DISCUSSION PAPER SERIES

IZA DP No. 10329

The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science

Jacques Poot
Omoniyi Alimi
Michael P. Cameron
David C. Maré

October 2016

Forschungsinstitut
zur Zukunft der Arbeit
Institute for the Study
of Labor

The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science

Jacques Poot
University of Waikato and IZA

Omoniyi Alimi
University of Waikato

Michael P. Cameron
University of Waikato

David C. Maré
Motu Economic and Public Policy Research and IZA

Discussion Paper No. 10329
October 2016

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0
Fax: +49-228-3894-180
E-mail: iza@iza.org

IZA Discussion Paper No. 10329
October 2016

ABSTRACT

The Gravity Model of Migration: The Successful Comeback of an Ageing Superstar in Regional Science*

For at least half a century, and building on observations first made a century earlier, the gravity model has been the most commonly-used paradigm for understanding gross migration flows between regions. This model owes its success to, firstly, its intuitive consistency with migration theories; secondly, ease of estimation in its simplest form; and, thirdly, goodness of fit in most applications. While fitting gravity models of aggregate migration flows started taking backstage to microdata analysis in the 1980s, a recent comeback has resulted from increasing applications to international migration and from the emergence of statistical theories appropriate for studying spatial interaction. In this paper we review the status quo and argue for greater integration of internal and international migration modelling. Additionally we revisit the issues of parameter stability and distance deterrence measurement by means of a New Zealand case study. We argue that gravity modelling of migration has a promising future in a multi-regional stochastic population projection system – an area in which the model has been to date surprisingly underutilised. We conclude with outlining current challenges and opportunities in this field.

JEL Classification: F22, J61, R23

Keywords: gravity model, internal migration, international migration, population projection

Corresponding author:

Jacques Poot
National Institute of Demographic and Economic Analysis
University of Waikato
Private Bag 3105
Hamilton
New Zealand
E-mail: jpoot@waikato.ac.nz

* This is a draft of an invited paper forthcoming in *Investigaciones Regionales / Journal of Regional Research*. This study has been supported by the 2012-2014 Nga Tangata Oho Mairangi project, funded by Ministry of Business Innovation and Employment grant CONT-29661-HASTR-MAU. Access to the data used in this study was provided by Statistics New Zealand (SNZ) under conditions designed to give effect to the security and confidentiality provisions of the Statistics Act 1975. All frequency counts using Census data were subject to base three rounding in accordance with SNZ's release policy for census data. We are grateful for comments received at the 8th International Conference on Population Geographies, University of Queensland, Brisbane, July 1-4, 2015; the LSE Population Seminar, London School of Economics and Political Science, November 19, 2015; and the 56th Congress of the European Regional Science Association, Vienna University of Economics and Business, August 23-26, 2016. The views, opinions, findings and conclusions or recommendations expressed in these papers are strictly those of the authors and do not necessarily represent, and should not be reported as, those of the organisations at which the authors are employed.

1. Introduction

One of the most pervasive empirical regularities in regional science is that any form of spatial interaction (migration, commuting, trade, information exchange, etc.) has the property of flows being positively related to stocks, whichever way measured, and inversely related to distance. Thus, the 'law' of spatial interaction in human behaviour (see also e.g. Anderson, 2011) resembles Newton's 1687 law of gravity. The idea of applying a physics law to population movement between two locations was first formally advocated by John Q. Stewart who established the 'social physics' school (Stewart, 1950). However, the gravity-like properties of internal migration flows had already been confirmed much earlier by Ravenstein (1885, 1889). There is of course no reason to expect that spatial interaction operates *exactly* as the gravity law of physics would dictate and Zipf (1946) already established that for US intercity movement of persons the flows were inversely related to distance and not to distance squared.

In its most commonly applied form, the gravity law of population migration states that

$$M_{ij} = G\frac{{P_i}^{\alpha} \times {P_j}^{\beta}}{D_{ij}^{\gamma}} \qquad (1)$$

in which: M_{ij} refers to the number of people resident in area j who at an earlier point in time (usually one or five years) resided in area i; P_i (P_j) refers to the population of i (j) usually measured at the beginning of the period over which migration is measured; D_{ij} is some measure of distance between i and j; α, β and γ are parameters to be estimated and G is a proportionality constant that is context specific (dependent on the geography, time dimension, etc.). The popularity of this simple model was undoubtedly related to the ease with which the model could be estimated by ordinary least squares after a transformation into logarithmic form:

$$\ln M_{ij} = \delta + \alpha \ln P_i + \beta \ln P_j - \gamma \ln D_{ij} + \varepsilon_{ij}, \qquad (2)$$

in which a zero-mean error term ε_{ij} has been added to the equation and the constant term ln G has been replaced by the parameter δ.[1] Historically, the absence of any migrants for certain specific origin-destination combinations (which is common in large and sparse gross migration matrices) was a cause for some concern, although easily ameliorated by substituting a small number such as 0.5 for such zeros. Count models (e.g., Biagi et al., 2011) or direct nonlinear estimation of migration model parameters (e.g., Fik and Mulligan, 1998) are nowadays quite common alternative approaches.

Parameter estimates of Eq. (2) vary across countries. An interesting recent project by Stillwell et al. (2014), called the IMAGE Studio, has been concerned with comparative modelling of internal migration in a wide range of countries. This project highlights the sensitivity of the distance decay parameter to the geography of the available data, specifically the boundaries and areas of the spatial units. In a UK application, Stillwell et al. (2014) find that the estimate of γ converges to around 1.5 to 1.6 once the country is carved up into 50 or more regions. In a New Zealand application, Alimi et al.

[1] The fact that the errors ε_{ij} are unlikely to be statistical 'white noise', i.e. independently and identically distributed, has been largely ignored in many applications. Curry (1972) was the first to tackle spatial correlation in the gravity model (of commuting) but major advances in statistical theory of spatial interaction modelling did not emerge until LeSage and Pace (2008). For a recent review of this literature, see e.g. Patuelli (2016).

(2015) find estimates of γ between 0.8 and 0.9 when the data refer to migration between 39 urban areas. As illustrated by these examples, distance decay in migration is generally less than 2, which is the value implied by Newton's law of gravity.

Estimates of α and β vary across applications as well. In the New Zealand case, estimates of α and β are commonly between 0.8 and 0.9 (Alimi et al., 2015). The two parameters are unlikely to be identical in the migration context. Given that $D_{ij} = D_{ji}$ in most applications and $E(\varepsilon_{ij}) = E(\varepsilon_{ji}) = 0$, the expected value of net migration between any origin and destination pair, $E(M_{ji} - M_{ij})$, is zero when $\alpha = \beta$. This is rather unrealistic given that there are in most countries regions that structurally gain population through internal migration while others lose population. We can use Eq. (1) to obtain an equation for net migration as follows:

$$M_{ij} - M_{ji} = M_{ij}\left[1 - \frac{M_{ji}}{M_{ij}}\right] = M_{ij}\left[1 - \left(\frac{P_i}{P_j}\right)^{\beta-\alpha}\right] \tag{3}$$

which shows that, when $\beta > \alpha$ and $P_j > P_i$, $M_{ij} - M_{ji} > 0$. The system has then a tendency for larger regions to be population gainers through internal migration while smaller regions lose population that way. This configuration could reflect agglomeration forces leading to, on average, expansion of the larger cities through net inward internal migration. One example is internal migration in New Zealand for the five-year periods between population censuses from 1981 until 2001, as will be shown in the next section.

Despite its simplicity, the gravity model fits internal migration data remarkably well – often yielding adjusted R^2 values of between 0.8 and 0.9. This makes the model useful for embedding in sub-national population forecasting procedures, as will be elaborated in Section 3. It is also possible to justify the gravity model of migration in terms of microeconomic foundations. For example, Poot (1995) shows that in a labour market in which workers can draw wage offers from distributions of jobs in each region among a set of potential destination regions, migration flows are positive related to the size of the labour force in origin and destination regions and inversely related to the cost of migrating from one to the other.

However, such a stylised description of migration as the gravity model provides is of limited use for those attempting to quantify the *processes* that drive population redistribution. The latter has been achieved over the last half century by many developments across a range of disciplines. Comprehensive reviews of modelling internal migration flows and propensities to migrate include Greenwood (1997). Clearly, the potential endogeneity of many determinants of internal migration flows remains a challenging issue for estimation. Suitable instruments are often difficult to find and it is common practice to use 'deep lagging' of right-hand side variables as a statistically acceptable practice in cross-section and panel models of gross migration flows.

