Explanation of the Do files

In the folder "Do files" we save the following material:

Figure1.do:

This program helps us generating **Figure1** which represents the eigenvalues of the Jacobian for different combinations of the parameter space (alfa, sigma). The do file generates artificially the truncated dependent variable and then imposes a very simple model that allows us to obtain all the elements of the Jacobian matrix. Finally, a reparametrization of the Jacobian is conducted (see technical details in the paper) from which the final eigenvalues are computed.

Table1_Figure2.do:

This do file includes the programming of the algorithm (FILS) for the simplest model in the paper and carries out a Monte Carlo simulation exercise to compare FILS with Tobit[1]. The idea is the following: we construct the latent variable (**ystar**) as the sum of a constant (**alfa**) and an stochastic error component (**u**) that follows a Normal distribution (0, **sigma**). Next we set the censoring level, which is **c**, and construct the truncated dependent variable named **y**. Once we have **y** created we carry out the two alternative estimation procedures, FILS and Tobit. This is done repeatedly **S**=100 number of simulations.

The estimators for the parameters of the model (alfa and sigma) are saved in each case. And finally the mean of both alfas and sigmas across Monte Carlo simulations is computed to check the distributional aspects of the estimators – this corresponds to **Table 1** in the paper. The different designs of the Table just include a change in the degree of censoring and a change in the sample size: in the do file this is done just by changing **N** or **c**. In addition, the kernel density of both estimators of the parameters that come from the FILS procedure are plotted – This corresponds to **Figure 2** in the paper.

Table1_two.sides.censoring.do:

This program is exactly the same as the previous one but extended to have censoring in both tails of the distribution of the dependent variable. In this case, we denote by **uc** the upper (or right) censoring and **lc** the lower (or left) censoring. The algorithm is adapted correspondingly and the same as before is executed.

Table2_AsymptoticStdErrors.do & **Table2_ FILSStdErrors.do**:

This do files derive all the computations required to obtain both the asymptotic standard errors and the FILS standard errors of the estimators when censoring is present. In the first case the asymptotic idea is proxied by setting a very large sample size (N=100000). In both

[1] If the logic of this do file is understood the rest is simpler because we just extend the idea to complex settings where the skeleton is virtually the same.

cases we compute the sandwich formula for the standard errors (that's why you will find Ham and Bread components in the program). Finally there are two ways of checking that things are done correctly – **check1** and **check2**, which are based on the variance components of the variance and covariance matrix being positive. By changing the level of censoring, which is **c**, we get all the rows of **Table 2**.

Table3_FILS.do & **Table3_Tobit.do**:

These two do files carry out the Monte Carlo simulation exercise that include worker dummies with a dependent variable that is right censored. The starting point is the do file Table1_Figure2.do where the time dimension is now included (**T**) to have a panel. In this case, we have a more complex model in which one time-varying explanatory variable is included (**age**). We set this variable to be in the range [18,65] years.

The estimation in this case is as follows: the truncated dependent variable is regressed onto a constant plus age – being **gamma** the coefficient associated to age. Therefore in the case of FILS we do Within Groups – This gives us the output of columns 1 and 3 of **Table 3**. And in the case of Tobit we need to include also the dummies for the individuals, N=1000 – This gives us the output of columns 2 and 4 of **Table 3**. The reason why we have two do separated do files is because we need to know the computation time in each case.

Table4and5_FILS.do & Table4and5_Tobit.do:

These two do files carry out the Monte Carlo simulation experiment that includes both worker and firm dummies with a dependent variable that is right censored. The starting point is the do files Table3_FILS.do and Table3_Tobit.do where firms are now constructed in a proper way (**nfirm** stands for the number of firms[2]). Assortative matching is imposed, which means that high wage workers are assigned to high wage firms. In practice, each worker is associated to an alfa and each firm is associated to a **beta**; the matching is then obvious: high alfas are matched with high betas. Additionally, each time there is a shock that can make individuals change firm – we need some movers to separately identify the parameters of the model. The scalar **tau** in the do file will be the responsible of having more or less percentage change of movers over time.

The estimation in this case is as follows: the truncated dependent variable is regressed onto a constant plus age plus the firm dummies. Therefore in the case of FILS we do Within Groups including firm dummies – This gives us the output of columns 1 and 3 of **Table 4**. And in the case of Tobit we need to include also the dummies for the individuals, N=1000 – This gives us the output of columns 2 and 4 of **Table 4**. Finally, for constructing Table 5 we just need to play with the scalars in the do files (**N, T, nfirm**).

[2] In the paper nfirm is defined as J. We could not name the number of firms as J in the do files because it is a command in mata to generate vectors.

Table6.do:

This do file is pretty similar to the previous ones but instead of setting the scalars in an arbitrary fashion here is calibrated with data from Abowd, Kramarz and Margolis (1999) – see description in the paper for more details. The only important thing that is different, apart from the parameter values, is the matching of workers to firms which in this case is random. We set the probability to remain in the same firm from one period to another equal to 0.55 (**probstay**) and generate a lottery that makes workers change firm (**shockmove**). In this way we can replicate the results from AKM and check whether taking into account the censoring made a difference in terms of consistency of the parameter estimates – see columns 1 and 2 of **Table 6**.

Note: In every do file we use stata and mata at the same time which is more efficient when having big matrices as it is the case here.