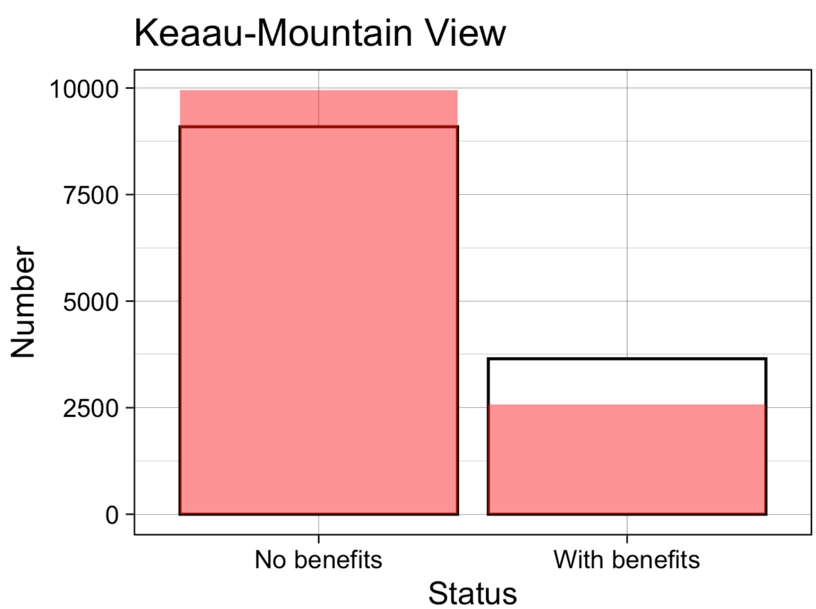


Figure 21. Public assistance and food stamp utilization—comparison of model and validation data [small geography example]



- There is no apparent systematic error associated with these data. In some cases, error may be significant enough to affect interpretation.

---

[15] Model source: experiments.covid_households, experiments.covid_individuals, sample_households, sample_individuals
Validation source: 2019 5-year ACS summary file, table B19058

Hours and Weeks Worked Among Workers[16]

Figure 22. Hours and weeks worked by worker—comparison of model and validation data [Hawaii]

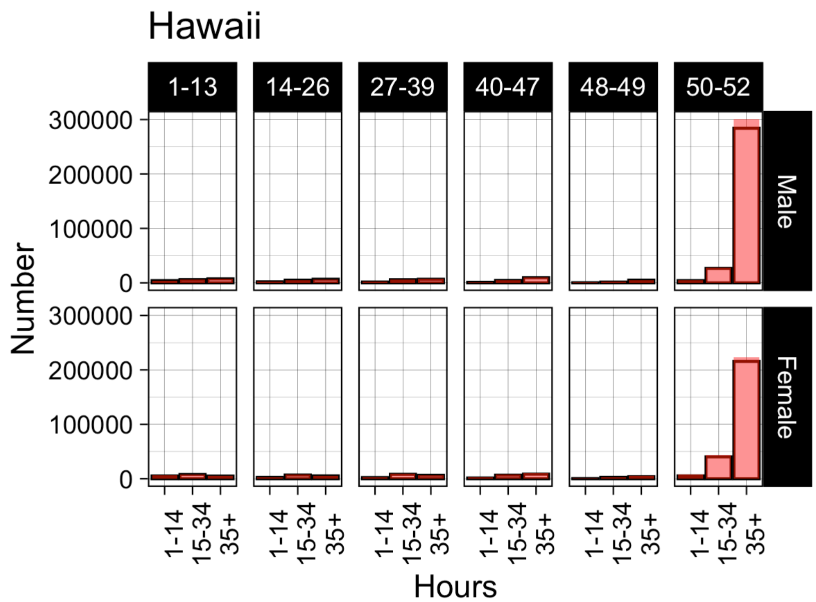
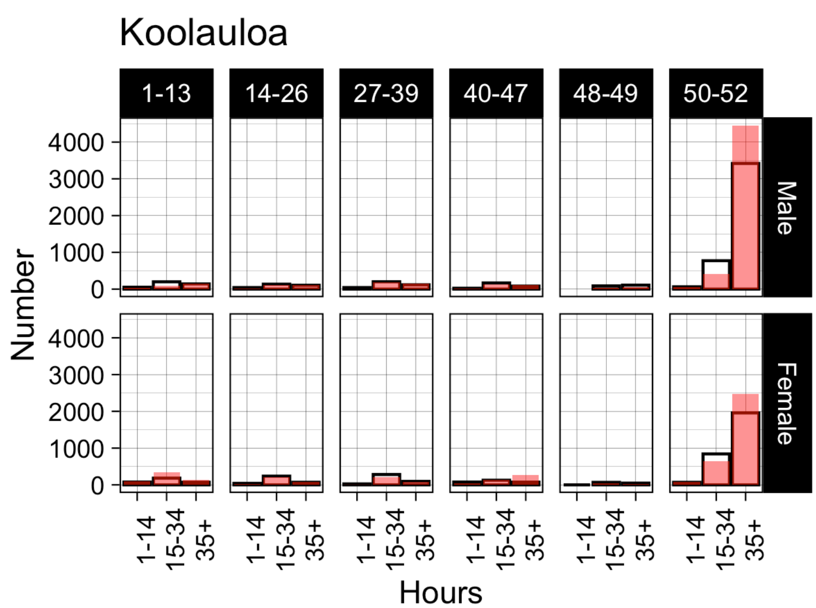


Figure 23. Hours and weeks worked by worker—comparison of model and validation data [small geography example]



- In general, there is an overestimation of full-time workers.

---

16 Model source: experiments.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table B23022

3. Temporal Shift

Assumption: ACS 2019 PUMS (5-year) data provide a reasonable estimate of the characteristics of people and households in Hawaii at the beginning of 2020. That is, person and household characteristics are relatively stable throughout the 2015–2019 data collection period.

Test: Review trends in characteristics of people and households throughout the ACS 2019 time period (i.e., 2015 to 2019).

Implications: *If the assumption is broken,* baseline measures are higher or lower than assumed. While this may not necessarily affect the direction or rate of change of trends in the modeled data; the magnitude will shift in the direction of the overestimate or underestimate.

Considerations: Issues in magnitude may occur among rapidly growing populations and/or periods of extreme currency inflation/deflation. These trends must be reviewed to determine the extent to which adjustments should be made to population totals and nominal values of currency.

Results: The **majority of variables reviewed were stable** throughout the reference period (2015–2019). Age, social security benefits, and housing tenure showed increasing or decreasing trends. Among the three variables, only age showed a significant difference between the mid-year estimate (representative of the target 5-year dataset) and the end-year estimate (representative of the beginning of 2020). The change amounts to a 1 percentage point increase in the retired-age population and a 1 percentage point decrease in the working-age population. **The synthetic population consists of a population that is slightly younger than the Hawaii population in 2020.**

3.1 Trends

Determine that 1) trends in years leading up to 2020 are flat. In cases where an upward or downward trend exists, determine that the 5-year aggregate time period 2015–2019 would not differ significantly from a hypothetical point estimate at the beginning of 2020.

Table 6. Trends in characteristics of the population, 2015–2019[17]

Variable	Level	Percent15	Percent16	Percent17	Percent18	Percent19
agegrp	1	21.8	21.5	21.4	21.3	21.2
agegrp	2	61.7	61.4	60.9	60.2	59.8
agegrp	3	16.5	17.1	17.8	18.5	19
SEX	1	50.4	50.4	50.2	50.1	50.2
SEX	2	49.6	49.6	49.8	49.9	49.8
MAR	1	40.1	40.6	40.5	40.9	40
MAR	2	5.1	5.1	5.1	5	4.9
MAR	3	7.7	7.1	8.1	7.2	7.6
MAR	4	1.1	1.1	0.8	1.1	1.1
MAR	5	46	46.1	45.4	45.8	46.4
HICOV	1	96.1	96.4	96.2	95.8	95.8
HICOV	2	3.9	3.6	3.8	4.2	4.2
ESR	1	57.5	58.4	57	58	57.2
ESR	2	1.2	1.3	1.8	1.5	1.5
ESR	3	3	2.6	2.3	2.2	2.3
ESR	4	3.7	3.4	4	3.6	4.2
ESR	6	34.6	34.3	34.9	34.7	34.9
industry	Ag	1.3	1.2	1.5	1.4	1.5
industry	Art	14.4	16.1	17.2	15.5	14.8
industry	Con	7.4	7.2	7.2	7.7	7.5
industry	Edu	18.7	19.6	18.2	20.3	19.5
industry	Fin	6.3	6.1	6.2	5.9	5.6
industry	Inf	1.9	1.4	1.3	1.1	1.2
industry	Man	3.1	2.5	2.9	2.8	2.7
industry	Mil	5.2	4.8	5.5	4.9	5.7
industry	Oth	4.6	4.1	4.2	4.8	4.2
industry	Pro	9.4	9.2	9.8	9.3	10.5
industry	Pub	8.3	8.3	7.8	7.8	7.5
industry	Ret	11	11	10.8	10.4	10.9
industry	Tra	5.1	5.8	4.8	5.9	5.8
industry	Une	0.8	0.9	0.7	0.4	0.7
industry	Who	2.4	1.9	2	1.6	1.9
occupation	man	33.2	32.3	31.6	32.8	33.9
occupation	mil	3.2	2.8	3.2	2.9	2.9

