Title

ppml_panel_sg - Fast Poisson Pseudo-Maximum Likelihood (PPML) regression for panel "gravity" models with time-varying origin and destination fixed effects and time-invariant pair fixed effects.

Syntax

ppml_panel_sg depvar [indepvars] [if] [in], exporter(*exp_id*) importer(*imp_id*) year(*time_id*) [options]

exp_id, *imp_id*, and *time_id* are variables that respectively identify the origin, destination, and time period associated with each observation.

Description

ppml_panel_sg enables faster computation of the many fixed effects required for panel PPML structural gravity estimation. In particular, it addresses the large number of "pair-wise" FEs needed to consistently identify the effects of time-varying trade policies such as regional trade agreements (see, e.g., Baier & Bergstrand, 2007; Dai, Yotov, & Zylkin, 2014). It also simultaneously absorbs the origin-by-time and destination-by-time FEs implied by theory.

Some options and features of interest:

1. Programmed to run in Mata, making it much faster than existing Stata Poisson commands for estimating the effects of trade policies.
2. Can store the estimated fixed effects in Stata's memory (but as a single column each, rather than as a large matrix with many zeroes)
3. In addition to the pair fixed effects, also readily allows for specifications with additional pair-specific linear time trends.
4. All fixed effects are also allowed to vary by industry, in case you wish to examine industry-level variation in trade flows.
5. Can be used with strictly cross-sectional regressions (i.e., without pair fixed effects).
6. Performs Santos Silva & Tenreyro (2010)'s recommended check for possible non-existence of estimates.

Main Options

nopair	Use origin-time and destination-time fixed effects only (do not include pair fixed effects).
trend	Add linear, pair-specific time trends.
symmetric	Assume pair fixed effects apply symmetrically to flows in both directions, as in, e.g., Anderson & Yotov (2016). Time trends, if specified, will also be symmetric.

industry(*ind_id*) *ind_id* is a varname identifying the industry associated with each observation. When an *ind_id* is provided, the origin-time fixed effects become origin-industry-time effects, the destination-time fixed effects become destination-industry-time effects, and pair-specific terms become origin-destination-industry-specific.

offset(*varname*) Include *varname* as a regressor with coefficient constrained to be equal to 1.

tolerance(#) The default tolerance is 1e-12.

maxiter(#) The default maximum number of iterations is 10,000.

cluster(*cluster_id*) Specifies clustered standard errors, clustered by *cluster_id*. The default is clustering by *exp_id*-*imp_id*[-*ind_id*], unless **nopair** is enabled.

robust Use robust standard errors. This is the default if **nopair** is enabled.

nosterr Do not compute standard errors (saves time if you only care about point estimates).

verbose(#) Show iterative output for every #th iteration. Default is 0 (no output).

Guessing and Storing Values

These options allow you to store results for fixed effects in memory as well as use information from memory to set up initial guesses. You may also utilize the "multilateral resistances" of Anderson & van Wincoop (2003).

olsguess Use reghdfe to initialize guesses for coefficient values.

guessB(*str*) Supply the name of a row vector with guesses for coefficient values.

guessD(*varname*) Guess initial values for the (exponentiated) set of pair fixed effects. Default is all 1s.

guessS(*varname*) Guess initial values for the (exponentiated) set of origin-time fixed effects. Default is the share of depvar within each [*ind_id*-]*time-id* associated with each *exp-id.*

guessM(*varname*) Guess initial values for the (exponentiated) set of destination-time fixed effects. Default is the share of depvar within each [*ind_id*-]*time-id* associated with each *imp-id.*

guessO(*varname*) Guess initial values for the set of "outward" multilateral resistances. Default is all 1s. Overrides **genS**.

guessI(*varname*) Guess initial values for the set of "inward" multilateral resistances. Default is all 1s. Overrides **genM**.

guessTT(*varname*) Guess initial values for pair time trends. These are not exponentiated. Default is all 0s.

genD(newvar), **genS**(newvar), **genM**(newvar), **genO**(newvar), **genI**(newvar), and **genTT**(newvar): These options store fixed effects and/or time trend parameters in memory as new variables.

