Notes on the case study data and case study in HRG (2005)

(and the file CaseStudy.doc received in DH’s March 2, 2025 email which is Chapter 9 of the book but with differing section, table and graph numbering)

1 Data 1

1.1 Level labelling of variable CN 1

1.2 Levels of TOLLPRED 2

1.3 Replicated data set descriptions 2

2 Supplementary regressions 2

# Data

## Mode labelling (variables ALTIJ, CN,…)

Appendix /B:

The variable altij contains the label of each alternative. Coding according to Appendix 9c in book is:

1 cart Car with toll

2 carnt car no toll

3 bus Bus

4 train Train

5 lr Light Rail

6 bw busway

This coding from *altij* for SP is not consistent with *spchoice*, the choice dummy variables *spcart,spcarnt,spbus,sptn,spbw*,*splr* and with the variable *cn* indicating options 3 and 4 in each choice situation: **the labels „lr” and „bw” should be swapped in altij**:

1 cart Car with toll

2 carnt car no toll

3 bus Bus

4 train Train

5 bw busway

6 lr light rail

This is also confirmed by checking that SPCHOICE has four identical values for each experiment number (combination of ID and SPEXP): the value of ALTIJ in the alternative (row) chosen by the respondent (CHOICE = 1).

If my assumption is correct that this modification is needed.the sentence “Given the method used to determine which public transport methods were present within a choice set the *bus and LR* and *train and busway* alternatives never appeared in the same choice set.” in section 9.4.1 (p.268) needs to be modified too by swapping LR and busway.

## Levels of TOLLPRED

Table 9.4 in HRG2005 (p.269) indicates three levels of the TOLLPRED variable (“Pay toll if you leave at this time (otherwise free)”

|  |  |
| --- | --- |
| Level code | Label |
| 0 | 06:00 - 10:00 |
| 1 | 06:30 – 08:30 |
| 2 | 06:30 – 09:00 |

A consistent ranking regarding restrictiveness would imply different labels, e.g. by exchanging the second and third. Which labels for levels 0, 1, and 2have actually been used in the choice cards?

## Cleaning

The assignment of choice cards is in some cases not in accordance with stated usual commuting time (SPLENGTH). Respondents stating SPLENGTH level 3 (30-45min) sometimes received choice cards with travel time attribute levels below 15 or above 45 minutes. I did not find any problems regarding the assignment of choice cards to respondents with SPLENGTH level 2 (<30).

## Replicated data set descriptions

The data set received has 2 observations more than indicated in the book. ID 6207 has one choice (4 rows) and two RP-rows. These are the only two rows with splength=”RP”. I try removing these two rows (in *01401a\_hensher\_imp\_repl.r*) to see if this entails consistency between data an book. It does exactly replicate descriptives such as the mean of variable ID.

# Multicollinearity checks

For including other alternatives (as suggested in the last paragraph of p. 295) requires data rearrangements (as explained on p. 247). Implemented using apollo\_longToWide() form apollo package (01402\_hensher\_descr.r).

Supplementary regressions on data with splength <30min and for first (=CART) alternative (p. 293-95 in the book) exactly replicated: two supplementary regressions suggest multicollinearity.

Regressions for second alternative (CARNT) find no collinearity.

Regressions for remaining alternatives (BUS, TR, BW,LR) indicate collinearity throughout.

This corresponds with correlation matrix:



Image 1: Correlations (SPLENGTH = <30min)

Is this in the data or due to a bug?

Large correlations occur within each of the public transport alternatives as well as between the attributes in Train and Busway alternatives and

For respondents with longer commuting times the pattern is similar:



Image 2: Correlations (SPLENGTH = 30-45min)



Image 3: Correlations (SPLENGTH = >45min)

# MNL estimation

## preparations

## results

* For the first model (6 modes, 5 ASC, one cost parameter, CST, for fuel and fare) HRG05 Ch. 10.3 (p.316) I get identical results for final LL and parameter estimates (except for LL of no-data base model and the resulting fit statistics): 01403\_hensher\_mnl1.r
* ASC-only model (p.329): 01403\_hensher\_mnl2
* LR-Test comparing two models (335ff): 01403\_hensher\_mnl3
* HRG05 10.4 (p.344) Dummy coded attributes (nonlinearity) 01403\_hensher\_mnl4
* HRG05 10.5 (p.352ff) interactions1 01403\_hensher\_mnl5  
  results in the book (p.353) result from the code on p.352. (with erroneously using the categorical variable *drivlic* (respondent’s drivers licence) rather than *ndrivlic* (number of licensed drivers licenses in household). The result from using *ndrivlic* (01403\_hensher\_mnl5.lis) suggest the interaction term vehlic non significant, however the LR-test indicates improvement over base model. Using *ndrivlic* rather than *drivlic* requires changes to Table 10.4 and Figure 10.11.
* HRG05 10.5 (p.355-57) Interactions w categorical vars: exact replication: 01403\_hensher\_mnl6
* HRG05 10.6 (p.357-59) WTP: