Help file for ppml_fe_bias

Title

`ppml_fe_bias` - Bias corrections for Poisson Pseudo-Maximum Likelihood (PPML) gravity models with two-way and three-way fixed effects.

Syntax

`ppml_fe_bias` depvar indepvars [if] [in], lambda(*varname*) i(*exp_id*) j(*imp_id*) t(*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. "lambda" is an input for the conditional mean of each observation. For more details, please see the "Background" section below.

Description

`ppml_fe_bias` implements analytical bias corrections described in Weidner & Zylkin (2020) for PPML "gravity" regressions with two-way and three-way fixed effects, as are commonly used with international trade data and other types of spatial flows. As shown in Weidner & Zylkin (2020), when the time dimension is fixed, the point estimates produced by the three-way PPML gravity model have an asymptotic incidental parameter bias of order $1/N$, where N is the number of countries, and the cluster-robust sandwich estimator that is typically used for inference itself has a downward bias that is also of order $1/N$. For the two-way PPML gravity model, only the standard errors are biased.

Main Options

These options allow you to store results for the bias corrections, select the type of gravity model being used (two-way or three-way), and stipulate whether an approximation should be used for the bias-corrected variance matrix.

bias(*name***)**	Store bias corrections for coefficients as a Stata matrix.
v(*name***)**	Store bias-corrected variance matrix as a Stata matrix.
w(*name***)**	Store the expected Hessian matrix ("W") as a Stata matrix.
b_term(*name***)**	Store the estimated "B"-component of bias in the score for the three-way model that is associated with the origin-time fixed effects.
d_term(*name***)**	Store the estimated "D"-component of bias in the score for the three-way model that is associated with the destination-time fixed effects.
beta(*name***)**	Pass the coefficients from a prior PPML estimation. When beta coefficients are provided, a new results table will be produced complete with bias-corrected coefficients and standard errors.

twoway Specifies that the estimated model is a two-way model with only origin(-time) and destination(-time) fixed effects. In this case, `ppml_fe_bias` will only compute a bias correction for the estimated variance.

nosterr Don't compute the bias correction for the variance.

approx If "approx" is enabled, the bias correction for the variance will be computed using an approximation. By default, this approximation is used whenever the number of origin-time and destination-time fixed effects exceeds 1000 in order to facilitate computation and to avoid running up against memory constraints. This approximation is only used with three-way models.

exact Use an exact method for computing the bias-corrected variance, even if the number of origin-time and destination-time fixed effects exceeds 1000.

notable Suppress results table.

Background

PPML gravity models with two-way and three-way fixed effects are very popular in the study of international trade and other similar applications that involve bilateral flows (such as urban commuting or interregional migration). The two-way PPML gravity model may be written as

$$y_{ijt} = \exp\left[\alpha_{it} + \gamma_{jt} + x'_{ijt}\beta\right]\omega_{ijt}, \tag{1}$$

where y_{ijt} is a flow from origin i to destination j at time t, α_{it} and γ_{jt} respectively are origin-time and destination-time fixed effects, β are the coefficients we want to estimate (typically variables that influence bilateral frictions, such as the distance between i and j), and $\omega_{ijt} \geq 0$ serves as an error term. The three-way model then adds a "country-pair" fixed effect, η_{ij}:

$$y_{ijt} = \exp\left[\alpha_{it} + \gamma_{jt} + \eta_{ij} + x_{ijt}\beta\right]\omega_{ijt}. \tag{2}$$

In the latter model, the advantage of the third fixed effect η_{ij} is that it absorbs all time-invariant determinants of flows between i and j. As discussed in Baier & Bergstrand (2007), the use of these fixed effects therefore allows us to identify the elements of β based on time-variation in trade within pairs after first conditioning on α_{it} and γ_{jt}. This approach is especially well suited for estimating the effects of bilateral trade agreements and other similar policy variables and is currently recommended for this purpose by several leading references on gravity estimation (e.g., Head & Mayer, 2014; Yotov, Piermartini, Monteiro, & Larch, 2016).