[17] Validation source: 2015 to 2019 1-year ACS PUMS

Variable	Level	Percent15	Percent16	Percent17	Percent18	Percent19
occupation	nat	10.5	9.8	10.1	10.1	9.7
occupation	pro	8.2	8.3	7.7	9.5	8.8
occupation	sal	23.8	23.8	23.8	21.2	21.5
occupation	ser	20.3	22.1	22.8	23	22.5
occupation	une	0.8	0.9	0.7	0.4	0.7
edattain	AA	8.4	8.2	7.9	8.8	8.4
edattain	BA	16	16.9	16.6	16.7	16.7
edattain	Grad	7.4	8.1	8	8.4	8.2
edattain	HS	23.3	23.9	24.8	23	24
edattain	LTHS	25.5	24.4	24.9	25.2	24.3
edattain	Some college	19.4	18.6	17.8	17.8	18.4
earnings	0	45.8	45.6	45.6	45.2	45.5
earnings	1	54.2	54.4	54.4	54.8	54.5
pubassis	0	98.6	98.3	98.7	98.9	98.8
pubassis	1	1.4	1.7	1.3	1.1	1.2
suppsec	0	98.1	98.2	98.2	97.8	98.4
suppsec	1	1.9	1.8	1.8	2.2	1.6
rtrmnt	0	90.5	90.6	90.9	90.6	87.3
rtrmnt	1	9.5	9.4	9.1	9.4	12.7
interest	0	88	88.5	88.1	88.9	87.6
interest	1	12	11.5	11.9	11.1	12.4
socsec	0	83.7	83.4	83.4	82.9	82.2
socsec	1	16.3	16.6	16.6	17.1	17.8
hawaiian	0	79.7	79.2	79.3	78.8	81.3
hawaiian	1	20.3	20.8	20.7	21.2	18.7
PUMA	100	16.5	16.6	16.7	16.9	17
PUMA	200	13.7	13.9	14.1	14.2	14.2
PUMA	301	7.9	7.6	7.8	7.4	6.6
PUMA	302	8	8.3	7.7	7.5	7.6
PUMA	303	7.3	7.3	7.3	7.3	7.5
PUMA	304	8.1	8.8	8.7	8	8.8
PUMA	305	9.3	7.9	8	8.5	7.8
PUMA	306	9.8	11.4	10.5	11.3	10.9
PUMA	307	10.9	10.5	10.2	9.8	10.9
PUMA	308	8.5	7.8	9.1	9.2	8.8
HHT	1	52.2	52.1	51.2	51.9	49.9
HHT	2	5.6	5.3	6	4.9	5.1
HHT	3	11.1	11.5	11.9	13.1	12.7
HHT	4	11.7	11.7	13	12	12.9
HHT	5	3.4	3.2	2.7	3.2	3.6
HHT	6	12	13.6	12.8	12.1	12.7
HHT	7	4.1	2.6	2.4	2.8	3.1
hhsize	1	23.7	25.4	25.8	24.1	25.6
hhsize	2	48.4	47.3	46.2	48.3	49.3

Variable	Level	Percent15	Percent16	Percent17	Percent18	Percent19
hhsize	3	23.7	22.9	24.2	24.1	21.3
hhsize	4	3.7	3.6	3.3	2.9	3.5
hhsize	5	0.5	0.8	0.5	0.6	0.4
TEN	1	37.7	37.8	37.1	37	38.6
TEN	2	19.3	19.6	21	20.7	22
TEN	3	40.6	39.9	39.1	39.9	37.5
TEN	4	2.4	2.7	2.8	2.4	1.9
FS	1	11.8	12	10.6	9.8	10.2
FS	2	88.2	88	89.4	90.2	89.8

- There is a decreasing proportion of working-age individuals (agegrp = 2); and increasing proportion of retirement-age individuals (agegrp = 3)
- There is a slightly larger proportion of individuals taking social security benefits (socsec = 1) at the end of the period compared with the beginning of the period.
- There is an increasing proportion of homes that are owned with a mortgage (TEN = 2); and smaller proportion of homes rented (TEN = 3).
- In all cases (except TEN = 2), the middle year (i.e., 2017) is the most representative of the estimates in the 2019 PUMS 5-year dataset (2018 is the most representative for TEN = 2).
- Only the change in age (i.e., agegrp) from 2017 to 2019 is significant. However, the magnitude of the change is relatively small (about 1 percentage point).

Model Constructs

There are two components of the model output that are critical to interpretation and utilization. *Household cost* is an estimate of the amount of income required to cover basic needs. *Income* is the total amount of money a household receives through earnings or other sources. These two components are used to calculate a ratio ($\frac{income}{cost}$), where any number greater than or equal to 1 indicates sufficient income to cover household costs and any number less than 1 indicates insufficient income to cover household costs.

This section tests the assumption that the ALICE framework can be used to understand income requirements to cover the cost of basic needs and that the framework is implemented appropriately in the synthetic population.

4. ALICE Framework

Assumption: The ALICE threshold is a reasonable proxy for the minimum level of income required to satisfy basic needs of individuals in households in Hawaii. Further, the ALICE framework has been appropriately applied to the synthesized data.

Test: Compare ALICE calculation results to alternative studies on economic sufficiency. Review implementation of the ALICE framework in the synthetic population.

Implications: *If the assumption is broken,* the ALICE framework cannot provide an estimate of need for the state. This may be because the threshold is not set at the correct level or because the generated synthetic data do not identify the correct households in need.

Considerations: The ALICE framework was developed in consideration of a "typical family household." However, as living arrangements and needs vary dramatically, the ALICE estimates serve only as a proxy of need and do not depict the complete picture of circumstance and need. Further, ACS data do, or are likely to, underrepresent individuals in "non-typical" living situations. As a result, both the ALICE framework and ACS data, on which the framework relies, are better representations of individuals and households in typical living situations.

As a practical example, the higher proportion of multigenerational households in Hawaii (compared with the rest of the United States) means that an estimate of need (based on the ALICE threshold) is lower than an estimate given where there are fewer multigenerational households. This is because many of the costs associated with occupying a separate household (e.g., housing, food) are avoided or minimized when multiple families live in the same household (both in a practical sense and in terms of an ALICE calculation).

Results: There are several improvements that can be made to the implementation of the ALICE criteria. For **housing**: efficiency sized (i.e., studio) housing units should be added for 1-person households; the number of bedrooms assigned to households as a cost should increase every two people, not with every additional person. For **childcare**: geography specific costs by age of child may be used; childcare costs should cover children up to age 17; and the determination of how many caregivers are in the household should exclude adults enrolled in school. For **food**: food costs may be adjusted based on household size, where cost savings are realized incrementally with additional people in households; food costs may be inflation adjusted using CPI data. For **transportation**: use public transportation cost when 8 percent or more of the population uses public transportation and the full cost of private transportation in all other geographies (this can be reviewed at the county subdivision level). For **healthcare**: tailor out-of-pocket healthcare costs for households, instead of using a flat rate. For **taxes**: taxes should be estimated using the marginal rates applicable to each household, taking into consideration the relevant deductions and exemptions.

ALICE based calculations are consistent with other studies that calculate the minimum level of income required to cover basic needs. When applied to the synthetic population at the county subdivision level, **most geographies underestimate the proportion of households that are below the threshold to be identified as having need.**

4.1 Housing[18]

ALICE Criteria

- 40th percentile of HUD FMR
- Gross rent = rent + utilities (electricity, gas, water/sewer, and trash)
- County adjusted rates
- 1-person = Studio
- 2-person = 1-bedroom
- 3-person+ = 2-bedroom

Model Discrepancies/recommendations

- Missing efficiency sized units in SQL database
- Increase "should have bedrooms" every two people starting at 2-bedrooms
- Consider using mortgage amount paid as housing cost for owner-occupied households

[18] Model source: experiments.hud_fmr, Alice base calculation.sql
Validation source:
https://www.huduser.gov/portal/datasets/fmr/fmrs/FY2020_code/2020state_summary.odn

4.2 Childcare[19]

ALICE Criteria

- Infants
- 4-year-old = 75th percentile
- School-age = 3/8 the cost of 4-year-old care

Table 7. Childcare costs

Geography	Infant	Age 4	School-age
East Hawaii	650	650	243.75
West Hawaii	850	733	274.88
Maui	700	700	262.50
Molokai	450	450	168.75
Lanai	-	-	
Oahu	800	800	300.00
Kauai	700	700	262.50

Model Discrepancies/recommendations

- Add "not in school" to determination of caregiver at home
- Increase school-age threshold from "under 15" to "17 years and under"
- Use geography specific childcare costs for infant, 4-year-old (3 and 4 years of age), and school-age (5 to 17 years of age) to compute total childcare costs; instead of a flat rate of $500 for < 5 year of age and (3/8)*500 for 6 to 14 years of age.

4.3 Food[20]

ALICE Criteria

- Thrifty level USDA food plan
  - adjusted by household composition;
  - month of June;
  - Hawaii adjusted

Model Discrepancies/recommendations

- Use household size adjustments as described in USDA food plan
- Use CPI data for inflation adjustment factor, instead of 2% rate.

