<u>Documentation for VMT Paper</u>

This file documents the data preparation and code estimation settings for the "Quantifying the relative contribution of factors to household vehicle miles of travel" paper. In this paper we present a holistic analysis to identify the relative contribution of socio-economic and demographic characteristics, built environment attributes, residential self-selection effects, and social and spatial dependency effects in explaining household VMT production. The modeling framework employs a simultaneous equations model of residential location (density) choice and household VMT generation. The analysis is performed using household travel survey data from the New York metropolitan region. The following sections provide details of the data preparation process, code settings for model estimation and post-processing of results. The documentation deals with three sections: the gauss code files (contprobitregression.gss, mnp.gss and jointvmtden.gss), the output file and the dataset (Final dataset.csv). The section on the gauss code file explains the code settings and the different options provided to the user. The Output file shows a sample output and the interpretation.

Dataset

This section highlights the data preparation and data coding procedures implemented to obtain the final datasets used for the modelling efforts. The data used in this study is derived from the 2010-2011 Regional Household Travel Survey (RHTS) conducted by the New York Metropolitan Transportation Council (NYMTC) and the North Jersey Transportation Planning Authority (NJTPA). The RHTS collected travel information for each household resident in the sample for one weekday. After extensive data cleaning, the household level data set included information for 14,790 households

STEP 1 Initial dataset 14790 observations: Initial Household level file derived from the NYMTC data set.

STEP 2 VMT_selectVarsHigh 4928 observations (or VMT_selectVarsMedium or VMT_selectVarsLow) are the files with high, medium and low density choices.

STEP 3 VMT_selectVarsHighSubset 1000 observations (or VMT_selectVarsMediumSubset or VMT_selectVarsLowSubset) are the files with 1,000 random observations derived from each of the files in Step 2.

STEP 4 VMT_selectVarscombinedSubset 3000 observations is the file compiled using the 1,000 observations from each of the files in Step 3.

STEP 5 Final dataset with all variables 3000 observations is the final file used for estimation with variables added based on household, individual and TAZ level characteristics. (This file is also converted to .xlsx format in the name of Final dataset.xslx)

The dependent variable of interest in this thesis is *weekday household vehicle miles of travel (VMT)*. Trip records provided by individual household members were used to derive VMT estimates at the household level. Household VMT is defined in this thesis as being exclusively arising from trips that are made by personal vehicle only. The household VMT was computed by aggregating distance traveled (in miles) across the personal vehicle trip records, *while explicitly ensuring that no trip was double-counted in the VMT calculation.* Thus, for example, if two household members travel together, only the mileage associated with the trip reported by the driver is counted towards calculating VMT. This was done to ensure that a clear distinction is drawn between vehicle miles of travel (VMT) and person miles of travel (PMT), and focus the analysis in this thesis exclusively on *household-level* VMT, which is naturally influenced by the extent to which household members travel jointly (rideshare or carpool). After calculating household VMT and appending the value to household records, data describing the traffic analysis zone (TAZ) of residence was also joined to the data set. Households were geo-located at the TAZ level, and data describing population and employment composition of the residence TAZ could be easily appended to the household level data set. For the current study, a random sample of 3000 households was extracted for analysis purposes. The sample here inculcates equal representation of low, medium and high residential location density observations to equally represent the categories under study. Comparisons were performed to ensure that the random sample is representative of the original 14,790 sample of households.

Gauss code files

contprobitregression.gss

Instructions to use the code:

1. Download contprobitregression.gss.
2. Copy this file to the folder you wish to run your simulation in.
3. Set the simulation settings as described in this section as per the simulation requirement. The code should be up and running.

The accompanying gauss code file (contprobitregression.gss) is a code written to evaluate the probit regression model for the VMT as the continuous dependent variable. The code offers the user the following options on the mentioned line numbers in the code:

- The chosen folder to save files and extract data from as “runpath” and “outpath” (Line 9-10)
- The number of individual observations (Line 13)
- The dataset (in the .dat format) that you’re working on (Line 39)
- The output file to save the results in as “logfile” (Line 41)
- The dataset must have a column of zeros (sero) and ones (uno) (Line 48 and 54)
- In Line 57, you put the probit regression specification. In the given file the dependent variable is “LogTime” as given in (Line 65), you must change this to your dependent variable
- Starting in (Line 74) you can put the starting values of your coefficients

mnp.gss

Instructions to use the code:

1. Download mnp.gss.
2. Copy this file to the folder you wish to run your simulation in.
3. Set the simulation settings as described in this section as per the simulation requirement. The code should be up and running.

