Two-way exponential-regression models

`twexp` and `twgravity`

Koen Jochmans and Vincenzo Verardi

University of Cambridge and Université de Namur

Setup

Double-indexed data $(y_{ij}, \boldsymbol{x}_{ij})$ of size $n \times m$.

Two-way fixed-effect model for non-negative outcomes:

$$y_{ij} = \exp(\alpha_i + \beta_j + \boldsymbol{x}_{ij}^\top \boldsymbol{\gamma})\, \varepsilon_{ij}, \qquad \mathbb{E}(\varepsilon_{ij} | \boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{nm}) = 1.$$

Object of interest is slope vector $\boldsymbol{\gamma}$.

Examples:

patent applications (or alike) with firm heterogeneity and aggregate time effects (common technological progress).

Trade volume with both importer and exporter heterogeneity (and other constant-elasticity models).

Focus on data sets where n, m are both 'large' — incidental parameters in both directions.

This covers large panels as well as cross-sections on bi-lateral interactions.

Differencing-out the nuisance parameters

Note that

$$\mathbb{E}\left(\frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\gamma})}\middle|\boldsymbol{x}_{11},\ldots,\boldsymbol{x}_{nm}\right)=\exp(\alpha_i+\beta_j)$$

for all (i,j).

Thus, when errors are uncorrelated,

$$\begin{aligned}\mathbb{E}\left(\frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^\top\boldsymbol{\gamma})}\frac{y_{i'j'}}{\exp(\boldsymbol{x}_{i'j'}^\top\boldsymbol{\gamma})}\middle|\boldsymbol{x}_{11},\ldots,\boldsymbol{x}_{nm}\right)&=\exp(\alpha_i+\beta_j)\exp(\alpha_{i'}+\beta_{j'})\\&=\exp(\alpha_i+\alpha_{i'}+\beta_j+\beta_{j'}),\end{aligned}$$

and

$$\begin{aligned}\mathbb{E}\left(\frac{y_{i'j}}{\exp(\boldsymbol{x}_{i'j}^\top\boldsymbol{\gamma})}\frac{y_{ij'}}{\exp(\boldsymbol{x}_{ij'}^\top\boldsymbol{\gamma})}\middle|\boldsymbol{x}_{11},\ldots,\boldsymbol{x}_{nm}\right)&=\exp(\alpha_{i'}+\beta_j)\exp(\alpha_i+\beta_{j'})\\&=\exp(\alpha_i+\alpha_{i'}+\beta_j+\beta_{j'}),\end{aligned}$$

for all i,i' and j,j'.

Consequently,

$$\mathbb{E}\left(\frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\gamma})} \frac{y_{i'j'}}{\exp(\boldsymbol{x}_{i'j'}^{\top}\boldsymbol{\gamma})} - \frac{y_{ij'}}{\exp(\boldsymbol{x}_{ij'}^{\top}\boldsymbol{\gamma})} \frac{y_{i'j}}{\exp(\boldsymbol{x}_{i'j}^{\top}\boldsymbol{\gamma})} \middle| \boldsymbol{x}_{11}, \ldots, \boldsymbol{x}_{nm} \right) = 0,$$

for all i, i' and j, j'.

This implies unconditional moments that can be used in a GMM framework.

See [Charbonneau, 2013] and [Jochmans, 2017].

This differencing approach is the two-way generalization of [Chamberlain, 1992].

In the one-way case, this nests 'pseudo-poisson' but in the two-way case, it does not.

Implications

Inference on γ can be separated from estimation of the incidental parameters.

Moment conditions are exactly unbiased and fixed in number.

This is not so for pseudo-poisson:

High-dimensional problem; [Guimarães, 2016], [Correia et al., 2019].

Estimated fixed effects introduce bias in standard errors; [Jochmans, 2017], [Pfaffermayer, 2019].

If the panel were short clustering could be used to obtain (conservative) standard errors. No such theory here.

Bootstrap/jackknife standard errors equally unavailable.

GMM1 and GMM2

Our first estimator, GMM1, solves

$$\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{m}\sum_{j'=1}^{m} \boldsymbol{x}_{ij}\left\{\frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\gamma})}\frac{y_{i'j'}}{\exp(\boldsymbol{x}_{i'j'}^{\top}\boldsymbol{\gamma})} - \frac{y_{ij'}}{\exp(\boldsymbol{x}_{ij'}^{\top}\boldsymbol{\gamma})}\frac{y_{i'j}}{\exp(\boldsymbol{x}_{i'j}^{\top}\boldsymbol{\gamma})}\right\} = \mathbf{0}.$$

Do not compute this by brute force but exploit the representation

$$\sum_{i=1}^{n}\sum_{j=1}^{m} \boldsymbol{x}_{ij}\left\{u_{ij}\overline{u} - \overline{u}_{i\cdot}\,\overline{u}_{\cdot j}\right\}, \qquad u_{ij} := \frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^{\top}\boldsymbol{\gamma})},$$

where bars indicate sample averages.

