INTRODUCTION

In this memo, details regarding the imputation analysis are presented. The rest of the appendix is organized as follows:

- In the next section, background for imputation analysis and methodologies that were explored in this study are presented.
- In the third section, the model development process is described.
- In the fourth section, model estimation results are presented.
- In the fifth section, the approach to imputing the missing income values is described.
- In the last section, the improvements made to the final imputation analysis is presented.

BACKGROUND

In surveys, valid values for some data items are missing because a respondent may not have provided a response – this is also referred to as item nonresponse or missing data. Data is often missing for sensitive items (e.g. income, race). Even in survey instruments that force respondents to provide responses, a valid skip is offered for such sensitive data items so as to avoid respondent drop-out and potentially risk not collecting responses for all other data items for which the respondent may be willing to provide an answer. Missing data are problematic and they impact subsequent survey data analyses.

In the CSTS, respondents were generally forced to provide a valid response for all but three data items: exact age, income, and ownership status. For these three data items, respondents were offered “prefer not to respond” as an option (not a valid value that can be used for subsequent analyses). For the age question, respondents were required to answer a follow up question with coarser categories, therefore valid age information is available (not missing) for all respondents. For income, respondents were offered (but not required to answer) a follow up question where they were asked to respond to the same question on a coarser scale in the hope that those hesitant to providing response on a detailed level may respond when the options were consolidated. The presence of follow up did improve valid responses, however, there were still missing values. For the ownership status, no follow up option was available. Thus, there are observations with missing home ownership status information. The focus of the imputation analysis presented in the appendix is on the income variable. The imputation analysis was limited only to the income variable because this was the only variable that was used for correcting sample bias as part of the weighting analysis.

In order to impute the income, two different imputation techniques were explored: stochastic regression technique and Markov Chain Monte Carlo multiple imputation technique[1]. A brief overview of the two methodologies is provided below (Stopher 2008, Enders 2010):

- In a regression imputation technique, values are imputed using predictions from a regression model. The regression model is specified using other observed variables in the dataset as explanatory variables. A common limitation of the regression technique is that it reduces variability in the data. An extension of this approach that avoids this issue is the stochastic regression imputation wherein an error term is added to the predicted value before assigning it as the imputed value. Thus, the stochastic regression imputation preserves variability in the data. While this approach overcomes the variability issue, it still suffers from the issue of overfitting the data.

[1] There are a number of other approaches to treating missing data including deletion methods, single point imputation techniques, and expectation maximization among others. However, the chosen methods are superior if not comparable in performance to these other methods. (Zimowski et al. 1997, Stopher 2008, Enders 2010)

- The intuition behind multiple imputation technique is to draw multiple missing values from an underlying distribution. This approach accurately reflects the uncertainty in predicting the true value instead of inferring a single value for the missing entries. Multiple realizations of the imputed dataset are created by using a different draw from the underlying distribution for each missing observation. The realizations are then treated as complete datasets and analyzed to come up with multiple estimates of the parameters of interest. For any given parameter, these multiple estimates can then be used to draw inferences about attributes of interest. There are different implementations of the multiple imputation framework based on the type of missing data and approach to imputation (e.g. parametric regression method, propensity score method, and Markov Chain Monte Carlo method to name a few) (Yang 2016).

In the CSTS, both methodologies were explored for imputing the income variable. In particular, for the stochastic regression approach, a multinomial logistic (MNL) model form was assumed to impute income. The MNL model formulation is appropriate for imputing income because income responses were collected on a discrete scale in the CSTS (as opposed to continuous scale). For the multiple imputation approach, Markov Chain Monte Carlo (MCMC) approach was applied. Unlike the MNL model, income was treated as a continuous variable in this approach. Comprehensive validation analysis was conducted to evaluate the two methods for imputing income values. Missing income observations were created artificially from valid records in the CSTS. The two methods were then applied and predictions for the artificially missing observations were compared against the actual values. Disaggregate and aggregate comparisons were performed to evaluate the two methods. It was found that the MNL model based stochastic regression imputation provided comparable results to the MCMC based multiple imputation. Additionally, MNL model did not require making any strong assumptions or manipulating the income variable to conform to the underlying model formulation. As a result, in the CSTS, MNL based stochastic regression model was applied to impute missing income information. In the next section, the model development process is described.