Weidner & Zylkin (2020)'s analysis identifies several econometric issues that arise with the three-way model, but their method for correcting the standard errors of the three-way model also can be adapted to address a similar issue with the two-way model. Focusing on the three-way model for now, notice that the first-order conditions of PPML allow us to "profile out" the pair fixed effect η_{ij} using $\exp(\eta_{ij}) = \sum_t y_{ijt} / \sum_t \exp\left[\alpha_{it} + \gamma_{jt} + x_{ijt}\beta\right]$. Applying this substitution then respectively gives us the following modified first-order conditions for the remaining parameters α_{it}, γ_{jt}, and β:

$$\sum_j \left[y_{it} - \frac{\mu_{ijt}}{\sum_{s=1}^T \mu_{ijs}} \sum_{s=1}^T y_{is} \right] = 0, \quad \sum_i \left[y_{it} - \frac{\mu_{ijt}}{\sum_{s=1}^T \mu_{ijs}} \sum_{s=1}^T y_{is} \right] = 0, \quad \sum_{i,j,t} x_{ijt} \left[y_{it} - \frac{\mu_{ijt}}{\sum_{s=1}^T \mu_{ijs}} \sum_{s=1}^T y_{is} \right] = 0,$$

where $\mu_{ijt} := \exp\left[\alpha_{it} + \gamma_{jt} + x_{ijt}\beta\right]$. In this way, the three-way FE-PPML is effectively re-expressed as a two-way FE-Multinomial model, such that we can be assured that the otherwise-large number of fixed effects does not pose an issue for the consistency of our estimate for β. This is an important observation because the typical alternatives to PPML for gravity estimation (e.g., OLS, Gamma PML) can be shown to be inconsistent in this setting. At the same time, the two-way representation of the model also brings to mind the results from Fernández-Val & Weidner (2016) for the asymptotic bias of two-way nonlinear models due to the incidental parameter problem. The three-way PPML gravity model is a more complicated model than the ones studied in Fernández-Val & Weidner (2016), but Weidner & Zylkin (2020) nonetheless demonstrate that an analogous result occurs for three-way PPML whenever the time dimension is fixed. Specifically, if I is the number of origins and J is the number of destinations, then β will have an asymptotic bias of the form

$$\frac{1}{I} B^\beta + \frac{1}{J} D^\beta,$$

that is, an asymptotic bias that vanishes only as both I and $J \to \infty$. Because the asymptotic standard error itself shrinks with $1/\sqrt{IJ}$ as I and J become large, we have the discomfiting result that uncorrected confidence intervals will be systematically off-center even in moderately large samples because of the large relative magnitude of the bias in relation to the standard error.

Weidner & Zylkin (2020) provide a more detailed derivation of the bias in β based on a second-order Taylor expansion of the expected profile score around the correct values of the α- and γ-parameters. This expansion also provides the basis for the analytical bias correction performed by this command. In addition to this correction, `ppml_fe_bias` also addresses a related issue that affects the cluster-robust standard errors that are typically used with the three-way model. The latter problem is effectively a version of the more general result that "heteroscedasticity-robust" standard errors tend to be downward-biased in small samples (cf., MacKinnon & White, 1985; Imbens & Kolesar, 2016), only this problem is exacerbated here by the slow convergence of the fixed effects, which causes the bias in the standard error to vanish at only the relatively slow rate of $1/\sqrt{I} \vee 1/\sqrt{J}$ as I and J become large. A similar issue also arises in the two-way PPML model (Egger & Staub, 2015; Jochmans, 2017; Pfaffermayr, 2019); thus, `ppml_fe_bias` has been programmed to provide bias-corrected standard errors for two-way models as well as three-way models (by making use of the "`twoway`" option in the former case). The β-coefficients obtained using the two-way PPML gravity model do not suffer from any asymptotic bias, however, as originally shown by Fernández-Val & Weidner (2016).

Examples

These examples follow the sample .do file included along with this command. The data set used in this .do file consists of a panel of 65 countries trading with one another over the years 1988-2004, using every 4 years. The trade data uses aggregate manufacturing trade flows from UN COMTRADE, with information on FTAs taken from the NSF-Kellogg database maintained by Scott Baier and Jeff Bergstrand and other covariates taken from the CEPII gravity data set created by Head, Mayer, & Ries (2010). The computation of the PPML regression relies on the ppmlhdfe Stata command created by Correia, Guimarães, & Zylkin (2020).

