Technical Documentation for Household Demographics Projection

REMI Household Forecast is a tool to complement the PI+ demographic model by providing comprehensive forecasts of a variety of household characteristics. Our forecast methodology extracts the county-level historical data from the ACS and decennial census to examine how demographic figures by sex, age and racial/ethnic group drive the household forecast in each category, and then uses the estimated parameters and the REMI projected demographic characteristics data to forecast the number of households in the different categories. The forecast methodology applied here is basically a regression-based forecast model. The data sets used in the regression approach are from decennial census or ACS estimates, depending on the availability of the data. Ideally a panel data regression approach is preferred to cross-sectional regression in modeling the time trend of household characteristics. However, panel data are not always available for all the regression models we ran. The details are given in the following documents.

The total numbers of households by county are first forecasted based on the regression approach. Household categorization at single dimension such as household size, household income brackets, number of workers, number of vehicles available, and race of householder are projected use the similar regression approach. In addition, in order to provide more detailed information about household characteristics, cross tabulation approach is applied to obtain two-dimension household forecast. The initial frequency distribution in cross tabulation is generated from Public Use Microdata Sample (PUMS) data. By assuming the multivariate frequency distribution is stable over the forecast period, combined with the forecast results from regression approach, REMI model can extend the household characteristics forecast to more detailed level.

I. Household Forecast at Single Dimension

1. Forecast the total number of households

1) Dataset construction:
   a. Data sources:
      - 2000 census
      - 2007-2009 3-year ACS estimates
      - 2010 census

   b. Variables: county level data
      Dependent variable: the ratio of households to total population (`hh_pop`)
      Explanatory variables: the share of non-hispanic white to total population (`Swhite_NH`), the share of non-hispanic black(`Sblack_NH`), the share of Hispanic(`Shispanic`), the share of age cohort 15_24(`ag15_24`), the share of age

cohort 25_64(ag25_64), the share of age cohort 65+(ag65p), the ratio of labor force to total population (SLF), and time (year)

2) Regression diagnostics and model selection

Adding the sex ratio variable (W_SEX) has increased the R square significantly.
There is negative trend when adding year as the explanatory variable.
Adding Interaction terms between race and age cohort do not improve the R^2 significantly. However, they make the coefficients for age and race difficult to interpret.
Checking the non-linearity issue by scatterplot of residuals and acprplot does not reveal suspicious problems.
The model we decide to use and the result is:

$$HH_{POP} = f(Swhite_{NH}\ Sblack_{NH}, Shispanic, ag_{15-24}, ag_{25-64}, ag_{65p,} SLF, W_{SEX,}\ year)$$

Source	SS	df	MS	Number of obs =	7873
				F(9, 7863) =	2452.58
Model	6.64660642	9	.738511824	Prob > F =	0.0000
Residual	2.36767473	7863	.000301116	R-squared =	0.7373
				Adj R-squared =	0.7370
Total	9.01428115	7872	.001145107	Root MSE =	.01735

hh_pop	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
Swhite_NH	.004466	.0025367	1.76	0.078	-.0005066	.0094387
Sblack_NH	-.0155384	.002787	-5.58	0.000	-.0210016	-.0100752
Shispanic	-.0213757	.0028769	-7.43	0.000	-.0270152	-.0157362
ag15_24	.3202235	.0096023	33.35	0.000	.3014004	.3390465
ag25_64	.5169481	.0100896	51.24	0.000	.4971698	.5367265
ag65p	.7347046	.007948	92.44	0.000	.7191245	.7502847
SLF	.1013187	.0042317	23.94	0.000	.0930234	.109614
W_SEX	.6019883	.0113619	52.98	0.000	.5797159	.6242608
year	-.000194	.0000453	-4.28	0.000	-.0002829	-.0001052
_cons	.0034615	.0895528	0.04	0.969	-.1720857	.1790087

2 Forecast the household number categorized by income brackets

1) We define households are categorized by the following income brackets:
    - Category 1: 0-$24,999
    - Category 2: $25,000- $49,999
    - Category 3: $50,000- $74,999
    - Category 4: $75,000- $99,999
    - Category 5: $100,000 and more