Background

As a typical application, consider the following PPML regression:

$$X_{ijt} = \exp\left[\ln S_{it} + \ln M_{jt} + \ln D_{ij} + b \times RTA_{ijt}\right] + e_{ijt}. \tag{1}$$

X_{ijt} are international trade flows. i, j, and t are indices for origin, destination, and time. The goal is to consistently estimate the average effect of RTA_{ijt}, a dummy variable for the presence of a regional trade agreement on trade flows, using a "structural gravity" specification. The origin-time and destination-time fixed effects—S_{it} and M_{jt}—ensure the theoretical restrictions implied by structural gravity are satisfied. The pair fixed effect—D_{ij}—then absorbs all time-invariant pair characteristics that may be correlated with the likelihood of forming an RTA.

Computationally, the biggest obstacle to estimating (1) is the pair fixed effect term D_{ij}. Because a unique D_{ij} must be computed for each pair, the number of D_{ij}'s increases rapidly with the number of locations. For a balanced international trade data set with 75 countries trading with each other over 10 years (not an especially large sample for trade data), there will be on the order of $75^2 = 5,625$ pair fixed effects that must be computed. In addition (ignoring collinearity), we will also require $75 \times 2 \times 10 = 1,500$ origin-time and destination-time effects. The total number of parameters needed to estimate (1) (around $7,000$) would normally require a long computing time in Stata, likely several hours at least. If we push the number of locations and/or years further, we will quickly approach Stata's matsize limits, beyond which estimation becomes infeasible.

To date, this is the only available Stata command that will perform "fast" estimation of specifications such as (1) using PPML. It works by manipulating the first order conditions of the Poisson to produce analytical expressions for each of the fixed effects that can be computed via simple iteration. In this way, it both adapts and extends existing procedures described in Guimarães & Portugal (2010) and Figueiredo, Guimarães, & Woodward (2015) for estimating Poisson models with high dimensional fixed effects. These works and others are recommended below for further reading.

Examples

To perform a basic panel estimation such as (1):

```
ppml_panel_sg trade rta, ex(iso_o) im(iso_d) y(year)
```

To add pair-specific time trends, i.e.,

$$X_{ijt} = \exp\left[\ln S_{it} + \ln M_{jt} + \ln D_{ij} + a_{ij} \times t + b \times RTA_{ijt}\right] + e_{ijt}, \tag{2}$$

you would input:

```
ppml_panel_sg trade rta, ex(iso_o) im(iso_d) y(year) trend
```

If you want your pair fixed effects to be symmetric (i.e., $D_{ij} = D_{ji}$), the syntax is:

```
ppml_panel_sg trade rta, ex(iso_o) im(iso_d) y(year) sym
```

To estimate coefficients of more traditional, time-invariant gravity variables, such as bilateral distance, use the **nopair** option:

```
ppml_panel_sg trade ln_dist colony language contiguity rta,
ex(iso_o) im(iso_d) y(year) nopair
```

Unlike the regressions with pair fixed effects, however, obtaining estimates for time-invariant regressors may not be noticeably faster than existing methods (e.g., glm, ppml) unless the number of origin-time and destination-time effects is sufficiently large. You may also exclude the year ID in this last specification if your data includes only 1 year.

Advisory

This estimation command is strictly intended for settings where the dependent variable is spatial flows from one set of locations to another (such as international trade or migration flows). It is not a generalized Poisson fixed effects command. For more general problems that require Poisson estimation, you may try: poisson, glm, ppml, xtpoisson, xtpqml, and/or poi2hdfe. For an OLS command that can compute similar "gravity" specifications using OLS, I recommend reghdfe.

As noted above, a useful feature of this command is that it will automatically drop any of your main covariates which do not satisfy the condition for guaranteeing the existence of estimates described in Santos Silva & Tenreyro (2010). This should ensure convergence in most cases. However, you may still encounter convergence issues in cases when linear time trends are specified and when the data contains many zeroes. Future versions of this command will seek to address this latter issue.

This is version 1.0 of this command. If you believe you have found an error that can be replicated, or have other suggestions for improvements, please feel free to contact me.

