PPML in Gretl

Th. Warin

2025

In a gravity setting, the central objective of PPML is to estimate a multiplicative conditional-mean model of the form ($\mathbb{E}(X_{ij,t}\mid Z_{ij,t})=\exp(Z_{ij,t}\beta)$) without log-transforming the dependent variable, thereby retaining zero trade flows and ensuring robustness to heteroskedasticity under correct mean specification (Santos Silva & Tenreyro, 2006). In Gretl, the most faithful way to do this for gravity data is to estimate the Poisson quasi-likelihood via a user-specified maximum-likelihood block that returns per-observation log-likelihood contributions, because Gretl's built-in `poisson` command is documented as requiring a non-negative integer dependent variable, which is often not the case for trade values measured in currency units. (Gretl)

To make the procedure concrete, the tutorial below assumes that your dataset contains, at minimum, a nonnegative dependent variable `trade` (possibly with many zeros), exporter and importer identifiers `exp` and `imp`, a time variable `year`, a bilateral distance variable `dist`, and any additional gravity covariates you wish to include (contiguity, common language, PTA, etc.). The workflow is deliberately written as a reproducible Gretl script (Hansl), because it gives you precise control over fixed effects and clustered inference.

Step 1 consists in importing the data, checking basic admissibility conditions, and constructing the standard transformations. In PPML gravity you do not take logs of `trade`, but you typically take logs of strictly positive cost shifters such as distance.

```
# Example: load a CSV (adjust the path and separators as needed)
open "gravity.csv" --preserve --quiet

# Basic checks: trade must be nonnegative; dist must be positive if
logged
smpl trade >= 0 --restrict
genr lndist = log(dist)
```

Step 2 consists in defining the panel structure. For gravity panels, it is natural to define the observational unit as a dyad (exporter–importer pair) and time as `year`. Gretl can impose a panel interpretation using `setobs unitvar timevar --panel-vars`. (Departamento de Economía UC3M)

```
# Create a dyad identifier (ensure exp and imp are numeric codes)
genr dyad = exp*100000 + imp

# Declare dyad-year as a panel
setobs dyad year --panel-vars
```

Step 3 consists in constructing the fixed effects that correspond to modern structural gravity practice. A widely used specification includes exporter–time and importer–time fixed effects (and often dyad fixed effects if you have true panel variation within dyads). In Gretl, the pragmatic approach is to create discrete interaction identifiers (e.g., `exp_year` and `imp_year`) and then expand them into dummy lists using `dummify`, omitting one category to avoid collinearity. (Gretl)

```
# Exporter-year and importer-year identifiers
genr exp_year = exp*10000 + year
genr imp_year = imp*10000 + year

# Mark as discrete so dummify treats them as categorical
discrete exp_year
discrete imp_year
discrete dyad

# Create dummy lists; drop one category from each set
list FE_expyr = dummify(exp_year) --drop-first
list FE_impyr = dummify(imp_year) --drop-first

# Optional: dyad fixed effects (can be heavy if many dyads)
# list FE_dyad = dummify(dyad) --drop-first
```

Step 4 consists in defining the regressor list. In PPML gravity, you typically include bilateral trade-cost proxies (distance, borders, language, agreements) and omit exporter/importer GDP terms when exporter–time and importer–time fixed effects are included, because those fixed effects absorb time-varying country-level sizes by construction.

```
# Example bilateral covariates (adapt to your variables)
list Xbilat = lndist contig comlang rta

# Combine covariates and fixed effects
list X = const Xbilat FE_expyr FE_impyr
# If you also add dyad FE, you would use: list X = const Xbilat
FE_expyr FE_impyr FE_dyad
```

Step 5 consists in writing the PPML log-likelihood contribution at the observation level. Gretl's documentation and example scripts use the Poisson log-likelihood in the familiar form $(-\exp(xb)+y\cdot xb-\ln\Gamma(y+1))$, which remains well-defined for nonnegative, non-integer (y) via the gamma function. (Gretl)

```
function series ln_ppml(series y, list X, matrix b)
    series xb = lincomb(X, b)
    return -exp(xb) + y*xb - lngamma(y+1)
end function
```

Step 6 consists in running maximum likelihood with robust and clustered inference. For gravity, clustering by dyad is common because it allows arbitrary within-dyad dependence over time; Gretl's `mle` command supports both `--robust`(QML "sandwich" covariance) and `--cluster=clustvar` when the log-likelihood is provided per observation. (Gretl)

```
# Initialize parameters
matrix b = zeros(nelem(X), 1)

# Estimate PPML with dyad-clustered QML standard errors
mle ll = ln_ppml(trade, X, b)
    params b
end mle --robust --cluster=dyad --hessian
```

Step 7 consists in diagnosing convergence and interpreting coefficients. In a log-link mean model, a one-unit increase in regressor (x_k) changes the conditional mean multiplicatively by a factor (\exp(\beta_k)); the proportional effect is (\exp(\beta_k)-1). When (x_k=\log(\text{distance})), (\beta_k) is interpreted as an elasticity of the conditional mean with respect to distance, in the same sense as in log-linear gravity, but without discarding zeros in the dependent variable.

Two practical cautions are worth stating explicitly. First, the inclusion of exporter–time and importer–time fixed effects is often high-dimensional; Gretl can handle this for moderate datasets, but very large dyad–year panels may become memory-intensive when expanded into dummy variables. Second, if you add dyad fixed effects on top of exporter–time and importer–time fixed effects, identification of bilateral-invariant regressors becomes limited in the expected way: any regressor that is constant within dyads over time is absorbed by dyad fixed effects, and Gretl will typically drop it due to collinearity.

References

Cottrell, A., & Lucchetti, R. (2026). *Gretl user's guide*. Wake Forest University. (Gretl)
Santos Silva, J. M. C., & Tenreyro, S. (2006). The log of gravity. *Review of Economics and Statistics, 88*(4), 641–658. Gretl Team. (n.d.). *Gretl Command Reference* (commands: `mle`, `dummify`, `poisson`). (Gretl)