How to use the files in reformEstimation.zip

Andrey Launov[(*)] and Klaus Wälde[(*), 1]

[(*)]University of Mainz, Université catholique de Louvain and CESifo

May 21, 2010

This file provides information on how to work with the matlab code contained in reformEstimation.zip.

Section 1 describes the contents of the replication package. Section 2 explains the structure of the data sets and solutions stored in .mat-workspaces. Section 3 explains how to estimate the model and repeat our results.

The description is relatively short. More detailed explanations can be provided should questions arise.

1 Contents of reformEstimation.zip

1.1 Technical preparation

Unzip the files by preserving the structure of the subdirectories. Lets assume the directory in which you save the files is called estimation. Open matlab and use the "File" menu, to add the entire directory \estimation with all subdirectories to matlab path library:

"File" → "Set Path ..." → "Add with subfolders" → [select the \estimation folder] → "Save".

1.2 Contents

There are three folders in the file reformEstimation.zip:

\matlab - contains the entire estimation code

\results - contains the workspaces with all the data and estimation results

\figures - contains the source files for figures related to estimation

Brief description follows.

1.2.1 Matlab Folder

Key subfolder is \toolbox, which contains the likelihood function and the solution code for the structural econometric model. The other subfolder - \auxies - has the auxiliary files and data transformation utilities, some of which will be automatically called during the estimation.

[1]Both authors are at the Mainz School of Management and Economics, University of Mainz, Jakob-Welder-Weg 4, 55131 Mainz, Germany. Fax + 49.6131.39-23827. andrey.launov@uni-mainz.de, phone + 49.6131.39-23233. klaus.waelde@uni-mainz.de, www.waelde.com, phone + 49.6131.39-20143.

toolbox subfolder

In the toolbox folder you will see two .m-files: "logl.m" and "logl_mix.m". These are the total likelihoods which we maximize in the estimation part of the paper. "logl.m" describes specifications I and II. "logl_mix.m" describes specification III. Both functions accept three arguments: 1) the parameter vector (**x** in "logl.m" or **x0** in "logl_mix.m"), 2) the data matrix with individual duration, benefits and earnings information (**dmxs**), and 3) the data matrix with individual characteristics (**dmxc**). Both these functions return the negative value of the total log-likelihood.

The subdirectory \indclf includes the files that solve block 1, solve for the path of the exit rate at any given time point and evaluate the individual contribution to the total log-likelihood. All the files in this subdirectory are subordinates to "logl.m" and "logl_mix.m".

_**auxies** subfolder

In the _auxies folder you will see three subdirectories: \datatran - contains auxiliary files for matrix transformations [during estimation of specifications I-III these will be called automatically]; \nonphfs - contains the code for computation of nonparametric hazard functions reported in fig. 1 [*does not* take part in estimation of specifications I-III]; \vtstrnon - contains the code for estimation of a competing model without spell effect for the vuong test [*does not* take part in estimation of specifications I-III].

1.2.2 Results Folder

Contains four subfolders.

The first three: \sp_i to \sp_iii are the subfolders in which the data and estimation results for specifications I to III are stored as matlab workspaces named "specification_(number).mat". In addition to that, in each of these subfolders there is a complete iteration log reporting the estimation history: convergence, gradient at the solution, estimated parameter vector, the significance test results for the estimated parameter vector and the value of the total log-likelihood at the maximum.

The fourth subfolder \v_test contains the estimation results of the specification III restricted to have no spell effect, and the results of the vuong test for equivalence of specification III with and without spell effect.

1.2.3 Figures Folder

Contains the plots of the estimated nonparametric hazard functions (fig. 1 in the paper) as well as the plots of the Kaplan-Meier estimates of the survivor function along with their confidence intervals and the survivor functions predicted by the structural model (fig. 7 in the paper).

2 Structure of the data

All saved workspaces with the initial data and the estimation results ["specification_i.mat"-"specification_iii.mat" and "specification_iii_restricted.mat"] have identical structure. The objects stored in these saved workspaces bear the same names. These are:

A: Data objects [identical across all workspaces!]

dmxs - data matrix with individual duration, benefits and earnings information

dmxc - data matrix with individual characteristics [*absent in "specification_i.mat" since specification I considers an unconditional model*]

B: Solution objects

b - vector that contains the maximizer of the likelihood function

cov - covariance matrix for **b**

f - value of the negative total log-likelihood at the maximum

grad - gradient evaluated at the maximizer

stats - matrix with the output of the tests for significance of elements in **b**

C: Objects related to optimization

xst - vector of starting values for the estimation

bds - matrix with the lowest and highest admissible values for each parameter, defining the range on which the solution is searched for

exit - exit flag of the optimization routine at the solution

out - matlab structural object with the information summary on the completed optimization

hess - matrix of second order derivatives of the objective function evaluated at the solution

lambda - matlab structural object with the information summary on possible constraint violation

opt_ml - matlab structural object with optimization defaults for the estimation

3 How to estimate the model

3.1 Start

To be able to replicate our results one first needs to see how can one compute the value of the total log-likelihood given the data and the vector of the starting values for the estimation. Open any of the results files and for a given specification enter the command:

Workspace file	*Command*	*Result*
specification_i:	fst=logl(xst,dmxs,[]);	fst = 2925.6470
specification_ii:	fst=logl(xst,dmxs,dmxc);	fst = 2859.5949
specification_ii:	fst=logl_mix(xst,dmxs,dmxc);	fst = 2825.8527

The result you should receive for 'fst' is given in the last column. Once this is done and the model at the initial vector is checked, one can proceed with the estimation.

3.2 Estimation

To maximize the total log-likelihood we use the internal matlab routine "fmincon" available with the matlab optimization toolbox. This would minimize the negative total log-likelihood. Optimization is constrained in a sense that we look for a solution on a pre-specified range, defined in **bds**, where: $^{a)}$ the range is wide enough to insure that the maximum is in the interior, and $^{b)}$ the range is not too wide to exclude unreasonably high or low values for certain parameters that will make equilibrium solution for certain individuals unfeasible.[2]

To start the optimization, simply use the syntax of "fmincon". For instance, to estimate specification III type in:

```
[b,f,exit,out,lambda,grad,hess]=fmincon('logl_mix',xst,[],[],[],[],bds(:,1),bds(:,2),[],opt_ml,dmxs,dmxc);
```

You can also copy-paste this command from the corresponding log-files.

All estimations were done in Matlab 7.8.0.347. Estimation of specifications II and III is extremely time consuming. For replication of our results it is strongly advisable to use computer cluster with parallel computing possibility to estimate specification II. For estimation of specification III using parallel computing is indispensable.

In all our estimations we were using 8 processors working in parallel.

[2] All MLE we get lie in the interior, which can also be checked by inquiring the **lambda**.