The accompanying gauss code file (mnp.gss) is a code written to evaluate the multinomial probit model for the nominal unordered variable (residential choice). The code offers the user the following options on the mentioned line numbers in the code:

- To compare numerical and analytical gradients you can put the value of “Grdient_Check” as 1 in (Line 16)
- The chosen folder to save files and extract data from as “runpath” and “outpath” (Line 25-26)
- “nind” for the number of individuals in the dataset in (Line 28)
- “nchocc” for the number of choice occasion per individual (Line 29)
- “nobs” for the total number of observations in the dataset (Line 30)
- “output_file_gen” should be made 1 if you want to write estimation results to a file (Line 46)
- The dataset (in the .dat format) that you’re working on (Line 100)
- The output file to save the results (Line 101)
- The column names in the dataset to go in “vnames” (Line 109)
- The dataset must have a column of zeros (sero) and ones (uno) (Line 166 and 170)
- Column indicating chosen alternative. Should be in the format 1,2,3. (Line 173)
- The MNP specification goes in after chosen alternative (Line 177-179)
- Starting in (Line 194) you can put the starting values of your coefficients
- Starting in (Line 212) you can put the Covariance matrix for random parameters (exogeneous variables)
- Starting in (Line 219) you can put the Covariance matrix for the differenced error term vector
- Make “Non_IID” as 1, if you are providing non IID errors and to estimate the variance, otherwise zero (Line 227)
- The “max_active” in (Line 256) is the part for the covariance matrix as " 0|ones((nCholErr-1),1);". The first variance is fixed of course. Then it is the covariance and the second variance. So you should try 0|0|1, 0|1|0, and 0|1|1. Right now is going directly for 0|1|1

- The code is a multi-threaded code and you can define the number of threads in (Line 302) and then must divide the code into equal number of parts and set it in “Data_Split” following the strict format (Line 303 onwards)

jointvmtden.gss

Instructions to use the code:

1. Download jointvmtden.gss.
2. Copy this file to the folder you wish to run your simulation in.
3. Set the simulation settings as described in this section as per the simulation requirement. The code should be up and running.

The accompanying gauss code file (jointvmtden.gss) is a code written to evaluate the joint model estimating the probit regression model for the VMT as the continuous dependent variable and the multinomial probit model for the nominal unordered variable (residential choice). The code offers the user the following options on the mentioned line numbers in the code:

- Don't transform the continuous variables on to logarithmic scale. The code does that internally.
- The dataset (in the .csv format) that you’re working on (Line 6)
- The chosen folder to save files and extract data from as “runpath” and “outpath” (Line 11-12)
- Change “nvar_mear_cont” as per the total number of continuous variables in the measurement equation (here it’s 1, Line 27)
- Change “nvar_mear_ordl” as per the total number of ordinal variables in the measurement equation (here it’s 0, Line 28)
- Change “nvar_mear_grup” as per the total number of grouped variables in the measurement equation (here it’s 0, Line 29)
- Change “nvar_mear_count” as per the total Number of count variables in the measurement equation (here it’s 0, Line 30)
- Specify the Number of Options for each Nominal Variable. If you have at least one nominal variable with three or more options, you must provide the number of options corresponding to that nominal variable first followed by other nominal variables. That is, you must first provides the number of options corresponding to nominal variables with three or more options followed by binary unordered variables. If you do not have any nominal variable with three or more options, simply provide the number of options for each of the binary unordered variable. (Line 50)
- Specify the Number of categories for each of the ordinal Variable. if you have ordinal variables in your measurement equation, provide the number of categories for each of the ordinal variable, otherwise leave the specification as it is. (Line 74)