One fundamental weakness of the basic gravity model is the absence of any systemic effects. This was first addressed by Wilson (1970) in the doubly-constrained spatial interaction model in which (1) is replaced by

$$M_{ij} = A_i M_{i.} B_j M_{.j} D_{ij}^{-\gamma} \tag{4}$$

with $M_{i.}$ referring to total out-migration from *i*, $M_{.j}$ referring to total in-migration into *j* and A_i and B_j are balancing factors that ensure that gross origin-destination migration flows add up to exogenous and pre-set out-migration and in-migration flows for each region. If structural equations are added to (4) that include macro-level determinants of $M_{i.}$ and $M_{.j}$, Alonso's (1978) general theory of movement results. The empirical estimation of this model gained some popularity during the 1980s (see e.g. De Vries et al., 2001, for a review and Poot, 1986, for a New Zealand application). While the Alonso model has also interesting theoretical properties in a dynamical setting (see Nijkamp and Poot, 1987), its nonlinearity complicates interpretation. Since the 1980s the internal migration literature has predominantly moved to micro-data analysis (Cushing and Poot, 2004). In contrast with that, there has been growing interest in more recent years in explaining gross international migration flows by gravity models (e.g. Mayda, 2010; Ramos, 2016). Recent econometric issues in gross migration modelling include the challenge of accounting for spatial spillovers in such flow models (e.g., LeSage and Pace, 2008; 2009). Another issue is that of spatial heterogeneity in the parameters (e.g., Peeters, 2012). A further interesting development has been the interpretation of migration flows as a weighted network, with applications both in internal migration (e.g., Mayer and Vyborny, 2008) and international migration (Tranos et al., 2015; Fagiolo and Mastrolillo, 2013; Davis et al., 2013).

Given its enduring popularity, we revisit in the next section several key issues in estimating conventional gravity models of migration. We firstly focus on the temporal stability of gravity model coefficients; secondly, on the best measurement of the distance deterrence effect; and thirdly, the extent to which reduced long-distance travel time and costs have spurred additional migration. Additionally, one of the main deficiencies of internal migration modelling to date is the common neglect of accounting for international migration flows. We therefore show that inter-urban migration flows can be easily embedded in an expanded gravity model that also includes international (and urban-rural) flows. Considering international migration flows in spatial population redistribution is nowadays particularly important given the rapid growth in the stock of foreign born in most developed countries. We use New Zealand data to look at each of the four specific issues stated above. New Zealand is an attractive case to consider given that geographical mobility is high and the foreign born account for about one quarter of the population.

It is well known that a model that describes the evolution of a multiregional population leads to biased forecasts when population change is modelled as a function of net migration rather than gross migration (Rogers, 2015). Yet it still remains common, when forecasting population change in a multiregional system, to use assumed age-specific net migration numbers for each region that are subsequently calibrated to ensure that total net migration in the system is zero (this applies to the assumed net international migration by country in UN global population projections too). In section 3 we briefly outline the possibility of developing a multi-regional population projection system that includes a gravity model of interregional migration.

Given that the gravity model of gross migration has returned to prominence as a tool for analysing and projecting multi-regional populations, we may expect a range of new development triggered by new types of data, such as 'big data' obtained by various electronic information systems and new techniques for statistical analysis of dyadic data generated by population movement. The final section of the paper, section 4, briefly elaborates on such potential developments.

2. Sensitivity of the gravity model to specification choices – a New Zealand case study

The distance variable included in Eq. (1), (2) and (4) of the previous section is open to a range of interpretations and measurements. It is usually thought of as a proxy of the cost of migration and measured in various ways. Most applications of the gravity model to migration usually select only one single measure of distance between origin and destination, such as the railroad distance (Fan, 2005), straight line distance (Lewer and van der Berg, 2008), airline miles travelled between origin and destination airports (Karemera, Oguledo & Davis, 2000) and road travel distance (Courchene, 1970). In this section we first consider the sensitivity of the gravity model to three different measures of distance, namely: straight line distance (distance as the crow flies), road travel distance in kilometres and road travel time in minutes. We use New Zealand data to test the sensitivity of the gravity model to these different measures of distance.

Some measures of distance change over time, for example due to changes in preferred transport modes, transport technology, new infrastructure or changes in the speed limit imposed on road use. Hence we also test the extent to which changes over time in distance between specific origin-destination pairs impact on the corresponding migration flows. Moreover, in today's world in which cross-border migration flows are increasingly important (e.g. Poot, 2015), we also consider the impact of modelling internal and international migration flows simultaneously. New Zealand is a good case study for this, given that about one quarter of the population is foreign born.[2]

Migration data are recorded in the New Zealand Census by means of a question on 'usual residence five years ago'. The census is held every five years, except for the 2011 census which was postponed until 2013 due to a major earthquake in February 2011 in Christchurch, where the Statistics New Zealand (SNZ) census division is located. We assembled data from six censuses, starting in 1986. We focus on each of the 40 areas identified by SNZ as main and secondary urban areas in New Zealand in 2013.[3] The population is restricted to individuals aged between 25 and 54 in order to model predominantly labour migration and exclude movements of students and retired persons. We also embed inter-urban migration in a population flows matrix that includes international migration and migration between urban and rural areas. Since the census includes only people who are actually in New Zealand at the time of the census, emigration from New Zealand is not recorded in census data but has been estimated by a residual method.[4]

Excluding international and rural-urban migration, the specification of the gravity model is identical to Eq. (2). To include international and rural-urban migration we first note that 'international' and

[2] The importance of considering the interactions between interregional and international gross migration was previously considered, for example in the United Kingdom case, by Raymer at al. (2012) and Lomax et al. (2013).

[3] Urban areas in New Zealand have a population of at least 1000 people, but population size is not the only criterion to classify urban areas — factors such as remoteness, economic activity and location of employment of the majority of the population are also used to define and further differentiate the type of urban area. Statistics New Zealand categorises three types of urban areas: main urban areas which have a population of at least 30,000 people; secondary urban areas are ones with a population of between 10,000 and 29,999 people; and minor urban areas have a population of between 1,000 and 9,999.

[4] Given that censuses are held at the same time every five years, cohorts can be followed over time. After accounting for immigration, internal migration and observed registrations of deaths, emigration can be calculated as the residual change in the size of a cohort. Of course the resulting numbers are measured with some error, due to census undercounting, etc.

'rural' do not have a specific location, so that the distance between these areas and the set of urban areas may be considered undefined. Similarly, we consider the population of the generic international and rural areas undefined (the total rural population is known, but it is a spatially dispersed rather than compact mass). We now define five dummies variables. Firstly, U_{ij} = 1 if and only if both the origin *i* and the destination *j* are urban areas and 0 otherwise. Secondly, E_{j} = 1 if and only if the origin *i* is an urban area and the destination *j* is abroad (i.e., these correspond to the emigration flows). Thirdly, I_{ij} = 1 if and only if the origin *i* is abroad and the destination *j* is an urban area (i.e., these correspond to the immigration flows). Fourthly, O_{ij} = 1 if and only if the origin *i* is an urban area and the destination *j* is rural (i.e., these correspond to the urban to rural flows). Finally, R_{ij} = 1 if and only if the origin *i* is rural and the destination *j* is an urban area (i.e., these correspond to rural to urban migration flows). The specification of the gravity model with international and rural-urban migration then becomes:

$$\begin{aligned}\ln M_{ij} = \delta_0 + \delta_1 U_{ij} + \delta_2 E_{ij} + \delta_3 O_{ij} + \delta_4 I_{ij} + \delta_5 R_{ij} \\ + \alpha_1 U_{ij} \ln P_i + \alpha_2 E_{ij} \ln P_i + \alpha_3 O_{ij} \ln P_i \\ + \beta_1 U_{ij} \ln P_j + \beta_2 I_{ij} \ln P_j + \beta_3 R_{ij} \ln P_j - \gamma U_{ij} \ln D_{ij} + \varepsilon_{ij}\end{aligned} \quad (5)$$

For interpretation, this gravity model equation can also be rewritten in the following form:

$$\begin{aligned}lnM_{ij} = \delta_0 + U_{ij}\left(\delta_1 + \alpha_1 lnP_i + \beta_1 lnP_j - \gamma lnD_{ij}\right) \\ +E_{ij}(\delta_2 + \alpha_2 lnP_i) + O_{ij}(\delta_3 + \alpha_3 lnP_i) \\ +I_{ij}\left(\delta_4 + \beta_2 lnP_j\right) + R_{ij}\left(\delta_5 + \beta_3 lnP_j\right) + \varepsilon_{ij}\end{aligned} \quad (6)$$

Eq. (6) shows that for exclusively inter-urban migration flows ($U_{ij} = 1, E_{ij} = O_{ij} = I_{ij} = R_{ij} = 0$) the model simply reduces to that of Eq. (2).

