2D Gravity Model Estimation on Tourism Panel Data: Case of New Zealand

By

Nikita Malykhin

Submitted to

Central European University

Department of Economics

In partial fulfillment of the requirements for the degree of Master of Arts in Economics

Supervisor: Professor Gábor Kőrösi

Budapest, Hungary

2017

Abstract

The main focus of the paper is to determine key factors influencing tourism flow's size to New Zealand. A gravity model, which is usually applied in trade theory, is used. Unlike most articles in the field using log-linearization on cross-sectional data, the study implements Poisson pseudo-Maximum Likelihood estimator for panel data on several types of tourism flows (arrivals, departures, bilateral flows). The PPMLE is employed with the multiplicative form of the gravity equation rather than its log-linearized version, due to recent critic of the latter approach by various researchers. Overall results support usual findings on positive relationship between economy sizes and tourism flows, while big distance is a strong barrier for visitors. Additionally, an evidence on significant differences between holiday and business trips preferences is presented.

Keywords: gravity model, tourism flows, PPML.

JEL classification: Z32.

Table of Contents

List of Tables

Introduction

Tourism industry becomes a "feeding" sphere of economy for some countries. With the overall growth of average people's wealth per capita, improvements and spread of technologies, better working conditions and a change in preferable leisure activities, tourism flows have increased significantly (Matias, 2004). Tourism sector importance is not limited to value added to countries' Gross Domestic Product (GDP). There are other direct and indirect impacts on the economy as well: from creating workplaces and supplying additional taxes to the government to overall improvement of society's wealth and intercultural exchange. Santana-Gallegoa, Ledesma-Rodriguez, and Perez-Rodriguez (2016) show that tourism industry growth positively influences various sides of economies. Hence, sometimes the optimal strategy for countries is to open borders and increase tourism inflow to get all the benefits of it (Alawin & Abu-Lila, 2016).

Tourism activity in New Zealand was growing with the world-average rates until the beginning of 00's. Between 1980 and 1990 several entertainments were introduced such as bungy jumping, sky diving, black water rafting. Ecotourism had also become popular among visitors, such as whale watch and Maori cultural experience. However, there were still low awareness around the World about opportunities in New Zealand, overall tourism flow was slightly more than 0.9 mln people a year. That has changed by the end of 90's: during a year of 2000 more than 1.7 mln people arrived and the growth rate of 7% a year continued for half a decade, compared to only 3% of world-average (McClure, 2010)(see Figure 1[1]). Those radical changes happened due to a change in Tourism New Zealand (government tourism agency) policy, when a usual "one country - one advertisement" approach was changed to a global campaign "100% Pure New Zealand".

The new concept implied reaching high level of publicity on one source - official tourism site of New Zealand. In the next several years mobile apps were developed, YouTube channel opened, a number of pages on international tourism sites were created, while all these sources

[1]The figure shows the number of guest nights by international and domestic visitors in January of each year. The statistics on total number of guest nights per year was not collected before 2007. Notice that the growth of domestic tourism flows was smaller than the international, which means that New Zealand's tourism agency policy on promoting the country as a resort place, made a change.

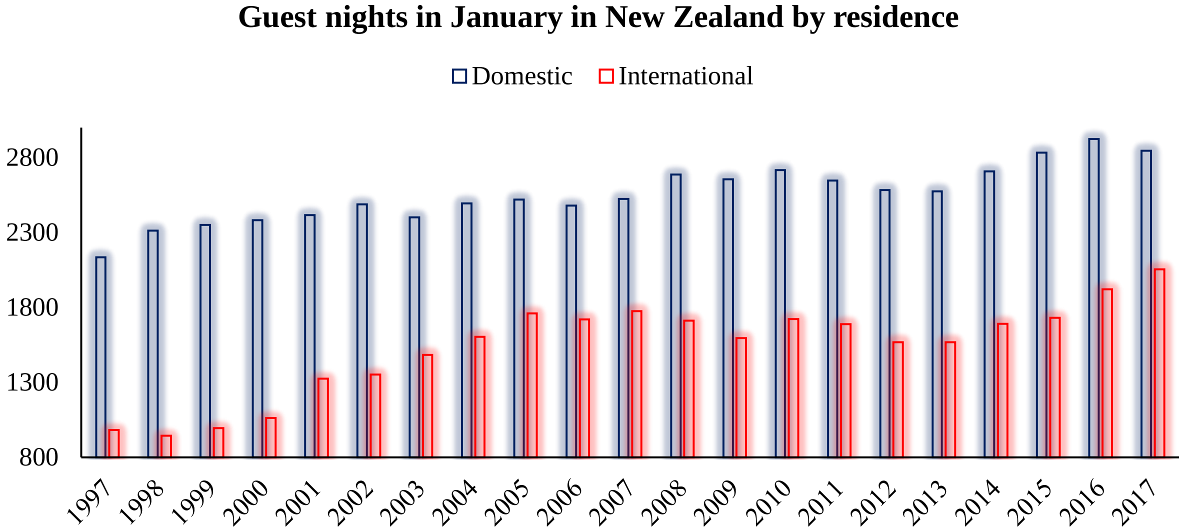

Figure 1: Tourism flows to New Zealand in January of each year (in '000 people)

were linked to one place - New Zealand official site. Additionally to that policy, tourism industry also benefited from establishing very successful the America's Cup yacht races and developing wine and film industries. A multiple Oscar award winner the *Lord of the rings* trilogy was completely shooted in New Zealand, which has created worldwide popularity for the country in 2001 – 2003 years. Due to all these factors the number of visits by 2010 year was already more than 2.5 mln, while in 2016 the flow increased to 3.5 mln people a year (see the biggest tourism markets on Figure 2 in the Appendix).

To more efficiently target tourist groups, an additional analysis is needed. With the development of data collection techniques and open data availability, tourism has become a popular research area in social sciences. Findings in these studies should help the government and tourism agencies to identify the most crucial factors influencing the size of tourist flows to the country, determine places of attraction, find existing problems in transportation network and infrastructure as a whole. This knowledge will be used for optimal tourism sector development, increasing its contribution to the economy.

In this paper I uncover the determinants of tourism flows to New Zealand and compare results with findings for other countries, which has not been done before. To my knowledge, besides governmental reports on tourism activity in New Zealand[2], there exist only three papers

[2] It has been a decade since the ministry of business, innovation and employment of New Zealand has developed tourism flows models aimed to help stakeholders analyze the impact of tourism growth in the country on its

studying tourism flows in the country. Law, Bryant, and Genc (2009) investigated migration flows to New Zealand, showing that it is positively linked to international trade activity and tourism flows with home-countries. Later, using the same data, a gravity model evaluating the size of the migration on tourism flows to New Zealand was proposed by Genç (2010). The findings of the author are consistent with the expectations driven by theoretical setting[3], but the work on paper was not continued. Unlike the former article, I apply different estimation technique, while contrast to the latter, I control for countries' heterogeneity and visa-agreements.[4]

In this study I apply gravity model to tourism panel data, while the majority of articles use this approach for cross-sectional data. Application of the model to cross-sectional data suffers from possible production of biased estimates because of the countries' heterogeneity. To deal with that problem a number of papers estimating the gravity model on panel data were published in the field of international trade (e.g.Burger, van Oort, & Linders, 2009; Egger, 2002; Linders & de Groot, 2006) as well as tourism research (e.g.Alawin & Abu-Lila, 2016; Santeramo & Morelli, 2015; Saray & Karagöz, 2010). The setting of gravity model I use in the paper is slightly different from usual three-dimensional panels, where a regression includes information on trade/flows/FDI/etc. varying on: 1. time scale, 2. exporter scale, 3. importer scale. The model I estimate in the paper is restricted to 2D type - in the network of tourism flows New Zealand represents "a central node" connected to all other countries, but there are no links between them. In other words, the data varies only in two scales - time and exporter (in case of arrivals estimation) or time and importer (in case of departures estimation). Notwithstanding such mutation of the model, the data can be characterized by the same features and standard estimation techniques are applicable as for the 3D case.

The usual way to estimate gravity equation is to get it's log-linearized transformation and employ classical estimation methods such as ordinary least squares (OLS) for cross-sectional data and fixed effects (FE), random effects (RE) for panel data. However, logarithmic transformation not only suffers from zero tourism flows, but may lead to biased estimation results due

infrastructure.

[3]The third paper estimated energy use in air travels (Becken, 2002) through the tourism flows and only indirectly relates to the field of this paper.

[4]Moreover, the authors interpolated observations for 20 out of 25 years in the data set they use, which may lead to possible bias in the estimates. In this thesis I do not have such shortcoming.

to heteroscedasticity in the error term. This conclusion comes from Jensen's inequality (Jensen, 1906), according to which the expected value of a logarithm of a random variable does not equal to the logarithm of expected value (Bobkova, 2012). Silva and Tenreyro (2006) show that the OLS assumptions on conditional expected value of the logarithm of an error term is generally violated if the non-specific heteroscedasticity takes place. To solve both problems described above, a Poisson pseudo-Maximum Likelihood (PPML) estimator can be applied directly to the multiplicative form of the gravity equation (Silva & Tenreyro, 2006). This approach was later proved to be more optimal than traditional log-linearization also by Westerlund and Wilhelmsson (2009). I will follow the same path in this thesis and estimate the model on several types of tourism flows (arrivals, departures, bilateral flows).

The paper is organized in the following way: in the next section most relevant literature in the field is covered. Then, section 3 provides the description of the methodology used for the analysis - the form of the equation and PPML estimator features. Afterwards a data description section comes with sources overview. The 5^{th} section uncovers the results of the estimation and robustness tests. Finally, conclusion is presented in the last section.

Literature overview

Approaches to estimate gravity equation

Since the introduction of gravity model concept in the beginning of the second half of the XX century by Tinbergen (1962) a significant number of papers were published in this subject. Most of the applications of the model are related to international trade relationships, where the amount of trade value between two countries is explained by their economy sizes (usually proxied by GDP) and the distance in between. Gravity model became popular due to its high explanatory power with relatively simple estimation procedure.

First micro-foundations for applying gravity equation in economics rather than physics under strong simplifying assumptions were proposed by Anderson (1979) (later the concept was developed in Anderson & Wincoop, 2003) and the model started to be widely used in trade theory. After that the equation found its applications as well in the fields of tourism (e.g. Khadaroo

& Seetanah, 2007; Massidda, Etzo, & Piras, 2015; Santeramo & Morelli, 2015), migration, foreign direct investment modeling (Morley, Rossello, & Santana-Gallego, 2014) and others.