MODEL DEVELOPMENT

MNL based stochastic regression model was used to impute detailed income information for the households – detailed income was collected on a 12 category scale. The model was estimated using those sample observations where households reported valid values for the detailed income. Out of 8,403 non-volunteer households that completed the travel diary, 1,443 households did not report the detailed income category. The MNL model was developed using the remaining 6,960 household observations.

The MNL model included two types of variables, namely, those related directly to the household with missing information and those related to other households in the vicinity of the household with missing information. Four types of variables were used in the model. Details regarding the variables are described below along with reasoning for their inclusion.

Employment and Education of Household Members

The first set of variables considered in the analysis include the counts of household members based on combination of employment status, the usual hours worked per week, and the level of educational attainment. All these variables are associated with income of a household. High number of workers in a household is associated with high level of household income. Similarly, low count of workers in a household will be associated with low household income. Additionally, the worker status (i.e. full time versus part time) as well as level of educational attainment is generally related to the salaries earned by employed individuals. Consequently, higher counts of such individuals will influence overall household income. Specific definitions for the employment status, hours worked per week, and level of education variables are listed below:

- Any member of the household who reported either being employed or self-employed in the CSTS was considered a worker; all other household members were considered as a non-worker.
- Level of education attained was grouped into three categories:
  - low education - completed high school but have not achieved a college degree
  - college education – completed bachelor or associate degrees, or have completed vocational or technical training
  - high education - completed post graduate education
- Each worker in the household was categorized into two categories based on employment status:
  - part-time - working fewer than 30 hours per week
  - full-time - working more than 30 hours per week

Only those household members who were workers were considered and categorized into one of the six types:

- Full-time low education worker
- Full-time college education worker
- Full-time high education worker
- Part-time low education worker
- Part-time college education worker
- Part-time high education worker

Each of these worker types reflect different combinations of employment status, hours worked per week and educational attainment.

Lifecycle Indicators

The second set of variables used for income imputation are related to the structure of the households. These are also referred to as lifecycle variables. There are a variety of ways in which household structure can be defined. In the CSTS, the structure is defined by combining counts of adult household members by retirement status and counts of children in the household. It is hypothesized that stage of the lifecycle of a household is closely related to the income. For example, single adult households make less income compared to households comprising of multiple adults. Additionally, older households are associated with higher income compared to younger households. Retired households would be associated with less income compared to those households who have not retired. Specific definitions of adult, retirement status, and children used for constructing the lifecycle variables are as noted below:

- Adults are defined as any household member who is 18 years or older. Based on counts of adults, households are grouped into two categories:
  - One adult households and
  - 2 or more adult households
- Household members who are retired are identified based on the employment variable. Based on retiree status of members of household, households are classified as:
  - Those households with retirees and
  - Those households without retirees.
- Children are defined into three categories based on the age range as noted below:
  - Children between 0 and 4 years old,
  - Children between 5 and 15 years old, and

- Children between 16 and 17 years old.

Further, based on presence of children of various age categories, households are grouped into four categories:

- No children,
- Youngest child between 0 and 4 years,
- Youngest child between 5 and 15 years, and
- Youngest child between 16 and 17 years.

Combining the category of the respondent households along the above three dimensions, the following 10 lifecycle indicators were defined:

- Household with one adult and no children
- Household with 2 or more adults and no children
- Household with one adult and youngest child between 0 and 4 years
- Household with 2 or more adults and youngest child between 0 and 4 years
- Household with one adult and youngest child between 5 and 15 years
- Household with 2 or more adults and youngest child between 5 and 15 years
- Household with one adult and youngest child between 16 and 17 years
- Household with 2 or more adults and youngest child between 16 and 17 years
- Household with one adult who is retired with no children
- Household with 2 or more adults who are retired with no children

Household Vehicle Ownership

The next set of variables used for income imputation are the count of the household vehicles. It is assumed that possession of high number of household vehicles is generally associated with high household income. On the other hand, lower vehicle ownership is generally associated with lower household income. The vehicle count variable was collected in the CSTS on an integer scale and included in the income model without any transformation.

