4

Alternative Gravity Model Estimators

4 Alternative Gravity Model Estimators

The previous section primarily used OLS as the estimation methodology for a variety of gravity models, both intuitive and theoretical. Just as the basic model has been subject to increasing scrutiny from a theoretical point of view, so too has OLS as the baseline estimator been subject to criticism from an econometric point of view. This section describes two alternative estimators from the literature – Poisson and Heckman – and discusses their application to the gravity model, as well as their respective advantages and disadvantages. The bottom line for applied policy researchers is that both estimators are now very commonly used in the literature, and it is therefore important to ensure that results obtained using OLS are robust to their application.

4.1 The Poisson Pseudo-Maximum Likelihood Estimator

Consider the nonlinear form of the Anderson and Van Wincoop gravity model with a multiplicative error term:

$$X_{ij}^k = \frac{Y_i^k E_j^k}{Y^k}\left(\frac{\tau_{ij}^k}{\Pi_i^k P_j^k}\right)^{(1-\sigma_k)} e_{ij}^k \quad (39)$$

Taking logarithms gives the standard gravity model in linearized form, but makes clear that the error term is in logarithms too:

$$\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + \log e_{ij}^k \quad (40)$$

The mean of $\log e_{ij}^k$ depends on higher moments of e_{ij}^k, thus including its variance. If e_{ij}^k is heteroskedastic, which is highly probable in practice, then the expected value of the error term depends on one or more of the explanatory variables because it includes the variance term. This violates the first assumption of OLS and suggests that the estimator may be biased and inconsistent. It is important to note that this kind of heteroskedasticity cannot be dealt with by simply applying a robust covariance matrix estimator, since it affects the parameter estimates in addition to the standard errors. The presence of heteroskedasticity under the assumption of a multiplicative error term in the original nonlinear gravity model specification requires adoption of a completely different estimation methodology.

Santos Silva and Tenreyro (2006) present a simple way of dealing with this problem. They show that under weak assumptions – essentially just that the gravity model contains the correct set of explanatory variables – the Poisson pseudo-maximum likelihood estimator provides consistent estimates of the original nonlinear model. It is exactly equivalent to running a type of nonlinear least squares on the original equation. Since we are dealing with a pseudo-maximum likelihood estimator, it is not necessary that the data be in fact distributed as Poisson. So although Poisson is more commonly used as an estimator for count data models, it is appropriate to apply it far more generally to nonlinear models such as gravity.

The Poisson estimator has a number of additional desirable properties for applied policy researchers using gravity models. First, it is consistent in the presence of fixed effects, which can be entered as dummy variables as in simple OLS. This is an unusual property of nonlinear maximum likelihood estimators, many of which have poorly understood properties in the presence of fixed effects. The point is a particularly important one for gravity modeling because most theory-consistent models require the inclusion of fixed effects by exporter and by importer.

Second, the Poisson estimator naturally includes observations for which the observed trade value is zero. Such observations are dropped from the OLS model because the logarithm of zero is undefined. However, they are relatively common in the trade matrix, since not all countries trade all products with all partners (see e.g., Haveman and Hummels, 2004). Although the issue has mainly arisen to date in the context of goods trade, it is also relevant for services trade (see further below). Dropping zero observations in the way that OLS does potentially leads to sample selection bias, which has become an important issue in recent empirical work (see further below). Thus the ability of Poisson to include zero observations naturally and without any additions to the basic model is highly desirable.

Third, interpretation of the coefficients from the Poisson model is straightforward, and follows exactly the same pattern as under OLS. Although the dependent variable for the Poisson regression is specified as exports in levels rather than in logarithms, the coefficients of any independent variables entered in logarithms can still be interpreted as simple elasticities. The coefficients of independent variables entered in levels are interpreted as semi-elasticities, as under OLS.