4.4 Transportation[21]

ALICE Criteria

---

19 Model source: Alice base calculation.sql, experiments.childcare_alice
Validation source: https://humanservices.hawaii.gov/bessd/files/2019/04/Hawaii-Child-Care-Market-Rate-Study-2018-final.pdf

20 Model source: experiments.foodcosts_alice
Validation source: https://fns-prod.azureedge.net/sites/default/files/media/file/AKHI1stHalf2019.pdf

21 Model source: Alice base calculation.sql

- Cost of public transportation when 8 percent or more of the population uses public transportation

or

- Gas, oil, maintenance, and minimum liability insurance for private transportation when less than 8 percent of the population used public transportation
- (Avg miles/day * cost/mile * 300) + license and fees + depreciation + insurance
- 1 car per family (car size increases with family size)

Model Discrepancies/recommendations

- From the PUMS data Town and Ewa are the only areas where public transportation is practical; all other areas should use private transportation cost. Use the summary file data to determine which geographies will be assigned public transportation costs and which geographies will be assigned private transportation costs
- Use full costs for private as outlined in ALICE methodology (where applicable)

4.5 Healthcare[22]

ALICE Criteria

- Premiums + out-of-pocket costs
- Premiums = employee contribution to employer sponsored plan
- Out-of-pocket cost = average out-of-pocket cost for families under 65 years of age by income

Model Discrepancies/recommendations

- Tailor out-of-pocket costs by household size

4.6 Technology[23]

ALICE Criteria

- Lowest cost cell phone plan (Consumer Reports)

Model Discrepancies/recommendations

- N/A

---

[22] Model source: Alice base calculation.sql, experiments.healthcosts_alice
Validation source:
https://meps.ahrq.gov/data_stats/summ_tables/insr/state/series_7/2018/tviic2.pdf;
https://meps.ahrq.gov/data_stats/summ_tables/insr/state/series_7/2018/tviid2.pdf;
https://meps.ahrq.gov/data_stats/summ_tables/insr/state/series_7/2018/tviie2.pdf;
https://www.bls.gov/cex/2018/CrossTabs/agebyinc/x45to54.PDF

[23] Model source: Alice base calculation.sql
Validation source: https://www.consumerreports.org/cell-phone-service-providers/best-low-cost-cell-phone-plans/

4.7 Miscellaneous[24]

ALICE Criteria

- 10 percent of the budget total (including taxes)

Model Discrepancies/recommendations

- N/A

4.8 Taxes[25]

ALICE Criteria

- Federal, state, and payroll taxes
- Includes federal and state deductions and exemptions
- Includes federal Child Tax Credit and Child and Dependent Care Credit

Model Discrepancies

- The tax formula getTaxes() used to compute taxes paid for each household is a simplified formula— $\frac{income_{household}}{size_{household}} * rate$ —for which the rate varies by income level and is applied for federal and state taxes. This formula underestimates the amount of taxes paid for Hawaii households and does not take into account deductions and tax credits for dependent children or tax filing status.
    - For example, a 3-person household with an annual household income of $77,271, pays $4,379 dollars in taxes in the model; one-quarter to one-third the amount of taxes typically paid by households of this size and income level.

---

24 Model source: Alice base calculation.sql

25 Model source: function getTaxes()

4.9 Comparison of ALICE to Census's poverty-to-income ratio variable[26]

Determine the adequacy of ALICE based calculations to represent economic need.

Figure 24. Conversion chart of ALICE to Census's Poverty-to-income ratio

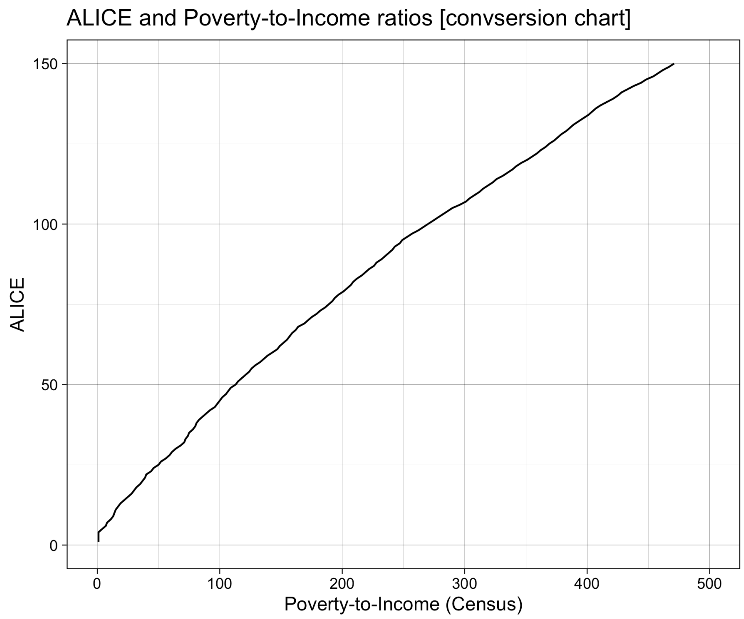
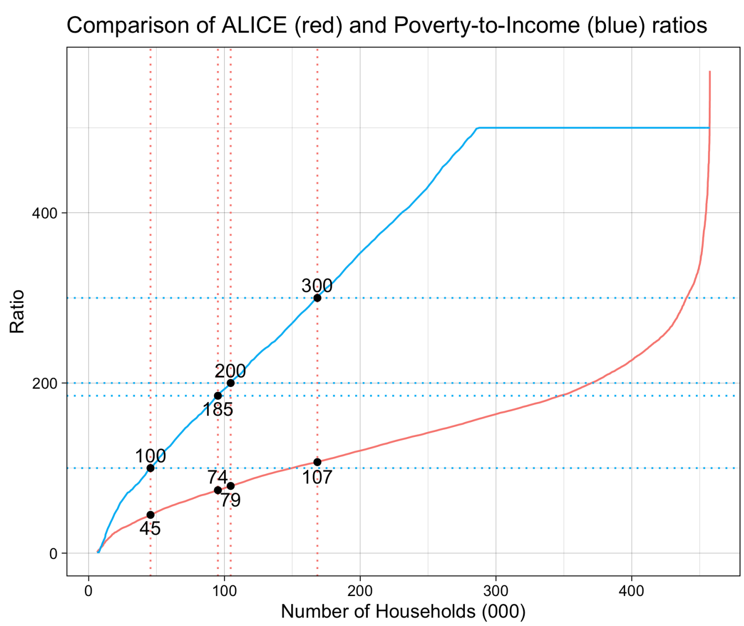


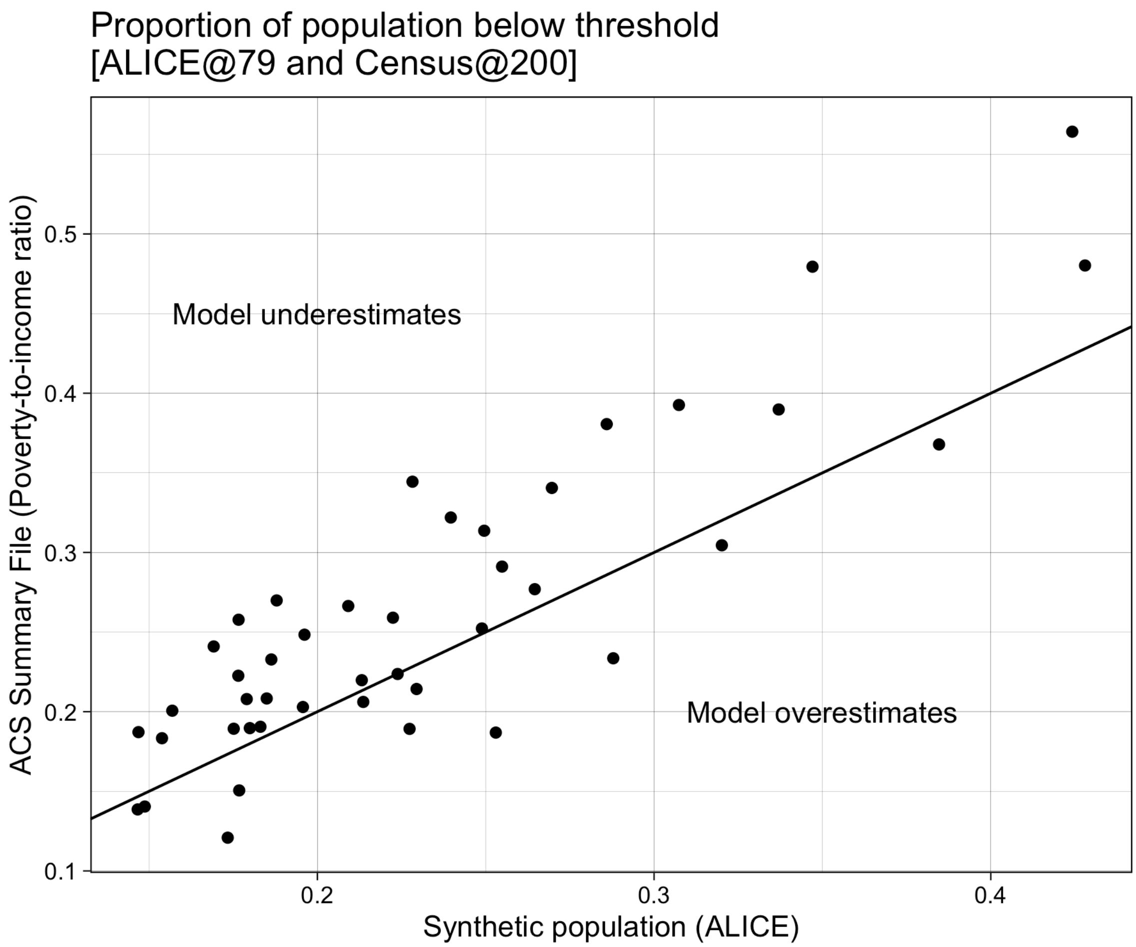
Figure 25. Comparison of ALICE and Poverty-to-income ratios, number of households below each level



---

26 Model source: experiments.alice_combined, experiments.covid_households, sample_individuals
Validation source: 2019 5-year ACS PUMS; summary file table C17002

Figure 26. Proportion of households below Broad-Based Categorical Eligibility (BBCE) [ALICE and Poverty-to-income ratio comparison]



- A previous study on income requirements to cover basic costs for Hawaii households (based off of the Economic Policy Institutes methodology) suggests that a minimum livable income is approximately equivalent to 300 percent of the federal poverty threshold. Comparing the poverty-to-income ratio variable in the ACS to the ALICE ratio of income to cost in the synthetic population, 300 percent is equivalent to 107 percent for ALICE. These results show consistency between methodologies in determining the minimum level of income required to cover basic needs.
- Comparing the proportion of households that are below 200 percent of the federal poverty level (i.e., SNAP eligible under Broad-Based Categorical Eligibility), using the poverty-to-income ratio variable, to the proportion of households below the ALICE equivalent for 200 percent of the federal poverty level (i.e., 79; Figure 25), the analysis finds that most Hawaii geographies at the county subdivision level underestimate the proportion of households below the threshold (Figure 26).