Even if this paper focuses on estimating coefficients of tourism flows structural factors, it uses the achievements on estimation techniques published in other fields as well, mostly - trade. There are three common approaches applied while working with gravity equation:

- log-linearization of the equation - the most widespread approach allowing to use straightforward estimation techniques such as *OLS* and *FGLS* (for example Baldwin & Di Nino, 2006; Linders & de Groot, 2006; Martin & Pham, 2008; Martinez-Zarzoso, Nowak-Lehmann, & Vollmer, 2007; Westerlund & Wilhelmsson, 2009), *Tobit* (for example Anderson & Marcouiller, 2002; Baldwin & Di Nino, 2006; Martin & Pham, 2008), *Heckman two-step procedure* (for example Linders & de Groot, 2006; Martin & Pham, 2008), Panel *FE* and *RE* (for example Andrews, Schank, & Upward, 2006; Egger, 2000; Mátyás, 1998);

- direct equation estimation from multiplicative form: *Non-Linear Least Squares* (NLS) (for example Silva & Tenreyro, 2006), *PPML* (for example Gomez-Herrera, 2013; Martinez-Zarzoso et al., 2007; Silva & Tenreyro, 2006; Westerlund & Wilhelmsson, 2009), *GPML* (for example Manning & Mulahy, 2001; Martinez-Zarzoso et al., 2007) are applied;

- combination of panel data techniques with time-series estimation techniques such as controlling for *GARCH* (Alawin & Abu-Lila, 2016; Simakova & Stavarek, 2015).

Each of the mentioned methods and techniques has its advantages and drawbacks. Due to its novelty there are not so many papers using the third approach yet. Hence, in the next subsections the discussion about applicability of first two approaches will take place.

Logarithmic transformation of gravity equation

In an overwhelming share of all papers using gravity equation a log-linearization is applied. This allows a researcher to get the equation in a linear form and use straightforward linear models to estimate the coefficients. However, to use a logarithm a "zero values" problem should be solved first. Moreover, a number of studies found such transformation threatening for obtaining consistent results in the presence of omitted variable bias and heteroscedasticity in the error term. Silva and Tenreyro (2006) test this by using Monte-Carlo simulations and comparing the results of applying basic OLS and PPML estimators. Even if the regression is correctly specified, under heteroscedasticity in the error term the results of the least squares estimator on log-linearized equation are severely inconsistent. The same result was later found by Westerlund and Wilhelmsson (2009).

As Silva and Tenreyro (2006) found out, a number of conditions have to be satisfied to obtain consistent estimator. First of all, they derived that for the multiplicative form of gravity equations the error term should be completely independent of other regressors. Secondly, the error term should be exponentially distributed with an exact degree, which is a strong assumption and generally does not hold in real data (more on this is discussed in the section Methodology). Then after log-linearization the results of OLS estimator will be consistent. Authors notice that, specifically independence is a crucial assumption for the error term, because otherwise after the log-linearization the logarithm of the error term will be correlated with regressors of the model.

Silva and Tenreyro (2006) test the influence of the presence of heteroscedasticity in the error term on estimates of different types of methods. They have generated pairwise trade data with the characteristics usually present in real observations. After they applied OLS, Tobit, NLS, PPML estimators, they have found that PPML estimates are remarkably different from those obtained with OLS or Tobit, but more or less close to NLS. Moreover, the results of PPML on the whole sample (zeroes included) and the sub-sample where countries' pairs with zero trade flows are dropped are almost the same[5], while the coefficients of OLS on those two samples[6] are significantly different. It means that the form of heteroscedasticity in the data, rather than truncation, is the key driver of differences between PPML and OLS.

Based on all the findings Silva and Tenreyro (2006) claim that it is not advisable to use log-linearization for estimation of gravity model and suggest estimating it directly from the multiplicative form with the use of either PPML or NLS. As it will be more precisely shown in the Methodology section, the NLS estimator puts more emphasis on those observations having

[5]Authors explain that truncation has little effect on results because for the pairs with zero flows the estimated trade flow by the model is close to zero. Hence, it will make very little effect on the coefficients values.

[6]To estimate the OLS on log-linearized equation with the zero trade flows the $ln(1+x)$ transformation were applied.

higher values. Unluckily, these observations are usually the ones having higher variance. Due to this imbalance the NLS estimator is expected to give very inefficient results in the presence of heteroscedasticity. Since the PPML gives the same weights to all observations, the coefficient estimates appear more efficient than the ones obtained with NLS. Additionally to that, if the conditional mean is indeed can be specified in the form of an exponential function then the estimator is consistent even if the data is not Poisson distributed.

Despite the very positive conclusion on using PPML with gravity model made by Silva and Tenreyro (2006), Martin and Pham (2008) express concerns about the optimality of that estimator. Authors estimate several models such as truncated OLS, Tobit models, Heckman two-step procedure on simulated data and argue that basic truncated OLS on log-linearized data provide better results than PPML estimator. However, later Silva and Tenreyro (2011) in their paper argue that the evidence Martin and Pham (2008) provided is not valid since the authors were using non-constant income elasticity model while generating data, meaning that the results of Silva and Tenreyro (2006) and Martin and Pham (2008) should not be compared. One more evidence on differences in favor of PPML estimates compared to OLS is given by Siliverstovs and Schumacher (2009), who analyze trade flows in OECD countries with both techniques. The findings are very close to those obtained by Silva and Tenreyro (2006).

The work of Burger et al. (2009) also suggests using PPML directly for the multiplicative form of gravity equation to estimate coefficients on structural factors. Additionally to Silva and Tenreyro (2006)'s approach, they argue in favor of using *negative binomial* and *zero-inflated* models (for example, used in Greene, 1994; Lambert, 1994; Long, 1997) of PPML estimator, which shortly are called as NBPML and ZIPPML, respectively. Authors claim that these approaches are viable alternatives to usual PPML in the two following cases:

- basic PPML assumes the equidispersion, i.e. the conditional mean of the dependent variable has to be equal to its conditional variance. However, in trade and tourism data conditional variance is usually higher than the conditional mean, which is called the overdispersion. Greene (1994) claims that usually the dispersion in the data is higher than it is predicted by the model because of the unobserved heterogeneity coming from omitted variables. PPML estimator does account only for observed heterogeneity. Luckily, in the presence of unobserved heterogeneity the estimator still gives unbiased results. However, the PPML gives inefficient estimation of the dependent variable in the sense that a researcher obtains spuriously big z-values and spuriously small p-values, i.e. PPML gives downward biased standard errors (Burger et al., 2009);

- trade and tourism data usually have a lot of zero values in the dependent variable. There are bunch of reason for that, people may not go to some country in tourism purposes because of: the historical enmity, remoteness, absence of infrastructure (there might be no flights while other travel types are too costly), political reasons, etc. Some of these zeroes may be explained by the model, while others not. Sometimes the situation occurs when the number of such zeroes in the data is greater than it is predicted by Poisson or negative binomial distribution. It comes from so called "non-Poissonnes" caused by different types of zeroes in the data: some zeroes in tourism flows may be explained by long distance and captured by the model, while others only by factors not included in the model (like political tensions or historical ties). As Greene (1994) states, even if the "excess zeroes" problem can imitate the over-dispersion problem, a researcher has to distinguish between them, because unlike the over-dispersion the "excess zeroes" problem comes from non-Poissonnes.

To solve these problems, the stated above methods were applied. Burger et al. (2009) uses NBPML to solve over-dispersion issues and ZIPPML to deal with "excess zero" problem. The authors confirm the estimation bias integrated in the OLS estimator while zeros in the data are transformed (logarithmic transformation is used and small positive value is added to all observations). Authors conclude that the best performance while estimating gravity equation is achieved by the zero-inflated Poisson pseudo-maximum likelihood estimator.

Head and Mayer (2013) later claim that the usage of NBPML estimator is not that beautiful as it may seem at first glance. The authors show that PPML estimator is consistent not only in case of equi-dispersion but in case of proportional dispersion as well. It means that the so-called "over-dispersion", when the variance exceeds the mean, is not a problem for PPML estimator at all. Moreover, they "urge researchers to resist the siren song of the Negative Binomial" (Head & Mayer, 2013, p. 45) because the estimates of it depend on the measurement units for the dependent variable (first noted by Boulhol & Bosquet, 2012). Authors demonstrate that by estimating models with trade counted in thousands and millions of dollars they obtain not only different sizes of estimates but also different signs.

Methodology

OLS estimator issues with log-linearization

The gravity model assumes that there is a possibility to model tourist flows between two countries (districts, living areas, etc.) on Earth as a function of population or "economic masses" of consumers and distance between the two destinations. Unlike, for example, physics, in economics this formula holds only in its average terms. In other words, it should be interpreted as the expected value of a dependent variable given regressors. The standard formula is given by (1).

$$F_{ijt} = G\frac{M_{it}^{\beta_1} M_{jt}^{\beta_2}}{D_{ij}^{\gamma}}, \tag{1}$$

where at time t: $F_{ijt} \geq 0$ is a tourist flow to destination i from destination j, G is a constant, M_{it} and M_{jt} are economic masses of tourists in countries and D_{ij} is a distance between them. Since it holds only on average, then there should be some deviation/error term $E\left[\omega_{ijt}|M_{it}, M_{jt}, D_{ij}\right] = 0$, which will equalize both sides of the equation[7]:

$$F_{ijt} = G\frac{M_{it}^{\beta_1} M_{jt}^{\beta_2}}{D_{ij}^{\gamma}} + \omega_{ijt} = exp\left[ln\left(G\right) + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(D_{ij}\right)\right] + \omega_{ijt} \tag{2}$$

As Silva and Tenreyro (2006) show, the stochastic form of the equation (1) can be written in as:

$$F_{ijt} = exp\left[ln\left(G\right) + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(D_{ij}\right)\right] \varepsilon_{ijt}, \tag{3}$$

where $\varepsilon_{ijt} = 1 + \omega_{ijt}/exp\left[ln\left(G\right) + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(D_{ij}\right)\right]$ is the error term

[7] To deal with negative sign of γ I assume that $\gamma = -\beta_3$.

for which the following should hold:

$$E\left[\varepsilon_{ijt} \mid M_{it}, M_{jt}, D_{ij}\right] = 1.$$

Then now, even assuming that F_{ijt} is strictly positive (which is not the case in real data), logarithms can be taken on both sides, leaving us with the following linear equation:

$$ln\left(F_{ijt}\right) = ln\left(G\right) + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(D_{ij}\right) + ln\left(\varepsilon_{ijt}\right) \tag{4}$$

In this form basic OLS estimator can be applied if we have cross-sectional data or FE and RE estimators usually used for panel data. Hence, to get consistent estimates of the parameters in the equation (2) estimating equation (4) by OLS a usual assumption should be imposed on the error term $ln\left(\varepsilon_{ijt}\right)$ - it should be statistically independent from M_{it}, M_{jt}, D_{ij}. This restriction comes from the fact that for some positive random variable the expected value of a logarithm of it depends both on the mean and the variance of this random variable (see A note on the expected value of a logarithm of a random variable for details). Hence, in case of equation (3) the error term ε_{ijt} is dependent on either of M_{it}, M_{jt}, D_{ij}, then the expected value of $ln\left(\varepsilon_{ijt}\right)$ will be also dependent on these regressors. In that case the consistency condition of OLS is violated (Wooldridge, 2009).

