**“Feasible Estimation of Robust Standard Errors in the Two-Way fixed effect model”**

This folder contains the Matlab functions necessary to run the algorithm described in the notes. It also contains a matlab script that generates a dataset and runs the code, and a Stata do file that runs the same procedure on the same data using xtreg and adding time dummies.

**Wrapper:**

reg2way.m estimates a 2-way fixed effect model absorbing the two set of dummies and reports standard errors. Syntax: [betaHat,aVarHat,yp,Xp,struc]=xtreg2way(y,X,iid,tid,w,struc,se,noise)

Description of arguments

-y (N-by-1) is the dependent variable

-X (N-by-K) is the matrix of covariates

-iid (N-by-1) is the group id

-tid (N-by-1) is the time id

-w (N-by-1) is a vector of weights. If w is omitted or set to [] then w=1 for all observations.

struc (structure) contains the results of the first step of the algorithm. struc can be omitted or set to [] in which case, it will be automatically calculated. Notice that the first step is computationally expensive. If many regressions are to be run using the same set of covariates it is possible to run the first step only once.

-se (possible values: 0,1,2,11) indicates the standard error estimate to be calculated. se==0 : standard errors assuming homoscedasticity and no within group correlation or serial correlation. se==1 : standard errors proposed by Arellano (1987) robust to heteroscedasticity and serial correlation. se==2 : standard errors robust to heteroscedasticity but assumes no correlation within group or serial correlation. se==11 : Arellano standard errors with a degree of freedom correction performed by Stata xtreg, fe. If se is omitted or set to [] then it is set to 1 and the Arellano (1987) estimator is computed.

-noise (possible values 0,1) If noise==0 results are not displayed. If noise==1 results are displayed. If noise is omitted or set to [], results will be displayed.

Description of output

betaHat (K-by-1) vector of estimated coefficients

aVarHat (K-by-K) estimate of the matrix of variances and covariance of the estimator.

yp (N-by-1) the residual of the projection of y on the two sets of dummies.

Xp (N-by-K) the residual of the projection of each column of X on the two sets of dummies.

struc (structure) results of the first step of the algorithm.

Alternative Usage:

If the first step was already performed and variables yp and Xp are the projected error of the original y and X on the two sets of dummies, then it is possible to avoid repeating the first two steps of the algorithm by calling:

[betaHat,aVarHat]=xtreg2way(yp,Xp,struc)

or

[betaHat,aVarHat]=xtreg2wayPost(yp,Xp,struc,se,noise)

**Matlab Functions:**

* projdummies.m computes the first step of the algorithm. If S=[D,H] where D is the matrix of individual effect dummies and H is the matrix of time dummies. It computes the inverse of (S’S) but returns a structure with the minimal information required to construct it. Syntax: struc=projdummies(hhid,tid,w). hhid is a vector with the individual effect identifier. tid is a vector with the time effect identifier and w is a vector of weights. Observations with weights equal to zero are dropped. The three vectors have to have the same length.
* projvar.m obtains or as described in the second step. It uses the structure generated by projdummies. Syntax: yp=projvar(y,struc), where yp is or a column of ,y is the original variable (or an original column of ).
* regress1.m performs an OLS (third step) and stores a structure with coefficients, residuals and the matrix . Syntax: reg=regress1(yp,Xp), where reg is the structure, yp is the result of projvar(y,invSS) and Xp is the result of stacking horizontally all vectors projvar(x,struc) for each x: column of . The structure reg has three fields: beta, res, and XX which stand for coefficients, residuals and the matrix .
* avar.m computes the asymptotic variance of the estimator (fourth step). Syntax: matCov=avar(X,e,group,J). X is the matrix of covariates: L-by-K. L num of obs, and K num of covariates. e is the vector of residuals: L-by-1; group is the cluster identifier: L-by-1 (it coincides with hhid in step one if standard errors are clustered at the individual level. If g has more or less than L elements it will be assumed that the user required the heteroscedasticity robust estimate. Otherwise, the clustered heteroscedasticity robust estimate is provided. If the matrix X'\*X was calculated before, it can be entered as the fourth argument. J ha to be a K-by-K matrix.

**Scripts:**

* CrossCheckStata.m (matlab script) generates and stores a dataset, and runs the fourth steps above.
* CrossCheckStata.do (stata do file) uses insheet to load the data generated above and runs the panel regression with time dummies. Both outcomes should be identical. The matlab code using the algorithm should be faster and able to handle larger datasets.