5. Construct Validity

Assumption: Although the individual level assignments of workers to workplaces and business to business transactions are 1) likely incorrect and 2) impossible to verify; at an aggregate level, these assignments are representative of economic activity within the state and at various levels of industry and geography, as defined by worker engagement and business operations such that modeled levels of unemployment and wages are representative of current/actual levels of employment and wages.

Test: Compare modeled levels of unemployment and wages with alternative estimates.

Implications: *If the assumption is broken,* assignment of worker to workplaces and business-to-business transactions as per the model design does not provide probable estimates of unemployment and wages among the population in Hawaii.

Considerations: Individual level comparisons of workers to workplaces (i.e., where individuals work) and business-to-business transactions (i.e., which businesses supply items to which businesses) cannot be verified. For the purpose of modeling economic activity in the state, only the levels of engagement (workers) and activity (businesses) for the various industries by geography at an aggregate level are necessary to derive probable estimates of employment and wages among the population in Hawaii.

Results: Some geographies show employment levels that are too high or too low, compared with source ACS data tables. **At a statewide level, the modeled decrease in unemployment is consistent with external data sources.** However, a **slight lag exists** in the model and the subsequent recovery is overestimated and decline underestimated. The current model appears **unable to produce reasonable estimates for subgroups** (including geographic and population). For example, the decline in employment within the Leisure and Hospitality sector was underestimated by about 37,000.

5.1 Size of Employed Population

Determine that employment size by geography matches between model and target data at the baseline.

Table 8. Employment size by geography—comparison of model and validation data [includes Armed Forces][27]

Geography	Model Estimate	Validation Estimate	Difference	Percent Difference
Spreckelsville	3469	105	3364	3203.8
Honokaa-Kukuihaele	1984	1624	360	22.2
East Molokai	1877	1562	315	20.2
Hana	795	688	107	15.6
Kalawao	57	52	5	9.6
Lihue	3927	3731	196	5.3
Kau	2833	2716	117	4.3
North Hilo	605	580	25	4.3
Koloa-Poipu	3181	3086	95	3.1
Koolauloa	9899	9614	285	3
Waialua	6776	6586	190	2.9
Lahaina	12805	12460	345	2.8
Pahoa-Kalapana	3944	3848	96	2.5
Kaumakani-Hanapepe	2174	2129	45	2.1
Wahiawa	22367	21960	407	1.9
Eleele-Kalaheo	4866	4815	51	1.1
Hilo	20970	20773	197	0.9
Makawao-Paia	10945	10880	65	0.6
Paauhau-Paauilo	1174	1171	3	0.3
Waianae	19555	19505	50	0.3
West Molokai	854	856	-2	-0.2
North Kohala	2928	2938	-10	-0.3
Kula	6603	6627	-24	-0.4
Ewa	180354	181433	-1079	-0.6
Hawaii	715951	721795	-5844	-0.8
Honolulu	210415	212406	-1991	-0.9
Koolaupoko	59465	60034	-569	-0.9
Keaau-Mountain View	13848	14039	-191	-1.4
Puhi-Hanamaulu	5579	5660	-81	-1.4
South Kohala	9473	9648	-175	-1.8
Kapaa	4091	4181	-90	-2.2

[27] Model source: experiments.covid_individuals, sample_individuals
Validation source: 2019 5-year ACS summary file, table B23001

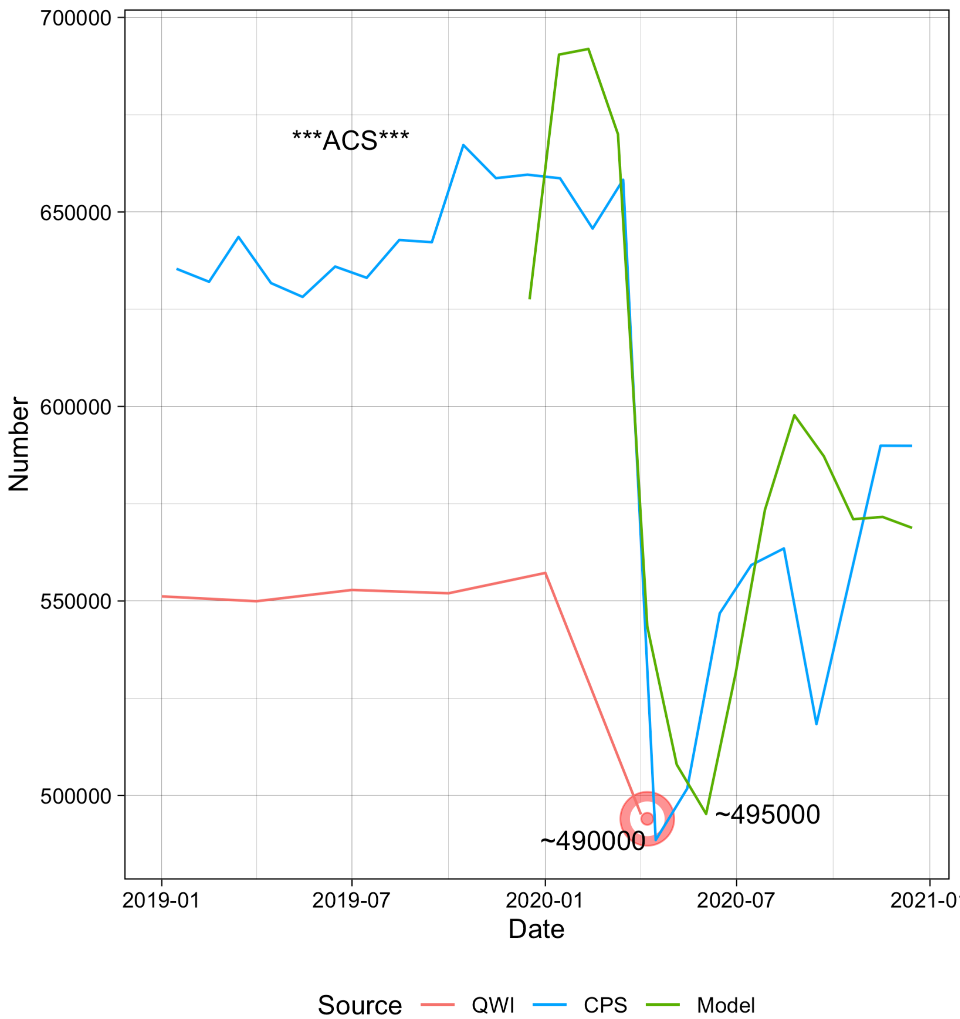
Geography	Model Estimate	Validation Estimate	Difference	Percent Difference
Wailuku	11775	12113	-338	-2.8
Kekaha-Waimea	2748	2830	-82	-2.9
Wailua-Anahola	6857	7166	-309	-4.3
Haiku-Pauwela	4925	5146	-221	-4.3
Waihee-Waikapu	3843	4059	-216	-5.3
Hanalei	2841	3002	-161	-5.4
Lanai	1266	1345	-79	-5.9
North Kona	22340	23897	-1557	-6.5
Kihei	12802	13845	-1043	-7.5
South Kona	4629	5147	-518	-10.1
Papaikou-Wailea	1740	1958	-218	-11.1
Kahului	12372	15580	-3208	-20.6

- Spreckelsville, Honokaa-Kukuihaele, East Molokai, and Hana have modeled employment levels greater than that observed in the ACS data.
- Kahului, Papaikou-Wailea, and South Kona, have modeled employment levels less than that observed in the ACS data.

5.2 Comparison of Modeled Unemployment and Wages to LEHD Data

Determine that modeled levels of unemployment are consistent with available sample data.

Figure 27. Model comparison of unemployment for the civilian, non-GQ population[28][29]



- Three data sources are used as comparisons to model unemployment.
  - Quarterly Workforce Indicators (QWI) data represent private sector jobs, reported quarterly.