One can notice that

$$\varepsilon_{ijt} = 1 + \omega_{ijt}/exp\left[ln\left(G\right) + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(D_{ij}\right)\right]$$

is independent from all regressors in case

$$\omega_{ijt} = exp\left[ln\left(G\right) + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(D_{ij}\right)\right] \upsilon_{ijt},$$

where υ_{ijt} is a random variable which is statistically independent from M_{it}, M_{jt}, D_{ij}. It will mean that $\varepsilon_{ijt} = 1 + \upsilon_{ijt}$ and $E\left[ln\left(\varepsilon_{ijt}\right) \mid M_{it}, M_{jt}, D_{ij}\right] = const$[8]. In this case, when ε_{ijt} is statistically

[8]Actually, if one needs to consistently estimate the intercept as well, then $E\left[ln\left(\varepsilon_{ijt}\right) \mid M_{it}, M_{jt}, D_{ij}\right] = 0$ should hold.

independent from all the regressors, the conditional variance of F_{ijt} (as well as ω_{ijt}) will be proportional to

$$exp\left[2\left(ln\left(G\right)+\beta_1 ln\left(M_{it}\right)+\beta_2 ln\left(M_{jt}\right)+\beta_3 ln\left(D_{ij}\right)\right)\right].$$

However in real data there is no ground to assume that the type of variance dependence is exactly of this form.

After running several Monte-Carlo simulations and applying different types of estimators on constant-elasticity models[9] Silva and Tenreyro (2006) show that OLS estimator, even if the dispersion of estimates is low, is extremely biased under most types of heteroscedasticity. Additionally to that, signs and the magnitudes of estimates vary considerably. Despite simulation procedures authors compare results of evaluating gravity equation with different estimators on real trade data. The outcome is the same as they have got during the theoretical investigations - OLS provides much different results compared to estimators being successful on simulated data.

Another issue a researcher has to deal with while applying log-linearization is zero value in tourism flows. In Newtonian gravity equation the gravitational force between two bodies in the space is assumed to be non-zero even if it is infinitely small (for the long-distanced objects). However, tourism flows between two countries may normally be equal to zero in some periods of time. It is easy to assume that, for example, nobody from Turkmenistan had a trip to New Zealand between 2010 and 2015, which is indeed the case in my data. To model zero flows a transformation $ln\left(X\right)\rightarrow ln\left(1+X\right)$ is usually applied if one intends to use OLS, RE, FE on log-linearized form of equation. Unfortunately, this method can lead to very biased estimates of coefficients, Flowerdew and Aitkin (1982) show that varying the constant that they add to zero values of the data between 0.01 and 1.00 the coefficient estimates decline as the constant grow. Moreover, few years later King (1988) demonstrated that a researcher can generate arbitrary coefficients estimate just by choosing a specific size of the constant. Thus, other methods are preferred like using Tobit estimator, Probit estimator or Maximum Likelihood family of

[9] Gravity model is just one type of that class of models

estimators[10].

Poisson pseudo-Maximum Likelihood estimator advantages

While linear Least Squares estimators cannot provide consistent results, NLS estimator does. It can be defined as:

$$\hat{\beta} = \arg\ \underset{b}{min} \sum_{i=1}^{n} \left[y_i - exp\left(x_i b\right)\right]^2, \tag{5}$$

leading to the following set of conditions:

$$\sum_{i=1}^{n} \left[y_i - exp\left(x_i \hat{\beta}\right)\right] exp\left(x_i \hat{\beta}\right) x_i = 0 \tag{6}$$

The drawback of such minimization problem is that it gives unequal weights to observations. Namely, observations with bigger value of $exp\left(x_i \hat{\beta}\right)$ are given bigger weights, because that's where the curvature of the conditional expectation is more noticeable. However, as it usually is in tourism data, those observations are also the ones with higher volatility, which implies that the NLS estimator assigns bigger weights to more volatile observations. This leads to the inefficiency of the estimator, especially when small number of observations are used. Silva and Tenreyro (2006) claim that if the form of the variance function were known, then the issue of inefficiency may be solved applying Weighted NLS estimator, but in practice it is not the case.

At this point, what is left to solve is the inefficiency drawback. One way of doing so is to apply non-parametric generalized least squares estimator used by Delgado (1992) and Delgado and Kniesner (1997), but doing so is very tedious, especially when having a lot of explanatory variables. Another approach is to obtain more efficient estimates of parameters by applying PML as suggest McCullagh and Nelder (1989), the set of first-order conditions for which is given by:

$$\sum_{i=1}^{n} \left[y_i - exp\left(x_i \tilde{\beta}\right)\right] x_i = 0 \tag{7}$$

[10]These estimators also have their features and restrictions, but I will not cover them in this paper. A long discussion on them can be found in Burger et al. (2009).

The reader can notice from (7) that PML gives the same weight to all observations in the data-set, unlike NLS. Giving the same weight to all observations seems to be more logical, the PML estimator from equation (7) gives more efficient results than NLS from equation (6).

Silva and Tenreyro (2006) also assure that the conditions defined in (7) are not numerically different from Poisson pseudo-Maximum Likelihood estimator. Moreover, from equation (7) we can see that for PPML estimator to be consistent only correct specification of the conditional mean is needed, so the data does not have to be distributed according to the Poisson type[11].

All in all, if the conditional mean of the dependent variable is correctly specified the PPMLE provides unbiased and consistent coefficient estimates while estimating constant elasticity models and does not suffer from the following issues:

- over-dispersion. As it was stated above, PPML is optimal if the conditional variance is proportional to the conditional mean. In other words, the estimator can deal not only with over-dispersion, but under-dispersion as well. Moreover, as Silva and Tenreyro state on their Log of Gravity page, even if the proportionality condition does not hold, the PPML will still be consistent;

- excess zeroes in the data. Silva and Tenreyro (2011) provide evidence that this is also is not a problem for the PPML estimator to provide correct results[12];

- non-Poisson distribution. That is true that generally the data on tourist flows is not Poisson distributed. However, as stated by Gourieroux et al. (1984) the data does not have to be Poisson distributed for the PPMLE to provide consistent results. As it is said in the beginning of this paragraph, all is needed is the conditional mean to be correctly specified.

Today, when a bunch of econometric and statistical packages are available, it is easy to estimate the gravity equation. I use the R package not only for data preparation stage, but for the econometric analysis as well. The functions applied will be described in the next section.

[11] This result mentioned first by Gourieroux, Monfort, and Trognon (1984)

[12] While estimating my model, I have also compared the results obtained with and without including zero tourism flows in the data. Similarly to the findings in trade by Silva and Tenreyro (2006), I have found no significant differences beween coefficient estimates of PPMLE in these two regression - the differences took place only in the second digit after the comma. It means that truncation is indeed not a problem for PPML, while according to Silva and Tenreyro (2006) and Westerlund and Wilhelmsson (2009) results of OLS regression are affeted by zeroes deletion.

Expected results

To overcome all the problems of log-linearization and OLS estimator described above, I apply PPML estimator in my paper. Since it can deal with the zero flows, I don't need to transform the dependent variable. The regression equation I am estimating can be written as follows:

$$F_{ijt} = exp\left[\beta_0 + \beta_1 ln\left(M_{it}\right) + \beta_2 ln\left(M_{jt}\right) + \beta_3 ln\left(W\_Dist_{ijt}\right) + \beta_4 ln\left(POP_{it}\right) + \beta_5 dEX\_RATE_{ijt} + \beta_6 S\_LANG_{ij} + \beta_7 VISA\_FREE_{ij} + \theta\Phi + \varepsilon_{ijt}\right], \tag{8}$$

- F_{ijt} is the value of tourism flow of country i to/from New Zealand (j) or total amount of tourists traveling between a pair. The main data I am using for this variable are the total amount of arrivals, total amount of departures and total tourism flow. However in the Robustness check section I also run my regression on the subsets of data representing arrivals and departures with the purpose of holidays and business;

- M_{it} is the economic mass variable for which I use the GDP per capita. However, again, in the Robustness check section I compare my main results for the case when I include GNP per capita instead as an economic mass proxy;

- W_DIST_{ijt} denotes the weighted distance between location i and j at time t (more on this is in the Data sources subsection);

- POP_{it} is a population of country i in year t;

- dEX_RATE_{ijt} stands for the percentage change in the exchange rate of the currency in country j to the currency of country i in year t compared to the previous year. It is calculated as $dEX\_RATE_{ijt} = ln\left(EX\_RATE_{ijt}/EX\_RATE_{ij(t-1)}\right)$. Since the j country is the same for every pair - New Zealand, then the variable represents change in exchange rates of 1 New Zealand Dollar to the currency of corresponding country (the source of data for that variable is described in subsection Data sources);

- to account for cultural closeness I use "share the same language" dummy in my regression (more on its construction is in Data sources subsection) - S_LANG_{ij}. It takes value 1 in case people in the departing country and in New Zealand speak the same language;
- the number of people traveling between two countries is expected to be affected by migration policies of their governments. I include $VISA\_FREE_{ij}$ dummy taking value 1 for countries the residents of which are not required to get visa to visit New Zealand[13];
- Φ is the vector of time and country fixed effects dummy variables, while θ includes list of coefficients for them.

The signs of coefficients may differ between regressions including different dependent variables. Since the directions of arrivals and departure tourist flows are opposite then some variables may have opposite signs of coefficients. However, some variables should have similar effect on flows since they represent overall economic incentives to travel. Hence, I expect to get positive beta coefficient for the economic mass of departure country since higher GDP/GNP per capita in general reflects higher net income available to individuals. This increases opportunities to purchase luxury goods like traveling. The sign of β_2 should also be positive because people are attracted more by wealthier countries, meaning that higher economic mass of New Zealand could also represent better tourism infrastructure development.

Obviously, anyone would expect negative sign on the distance coefficient β_3. Long geographical distances result in high cost of traveling, which includes not only expensive flight/ship tickets but also durability and stress. These factors make people choose less distant places to travel to. It is natural to assume that the coefficient on population variable β_4 will be positive both for arrivals and departures equations. More populated countries, holding all the other factors fixed, should have bigger number of people traveling abroad.

The coefficient on change in exchange rate of New Zealand's dollar to the currency of another country should have negative sign in regressions with arrivals to New Zealand and positive sign in regressions with departures from New Zealand. This intuition simply comes from the budget shrinking happening when the currency that a visitor uses weakens.

[13] At the same time, according to the The Henley & Partners Visa Restrictions Index 2017 New Zealand is placed 5^{th} in the rating of countries in terms of freedom of travel with the score of 172 together with Ireland and Japan. New Zealand citizens are free to visit more than 80% of countries.

I cannot tell what is the most probable sign of β_6 estimate because sharing the same language induces two-sided effect on tourism flows. From the one point of view, in the country where most of the population does not speak the same language as you do it is a barrier for visitors to commute and communicate freely . On the other side, it increases the interest to another culture and stimulates tourism flows.

Visa free agreements remove a barrier between countries making visiting less costly in terms of time and finance. Such arrangements also create a trustworthy relationship to the tourists from the side of community, stimulating visits of the former even more. The sing of β_7 coefficient is expected to be positive then.

Data

Data sources

The data for the analysis was collected from several sources either manually or with the use of R language.