Three-way example. Using ppmlhdfe, the appropriate syntax for specifying a three-way gravity model with exporter-time, importer-time, and exporter-importer fixed effects and standard errors that are clustered by pair would be

```
ppmlhdfe trade fta, a(imp#year exp#year imp#exp) cluster(imp#exp) d
```

The "d" option is needed to facilitate obtaining values for the conditional mean of the dependent variable, which `ppml_fe_bias` will need in order to construct the necessary expressions for the bias corrections. We next create a new variable "lambda" containing the conditional mean from the regression as well as a matrix "beta" for the estimated coefficient on `fta`. We then pass these results along with the data to `ppml_fe_bias`:

```
predict lambda
matrix beta = e(b)
ppml_fe_bias trade fta, i(expcode) j(impcode) t(year) lambda(lambda) beta(beta)
```

The resulting output shows there is an upward bias of about 0.008 in the estimated PPML coefficient for `fta`. While the bias-corrected estimate is not overly different from the original uncorrected one (0.170 vs 0.178), it is important to keep in mind this bias is about 22% of the estimated standard error, which is easily large enough to make a meaningful difference for hypothesis testing in general cases. Consistent with the results of Weidner & Zylkin (2020), the estimated standard error is also found to be biased: the bias-corrected standard error is itself about 20% larger than the uncorrected standard error, and the lower bound of the estimated 95% confidence interval after both of these corrections are applied is substantially lower than what would be found otherwise (0.086 versus 0.109). While these results were obtained for a moderate number of countries ($I = J = 65$), it is important to note that Weidner & Zylkin (2020)'s results imply that the magnitudes of these biases are likely to depend on the distribution of the data, even for ostensibly large data sets. Thus, it is recommended that researchers implement these checks whenever feasible as a matter of good practice.

Two-way example. Using the same example data set and .do file, a typical two-way gravity regression would be

```
ppmlhdfe trade ln_distw contig colony comlang_off comleg fta, ///
a(imp#year exp#year) cluster(imp#exp) d
```

where we now include some covariates that would otherwise be absorbed by the exporter-importer fixed effect from the three-way model, such as the log of bilateral distance and the sharing of a colonial relationship. The code to compute bias-corrected standard errors and to present the results in a nicely formatted table would be

```
predict lambda_2way
matrix beta_2way = e(b)
ppml_fe_bias trade ln_distw contig colony comlang_off comleg fta, ///
i(expcode) j(impcode) t(year) lambda(lambda_2way) beta(beta_2way) twoway
```

As the results show, the implied downward biases in the standard error for the two-way model are similar to the corresponding bias found for `fta` coefficient in the three-way model, generally ranging between 20% and 26%. As shown by Fernández-Val & Weidner (2016), the coefficients themselves are unbiased in this case.

Storing results. So long as an input is given for the "beta" matrix, `ppml_fe_bias` will store results for the bias-corrected coefficients and variance matrix using `ereturn post`. This allows users to access these results using the post-estimation operators `e(b)` and `e(V)`. It also makes it possible to use `ppml_fe_bias` in conjunction with standard table-formatting commands such as `estout` or `estimates table`.

Approximation for the adjusted variance. Depending on the size of the data, it may be necessary to use an approximation method to compute the necessary bias correction for the cluster-robust variance matrix in the three-way model. Because the details behind this approximation are mathematically complex, they have been left for the end of this document.

Advisory

This is an advanced technique that requires a basic understanding of PPML estimation and of the three-way gravity model. I would recommend either Yotov, Piermartini, Monteiro, & Larch (2016) or Larch, Wanner, Yotov, & Zylkin (2019) for further reading on these topics. For essential reading on two-way gravity estimation, see Head & Mayer (2014).

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.

Suggested citation

If you are using this command in your research, I would appreciate if you would cite

- Weidner, Martin and Thomas Zylkin. "Bias and Consistency in Three-way Gravity Models." *arXiv preprint arXiv:1909.01327* (2019).

The code used in this command implements the bias corrections described in Section 3 of our paper for the three-way model. The bias correction for the variance of the two-way model is discussed in the appendix. Depending on interest, future versions of this command could add further options for multi-way clustering, multi-industry models, and/or dynamic models.

Acknowledgements

The computations needed to construct these bias corrections have been greatly facilitated by three outstanding Stata packages: rowmat_utils by Matthew J. Baker, gtools by Mauricio Caceres Bravo, and hdfe by Sergio Correia. The code for displaying results uses the frmttable command by John Luke Gallup. Thanks to each of these creators for these incredibly useful tools!

Further Reading

- Gravity estimation: Head & Mayer (2014); Yotov, Piermartini, Monteiro, & Larch (2016)
- Bias corrections for two-way models: Fernández-Val & Weidner (2016)
- Bias corrections for other three-way models aside from PPML: Fernández-Val & Weidner (2018); Hinz, Stammann, & Wanner (2019)
- Fast methods for computing PPML models with multiple levels of fixed effects: Correia, Guimarães, & Zylkin (2020); Larch, Wanner, Yotov, & Zylkin (2019); Stammann (2018)

References

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., Guimarães, P., & Zylkin, T. (2020), "Fast Poisson Estimation with High-dimensional Fixed Effects", *Stata Journal* **20**(1), 95–115.