This is immediate in `Mata`.

Similar ‘tricks’ can be used for calculating the covariance matrix.

Details on covariance matrix are in [Jochmans, 2017].

Our second estimator, GMM2, solves

$$\sum_{i=1}^{n}\sum_{i'=1}^{n}\sum_{j=1}^{m}\sum_{j'=1}^{m} \tilde{\boldsymbol{x}}_{ij} \left\{ \frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\gamma})} \frac{y_{i'j'}}{\exp(\boldsymbol{x}_{i'j'}^\top \boldsymbol{\gamma})} - \frac{y_{ij'}}{\exp(\boldsymbol{x}_{ij'}^\top \boldsymbol{\gamma})} \frac{y_{i'j}}{\exp(\boldsymbol{x}_{i'j}^\top \boldsymbol{\gamma})} \right\} = \mathbf{0}$$

for

$$\tilde{\boldsymbol{x}}_{ij} := \frac{\boldsymbol{x}_{ij}}{\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\gamma})\ \exp(\boldsymbol{x}_{i'j'}^\top \boldsymbol{\gamma})\ \exp(\boldsymbol{x}_{i'j}^\top \boldsymbol{\gamma})\ \exp(\boldsymbol{x}_{ij'}^\top \boldsymbol{\gamma})}$$

(with some abuse of notation).

Computational burden is again non-existent.

Behaves similar to pseudo-poisson but with more accurate standard errors.

twexp

The command **twexp** is designed for (balanced) $n \times m$ panel data sets.

twexp *depvar* [*indepvars*] , **indn**(varname) **indm**(varname) **model**(option) **init**(vec)

indn(varname) declares the cross-sectional dimension of the panel.

indm(varname) declares the time-series dimension of the panel.

model(option) determines whether $GMM1$ or $GMM2$ is implemented.

init(vec) specifies the starting value for the numerical optimization.

```
ssc install twexp
```

twgravity

The command **twgravity** is designed for a cross-sectional data on dyadic interactions between n agents.

Agents do not interact with themselves, so y_{ii} and $\boldsymbol{x}_{ii}$ are not defined.

The syntax is the same as for **twexp**.

twgravity *depvar* [*indepvars*] , **indn**(varname) **indm**(varname) **model**(option) **init**(vec)

indn(varname) identifies the first agent in the dyad.

indm(varname) identifies the second agent in the dyad.

model(option) determines whether $GMM1$ or $GMM2$ is implemented.

init(vec) specifies the starting value for the numerical optimization.

```
ssc install twgravity
```

Trade illustration

Country-level trade data from

`http://personal.lse.ac.uk/tenreyro/lgw.html`

[Santos Silva and Tenreyro, 2006].

Descriptive statistics:

Variable	Obs	Mean	Std. Dev.	Min	Max
trade	18,360	172129.5	1829058	0	1.01e+08
ldist	18,360	8.785508	.7416775	4.876723	9.898691
border	18,360	.0196078	.1386522	0	1
comlang	18,360	.209695	.407102	0	1
colony	18,360	.1704793	.3760636	0	1
comfrt_wto	18,360	.0250545	.1562948	0	1

`twgravity trade ldist border comlang colony comfrt_wto`, indn(s2_ex) indm(s1_im) model(GMM1)

Number of obs = 18360

trade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ldist	-.8165761	.188396	-4.33	0.000	-1.185825	-.4473268
border	.4873677	.2339165	2.08	0.037	.0288999	.9458356
comlang	.2594789	.2119004	1.22	0.221	-.1558382	.674796
colony	.1648687	.1955009	0.84	0.399	-.2183059	.5480433
comfrt_wto	.3064196	.217326	1.41	0.159	-.1195316	.7323707

Completes in .81 seconds on my laptop.

Poisson takes 16 seconds, 3.87 seconds, or 1.65 seconds, depending on the routine used.

`twgravity trade ldist border comlang colony comfrt_wto`, indn(s2_ex) indm(s1_im) model(GMM2)

Number of obs = 18360

trade	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
ldist	-.7509313	.0567805	-13.23	0.000	-.8622191	-.6396436
border	.1490604	.0771748	1.93	0.053	-.0021994	.3003202
comlang	.4909294	.0929732	5.28	0.000	.3087052	.6731536
colony	.2128996	.1212684	1.76	0.079	-.0247821	.4505813
comfrt_wto	.3298556	.1249293	2.64	0.008	.0849987	.5747126

Completes in 1.85 seconds on my laptop.

Patents and R&D illustration

Panel data on 346 firms from 1970–1979, taken from

`http://cameron.econ.ucdavis.edu/mmabook/mmaprograms.html`,

[Hall et al., 1986].