The data on dependent variables, i.e. tourist flows in thousands of people between New Zealand and other countries from 1996 to 2015[14], was downloaded from the official site of statistics of the country - Infoshare. Easily browsing the data categories, a researcher can access national statistics of the country for any industry. For the purposes of the analysis the data on both arrivals from and departures to of New Zealand visitors/citizens was collected. It is worth noting that for the "departures" category the government of the country can only record the main destination of a traveler. In other words, if the trip of a traveler assumes to include several countries, which is often the case[15], it is impossible for New Zealand's tourism agency to record all the destinations. As for the arrivals, according to New Zealand's law, those visitors who stayed in the country for less than 12 month were counted as tourists and statistics on them was

[14]Initial sample was bigger, but since i use other regressors such as GDPPC and exchange rates, the data on which before 1996 and after 2015 is not easily available for most countries, I have restricted the sample to 20 years.

[15]According to several studies, tourists tend to visit multiple destinations at one trip (Hwang & Fesenmaier, 2003; Koo, Wu, & Dwyer, 2012; L. Wu, Zhang, & Fujiwara, 2011; Yang, Fik, & Zhang, 2013). Experiencing a number of destinations/attractions or activities at the same time a tourist increases her utility from the trip significantly. Additionally to that, visiting a region of closely located attractions helps to cut the expenses of the trip, tourists take advantage on this and spend more money on attractions enjoying diversity (Smith, 1983).

collected in the "arrival" category[16].

The data on overall economic wealth of tourists was collected with a package of R language allowing to access World Bank APIs[17]. Visitor's wealth is used in the numerator in the gravity model equation. The wealth of tourists are usually proxied by countries' GDP or National Income value. I use both of the variables to compare the results.

Tourism flows are highly dependent on cost of traveling. When a market shock occurs, the economy needs time to reach new equilibrium price level. Hence, at times of national currency exchange rate fall the opportunity to get relatively cheaper holidays for foreigners arises. To control for those opportunities I include exchange rate of New Zealand's Dollar to departure country into the regression. The list of currencies and countries where its applied was taken from FXtop site, while the data for 26 years and 148 unique currencies were collected manually from Oanda site, that provides yearly average exchange rates for a chosen pair of currencies.

To control for language barriers for visitors, I include the dummy variable of "share the same language". There are two main languages people in New Zealand use to communicate - English (96.1% of population) and Maori (3.7% of population). With the usage of Infoplease site, for each country in the sample, the data on the 3 most widely used languages was recorded and the dummy variable was constructed taking 1 for those countries that use either English or Maori language and 0 otherwise (see the list of all countries, its currencies and languages the data on which were used in the regression analysis in). One can notice that there are no countries where Maori is used as one of the official languages except New Zealand. Thus, in the context of the data the "same language" dummy is similar to "uses English" dummy.

The most non-trivial variable in the gravity equation is the distance. From a theoretical point of view it serves as a proxy for traveling costs (in tourism) and shipping/transportation costs (in trade). That is why, to account for other possible barriers to trade/travel, in literature other variables are used, such as access to sea, "contiguity", common language, tariffs, colonial history, etc., some of which I include in my regression. Anyway, the highest barrier for open economies to trade/travel proved to be the geographical distance (H. Wu, 2015). Modeling

[16]Those who stayed for more than 12 month are considered in "migration" category.

[17]A WDI (World Development Indicators) package allows to download World Bank data directly to the working session of an R-studio. For fast searching, the WDI package lists all available data series on a search request. This local list can be updated to the latest version using the WDIcache function. More on this is here.

distance in panel data is not that trivial, since this is something not changing over time.

To introduce some variation into the variable, the distance is usually weighted by the country's economic mass either through the time dimension, space dimension or both dimensions. All approaches lead to almost the same "trends" in the weighted distance variables: increasing/decreasing GDP of a country leads to a bigger weight and, hence, bigger distance for that time-period.

The approach I use in the paper is the following:

$$W\_Dist_{ijt} = \frac{GDP_{it} \cdot Dist_{ij}}{\sum_{k}^{n} GDP_{kt}}, \tag{9}$$

where W_Dist_{ijt} is now the weighted distance from location i to location j (which is New Zealand for arrival regression) at time t, GDP_{it} is the country's i GDP at time t, $Dist_{ij}$ stands for the geographical distance between two locations calculated by the Haversine formula (see Haversine formula for distance computation for details) and $\sum_{i}^{n} GDP_i$ is the sum of GDPs of all countries at the given year. The weighted distance of country i in year t is then proportional to it's GDP share in the total wealth of all countries in the data set in the current period.

The data on $VISA\_FREE_i$ variable is created using the list of countries having visa-free entry to New Zealand, which is published on the official immigration site of the country and includes 61 destinations.

Descriptive statistics and data features

To estimate the gravity equation I used the data for 189 countries from 1996 to 2015 year. The sources of data are described above, whereas in this subsection I focus on descriptive statistics of data. In Appendix the table with explanatory variables characteristics is provided. Regressions including GDP as economic mass variable have around 3400 observations, whereas using GNP as economic mass reduces the sample to only 2300 year-country pairs due to unavailability of data. The average and median distance from New Zealand to all destinations is $14'000$ km. During the sample period GDP per capita in the country varied between 29 and 34 thousand dollars, which is twice as much the mean of visitors' home countries. The distribution of population is far from normal: mean value of 7, high kurtosis and positive skewness, together with standard deviation being 4 times bigger than the mean, signalize that tourists mostly come from countries with the population less than $10-15$ mln people. About a half of all countries share the language with New Zealand. Visitors from about a third of all countries are not obliged to obtain a visa to enter New Zealand.

Table 5 with descriptive statistics is shown in the Appendix and gives intuition on the distribution of the dependent variables and logarithmized variables, which are used in regression analysis. The *Total flows* variable represents total tourist flows between New Zealand and a destination country, which is calculated simply as the sum of total arrivals from the country in a pair (denoted as *Arrivals (Total)*) and departures to the country in a pair (denoted as *Departures (Total)*). From the data one may assume that on average around 12.3 thousand people arrive to New Zealand and around 8.9 thousand people departure to each country each year. However the skewness of tourism flows is very high and positive, signalizing that a considerable share of all flows are close or equal to zero. Luckily, the PPML estimator does not suffer from the "excess zeros" problem. Besides total flows, I use tourism flows with business and holiday purposes as dependent variables in the Robustness check subsection. The descriptive statistics on arrivals and departures by type of a travel are also presented in the table. As expected, the logarithmized variables are much less skewed than their counterparts in levels. The kurtosis has also decreased significantly.

Estimation results

Factors determining tourism flows size

In the baseline specification I estimate the equation (8) using total tourism flows, total arrivals and total departures as the dependent variable (denoted as F_{ijt} in the regression (8)). Estimation results of the gravity model are presented in the Table 1. Since the sample for total flows between New Zealand and a pair country is constructed as the sum of arrivals to and departures from New Zealand, then it is logical to expect that the size of estimated coefficients in the "Total flows" regression can be represented as weighted by the number of tourists sum of coefficients in "Arrivals (total)" and "Departures (total)" regressions. Mostly, the signs on significantly estimated coefficients are consistent with the expectations and results of other studies of tourism flows (see for example Culiuc, 2014; Artal-Tur, Pallardó-López, & Requena-Silvente, 2016).

All three regressions reveal positive signs on GDP per capita and population size, while long distances are indeed high barriers for tourists to visit the country. Significantly higher number of tourists arrive to New Zealand from countries using English as one of their main languages. At the same time, New Zealand's citizens seem to not take into account language similarities when travel abroad. The dummy on existence of visa-free agreement with countries of incoming tourists is highly significant and positive for first two regression, while takes negative sign in the "departures" regression. Some explanations on these non-straightforward results of "departures" regression is given in the Robustness check subsection. Surprisingly, estimates of the coefficient for the New Zealand's GDPPC are negative. The result may appear due to a bias caused by small variation in the data - there are only 20 out of ~3400 unique observations for the variable.

The form of the estimated regression (8) allows to interpret coefficients in the same way as it would have been done while applying log-linearization to gravity equation. Coefficients on logarithmized variables on the RHS represent tourism flow elasticities. Thus, holding all the other factors fixed, the amount of visits to New Zealand from countries that have 1% higher GDP per capita are on average 2.2% bigger. Similarly, one can see that 1% longer distance is associated with about 1.4% smaller number of tourists. New Zealand citizens are even more susceptible to decrease the amount of trips on longer distances.

Tourist flows to New Zealand are more elastic in population size of a country than in per capita GDP: there are on average 2.8% more visitors come from countries with 1% bigger population. Slightly bigger size of the coefficient is obtained for language dummy. Results on arrivals are consistent with findings of Neumayer (2010) and Artal-Tur et al. (2016), who obtained significantly negative signs on visa restrictions dummy. Not surprisingly that islands invite considerably more people from countries with visa-free agreement[18].

[18]The increase in 5% is much smaller than the one found by Neumayer (2010) - 60% lower tourism flows come from countries with visa restrictions. But, actually, these two estimates are hardly comparable. First of all, Neumayer (2010) estimated the coefficient size for 3D Gravity model, i.e. 60% decrease is a "worldwide" effect, while 5% i have obtained is valid only for New Zealand arrivals and thus may be different. Secondly, the author

Table 1: Estimation results of tourism flows to New Zealand by direction of a travel

	Total flows	Arrivals (Total)	Departures (Total)
Log GDPPC	2.70***	2.15***	3.72***
	(0.15)	(0.20)	(0.12)
Log GDPPC in New Zealand	-2.04***	-1.44*	-3.07***
	(0.60)	(0.76)	(0.45)
Log weighted distance	-1.38***	-0.64**	-2.75***
	(0.23)	(0.29)	(0.19)
Log population	2.07***	2.84***	0.68***
	(0.20)	(0.30)	(0.16)
Exchange rate difference	0.06	0.15	-0.07
	(0.07)	(0.12)	(0.06)
Share the same language	2.31***	3.58***	-0.05
	(0.34)	(0.47)	(0.26)
Visa-free agreement	1.41**	3.69***	-2.69***
	(0.56)	(0.83)	(0.44)
Constant	-8.50***	-10.41***	-7.36***
	(1.82)	(2.37)	(1.32)
Country and time fixed effects	YES	YES	YES
Observations	3,386	3,386	3,386
Degrees of freedom	3,178	3,178	3,178
pseudo-R^2	0.12	0.13	0.09

Note: White's robust standard errors in parenthesis.
*, **, *** represent 10%, 5% and 1% significance levels, respectively.

Robustness check

To check the stability of results obtained in Table 1, I run regression (8) using GNP per capita as the economic mass variables M_{it}, M_{jt}. Due to a lot of missing observations for the GNP variable, the size of the sample decreased to only 2271 observations. The results of estimation are presented in the Table 2.

Coefficient estimates for the economic mass variables, weighted distance, population and exchange rate difference are of the same signs and sizes (taking into account standard deviation of estimates). However, the values of constant and dummies' estimates have changed

did not include country fixed effects into regression, which would decrease the estimate to 20% as it is estimated by Artal-Tur et al. (2016).

significantly and have become very noisy. It can hardly be the consequence of sample size reduction. What is more probable, is that the number of observations from reference groups of language and visa-free dummies decreased drastically, when GNP missing values were removed. It might have happened since there are no data on less developed countries, which are also those that have no visa-free agreement and do not use English language to communicate.