Egger, P. H. & Staub, K. E. (2015), "GLM estimation of trade gravity models with fixed effects", *Empirical Economics* **50**(1), 137–175.

Fernández-Val, I. & Weidner, M. (2016), "Individual and Time Effects in Nonlinear Panel Models with Large N, T", *Journal of Econometrics* **192**(1), 291–312.

Fernández-Val, I. & Weidner, M. (2018), "Fixed Effect Estimation of Large T Panel Data Models", *Annual Review of Economics* .

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

Head, K., Mayer, T., & Ries, J. (2010), "The erosion of colonial trade linkages after independence", *Journal of International Economics* **81**(1), 1–14.

Hinz, J., Stammann, A., & Wanner, J. (2019), "Persistent Zeros: The Extensive Margin of Trade", *Unpublished manuscript* .

Imbens, G. W. & Kolesar, M. (2016), "Robust standard errors in small samples: Some practical advice", *Review of Economics and Statistics* **98**(4), 701–712.

Jochmans, K. (2017), "Two-Way Models for Gravity", *Review of Economics and Statistics* **99**(3), 478–485.

Kline, P., Saggio, R., & Sølvsten, M. (forthcoming), "Leave-out estimation of variance components", *Econometrica* .

Larch, M., Wanner, J., Yotov, Y. V., & Zylkin, T. (2019), "Currency Unions and Trade: A PPML Re-assessment with High-dimensional Fixed Effects", *Oxford Bulletin of Economics and Statistics* **81**(3), 487–510.

MacKinnon, J. G. & White, H. (1985), "Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties", *Journal of econometrics* **29**(3), 305–325.

Pfaffermayr, M. (2019), "Gravity models, PPML estimation and the bias of the robust standard errors", *Applied Economics Letters* pp. 1–5.

Stammann, A. (2018), "Fast and feasible estimation of generalized linear models with high-dimensional k-way fixed effects", *arXiv preprint arXiv:1707.01815* .

Weidner, M. & Zylkin, T. (2020), "Bias and Consistency in Three-Way Gravity Models", *arXiv preprint arXiv:1909.01327* .

Yotov, Y. V., Piermartini, R., Monteiro, J.-A., & Larch, M. (2016), "An Advanced Guide to Trade Policy Analysis: The Structural Gravity Model", *World Trade Organization, Geneva* .

Notes on approximation

As is apparent from equation (14) from Weidner & Zylkin (2020), constructing the bias correction for the variance in the three-way model requires computing objects of the form $\bar{H}_{ij}d_{ij}W^{(\phi)-1}d'_{ij}$, where $\bar{H}_{ij}$ captures the second derivative of the multinomial likelihood with respect to the the α- and γ-parameters associated with pair ij, d_{ij} is an $T \times [dim(\alpha)+dim(\gamma)]$ matrix of dummy variables, and $W^{(\phi)} := [I(J-1)]^{-1}\sum_{i,j} d'_{ij}\bar{H}_{ij}d_{ij}$ gives us the expected Hessian associated with these parameters.

From a computational perspective, the obvious problem here is that the inverse expected Hessian matrix $W^{(\phi)-1}$ could be difficult to calculate if the number of α- and γ-parameters is large. For example, for a trade data set with 150 countries and 10 years, we would have $[dim(\alpha)+dim(\gamma)] \approx 10\times150\times2 = 3{,}000$, and the dimension of $W^{(\phi)}$ would be roughly 3,000×3,000. While this is not necessarily "large" in the context of what modern techniques for large matrix inversion can theoretically handle, the series of calculations needed to compute the adjusted variance can nonetheless become sufficiently complex that Stata can and will run up against memory constraints and may even exit with an error.

As a consequence, `ppml_fe_bias` uses an approximation based on the method of alternating projections to compute the bias-corrected variance when the data is sufficiently large. The idea behind this approximation follows from the fact that the $\bar{H}_{ij}d_{ij}W^{(\phi)-1}d'_{ij}$ terms can instead be obtained by first computing the HDFE-PPML "annihilator" matrix $I - M'D(D'\bar{H}D)^{-1}D'\bar{H}$, where $\bar{H}$ is an $[dim(\alpha)+dim(\gamma)]\times[dim(\alpha)+dim(\gamma)]$ matrix with the $\bar{H}_{ij}$ expected Hessian terms arranged along its diagonal, D is an $IJT\times[dim(\alpha)+dim(\gamma)]$ matrix that stacks each of the d_{ij} matrices in the appropriate order, and M is a weighted demeaning operator such that $M\Lambda = \bar{H}$, with $\Lambda := diag(\lambda_{ijt})$ a diagonal matrix with the conditional mean λ_{ijt} along its diagonal. Each of the $T\times T$ blocks that lie along the diagonal of $M'D(D'\bar{H}D)^{-1}D'\bar{H}$ can be shown to be equal to $[\overline{H}_{ij}d_{ij}W_N^{(\phi)-1}d'_{ij}M_{ij}]'$, with the normalization matrix M_{ij} playing an innocuous role.