Stata contains a built in *poisson* command that can easily be applied to the gravity model, but it suffers from a number of numerical issues that result in sometimes unstable or unreliable results. A better option for applied researchers is to use the *ppml* command developed by Santos Silva and Tenreyro (2011b). The command can be installed by typing *findit ppml* and following the prompts. The *ppml* command automatically uses the *robust* option for estimation, so it is not necessary to specify it. Clustering can be corrected for in the usual way, i.e. by specifying the *cluster(dist)* option. If the command experiences estimation problems, it is sometimes possible to work around them by expressing the dependent variable as trade in thousands or millions of dollars, rather than in dollars: large values of the dependent variable are more difficult to treat numerically, and dividing through by a constant does not make any difference to the final result, as the estimator is scale-invariant.

Table 12 presents results for a fixed effects gravity model estimated using the *ppml* command. The first point to note is that, as expected, the number of observations is greater using Poisson than using OLS: 6580 compared with 3884. This difference shows that there is a large number of zero observations in the dataset, which is typical for gravity data. Those observations were dropped from the OLS estimates because the dependent variable was in logarithms, but they can be included naturally by Poisson.

Table 12: Poisson estimates of a fixed effects gravity model

```
. ppml trade ln_distance contig comlang_off colony comcol exp_dum_* imp_dum_* if
sector=="SER", cluster(dist)
note: checking the existence of the estimates
note: starting ppml estimation
note: exp_dum_218 omitted because of collinearity
note: imp_dum_218 omitted because of collinearity
note: trade has noninteger values

Iteration 1:   deviance =  919576.3
Iteration 2:   deviance =  571271.1
Iteration 3:   deviance =    505370
Iteration 4:   deviance =  494184.4
Iteration 5:   deviance =  492564.3
Iteration 6:   deviance =    492333
Iteration 7:   deviance =  492304.4
Iteration 8:   deviance =  492302.2
Iteration 9:   deviance =  492302.1
Iteration 10:  deviance =  492302.1
Iteration 11:  deviance =  492302.1

Number of parameters: 407
Number of observations: 6580
Number of observations dropped: 148
Pseudo log-likelihood: -256224.74
R-squared: .86676832
                                (Std. Err. adjusted for 3357 clusters in dist)
------------------------------------------------------------------------------
             |               Semirobust
       trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 ln_distance |    -.55767   .0483891   -11.52   0.000    -.6525108   -.4628292
      contig |   .2205841   .1670074     1.32   0.187    -.1067443    .5479126
 comlang_off |   .4592715   .1174326     3.91   0.000      .229108    .6894351
      colony |   .1420645   .1153311     1.23   0.218    -.0839803    .3681094
      comcol |  -.5527961   .3745067    -1.48   0.140    -1.286816    .1812235
   exp_dum_1 |  -.2121395   .8042537    -0.26   0.792    -1.788448    1.364169
   exp_dum_2 |  -1.537334   .6055117    -2.54   0.011    -2.724115   -.3505525
   exp_dum_3 |    2.25762   .5087708     4.44   0.000     1.260448    3.254793
```

It is also notable that the Poisson model fits the data much better than does the original OLS model. R2 for the former is around 87 per cent, compared with 77 per cent for OLS. Given that the set of explanatory variables is the same in both cases, this difference suggests that the change in estimator is important in order to pick up significant features of the data, most likely heteroskedasticity of the type outlined above.

Finally, the coefficient estimates are significantly different under Poisson compared with OLS. In particular, the distance coefficient is smaller in absolute value. This result is typical of Poisson gravity regressions, and largely reflects the impact of heteroskedasticity on the original OLS estimates (Santos Silva and Tenreyro, 2006).

The Poisson estimator is becoming steadily more popular in the literature, but it is not free from divergent opinions. Applied researchers need to be aware of some of the issues that have been highlighted in relation to the Poisson estimator, and of the additional properties that have been demonstrated for it as a result.

On the one hand, some researchers have used alternative count data models in place of Poisson, such as the negative binomial model, on the assumption that trade data are likely to exhibit over-dispersion (variance greater than the mean). However, this approach is erroneous for two reasons. First, Poisson is consistent as a pseudo-maximum likelihood estimator regardless of how the data are in fact distributed. The only improvement that could come from allowing for over-dispersion would be in terms of efficiency. For the efficiency gain to be real, moreover, the exact nature of the over-dispersion would need to be known, which it usually is not. Second, the negative binomial estimator has an undesirable property in a trade context: it is not scale invariant. Thus, results from a model with trade in dollars as the dependent variable will be different from those obtained with trade in millions of dollars as the dependent variable. This feature of the negative binomial model is not problematic in its usual count data setting, but becomes worrying in the gravity modeling context. Applied researchers should therefore avoid the negative binomial model in practice.