Table 2: Estimation results of tourism flows to New Zealand using GNP as economic mass variable

	Total flows	Arrivals (Total)	Departures (Total)
Log GNPPC	2.48***	2.15***	3.36***
	(0.19)	(0.26)	(0.18)
Log GNPPC in New Zealand	-2.45**	-1.53	-4.16***
	(0.97)	(1.50)	(0.81)
Log weighted distance	-1.44***	-1.20***	-2.41***
	(0.31)	(0.45)	(0.29)
Log population	2.81***	4.01***	0.36
	(0.31)	(0.48)	(0.25)
Exchange rate difference	-0.01	0.01	-0.08
	(0.07)	(0.09)	(0.09)
Share the same language	19.27	19.31	17.95
	(86.78)	(26.60)	(13.66)
Visa-free agreement	19.57	21.78	14.3
	(92.39)	(32.16)	(31.82)
Constant	-25.83	-31.2	-18.58
	(97.81)	(116.07)	(81.9)
Country and time fixed effects	YES	YES	YES
Observations	2,271	2,271	2,271
Degrees of freedom	2,069	2,069	2,069
pseudo-R^2	0.12	0.13	0.10

Note: White's robust standard errors in parenthesis.
*, **, *** represent 10%, 5% and 1% significance levels, respectively.

Robustness of the results was also checked via estimation of equation (8) on tourist flows by type. Two categories were used both for arrivals and departures: business travels and holiday travels. As one can see from Table 3, except for the Departures (Holiday) regression, the estimates of all other regressions are consistent with results of the baseline model.

One can notice that among incoming tourists only those who travel to do business are affected by distance covered. Holiday arrivals to New Zealand are, on the other hand, characterized by complete independence on the covered by visitors distance. This result is controversial to common economic sense and results of other studies, confirming that distance is the strongest barrier for travelers.

The number of arriving holiday tourists, on the other hand, is highly correlated with the language of home country and whether a visa-free agreement with New Zealand exists (visitors coming to the country with the purpose of spending there their holidays are about five times more concerned about communication in English than businessmen). It is a natural results that business travels are less dependent on visa-free agreements since all the paper work on receiving visa for an employee is done by the company. Hence, a decision to meet with a partner from another country is less affected by bureaucratic barriers. Same results are in the business departures regression. This is simply explained by the fact that business ties are usually two-way, meaning that parties from both countries travel to their partner's office with the same frequency and guided by the same business logic.

Most controversial results of estimated regression (8) are obtained for holiday departures category. Unlike expected positive signs on language and visa dummy, the estimates of these coefficients are negative. In addition, countries of 1% smaller population are on average expected to attract 0.5% less tourists from New Zealand, holding other factors fixed. Based on all being said, these results may only be explained by country-specific preferences on traveling abroad. New Zealand citizens prefer spending their holidays in small countries with contrasting culture, most of which have no visa-free entry to New Zealand. Strong negative estimates of regressors for holiday departures result in insignificant language dummy and negative visa-free agreement dummy in the baseline model.[19]

[19] This is easily explained by the fact that "holiday trips" is the most populous category of tourism flows. Thus, sizes and signs of coefficients obtained from "total" departures are highly dependent on features of "holiday" category estimates. That is why, for example, even if estimates of business departures (see Table 3) are consistent with the expectations, the estimate of *log(population)* in total departures (see Table 1) regression is small - strong positive coefficient of business travels was overweighted by negative holiday departures coefficient (see Table 3).

Table 3: Estimation results of tourism flows to New Zealand by type

	Arrivals (Business)	Arrivals (Holidays)	Departures (Business)	Departures (Holidays)
Log GDPPC	4.09***	1.71***	3.52***	4.54***
	(0.18)	(0.31)	(0.17)	(0.23)
Log GDPPC in New Zealand	-4.58***	-0.60	-2.92***	-4.41***
	(0.45)	(0.88)	(0.50)	(0.63)
Log weighted distance	-3.74***	0.29	-2.46***	-3.55***
	(0.25)	(0.40)	(0.22)	(0.31)
Log population	2.14***	3.02***	1.73***	-0.49*
	(0.15)	(0.45)	(0.15)	(0.28)
Exchange rate difference	0.09	0.20	0.01	-0.17
	(0.09)	(0.17)	(0.05)	(0.11)
Share the same language	0.93***	4.71***	1.64***	-1.73***
	(0.31)	(0.70)	(0.33)	(0.46)
Visa-free agreement	2.00***	4.70***	1.34***	-6.29***
	(0.44)	(1.23)	(0.45)	(0.77)
Constant	-13.95***	-11.14***	-12.49***	-4.10**
	(1.36)	(3.00)	(1.58)	(1.80)
Country and time fixed effects	YES	YES	YES	YES
Observations	3,386	3,386	3,386	3,386
Degrees of freedom	3,178	3,178	3,178	3,178
pseudo-R^2	0.03	0.13	0.06	0.05

Note: White's robust standard errors in parenthesis.
*, **, *** represent 10%, 5% and 1% significance levels, respectively.

It has been already shown in the literature on gravity model that including fixed effects is crucial for obtaining unbiased estimates for structural coefficients (see for example Mátyás, 1997; Mátyás, 1998; Head & Mayer, 2013). In the Appendix I also show the results of regressions estimated without including country and time dummies into regression (see Tables 6–8). As expected, the values of estimates for structural coefficients changed substantially, while dummies have become completely insignificant. It worth noting that overwhelming part of fixed effects dummies from regressions in Tables 1–3 is significant [20].

[20] Besides using FE specifications, I have run several regressions with explanatory variables in differences. However, the estimates of these specifications are too noisy, making the results insignificant. Together with this, the explanatory power of the regression falls to almost 0.

Concluding remarks

This paper studies what structural factors significantly affect the size of tourism flows to New Zealand. A gravity model frequently used in international trade literature is applied. Main results are consistent with the theoretical expectations: the number of visits to New Zealand are positively associated with the GDP/GNP per capita value in the home country and its population. The value of tourism flows decrease with the distance to the islands, which is a widely accepted result in the literature.

One of the findings also supported in the literature is that visa-free agreements are positively associated with the number of travelers coming to the country. Removing such a barrier leads to a decrease in financial and time costs of tourists preparing to a trip. Plus, much tighter connections are found with countries using English as one of their main languages. This discovery generally holds for country-pairs using the same language, hence positive estimate on "uses English language" dummy goes in line with results in other empirical paper studying tourism flows.

Another issue the thesis dealt with was application of PPML estimator for gravity model. The study supports the finding of Silva and Tenreyro (2006) – PPMLE estimates are robust to zeros deletion.

References

Alawin, M., & Abu-Lila, Z. (2016). Uncertainty and gravity model for international tourism demand in jordan: evidence from panel-grach model. *Applied Econometrics and International Development*, *16*(1), 131-146.

Aldaz, J. (2009). Self-improvement of the inequality between arithmetic and geometric means. *Journal of Mathematical Inequalities*, *3*(2), 213-216.

Anderson, J. (1979). A theoretical foundation for the gravity equation. *American Economic Review*, *69*(1), 106-116.

Anderson, J., & Marcouiller, D. (2002). Insecurity and the pattern of trade: an empirical investigation. *Review of Economics and Statistics*, *84*, 342-352.

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

Andrews, M., Schank, T., & Upward, R. (2006). Practical fixed effects estimation methods for the three-way error components model. *Stata Journal*, *6*, 461-481.

Artal-Tur, A., Pallardó-López, V. J., & Requena-Silvente, F. (2016, December). Examining the impact of visa restrictions on international tourist flows using panel data. *Estudios de Economia*, *43*(2 Year 20), 265-279. Retrieved from https://ideas.repec.org/a/udc/esteco/v43y2016i2p265-279.html

Baldwin, R., & Di Nino, V. (2006). Euros and zeros: The common currency effect on trade in new goods. *Center for Economic Policy Research Discussion Papers no 5973*.

Becken, S. (2002). Analysing international tourist flows to estimate energy use associated with air travel. *Journal of Sustainable Tourism*, *10*(2), 114-131. Retrieved from http://dx.doi.org/10.1080/09669580208667157 doi: 10.1080/09669580208667157

Bobkova, B. (2012). Gravity model estimation using panel data - is logarithmic transformation advisable? *Charles University in Prague Master thesis*, 58.

Boulhol, H., & Bosquet, C. (2012). Applying the glm variance assumption to overcome the scaledependence of the negative binomial qgpml estimator. , *33*(7), 772-784.

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

Culiuc, A. (2014). Determinants of international tourism. *IMF Working Paper No. 14/82*, 46. Retrieved from https://ssrn.com/abstract=2445467

Delgado, M. (1992). Semiparametric generalized least squares estimation in the multivariate nonlinear regression mode. *Econometric Theory*, *8*, 203-222.

Delgado, M., & Kniesner, T. (1997). Count data models with variance of unknown form: An application to a hedonic model of worker absenteeism. *The Review of Economics and Statistics*, *79*, 41-49.

Egger, P. (2000). A note on the proper econometric specification of thegravity equation. *Economics Letters*, *66*, 25-31.

Egger, P. (2002). An econometric view on the estimation of gravity models and the calculation of trade potentials. *The World Economy*, *25*(2), 297-312.

Flowerdew, R., & Aitkin, M. (1982). A method of fitting the gravity model based on the poisson distribution*. *Journal of Regional Science*, *22*(2), 191–202. Retrieved from http://dx.doi.org/10.1111/j.1467-9787.1982.tb00744.x doi: 10.1111/j.1467-9787.1982.tb00744.x

Genç, M. (2010). *Migration and Tourism Flows to New Zealand* (EcoMod2010 No. 259600062). EcoMod. Retrieved from https://ideas.repec.org/p/ekd/002596/259600062.html

Gomez-Herrera, E. (2013). Comparing alternative methods to estimate gravity models of bilateral trade. *Empirical Economics*, *44*, 1087-1111.

Gourieroux, C., Monfort, A., & Trognon, A. (1984). Pseudo maximum likelihood methods: Applications to poisson models. *Econometrica*, *52*, 701-720.

Greene, W. (1994). Accounting for excess zeros and sample selection in poisson and negative binomial models. *Stern School of Business, New York University, Working Paper 94-10*.

Head, K., & Mayer, T. (2013). Gravity equations: workhorse, toolkit, and cookbook. , 68.

Hwang, Y., & Fesenmaier, D. (2003). Multidestination pleasure travel patterns: Empirical evidence from the american travel survey. *Journal of Travel Research*, *42*(2), 166-171.

Jensen, J. (1906). Sur les fonctions convexes et les inégalités entre les valeurs moyennes. *Acta Mathematica*, *30*(1), 175-193. doi: 10.1007/BF02418571

Khadaroo, J., & Seetanah, B. (2007). Transport infrastructure and tourism development. *Annals of Tourism Research*, *34(4)*, 1021-1032.

