UNIVERSITY OF WAIKATO

Hamilton
New Zealand

The Estimation and Interpretation of Coefficients in Panel Gravity Models of Migration

Michael P. Cameron and Jacques Poot

Working Paper in Economics 1/18

January 2018

Corresponding Author

Michael P. Cameron
School of Accounting, Finance and Economics
University of Waikato
Private Bag 3105
Hamilton
NEW ZEALAND, 3240

Email: mcam@waikato.ac.nz

Jacques Poot
University of Waikato

Email: jacques.poot@waikato.ac.nz

Abstract

We demonstrate that the conventional OLS and fixed effects estimators of gravity models of migration are biased, and that the interpretation of coefficients in the fixed effects model is typically incorrect. We present a new best linear unbiased estimator for gravity models of migration.

Keywords

gross migration flows
gravity model
New Zealand

JEL Classification

O15; R23

Acknowledgements

This research was funded by the Ministry of Business, Innovation and Employment (Grant numbers UOWX1404 and MAUX1206) and by the Royal Society of New Zealand Marsden Fund (Grant number UOW1302). The authors are grateful to Sialupapu Siameja and Tobias Dean for excellent research assistance. The usual disclaimer applies.

1. INTRODUCTION

Greater availability of temporal gross flow data in many types of spatial interaction (trade, migration, tourism, etc.) in recent years has led to fixed effects (FE) panel models becoming the standard approach for the estimation of gravity models, following Anderson and van Wincoop (2003). This preference for FE models can be attributed to the estimated coefficients in a FE model being consistent and estimable without recourse to the inclusion of multilateral resistance terms (Anderson and van Wincoop, 2003). Moreover, when compared with ordinary least squares (OLS), FE models eliminate bias arising from standard error clustering when some variables apply only to the origin or the destination, and not both (Feenstra, 2004). Peeters (2012) has critiqued the OLS and standard FE estimators for their dissimilarity in coefficient point estimates, but despite this critique and those of others (e.g., Etzo, 2011), FE gravity models continue to be widely used (Lewer and Van den Berg, 2008; Aldashev and Dietz, 2012).

In this paper, we first demonstrate that both the conventional OLS estimator of the gravity model and the conventional FE estimator are biased estimators of a spatial interaction data generating process (DGP) with origin and destination time invariant effects. We then show that comparisons of the population size coefficients between OLS and panel FE gravity models are not straightforward. We demonstrate that there should be no expectation that the signs and magnitudes of the coefficients on these models be similar. We then introduce a best linear unbiased estimator for the FE DGP that can be used in gravity model specifications. We illustrate the issue of gravity model estimation with the example of five-yearly internal migration flows between regions of New Zealand over the period 1991-2013.

2. MODEL SPECIFICATIONS

Although the gravity model can be augmented to include factors that differ between different regions and over time (Lewer and Van den Berg, 2008), we assume a specification that is expressed in log-linear form and supplemented with origin and destination fixed effects:

$$\ln M_{ij,t} = \beta_1 \ln P_{i,t} + \beta_2 \ln P_{j,t} + \beta_3 \ln D_{ij,t} + \gamma_i + \varphi_j + \varepsilon_{ij,t};\ i \neq j \qquad (1)$$

where $M_{ij,t}$ is the instantaneous flow or force of gross migration from area i (the origin) to area j (the destination) at time t, $i,j = 1,2,\ldots R$, $P_{i,t}$ and $P_{j,t}$ the corresponding population stocks in areas i and j respectively, $D_{ij,t}$ is the distance between i and j, and γ_i and φ_j are time-invariant origin and destination-specific fixed effects. The β_k (k=1,2,3) and origin and destination fixed effects are to be estimated, and $\varepsilon_{ij,t}$ is a white noise error term. If data were available on instantaneous flows of migration, the Best Linear Unbiased Estimator (BLUE) of Equation (1) is OLS with 2R-1 binary dummy variables representing the origin and destination fixed effects. However, considering explicitly the discrete time interval over which migration is measured, Equation (1) becomes:

$$\ln M_{ij,t\to(t+1)} = \beta_1 \ln P_{i,t\to(t+1)} + \beta_2 \ln P_{j,t\to(t+1)} + \beta_3 \ln D_{ij,t\to(t+1)} + \gamma_i + \varphi_j + \varepsilon_{ij,t\to(t+1)};\ i \neq j \tag{2}$$

Hence, Equation (1) is the cross-sectional limit case, i.e. $t\to 0$, of Equation (2). Now assume, realistically, that the populations of areas i and j are growing at time-varying growth rates. Additionally, we assume that $D_{ij,t\to(t+1)}$ is time-invariant (D_{ij}). A major source of bias now arises from estimating regression Equation (2) with cross-sectional or pooled (t=1,2,…,T) migration flow data $M_{ij,t\to(t+1)}$, in which each migration flow matrix corresponds to a relatively long period, while the populations P_{it} and P_{jt} are measured at the start of the period. Because the size of the origin and destination populations change over the course of the period, the estimated coefficients of Equation (2) are biased estimates of Equation (1). The degree of bias will be relatively small for annual time-steps, but larger for longer time-steps that are common in the migration literature. To reduce this bias, it is convenient to substitute the geometric average of the population over the time period, that is:

$$P_{k,t\to(t+1)} = \sqrt{P_{k,t}P_{k,(t+1)}};\ k \in \{i,j\} \tag{3}$$

Now, let $g_{k,t\to(t+1)}$ refer to the time-varying growth rate of population k. Then:

$$P_{k,(t+1)} = P_{k,t}\left(1 + g_{k,t\to(t+1)}\right);\ k \in \{i,j\} \tag{4}$$