A second argument that has been made is that it other estimators may be superior to Poisson because they allow for a greater proportion of zeros in the observed trade matrix. However, the response to this argument is identical to the previous one: Poisson is consistent regardless of how the data are in fact distributed, assuming only that the zero and non-zero observations are produced by the same data generating process (see further below). In any case, recent simulation evidence (Santos Silva and Tenreyro, 2011a) shows that Poisson performs strongly even in datasets with large numbers of zeros.

Taking all of these points together, there is a strong argument for using Poisson as the workhorse gravity model estimator. From an applied policy research point of view, the desirable properties of Poisson suggest that estimates of policy impacts should generally be based on Poisson results rather than OLS.[4] In any case, Poisson results should always be presented for comparative purposes or as a robustness check.

[4] The choice between OLS and Poisson is, of course, an empirical one. Santos Silva and Tenreyro (2006) present a test for determining whether the OLS estimator is appropriate, and another for determining whether Poisson or another pseudo-maximum likelihood estimator is likely to be efficient. However, a detailed presentation of these tests is outside the scope of the current user guide.

4.2 The Heckman Sample Selection Estimator

As noted above, zero trade flows are relatively common in the bilateral trade matrix. As the level of product disaggregation becomes greater, so do zeros become more frequent. Even using aggregate trade data, Helpman et al. (2008) report that around half of the bilateral trade matrix is filled with zeros. Dropping these observations – as OLS automatically does because the logarithm of zero is undefined – immediately gives rise to concerns about sample selection bias. The sample from which the regression function is estimated is not drawn randomly from the population (all trade flows), but only consists of those trade flows which are strictly positive. One way of thinking of this problem is that the probability of being selected for the estimation sample is an omitted variable in the standard gravity model. To the extent that that variable is correlated with any of the variables included in the model – which it certainly is, since the probability of trading no doubt depends on trade costs – then a classic case of omitted variable bias arises. This is a violation of the first OLS assumption, which can lead to biased and inconsistent parameter estimates.

One way of dealing with this problem is to use the sample selection correction introduced by Heckman (1979). Helpman et al. (2008) developed a model of international trade that yields a gravity equation with a Heckman correction combined with an additional correction for firm heterogeneity. We explore only the first part of their model here (the Heckman correction), not the second.

To apply the Heckman sample selection model to the gravity model, we first need to split it into two parts: an outcome equation, which describes the relationship between trade flows and a set of explanatory variables, and a selection equation, which describes the relationship between the probability of positive trade and a set of explanatory variables. The outcome equation takes the form of the standard gravity model, but makes clear that it only applies to those observations within the estimation sample:

$$\left.\begin{aligned} &\log X_{ij}^k = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + e_{ij}^k \; if \; p_{ij}^k > 0 \\ &\log X_{ij}^k = missing \; if \; p_{ij}^k \leq 0 \end{aligned}\right\} \quad (41)$$

The variable p_{ij}^k is a latent (unobserved) variable that can be interpreted as the probability that a particular data point is included in the estimation sample. The selection equation relates the latent variable to a set of observed explanatory variables. That set must include all variables in the outcome equation, and preferably at least one additional variable that affects the probability that two countries engage in trade, but not the volume of such trade once it takes place.[5] One possible candidate is the cost of market entry of the exporter and the importer from the World

[5] Strictly speaking, the Heckman model is just-identified when the two sets of explanatory variables are the same. However, identification is only achieved due to the fact that the inverse Mill's ratio is a nonlinear function of the explanatory variables. Results therefore tend to be more stable when an additional over-identifying variable is included in the selection model. Applied researchers should try to specify an over-identified model whenever possible.