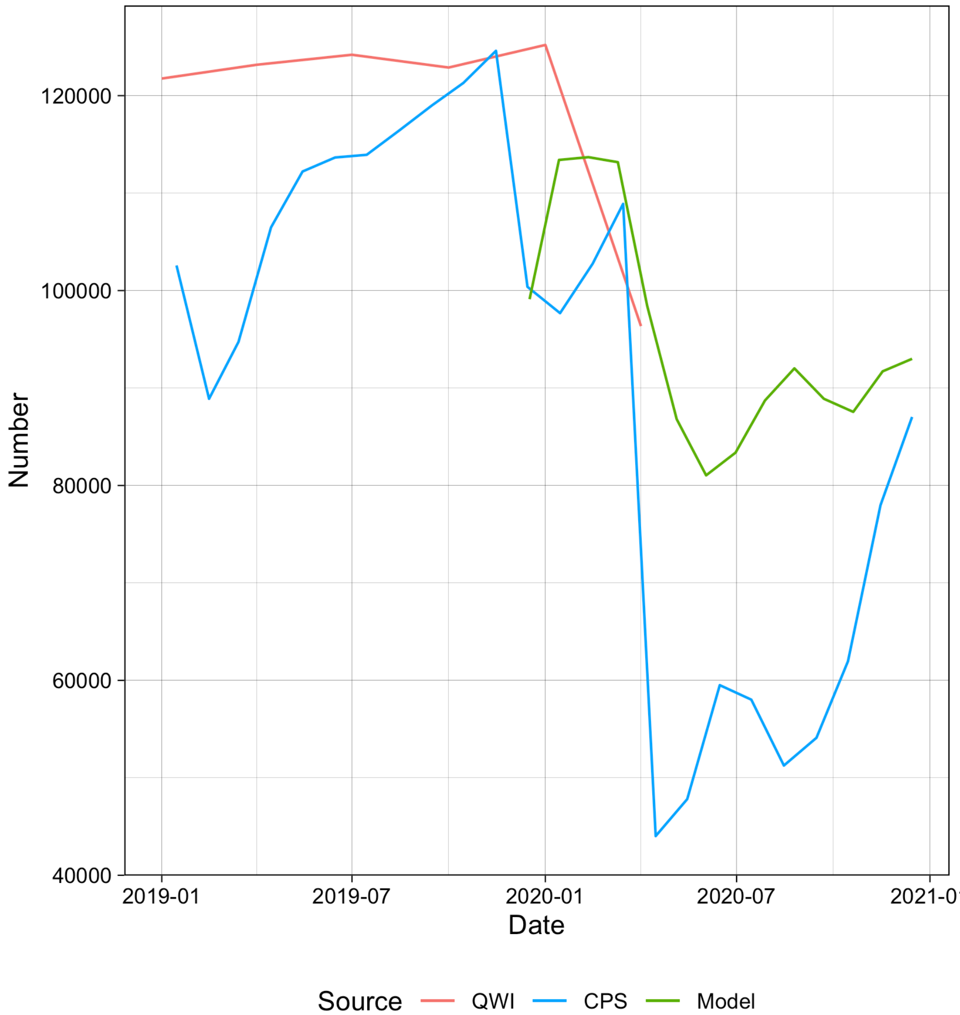
---

[28] Model source: covid_individuals_analytics_instantaneous, sample_households, synthesized households, sample_individuals, synthesized_individuals
Validation source: Quarterly Workforce Indicators file qwi_hi_rh_f_gc_ns_oslp_u, Current Population Survey basic monthly January 2019 to December 2020, 2019 1-year ACS PUMS

[29] The model results shown here are based off of the SQL database for the week of February 22, 2021; the subsequent update to the database showed results that were less congruent with external data sources.

    - Current Population Survey (CPS) data represent employment for the civilian population, reported monthly.
    - American Community Survey (ACS) data are subset to represent employment among the civilian population for the 2019 time period.
    - Modeled data (Model) uses employment scenario 3 to show changes in employment every 4 weeks.
- The magnitude and timing of reductions in modeled employment are relatively consistent with other data sources for the total population, statewide.
    - CPS data show a minimum employment level of about 490,000 in April 2020; compared with a minimum employment level of about 495,000 in the modeled data for May/June.
        - The lag in modeled data may be due to the amount of time to realize the economic impacts from decreases in visitors, which may be adjusted.
- An initial recovery of employment to a total of 563,000 occurred by August 2020, according to CPS data. In the modeled data, a larger recovery to 598,000 was estimated.

Figure 28. Model comparison of unemployment for the civilian, non-GQ population [Leisure and Hospitality][30][31]



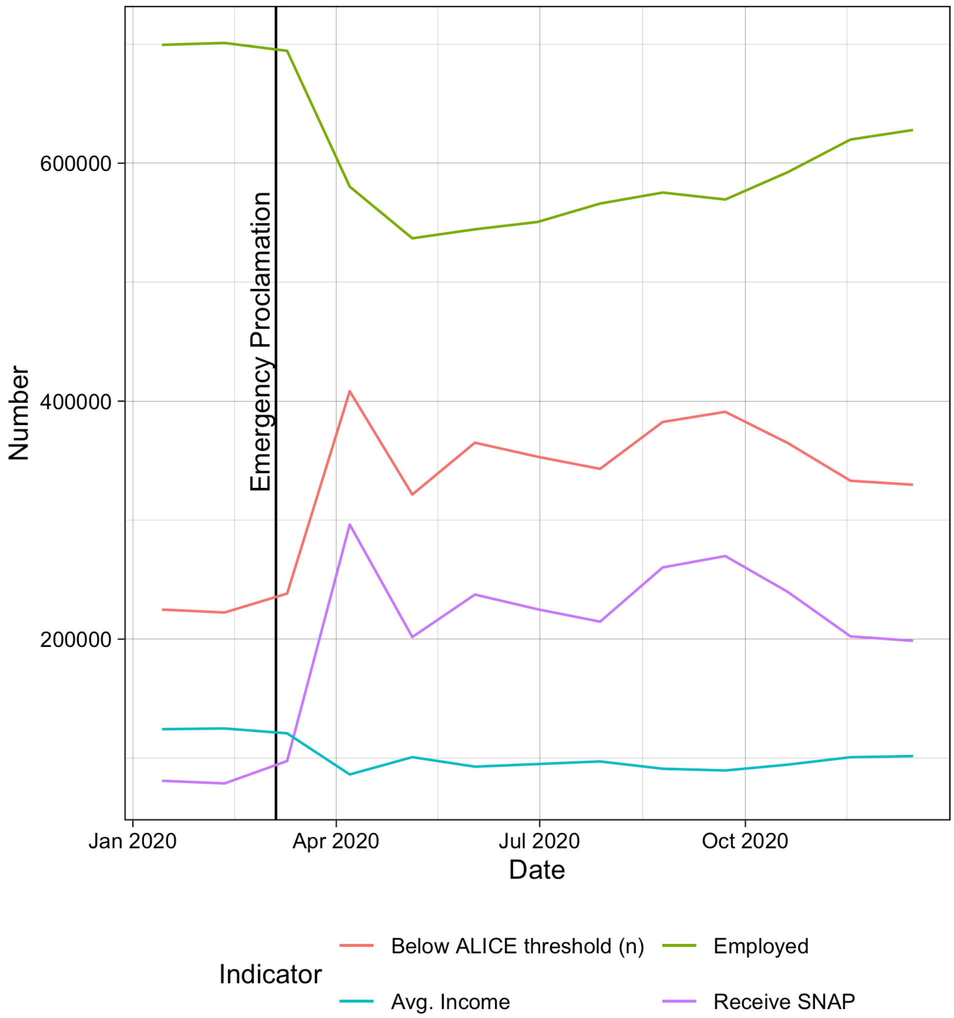
- According to CPS data, minimum employment reached roughly 44,000 in the Leisure and Hospitality sector in April 2020. Modeled estimates showed minimum employment at about 81,000 in June 2020 for the Leisure and Hospitality sector—a difference of 37,000.

---

30 Model source: covid_individuals_analytics_instantaneous, sample_households, synthesized households, sample_individuals, synthesized_individuals
Validation source: Quarterly Workforce Indicators file qwi_hi_rh_f_gc_ns_oslp_u, Current Population Survey basic monthly January 2019 to December 2020

31 The model results shown here are based off of the SQL database for the week of February 22, 2021; the subsequent update to the database showed results that were less congruent with external data sources.

Figure 29. Comparison of model components[32]



- Average incomes of individual are too high—over $100,000 before the pandemic and around $100,000 during the pandemic.
- SNAP recipiency is aligned with ALICE individuals.

[32] Model source: covid_individuals_analytics_instantaneous, sample_households, synthesized households, sample_individuals, synthesized_individuals

Data Inputs

External data are used to develop scenarios that output estimates of unemployment, income, and household need (measured using the ALICE framework). The accuracy and appropriate incorporation of these data are critical to produce reasonable estimates. This section reviews the sources of data used by the model and the extent to which the integration of these data into the model seems reasonable.

6. Accuracy of Data Inputs

Assumption: Visitor arrivals, expenditures, and resident transfer (i.e., government provided assistance) data are accurate and current *and* are representative of Hawaii.

Test: Review data sources and model implementation.

Implications: *If the accuracy assumption is broken,* externally derived visitor, and transfer data cannot be used to develop model scenarios of economic impact. *If the currency assumption is broken,* the most recent reported data do not reflect the current economic state and projected figures may be higher or lower than otherwise expected. *If the representativeness assumption is broken*, then the model represents a smaller fraction of or different portion of the intended population.

