

Memorandum

TO: Birat Pandey, BMC
FROM: Thomas Rossi
DATE: August 26, 2013
RE: Model estimation software

This memorandum summarizes options for model estimation software for the new BMC activity based model components. We evaluated the following options for the software to be used for the project:

- ALOGIT (Rand Europe)
- Biogeme (Bierlaire - open source)
- ELM (ELM-Works, Inc.)
- R (The R Project for Statistical Computing)
- SAS (SAS Institute)
- STATA (StataCorp)

Note that while there are some reasons for CS staff and BMC staff to use the same model estimation software (common data set formats, ability for CS to more easily assist BMC staff), it is not strictly necessary to do so. Since in general CS will be estimating the more complex models, what is most efficient for CS may not be best for BMC.

Based on our initial evaluation, we do not recommend the use of ELM, STATA, or SAS for the model estimation for this project:

- ELM is relatively new software, and it is not free. CS does not have much experience with it, and so we are not sure how well it will do in estimating the more complex models in this project. There does not seem to be much on-line help at present (recent attempts to access the ELM-Works web site have been unsuccessful). Since there would be a learning curve for CS, and that would also limit how much we could help BMC, we feel that it is preferable to use something with which we are more familiar.
- SAS is a well-known statistical analysis package. It is relatively expensive, but CS already has a license for it (if BMC has a license already, that would reduce the cost). We use it for statistical analysis and data set preparation, and so there would be an advantage in being able to use it to create estimation data sets that would be consistent with the requirements of the estimation software if SAS were used for estimation. However, despite this advantage and SAS's availability to us, we have not used it for model estimation. Part of the reason is our lack of familiarity with the model estimation

functions, but there is also concern about the ability to do the specialized estimation we need for the ABM.

- STATA has been used by one CS staff member prior to his work at CS. STATA is not free but it is user friendly. There are information and helpful websites to make model estimation easier. The disadvantage with this software may be the necessary data massaging/preparation before model estimation. The data for choice models must be inputted in a different type of format what we are used to using with software such as ALOGIT or Biogeme. Because of these reasons and that we have only one experienced CS user, meaning there would be a learning curve for both CS and BMC, we are not recommending the use of STATA for this project.

We therefore recommend discussing the following options with BMC:

- ALOGIT has been used for this type of model estimation more than any other software. It is not free, but CS has a license (BMC would likely have to purchase a license to use it). It has proven very flexible for estimating these types of models. ALOGIT is particularly good with very large datasets/choice sets. It may not be the most user friendly package, but all CS model estimation staff have experience with it and could help train and assist BMC staff.
- Biogeme is a free open source package that was developed by Michel Bierlaire, a well-known choice modeling researcher in Europe. BIOGEME is generally user friendly. There is very good documentation, a pretty good user forum (e-mail list) for questions, and a lot of examples in documentation for testing. Our experience has been somewhat mixed, with some model estimations using earlier versions taking a very long time to converge (we ended up using ALOGIT in these cases). Some CS staff have experience, but not as many as with ALOGIT.
- R is a free open source programming language that is well suited for model estimation. Our experience is somewhat limited, but it has been used in some other model applications in the U.S., including Portland, Oregon (not just for model estimation). There is a large user community that could be drawn on for help. R is supposed to be quite powerful.

The next steps are to discuss with you any experience your staff has with model estimation and the use of any of these software packages. We should have a conference call with everyone at BMC and CS who would be involved in model estimation (and in “coaching” from our end) to come to a conclusion. This can take place in September or October, depending on schedules.