2) Dataset construction:
   Data sources:
    - 2000 census
    - 2006-2010 5-year ACS estimates

   Variables: county level data.
   Dependent variable: the ratio of households in different household income brackets to total household number (`inc0`, `inc1`, `inc2`, `inc3`, and `inc4`).
   Explanatory variables: the share of non-hispanic white to total population (`Swhite_NH`), the share of non-hispanic black (`Sblack_NH`), the share of Hispanic (`Shispanic`), the share of age cohort 15-24 (`ag15_24`), the share of age cohort 25-64 (`ag25_64`), the ratio of labor force to total population (`SLF`), year, per capita income (`PCIN`)

3) Regression diagnostics and model selection

Adding Interaction terms between race and age cohort, between year and race, between per capita income and race do not improve the R^2 significantly.
The influence from "year" causes the overarching increasing pattern in the regression related to `hh_inc0`, so we leave out "`year`" in the regression.
Issue: it seems that the shares are too closely related to the per capita income and have very obvious pattern of decrease and increase.

The model we decide to use and the result is:

$$HH_{INC}^{i} = f\left(Swhite_{NH}\ Sblack_{NH}, Shispanic, ag_{15-24}, ag_{25-64}, ag_{65p,}\ PCIN, SLF, W_{SEX}\ \right)$$

Variable	inc0	inc1	inc2	inc3	inc4
W_SEX	.49412038	-.20429395	-.23680467	-.12164681	.06981907
Swhite_NH	-.05578525	.03139796	.05215438	.01338681	-.04281987
Sblack_NH	.03076465	-.01269515	.0119284	.00131651	-.03319446
Shispanic	.03396231	-.01940497	.00203382	-.00785045	-.01053167
ag15_24	.42284714	-.04896727	-.22554536	-.18816839	.04145277
ag25_64	.17179167	.11651452	-.08164753	-.14707935	-.05457363
ag65p	.65403297	.29627756	-.25154147	-.31769693	-.377905
PCIN	-10.482351	-6.4936538	.29457212	3.7649063	12.895983
SLF	-.54773756	.26359788	.33774737	.14901591	-.2040631
_cons	.35529807	.29285375	.197594	.13616971	.01762171

3 Forecast the household number categorized by number of workers in household

1) We define households are categorized by number of workers in the following way
   - Category 1: No workers
   - Category 2: 1 worker
   - Category 3: 2 workers
   - Category 4: 3-or-more workers

2) Dataset construction:
   c. Data sources:
   - 2000 census
   - 2006-2010 5-year ACS estimates

Variables: county level data.

Dependent variable: the ratio of households in each category to total household number (`hh_w0, hh_w1, hh_w2, hh_w3`).

Explanatory variables: the share of non-hispanic white to total population (`Swhite_NH`), the share of non-hispanic black (`Sblack_NH`), the share of Hispanic (`Shispanic`), the share of age cohort 15_24 (`ag15_24`), the share of age cohort 25_64 (`ag25_64`), the share of age cohort 65+(`ag65p`), the ratio of labor force to total population (`SLF`), year, per capita income (`PCIN`)

3) Regression diagnostics and model selection

The high correlation between `ag25-64` and `SLF` introduced multicollinearity problem in the model. We decide to drop the variable for labor force.

Decide not to include any interaction terms due to the small increase in R square.

Checking non-linearity by residual plots and acprplot did not reveal severe problems.

The model we decide to use and the result is as follows.

$$HH_{LF}^{i} = f\left(Swhite_{NH}\ Sblack_{NH}, Shispanic, ag_{15-24}, ag_{25-64}, ag_{65p,} PCIN, W_{SEX}\right)$$

Variable	hh_w0	hh_w1	hh_w2	hh_w3
W_SEX	0.211868	-0.02571	-0.20832	0.022153
Swhite_NH	-0.00769	-0.06719	0.068474	0.006405
Sblack_NH	0.024644	0.010381	-0.02062	-0.0144
Shispanic	-0.08882	0.007419	0.062707	0.018693
ag15_24	-0.36446	0.083058	0.2141	0.067305
ag25_64	-0.8846	0.052122	0.711707	0.120769
ag65p	0.5387	-0.23939	-0.17675	-0.12256
PCIN	-1.48E-06	1.04E-06	5.61E-08	3.76E-07
_cons	0.549294	0.411632	0.050135	-0.01106