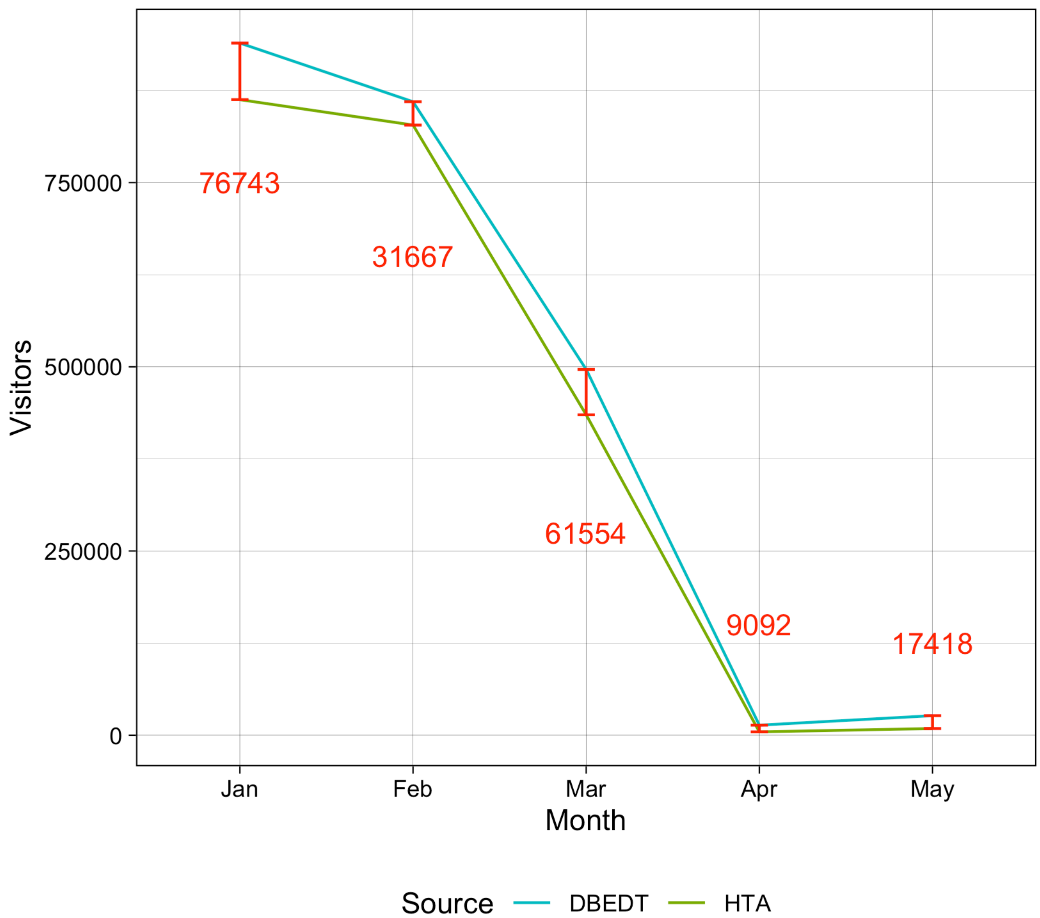
Considerations: While the accuracy of data is important at all stages in model development—including accuracy of ACS data and ALICE data—ACS and ALICE data 1) have a longer history of development and use and 2) were developed by highly reputable organizations. Specific to the synthetic population model, the accuracy of visitor and transfer data is of particular importance, as these data are fundamental to the simulation of employment and wage changes in the state for periods beyond collected sample data.

Results: There is a substantial difference in passenger arrivals (used by the model) and visitor arrivals (as reported by HTA). Visitor expenditure data based on DBEDT's input-output table and used in the model **is consistent** with data reported by HTA. **These data are incorporated into the model at the state level.** Similarly, expenditure data for households and businesses come from the input-output table at the state level. Unemployment insurance claims data is incorporated into the model at the state level, though data for lower level geographies are available. Characteristics of individuals applying for UI are available by sex, age, race, industry, and occupation, but are not included in the model as inputs.

6.1 Visitor Arrivals

Determine that number of visitors to the state during the COVID-19 time period match model input values.

Figure 30. Visitor monthly arrivals—comparison of DBEDT and HTA data[33]



- Daily passenger counts include: returning residents, intended residents, and visitors.
    - these counts exclude flights from Canada.
- Domestic flight passenger counts are estimated using Hawaii Department of Agriculture data.
- International flight passenger counts are estimated using Hawaii Department of Transportation data.
- Data come from the Air Traffic Summary Report, which airlines submit to the DOT each month. The report includes domestic and international in-transit passengers.

[33] Validation source: https://dbedt.hawaii.gov/visitor/daily-passenger-counts, Hawaii Tourism Authority

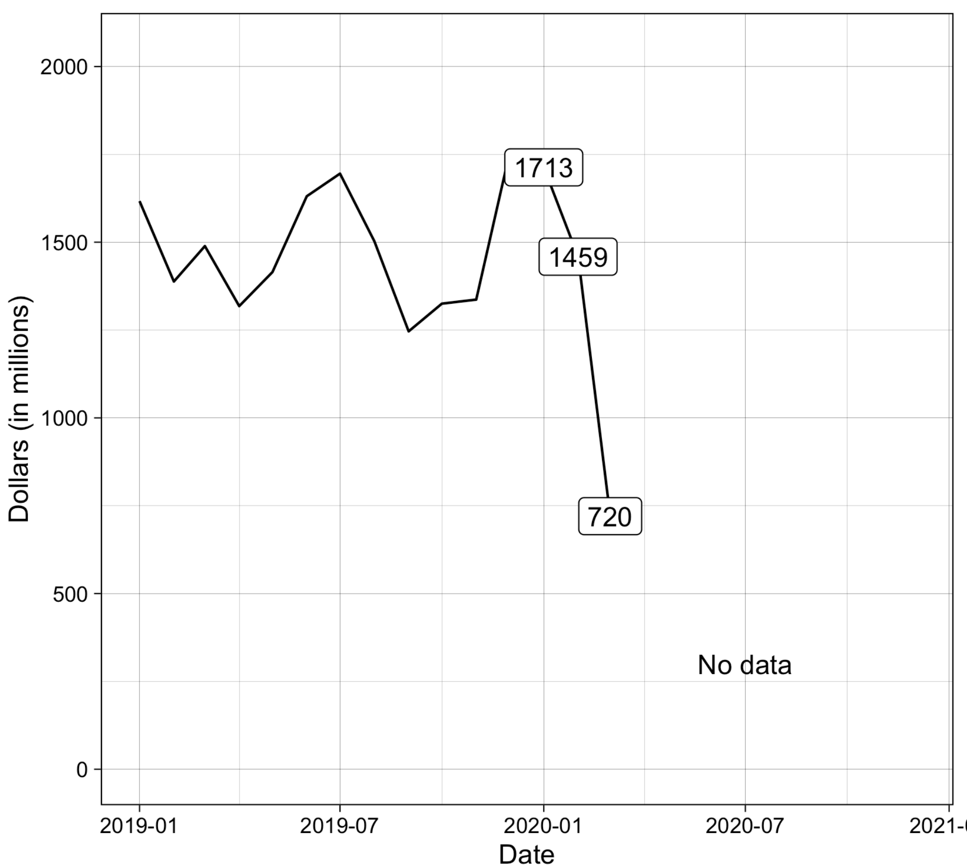
- Arrivals by air is derived from the Domestic In-flight Survey and the International Departure Survey.
  - Arrivals by cruise ship is derived from Department of Transportation Harbors Division and from the Cruise Visitor Survey.
- Visitors are calculated as:

  Airline passenger counts
  – (in-transit passengers + returning residents + intended residents)
  + (passengers on-board foreign-flagged cruise ships – Hawaii residents)

- There is a substantial difference between visitor arrivals (reported by HTA) and passenger arrivals (reported by DBEBT). Passenger arrivals include non-visitors, which became a greater proportion of passenger arrivals during the early COVID period (Figure 30).

6.2 Visitor Expenditures

Figure 31. Visitor monthly expenditures[34]



[34] Validation source: Hawaii Tourism Authority

- March 2020 was the last month visitor expenditure data were reported by the HTA.
- In the model, visitor expenditures are set by number of visitors and distributed across industries using The Hawaii State Input-Output Study: 2012 Benchmark report.
    - These data are specific to Hawaii; however, the most recently available data are for 2012.
    - According to the report, total visitor expenditures across all industries in 2012 was $16.1 billion. This amount—inflation adjusted—is $18.0 billion, which is nearly equivalent to the amount reported by the Hawaii Tourism Authority ($17.7 billion) for 2019.
- Visitor expenditures for small geographies are based off of the statewide number of visitors (i.e., not geography specific).

6.3 Input-Output Table

- The Hawaii Input-Output Study: 2012 Benchmark Report describes the economic inputs and outputs for, and their inter-relationships among, Hawaii's industries and consumers (including visitors and households) for 2012.
    - The study represents our best understanding of Hawaii's economy and is based off of numerous data sources and methodologies to derive the resultant input-output table.
        - As Honolulu's economy is dominant in the state, these tables are more reflective of Honolulu's economy than that of other islands.
    - Personal consumption expenditures (i.e., household expenditures) are based off of the U.S. Bureau of Labor Statistics Consumer Expenditure Survey (available to the state level) and receipts data from 2012 Economic Census (available to the county level for most industries and place and zip code level for some industries).
    - Visitor expenditure data are based DBEDT's visitor expenditure surveys[35] (available to the island level).
        - The model currently uses state level expenditure data.