Note that for estimating Eq. (5) the values assigned to the population abroad, the population of rural areas and the distances between urban areas and abroad, or between urban areas and rural areas, are irrelevant.[5] Figure 1 below shows the origin-destination matrix and the dummy variables accounting for each type of flow.

Figure 1 about here

An important limitation of the current specification of the gravity model is the treatment of zero flows, given the specification of the model in logarithms. Here we set $M_{ij} = 0.5$ where the reported migration flow is 0.[6] Alternative methods, such as excluding zero flows or using count data models,

[5] In the estimation in Stata we have set these values to one. Also note that perfect collinearity is avoided by defining international migration from and to rural areas as the benchmark category of the migration matrix.
[6] To preserve confidentiality, New Zealand census counts are rounded to multiples of 3: an actual count of 0 is reported as such, but an actual count of 1 is rounded down to 0 with probability 2/3 and rounded up to 3 with probability 1/3, with the reverse probabilities for rounding a count of 2. If the low frequencies were uniformly distributed, a rounded value of 0 is therefore reported in 2/3 of the cases rather than 1/3. However, the distribution of low frequencies is unlikely to be uniform, with 0 likely to be much more common than 1 or 2, particularly in migration matrices referring to small areas or relatively small groups. In this case count models that allow for excessive zeros, such as the zero-inflated Poisson model (see e.g. Bohara and Krieg, 1996) would be more appropriate than the simple gravity model.

are also possible to address this problem, but zero flows in our data form only around 5 percent of the total flows in all Census periods. We checked the sensitivity of the results by excluding zero flows from the regressions and found that none of the main results were sensitive to the way we treated zero flows.

Three different measures of distance D_{ij} are used in this study. The first is the amount of time in minutes it would take to travel from the city centre of an origin to the city centre of a destination. This measure is referred to as D_{ij}^{Min}, with the data corresponding estimates of travel time obtained in 2013 and 1984 from Google Maps and the 1984 Mobil Map respectively.[7] There are 1560 (40x39) origin-destination pairs available to estimate distance deterrence with 2013 travel time data. However, the 1984 measures were available only for 25 urban areas, reducing the origin-destination pairs to 600 (25x24). The second measure of distance is distance by road, in kilometres, between the city centre of an origin urban area and destination urban area, as represented by D_{ij}^{Km}. Again, estimates for 2013 and 1984 of this variable were obtained using Google Maps and the 1984 Mobil Map, respectively. The final distance measure is the straight line distance between urban areas, denoted by D_{ij}^{Skm} and calculated as the straight-line distance between population-weighted centroids in origin and destination areas.[8] A description of the variables and summary statistics can be found in Table 1.

Table 1 about here

Three different gravity model specifications, each based on a different measure of distance, were estimated for each of the six censuses from 1986 to 2013 to make a total of 18 regressions. Figure 2 plots the change over time in the distance elasticity of inter-urban migration. The full regressions are reported in Table 2. All the variables are significant at the one per cent level of significance for all time periods. The results from the models show that the specifications with distance measured in minutes yield the best fit. This is plausible given that travel time is economically a better measure of travel cost than travel distance in kilometres. Travel time will reflect the opportunity cost of using that time for other activities. The specification with straight-line distance has the worst fit in terms of migration modelling in all periods, as we would expect.

Figure 2 about here

Table 2 about here

In all periods, the absolute value of the distance elasticity of migration is largest when distance is measured in kilometres. However, distance in minutes yields almost identical distance elasticities. Using straight-line distance suggests elasticities that are about 0.1 lower, again as expected. Regardless of the measure of distance, estimates of the distance elasticity of migration show an increase in all pre-census five-year periods between 1986 and 2001 before decreasing subsequently in 2006 and 2013. The rise in the distance elasticity of migration is against the expectation that the

[7] Manukau city centre was the reference point for the South Auckland urban area, Henderson for West Auckland, North Shore Information centre for North Auckland, Kapiti Coast District Council for Kapiti and Auckland city centre for the Central Auckland urban area.

[8] Geographic centroids were calculated for each meshblock, derived from 2006 meshblock shape files available on the Statistics New Zealand website. Population-weighted means of longitude and latitude were then calculated for each urban area to give the representative location of the urban area.

greater connectivity between urban areas would have reduced the distance deterrence effect over time. We would expect the declining cost and increasing quality of internet-based information exchange to have lowered job and housing search costs and also the psychic cost of being away from one's family and friends. The distance elasticity of migration did decrease between 2001 and 2013, but over the whole 1986-2013 period the elasticity increased in absolute value for all three measures. Two effects may work here in opposite direction. There are strong agglomeration forces at work in New Zealand that have led to relatively fast population growth in the largest city, Auckland, which accounts for about one third of the population. Hence the growth in Auckland's share of the total population would have increased the estimated distance deterrence effect. On the other hand, relatively fast income growth outside Auckland and Wellington after 2001 (see e.g. Alimi et al., 2016) could be responsible for the decline in the estimated elasticity post 2001.

As shown earlier, distance measured in minutes provides the best fit in the gravity model compared with the other two measures. However, travel time and road distance between places are not constant over time. Improvements in transportation technology, new roads and changes in government legislation, such as maximum speed limits, do affect travel time and road distance between places. We examine evidence on the effect of changing distances between specific origin-destination pairs over time, using historical and current road travel distance information. During the decades that correspond to the available migration data there have in fact been significant changes in some road distances and travel times. For example, the 793 km journey from Whangarei to Wellington which currently takes around 9 hours 23 minutes (based on 2013 information) was an 839 km journey that took 15 hours 5 minutes in 1984. For the 600 origin-destination pairs for which we have comparable data, there was about a 30 percent average decline in travel time between 1984 and 2013. It is important to see whether such improvements actually matter for migration. Pooling the 1986-1991 and 2008-2013 migration flows data, as well as the roughly corresponding 1984 and 2013 distance data, we have 1200 observations with which we can run the following two-wave fixed effects panel model regressions (one for distance in time and one for distance in kilometres):

$$\ln M_{ijt} = \delta_* + \alpha_* \ln P_{it} + \beta_* \ln P_{jt} - \gamma_* \ln D_{ijt} + \theta_{ij} + \mu_t + \varepsilon_{ijt} \tag{7}$$

The results from these two regressions are presented in Table 3.

Table 3 about here

The results do not show evidence that reductions in distance (time or kilometres) have increased migration flows. The coefficient on both time-varying measure of distance is even positive, albeit not statistically significant at the 5 per cent level. These results imply that the improvements in connectivity brought about by factors such as upgraded road infrastructure, advances in transportation technology and increases in highway speed limits have not led to increased migration flows. This result is indicative that there could be other factors at work, such as changes in commuting behaviour – with improved connectivity leading to increased long-distance commuting instead of encouraging migration. In any case, the positive distance coefficients in Table 3 are consistent with the upward trend in distance deterrence in Figure 2.

Given the rapid growth in international migration throughout the world, an increasingly important question is the extent to which such gravity models of internal migration are affected by concurrent international migration flows. As noted earlier, New Zealand provides a good case study to investigate this given that one quarter of its population is foreign born. To ensure that all forms of migration are accounted for, we also simultaneously consider migration from and to rural areas. We estimate the effect of urban population push and pull on international and rural-urban migration by means of Eq. (5). The results are reported in Table 4. For brevity, we restrict this estimation to the case of travel distance measured in minutes.[9] Note that, by design, the coefficients related to inter-urban flows, i.e. the coefficients of $U_{ij}\ln P_i$, $U_{ij}\ln P_j$ and $U_{ij}\ln D_{ij}^{Min}$ are the same as the corresponding ones in Table 2 (see also Eq. (6)). These are therefore not further discussed here.