4. Forecast the household number categorized by household sizes

1) We define households are categorized by household size in the following way:
   - Category 1: 1-person household
   - Category 2: 2-person household
   - Category 3: 3-person household
   - Category 4: 4-or-more-person household

2) Dataset construction:
   a. Data sources:
      - 2000 census
      - 2006-2010 5-year ACS estimates

   Problem: there is an issue of overlapping between the 2007-2009 3-year ACS estimates and 2006-2010 5-year ACS estimates. So I only keep the 5-year ACS estimates in 2010 instead of using both 3-year and 5-year estimates.

   b. Variables: county level data.

   Dependent variable: the ratio of households in each household size category to total household number (`hh_sz1, hh_sz2, hh_sz3, hh_sz4`)

   Explanatory variables: the share of non-hispanic white to total population (`Swhite_NP`), the share of non-hispanic black(`Sblack_NP`), the share of Hispanic(`Shispanic`), the share of age cohort 15_24(`ag15_24`), the share of age cohort 25_64(`ag25_64`), the share of age cohort 65+(`ag65p`), the ratio of labor force to total population (`SLF`), and time (`year`).

3) Regression diagnostics and model selection

Year is not a significant predictor, so we will use cross-sectional model.
After checking the interaction terms and non-linearity problems, we decide to stick to the simple model as follows.

$$HH_{SZ}^{i} = f\left(Swhite_{NH}\ Sblack_{NH}, Shispanic, ag_{15-24}, ag_{25-64}, ag_{65p,} SLF, W_{SEX}\ \right)$$

Variable	hh_sz1	hh_sz2	hh_sz3	hh_sz4
W_SEX	.0728906	-.13922043	.31598616	-.24965633
Swhite_NH	-.09716579	.07750908	.05002047	-.03036377
Sblack_NH	-.01752391	.00225599	.05556941	-.04030149
Shispanic	-.07332522	.02197951	.03322654	.01811917
ag15_24	.74758747	.33709012	.00893984	-1.0936174
ag25_64	.61478032	.3993068	.13458539	-1.1486725
ag65p	1.0866503	.7240057	-.40197433	-1.4086816
SLF	.14260691	.04747334	-.13861565	-.05146459
_cons	-.3396393	-.02159852	.00802821	1.3532096

5 forecast the household number categorized by vehicles available

1) We defined households are classified into 4 categories by number of vehicles available in the following way.
    - Category 1: No vehicles
    - Category 2: 1 vehicle
    - Category 3: 2 vehicles
    - Category 4: 3-or-more vehicles

2) Dataset construction:
   Data sources:
    - 2000 census (households by number of workers in household is not available)
    - 2006-2010 5-year ACS estimates

Variables: county level data.

Dependent variable: the ratio of households in each category to total household number (`hh_veh0, hh_veh1, hh_veh2, hh_veh3p`)

Potential explanatory variables: the share of non-hispanic white to total population (`Swhite_NP`), the share of non-hispanic black (`Sblack_NP`), the share of Hispanic (`Shispanic`), the share of age cohort 15_24 (`ag15_24`), the share of age cohort 25_64 (`ag25_64`), the share of age cohort 65+ (`ag65p`), the ratio of labor force to total population (`SLF`), year, per capita income (`PCIN`)

3) Regression diagnostics and model selection

Year is significant in most of the 4 equations. However, according to our following forecast, the sigh for the variable year has the overarching effect of making the forecast increase or decrease. We decide to drop it in the model.
Per capita income is not included due to the weak explanatory power.
Checking non-linearity problem does not reveal suspicious problems.

$$HH_{VEH}^{i} = f\left(Swhite_{NH}\ Sblack_{NH}, Shispanic, ag_{15-24}, ag_{25-64}, ag_{65p,} SLF, W_{SEX}\ \right)$$