King, G. (1988, August). Statistical models for political science event counts: Bias in conventional procedures and evidence for the exponential poisson regression model. *American Journal of Political Science*, *32*, 838-863.

Koo, T., Wu, C., & Dwyer, L. (2012). Dispersal of visitors within destinations: Descriptive measures and underlying drivers. *Tourism Management*, *33*(5), 1209-1219.

Lambert, D. (1994). Zero-inflated poisson regression with an application to defects in manufacturing. *Technometrics*, *34*, 1-14.

Law, D., Bryant, J., & Genc, M. (2009, June). *Trade, diaspora and migration to New Zealand* (NZIER Working Paper No. 2009/4). New Zealand Institute of Economic Research. Retrieved from https://ideas.repec.org/p/ris/nzierw/2009_004.html

Linders, G., & de Groot, H. (2006). Estimation of the gravity equation in the presence of zero flows. *Tinbergen Institute Discussion Papers*, *06-072/3*.

Long, J. (1997). Regression models for categorical and limited dependent variables. *Sage, Thousand Oaks.*.

Manning, W., & Mulahy, J. (2001). Estimating log models: to transform or not to transform? *Journal of Health Economics*, *20*, 461-494.

Martin, W., & Pham, C. (2008). Estimating the gravity equation when zero trade flows are frequent. *MPRA Working Paper 9453, University Library of Munich.*

Martinez-Zarzoso, I., Nowak-Lehmann, F., & Vollmer, S. (2007). The log of gravity revisited (in revision). *CEGE Discussion Paper 64, University of Gottingen.*

Massidda, C., Etzo, I., & Piras, R. (2015). Migration and inbound tourism: an italian perspective. *Current Issues in Tourism*, 1-20.

Matias, A. (2004). Economic geography, gravity and tourism trade: The case for portugal. A Paper Submitted to the I. Congress on Tourism Economics, Universitat de les Illes Balears, Palma, 28-29 May 2004.

Mátyás, L. (1997). Proper econometric specification of the gravity model. *World Economy*, *20*(3), 363–368. Retrieved from http://dx.doi.org/10.1111/1467-9701.00074 doi: 10.1111/1467-9701.00074

Mátyás, L. (1998). The gravity model: Some econometric considerations. *World Economics*, *21*, 397-401.

McClure, M. (2010). *Te ara - the encyclopedia of new zealand.* Retrieved from http://www.teara.govt.nz/en/tourist-industry/page-6

McCullagh, P., & Nelder, J. (1989). *Generalized linear models* (2nd ed.). Chapman and Hall/CRC.

Morley, C., Rossello, J., & Santana-Gallego, M. (2014). Gravity models for tourism demand: theory and use. *Annals of Tourism Research*, *48*, 1-10.

Neumayer, E. (2010). Visa restrictions and bilateral travel. *The Professional Geographer*, *62*(2), 171-181.

Santana-Gallegoa, M., Ledesma-Rodriguez, F., & Perez-Rodriguez, J. (2016). International trade and tourism flows: An extension of the gravity model. *Economic Modelling*, *52*, 1026-1033.

Santeramo, F., & Morelli, M. (2015). Modeling tourism flows through gravity models: A quantile regression approach. *Munich Personal RePEc Archive*, 1-15.

Saray, M., & Karagöz, K. (2010). Determinants of tourist inflows in turkey: evidencefrom panel gravity model. *ZKU Journal of Social Sciences*, *6*(11), 33-46.

Siliverstovs, B., & Schumacher, D. (2009). Estimating gravity equations: to log or not to log? *Empir Econ*, *36*, 645-669.

Silva, J., & Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, *88*(4), 641-658.

Silva, J., & Tenreyro, S. (2011). Further simulation evidence on the performance of the poisson pseudo-maximum likelihood estimator. *Economics Letters*, *112*(2), 220-222.

Simakova, J., & Stavarek, D. (2015). An empirican sector-specific gravity model for hungarian international trade. *Acta Universitatis Agriculturae et Silviculturae Mendelianae Brunensis*, *63*(6), 2145-2150.

Smith, S. (1983). *Recreation geography*. New York: Longman.

Teh, Y., Newman, D., & Welling, M. (2006). A collapsed variational bayesian inference algorithm for latent dirichlet allocation. *Neural Information Processing Systems*, *19*, 1353-1360.

Tinbergen, J. (1962). *Shaping the world economy: Suggestions for an international economic policy*. New York: Twentieth Century Fund.

Westerlund, J., & Wilhelmsson, F. (2009). Estimating the gravity model without gravity using panel data. *Applied Economics*, *43*(6), 641-649.

Wooldridge, J. M. (2009). *Introductory econometrics: A modern approach*. Australia: South-Western College Pub.

Wu, H. (2015). Re-visiting the distance coefficient in gravity model. *Cornell University*

Library, 20.

Wu, L., Zhang, J., & Fujiwara, A. (2011). A tourist's multi-destination choice model with future dependency. *Asia Pacific Journal of Tourism Research*, *17*(2), 121-132.

Yang, Y., Fik, T., & Zhang, J. (2013). Modelling sequential tourist flows: Where is the next destination? *Annals of Tourism Research*, *43*, 297-320.

Appendix

Arrivals to New Zealand in 2015

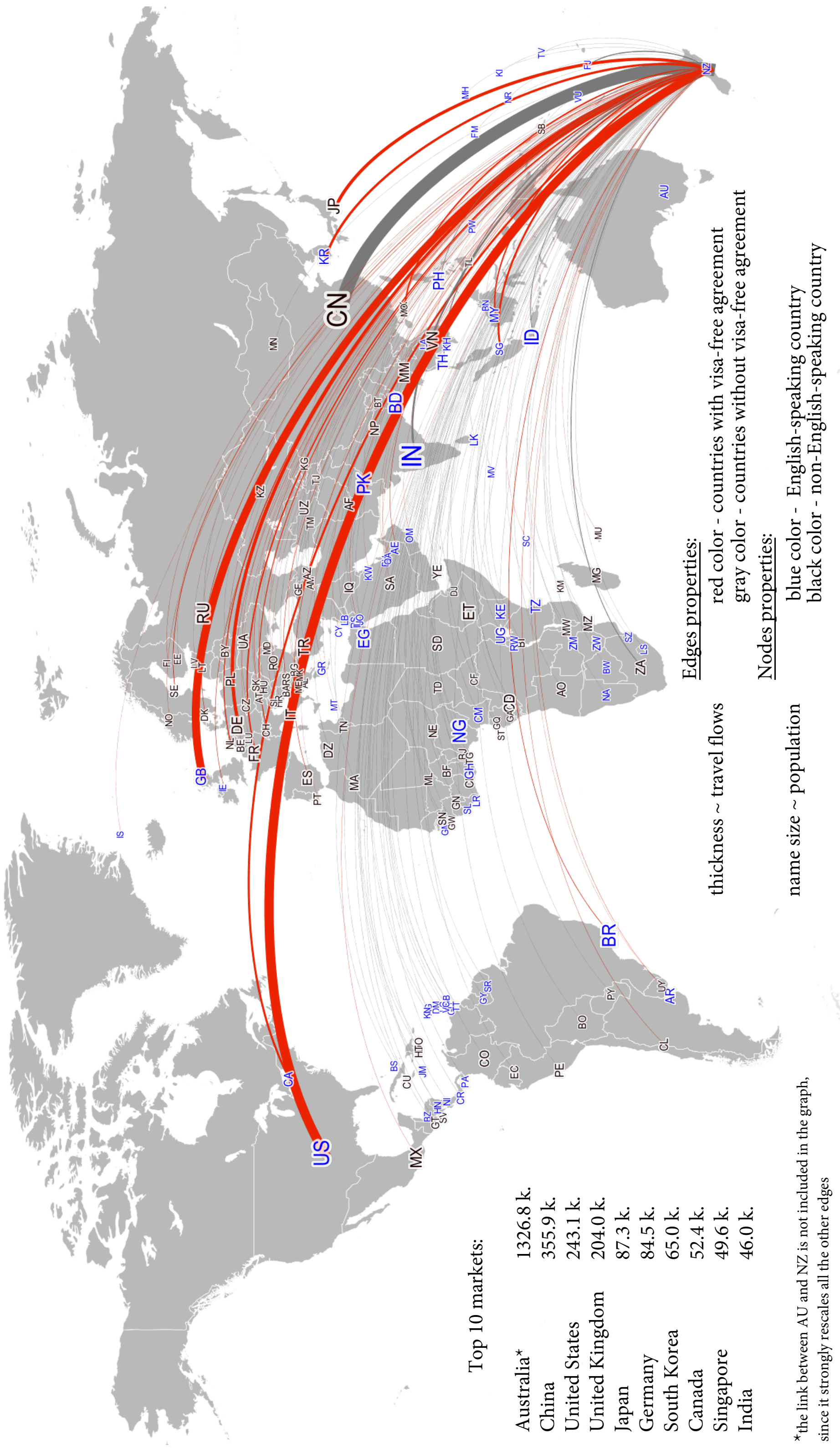

Figure 2: Arrivals to New Zealand in 2015

A note on the expected value of a logarithm of a random variable

Suppose we have an approximately normally distributed random variable X, concentrated around some positive mean value and the density of which near to 0 is approximately 0[21]. Then, following Jensen (1906), since the log-function is concave, we can prove that $E\left[ln\left(X\right)\right] \leq ln\left(E\left[X\right]\right) - ln\left(E\left[X\right]/E^2\left[X\right]\right)$. One of the proofs is described by Aldaz (2009), which states that by using the inequality of arithmetic and geometric means (AM-GM inequality) and the law of large numbers to the square roots of i.i.d. X_i, that is:

$$\sqrt[n]{\sqrt{X_1}\ldots\sqrt{X_n}} \leq \frac{1}{n}\left(\sqrt{X_1}+\cdots+\sqrt{X_n}\right)$$

and taking the limit $n \longrightarrow \infty$the law of large numbers gives:

$$exp\left(E\left[ln\sqrt{X}\right]\right) \leq E\left[\sqrt{X}\right]$$

Hence, taking logs on both sides:

$$E\left[lnX\right] \leq 2ln\left(E\left[\sqrt{X}\right]\right).$$

Now, plugging the $E\left[\sqrt{X}\right]$:

$$E\left[lnX\right] \leq ln\left(E\left[X\right]\right) - \delta\left(X\right), \quad \text{where } \delta(X) = ln\left(\frac{E\left[X\right]}{E^2\left[X\right]}\right).$$

Note that the same result can be obtained by applying the Taylor's approximation. One of the proofs is given by Teh, Newman, and Welling (2006), stating that for some transformation $g\left(.\right)$ and its expansion $g\left(\mu_x + h\right) = g\left(\mu_x + \left[X - \mu_x\right]\right)$, after computing the expectation and variance of the expansion we get:

[21]This assumption is needed to easily apply the $ln\left(.\right)$ function and leave behind all the cases when the argument is zero or even negative.

$$
\begin{aligned}
E[g(X)] &\approx g(\mu_x) + \frac{g''(\mu_x)}{2}\sigma_x^2 \\
Var[g(X)] &\approx [g'(\mu_x)]^2 \sigma_x^2
\end{aligned}
$$

Now, substituting $g(.)$ with $ln(.)$ we see that:

$$E[ln(X)] \approx ln(\mu_x) - \frac{\sigma_x^2}{2\mu_x^2}$$

Therefore, the expected value of a logarithm of a random variable is indeed a function of its mean and variance.