Income Distribution of Surrounding Households

In addition to the above mentioned variables related to the respondent household, information regarding the income distribution of households in the vicinity of the respondent household were also used. It is posited that the income distribution of the households in the vicinity will be closely related to the income of the respondent household. Information about income of households in the vicinity of the responding household was collected from the 2009-2013 American Community survey (ACS) Summary Files. In particular, the income distribution of households in the same blockgroup as the respondent household, measured as proportion of households belonging to various income categories, was used. In the following section, the model estimation results are presented.

MODEL ESTIMATION RESULTS

A MNL model was estimated using the above explanatory variables. A total of 12 income categories were considered as noted below:

- Less than $10,000
- $10,000 - $14,999
- $15,000 - $24,999
- $25,000 - $34,999
- $35,000 - $49,999
- $50,000- $59,999
- $60,000 - $74,999
- $75,000 - $99,999
- $100,000 - $149,999
- $150,000 - $199,999
- $200,000 - $249,999
- $250,000 or more

The model estimation summary is presented in Table J.1. Note that all parameters in the model were treated to be alternative specific. Income category 6 (i.e. income between $50,000- $59,999) was set as the base alternative. The final model was estimated on 6959 households out of 6960 households who reported detailed income information; one record was excluded from the model estimation because the income distribution of surrounding households could not be identified. The final model consists of 155 parameters; 11 parameters are constants associated with various income categories (other than baseline alternative). The model estimation results are plausible and consistent with our expectations of the influence of the different variables in explaining household income. Log-likelihood ratio test confirmed that the model is statistically significant. Final model estimation results are presented in Table J.2.

Employment and Education of Household Members

The parameters associated with the composite variables combining employment and education of household members show that as the count of employed individuals increase, the household tends to belong to a higher income category and vice versa. This can be seen from the negative coefficients for income categories 1 through 5 and positive coefficients for income categories 7 through 12. One exception was the variable that indicates the number of workers who are part time with low education – as the count of the workers in the category increases, the household tends to be associated with lower income. This is consistent with expectation that individuals with higher education are generally associated with higher levels of income and vice versa for those with lower education attainment. Also, the opposite influence can be seen for high income category for this variable. Overall, the impact of counts of full-time workers (across all levels of educational) was found to be more significant than the impact of the counts of part-time workers.

Lifecycle Indicators

Households with one adult were found to be associated with lower income categories (as indicated by positive values of the parameters in low income category alternatives) and lower probability to belong to the high income categories (as indicated by the negative values of the parameters in high income category alternatives). However, this trend was limited to households with no children (i.e. households with one adult and no children). For other households with one adult with children of various ages, they mostly showed higher probability to belong to lower income categories but there were also instances where they were associated with some high income categories. The retired households with 2 or more adults show higher propensity to belong to high income categories and show lower propensity to belong to the low income categories.

Household Vehicle Ownership

The vehicle count variable was found to be very significant and contributed negatively towards the household's propensity to belong to lower income categories. On the other hand, vehicle count contributed positively to the household's propensity to belong to higher income categories.

Income Distribution of Surrounding Households

The coefficient estimates associated with the influence of income distribution of surrounding households also aligned with the prior hypothesis about the potential association of these variables with household income category. The proportion of surrounding households of lower income categories was found to be significantly positively associated with low income categories for respondents. Also, proportion of surrounding households of higher income categories was found to be significantly negatively associated with low income categories. It was interesting to note that the impact of the surrounding households' income distribution on the income category of the respondent household was more pronounced for the lowest and highest income distributions. For the intermediate income distributions, the associated parameters were found to be significantly positive only in and around the respective income categories. In the next section, the approach to applying the model to impute the income values is presented.