Variable	hh_veh0	hh_veh1	hh_veh2	hh_veh3p
W_SEX	.20462738	.25407976	-.00042115	-.45828599
Swhite_NH	-.2086197	-.11015956	.18717426	.131605
Sblack_NH	-.09551997	.01046228	.06422597	.02083172
Shispanic	-.18680394	-.01475192	.16016451	.04139135
ag15_24	.1729468	.42700744	-.13593375	-.46402049
ag25_64	.10441309	.12319285	.0507937	-.27839963
ag65p	.05037656	.32692938	-.31099345	-.06631249
SLF	-.19573625	-.18435922	.05468462	.32541085
_cons	.17171857	.19422531	.23819865	.39585746

6 Forecast the household number by race of householder

1) We define the classification of "race of householder" in the following way to keep it consistent with the current REMI model.
    - Category 1: non-Hispanic White householder
    - Category 2: non-Hispanic black householder
    - Category 3: Hispanic householder
    - Category 4: non-Hispanic other

2) Dataset construction

Data sources:

- 2000 census
- 2010 census

Variables: county level data.

- Dependent variable: the ratio of households in each category of race of householder to total household number (SH_white_NH, SH_black_NH, SH_hispanic, SH_other_NH)
- Potential explanatory variables: the share of non-hispanic white to total population (Swhite_NP), the share of non-hispanic black (Sblack_NP), the share of Hispanic (Shispanic), the share of age cohort 15-24 (ag15_24), the share of age cohort 25-64 (ag25_64), the share of age cohort 65+ (ag65p), the ratio of labor force to total population (SLF), year, per capita income (PCIN)

3) Methodology selection

According to the pairwise correlation, we can find the ratio of households by race and Hispanic origin of householder is highly correlated with the ratio of population in each category. Thus, we use the simple linear regression to do the forecast. The dependent variable is the ratio of households with the householder belonging to one specific race and Hispanic origin category (Category 1- 4). The independent variable is the population ratio of one specific race and Hispanic origin category.

$$HH^{i}_{RACE} = f(Swhite_{NH}\ Sblack_{NH}, Shispanic)$$

	SH_whi~H	SH_bla~H	SH_his~c	SH_oth~H	W_SEX	Swhite~H	Sblack~H	Shispa~c	ag15_24	ag25_64
SH_white_NH	1.0000									
SH_black_NH	-0.6734	1.0000								
SH_hispanic	-0.5304	-0.0989	1.0000							
SH_other_NH	-0.3709	-0.0891	0.0098	1.0000						
W_SEX	-0.0287	0.2008	-0.1123	-0.1350	1.0000					
Swhite_NH	0.9924	-0.6437	-0.5458	-0.3847	0.0341	1.0000				
Sblack_NH	-0.6665	0.9959	-0.1014	-0.0940	0.1505	-0.6431	1.0000			
Shispanic	-0.5237	-0.1038	0.9900	0.0168	-0.1420	-0.5516	-0.1059	1.0000		
ag15_24	-0.2126	0.1552	0.0990	0.0767	-0.0439	-0.2156	0.1563	0.1000	1.0000	
ag25_64	0.1324	-0.0162	-0.1546	-0.0617	-0.2854	0.1248	-0.0016	-0.1481	-0.4658	1.0000
ag65p	0.3348	-0.1914	-0.1479	-0.2261	0.2262	0.3429	-0.1907	-0.1562	-0.4459	-0.3320
PCIN	0.2021	-0.1432	-0.1155	-0.0508	0.0482	0.1899	-0.1527	-0.0901	-0.1601	0.4965
SLF	0.3893	-0.2830	-0.2186	-0.0907	0.1147	0.3986	-0.3080	-0.2089	0.0821	0.2366

	ag65p	PCIN	SLF
ag65p	1.0000		
PCIN	-0.2022	1.0000	
SLF	-0.2149	0.6660	1.0000

The regression results from the simple linear model are as follows:

	NH-white householder	NH-black householder	hispanic householder
Swhite_NH	0.902022		
Sblack_NH		0.933681	
Shispanic			0.854001
constant	0.114742	-0.00276	-0.00872

II. Two-dimension Household Characteristics Forecast

Methodology selection:

Cross Tabulation: This method is based on the forecasting results from the regression methods above. The household forecast method provides household number forecasts at the county level by household characteristics, such as, the number of households by household size, the number of households by household income, the number of households by vehicles available, the number of households by number of workers in household, and the number of households by race of householder. In order to forecast the household characteristics in more detail, or at more dimensions, we use the PUMS data to create contingency tables or the existing cross tabulation from ACS to do the forecast for every cell in the contingency tables.