Substituting (4) into (3) and then (3) into (2) and using that for small x, ln (1+x) ≈ x, leads to:

$$\ln M_{ij,t\to(t+1)} = \beta_1 \ln P_{i,t} + \beta_2 \ln P_{j,t} + \beta_3 \ln D_{ij} + \gamma_i + \tfrac{1}{2}\beta_1 g_{i,t\to(t+1)} + \varphi_j + \tfrac{1}{2}\beta_1 g_{j,t\to(t+1)} + \varepsilon_{ij,t\to(t+1)};\ i \neq j \tag{5}$$

We can now compare this with two common estimators of the gravity model. The first is the OLS estimator, which assumes the specification:

$$\ln M_{ij,t\to(t+1)} = \beta_1^* \ln P_{i,t} + \beta_2^* \ln P_{j,t} + \beta_3^* \ln D_{ij} + \varepsilon_{ij,t\to(t+1)}^* \tag{6}$$

and the second is the FE estimator which assumes the specification:

$$\ln M_{ij,t\to(t+1)} = \beta_1' \ln P_{i,t} + \beta_2' \ln P_{j,t} + \beta_3' \ln D_{ij} + \gamma_i' + \varphi_j' + \varepsilon_{ij,t\to(t+1)}' \tag{7}$$

It is clear that both estimators are biased estimators of the parameters of the "true" DGP described by Equation (5). In fact, given that $\varepsilon_{ij,t\to(t+1)}$, the geometric average of $\varepsilon_{ij,t}$ and $\varepsilon_{ij,t+1}$ is also a white noise error term (in the assumed absence of temporal autocorrelation), it is straightforward to directly estimate Equation (5) by restricted least squares (RLS). This simply requires the assumption that the time-varying population growth rates are exogenous. This assumption is plausible because there is no correlation between gross migration *levels* and net migration or population growth *rates*, even though gross inward (outward) migration is positively (negatively) correlated with net migration, see Vias (2001). Given the assumption of white noise errors, the RLS estimator is BLUE (Greene, 2017).

The (log) population at any point in time is a function of the initial population and the population growth rate:

$$\ln P_{k,t} = \ln P_{k,0} + \sum_{\tau=0}^{(t-1)} \ln\left(1 + g_{k,\tau\to(\tau+1)}\right); k \in \{i, j\} \tag{8}$$

Substituting (8) into (5) gives:

$$\ln M_{ij,t\to(t+1)} = \beta_1 \ln P_{i,0} + \beta_2 \ln P_{j,0} + \beta_3 \ln D_{ij} + \gamma_i + \beta_1 \left[\sum_{\tau=0}^{(t-1)} g_{i,\tau\to(\tau+1)} + \frac{1}{2} g_{i,t\to(t+1)}\right] + \varphi_j + \beta_2 \left[\sum_{\tau=0}^{(t-1)} g_{j,t\to(t+1)} + \frac{1}{2} g_{j,t\to(t+1)}\right] + \varepsilon_{ij,t\to(t+1)};\ i \neq j \tag{9}$$

Since $P_{i,0}$ and $P_{j,0}$ in Equation (9) are fixed and time-invariant, they form part of the fixed effects in the FE regression. Therefore the coefficients β_1 and β_2 in Equation (9) should be interpreted as relating to the effect of the population growth rates g_i and g_j. Because the choice of the base year is arbitrary, this shows that the effect of population scale in Equation (7) is also subsumed in the fixed effect and the estimated coefficients β_1' and β_2' reflect the impact of the population growth rates. This is in contrast with the OLS model of Equation (6) where the corresponding coefficients relate to the effect of population levels. This highlights that the coefficients in the OLS model are *not* directly comparable to the coefficients from a FE model.

In the simple OLS specification of the gravity model [Equation (6)], the expected sign on the coefficients β_1 and β_2 is positive. However, in the FE specification in Equation (7), there is no *a priori* reason to believe that the estimates of coefficients β_1' and β_2' should both be positive. For a given propensity to migrate (e.g. linked to the age structure of the population), faster population growth will imply higher outward migration levels, i.e. a positive estimate of β_1'. This is mostly a demographic effect. On the other hand, faster population growth may not necessarily imply relatively higher inward migration levels. The outcome would depend on a range of economic factors such as job growth, resource constraints and the corresponding prices (particularly of housing), and on the source of the population growth (natural increase, international migration, etc.). Consequently, the estimate of β_2' is not necessarily positive.