6.4 Transfer Data

- Unemployment insurance, new and weeks claimed[36]
    - These data are reported by the Department of Labor and Industrial Relations. New and weekly claims are available on a weekly basis to sub-county geographies.
        - The model currently uses statewide claims data.
        - Data on the characteristics of claimants are available (but not incorporated into the model). These data are useful to better

[35] see https://dbedt.hawaii.gov/visitor/tourismdata/

[36] https://labor.hawaii.gov/rs/home/unemployment/unemployment-claims-data/

reflect the impacts of COVID-19 on sub-populations, such as by sex, age, race (including Hawaiian), industry, and occupation.[37]

- SNAP data[38] are reported by the Department of Human Services with tabulated frequency on a daily basis. The data are broken down by county (Hawaii county is broken down into East and West Hawaii). These are not incorporated into the model.
- Other sources of social benefits information/data in response to COVID-19 include:
    - CARES[39]
        - Period covered: April 1, 2020 – July 31, 2020
        - Amount:
            - $1,200 Individuals
            - $2,400 Joint filiers
            - $500 per qualifying child
        - Reductions for income above:
            - $150,000 for joint return
            - $112,500 for head of household
            - $75,000 taxpayer not described above
        - Model description:
            - "One-time 1,200 USD per adult, 700 USD per dependent. Tapers off for individuals with gross annual income above 75,000 USD."
    - Paycheck Protection Program (PPP)
        - PPP data are available for each business entity receiving funds[40]
        - The most recent announcement regarding PPP loans: "In order to reach the smallest businesses, SBA will offer PPP loans to businesses with fewer than 20 employees and sole proprietors only from Wednesday, February 24 through Tuesday, March 9, 2021 at 5pm ET. President Biden has also announced additional program changes to make access to PPP loans more equitable."[41]
        - Model description:
            - "Full reimbursement of eight weeks of payroll in return for not reducing headcount."
    - Families First Coronavirus Response Act
        - Expands WIC and TEFAP

---

[37] https://labor.hawaii.gov/rs/home/unemployment/characteristics-of-the-insured-unemployed/

[38] https://humanservices.hawaii.gov/communications-2/

[39] https://www.congress.gov/bill/116th-congress/house-bill/748/text

[40] https://www.sba.gov/funding-programs/loans/coronavirus-relief-options/paycheck-protection-program/ppp-data#section-header-2

[41] https://www.sba.gov/funding-programs/loans/coronavirus-relief-options/paycheck-protection-program

        - The model does not appear to incorporate these data
    - The following are incorporated into the model:
        - Island unemployment insurance (SUI) – up to $625 dollars per week for up to 26 weeks
        - Pandemic unemployment assistance (PUA) – Additional $600/week on top of SUI paid before July 31
        - Unemployment insurance relief – Additional 13 weeks of SUI paid after primary SUI period

Model Theory

This section reviews the foundational model theory that levels of employment and wages are mainly affected by the number of visitors to Hawaii during the COVID-19 pandemic. This theory is critical to projecting employment, income, and need of the population—based off of the ALICE framework—beyond the time period of the most recently available data.

This section also tests the assumption that the current model may be used as a proxy for directly observable population data by simulating changes in the population in response to key drivers.

7. Key Drivers of Employment and Wages

Assumption: Visitor volume is the primary driver and business revenue is the secondary driver of employment and wage cut levels for Hawaii residents during the COVID-19 time period.

Test: Examine face validity of key drivers assumption through review of employment levels by industry.

Implications: *If the assumption is broken,* visitor volume and business revenue are not the main drivers of employment and wages among Hawaii residents. The model requires a direct measure of employment and wages or an alternative driver of employment and wages to develop modeled scenarios.

Considerations: The primary driver in the modeled scenarios is visitor volume, which affects business revenue (secondary driver). These two key drivers determine the level of employment and wages among Hawaii's workforce in the model.

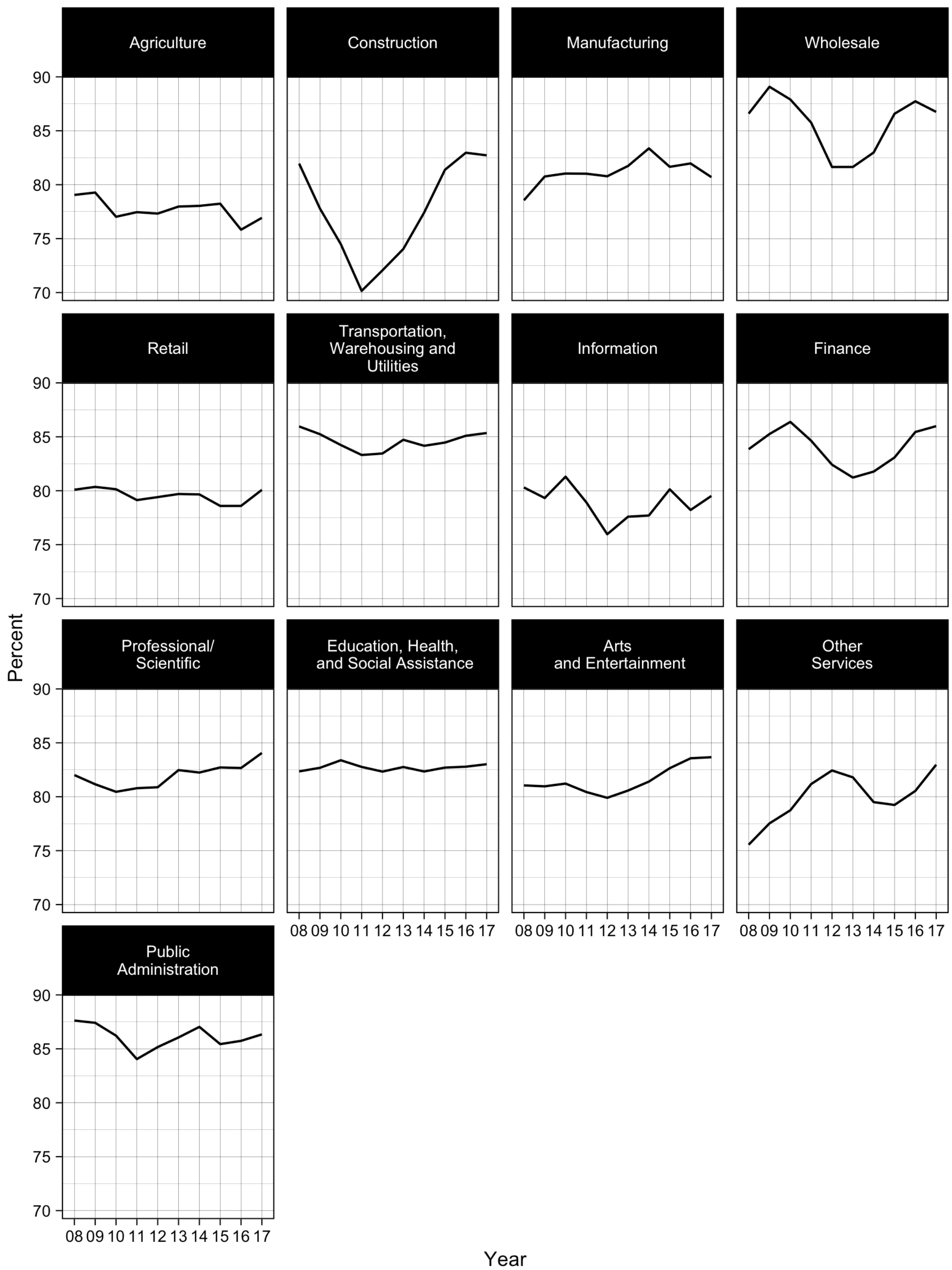
In the model, visitor flows are applied at the state level.

Results: Similar to the Great Recession of 2008 and 2009, COVID-19 has resulted in a decrease in employment among the population in Hawaii. However, different industries have been affected during the pandemic than those of the recession. During the COVID-19 pandemic**, the number of visitors to the islands has decreased, disproportionately affecting the Accommodation and Food Services industry in terms of level of employment and income**.

7.1 Comparison of industry impacts—Great Recession and COVID-19

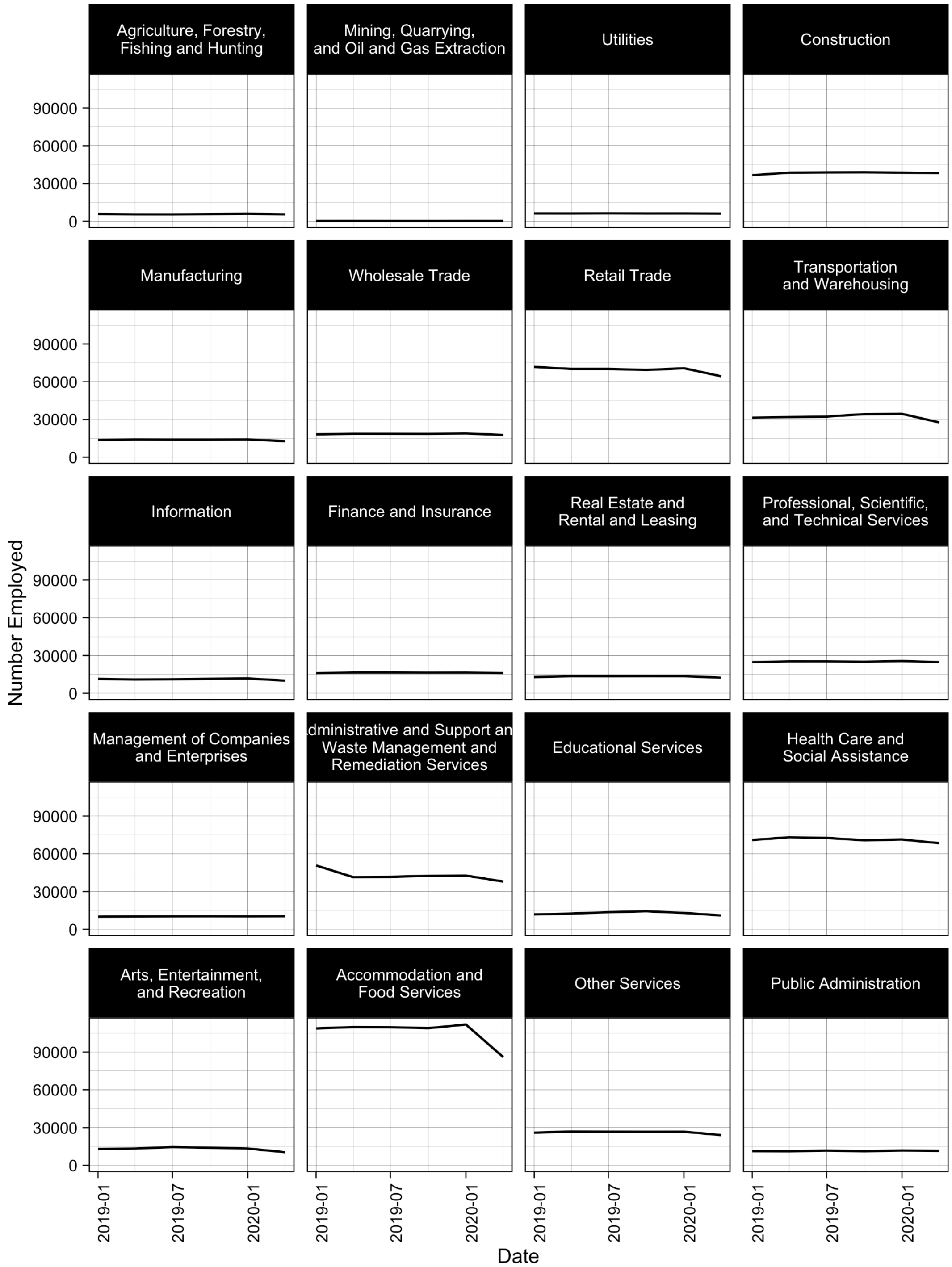
Compare changes in levels of employment due to the Great Recession and COVID-19.