If the tabulation is available from ACS 2010, for example, household size by vehicles available, we can use the tabulation directly from ACS. However, if the tabulation is not available from ACS, for example, household income by number of workers in household, we need to go to the PUMS data for tabulation.

The baseline contingency tables are generated using 2006-2010 PUMS data at the state level. We ignored selecting the baseline contingency table for each county because PUMS data cannot provide the complete data for every county.

Data set constructions

- 2006-2010 5-year ACS estimates
- 2006-2010 PUMS data

1. Household size by vehicles available

Household numbers	1-person household:	2-person household:	3-person household:	4-or-more-person	TOTAL
1 vehicle available					
2 vehicles available					
3vehicles available					
4 or more vehicles available					

The contingency tables for household size by vehicles available can be obtained from 2007, 2008, 2009 ACS estimates. The baseline data we choose is the 2010 5-year estimates as shown in the following table.

Household numbers	1-person household:	2-person household:	3-person household:	4-or-more-person	TOTAL
No vehicle available	87762 **14.09%**	31509 **3.85%**	13652 **4.08%**	18630 **3.39%**	151553 **6.51%**

1 vehicle available	443966 **71.25%**	244037 **29.82%**	77537 **23.15%**	100316 **18.24%**	865856 **37.22%**
2 vehicles available	73266 **11.76%**	435479 **53.21%**	144445 **43.13%**	246847 **44.88%**	900037 **38.69%**
3vehicles available	13203 **2.12%**	84092 **10.28%**	78265 **23.37%**	118874 **21.61%**	294434 **12.66%**
4 or more vehicles available	4873 **0.78%**	23271 **2.84%**	21035 **6.28%**	65409 **11.89%**	114588 **4.93%**
total	623070 **100.00%**	818388 **100.00%**	334934 **100.00%**	550076 **100.00%**	2326468 **100.00%**

The forecast of household size by vehicles available can be calculated based on the frequency distribution above, as well as the forecast for total number of households and households by household size from the regression approach.

2. Household size by number of workers in household

Household numbers	1-person household:	2-person household:	3-person household:	4-or-more-person	TOTAL
No workers					
1 worker					
2 workers					
3 or more workers					

The contingency tables between household size by number of workers in household are available from 2007, 2008, 2009 ACS estimates. The baseline data we choose are from the 2010 5-year ACS estimates as follows.

Household numbers	1-person household:	2-person household:	3-person household:	4-or-more-person	TOTAL
No workers	286968 **46.06%**	280150 **34.23%**	36065 **10.77%**	39923 **7.26%**	643106 **27.64%**
1 worker	336102 **53.94%**	264920 **32.37%**	125899 **37.59%**	202283 **36.77%**	929204 **39.94%**
2 workers	0 **0.00%**	273318 **33.40%**	132459 **39.55%**	216247 **39.31%**	622024 **26.74%**
3 or more workers	0 **0.00%**	0 **0.00%**	40511 **12.10%**	91623 **16.66%**	132134 **5.68%**
total	623070	818388	334934	550076	2326468

The forecast of household size by number of workers in household is then calculated based on the frequency distribution above, as well as the forecast for total number of households and households by household size from the regression approach.