Bank's Doing Business dataset, as used by Helpman et al. (2008) in robustness checks. Using this variable, the selection equation takes the following form, where p_{ij}^{k} is the (unobserved) probability of selection, and d_{ij}^{k} is an (observed) dummy variable equal to unity for those observations that are in the sample, and zero for those that are not.

$$p_{ij}^{k} = \log Y_i^k + \log E_j^k - \log Y^k + (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + b \log entry_{ij} + w_{ij}^k \quad (42)$$

$$\left.\begin{array}{l} d_{ij}^k = 1\ if\ p_{ij}^k > 0 \\ d_{ij}^k = 0\ if\ p_{ij}^k \leq 0 \end{array}\right\} \quad (43)$$

where w_{ij}^{k} is a standard error term.

Another way of looking at the sample selection problem is that it creates bias if the error terms in the selection and outcome equations are correlated. In the trade context, we have good reason to believe that such correlation will be significant, in light of the tendency of firms to self-select into export status (Helpman et al., 2008).

Intuitively, the solution proposed by Heckman (1979) amounts to a two-step procedure. The first step is to estimate the probability that a particular observation is included in the gravity model sample, using a probit estimator. We therefore estimate:

$$Prob\left(d_{ij}^k = 1\right) = \Phi\Big(\log Y_i^k + \log E_j^k - \log Y^k$$

$$+ (1-\sigma_k)\left[\log \tau_{ij}^k - \log \Pi_i^k - \log P_j^k\right] + b \log entry_{ij}\Big) \quad (44)$$

The probit estimates are then used to calculate the inverse Mill's ratio $\left(\frac{\phi}{\Phi}\right)$, which corresponds to the probability of selection variable omitted from the original equation.[6] Inclusion of the additional variable solves the omitted variable bias and produces estimates that are consistent in the presence of non-random sample selection.

There are a number of technical drawbacks to actually performing the Heckman (1979) correction as a two-step procedure, however. Most researchers therefore use a maximum likelihood procedure in which the selection and outcome equations are estimated simultaneously. That approach is implemented by default in Stata's *heckman* command. The format for *heckman* is similar to that for *regress*, but the option *select(variables)* must always be specified. This option tells Stata which variables to include in the selection equation. As noted above, that list must always include the full set of variables from the original gravity model, and preferably one additional variable, such as entry costs, that effects the probability that two countries engage in trade, but not the volume of trade conditional on the existence of a trading link.

[6] In technical terms, the inverse Mill's ratio is the ratio of the probability density function to the cumulative distribution function.

Table 13 presents results for a Heckman sample selection model using our services data. The first part of the output is the outcome equation, i.e. the usual gravity model. Coefficients are quite close to their OLS counterparts, except for the common colonizer dummy, which has the expected positive sign in the Heckman results but is statistically insignificant.