Table 4 about here

By design, the case of all dummy variables being set to zero refers to the migration from abroad to rural areas and vice versa. For example, the predicted value of these two flows is about $e^{9.784}$ = 17,748 over the 1986-1991 period.[10] $E_{ij}\ln P_i$ and $O_{ij}\ln P_i$ measure the *push* effect of origin urban population on emigration and urban to rural migration respectively. Similarly, $I_{ij}\ln P_j$ and $R_{ij}\ln P_j$ estimate the *pull* effect of destination urban population on immigration and rural to urban migration respectively. The coefficients of $O_{ij}\ln P_i$ (urban to rural migration) are much larger than their inward migration counter-parts (i.e. the coefficients of $R_{ij}\ln P_j$). The coefficients of $E_{ij}\ln P_i$ (emigration from urban areas) and $I_{ij}\ln P_j$ are roughly the same, except between 1991 and 1996 and between 2008 and 2013.[11] The urban population elasticities of migration are shown in Figure 3.

Insert Figure 3 here

Several conclusions can be drawn from Figure 3. Firstly, comparing large urban areas with small urban areas (in terms of population), the former experience relatively more migration *to* rural areas and relatively less migration *from* rural areas than the latter (because the estimated coefficients of origin urban population in urban to rural flows are always much larger than those of destination urban population in rural to urban flows). Secondly, the population elasticities of inter-urban migration are always larger than those of urban-rural migration. Hence, larger urban areas generate relatively more inter-urban than urban-rural migration than smaller urban areas. Thirdly, international migration is even more selective of population size, with the largest urban areas generating relatively much more international migration than inter-urban (or rural-urban) migration. This selectivity of migration, with the greatest cross-border mobility rates observed in the large metropolitan areas is a well-known phenomenon globally (e.g. Gorter et al., 1998).

[9] Results of the regressions for the other measures of distance are available from the authors upon request.
[10] The predicted value of net migration between abroad and rural areas is in this model zero.
[11] This means that emigration has been disproportionally more common among the larger urban areas in 1991-1996 and 2008-2013 than among the smaller urban areas. While we should avoid a temporal interpretation of changes in coefficients estimated with successive waves of cross-sectional migration data, it is worthwhile to note that New Zealand has experienced much greater temporal volatility in emigration, relative to population, than in immigration (see e.g. Cochrane and Poot, 2016).

3. Gravity model and subnational population projections

We saw in the previous section that, once correctly calibrated, a simply gravity model is a very effective means of describing patterns in observed gross migration flows. It links populations of subnational areas at the beginning of a period with subsequent inward and outward migration. It is therefore not surprising that the gravity model has been increasingly applied to international migration flows (e.g., Ramos, 2016; Beine et al., 2016; Karemera et al., 2000) while the model continues to be applied to internal migration flows (e.g., Etzo, 2011; Peeters, 2012). While the literature makes it clear that there continue to be econometric challenges once researchers move to more advanced versions of the model that include systemic and dynamic effects combined with many push and pull factors, and spatial spillovers, a relatively underexplored topic is the role which the model might have in a multi-regional population projections methodology. In this section we argue that the gravity model of gross internal migration may be helpful to improve subnational population projection methodologies. International flows and rural-urban flows can also be taken into account as specified in Eq. (5).

The most common method employed for projecting the population (at both the national and subnational levels) is the cohort component model, dating back to the work of Whelpton (1928). The cohort-component model is a stock-flow model that is based on the following fundamental 'accounting identity' of population growth:

usually resident population in area *i* at the *end* of year *t*
= usually resident population in area *i* at the *beginning* of year *t*
+ births to mothers residing in area *i during* year *t*
– deaths of residents of area *i during* year *t*
+ inward migration from other regions and from overseas into region *i during* year *t*
– outward migration of residents from area *i* to other regions or to overseas *during* year *t* (8)

By applying assumed fertility, mortality and migration rates for each of the components (usually by age and sex), the model is then run sequentially one year at a time to project the future population of area *i*. Cohort component models are widely employed because of their simplicity and because they require only projections of future fertility (usually based on an assumed future overall total fertility rate and age-specific fertility profiles), future mortality (usually based on assumed future life expectancy and age-specific survivorship rates), and future migration (either in-migration and out-migration separately, or net migration – and in both cases with either internal and international migration separately or combined).

With respect to migration, the 'conventional' method for projecting future migration at both the national and subnational levels – used by Statistics New Zealand and by many other national statistics agencies – is to assume a certain *level* of net migration in each area in each future year (or five- or ten-year period). This level, plus an assumed age-sex distribution, can be varied across several population projection scenarios. A net migration level approach helps to ensure that net migration in subnational areas satisfies the 'adding-up constraint', i.e. that the sum of net migration of all areas considered is equal to the overall net migration at the national level (which is, by definition, zero for net internal migration in a country). This constraint is necessary because statistical agencies usually adopt a top-down projections approach, whereby the national-level population projection is estimated first, before the set of subnational projections. This approach ensures that the sum of projected subnational populations is indeed equal to the projected national population.

The disadvantage of this conventional method is that it does not take account of the fact that the *volume* of net migration is likely to be related to the *size* of the population, as can be easily seen by combining Eqs. (1) and (3) above. An alternative approach, which explicitly captures this effect, is to project net migration in the form of net migration rates, either at the level of total population or at the level of individual age groups (e.g. Cameron and Poot, 2010; 2011; Cameron et al., 2007). However, the use of net migration rates is problematic for at least two reasons. First, the denominator in the net migration rate is the population of the projected area – for out-migrants this is their origin, but for in-migrants it is their destination – which presents a problem of theoretical inconsistency given that origins and destinations have different roles to play in migration processes, as we saw in the previous section. Second, the net migration rate is silent as to the sources of in-migration (or destinations of out-migration) and is insensitive to changes in surrounding populations which might be expected to impact on migration to and from the area of interest. It is precisely for such reasons that Rogers (1968) developed multiregional mathematical demography with a focus on events and populations that are exposed to the *risk* of experiencing them. Thus, Rogers (2015, p.111) notes that "there is no such individual as a net migrant, and attempts to explain the behaviour of net migrants are likely to lead to misspecified models and biased findings".

Gravity models offer a way of explicitly capturing the influence of the source and destination of in-migration and out-migration respectively. This is important because end users of population projections are increasingly concerned about the 'black box' nature of forecasting, and want reassurance that models are capturing the underlying dynamics of population change in their areas. In our experience, both the conventional method and the net migration rates method have been unable to fully satisfy end users in this regard.

Integrating a gravity model within a population projection modelling framework is a relatively straightforward exercise. Historical data are first used to parameterise the model, as in the previous section. In the simplest of gravity models, this requires only data on origin-destination migration flows, population, and distance. The model can also be extended to include other variables of interest known to influence migration flows, such as income, unemployment rates, migrant stocks, climate, and so on (e.g., see Piras, 2016; Aldashev et al., 2014). The key constraint to using such additional variables in an augmented gravity model is that in order to derive population projections from the model, forecasts (or assumptions about the future values) of these additional variables will be required.

Then, in each time-step of the projections, the origin and destination populations and distance, as well as any other variables included in an augmented gravity model, are used to project dyadic migration flows (from origin to destination). The sum of in-migration flows minus the sum of out-migration flows obtained from the gravity model is net migration. The ability to project directional migration flows, as well as to pinpoint the specific push and pull factors driving those flows (through the parameters in an augmented gravity model), is useful in achieving 'buy-in' from the end users of population projections, since it avoids to some extent the 'black box' problem noted earlier.

International migration should also be incorporated explicitly in the multi-regional population projection system. The internal migration matrix is then augmented by a row of regional immigration and a column of regional emigration. The projected gross internal migration matrix has by definition the property that when adding across all rows and columns, net internal migration is zero for the country. By contrast, the cross-regional sum of emigration in developed countries is of course usually much less than the cross-regional sum of immigration, given that most developed countries gain population through international migration. In a multi-regional projection system immigration can either be exogenously set (which is reasonable when there are strict controls of external borders) or modelled in some way.[12] Emigration can be projected by sub-group specific rates.

To account for differences in migration propensities and spatial distribution by age groups, two approaches are possible. The first is to estimate a gravity model for the entire population and then apply an assumed age profile to inward and outward migration in each of the regions. The age profile is likely to be different for in-migration as compared with out-migration (e.g., consider the difference between the ages of in-migrants and out-migrants in a university town), or might be different for different origin-destination combinations (e.g. consider the difference between the typical ages of rural-urban migrants and rural-rural migrants). The age profiles need not be static over time and may instead be projected as well.