Acknowledgements

I have adapted parts of my code from several other related commands. These include: poi2hdfe, by Paulo Guimarães, SILS, by Keith Head and Thierry Mayer, and reghdfe by Sergio Correia. I give the utmost credit to each of these authors for creating these programs and making their code available.

I also thank Yoto Yotov, Davin Chor, Sergio Correia, Paulo Guimarães, and João Santos Silva for kindly taking the time to offer feedback and suggest improvements.

Further Reading

- Structural gravity: Anderson & van Wincoop (2003); Head & Mayer (2014)
- On the use of PPML to estimate gravity equations: Santos Silva & Tenreyro (2006, 2011); Arvis & Shepherd (2013); Egger & Staub (2014); Fally (2015)

- Possible non-existence of Poisson MLE estimates: Santos Silva & Tenreyro (2010)
- Consistently estimating the effects of trade policies: Baier & Bergstrand (2007); Dai, Yotov, & Zylkin (2014); Anderson & Yotov (2016); Piermartini & Yotov (2016)
- Estimating models with high dimensional fixed effects: Guimarães & Portugal (2010); Gaure (2011); Figueiredo, Guimarães, & Woodward (2015); Correia (2016)
- Multi-variate Steffenson's acceleration (used to accelerate convergence): Nievergelt (1991)

References

Anderson, J. E. & van Wincoop, E. (2003), "Gravity with Gravitas: A Solution to the Border Puzzle", *American Economic Review* **93**(1), 170–192.

Anderson, J. E. & Yotov, Y. V. (2016), "Terms of Trade and Global Efficiency Effects of Free Trade Agreements, 1990-2002", *Journal of International Economics* **(in press)**.

Arvis, J.-F. & Shepherd, B. (2013), "The Poisson quasi-maximum likelihood estimator: a solution to the 'adding up' problem in gravity models", *Applied Economics Letters* **20**(6), 515–519.

Baier, S. L. & Bergstrand, J. H. (2007), "Do free trade agreements actually increase members' international trade?", *Journal of International Economics* **71**(1), 72–95.

Correia, S. (2016), "REGHDFE: Stata module to perform linear or instrumental-variable regression absorbing any number of high-dimensional fixed effects", .

Dai, M., Yotov, Y. V., & Zylkin, T. (2014), "On the trade-diversion effects of free trade agreements", *Economics Letters* **122**(2), 321–325.

Egger, P. H. & Staub, K. E. (2014), "GLM estimation of trade gravity models with fixed effects", Mimeo, ETH Zurich.

Fally, T. (2015), "Structural gravity and fixed effects", *Journal of International Economics* **97**(1), 76–85.

Figueiredo, O., Guimarães, P., & Woodward, D. (2015), "Industry localization, distance decay, and knowledge spillovers: Following the patent paper trail", *Journal of Urban Economics* **89**(C), 21–31.

Gaure, S. (2011), "OLS with Multiple High Dimensional Category Dummies", Memorandum 14/2010, Oslo University, Department of Economics.

Guimarães, P. & Portugal, P. (2010), "A simple feasible procedure to fit models with high-dimensional fixed effects", *Stata Journal* **10**(4), 628–649.

Head, K. & Mayer, T. (2014), "Gravity equations: Workhorse, toolkit, and cookbook", *Handbook of International Economics* **4**, 131–196.

Nievergelt, Y. (1991), "Aitken's and Steffensen's accelerations in several variables", *Numerische Mathematik* **59**(1), 295–310.

Piermartini, R. & Yotov, Y. (2016), "Estimating Trade Policy Effects with Structural Gravity", School of Economics Working Paper Series 2016-10, LeBow College of Business, Drexel University.

Santos Silva, J. M. C. & Tenreyro, S. (2006), "The Log of Gravity", *Review of Economics and Statistics* **88**(4), 641–658.

Santos Silva, J. M. C. & Tenreyro, S. (2010), "On the existence of the maximum likelihood estimates in Poisson regression", *Economics Letters* **107**(2), 310–312.

Santos Silva, J. M. C. & Tenreyro, S. (2011), "Further simulation evidence on the performance of the Poisson pseudo-maximum likelihood estimator", *Economics Letters* **112**(2), 220–222.