R- Package for Vehicle Fleet Composition (VFC) Model System

Disclaimer: Please note that the VFC package code is provided with an intent to share the knowledge developed as a part of the research efforts indicated below. The package as such does not come with any 'guarantees' and it is assumed that the user is familiar with the MDCEV modeling method.

If you use the R-code for VFC model system (as a whole or in part) in your research, please do acknowledge so in you work. Cite the following papers in any and all of the efforts that utilize this code:

You, D., Garikapati, V. M., Pendyala, R. M., Bhat, C. R., Dubey, S., Jeon, K., & Livshits, V. (2014). Development of Vehicle Fleet Composition Model System for Implementation in Activity-Based Travel Model. *Transportation Research Record: Journal of the Transportation Research Board*, (2430), 145-154.

Garikapati, V. M., You, D., Pendyala, R. M., Jeon, K., Livshits, V., & Bhat, C. R. (2015). Development of a Vehicle Fleet Composition Model System: Results from an Operational Prototype. Working Paper, School of Sustainable Engineering and the Built Environment, Arizona State University, Tempe, AZ.

Papers corresponding to the parent MDCEV estimation/forecasting codes:

Bhat, C. R. (2005). A Multiple Discrete–Continuous Extreme Value Model: Formulation and Application to Discretionary Time-Use Decisions. *Transportation Research Part B: Methodological*, *39*(8), 679-707.

Bhat, C. R. (2008). The Multiple Discrete-Continuous Extreme Value (MDCEV) Model: Role of Utility Function Parameters, Identification Considerations, and Model Extensions. *Transportation Research Part B: Methodological*, *42*(3), 274-303.

Pinjari, A. R., and Bhat, C. R. (2009). Computationally Efficient Forecasting Procedures for Kuhn-Tucker Consumer Demand Model Systems: Application to Residential Energy Consumption Analysis. Technical paper, Department of Civil & Environmental Engineering, University of South Florida, July 2009, 1st revision June 2010, 2nd revision July 2011.

VFC Model System

1. **Installing R, RStudio and built-in packages required to run the VFC Model System (see the appended presentation titled 'Installation_Screenshots_VFC.pptx' for step-by-step instructions)**
   a) Download and install R package for Windows from http://cran.r-project.org/. Install the base package → instructions in slides 3-14.
   b) Use default options during the installation process

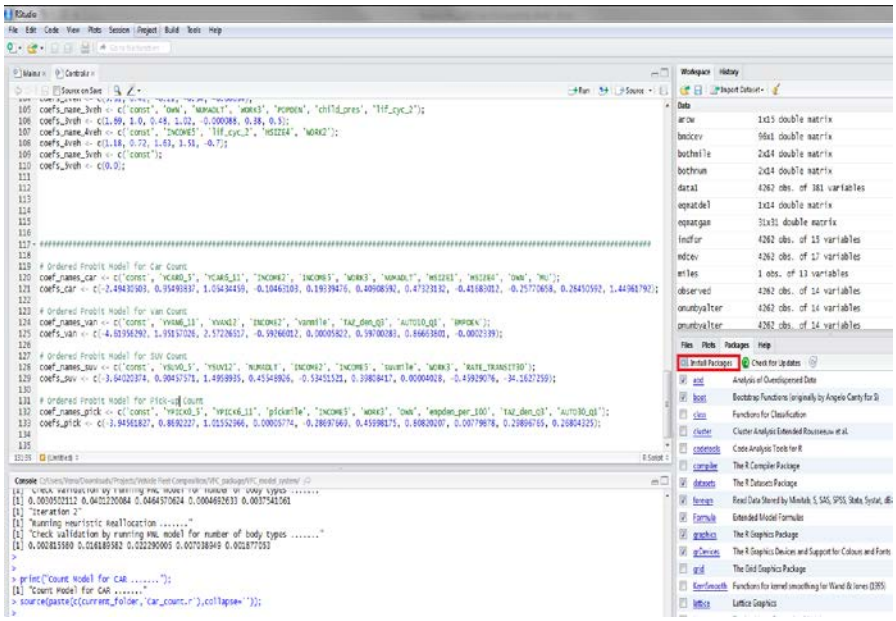

Figure 1. RStudio Dashboard

Note: Please be sure to install R before installing RStudio

   c) In the select components window that appears during the installation process, check 'Core files' and the configuration of your system (either 32 or 64-bit system).
   d) After installing a working version of R on your computer, download and install R-Studio from: http://www.rstudio.com/products/rstudio. RStudio is a user interface built for R that makes editing and understanding of R-code easy. Download the package that corresponds to the configuration of your system (Windows XP/Vista/7/8) → instructions in slides 15-22.
   e) The vehicle fleet composition model systems uses several inbuilt packages in R. Please make sure that all the packages are installed before you run the R-codes. The packages can be installed from the lower right sub-pane in the RStudio window named 'Install Packages' (see Figure 1) → instructions in slides 26-30.