The alternative approach is to estimate gravity models for separate age-sex groups. To our knowledge, this approach has not been attempted. Given that it is likely that different age-sex groups are influenced differently by push and pull factors, this approach offers some promise for improvements in the quality of population projections. Of course, the greater the level of disaggregation of population, the greater the need to account for zero flows in a statistically satisfactory manner.

A number of further extensions are possible. Projecting migration flows by education level is certainly possible, and given that fertility rates are closely related to education levels, projections of migration flows by education level are potentially important (e.g. see the recent work on international migration by level of educational attainment by Samir and Lutz, 2015; and Samir et al., 2010). Push and pull factors are likely to differ between groups with different levels of educational attainment, so augmented gravity models of education-level-specific migration flows offer much promise.

Another promising extension of gravity models is the projection of multi-regional ethnic populations. Ethnic population projections present a challenging case because, unlike race, ethnicity is a fluid concept (Burton et al., 2010). Hence in this case both spatial migration (from origin to destination) as well as inter-ethnic mobility (as people change their ethnic affiliation) should be taken into account in principle. In fact the gravity property may even have some predictive power in the context of inter-ethnic flows (although the situation is different from that of spatial movement in that individuals may have multiple ethnicities but are usually assigned to only one location). In any case, the appropriate measurement of the 'distance' between different ethnicities remains a major

[12] See e.g. Gorbey et al. (1999) for a New Zealand application. The assumption of exogenous levels of immigration is of course relatively more realistic for a remote island nation like New Zealand than for, e.g. the case of the European Union, as the situation regarding refugee migration in recent years has made very clear.

challenge, although some progress has been made in recent years (e.g. Wang et al., 2016). Of course, explicit measures of either spatial or ethnic distance do not need to be observed if such distances are time invariant and successive flow matrices are observed. In that case transition probabilities can be calculated either deterministically or by means of fixed effects models. Using the latter approach, the proposed methods have much in common with those of multiregional mathematical demography (Rogers, 2015). Overall, gravity models offer a highly promising avenue for improving population projection methodology. In our experience, end users appreciate the greater depth of understanding that these models provide.

4. Retrospect and prospect

In this paper we have reviewed how a common workhorse of regional science, namely the gravity model of spatial interaction, continues to be a very effective means of describing gross flows of human migration. Indeed, the gravity model of migration flows has seen somewhat of a comeback in recent years after it was relegated to a less prominent role during the years in which microdata on population mobility became more readily and comprehensively available. A quick check with Google Scholar shows that the number of papers with 'gravity model', or 'gravity' and 'migration', in the title increased by more than 40 percent in the first half of this decade, as compared with the second half of the previous decade.[13]

This increasing interest in the gravity model is not surprising given the growth in availability of dyadic flow data, for example in international migration, but also in long-distance commuting, temporary worker flows and student mobility. Statistically, the development of estimation techniques that account for spatial correlation in spatial interaction matrices (as reviewed e.g. by Patuelli, 2016) may have provided an impetus for new work in this area too.

When estimating an internal migration gravity model for New Zealand, we have found that, perhaps counter-intuitively, distance deterrence has increased over time and we hold agglomeration forces predominantly responsible for this result. At the international level, the trend in the distance deterrence parameter is yet to be assessed. It may well be that the impact of the sharply declining real cost of air travel and communications leads to a growing global dispersion of international migrants and hence, a smaller value of the parameter. On the other hand, regionalism in spatial interaction generally and regional concentration of migration in Europe and the Middle East in particular – such as resulting from the explosive growth in refugee migration in recent years – may well result in an increasing value of the distance deterrence parameter.

As we have argued in this paper, there are three areas where future developments of the gravity model of migration would appear to be particularly promising. One is the linking of migration matrices of cross-border flows and internal flows. The second is the embedding of gravity models in multi-regional population projection systems. The third is the further development of spatial econometric interaction models.

[13] Of course, some of these papers focus on trade or investment flows rather than migration. The increasing popularity of gravity models in economics undoubtedly applies to all kinds of flows (see also Ramos, 2016).

Nonetheless, significant challenges remain. Firstly, the growing complexity of spatial mobility, in which individuals may have more than one residence (think e.g. of children in families with separated parents or older couples with a second home abroad) requires a fresh approach to the notions of residence, mobility and transitions (see also e.g. Poot, 2015). Secondly, the greater availability of very rich mobility data (both in terms of personal characteristics and temporal-spatial patterns of movement) begs the question of the desirable level of disaggregation in gravity modelling. Clearly, disaggregation by migrant type and for small areas would lead to very large migration matrices that may contain many cells with a migration count of zero. Count models that explicitly allow for an inflated number of zeros are essential in that context (see e.g. Burger et al., 2009). However, we would not expect such a model to be appropriate for embedding in small area population projection methodologies, for which a range of other approaches are available such as microsimulation and time-series modelling (see e.g. Wilson and Bell, 2011). In a sense, we could argue that – as in physics – there is as yet no unified theory that captures behaviour both at the very small level as well as at the macro level!

A third challenge will be to revisit systemic approaches, both from the theoretical perspective and from the statistical perspective. Path-breaking exploratory work with spatial econometric interaction models can already be found in LeSage and Pace (2009), but the extension of this work to systemic models, such as Alonso's (1978) theory of movements, are still to be explored, particularly in a dynamical setting. A difficulty in this area is the high dimensionality of the parameter space, given that n dyadic flows may generate up to n^2 potential spatial spillover terms even in the simplest setting. Applications of spatial econometric interaction models are likely to take off once estimation techniques become embedded in common statistical software such as *Stata* or *R*. In this context, the increased availability of network flow measures, such as internet or phone traffic, may open up opportunities to examine dyadic factors other than distance. Migration flows may vary with the strength of interactions, or socio-economic similarity, in ways not captured by standard distance metrics (Beine et al, 2016).

In conclusion, the gravity model of migration may be expected to have many more years of vitality left – both in terms of contributing to a better understanding of human mobility processes, as well contributing to enhanced population projection procedures.

References

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

Alimi, O., Maré, D.C. and Poot, J. (2015) Does distance still matter for internal migration and, if so, how? Evidence from 1986 to 2006. In: PS Morrison (ed.) *Labour, Employment and Work in New Zealand - Proceedings of the LEW16 Conference, November 27&28 2014*. Wellington: School of Geography, Environment and Earth Sciences, Victoria University of Wellington.

Alimi, O., Maré, D.C. and Poot, J. (2016) Income inequality in New Zealand regions. In: P. Spoonley (ed.) *Rebooting the Regions*. Massey University Press, pp. 177-212.

Alonso, W. (1978) A theory of movements. In N.M. Hansen (ed.) *Human Settlement Systems*. Cambridge, Mass.: Ballinger, pp. 197-211.

Anderson, J.E. (2011) The gravity model. *Annual Review of Economics*, 3: 133-160.

Beine, M., Bertoli, S., Fernandez-Huertas Moraga, J. (2016). A practitioners' guide to gravity models of international migration, *The World Economy*, 39(4), 496-512.

Biagi, B., Faggian, A. and McCann, P. (2011) Long and short distance migration in Italy: the role of economic, social and environmental characteristics. *Spatial Economic Analysis* 6(1): 111-131.

Bohara, A.K. and Krieg, R.G. (1996) A zero-inflated Poisson model of migration frequency. *International Regional Science Review* 19(3): 211-222.

Burger, M., Van Oort, F., and Linders, G.J. (2009) On the specification of the gravity model of trade: zeros, excess zeros and zero-inflated estimation. *Spatial Economic Analysis* 4(2): 167-190.

Burton, J., Nandi, A., and Platt, L. (2010). Measuring ethnicity: challenges and opportunities for survey research, *Ethnic and Racial Studies*, 33(8), 1332-1349.

Cameron, M.P. and Poot, J. (2010). A Stochastic Sub-national Population Projection Methodology with an Application to the Waikato Region of New Zealand, *Population Studies Centre Discussion Paper No. 70*, Hamilton: Population Studies Centre, University of Waikato.

Cameron, M.P., and Poot, J. (2011). Lessons from stochastic small-area population projections: The case of Waikato subregions in New Zealand, *Journal of Population Research* 28(2-3), 245-265.