- Change the value of “gradient_limit” as per your requirement of the gradient limit for convergence (Line 117)
- Change the value of “gradient_limit” as per your requirement of the gradient limit for convergence (Line 117)
- The column names in the dataset to go in “vnames1” (Line 129)
- The dataset must have a column of zeros (sero) and ones (uno) (Line 201 and 207)
- The regression specification for the continuous variable goes in Line 209
- The dependent variable for the continuous regression goes in Line 216 (in place of “vmtone”)
- The utility specification for the multinomial probit model for the nominal unordered variable (residential choice) starts at Line 246
- The dependent variable for the MNP model goes in Line 285 (in place of “denchoc”). The MNP variable should be coded as 1,2,3 with 1 being the base alternative.
- Manage “iteration_limit” as per computation needs for the maximum number of iterations (Line 298)
- You can make “Force_converegnce” as 1 if you want to stop the estimation based on maximum number of iterations (Line 299)
- Starting in (Line 302) you can put the starting values of your coefficients for the regression specification for the continuous variable
- Starting in (Line 320) you can put the starting values of your coefficients for the multinomial probit model for the nominal unordered variable (residential choice)
- Some more clarifications about Lambda_error (Line 338):
  - Element 1,1 is the variance of the error term of the continuous variable
  - Element 2,2 is the variance of the utility difference between the second alternative and the base alternative. For identification this element should be always equal to 1 and should be always be fixed in your estimation.
  - Element 3,3, is the variance of the utility difference between the third alternative and the base alternative.
  - The other parameters are the correlation (between alternatives or between the VMT and the alternatives). The last row of this matrix does not count, ignore it.
- To include the correlation matrix between the two variables (joint estimation) you should modify the _max_active vector to see which elements of the matrix you want to estimate (starts in line 499, again.. please ignore the last row).

MATLAB code files

covtocorr.m used to convert co-variances to correlations

pos2dist.m used to convert position co-ordinates to distance values (in miles)

spatial_distance_matrix.m used to create a spatial matrix for the purpose of evaluating the estimation of spatial model.

Output file

Here's a sample output which gives the estimate of the joint model. The "Gamma" give the output for the VMT as the continuous dependent variable and "Beta" give the estimate coefficients for the multinomial probit model for the nominal unordered variable (residential choice)

```
===========================================================================
 MAXLIK Version 4.0.26                    7/31/2017   7:57 pm
===========================================================================


return code =    0
normal convergence


Mean log-likelihood        -3.41210
Number of cases     3000


Covariance matrix of the parameters computed by the following method:
Cross-product of first derivatives


Parameters    Estimates     Std. err.  Est./s.e.  Prob.    Gradient
------------------------------------------------------------------
Gamma01         0.7276       0.1008    7.218   0.0000      0.0000
Gamma02        -0.1270       0.0681   -1.866   0.0310      0.0000
Gamma03        -0.2385       0.0576   -4.142   0.0000      0.0000
Gamma04        -0.6025       0.0576  -10.456   0.0000      0.0000
Gamma05         1.6620       0.0746   22.293   0.0000      0.0000
Gamma06         2.6015       0.0803   32.399   0.0000      0.0000
Gamma07         0.1497       0.0321    4.658   0.0000      0.0000
```

Gamma08	0.2325	0.0747	3.112	0.0009	0.0000
Gamma09	0.2016	0.0741	2.722	0.0032	0.0000
Gamma10	0.1376	0.0548	2.513	0.0060	0.0000
Beta01	-0.1092	0.0336	-3.246	0.0006	0.0000
Beta02	-0.2184	0.0361	-6.053	0.0000	0.0000
Beta03	0.1973	0.0686	2.877	0.0020	0.0000
Beta04	0.1967	0.0730	2.695	0.0035	0.0000
Beta05	-0.3035	0.0875	-3.467	0.0003	0.0000
Beta06	0.4322	0.1081	3.997	0.0000	0.0000
Beta07	0.6283	0.1074	5.850	0.0000	0.0000
Beta08	0.2093	0.0520	4.028	0.0000	0.0000
Beta09	0.3170	0.0845	3.753	0.0001	0.0000
Beta10	0.4046	0.0846	4.783	0.0000	0.0000
Beta11	0.1863	0.0853	2.183	0.0145	0.0000
Beta12	-0.0919	0.0524	-1.754	0.0397	0.0000
Delta	0.0000	.	.	.	0.0000
Err1_01	1.2480	.	.	.	0.0000
Err1_02	0.0000	.	.	.	0.0000
Err1_03	1.0000	.	.	.	0.0000
Err1_04	0.0000	.	.	.	0.0000
Err1_05	0.4437	.	.	.	0.0000
Err1_06	0.8973	.	.	.	0.0000
Err1_07	0.0000	.	.	.	0.0000
Err1_08	0.0000	.	.	.	0.0000
Err1_09	0.0000	.	.	.	0.0000
Err1_10	1.0000	.	.	.	0.0000