

Final Exam

Topics in Advanced Econometrics (ResEcon 703)
University of Massachusetts Amherst

Due: December 19, 11:59 pm ET

Rules

Email two files — a .R file of your code and a .pdf file of your writeup, code, and output — to `mwoerman@umass.edu` by the date and time above. This final exam is solo-authored, meaning you cannot collaborate or consult with anyone, including (but not limited to) your classmates. The only person with whom you may discuss the content of this exam is Matt Woerman, who will answer only clarifying questions. You may refer to any pre-existing materials in print or online, including (but not limited to) lecture slides, textbooks, your notes, official R help documentation and package vignettes, R tutorials, and R forums and message boards; but you cannot post your own questions on these forums. Any violation of these rules will result in a failing grade for the course and appropriate disciplinary action.

You have two options for answering problems 2 and 3 of this final exam: using canned routines or writing your own estimation code.

Canned routines If you use canned routines to answer these problems, the highest course grade you can receive is A-. For example, even if you have a final score of 100% for the course, you will be assigned a course grade of A- if you use canned routines to answer problems 2 and 3. This option is easier but may limit your final course grade.

Hand-coded estimation If you write your own estimation code, there is no such limit to the course grade you can receive. This option is more difficult but will not limit your grade. In order to receive an A for the course, you must write your own estimation code.

Clearly indicate which option you have chosen. Any code corresponding to the other option will be ignored when grading. For example, if you indicate you are writing your own estimation code, any use of the `mlogit()` function will be ignored.

Data

Download the file `exam_dataset.zip` from the course website (github.com/woerman/ResEcon703). This zipped file contains the dataset, `nox.csv`, that you will use for the final exam. See the file `data_descriptions.txt` for descriptions of the variables in the dataset.

Paper

For this final exam, you will replicate the main estimation of the following paper:

Meredith Fowlie. 2010. "Emissions Trading, Electricity Restructuring, and Investment in Pollution Abatement." *American Economic Review* 100 (3): 837–869

This paper is available for download from the AER website (doi.org/10.1257/aer.100.3.837).

Problem 1: Summary Statistics

Before trying to replicate the estimation of the paper, first make sure you can recreate the summary statistics presented in Figure 2 and Tables 1–3.

- a. The main point of this paper is to show that different regulatory regimes lead plant managers to make different choices when complying with NO_x Budget Program. Figure 2 nicely summarizes this finding by depicting, for each regulatory regime, the proportion of installed capacity with each category of compliance choices. Calculate these proportions shown in Figure 2. Do not plot your results; report the proportions in a table or well-organized R output.
- b. One potential concern with this analysis is that units in different regulatory regimes have different characteristics, and these different unit characteristics drive the compliance choices. Table 1 shows this is not the case because characteristics are broadly similar across regulatory regimes. Calculate these unit summary statistics reported in Table 1. Note that these summary statistics, and all future calculations, include only the units with sufficient data, as given by the `full_data` variable.
- c. A related concern is that the costs of compliance choices may differ for units in different regulatory regimes, which would clearly drive differences in compliance choices. Table 2 shows this is not the case because compliance costs are broadly similar across regulatory regimes. Calculate these compliance cost summary statistics reported in Table 2. Note that each row of this table reports costs for a specific compliance alternative (not for a category of compliance alternatives, as in Figure 2); in descending order, the rows correspond to compliance alternatives 2 (CM), 7 (LNB), 15 (SN), 14 (SC), and 10 (N), respectively.
- d. One final concern is that all of the compliance alternatives are not feasible at every unit because of unit characteristics, and differences in choice sets could drive difference in compliance choices. Table 3 shows this is not the case because compliance choice sets are broadly similar across regulatory regimes. Calculate these choice set summary statistics reported in Table 3. Note that, in the bottom panel of this table, the first two rows correspond to categories of compliance alternatives — categories `cm` and `lnb` — and the final two rows correspond to individual compliance alternatives — alternatives 15 (SN) and 14 (SC). You may not be able to replicate these summary statistics as closely as the previous summary statistics, but you should find similar patterns in the data.