2. **Package Structure**

   This package consists of four sub-folders: Software_R_Libraries, Halton_R, Model_Estimation, and Model_Application. Contents of each of these sub-folders is explained below:

2.1. Software_R_Libraries

This sub-folder contains all of the software required to carry out and *off-line* installation of R, RStudio and all of the libraries required to use the VFC package. The software is provided as a back-up only. The users are advised to download and install the latest and versions of all the required software components.

2.2. Halton_R

This sub-folder consists of four files that are used to generate a scrambled Halton sequence (error term) distribution to be used in the forecasting code for the VFC model system.

Table 1. Contents of the sub-folder 'Halton_R'

File Name	Description
Halton_Sequence.r	R code to generate Halton sequence numbers
xbrat.csv	Input file to Halton_Sequence.r
xiden.csv	Input file to Halton_Sequence.r
haltbrat_vfc101.csv	Output from Halton_Sequence.r. This file is used by the MDCEV model as input (error distribution).

2.3. Model_Estimation

This sub-folder contains R codes used to estimate various models (MNL, MDCEV, and ordered probit) in the VFC model system.

Table 2. Contents of the sub-folder 'Model_Estimation'

Name of the R Code	Model Type	Input File Name
1_PTR_mileage_prediction.r	Power transformed regression for mileage prediction	az_hhld_mileage_estimation_data.csv
2_MNL_Number_of_alternatives.r	MNL model for number of vehicle alternatives	az_hhld_vfc_cleaned_final.csv
3_MDCEV_Vehicle_fleet_mix.r	MDCEV model for vehicle fleet mix	num_alt_r.dat
5_MNL_Number_of_bodytypes.r	MNL model for number of vehicle body types	az_hhld_vfc_cleaned_final_restructured.dat
6.1_OP_Car_count.r	OP model for car count	cars.dat
6.2_OP_Van_count.r	OP model for van count	vans.dat
6.3_OP_SUV_count.r	OP model for SUV count	suvs.dat
6.4_OP_Pickup_count.r	OP model for pick-up count	picks.dat

A separate input file is provided for each R code to estimate a corresponding model as each of these models require a different data format. It should however be noted that all the data

files prepared for different model estimations essentially originate from the same cleaned data source which contains 4,262 households (az_hhld_vfc_cleaned_final.csv). Table 2 lists the R codes and corresponding input file names used by each model. These estimation files can be used to update/extend the models in VFC model system for new data. To run any model estimation code, double click on the code to open it in RStudio. Once opened, select the entire code (Ctrl + A) and run (Ctrl + R)

2.4. Model_Application

This sub-folder contains the files required for applying the VFC model system using a set of estimated models (using the codes given above). The forecasting code can also be used for calibrating the model by modifying the coefficients of certain models in the VFC graphic user interface (GUI). After the forecasting code is run, the user will see a set of graphs that show observed and predicted patterns from different models in the fleet composition model system. Set of codes that comprise the VFC model application are presented in Table 3.

Table 3. Contents of the sub-folder 'Model_Application '

Name of the R Code	Description	Editable
0_Main.r	Imports libraries, and reads dataset and Halton sequence	Yes
1_Mileage_prediction.r	Predicts annual mileage budget of the household	No
2_Number_of_alternatives.r	Predicts the number of vehicle alternatives owned by the household	No
3_Vehicle_fleet_mix.r	Predicts the fleet mix owned by the household	No
4_Heuristic_mileage_reallocation.r	Reallocates the annual mileage using a random assignment algorithm	No
5_Number_of_bodytypes.r	Predicts the number of body types owned by a household (for tolerance check)	No
6.1_Car_count.r	Predicts no. of cars owned by the household in each body type-vintage category	No
6.2_Van_count.r	Predicts no. of vans owned by the household in each body type-vintage category	No
6.3_SUV_count.r	Predicts no. of SUVs owned by the household in each body type-vintage category	No
6.4_Pickup_count.r	Predicts no. of pick-ups owned by the household in each body type-vintage category	No

In addition to these files, there is an input file named '**az_hhld_vfc_cleaned_final.csv**' which also resides in this folder.

3. How to Run the VFC Model System in R

The fleet composition model system application in R consists of two steps. In the first step, a scrambled Halton sequence distribution is generated. This is an essential input for the MDCEV forecasting algorithm. For applying the current version of the model system, this step has already been run and the output file is provided with this package. This is done because the generation of scrambled Halton sequence takes considerable amount of time and might slow down the process of testing the code. It is suggested that the users utilize the file already provided (and skip this step) for the current application of the model system. This step can be run for a future application as necessary. In the second step, entire array of models in the VFC model system are applied in a sequence finally resulting the distribution of household vehicle fleet composition by body type-age category along with the count of vehicles in each category.