Cameron, M.P., Cochrane, W., and Poot, J. (2007). *End-user informed demographic projections for Hamilton up to 2041*, research report commissioned by Hamilton City Council, Population Studies Centre Discussion Paper No. 66, Hamilton: Population Studies Centre, University of Waikato.

Cochrane, W. and Poot, J. (2016) *Past Research on the Impact of International Migration on the Housing Market: Implications for Auckland*. Report commissioned by MBIE. Wellington: Ministry of Business, Innovation and Employment. http://www.mbie.govt.nz/publications-research/publications/housing-and-property/nidea-report-immigration-housing-literature-review.pdf

Courchene, T. J. (1970). Interprovincial migration and economic adjustment. *Canadian Journal of Economics*, 3(4), 550–576.

Curry L (1972) A Spatial Analysis of Gravity Flows. *Regional Studies* 6 (2):131-47

Cushing, B. and Poot, J. (2004) Cross boundaries and borders: regional science advances in migration modelling. *Papers in Regional Science* 83(1): 317-338.

Davis, K.F., D'Odorico, P., Laio, F. and Ridolfi, L. (2013) Global spatio-temporal patterns in human migration: A Complex Network Perspective. *PLoS ONE* 8(1): e53723.doi: 10.1371 / journal. pone. 0053723.

De Vries, J.J., Nijkamp, P. and Rietveld, P. (2001) Alonso's theory of movements: developments in spatial interaction modeling. *Journal of Geographical Systems* 3(1): 233-256.

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.

Fagiolo, G. and Mastrolillo, M. (2013) International migration network: topology and modelling. *Physical Review E* 88: 012812/1-012812/11.

Fan, C. C. (2005). Modeling interprovincial migration in China, 1985-2000. *Eurasian Geography and Economics*, 46(3), 165–184.

Fik, T.J. and Mulligan, G.F. (1998) Functional form and spatial interaction models. *Environment and Planning A* 30(8): 1497-1507.

Gorbey, S., James, D. and Poot, J. (1999). Population forecasting with endogenous migration: an application to Trans-Tasman migration. *International Regional Science Review* 22(1): 69-101.

Gorter, C., Nijkamp, P. and Poot, J. (eds) (1998) *Crossing Borders: Regional and Urban Perspectives on International Migration*. Ashgate, Aldershot UK. Pp. 376.

Greenwood, M.J. (1997) Internal migration in developed countries. In M.R. Rosenzweig and O. Stark (eds.) *Handbook of Population and Family Economics*, *Volume 1B*. Amsterdam: Elsevier.

Karemera, D., Oguledo, V. I., & Davis, B. (2000). A gravity model analysis of international migration to North America. *Applied Economics*, 32(13), 1745–1755.

LeSage, J. and Pace, R.K. (2008) Spatial econometric modelling of origin-destination flows. *Journal of Regional Science* 86(3): 393-421.

LeSage, J. and Pace, R.K. (2009) Spatial econometric interaction models. Chapter 9 in *Introduction to Spatial Econometrics*. Boca Raton: CRC Press.

Lewer, J. J., & H. Van den Berg (2008). A gravity model of immigration. *Economics Letters* 99(1), 164–167.

Lomax, N., Norman, P., Rees, P. and Stillwell, J. (2013) Subnational migration in the United Kingdom: producing a consistent time series using a combination of available data and estimates. *Journal of Population Research* 30: 265-288.

Mayda, A.M. (2010) International migration: a panel data analysis of the determinants of bilateral flows. *Journal of Population Economics* 23: 1249-1274.

Mayer, G. and Vyborny, M. (2008) Internal migration between US States: A social network analysis. In J. Poot, B. Waldorf and L. van Wissen (eds) *Migration and Human Capital*. Cheltenham UK: Edward Elgar.

Newton, I.S. (1687) *Philosophiæ Naturalis Principia Mathematica*. London: Royal Society Press.

Nijkamp, P. and Poot, J. (1987) Dynamics of generalised spatial interaction models. *Regional Science and Urban Economics* 17(3): 367-390.

Patuelli, R. (2016) Spatial autocorrelation and spatial interaction. In: S. Shekhar, H. Xiong and X. Zhou (eds.) *Encyclopedia of GIS*. Springer Verlag.

Peeters, L. (2012) Gravity and spatial structure: the case of interstate migration in Mexico. *Journal of Regional Science* 52(5): 819-856.

Piras, R. (2016 in press). A long-run analysis of push and pull factors of internal migration in Italy. Estimation of a gravity model with human capital using homogeneous and heterogeneous approaches, *Papers in Regional Science*, online early.

Poot, J. (1986) A system approach to modelling the inter-urban exchange of workers in New Zealand. *Scottish Journal of Political Economy* 33(3): 249-274.

Poot, J. (1995) Do borders matter? A model of interregional migration in Australasia. *Australasian Journal of Regional Studies* 1(2): 159-182.

Poot, J. (2015) Cross-border migration and travel: a virtuous relationship. *IZA World of Labor*, doi: 10.15185/izawol.209.

Ramos, R. (2016) Gravity models: a tool for migration analysis. *IZA World of Labor*, doi: 10.15185/izawol.239

Ravenstein, E.G. (1885, 1889) The laws of migration, Part 1 and Part 2. *Journal of the Royal Statistical Society*, 48(2): 167-235 and 52(2): 241-305.

Raymer, J., Abel, G.J. and Rogers, A. (2012) Does specification matter? Experiments with simple multiregional probabilistic population projections. *Environment and Planning* A 44(11): 2664-2686.

Rogers, A. (1968) *Matrix Analysis of Interregional Population Growth and Distribution*. University of California Press.

Rogers, A. (2015) *Applied Multiregional Demography: Migration and Population Redistribution*. Berlin: Springer Verlag.

Samir, K.C., and Lutz, W. (2015 in press). The human core of the shared socioeconomic pathways: Population scenarios by age, sex and level of education for all countries to 2100, *Global Environmental Change*, online early.

Samir, K.C., Barakat, B., Goujon, A., Skirbekk, V., Sanderson, W., and Lutz, W. (2010). Projection of populations by level of educational attainment, age, and sex for 120 countries for 2005-2050, *Demographic Research* 22(15), 382-472.

Stewart, Q.J. (1950) The development of social physics. *American Journal of Physics* 18: 239-253.

Stillwell, J., Daras, K., Bell, M. and Lomax, N. (2014) The IMAGE Studio: a tool for internal migration analysis and modelling. *Applied Spatial Analysis and Policy* 7(1): 5-23.

Tranos, E., Gheasi, M. and Nijkamp, P. (2015) International migration: a global complex network. *Environment and Planning B: Planning and Design* 42: 4-22.

Wang, Z., De Graaff, T., and Nijkamp, P. (2016) Cultural Diversity and Cultural Distance as Choice Determinants of Migration Destination. *Spatial Economic Analysis* 11(2): 176-200.

Whelpton, P. K. (1928). Population of the United States, 1925 to 1975, *American Journal of Sociology*, 34(2), 253-270.

Wilson, A.G. (1970) *Entropy in Urban and Regional Modelling*. London: Pion.

Wilson, T. and Bell, M. (2011) Editorial: advances in local and small area demographic modelling. *Journal of Population Research* 28: 103-107.

Zipf, G.K. (1946) The P1 P2/D hypothesis: On the intercity movement of persons. *American Sociological Review*, vol. 11, Oct.