Haversine formula for distance computation

To compute the actual distance between two points on Earth the Haversine formula for distance is applied (used to calculate distances on spherical objects, very important in navigation). The form of the formula is given by:

$$d = 2 \cdot r \cdot arcsin\sqrt{sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + cos(\phi_1)\,cos(\phi_2)\,sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)},$$

where:

- d is the Haversine distance;
- r is the radius of the sphere, which is equal to 6371 km in my case;
- ϕ_1,ϕ_2 are latitudes of points 1 and 2, respectively, in radians;
- λ_1,λ_2 are longitudes of points 1 and 2, respectively, in radians.

To calculate Haversine distances in R package, the following function can be used:

```
dist_haversine ←function(lat_from, lon_from, lat_to, lon_to, r = 6.371){
    radians ← pi/180
    lat_to ← lat_to * radians
    lat_from ← lat_from * radians
    lon_to ← lon_to * radians
    lon_from ← lon_from * radians
    dLat ← (lat_to - lat_from)
    dLon ← (lon_to - lon_from)
    a ← (sin(dLat/2)^2) + (cos(lat_from) * cos(lat_to)) * (sin(dLon/2)^2)
    return(2 * atan2(sqrt(a), sqrt(1 - a)) * r)
}
```

The data on countries' coordinates is available on Maxmind site or on a lot of other sources.

NOTE: since the Earth is not a perfect sphere, the Vincenty's formulae can be used as well, which allows to calculate distances on oblate spheroids.

Descriptive statistics of data

Table 4: Descriptive statistics of variables in levels

Variable	*n*	*mean*	*sd*	*median*	*min*	*max*	*skew*	*kurtosis*	*Units of measurement*
GDPPC	3386	16.62	19.78	9.65	0.26	137.16	2.32	6.96	'000 of $ per capita
GDPPC in New Zealand	3386	31.31	2.63	32.16	26.28	35.16	-0.61	-0.75	'000 of $ per capita
GNPPC	2271	17.39	18.83	10.85	0.33	126.53	2.06	5.81	'000 of $ per capita
GNPPC in New Zealand	2271	30.03	2.06	30.57	25.23	33.21	-0.71	-0.14	'000 of $ per capita
Distance	3386	13.57	3.95	14.01	2.56	19.81	-0.86	0.37	'000 kilometers
Population	3386	36.93	137.26	6.98	0.01	1371.22	7.95	67.01	mln people
Share the same language	3386	0.45	0.5	0	0	1	1	0.21	-
Exchange rate difference	3386	0.03	0.39	0.04	-10.55	4.61	15.16	-8.39	%
Visa-free agreement	3386	0.3	0.46	0	0	1	1	0.87	-

Table 5: Descriptive statistics of tourism flows and variables in logs

Variable	*n*	*mean*	*sd*	*median*	*min*	*max*	*skew*	*kurtosis*	*Units of measurement*
Total flows	3386	21.21	140.96	0.36	0	2463.7	2463.7	12.57	‘000 of people
Arrivals (Business)	3386	1.32	10.89	0.01	0	175.49	175.49	12.79	‘000 of people
Arrivals (Holidays)	3386	6.17	32.25	0.05	0	516.02	516.02	9.71	‘000 of people
Arrivals (Total)	3386	12.32	75.01	0.16	0	1326.8	1326.8	11.91	‘000 of people
Departures (Business)	3386	1.42	11.39	0.03	0	169.24	169.24	12.65	‘000 of people
Departures (Holiday)	3386	3.72	27.86	0.04	0	420.24	420.24	12.17	‘000 of people
Departures (Total)	3386	8.89	67.5	0.15	0	1136.9	1136.9	12.87	‘000 of people
Log GDPPC	3386	2.14	1.25	2.27	-1.34	4.92	6.26	-0.14	-
Log GDPPC in New Zealand	3386	3.44	0.09	3.47	3.27	3.56	0.29	-0.72	-
Log GNPPC	2271	2.24	1.23	2.38	-1.11	4.84	-0.33	-0.75	-
Log GNPPC in New Zealand	2271	3.4	0.07	3.42	3.23	3.5	-0.85	0.04	-
Log population	3386	1.7	2.16	1.94	-4.68	7.22	11.91	-0.42	-
Log accommodation	3386	4.88	0.09	4.91	4.68	4.98	0.3	-0.82	-

Estimation results without including FE

Table 6: Estimation results of tourism flows to New Zealand by direction of a travel

	Total flows	Arrivals (Total)	Departures (Total)
Log GDPPC	3.68***	3.68***	3.73***
	(2.58,4.78)	(2.41,4.94)	(2.69,4.77)
Log GDPPC in New Zealand	-2.99***	-3.03***	-2.87***
	(-4.64,-1.33)	(-4.9,-1.16)	(-3.74,-2.01)
Log weighted distance	-2.51**	-2.22**	-2.87***
	(-4.43,-0.6)	(-3.91,-0.53)	(-4.45,-1.28)
Log population	0.65***	0.73***	0.56***
	(0.47,0.83)	(0.54,0.93)	(0.35,0.76)
Exchange rate difference	-0.08	-0.04	-0.15
	(-0.26,0.09)	(-0.22,0.15)	(-0.33,0.03)
Share the same language	0.58	0.44	0.84
	(-0.76,1.93)	(-1.06,1.93)	(-0.31,1.99)
Visa-free agreement	0.16	0.04	0.12
	(-0.67,0.98)	(-0.87,0.95)	(-0.72,0.96)
Constant	-6.59**	-6.32	-9.01***
	(-11.62,-1.56)	(-13.81,1.18)	(-11.71,-6.3)
Country and time fixed effects	NO	NO	NO
Observations	3,386	3,386	3,386
Degrees of freedom	3,378	3,378	3,378
pseudo-R^2	0.14	0.14	0.15

Note: bootstrapped clustered by country 95% c.i. in parenthesis.
*, **, *** represent 10%, 5% and 1% significance levels, respectively.

Table 7: Estimation results of tourism flows to New Zealand using GNP as economic mass variable

	Total flows	Arrivals (Total)	Departures (Total)
Log GNPPC	3.83***	3.81***	3.83***
	(2.72,4.93)	(2.30,5.33)	(3.09,4.56)
Log GNPPC in New Zealand	-3.13***	-3.31**	-2.83***
	(-4.48,-1.79)	(-5.53,-1.08)	(-3.98,-1.68)
Log weighted distance	-2.16***	-2.04**	-2.35***
	(-3.63,-0.69)	(-3.87,-0.21)	(-3.32,-1.38)
Log population	0.69***	0.71***	0.66***
	(0.42,0.97)	(0.31,1.11)	(0.50,0.81)
Exchange rate difference	-0.09	-0.07	-0.12
	(-0.25,0.08)	(-0.23,0.10)	(-0.32,0.08)
Share the same language	0.70	0.59	0.97*
	(-0.68,2.09)	(-1.2,2.38)	(-0.18,2.12)
Visa-free agreement	-0.42	-0.26	-0.69
	(-1.34,0.50)	(-1.27,0.76)	(-1.62,0.23)
Constant	-5.83	-5.42	-8.32***
	(-14.15,2.49)	(-20.6,9.75)	(-10.65,-6.00)
Country and time fixed effects	NO	NO	NO
Observations	2,271	2,271	2,271
Degrees of freedom	2,263	2,263	2,263
pseudo-R^2	0.13	0.14	0.13

Note: bootstrapped clustered by country 95% c.i. in parenthesis.
*, **, *** represent 10%, 5% and 1% significance levels, respectively.

Table 8: Estimation results of tourism flows to New Zealand by type

	Arrivals (Business)	Arrivals (Holidays)	Departures (Business)	Departures (Holidays)
Log GDPPC	4.22***	3.70***	4.08***	3.75***
	(3.27,5.17)	(2.59,4.81)	(3.33,4.83)	(2.61,4.89)
Log GDPPC in New Zealand	-4.35***	-3.38***	-4.89***	-3.53***
	(-5.54,-3.15)	(-5.30,-1.46)	(-6.07,-3.72)	(-4.57,-2.49)
Log weighted distance	-2.63***	-2.08***	-2.99***	-3.02***
	(-3.81,-1.46)	(-3.36,-0.81)	(-4.11,-1.87)	(-4.77,-1.28)
Log population	0.80***	0.82***	0.62***	0.53***
	(0.57,1.02)	(0.45,1.20)	(0.43,0.80)	(0.33,0.73)
Exchange rate difference	-0.02	0.00	-0.07	-0.16*
	(-0.17,0.14)	(-0.24,0.24)	(-0.26,0.12)	(-0.35,0.04)
Share the same language	0.36	0.17	0.48	1.01*
	(-0.82,1.54)	(-1.28,1.62)	(-0.55,1.51)	(-0.15,2.17)
Visa-free agreement	-0.43	0.14	0.17	0.24
	(-1.33,0.47)	(-0.97,1.25)	(-0.52,0.87)	(-1.20,1.68)
Constant	-7.03***	-5.72	-5.30***	-8.25***
	(-10.87,-3.20)	(-13.30,1.86)	(-8.13,-2.47)	(-11.77,-4.73)
Country and time fixed effects	NO	NO	NO	NO
Observations	3,386	3,386	3,386	3,386
Degrees of freedom	3,378	3,378	3,378	3,378
pseudo-R^2	0.11	0.16	0.14	0.17

Note: bootstrapped clustered by country 95% c.i. in parenthesis.
*, **, *** represent 10%, 5% and 1% significance levels, respectively.