Figure 32. Employment rates by industry for period covering the Great Recession[42]



42 Validation source: 2006 to 2017 1-year ACS PUMS (3-year averages)

Figure 33. Employment number by industry for period leading up to and including initial economic shock from COVID-19[43]



43 Validation source: Quarterly Workforce Indicators, file qwi_hi_rh_f_gc_ns_oslp_u

- The economic effects of the financial and housing market crisis that led to the Great Recession shows up in employment data for the construction industry and the wholesale and finance industries, subsequently. The percentage of individuals employed in the construction industry fell nearly 12 percentage points from 2008 to 2011 (Figure 32).
- From January 1, 2020 to April 1, 2020, the construction industry had 345 fewer employees.
- All industries saw a decrease in the number employed between January 1, 2020 and April 1, 2020 (except for Management of Companies and Enterprises).
- The Accommodation and Food Services industry had the greatest decrease in the number and proportion of individuals employed between January and April (roughly 26,000 fewer jobs—a 30 percent loss).
    - The Arts, Entertainment and Recreation; Transportation and Warehousing; Educational Services; and Information industries were among the top 5 for largest proportion of employment lost.
    - The Transportation and Warehousing; Retail Trade; Administrative and Support and Waste Management and Remediation Services; and Arts, Entertainment, and Recreation industries were among the top 5 for number of employees lost.
- In January 2020, there were 852,037 visitors to Hawaii. In April 2020, the number of visitors dropped 99.5 percent to 4,564 (Hawaii Tourism Authority).
    - Between February and March 2020, visitor arrivals decreased from 828,056 to 434,856 and expenditure went from $1,458,700,000 to $720,200,000.
    - Between the same period, Hawaii state tax base fell from $8,588,599,121 to $8,378,777,553 . Transient accommodations tax base went from $543,382,856 to $370,194,216.
    - By April 2020, the Hawaii tax base was down to $5,944,215,248 (down from $8,811,781,854 the previous year in April). Transient accommodations was down to $21,189,838 in April 2020 (compared to $438,470,265 the previous year in April).
- The model applies visitor flows at the state level and not sub-geography level.
    - Some communities in Hawaii are more reliant on tourism than others. This may show up in the model through the industry variable, but distinctions are not made within the simulation itself (i.e., using data inputs or differential calibrations).

8. Scenarios as Estimates

Assumption: Modeled scenarios of the economic impact of COVID-19 provide reasonable estimates of person and household characteristics for time periods beyond that of the sample data collection time period.

Test: Synthesize assumptions 1 through 7 to determine if modeled scenarios may serve as a proxy for sample design estimates.

Implications: *If the assumption is broken,* the model provides results that are disjointed from the observable economic impacts of COVID-19 in Hawaii. The model may not be used to provide any reasonable estimate of person or household characteristics for the state. The model's primary use is as a medium to conduct "thought experiments" for understanding downstream effects of changes in the model drivers given other valid assumptions.

Considerations: While modeled scenarios are treated and termed as "scenarios," incoming requests and analyses of model results may treat such results as "estimates" for the purposes of program implementation and/or decision-making. As such, it is important to consider this assumption, while acknowledging the obvious limitation that these data are not empirical, but provide a reasonable projection of what an estimate may be if more current estimates were available.

The synthetic population model is designed as an economic model. It does not take epidemiological considerations directly into account.

The informal economy is sparsely represented in the data.

The intended model can answer questions about the characteristics of people and households, but not relationships between those characteristics. For example, we might be able to conclude that unemployment in a region has increased 10 percentage points from 5 percent to 15 percent; however, we are not able to conclude that high levels of educational attainment in the region helped obviate higher unemployment

Results: The synthetic population is representative of people and households in Hawaii at the state level and larger geographies (with a few exceptions). Most notably, Kahului is consistently underrepresented. For many smaller geographies, the synthetic population is not representative of people and households. These misrepresentations can be observed for critical analysis and matching variables, such as household income, household type and size, and number of earners in households as well as for variables of potential interest, such as race. In instances where variables are not used to conduct distribution matching and are not highly correlated with matching variables, those variables have a tendency toward the overall distribution. This tendency is easily viewed when comparing the distributions of travel time to work. The assignment of workers to work places in the model is done with minimal consideration of geographic boundaries and transportation infrastructure. This results in a large number of workers commuting to work outside reasonable boundaries—for example, commuting from Puna (Hawaii Island) to Hanalei (Kauai). At a state level, and assuming equal economic impacts of COVID-19 on the state (i.e., no geographic differentiation), this approach does not affect analysis results. However, any analysis at a sub-state level or an assumption of differential economic impact requires 1) workers to be confined within their reasonable geographic boundaries and 2) modeling of the differential economic impacts of COVID-19 on various communities.

Region/community specific impacts are important to consider in the model in order to derive estimates of need (according to the ALICE framework) for those particular communities. Currently, the model uses inputs—such as unemployment insurance claims, SNAP recipiency, and visitor arrivals—at the state level. This assumes that these indicators can be spread out proportionately to the population and/or businesses. As some areas are more popular tourist destinations, these areas will realize greater negative effects from the decline in tourism in Hawaii. The model should include as much geographic and demographic detail for critical input variables to ensure that simulation results show greater impact for those most likely to be impacted (see Figure 28 for example).

The model relies on two key components to estimate level of need[44] among the population: costs and income. The ALICE framework is used to estimate the amount of income required to cover basic costs. This framework is consistent with other frameworks and research working to define the same concept. The current model appears to underestimate the basic costs that should be assigned to households. Several small adjustments to the cost formula should assign more adequate cost thresholds to the synthetic population. Further, income levels in the synthetic population appear to be too high. While modeled decreases in unemployment statewide are consistent with external sources of survey data, these decreases will likely underestimate the number of households in need as 1) assigned income is high and 2) assigned cost thresholds are low. In addition, because the baseline synthetic population underrepresents the number of 0- and 1-earner households, the underestimation of need will likely be exacerbated.

In its current state, the synthetic population model can provide reasonable estimates of unemployment at the state level and lower bound estimates of household economic need at the state level. In order to provide region specific estimates for both unemployment and economic need, the following is suggested:

1) Combine small geographies to create more reliable estimates in target data.
    a) A reasonable aggregation may be: East Hawaii, West Hawaii, Maui, Molokai, Lanai, Honolulu, Central Oahu, Waianae, North Shore (Waialua and Koolauloa), Koolaupoko.

[44] Defined as having income less than what is necessary to cover basic standard of living costs.

2) Use the "PUMA" variable from the ACS as part of the matching step to confine records to their most probable geographies.
3) Include only relevant analysis and contextual variables in the synthetic population dataset to discourage improper use of variables not included as part of the model design.
4) Enforce geographic restrictions to the assignment of workers to work places.
5) Use as much geographic and demographic detail as possible for model inputs (e.g., unemployment claims by sex, age, industry and visitors by island)
6) Use visitor arrivals instead passenger arrivals as a model input.
7) Review and implement ALICE framework as defined in ALICE methodology, paying particular attention to the taxes formula for the state of Hawaii.
8) Review total income assigned to people and households to ensure totals are reasonable.

Implementation of the aforementioned suggestions are intended to improve the model to answer the following questions: 1) who is impacted economically by COVID-19 (e.g., race/ethnicity, industry, educational attainment, household type); 2) with what are individuals and/or households struggling (within the ALICE framework); 3) and what is the extent to which various regions have been impacted?

In consideration of limited resources and/or time to implement the suggestion above, a simplified approach may be to:

1) Utilize the ACS PUMS data directly, instead of synthesizing the population. Geographic detail is still available at the PUMA level.
2) Enforce geographic restrictions for worker to work place assignments.
3) Update implementation of ALICE framework.
4) Use geographic and demographic detail for data inputs.

Finally[45], as employment, income, and expenses are dynamic variables contingent on external factors and on each other, these variables should be dynamically generated in the model as events change/occur. For example, an employed individual who becomes unemployed in the model scenario should have ALICE costs recomputed such that taxes, transportation, childcare, and healthcare costs reflect the current (modeled) need. The duration of employment may also be an important factor in this calculation (especially for healthcare).

[45] Other recommendations based off of informal discussion (not direct review and analysis) consist of the following: 1) calculate and use ALICE costs on a monthly basis; instead of projecting income for the year, 2) use observed earnings and other income from the sample data; instead of applying a 0.9/.1 split of earnings and income for all households.