Table J.1: Model Estimation Summary

Statistics	Value
Number of Observations	6959
Log-likelihood value at convergence, final model	-13242.559
Log-likelihood value at convergence, constant only model	-16513.707
Number of Parameters, final model	155
Number of Constants	11
Likelihood ratio	6542.296
Degrees of freedom (DOF)	144
Critical Chi-square value (DOF = 144, level of significance = 0.001)	202.184

APPLICATION OF THE MODEL TO IMPUTE MISSING INCOME

The MNL model was then applied to impute the income values for the 1,443 households with missing income information. The steps applied to impute the income values for each household are noted below:

1. If the household answered the follow up income question wherein they provided income information on a coarse scale, the applicable income categories for imputation were limited to only those that fall within the broad reported income. If they did not provide a valid answer to the follow up income question, then all income categories are applicable for imputation.
2. The probability that the missing value belongs to each applicable category of income is estimated using the MNL model estimated by utilizing the attributes of the household.

3. A Monte Carlo procedure is applied to predict a single income category value based on the estimated probabilities[2].

Table J.3 shows the comparison of the prediction using the MNL model and the actual values. This table shows the ability of the model to replicate the results for observations that were used to estimate the model. As can be seen, the model does well in predicting the income category. For most observations, the predicted value matches the actual value. Even when the predictions do not match the actual value, the predictions are distributed around the actual value. It is also interesting to note that the predictions are limited to a narrow band around the actual value. This is by design owing to the first step described above wherein applicable income categories are limited based on the response to the coarse income category. In the replication analysis, it was assumed that response to coarse income category question was available for all observations. This in turn aids the prediction performance. However, for the observations with missing income data, only a small percentage of the observations have the coarse income response available which could affect the performance.

Table J.4 shows the comparison of income values between the original income and imputed income variables. The missing values in the original dataset are denoted by “Prefer not to answer” category. As can be seen, all the valid income values in the original variable were retained in the imputed income variable whereas the missing income values (i.e. “Prefer not to answer”) are assigned a valid income value in the imputed variable. The highest counts of imputed income fall in the income category “$100,000 - $149,999”. This is also the category with the highest number of observations in the original income variable. The least counts of imputed income are for the lowest income category i.e. “less than $10,000”.

[2] The Monte Carlo proceeds by first calculating cumulative probability distribution that provides the range of probability for each income value. After this step, a random number is drawn between 0 and 1. Based on the probability range to which the random draw belongs, the income category value is assigned.

