8 May 2016

Fedora Submission

Benjamin M. Craig

Model Description

1. Describe your choice of software and reasons underlying your choice (e.g., Stata)

For this competition, I used the STATA MP, specifically version 14.1 with 12 cores. From my perspective, this software has provided me with a suitable balance between packaged commands and the ability to program and publish my own commands (e.g hyreg). Knowing what the estimator is doing at each step in the analysis has provided me with greater insights in the limitation of alternative approaches (such as the importance of initial values). Although I read the descriptions provided in the software handbooks, I also appreciate seeing how the code operates (line by line), which is not possible in many canned packages. At times, the small numerical issues (such as rounding errors) can have substantial consequences in the interpretation. I am trained to program in other languages, but some require more effort for simple commands (e.g., Gauss) or don't show the underlying code (SAS). Other reasons for using this software package includes its handling of large datasets, its accessibility for novice programmers, its widespread community of users, and its graphics.

2. Describe your choice of estimation technique and reasons underlying your choice (e.g., Bayesian)

As a frequentist, I typically begin modeling by specifying a decision rule to either maximize or minimize. In the case of binomial analysis, I choose weighted least squares (WLS), because it has the same first derivative as the maximum likelihood (ML) function, but it allows for unanimous predictions (e.g., 0 and 1). The weights are typically based on the predicted probability, not the empirical probability (See response #5). WLS has a limitation that if the predictions are unanimous, a weight correction is required (e.g., Berkson weights); otherwise, it has served me well past analyses. This estimation technique is also known as minimized chi-square, Urban's Normit, or GLM. Basically, it has the advantage of minimizing chi square, which is the basis of this competition. If it is the winner, I may explore alternative estimation techniques that allow preference heterogeneity, which will require switching either to ML or Bayesian approaches (i.e., more assumptions).

3. Describe your choice of functional form and reasons underlying your choice (e.g., Logit)

The functional form has two components. The first is the cumulative density function (CDF). For this competition, I choose to use the Bradley-Terry model (i.e., A/(A+B)). Under this CDF, scaling terms that are common to A and B cancel. Second is the value specification. For this, I choose Value = lifespan^{alpha} - problems × duration^{beta} where problems includes the 5 attributes of the EQ-5D description. This functional form was identified semi-parametrically in a previous study. In this competition (unlike the previous study), all problems have the same duration as lifespan (i.e., lifespan=duration), but it seems appropriate to allow for different time preferences (alpha and beta). If I were to choose an alternative functional form, I would consider allowing the beta to vary by health problem (i.e., the effect of duration of slight problems may not be the same as severe problems).

4. Describe your choice of variables and reasons underlying your choice (e.g., 20 effects-coded variables)

The regression model includes only the 20 effects-coded dummy variables. These are standard in most EQ-5D valuation studies. In this parsimonious model, each coefficient represents the loss in QALYs incurred by an increase in a domain by 1 increment of level (e.g., going from level 2 (slight problems) to 3 (moderate problems) on Mobility). I did not include any interaction terms, adjustments for scale, adjustment for temporal units, adjustments for pair types (i.e., TTO pair vs. efficient pairs), or behavioral parameters (e.g., left/right, sequence), which might have improved fit. This is the simplistic approach (20 regression parameters and 2 time effect parameters [alpha and beta]).

Modeling Recommendations:

5. Did you have difficulty modeling the 2 pair types (TTO pairs [quantity vs. quality] and efficient pairs [all attributes])? Did you have difficulty with the 4 temporal units (days, weeks, months, years)?

No, I did not. Studies on preferences between health-related goods and services (e.g., choice-based conjoint) tend to be small and focused. For example, the value of night in the hospital after knee surgery. In health valuation, we attempt to understand preferences on all health outcomes (e.g., imagine all possible durations and experiences in a hospital). Sometimes we asks about preferences between improved quality of life and extended lifespan (i.e., QALYs using TTO pairs) and others we ask about preferences between 2 entirely different health outcomes (i.e., efficient pairs). Both are relevant; therefore, we need a unifying model (e.g., Fedora).

6. Does you believe that you would have been able to predict choice probabilities better had you received data on the respondent characteristics as part of the exploratory dataset (e.g., age)? Why?

I do not believe so. Preference heterogeneity exists, but it is difficult to separate preference heterogeneity from blocking in the pair allocation (e.g., respondents asked similar pairs may seem to have similar preferences). Little evidence suggests that there are substantial differences in preference weights across HRQOL domains, but there might be differences in time preferences (alpha and beta) by respondent age and health.

Most estimation techniques applied to identify preference heterogeneity are assumption laden and dependent. A natural next step after identifying the merits of alternative modeling approaches would be to examine which models predict heterogeneity the best (similar to a confirmatory factor analysis). This outside the scope of the competition at this time.

7. Did you change your model's functional form or variables based on the estimation results (i.e., data mining)? If so, why and how? If not, why not?

Yes, I changed the functional form in two ways. First, I originally planned to include a theta term to adjust for non-traders in the analysis (i.e., the proportion of respondents who always choose the longer lifespan); however, the parameter added little to the predictions and was dropped. Second, I originally ran WLS using the empirical probabilities in the sample weights (see example code), but replaced these with the predicted probabilities, because the revised estimation is more efficient, when correct. This second change had little to do with the results, but was based on a discussion with Mark Oppe. The parameters and variables were not changed based on the results.

8. If your model wins, why do you believe it predicted better than the other models? If your model loses, why do you believe it did not predict better than the other models?

The top three reasons that Fedora might win are: (1) it does not use a logit; therefore, is not as susceptible to proportional scaling issues; (2) it relaxes the constant proportionality assumption by incorporating time preferences (alpha and beta); and (3) it minimizes chi square, which is the primary metric for model comparison.

The top three reasons that Fedora might lose are: (1) it does not data mine and a model with more parameters may predict better (e.g., N3 term); (2) it does not take into account behavioral effects, such as left/right bias, sequence bias, or non-trading, which could reduce error and improve prediction; and (3) it does not separately model the TTO pairs. The TTO pairs have the same number of attributes, but requires less information, because one description is always in full health. Modeling these choices may be intrinsically different than modeling the more complex paired comparisons. As discussed here, Fedora is limited to just preference attributes, and does not incorporate behavioral characteristics, which might predict choice beyond preference.

9. Based on your expertise and experience, what are the primary econometric advances needed to improve predictive modeling (not design)?

The top three advances to improve predictive modeling are: (1) we need a better understand of the correlation between choices (i.e., two choices from 1 person might provide more information than two choices from 2 people [e.g., intervals]); (2) we need greater diagnostic tests and investigation in interaction effects; and (3) we need to examine angular- and ratio-based approaches to modeling choice (e.g., log Cauchy, Bradley-Terry), which may address censoring and scaling issues.

Competition Recommendations:

10. What recommendations do you have to improve the competition?

I wish that someone else would run the competition and that I did not have to share my entry in advance (Joke). This is our first time running such a competition and Kim and I are quite pleased with how it is going thus far. We are particularly grateful for the wealth of support from the teams. Nevertheless, I would recommend for next time that (1) we have multiple rounds, kind of like playoffs; that (2) we improve the design based on the victorious model (20/20 hindsight); and that (3) we have more prizes (e.g., most elegant model), maybe even a prize for the worst model (i.e., Lanterne Rouge). Mostly, I want us to have geeky fun.