Descriptive statistics:

Variable	Obs	Mean	Std. Dev.	Min	Max
PAT	3,460	36.28439	74.46563	0	608
LOGR	3,460	1.229807	1.970524	-3.84868	7.06524

Include fixed effects to control for firm heterogeneity and time effects for common technological progress and other macro shocks.

```
twexp PAT LOGR, indn(id) indm(year) model(GMM1)
```

Number of obs = 3460

PAT	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LOGR	.4084421	.0457615	8.93	0.000	.3187521	.498133

```
matrix start = e(b)

twexp PAT LOGR, indn(id) indm(year) model(GMM2) init(start)
```

Number of obs = 3460

PAT	Coef.	Std. Err.	z	P>|z|	[95% Conf.	Interval]
LOGR	.3241356	.0635514	5.10	0.000	.1995772	.448694

Monte Carlo

No fixed effects,

Two binary regressors with unit coefficients,

Success probabilities are .05 and .50, respectively,

Independent log-normal errors,

Sample size is $n = 25$.

Ratio of average standard error to Monte Carlo standard deviation:

GMM1: .8654 and 1.0145,

GMM2: .8457 and 1.0319,

PMLE: .6761 and 0.9125.

Variable	Obs	Mean	Std. Dev.	Min	Max
BGMM11	5,000	.9542519	.3584049	-.2982407	2.925951
BGMM12	5,000	1.002699	.109982	.5822676	1.549646
BGMM21	5,000	.9396433	.3955723	-.3508639	3.290722
BGMM22	5,000	.9997944	.1121814	.5487244	1.508134
BPPML1	5,000	.940787	.3754578	-.3382381	2.783273
BPPML2	5,000	1.002575	.1124283	.5688691	1.547408

Variable	Obs	Mean	Std. Dev.	Min	Max
SEGMM11	5,000	.310138	.0805269	.1471121	.7484761
SEGMM12	5,000	.1115835	.0143905	.0828566	.2427083
SEGMM21	5,000	.3345285	.0903741	.1373527	.8006971
SEGMM22	5,000	.1157641	.0168373	.0827409	.4340205
SEPPML1	5,000	.2538752	.0547559	.1251421	.5346598
SEPPML2	5,000	.1025859	.0128109	.0756624	.2152773

Extensions: Overidentification

For now the code deals with the 'just-identified' setting, where the number of moments is equal to the number of parameters.

Overidentification is easily dealt with but not (yet) implemented.

This is useful for:

Approximating 'optimal' unconditional moments,

Instrumental-variable problems.

Extensions: Instrumental variable model

The approach extends straightforwardly to

$$y_{ij} = \exp(\alpha_i + \beta_j + \boldsymbol{x}_{ij}^\top \boldsymbol{\gamma})\, \varepsilon_{ij}, \qquad \mathbb{E}(\varepsilon_{ij} | \boldsymbol{z}_{11}, \ldots, \boldsymbol{z}_{nm}) = 1,$$

where $\boldsymbol{z}_{ij}$ are instrumental variables.

In the same way as before we get

$$\mathbb{E}\left(\frac{y_{ij}}{\exp(\boldsymbol{x}_{ij}^\top \boldsymbol{\gamma})} \frac{y_{i'j'}}{\exp(\boldsymbol{x}_{i'j'}^\top \boldsymbol{\gamma})} - \frac{y_{ij'}}{\exp(\boldsymbol{x}_{ij'}^\top \boldsymbol{\gamma})} \frac{y_{i'j}}{\exp(\boldsymbol{x}_{i'j}^\top \boldsymbol{\gamma})} \middle| \boldsymbol{z}_{11}, \ldots, \boldsymbol{z}_{nm} \right) = 0,$$

for all i, i' and j, j'.

An example in the trade context would be the potential endogeneity of trade agreements; [Egger et al., 2011].

Chamberlain, G. (1992).
Comment: Sequential moment restrictions in panel data.
Journal of Business & Economic Statistics, 10:20–26.

Charbonneau, K. B. (2013).
Multiple fixed effects in theoretical and applied econometrics.
PhD thesis, Princeton University.

Correia, S., Guimarães, P., and Zylkin, T. (2019).
PPMLHDFE: Stata module for Poisson pseudo-likelihood regression with multiple levels of fixed effects.
Mimeo.

Egger, P. H., Larch, M., Staub, K. E., and Winkelmann, R. (2011).
The trade effects of endogenous preferential trade agreements.
American Economic Journal: Economic Policy, 3:113–143.

Guimarães, P. (2016).
POI2HDFE: Stata module to estimate a Poisson regression with high-dimensional fixed effects.
Mimeo.

Hall, B., Griliches, Z., and Hausman, J. (1986).
Patents and r&D: Is there a lag?
International Economic Review, 27:265–283.

Jochmans, K. (2017).
Two-way models for gravity.
Review of Economics and Statistics, 99:478–485.

Pfaffermayer, M. (2019).
Gravity models, PPML estimation and the bias of the robust standard errors.
Forthcoming in *Applied Economics Letters*.

Santos Silva, J. M. C. and Tenreyro, S. (2006).
The log of gravity.
Review of Economics and Statistics, 88:641–658.