3. Household size by household income

Household numbers	1-person household:	2-person household:	3-person household:	4-or-more-person	TOTAL
<$25k					
[$25k, $50k)					
[$50k, $75k)					
[$75k, $100)					
[$100k, ∞)					

There is no data available from ACS to generate the contingency table between household size and household income. We use the 2010 5-year PUMS data to build the tabulation between the cross tabulation as follows.

```
. tab HHIN HHSZ [fweight=WGTP], column
```

format	1-person	2-person	3-person	4-or-more	Total
min-$24,999	280,817	137,897	55,123	84,178	558,015
	45.09	**16.84**	**16.43**	**15.33**	**23.99**
$25,000-$49,999	195,374	225,756	77,608	130,698	629,436
	31.37	**27.56**	**23.13**	**23.8**	**27.06**
$50,000-$74,999	83,089	181,001	72,445	114,913	451,448
	13.34	**22.1**	**21.6**	**20.92**	**19.4**
$75,000-$99,999	31,425	107,267	51,166	85,001	274,859
	5.05	**13.1**	**15.25**	**15.48**	**11.81**
100,000 and more	32,084	167,088	79,121	134,413	412,706
	5.15	**20.4**	**23.59**	**24.47**	**17.74**
Total	622,789	819,009	335,463	549,203	2,326,464
	100	100	100	100	100

The household income data from 2010 5-year ACS is as follows. This is fairly consistent with the tabulation from PUMS data, which confirm the consistency between PUMS and ACS data.

Household numbers	TOTAL	percentage
<$25k	534,150	0.229597
[$25k, $50k)	618,690	0.265935
[$50k, $75k)	446,445	0.191898
[$75k, $100)	286,189	0.123014

[$100k, ∞)	440,994	0.189555
total	2,326,468	

The forecast of household size by household income is then calculated based on the frequency distribution above, as well as the forecast for total number of households and households by household size from the regression approach.

4. Number of workers in household by vehicles available

Household numbers	No workers	1 workers	2 workers	3 or more workers	TOTAL
0 vehicle available					
1 vehicle available					
2 vehicles available					
3vehicles available					
4 or more vehicles available					

The contingency tables for number of workers in household by vehicles available can be obtained from 2007, 2008, 2009 ACS estimates. The baseline data we choose is the 2010 5-year ACS estimates as follows.

Household numbers	No workers	1 workers	2 workers	3 or more workers	TOTAL
0 vehicle	89605	48950	10790	2208	151553
	13.93%	**5.27%**	**1.73%**	**1.67%**	**6.51%**
1 vehicle available	340560	442473	75236	7587	865856
	52.96%	**47.62%**	**12.10%**	**5.74%**	**37.22%**
2 vehicles available	171321	329422	370906	28388	900037
	26.64%	**35.45%**	**59.63%**	**21.48%**	**38.69%**
3 or more vehicles available	41620	108359	165092	93951	409022
	6.47%	**11.66%**	**26.54%**	**71.10%**	**17.58%**
total	643106	929204	622024	132134	2326468
	100%	100%	100%	100%	100%

The forecast of number of workers in household by vehicles available is then calculated based on the frequency distribution above, as well as the forecast for total number of households and households by number of workers in household from the regression approach.

5 number of workers in household by household income

Household numbers	No workers	1 worker	2 workers	3 or more workers	TOTAL
<$25k					

[$25k, $50k)					
[$50k, $75k)					
[$75k, $100)					
[$100k, ∞)					

There is no data available from ACS to generate the contingency table between number of workers in household and household income. We use the 2010 5-year PUMS data to build the cross tabulation as follows.

```
. tab  HHIN HHLF [fweight=WGTP], column
```

Household numbers	0 workers	1 worker	2 workers	3 or more	Total
min-$24,999	194,481	266,611	84,315	12,608	558,015
	51.25%	**31.76%**	**9.72%**	**5.26%**	**23.99%**
$25,000-$49,999	105,260	287,070	197,432	39,674	629,436
	27.74%	**34.2%**	**22.75%**	**16.55%**	**27.06%**
$50,000-$74,999	43,737	143,617	211,349	52,745	451,448
	11.52%	**17.11%**	**24.35%**	**22%**	**19.4%**
$75,000-$99,999	15,825	63,550	146,589	48,895	274,859
	4.17%	**7.57%**	**16.89%**	**20.39%**	**11.81%**
100,000 and more	20,194	78,559	228,134	85,819	412,706
	5.32%	**9.36%**	**26.29%**	**35.8%**	**17.74%**
Total	379,497	839,407	867,819	239,741	2,326,464
	100%	100%	100%	100%	100%

The forecast of number of workers in household by household income is based on the frequency distribution above, as well as the forecast for total number of households and households by number of workers in household from the regression approach.