Table 1: Summary statistics by census period

Census Migration Period	Variable	Number of observations	Mean	Std. Dev.	Min	Max
1981-1986	Population (P_i and P_j)	40	23,890	28,335	1,821	113,250
	Inter-urban migr. (M_{ij})	1,560	81	311	0	6,648
	All migration (M_{ij})	NA	NA	NA	NA	NA
1986-1991	Population (P_i and P_j)	40	26,373	32,003	2,484	129,483
	Inter-urban migr. (M_{ij})	1,560	89	364	0	7,647
	All migration (M_{ij})	1,722	259	1,177	0	24,846
1991-1996	Population (P_i and P_j)	40	29,614	37,318	3,879	155,658
	Inter-urban migr. (M_{ij})	1,560	86	361	0	7,197
	All migration (M_{ij})	1,722	250	1,196	0	26,814
1996-2001	Population (P_i and P_j)	40	31,138	40,337	4,185	168,867
	Inter-urban migr. (M_{ij})	1,560	95	400	0	8,289
	All migration (M_{ij})	1,722	309	1,550	0	29,817
2001-2006	Population (P_i and P_j)	40	33,037	44,137	4,098	184,146
	Inter-urban migr. (M_{ij})	1,560	99	436	0	9,393
	All migration (M_{ij})	1,722	335	1,738	0	37,551
2008-2013	Population (P_i and P_j)	40	33,550	45,589	3,732	193,188
	Inter-urban migr. (M_{ij})	1,560	92	440	0	8,937
	All migration (M_{ij})	1,722	330	1,744	0	32,430
Time invariant measures						
	Straight line distance in kilometres D_{ij}^{Skm}	1560	417	284	12	1288
Time variant measures						
	2013 Travel time in minutes D_{ij}^{Min13}	1,560 (600)	481 (553)	346 (357)	14 (41)	1,440 (1440)
	2013 Road Travel distance in kilometres D_{ij}^{Km13}	1,560 (600)	568 (649)	391 (406)	10 (48)	1,784 (1784)
	1984 Travel time in minutes D_{ij}^{Min84}	600	757	465	55	2,035
	1984 Road Travel distance in kilometres D_{ij}^{Km84}	600	664	414	50	1,852

Observations for population are those for 40 urban areas. For migration flows and distance measures, there are 1,560 (=40*39) observations, i.e. the origin-destination pairs formed from these 40 areas. Current travel time in minutes and current road travel distance in kilometres are distance measures between origin and destination obtained from Google Maps in 2013 (except for Queenstown and Rangiora which were obtained in 2016). The 1984 distance measures are the travel time in minutes and road travel distance in kilometres between origin and destination obtained from the 1984 Mobil map, with data only available for 25 urban areas. Current travel time and distance in parentheses are the corresponding current travel time and distance for the journeys for which travel time and distance were available in 1984.

Table 2: Estimated gravity models (three measures of current distance, six periods)

Variables	1981-1986			1986-1991			1991-1996		
	D_{ij}^{Min}	D_{ij}^{Km}	D_{ij}^{Skm}	D_{ij}^{Min}	D_{ij}^{Km}	D_{ij}^{Skm}	D_{ij}^{Min}	D_{ij}^{Km}	D_{ij}^{Skm}
ln P_i (Population of	0.913***	0.919***	0.930***	0.899***	0.906***	0.918***	0.921***	0.928***	0.940***
origin)	[0.0182]	[0.0184]	[0.0201]	[0.0178]	[0.0180]	[0.0200]	[0.0182]	[0.0185]	[0.0208]
ln P_j (Population of	0.914***	0.921***	0.931***	0.936***	0.944***	0.954***	0.922***	0.930***	0.941***
destination)	[0.0182]	[0.0184]	[0.0201]	[0.0178]	[0.0180]	[0.0200]	[0.0182]	[0.0185]	[0.0208]
ln D_{ij}^{Min} (Travel	-0.706***			-0.775***			-0.813***		
time in minutes)	[0.0205]			[0.0200]			[0.0207]		
Ln D_{ij}^{Km} (Travel		-0.714***			-0.784***			-0.819***	
time in kilometres)		[0.0213]			[0.0209]			[0.0216]	
Ln D_{ij}^{Skm} (Straight			-0.623***			-0.690***			-0.703***
line distance)			[0.0237]			[0.0236]			[0.0249]
Constant	-10.37***	-10.31***	-11.25***	-10.17***	-10.10***	-11.10***	-10.31***	-10.27***	-11.39***
	[0.286]	[0.291]	[0.314]	[0.281]	[0.287]	[0.314]	[0.290]	[0.297]	[0.330]
Observations	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560
R-squared	0.808	0.804	0.766	0.823	0.818	0.776	0.819	0.812	0.762

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 2 continued

Variables	1996-2001			2001-2006			2008-2013		
	D_{ij}^{Min}	D_{ij}^{Km}	D_{ij}^{Skm}	D_{ij}^{Min}	D_{ij}^{Km}	D_{ij}^{Skm}	D_{ij}^{Min}	D_{ij}^{Km}	D_{ij}^{Skm}
ln P_i (Population of	0.883***	0.890***	0.900***	0.931***	0.937***	0.947***	0.944***	0.950***	0.959***
origin)	[0.0173]	[0.0176]	[0.0197]	[0.0166]	[0.0167]	[0.0190]	[0.0168]	[0.0168]	[0.0188]
ln P_j (Population of	0.937***	0.944***	0.954***	0.831***	0.838***	0.846***	0.844***	0.851***	0.858***
destination)	[0.0173]	[0.0176]	[0.0197]	[0.0166]	[0.0167]	[0.0190]	[0.0168]	[0.0168]	[0.0188]
ln D_{ij}^{Min} (Travel	-0.834***			-0.786***			-0.757***		
time in minutes)	[0.0201]			[0.0197]			[0.0200]		
Ln D_{ij}^{Km} (Travel		-0.845***			-0.804***			-0.775***	
time in kilometres)		[0.0209]			[0.0203]			[0.0206]	
Ln D_{ij}^{Skm} (Straight			-0.747***			-0.687***			-0.673***
line distance)			[0.0239]			[0.0236]			[0.0235]
Constant	-9.928***	-9.840***	-10.85***	-9.639***	-9.510***	-10.60***	-10.24***	-10.11***	-11.08***
	[0.278]	[0.284]	[0.315]	[0.268]	[0.271]	[0.305]	[0.270]	[0.273]	[0.302]
Observations	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560	1,560
R-squared	0.831	0.826	0.780	0.829	0.828	0.776	0.827	0.826	0.783

Standard errors in brackets
*** p<0.01, ** p<0.05, * p<0.1

Table 3: Fixed effect model estimating the effect of time variation in distance measures

Variables	Log of migration flows	Log of migration flows
ln P_i (Population of origin)	1.132***	1.132***
	[0.0913]	[0.0914]
ln P_j (Population of destination)	0.695***	0.694***
	[0.0913]	[0.0913]
ln D_{ij}^{Min} (Time varying measure of	0.111	
distance-Minutes)	[0.189]	
Time dummy (μ_t)	-0.533***	-0.570***
	[0.0748]	[0.0375]
ln D_{ij}^{km} (Time varying measure of		0.0231
distance-Kilometres)		[0.255]
Constant	-15.06***	-14.49***
	[1.765]	[2.077]
Observations	1,200	1,200
R-squared	0.312	0.312
Number of id	600	600
Fixed effect	Yes	Yes
Standard errors in brackets. *** p<0.01, ** p<0.05, * p<0.1		

Table 4: The impact of incorporating international and rural-urban flows

VARIABLES	1986-1991	1991-1996	1996-2001	2001-2006	2008-2013
$U_{ij} \ln P_i$ (Origin pop. in	0.899***	0.921***	0.883***	0.931***	0.944***
inter-urban migration)	[0.0172]	[0.0195]	[0.0168]	[0.0161]	[0.0174]
$E_{ij} \ln P_i$ (Origin pop. in	1.330***	2.073***	1.328***	1.301***	1.726***
emigration)	[0.107]	[0.121]	[0.105]	[0.101]	[0.108]
$O_{ij} \ln P_i$ (Origin pop. in	0.719***	0.717***	0.717***	0.679***	0.648***
urban to rural migr.)	[0.107]	[0.121]	[0.105]	[0.101]	[0.108]
$U_{ij} \ln P_j$ (Destination	0.936***	0.922***	0.937***	0.831***	0.844***
pop. in inter-urban mig)	[0.0172]	[0.0195]	[0.0168]	[0.0161]	[0.0174]
$I_{ij} \ln P_j$ (Destination	1.286***	1.272***	1.332***	1.231***	1.211***
pop. in immigration)	[0.107]	[0.121]	[0.105]	[0.101]	[0.108]
$R_{ij} \ln P_j$ (Destination	0.555***	0.547***	0.564***	0.516***	0.517***
pop. in rural to urban)	[0.107]	[0.121]	[0.105]	[0.101]	[0.108]
$U_{ij} \ln D_{ij}^{Min}$ (Travel	-0.775***	-0.813***	-0.834***	-0.786***	-0.757***
time in minutes)	[0.0194]	[0.0222]	[0.0194]	[0.0191]	[0.0207]
Intercept terms					
Constant	9.784***	9.629***	9.788***	9.621***	9.878***
	[0.498]	[0.568]	[0.497]	[0.489]	[0.531]
U_{ij} (Inter-urban	-19.95***	-19.94***	-19.72***	-19.26***	-20.12***
dummy)	[0.568]	[0.647]	[0.565]	[0.554]	[0.600]
E_{ij} (Emigration from	-15.96***	-24.86***	-15.77***	-15.64***	-20.19***
urban area dummy)	[1.151]	[1.313]	[1.137]	[1.102]	[1.187]
O_{ij} (Urban to rural	-9.720***	-9.628***	-9.823***	-9.160***	-9.257***
migration dummy)	[1.151]	[1.313]	[1.137]	[1.102]	[1.187]
I_{ij} (Immigration to	-15.53***	-15.12***	-15.93***	-14.32***	-14.54***
urban area dummy)	[1.151]	[1.313]	[1.137]	[1.102]	[1.187]
R_{ij} (Rural to urban)	-8.290***	-8.262***	-8.450***	-7.731***	-8.178***
migration dummy	[1.151]	[1.313]	[1.137]	[1.102]	[1.187]
Observations	1,722	1,722	1,722	1,722	1,722
R-squared	0.880	0.850	0.887	0.888	0.874
Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1					