Table J.2: Model Estimation Results

Exploratory Variable Names	Income Category Alternatives										
	Less than $10,000	$10,000 - $14,999	$15,000 - $24,999	$25,000 - $34,999	$35,000 - $49,999	$60,000 - $74,999	$75,000 - $99,999	$100,000 - $149,999	$150,000 - $199,999	$200,000 - $249,999	$250,000 or more
Constant	2.036 (4.57)	1.231 (3.99)	0.156 (0.65)	-0.363 (-1.26)	0.969 (7.39)	-0.326 (-3.45)	-1.538 (-6.81)	-2.144 (-8.99)	-3.947 (-12.96)	-4.779 (-12.42)	-4.954 (-17.75)
Employment and Education of Household Members											
# of Full time low educated worker	-2.451 (-9.18)	-1.444 (-6.82)	-0.851 (-6.47)			0.323 (3.61)	0.36 (4.16)	0.459 (5.11)	0.491 (3.99)	0.54 (3.13)	
# of Full time college educated worker	-3.56 (-9.69)	-2.894 (-8.84)	-2.112 (-10.35)	-0.682 (-4.85)	-0.214 (-1.85)	0.446 (4.02)	0.78 (7.49)	1.152 (10.67)	1.455 (11.47)	1.466 (9.16)	1.141 (7.9)
# of Full time post graduate worker	-3.893 (-6.47)	-4.578 (-4.52)	-2.226 (-6.88)	-1.238 (-4.88)		1.066 (7.64)	1.639 (13.02)	2.301 (17.83)	2.927 (19.81)	3.2 (17.93)	3.076 (19.77)
# of Part time low educated worker		0.325 (1.57)	0.554 (3.66)	0.537 (3.44)						-0.819 (-2.11)	-0.833 (-2.45)
# of Part time college educated worker	-0.62 (-1.91)	-0.445 (-1.38)								0.51 (2.28)	
# of Part time post graduate worker	-2.507 (-3.41)	-1.278 (-2.68)	-0.871 (-2.48)				0.617 (3.04)	0.807 (4.05)	1.108 (4.63)	1.263 (4.26)	1.113 (4.28)
Lifecycle Indicators											
HH with 1 adult, no children	0.365 (1.95)	0.288 (1.44)					-0.26 (-2.54)	-0.573 (-5)	-1.166 (-5.78)	-1.501 (-4.27)	-0.539 (-2.32)
HH with 1 adult, youngest child age 0-4	2.157 (3.31)	1.571 (2.22)	0.978 (1.41)	1.269 (2.12)	1.307 (2.39)						
HH with 2+ adults, youngest child age 0-4	0.672 (1.42)	0.633 (1.41)			0.226 (1.08)						
HH with 1 adult, youngest child age 5-15	1.001 (2.17)	1.125 (2.49)	1.327 (3.76)	1.016 (2.86)	0.717 (2.02)		0.827 (2.61)	0.556 (1.68)			
HH with 2+ adult, youngest child age 5-15			0.472 (1.86)		0.295 (1.66)						0.372 (2.41)
HH with 2+ adult, youngest child age 16-17			1.19 (2.63)								
HH with 1 retired adult, no children	-1.341 (-6.34)	-0.44 (-2.01)						-0.597 (-2.86)	-0.731 (-1.86)	-0.467 (-0.84)	-1.158 (-2.11)
HH with 2+ retired adults, no children		-0.916 (-2.15)	-0.992 (-3.15)	0.225 (0.97)	0.504 (2.52)	1.057 (5.37)	0.868 (4.38)	1.477 (7.41)	1.273 (4.68)	1.383 (3.71)	0.943 (2.9)

Exploratory Variable Names	Income Category Alternatives										
	Less than $10,000	$10,000 - $14,999	$15,000 - $24,999	$25,000 - $34,999	$35,000 - $49,999	$60,000 - $74,999	$75,000 - $99,999	$100,000 - $149,999	$150,000 - $199,999	$200,000 - $249,999	$250,000 or more
Household Vehicle Ownership											
HH Vehicle Count	-1.749 (-12.78)	-1.446 (-10.97)	-0.747 (-8.1)	-0.63 (-7.7)	-0.334 (-5)		0.099 (1.75)	0.272 (5.02)	0.42 (6.33)	0.502 (5.91)	0.664 (8.69)
Income Distribution of Surrounding Households											
% of HH in the BG with income < $10,000	2.862 (3.32)	2.382 (3.1)	2.471 (3.77)	1.13 (1.65)						-2.101 (-1.33)	
% of HH in the BG with income $10,000- $14,999	2.174 (1.84)	3.29 (3.01)	3.068 (3.18)	3.692 (3.86)			1.998 (2)	0.821 (0.81)	1.892 (1.33)		
% of HH in the BG with income $15,000-$24,999	2.723 (2.6)	3.314 (3.52)	3.503 (4.65)	2.531 (3.31)							
% of HH in the BG with income $25,000-$34,999			1.367 (1.73)	1.819 (2.27)							
% of HH in the BG with income $35,000-$49,999	1.699 (1.53)		2.297 (2.97)	2.299 (2.94)							
% of HH in the BG with income $60,000-$74,999			1.574 (1.68)	1.888 (2.01)			2.469 (3.57)	2.264 (3.38)	1.399 (1.44)		
% of HH in the BG with income $75,000-$99,999							2.469 (4.5)	1.431 (2.72)			
% of HH in the BG with income $100,000-$149,999	-3.523 (-2.72)				-2.239 (-4.26)		0.931 (1.86)	1.475 (3.09)	2.346 (3.61)	1.088 (1.19)	
% of HH in the BG with income $150,000-$199,999	-4.012 (-1.97)	-3.367 (-1.99)					1.644 (2.25)	3.191 (4.61)	3.112 (3.38)	3.264 (2.72)	2.59 (2.38)
% of HH in the BG with income $200,000-$249,999				1.891 (2.64)			2.866 (6.67)	3.559 (8.76)	5.431 (11.58)	6.193 (11.89)	8.82 (20.47)