6. Household income by vehicles available

Household numbers	<$25k	[$25k, $50k)	[$50k, $75k)	[$75k, $100)	[$100k, ∞)	TOTAL
0 vehicle available						
1 vehicle available						
2 vehicles available						
3 or more vehicles available						

This tabulation of household income by vehicles available is not available from ACS. We use the 2010 5-year PUMS data to build the contingency table.

```
. tab HHVE HHIN [fweight=WGTP], column
```

Household numbers	min-$24,9	$25,000-$	$50,000-$	$75,000-$	100000 above	Total
0 vehicle	106,190	28,887	9,053	3,145	4,253	151,528
	19.03%	**5%**	**2.01%**	**1.14%**	**1.03%**	**6.51%**
1 vehicle	314,407	309,358	137,969	52,906	52,059	866,699
	56.34%	**49%**	**30.56%**	**19.25%**	**12.61%**	**37.25%**
2 vehicles	109,318	222,394	216,006	142,869	209,557	900,144
	19.59%	**35%**	**47.85%**	**51.98%**	**50.78%**	**38.69%**
3 vehicles or more	28,100	68,797	88,420	75,939	146,837	408,093
	5.04%	**11%**	**19.59%**	**27.63%**	**35.58%**	**17.54%**
Total	558,015	629,436	451,448	274,859	412,706	2,326,464
	100	100	100	100	100	100

Households by vehicles available from 2010 5-Year ACS are as follows. This is very close to the tabulation result from 2010 5-year PUMS, which confirms the consistency between PUMS data and ACS data.

Household numbers	households	percentage
No vehicle available	151553	0.06514295
1 vehicle available	865856	0.37217619
2 vehicles available	900037	0.38686842
3 vehicles available	409022	0.17581243

The forecast of household income by vehicles available is based on the frequency distribution above, as well as the forecast for total number of households and households by household income from the regression approach.

7. Household income by race of householder

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
<$25k					
[$25k, $50k)					
[$50k, $75k)					
[$75k, $100)					
[$100k, ∞)					

This tabulation of household income by race of household is not available from ACS. We use the 2010 5-year PUMS data to build the contingency table.

```
. tab  HHIN HHLF [fweight=WGTP], column
```

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
min-$24,999	47,262	333,364	25,431	151,958	558,015
	30.77%	**20.75%**	**30.95%**	**31.37%**	**23.99%**
$25,000-$49,999	39,465	409,998	23,155	156,818	629,436
	25.7%	**25.52%**	**28.18%**	**32.38%**	**27.06%**
$50,000-$74,999	26,768	320,512	15,481	88,687	451,448
	17.43%	**19.95%**	**18.84%**	**18.31%**	**19.4%**
$75,000-$99,999	15,922	207,232	7,963	43,742	274,859
	10.37%	**12.9%**	**9.69%**	**9.03%**	**11.81%**
100,000 and more	24,158	335,236	10,139	43,173	412,706
	15.73%	**20.87%**	**12.34%**	**8.91%**	**17.74%**
Total	153,575	1,606,342	82,169	484,378	2,326,464
	100%	100%	100%	100%	100%

The forecast of household income by race of householder is based on the frequency distribution above, as well as the forecast for total number of households and households by race of householder.

8. Household size by race of householder

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
1-person household:					
2-person household:					
3-person household:					
4-or-more-person					

The cross tabulation between household size and race of householder is not available from ACS. We use the 2010 5-year PUMS data to build the contingency table.

```
. tab HHSZ HH_RACE [fweight=WGTP], column
```

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total

1-person household	34,957	481,472	26,136	80,224	622,789
	22.76%	**29.97%**	**31.81%**	**16.56%**	**26.77%**
2-person household	41,178	648,703	22,644	106,484	819,009
	26.81%	**40.38%**	**27.56%**	**21.98%**	**35.2%**
3-person household	26,283	207,582	13,255	88,343	335,463
	17.11%	**12.92%**	**16.13%**	**18.24%**	**14.42%**
4-or-more-person household	51,157	268,585	20,134	209,327	549,203
	33.31%	**16.72%**	**24.5%**	**43.22%**	**23.61%**
Total	153,575	1,606,342	82,169	484,378	2,326,464
	100%	100%	100%	100%	100%

The forecast of household size by race of householder is based on t the frequency distribution above, as well as he forecast total number of households and households by race of householder from the regression approach.