Table 13: Heckman estimates of a fixed effects gravity model

```
. heckman ln_trade ln_distance contig comlang_off colony comcol i.exporters
i.importers if sector == "SER", select(ln_distance contig comlang_off colony comcol
ent_cost_both i.ex
> porters i.importers) robust cluster(dist)

Iteration 0:   log pseudolikelihood = -6436.0627
Iteration 1:   log pseudolikelihood = -6433.5323
Iteration 2:   log pseudolikelihood = -6433.5284
Iteration 3:   log pseudolikelihood = -6433.5284

Heckman selection model                         Number of obs      =      5164
(regression model with sample selection)        Censored obs       =      1681
                                                Uncensored obs     =      3483

                                                Wald chi2(322)     =         .
Log pseudolikelihood = -6433.528                Prob > chi2        =         .

                             (Std. Err. adjusted for 2617 clusters in dist)
------------------------------------------------------------------------------
             |               Robust
    ln_trade |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ln_trade     |
 ln_distance |   -1.05512   .0461723   -22.85   0.000    -1.145617   -.9646245
      contig |   .2037951   .2015601     1.01   0.312    -.1912554    .5988456
 comlang_off |    .461497   .0964098     4.79   0.000     .2725372    .6504568
      colony |   1.115957    .115507     9.66   0.000     .8895671    1.342346
      comcol |   .2982764   .2700312     1.10   0.269     -.230975    .8275277
             |
   exporters |
           3 |   2.565928   .6870591     3.73   0.000     1.219317    3.912539
           5 |  -1.370435   .8193477    -1.67   0.094    -2.976327    .2354564
           8 |    3.77945   .4426346     8.54   0.000     2.911903    4.646998


-------------+----------------------------------------------------------------
select       |
 ln_distance |   -.6646448   .1862166    -3.57   0.000    -1.029623   -.2996671
      contig |    .9822229   .4741379     2.07   0.038     .0529297    1.911516
 comlang_off |    .4065678   .1439841     2.82   0.005     .1243641    .6887715
      colony |    .8266087   .2456122     3.37   0.001     .3452177       1.308
      comcol |    .1321516   .2518237     0.52   0.600    -.3614138    .6257171
ent_cost_both |  -.2595793   .2237018    -1.16   0.246    -.6980268    .1788682


-------------+----------------------------------------------------------------
     /athrho |  -.1166728   .0709321    -1.64   0.100    -.2556972    .0223515
    /lnsigma |   .0552157   .0166455     3.32   0.001     .0225912    .0878402
-------------+----------------------------------------------------------------
         rho |  -.1161463   .0699752                     -.2502665    .0223478
       sigma |   1.056769   .0175904                      1.022848    1.091814
      lambda |  -.1227397   .0744213                     -.2686028    .0231234
------------------------------------------------------------------------------
Wald test of indep. eqns. (rho = 0): chi2(1) =     2.71   Prob > chi2 = 0.1000
```

The second part of Table 13 presents results for the selection equation, i.e. the probit model of export participation. In line with intuition, distance has a negative and 1 per cent statistically significant impact on the probability that two countries engage in trade. The historical and cultural variables all have the expected positive signs and are at least 5 per cent statistically significant, except for the common colonizer dummy, which is statistically insignificant. We can therefore see that geography as well as cultural and historical ties do not just influence the volume of trade between countries (outcome equation), but also the probability that two countries engage in trade at all (selection equation). Finally, the entry cost variable – which we have transformed to vary bilaterally by taking the product of the exporter and importer values – has the expected negative sign, but its coefficient is statistically insignificant. In this case, further research would be required with alternative over-identifying variables to try and deal with this problem. To produce stable and robust results, researchers should be careful to choose an over-identifying variable that both has a coefficient that accords with intuition, and one which is statistically significant at conventional levels.

The final part of the Stata output provides information on the relationship between the outcome and selection equations. As noted above, sample selection only creates bias if the error terms of the two equations are correlated. That information is contained in Stata's estimate of the parameter rho: an estimate that is large in absolute value (up to a maximum of one) and statistically significant suggests that sample selection is a major problem in a given dataset. In fact, the final line of Stata's output is a test of the null hypothesis that rho is equal to zero, i.e. that the two error terms are uncorrelated. If the null hypothesis is rejected, we conclude that sample selection is a serious issue. In this case, the test statistic is marginal at the 10 per cent level. It would be dangerous to conclude, however, that sample selection is not an issue: use of a better over-identifying variable may well give different results. Indeed, evidence from goods markets suggests that sample selection can indeed create significant bias (Helpman et al., 2008).

Like the Poisson estimator, the Heckman model provides a natural way of including zero trade observations in the dataset. As yet, the literature does not provide any decisive guidance on which model should be preferred in applied work. Each has its own set of advantages and disadvantages. For example, Poisson deals well with heteroskedasticity, but Heckman does not. Similarly, fixed effects Poisson models are well understood and have desirable statistical properties, but fixed effects probit models suffer from a technical issue – the incidental parameters problem – that introduces bias and inconsistency into the estimates, but the empirical extent of that issue is still unclear. On the other hand, Heckman allows for separate data generating processes for the zero and non-zero observations, whereas Poisson assumes that all observations are drawn from the same distribution. For the moment, then, applied researchers generally need to present both Poisson and Heckman results in the interests of showing that their results are robust to the use of different, but commonly used, estimators.