Table J.3: Comparison of the Actual and Imputed Income Values for the Estimation Sample

		Imputed Household Income												Total
		Less than $10,000	$10,000-$14,999	$15,000-$24,999	$25,000-$34,999	$35,000-$49,999	$50,000-$59,999	$60,000-$74,999	$75,000-$99,999	$100,000-$149,999	$150,000-$199,999	$200,000-$249,999	$250,000 or more	
Original Household Income	Less than $10,000	127	75	108	0	0	0	0	0	0	0	0	0	**310**
	$10,000-$14,999	80	93	118	0	0	0	0	0	0	0	0	0	**291**
	$15,000-$24,999	113	122	243	0	0	0	0	0	0	0	0	0	**478**
	$25,000-$34,999	0	0	0	236	251	0	0	0	0	0	0	0	**487**
	$35,000-$49,999	0	0	0	254	401	0	0	0	0	0	0	0	**655**
	$50,000-$59,999	0	0	0	0	0	247	313	0	0	0	0	0	**560**
	$60,000-$74,999	0	0	0	0	0	311	372	0	0	0	0	0	**683**
	$75,000-$99,999	0	0	0	0	0	0	0	986	0	0	0	0	**986**
	$100,000-$149,999	0	0	0	0	0	0	0	0	764	280	114	136	**1294**
	$150,000-$199,999	0	0	0	0	0	0	0	0	294	119	55	88	**556**
	$200,000-$249,999	0	0	0	0	0	0	0	0	101	74	27	62	**264**
	$250,000 or more	0	0	0	0	0	0	0	0	138	79	58	120	**395**
Total		320	290	469	490	652	558	685	986	1297	552	254	406	6959

Table J.4: Comparison Between Values for the Original and Imputed Income Variables

		Imputed Household Income												Total
		Less than $10,000	$10,000-$14,999	$15,000-$24,999	$25,000-$34,999	$35,000-$49,999	$50,000-$59,999	$60,000-$74,999	$75,000-$99,999	$100,000-$149,999	$150,000-$199,999	$200,000-$249,999	$250,000 or more	
Original Household Income	Less than $10,000	310	0	0	0	0	0	0	0	0	0	0	0	**310**
	$10,000-$14,999	0	291	0	0	0	0	0	0	0	0	0	0	**291**
	$15,000-$24,999	0	0	478	0	0	0	0	0	0	0	0	0	**478**
	$25,000-$34,999	0	0	0	487	0	0	0	0	0	0	0	0	**487**
	$35,000-$49,999	0	0	0	0	656	0	0	0	0	0	0	0	**656**
	$50,000-$59,999	0	0	0	0	0	560	0	0	0	0	0	0	**560**
	$60,000-$74,999	0	0	0	0	0	0	683	0	0	0	0	0	**683**
	$75,000-$99,999	0	0	0	0	0	0	0	986	0	0	0	0	**986**
	$100,000-$149,999	0	0	0	0	0	0	0	0	1294	0	0	0	**1294**
	$150,000-$199,999	0	0	0	0	0	0	0	0	0	556	0	0	**556**
	$200,000-$249,999	0	0	0	0	0	0	0	0	0	0	264	0	**264**
	$250,000 or more	0	0	0	0	0	0	0	0	0	0		395	**395**
	Prefer not to answer	51	56	87	83	134	116	163	174	253	145	61	120	**1443**
Total		**361**	**347**	**565**	**570**	**790**	**676**	**846**	**1160**	**1547**	**701**	**325**	**515**	**8403**