Problem 2: Model Estimation

Table 4 reports the main estimation results of this paper. Six models are estimated, three logit models (or conditional logit models) and three mixed logit models (or random coefficient logit models); these models are described in Sections IV and V of the paper. Estimate each of these six models. Report the parameter estimates, standard errors, p-values, and log-likelihood value for each model. You should be

able to closely replicate the parameter estimates and log-likelihood values reported in Table 4, but the standard errors and p-values of models 2, 3, 5, and 6 will likely be smaller than those reported in the paper. See these additional comments for each exam option:

Canned routines Use the `mlogit()` function from the `mlogit` package to estimate these models. A few notes on using the `mlogit` package for these models:

- You should only include in your estimation the compliance alternatives that are available to each unit. Use the `subset` argument in the `mlogit()` function to limit the estimation to these available compliance alternatives.
- Models 1 and 4 include scale parameters for deregulated and public units. To estimate scale parameters, specify a fourth “bin” in your formula within the `mlogit()` function; the variable in this fourth “bin” must be a factor variable. For examples of estimating scale parameters, see the `mlogit` vignettes (cran.r-project.org/web/packages/mlogit/index.html).
- We observe choices for each unit, but the paper assumes choices are being made by the facility manager. So we observe multiple choices (units) for each individual (facility manager), which is effectively panel data. To implement panel data in the `mlogit` package, use the `id.var` argument in the `mlogit.data()` function and the `panel` argument in the `mlogit()` function.
- When estimating the mixed logit models, set a seed so your results can be replicated and use at least 1000 random draws to most accurately simulate choice probabilities.
- You may have difficulty getting some of these models to converge to a global maximum. If you encounter this problem, try different starting values (the `start` argument in the `mlogit()` function) and convergence methods (the `method` argument in the `mlogit()` function). Some good method options include ‘`nr`’, ‘`bfgs`’, and ‘`bhhh`’.

Hand-coded estimation Write your own functions to calculate log-likelihood, and then use the `optim()` function to find the set of parameters that maximizes the log-likelihood for each model. A few notes on writing your own estimation code for these models:

- You should only include in your estimation the compliance alternatives that are available to each unit.
- Models 1 and 4 include scale parameters for deregulated and public units, as defined in Equation (6) of the paper. These parameters scale cost for all units within that regulatory regime to account for differences in the error term variances for the different regimes.
- We observe choices for each unit, but the paper assumes choices are being made by the facility manager. So we observe multiple choices (units) for each individual (facility manager), which is effectively panel data. These panel choice probabilities are given by Equations (3) and (4) in the paper.
- When estimating the mixed logit models, set a seed so your results can be replicated and use at least 1000 random draws to most accurately simulate choice probabilities.
- You may have difficulty getting some of these models to converge to a global maximum. If you encounter this problem, try different starting values and convergence methods in the `optim()` function.

Problem 3: Manager-Specific Coefficient Distributions

Table 5 summarizes manager-specific coefficient distributions for the random coefficients in models 4–6. For each model, summary statistics for both the population (or unconditional) parameters and the

conditional parameters are reported, and these statistics are reported separately by regulatory regime. Calculate these manager-specific coefficient distribution summary statistics reported in Table 5. You may not be able to perfectly replicate all of the values in this table, but they should generally be close; you should also find the same patterns: means are roughly equal for comparable unconditional and conditional distributions, but standard errors are smaller for the conditional distributions. See these additional comments for each exam option:

Canned routines Unconditional (or population) distribution summary statistics can be calculated using only the model results. To calculate the conditional distribution summary statistics, first find manager-specific mean conditional coefficients using the `fitted()` function with argument `type = 'parameters'`.

Hand-coded estimation To calculate these distribution summary statistics, use the same random draws that you used when estimating the corresponding model.