Origin-Destination Matrix	Destination							
Origin	UA 1	UA 2	UA 3	UA 4	……..	UA 40	Rural	International
UA 1	X						Urban to rural migrants (O_{ij}=1)	Emigrants from UA ($E_{ij} = 1$)
UA 2		X						
UA 3	Inter-urban flows (U_{ij}=1)							
UA 4								
⋮								
UA 40						X		
Rural	Rural to urban migrants ($R_{ij} = 1$)						X	Emigration from rural areas $U_{ij} = O_{ij} = E_{ij} = R_{ij} = I_{ij} = 0$
International	Immigrants to UA ($I_{ij} = 1$)						Immigration to rural areas $U_{ij} = O_{ij} = E_{ij} = R_{ij} = I_{ij} = 0$	X

Note: X refers to the main diagonal, which represents intra-area mobility. This is not observed and excluded from the analysis.

Figure 1: Origin-destination matrix and dummy variables signalling inter-urban, urban-rural and international migration flows

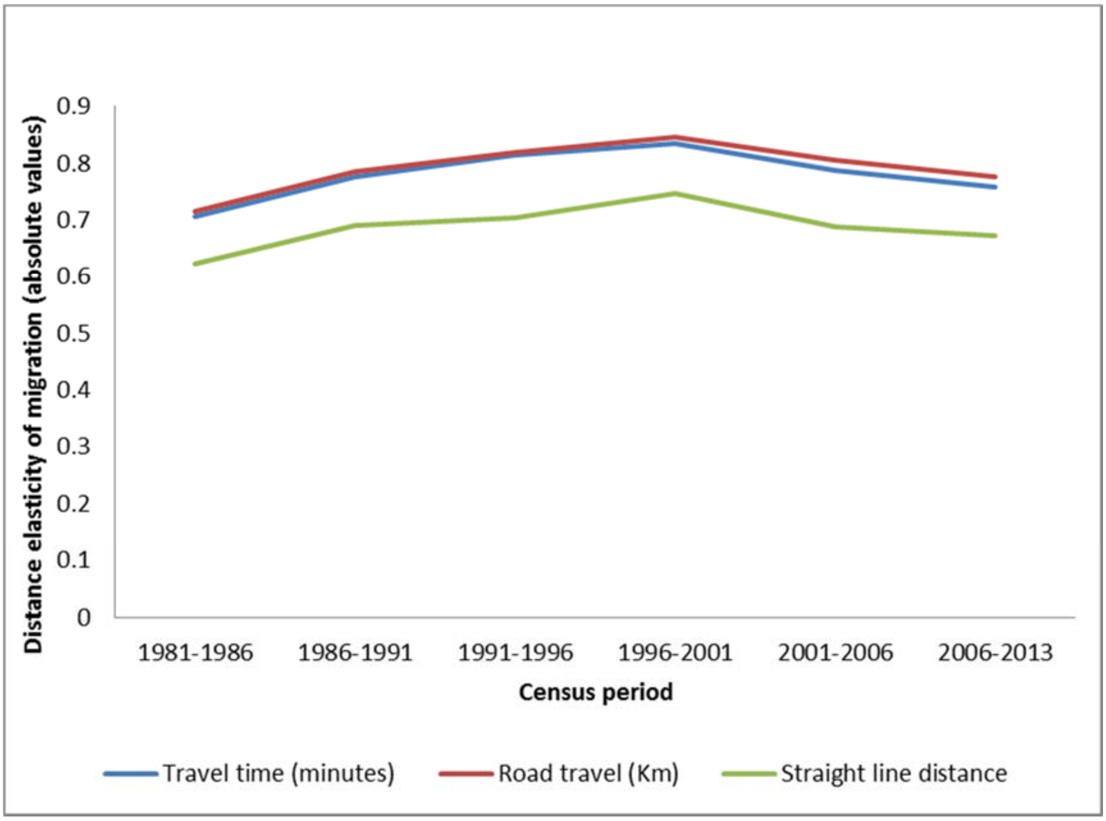

Figure 2: Estimates of the distance elasticity of inter-urban migration for each five-year period preceding the New Zealand population censuses between 1986 and 2013

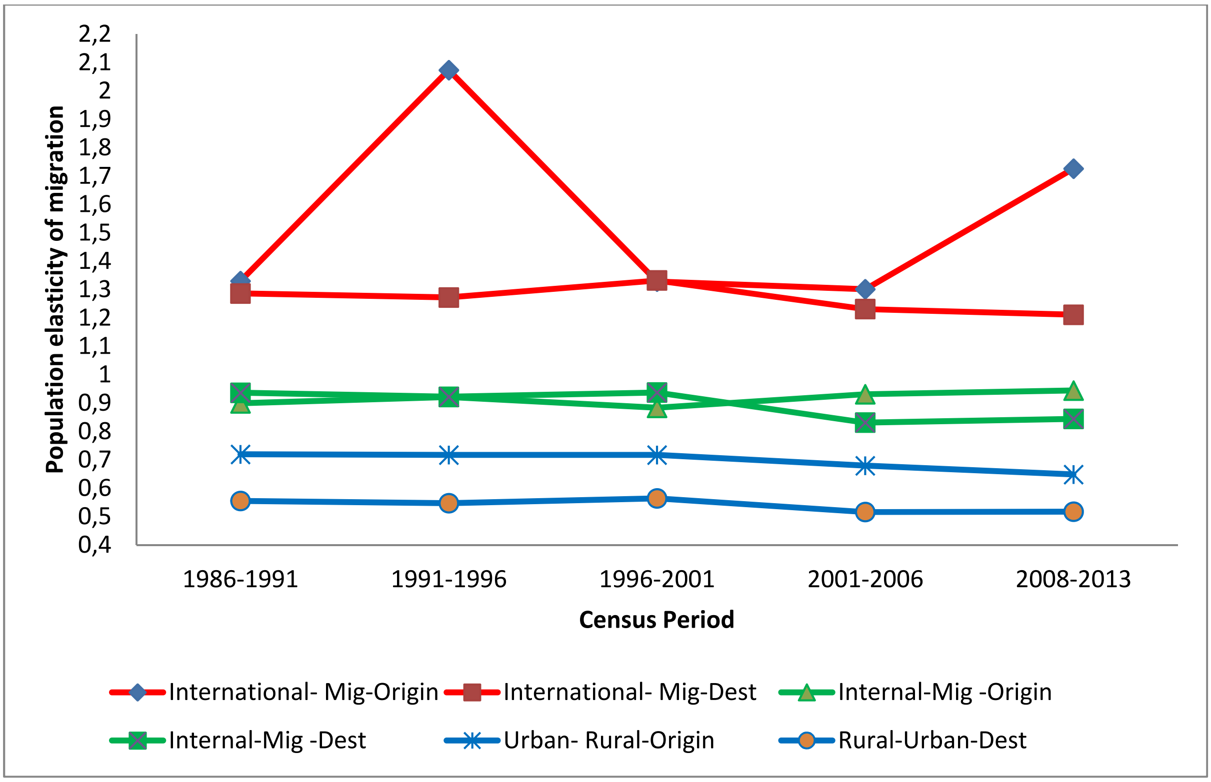

Figure 3: Origin and destination urban population elasticity of migration for international, inter-urban and urban-rural flows