IMPROVEMENTS TO IMPUTATION IN FINAL ANALYSIS

The study team proposed an update to the imputation analysis documented above to address the issue of stochasticity during imputation. As noted above, a MNL based stochastic regression technique was used to impute detailed income information for the households with missing income information. To impute the income value, a Monte Carlo procedure was applied. Questions abound on the best approach to prediction using the Monte Carlo procedure. In order to address the above issue, an iterative stochastic regression approach was applied (Gold et al. 2000). In this approach, a multinomial logit (MNL) model is estimated and applied iteratively to impute the income information until convergence is achieved. The process begins by estimating the MNL model first using the observations with valid income information. These first set of MNL model coefficient estimates are then used to stochastically impute the missing income values. In the second step, the observations with imputed income information are combined with observations with valid income information. The MNL model is re-estimated using this combined dataset (including observed and imputed income values). The income values are imputed again using the new coefficient estimates. This second step is repeated until convergence is achieved. In this project, stability in the parameter estimates (i.e. no change in parameter estimates across iterations) was used as the convergence criterion.

The coefficient estimates and the imputed income values are very similar to the previous earlier model estimation results. Therefore, the presentation of the revised model estimation is limited to the results. The coefficient estimates are presented in Table J.5. The imputed income values are predicted in the same way as before using the Monte Carlo procedure. The comparison of the original and imputed income variables using the revised model estimation are presented in Section 3.7.4.

Table J.5: Revised Model Estimation Results using Iterative Stochastic Regression Approach

Exploratory Variable Names	Income Category Alternatives										
	Less than $10,000	$10,000 - $14,999	$15,000 - $24,999	$25,000 - $34,999	$35,000 - $49,999	$60,000 - $74,999	$75,000 - $99,999	$100,000 - $149,999	$150,000 - $199,999	$200,000 - $249,999	$250,000 or more
Constant	2.134 (5.34)	1.163 (4.15)	0.227 (1.04)	-0.282 (-1.07)	1.054 (8.74)	-0.288 (-3.38)	-1.444 (-6.89)	-2.045 (-9.36)	-3.713 (-13.55)	-4.704 (-13.63)	-4.934 (-19.67)
Employment and Education of Household Members											
# of Full time low educated worker	-2.363 (-9.73)	-1.459 (-7.25)	-0.822 (-6.6)			0.337 (4.06)	0.371 (4.55)	0.478 (5.79)	0.456 (4.05)	0.451 (2.8)	
# of Full time college educated worker	-3.768 (-10.4)	-3.109 (-9.68)	-2.102 (-11.23)	-0.728 (-5.63)	-0.302 (-2.83)	0.332 (3.28)	0.713 (7.54)	1.058 (10.93)	1.332 (11.84)	1.258 (8.86)	1.101 (8.72)
# of Full time post graduate worker	-4.145 (-6.92)	-4.801 (-4.74)	-2.27 (-7.36)	-1.199 (-5.18)		0.983 (7.59)	1.585 (13.61)	2.243 (19.01)	2.807 (21.01)	3.172 (19.99)	3.013 (21.53)
# of Part time low educated worker		0.305 (1.57)	0.539 (3.77)	0.592 (4.14)						-0.696 (-2.12)	-0.799 (-2.66)
# of Part time college educated worker	-0.792 (-2.57)	-0.549 (-1.83)								0.503 (2.54)	
# of Part time post graduate worker	-2.7 (-3.69)	-1.32 (-3.01)	-0.932 (-2.81)				0.505 (2.77)	0.681 (3.87)	0.878 (4.08)	1.061 (4.02)	1.035 (4.6)
Lifecycle Indicators											
HH with 1 adult, no children	0.362 (2.15)	0.396 (2.15)					-0.149 (-1.57)	-0.54 (-5.11)	-1.068 (-5.99)	-1.619 (-4.89)	-0.498 (-2.42)
HH with 1 adult, youngest child age 0-4	1.931 (3.18)	1.302 (1.9)	0.71 (1.05)	1.187 (2.16)	1.062 (2.05)						
HH with 2+ adults, youngest child age 0-4	0.499 (1.12)	0.633 (1.48)			0.117 (0.57)						
HH with 1 adult, youngest child age 5-15	1.062 (2.45)	1.312 (3.1)	1.384 (4.11)	0.967 (2.8)	0.63 (1.82)		0.829 (2.73)	0.636 (2.09)			
HH with 2+ adult, youngest child age 5-15			0.572 (2.41)		0.22 (1.28)						0.346 (2.57)
HH with 2+ adult, youngest child age 16-17			1.28 (3.29)								
HH with 1 retired adult, no children	-1.485 (-7.74)	-0.397 (-2.01)						-0.622 (-3.36)	-1.021 (-2.8)	-0.666 (-1.35)	-1.416 (-2.88)