3. RESULTS

We test our interpretations using data from migration flows between the sixteen regions of New Zealand. Inter-regional migration data were obtained from the Census of Population and Dwellings (1996, 2001, 2006 and 2013), based on self-reported region of residence five years previously. Responses that were unidentifiable or not elsewhere classified were distributed proportional to valid responses (including non-movers), while the few zero-count flows were increased by one. This provides 960 observations of the dependent variable. Population numbers were taken from the Estimated Usually Resident Population at 30 June in each year. Distances between each region were population-weighted straight line distances (based on the 2013 population distribution). Additional dummy variable controls were included for contiguity of regions, and for flows between the two main islands of New Zealand. We implement BLUE, OLS, and FE regressions, equivalent to Equations (5), (6), and (7) respectively, as well as FE re-specified in growth rates as in Equation (9).

Table 1 presents the results of our regression models, excluding control variables and fixed effects. Column (1) contains the results for the OLS specification, equivalent to Equation (6), Column (2) is the FE specification of Equation (7), Column (3) corresponds to the growth rates specification of Equation (9), and Column (4) is the BLUE estimator (given the assumed DGP) of Equation (5).

Table 1: Regression Results

Model	(1) OLS	(2) FE	(3) FE (Growth Rates)	(4) BLUE (RLS)
lnP_i	0.818^{***} (0.016)	0.974^{***} (0.206)	-	0.666^{***} (0.141)
lnP_j	0.803^{***} (0.016)	-0.782^{***} (0.214)	-	-0.531^{***} (0.147)
g_i	-	-	1.051^{***} (0.228)	-
g_j	-	-	-0.824^{***} (0.228)	-
lnD_{ij}	-0.503^{***} (0.035)	-0.782^{***} (0.038)	-0.782^{***} (0.033)	-0.782^{***} (0.038)
Adj. R^2	0.888	0.948	0.948	-

Note: Robust standard errors in parentheses; n=960; *** p<0.01; ** p<0.05; * p<0.1.

All coefficients in Table 1 are highly statistically significant. Comparing the coefficients between the OLS (1) and FE (2) models highlights one substantial difference – the coefficient on the destination population is positive and statistically significant in the OLS specification, but negative and statistically significant in the FE specification. This change in coefficients could be construed as demonstrating a lack of robustness in the estimates. However, based on the exposition of our specification earlier, it is clear that the coefficients on population variables cannot be compared directly. The negative sign on population in the FE specification simply suggests that higher population growth rates in the destination are associated with smaller migration flows. As described above, our results are consistent with inward migration being constrained, perhaps by unavailability of a suitable quantity of affordable housing for migrants, or competition between internal and international in-migrants for the housing that is available. Moreover, this result is consistent with the model in Column (3), where the model is re-specified in growth rates. The results in column (4) for the BLUE estimator demonstrate the degree of bias in the OLS and FE estimators for the DGP in Equation (5) – the coefficients on population are much smaller with the BLUE estimator, suggesting that other estimators substantially over-state the effect of population on migration.

4. CONCLUSION

Peeters (2012) critiqued the OLS and FE estimators for their dissimilarity in the point estimates of the population coefficients. However, as demonstrated in this paper, there is no a priori reason to expect these coefficients to hold the same sign. The coefficients in OLS and FE models must be interpreted differently. This misinterpretation is relatively common. For instance, the unexpected sign on employment in Aldashev and Dietz (2014), and the change in coefficient signs between OLS and FE models in Ramos and Surinach (2017) can be explained in this way. Moreover, we have shown that the standard OLS and FE models lead to biased coefficients on population compared with a BLUE estimator of a DGP with fixed effects. These results have significance not only for the estimation and interpretation of the coefficients of gravity models in the migration literature, but also in the literature on trade.

References

Aldashev, A., and Dietz, B. (2014). Economic and spatial determinants of interregional migration in Kazakhstan, *Economic Systems* 38(3), 379-396.

Anderson, J.E, and van Wincoop, E. (2003). Gravity with gravitas: A solution to the border puzzle, *American Economic Review* 93(1), 170-192.

Etzo, I. (2011). The determinants of the recent interregional migration flows in Italy: A panel data analysis, *Journal of Regional Science* 51(5), 946-966.

Feenstra, R.C. (2004). *Advanced international trade: Theory and evidence*. Princeton, NJ: Princeton University Press.

Greene, W.H. (2017). *Econometric Analysis. 8th edition*. New York: Pearson.

Lewer, J.J., and Van den Berg, H. (2008). A gravity model of immigration, *Economics Letters* 99, 164-167.

Peeters, L. (2012). Gravity and spatial structure: The case of interstate migration in Mexico, *Journal of Regional Science* 52(5), 829-856.

Ramos, R., and Surinach, J. (2017). A gravity model of migration between ENC and EU. *Tijdschrift voor Economische en Sociale Geografie* 108(1), 21-35.

Vias, A.C. (2001). Reevaluating the relationship between in-, out-, and net migration for non-metropolitan counties: an update on Beale's u-shaped curve, *Geographical Analysis* 33(3), 228-246.