Internal migration and natural disasters in Mexico.

A spatial modeling approach

By Paola Kim-Blanco
Master's Thesis
Quantitative Methods in the Social Sciences
Graduate School of Arts and Sciences
Columbia University in the City of New York

Thesis Advisers: Susana Adamo, PhD and Jeremy Porter, PhD.

April 30, 2016.

1. Abstract

This study aims to investigate whether the frequency of natural disasters, or reported damages from disastrous events in rural areas, had an effect on internal migration in Mexico between 2010 and 2015. Spatial regression models are used to explain the associations between migration and explanatory factors. The results suggest that the frequency of natural disasters have a significant, spatial effect on internal migration. The models with the best fit for both in-migration and out-migration consider hazardous events as aggregates rather than individual events. This finding is similar to previous studies in emigration from 1990 and 2000 (Saldaña-Zorrilla et al., 2009). The data from reported damages of disastrous events in rural areas is not appropriate for spatial modeling, so no meaningful results were obtained in that regard.

2. Objective and justification

Mexico has a long-standing tradition on migration. Several authors argue that the presence of shocks influence migration (Adamo and de Sherbinin, 2011; Munshi, 2003; Massey, Axinn, et al., 2010; Hunter, Murray, et al., 2013; Nawrotzki, Riosmena, et al., 2013; Hunter, Luna et al., 2015). In Mexico specifically, empirical studies have shown that droughts and floods have influenced international migration; and the presence of natural disasters have contributed to Mexico's social vulnerability (Rodriguez-Oreggia, de la Fuente et al., 2013). This study will build upon previous work conducted by Saldaña-Zorrilla and Sandberg (2009), with two main differences: 1) the methodology will be applied to understand internal migration, instead of international emigration; and 2) data from 2010-2015 will be used, instead of the 1990-2000 timeframe.

3. Research questions

This study will answer three questions:

1. Does the frequency of natural disasters have an effect on in-migration and out-migration rates within Mexico?
2. Are the same effects identified when data from reported damages (i.e., perception data) is used?
3. Are there any differences on how the frequency of natural disasters may affect the rates of in-migration and out-migration?

4. Theoretical framework

Migration is a complex demographic process, in which underlying forces intervene at various scales and levels, and is dependent on contextual factors (Massey, Arango, et al., 1993; Renaud, Bogardi et al., 2007; Laczko, 2010; Warner, 2010; Black, Adger et al., 2011; Renaud, Dun et al., 2011; Black, Arnell et al., 2013; Hunter, Luna et al., 2015). The act of migrating may be personal though the decision to migrate may be household-based. Likewise, the motivation to migrate may be purely income-based, though issues related to families' incomes may include labor demand at destination, institutional and legal mechanisms that allow the transfer of monetary resources from one place to the other, the ability to cope under foreign circumstances, among others. Hence, migration is not an easy process to model.

Lee's view of the migratory process involves an origin, a destination, and an intervening set of obstacles. Among the set of intervening obstacles, the distance of the move is included (Lee, 1965). At origin, there may be a set of negative factors that may

push individuals to relocate; at the same time, there may be a set of positive features that may pull individuals to move in (idem). In that sense, the interaction between push and pull factors may act as the underlying force on how people distribute across space. Though one important distinction is that features may have different effects on different individuals. Lee explains that what may constitute pull factors for some, may represent a push factor for others (idem). Costs of relocation and social ties with previous migrants may have significant effects on how far individuals may move.

4.1. Which factors could potentially act as pull/push factors to model migration?

Proponents of theories of (mostly international) migration have outlined the main theoretical lines of thought from which migration processes could be modeled and explained. Some of these include the neoclassical economic approach (at macro and micro levels), which takes into account wage differentials between urban and rural areas[1] as the main driver for migration (Massey et. al., 1993; Todaro, 1969; Harris and Todaro, 1970). A second theory refers to the new economics of migration, which assumes not only income maximization at the individual level but also risk reductions at the household level (Stark and Bloom, 1985 in Massey et. al., 1993). The new economics of migration theory is based on the premise that income sources diversification and insurance provisions could act as poverty reduction strategies. A third one refers to the dual labor market theory in which Piore (1979) argues that international migration is caused by the permanent demand of labor (Piore, 1979 in Massey et. al., 1993), which in other words argues in favor of cheap labor, a common immigrants' characteristic, at the

[1] Wage differentials theory could be applied within urban-urban contexts as well.

place of destination[2]. In all three cases, the expectation of better incomes (for any applicable purpose) among migrants is a common denominator. As Todaro (1969) and Harris and Todaro (1970) suggest, the higher the difference between the current and the expected income is, the higher the probability a person would migrate (Todaro, 1969; Harris and Todaro, 1970). Therefore, it is possible to use income as a positive predictor to explain migration.