List of countries used in the sample

Table 9: List of countries used in the sample, its currencies and languages

Country name	Code	Currency	Language 1	Language 2	Language 3
AFGHANISTAN	AF	AFN	Dari	Pashtu	Persian
ALBANIA	AL	ALL	Albanian	Greek	
ALGERIA	DZ	DZD	Arabic	French	Berber
ANGOLA	AO	AOA	Portuguese	Bantu	
ANTIGUA AND BARBUDA	AG	XCD	English		
ARGENTINA	AR	ARS	Spanish	English	Italian
ARMENIA	AM	AMD	Armenian	Yezidi	Russian
ARUBA	AW	AWG	Dutch		
AUSTRALIA	AU	AUD	English		
AUSTRIA	AT	EUR	German	Croatian	Hungarian
AZERBAIJAN	AZ	AZN	Azerbaijani	Russian	Armenian
BAHAMAS	BS	BSD	English	Creole	
BAHRAIN	BH	BHD	Arabic	English	Farsi
BANGLADESH	BD	BDT	Bangla	English	
BARBADOS	BB	BBD	English		
BELARUS	BY	BYR	Belorussian	Russian	
BELGIUM	BE	EUR	Dutch	French	German
BELIZE	BZ	BZD	English	Spanish	Mayan
BENIN	BJ	XOF	French	Fon	Yoruba
BERMUDA	BM	BMD	English		
BHUTAN	BT	BTN	Dzongkha	Nepalese	
BOLIVIA	BO	BOB	Spanish	Quechua	Aymara
BOSNIA AND HERZEGOVINA	BA	BAM	Bosnian	Croatian	Serbian
BOTSWANA	BW	BWP	English	Setswana	Kalanga
BRAZIL	BR	BRL	Portuguese	Spanish	English
BRUNEI DARUSSALAM	BN	BND	Malay	English	Chinese
BULGARIA	BG	BGN	Bulgarian	Turkish	Roma
BURKINA FASO	BF	XOF	French	Sudanic	
BURUNDI	BI	BIF	Kirundi	Swahili	French
CAMBODIA	KH	KHR	Khmer	French	English
CAMEROON	CM	XAF	French	English	
CANADA	CA	CAD	English	French	
CAYMAN ISLANDS	KY	KYD	English		
CENTRAL AFRICAN REPUBLIC	CF	XAF	French	Sangho	
CHAD	TD	XAF	French	Arabic	Sara
CHILE	CL	CLP	Spanish		
CHINA	CN	CNY	Chinese	Yue	Wu
COLOMBIA	CO	COP	Spanish		
COMOROS	KM	KMF	Arabic	Shikomoro	French
CONGO, DEMOCRATIC REPUBLIC OF THE	CD	CDF	French	Lingala	Kingwana
CONGO, REPUBLIC OF	CG	XAF	French	Lingala	Monokutuba
COSTA RICA	CR	CRC	Spanish	English	
COTE D'IVOIRE	CI	XOF	French		
CROATIA	HR	HRK	Croatian	Hungarian	Czech
CUBA	CU	CUC	Spanish		
CYPRUS	CY	EUR	Greek	Turkish	English
CZECH REPUBLIC	CZ	CZK	Czech		
DENMARK	DK	DKK	Danish	Faroese	Greenlandic
DJIBOUTI	DJ	DJF	French	Somali	Afar

DOMINICA	DM	XCD	English	French	
DOMINICAN REPUBLIC	DO	DOP	Spanish		
ECUADOR	EC	USD	Spanish	Quechua	
EGYPT	EG	EGP	Arabic	English	French
EL SALVADOR	SV	SVC	Spanish	Nahua	
EQUATORIAL GUINEA	GQ	XAF	Spanish	French	Fang
ESTONIA	EE	EUR	Estonian	Russian	
ETHIOPIA	ET	ETB	Amharic	Tigrigna	Orominga
FIJI	FJ	FJD	English	Fijian	Hindustani
FINLAND	FI	EUR	Finnish	Swedish	
FRANCE	FR	EUR	French	Breton	Alsatian
GABON	GA	XAF	French	Fang	Myene
GAMBIA	GM	GMD	English	Mandinka	Wolof
GEORGIA	GE	GEL	Georgian	Russian	Armenian
GERMANY	DE	EUR	German		
GHANA	GH	GHS	English	Moshi-Dagomba	
GREECE	GR	EUR	Greek	English	French
GRENADA	GD	XCD	English	French	
GUATEMALA	GT	GTQ	Spanish	Quiche	
GUINEA	GN	GNF	French	Susu	Fulani
GUINEA-BISSAU	GW	XAF	Portuguese	Criolo	
GUYANA	GY	GYD	English		Creole
HAITI	HT	HTG	Creole	French	
HONDURAS	HN	HNL	Spanish	English	
HONG KONG (SPECIAL ADMINISTRATIVE REGION)	HK	HKD	Chinese		
HUNGARY	HU	HUF	Hungarian		
ICELAND	IS	ISK	Icelandic	English	German
INDIA	IN	INR	Hindi	English	Bengali
INDONESIA	ID	IDR	Bahasa Indonesia	English	Dutch
IRAN	IR	IRR	Persian	Turkic	Kurdish
IRAQ	IQ	IQD	Arabic	Kurdish	Assyrian
IRELAND	IE	EUR	English	Irish	
ISRAEL	IL	ILS	Hebrew	Arabic	English
ITALY	IT	EUR	Italian	French	Slovak
JAMAICA	JM	JMD	English	Jamaican Creole	
JAPAN	JP	JPY	Japanese		
JORDAN	JO	JOD	Arabic	English	
KAZAKHSTAN	KZ	KZT	Kazak	Russian	
KENYA	KE	KES	English	Swahili	
KIRIBATI	KI	AUD	English	Kiribati	
KOREA, SOUTH	KR	KRW	Korean	English	
KUWAIT	KW	KWD	Arabic	English	
KYRGYZSTAN	KG	KGS	Kyrgyz	Russian	
LAOS	LA	LAK	Lao	French	English
LATVIA	LV	EUR	Latvian	Russian	Lithuanian
LEBANON	LB	LBP	Arabic	French	English
LESOTHO	LS	LSL	English	Sesotho	Xhosa
LIBERIA	LR	LRD	English		
LIBYA	LY	LYD	Arabic	Italian	English

LITHUANIA	LT	LTL	Lithuanian	Russian	Polish
LUXEMBOURG	LU	EUR	Luxermbourgish	German	French
MACAU (SPECIAL ADMINISTRATIVE REGION)	MO	MOP	Chinese		
MACEDONIA	MK	MKD	Macedonian	Albanian	Turkish
MADAGASCAR	MG	MGA	Malagasy	French	
MALAWI	MW	MWK	Chichewa	Chinyanja	Chiyao
MALAYSIA	MY	MYR	Malay	English	Chinese
MALDIVES	MV	MVR	Maldivian Dhivehi	English	
MALI	ML	XOF	French	Bambara	
MALTA	MT	EUR	Maltese	English	
MARSHALL ISLANDS	MH	USD	Marshallese	English	Japanese
MAURITANIA	MR	MRO	Arabic	Pulaar	Soninke
MAURITIUS	MU	MUR	Creole	Bojpoori	French
MEXICO	MX	MXN	Spanish	Mayan	Nahuatl
MICRONESIA	FM	USD	English	Chukese	Pohnpeian
MOLDOVA	MD	MDL	Romanian	Russian	Gagauz
MONGOLIA	MN	MNT	Mongolian	Turkic	Russian
MONTENEGRO	ME	EUR	Serbian		
MOROCCO	MA	MAD	Arabic	Berber	French
MOZAMBIQUE	MZ	MZN	Portuguese	Emakhuwa	Xichangana
MYANMAR	MM	MMK	Burmese		
NAMIBIA	NA	NAD	English	German	Herero
NAURU	NR	AUD	Nauruan	English	
NEPAL	NP	NPR	Nepali	Maithali	Bhojpuri
NETHERLANDS	NL	EUR	Dutch	Frisian	
NICARAGUA	NI	NIO	Spanish	English	
NIGER	NE	XOF	French	Hausa	Djerma
NIGERIA	NG	NGN	English	Hausa	Yoruba
NORWAY	NO	NOK	Norwegian	Norwegian	
OMAN	OM	OMR	Arabic	English	Baluchi
PAKISTAN	PK	PKR	Urdu	English	Punjabi
PALAU	PW	USD	Palauan	English	Sonsoralese
PALESTINIAN STATE (PROPOSED)	PS	ILS	Arabic	Hebrew	English
PANAMA	PA	PAB	Spanish	English	
PAPUA NEW GUINEA	PG	PGK	TokPisin		Hiri Motu
PARAGUAY	PY	PYG	Spanish	Guarani	
PERU	PE	PEN	Spanish	Quéchua	Aymara
PHILIPPINES	PH	PHP	Filipino	English	Cebuano
POLAND	PL	PLN	Polish		
PORTUGAL	PT	EUR	Portuguese	Mirandese	
PUERTO RICO	PR	USD	Spanish		
QATAR	QA	QAR	Arabic	English	
ROMANIA	RO	RON	Romanian	Hungarian	German
RUSSIA	RU	RUB	Russian		
RWANDA	RW	RWF	Kinyarwanda	French	English
SAMOA	WS	WST	Samoan	English	
SAO TOME AND PRINCIPE	ST	STD	Portuguese		
SAUDI ARABIA	SA	SAR	Arabic		
SENEGAL	SN	XOF	French	Pulaar	Jola

SERBIA	RS	RSD	Serbian	Hungarian	Slovak
SEYCHELLES	SC	SCR	Seselwa Creole	English	French
SIERRA LEONE	SL	SLL	English	Mende	Temne
SINGAPORE	SG	SGD	Mandarin	English	Malay
SLOVAKIA	SK	EUR	Slovak	Hungarian	Roma
SLOVENIA	SI	EUR	Slovenian	Croatian	Serbian
SOLOMON ISLANDS	SB	SBD	Melanesian		
SOUTH AFRICA	ZA	ZAR	IsiZulu	IsiXhosa	Afrikaans
SPAIN	ES	EUR	Spanish	Galician	Basque
SRI LANKA	LK	LKR	Sinhala	Tamil	English
ST. KITTS AND NEVIS	KN	XCD	English		
ST. LUCIA	LC	XCD	English	French	
ST. VINCENT AND THE GRENADINES	VC	XCD	English	French	
SUDAN	SD	SDG	Arabic	Nubian	Ta Bedawie
SURINAME	SR	SRD	Dutch	Surinamese	English
SWAZILAND	SZ	SZL	English	siSwati	
SWEDEN	SE	SEK	Swedish		
SWITZERLAND	CH	CHF	German	French	Italian
TAJIKISTAN	TJ	TJS	Tajik	Russian	
TANZANIA	TZ	TZS	Swahili	English	Arabic
THAILAND	TH	THB	Thai	English	
TIMOR-LESTE	TL	USD	Tetum		
TOGO	TG	XOF	French	Ewe	Mina
TONGA	TO	TOP	Tongan	English	
TRINIDAD AND TOBAGO	TT	TTD	English	Hindi	French
TUNISIA	TN	TND	Arabic	French	
TURKEY	TR	TRY	Turkish	Kurdish	Dimli
TURKMENISTAN	TM	TMT	Turkmen	Russian	Uzbek
TUVALU	TV	AUD	Tuvaluan	English	Samoan
UGANDA	UG	UGX	English	Ganda	Luganda
UKRAINE	UA	UAH	Ukrainian	Russian	Romanian
UNITED ARAB EMIRATES	AE	AED	Arabic	Persian	English
UNITED KINGDOM	GB	GBP	English	Welsh	Scots Gaelic
UNITED STATES	US	USD	English	Spanish	
URUGUAY	UY	UYU	Spanish	Portunol	Brazilero
UZBEKISTAN	UZ	UZS	Uzbek	Russian	Tajik
VANUATU	VU	VUV	Bislama	English	French
VENEZUELA	VE	VEF	Spanish	English	
VIETNAM	VN	VND	Vietnamese	Chinese	Khmer
YEMEN	YE	YER	Arabic		
ZAMBIA	ZM	ZMW	English	Kaonda	Lozi
ZIMBABWE	ZW	USD	English	Shona	Ndebele