Exploratory Variable Names	Income Category Alternatives										
	Less than $10,000	$10,000 - $14,999	$15,000 - $24,999	$25,000 - $34,999	$35,000 - $49,999	$60,000 - $74,999	$75,000 - $99,999	$100,000 - $149,999	$150,000 - $199,999	$200,000 - $249,999	$250,000 or more
HH with 2+ retired adults, no children		-0.754 (-2.11)	-1.077 (-3.71)	0.139 (0.66)	0.566 (3.27)	1.156 (6.82)	0.906 (5.21)	1.466 (8.42)	1.129 (4.81)	1.044 (3.15)	0.674 (2.38)
Household Vehicle Ownership											
HH Vehicle Count	-1.768 (-14.29)	-1.478 (-12.34)	-0.789 (-9.22)	-0.624 (-8.32)	-0.336 (-5.49)		0.112 (2.13)	0.288 (5.81)	0.421 (6.95)	0.522 (6.75)	0.636 (9.13)
Income Distribution of Surrounding Households											
% of HH in the BG with income < $10,000	2.917 (3.7)	2.336 (3.3)	2.258 (3.71)	1.462 (2.34)						-1.793 (-1.24)	
% of HH in the BG with income $10,000- $14,999	1.825 (1.66)	3.393 (3.36)	3.372 (3.8)	3.189 (3.56)			1.611 (1.72)	0.657 (0.7)	1.412 (1.07)		
% of HH in the BG with income $15,000-$24,999	2.708 (2.8)	3.637 (4.18)	3.501 (4.96)	2.388 (3.32)							
% of HH in the BG with income $25,000-$34,999			1.209 (1.65)	1.502 (2.01)							
% of HH in the BG with income $35,000-$49,999	2.114 (2.09)		2.489 (3.48)	2.4 (3.31)							
% of HH in the BG with income $60,000-$74,999			1.172 (1.35)	1.679 (1.94)			2.05 (3.21)	1.831 (3)	0.981 (1.11)		
% of HH in the BG with income $75,000-$99,999							1.904 (3.72)	1.206 (2.51)			
% of HH in the BG with income $100,000-$149,999	-4.078 (-3.52)				-2.551 (-5.3)		0.846 (1.84)	1.547 (3.58)	2.107 (3.57)	1.381 (1.67)	
% of HH in the BG with income $150,000-$199,999	-2.91 (-1.62)	-1.83 (-1.27)					1.825 (2.76)	3.12 (5.03)	3.08 (3.71)	2.838 (2.61)	3.825 (4.01)
% of HH in the BG with income $200,000-$249,999				1.563 (2.45)			2.677 (7.11)	3.43 (9.75)	5.411 (13.32)	6.261 (13.79)	8.806 (23.74)