The annihilator matrix is an important theoretical construct from the literature on high-dimensional fixed effects estimation, where it is known, for example, that $\widetilde{x}^{*}_{ijt} := [(I - M'D(D'\bar{H}D)^{-1}D'\bar{H})X]_{ijt}$ gives us an appropriately "within transformed" version of the regressor x_{ijt} that purges it of any partial correlation with the fixed effects when the estimator is PPML. One way to obtain this matrix is via the method of alternating projections (see Stammann, 2018), an iterative process in which the identity matrix is repeatedly differenced by a series of weighted differencing operators reflecting the different fixed effects dimensions in the model (it, jt, and ij). The values for $M_{ij} d_{ij} W_N^{(\phi)-1} d'_{ij} \bar{H}_{ij}$ that are left after the first several iterations in this process are easy to calculate without involving any large matrix operations, and cutting off the calculation of the annihilator matrix at this point generally results in a reasonable approximation for the adjusted variance that still leads to significantly improved inferences.

To be more precise, define $M^{(it)}$ as $M^{(jt)}$ as $NNT \times NNT$ differencing operators such that $M^{(it)}X$ and $M^{(jt)}X$ respectively demean X with respect to $\overline{X}^{(it)}$ and $\overline{X}^{(jt)}$, where

$$\overline{X}^{(it)} = \sum_j \frac{\lambda_{ijt}}{\sum_n \lambda_{int}} X_{ijt} \forall i,t, \qquad \overline{X}^{(jt)} = \sum_i \frac{\lambda_{ijt}}{\sum_m \lambda_{mjt}} X_{ijt} \forall j,t$$

are weighted means across it and jt. It is also useful to note that the previously-defined matrix M itself is a difference operator that will difference X with respect to its weighted mean across ij. With these objects in hand, we can obtain the annihilator matrix by repeatedly differencing X with respect to the three fixed effect dimensions:

$$I_{NNT} - M'D(D'\bar{H}D)^{-1}D'\bar{H} = \lim_{r\to\infty}\left(M^{(it)}M^{(jt)}M\right)^r.$$

To formulate a convergent sequence for solving for the annihilator, we can express the current value after having performed $3r-1$ matrix multiplications as

$$A^{(r)} = M^{(it)}M^{(jt)}MA^{(r-1)} \text{ for } r \geq 1$$

with $A^{(0)} = I_{NNT}$. A seemingly natural algorithm would then involve repeatedly iterating on $A^{(r)}$ until convergence. However, because this algorithm would require repeatedly multiplying large ($NNT \times NNT$) matrices by one another, obtaining an exact solution in this manner is likely to be slow or even impractical. Instead, a feasible approximation can be constructed by taking the result after using just the four differencing terms, i.e.,

$$I_{NNT} - M'D(D'\bar{H}D)^{-1}D'\bar{H} \approx MA^{(1)} = MM^{(it)}M^{(jt)}M.$$

A key advantage of this approximation is that the $T \times T$ blocks needed for the variance adjustment that lie along the diagonal of $MA^{(1)}$ can be computed directly without having to perform any large matrix multiplications.[1] Despite the simplicity of this approach, it can be shown to work reasonably well for improving inferences, especially when compared to the alternative of using an uncorrected variance, which should generally be avoided. In the exercises included in the sample .do file, for example, the above approximation yields a standard error for `fta` of 0.0429, whereas an exact calculation gives a standard error of 0.0420. The original, uncorrected standard error is only 0.0350.[2]

[1] While these operators could have been applied in a different order instead, this order was chosen based on some simple testing and based on the constraint that full matrix multiplication was to be avoided.

[2] Of course, it would be ideal if a method could be devised for recovering the variance correction terms exactly without sacrificing too much computational efficiency. At the moment, the problem of how best to recover these types of objects in high-dimensional settings remains an area of active research. See Kline, Saggio, & Sølvsten (forthcoming) for another approximation method that could potentially

be applied here.