9. Number of workers in household by race of householder

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
No workers					
1 worker					
2 workers					
3 or more workers					

The cross tabulation between number of workers and race of householder is not available from ACS. We use the 2010 5-year PUMS data to build the contingency table.

```
. tab HHLF HH_RACE [fweight=WGTP], column
```

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
0 workers	19,287	311,709	10,463	38,038	379,497
	12.56%	**19.4%**	**12.73%**	**7.85%**	**16.31%**
1 worker	55,891	568,342	38,033	177,141	839,407
	36.39%	**35.38%**	**46.29%**	**36.57%**	**36.08%**
2 workers	58,071	593,187	26,851	189,710	867,819
	37.81%	**36.93%**	**32.68%**	**39.17%**	**37.3%**
3 or more workers	20,326	133,104	6,822	79,489	239,741
	13.24%	**8.29%**	**8.3%**	**16.41%**	**10.3%**
Total	153,575	1,606,342	82,169	484,378	2,326,464
	100%	100%	100%	100%	100%

The forecast of number of workers in household by race of householder is then calculated based on the frequency distribution above, as well as the forecast for total number of households and households by race of householder from the regression approach.

10. Vehicles available by race of householder

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
0 vehicle available					
1 vehicle available					
2 vehicles available					
3 or more vehicles available					

The cross tabulation between vehicles available and race of householder is not available from ACS. We use the 2010 5-year PUMS data to build the contingency table.

```
. tab HHVE HH_RACE [fweight=WGTP], column
```

Household numbers	other non-Hispanic	White non-Hispanic	Black non-Hispanic	Hispanic	Total
0 vehicle	18,618	81,088	12,153	39,669	151,528
	12.12%	**5.05%**	**14.79%**	**8.19%**	**6.51%**
1 vehicle	60,451	600,400	36,990	168,858	866,699
	39.36%	**37.38%**	**45.02%**	**34.86%**	**37.25%**
2 vehicles	51,370	644,183	23,446	181,145	900,144
	33.45%	**40.1%**	**28.53%**	**37.4%**	**38.69%**
3 vehicles or more	23,136	280,671	9,580	94,706	408,093
	15.06%	**17.47%**	**11.66%**	**19.55%**	**17.54%**
Total	153,575	1,606,342	82,169	484,378	2,326,464
	100%	100%	100%	100%	100%

The forecast of vehicles available in household by race of householder is then calculated based on the frequency distribution above, as well as the forecast for total number of households and households by race of householder from the regression approach.

Appendix: Dependent variables and Explanatory variables in the regression approach

Dependent Variable	Description
HH_{POP}	the ratio of total households to total population
HH^{i}_{INC}	the share of households group i by household income, $i = 0\ to\ 4$
HH^{i}_{LF}	the share of household group i by number of workers, $i = 0\ to\ 3$
HH^{i}_{SZ}	the share of household group i by household size, $i = 1\ to\ 4$,
HH^{i}_{VEH}	the share of household group i by household vehicle availability, $i = 1\ to\ 5$,
HH^{i}_{RACE}	the share of households group i by the race of householder, $i = 1\ to\ 5$
Explanatory Variables	
$Swhite_{NH}$	the share of Non-Hispanic Whites to total population
$Sblack_{NH}$	the share of Non-Hispanic Blacks to total population
$Shispanic$	the share of Hispanics to total population
ag_{15-24}	the share of 15-24 age cohort to total population
ag_{25-64}	the share of 25-64 age cohort to total population
$ag_{65p,}$	the share of ≥65 years old to total population
SLF	the ratio of labor force to total population
W_{SEX}	the ratio of female population
year	year
$PCIN$	per capita personal income