3.1. Generating a scrambled Halton sequence (not required for each run)

This step is not mandatory each time the VFC model system is run. When this code is run, a new scrambled Halton sequence distribution is generated. The R-code and the files provided already have a Halton sequence generated for the purposes of running the forecasting code. Follow these steps to generate a new scrambled Halton sequence:

a) Open 'Halton_Sequence.r' in the folder Halton_R in RStudio
b) Change working directory to the one in which this code resides. Please note that folder separators in r-code are forward slashes '/', but not backward slashes '\'
c) Fix values of the following variables
    a. file_name → Provide an output file name of your preference
    b. draw → How many Halton numbers do you intend to generate. This should be equal to the size of your sample * number of error terms you want to generate for each record in your sample
    c. dim → How many alternatives are there in your model? (including outside goods)
d) Save the code (Ctrl-S)
e) Select entire code (Ctrl-A)
f) Run (Ctrl-R)

An output file will be created with the name that you have set in the folder **Halton_R**.

3.2. Running the fleet composition model system (all models)

a) Open '0_Main.r' from the sub-folder 'Model_Application' in RStudio
b) There are two ways to invoke the VFC model system:
    i. Type the following command in RStudio console: 'source("C:/workspace/VFC_Package_Beta/Model_Application/0_Main.r")' and hit enter

Or

ii. Select the entire '0_Main.r' code (Ctrl + A) and run (Ctrl + R)

c) A GUI shown in Figure 2 will pop-up

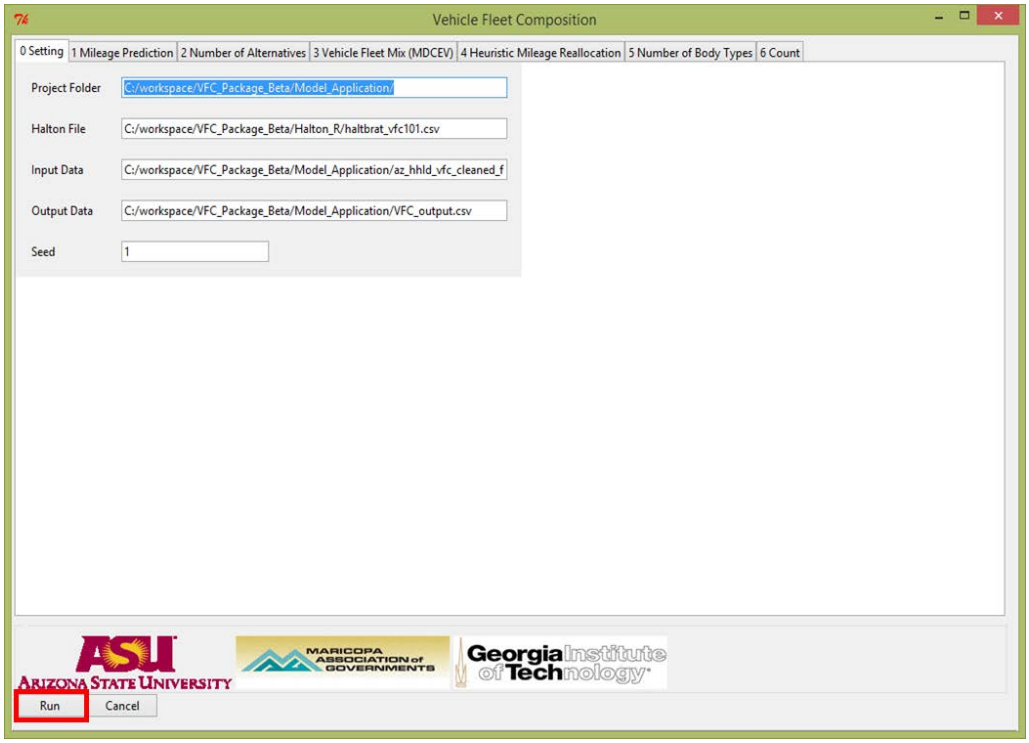

Figure 2. Vehicle Fleet Composition GUI

d) Change the name and/or location of 'Output Data' field as per your requirements. The 'Seed' number can also be changed if you intend to generate and use a different set of random numbers (for various model components) for model application than the ones used in the model system. If you have generated a new scrambled Halton sequence, update the 'Halton File' field accordingly. Otherwise, leave this field as is (Please do not change the location and/or name of 'Project Folder', and 'Input Data' as is)

Click the **Run** button located at the bottom-left corner of the GUI (see Figure 2). This step will run the VFC model system from start to finish in a sequential fashion. Comparison charts showing results from various components of the model system will get populated in the 'plots' pane of RStudio dashboard. Messages pertaining to the model run will be shown in the RStudio console.

Correspondence: For any questions/clarifications regarding the code, please contact

- Daehyun You (daehyun.you@gatech.edu)
- Venu Garikapati (venu.garikapati@gatech.edu)
- Ram Pendyala (ram.pendyala@ce.gatech.edu)