Related to income is the expectation of employment at destination. Usually well-developed and diversified economies may provide more labor opportunities to migrants than closed economies. Places active in all economic sectors, probably with higher levels of urbanization, may attract more migrants than those with limited economic options. Hence, size and economic diversification may have an explanatory effect to migration as well.

Levy and Wadycki (1974) observed a two-fold effect with respect to education and migration. On one hand, a decreased probability to out-migrate was observed as educational levels increased. However, an increasing trend toward out-migration was also observed when 'expected' educational opportunities at the place of destination were higher than the ones at the place of origin (Levy and Wadycki, 1974). Interestingly, the propensity to move longer distances increased among the most educated, and decreased among the least educated. One potential explanation is that the costs of moving are more affordable among the most educated than the rest (idem). In that sense, average educational attainment at places of destination could provide useful insights on the 'expected' educational opportunities for migrants.

[2] However, these labor opportunities are always better than the ones available at the place of origin.

Specifically, for the Mexican context, unfortunately, violence is one factor that may play a significant role in terms of people's mobility. Between 2006 and 2011, 1.6 million Mexicans have left their homes due to the ongoing drug violence and threats (Fausset, 2013 in Chi, et al., 2013). Fausset estimates that most of these individuals have left behind 20 "ghost pueblos", to relocate in cities where access to jobs or services is unknown (idem). Therefore, the level of violence can act as a push factor in Mexico.

4.2. Where does the migration-natural disasters nexus come from? Can natural disasters explain migration?

There is global consensus among experts that human migration is connected to environmental change (Moriniere and Hamza, 2012). Moreover, it has been argued that environmental factors in the form of natural disasters, environmental change or environmental degradation, have a multiplier effect over other drivers of migration (Adamo and de Sherbinin, 2011; Warner, Ehrhart et. al., 2009). Prolonged water-stress combined with high-temperature events such as in droughts, as well as floods from excess rainfall, tsunamis provoked by tectonic plates moving, or destructive winds from hurricanes or tornadoes are some examples of natural disasters associated to human migration.

Within the environmental migration literature, natural disasters are often referred to as 'rapid onset hazards'[3] (Renaud, Dun, et al., 2011). Though depending on the hazard, the duration of these events may range from minutes (earthquakes) to weeks (frost events). The way in which natural hazards are often related with human migration is that

[3] Droughts are usually considered a special case (though still a hazard), associated with long-term adverse effects on many fronts: depletion of aquifers, land degradation, loss of biodiversity, among other long-term environmental effects.

they may act as triggers: before prolonged droughts, rain-fed agriculture fails, and farmers are left with limited choices on how to move forward; migration may be explored as an alternative option and remittances are considered as a source of livelihood (Munshi, 2003; Massey, Axinn, et al., 2010; Hunter, Murray, et al., 2013; Nawrotzki, Riosmena, et al., 2013; Hunter, Luna et al., 2015). Alternatively, devastating events such as hurricanes and flooding may cause irreparable damages to crops, assets, or production equipment that may force people to relocate (Winsemius, Jongman, et al., 2015). Overall, there is a strong likelihood that extreme weather and climatic events intensify and augment in frequency by 2050 (IPCC, 2013). Specifically, in the case of Mexico City, the earthquake in 1985 had an important effect on people's distribution, who were deterred of moving to the city for some time (Izazola, 2004). The same effect was observed after the earthquake in Haiti, in 2010 (UN OCHA, 2011).

Environmental migration is often perceived as an adaptive response to changes in the environment (Cerrutti and Massey, 2001; Munshi, 2003; Saldaña- Zorrila, 2008; Massey, Axinn et al., 2010; Warner, 2010; Nawrotzki, Riosmena et al., 2015). However, it has also been argued that migration can be the result of non-adaptive mechanisms to changes happening at the places of origin (Laczko, 2010; Warner, 2010). Poverty is commonly associated as an impediment to cope with stressors, including natural hazards or other environmental changes. Evidence from 52 countries demonstrates that the poor are overexposed to droughts and urban floods (Winsemius, Jongman, et al., 2015). The inability to resist and act upon impacts due to economic hardship adds another layer of complexity when trying to understand migration. These social layers will not be studied here, but will be proposed as future lines of research.

4.3. Why Mexico?

Mexico has a long, well-documented history of migratory processes that have shaped the way the country is today. Internal migration has been the main demographic determinant on the distribution of population in Mexico within the XX century (Partida Bush, 2001). Almost 1 out of 100 individuals have crossed state-level boundaries to change his/her place of residence, at least once in their lifetime (idem). Partida Bush has observed this phenomenon in Mexico since 1950. Urban decentralization policies, insecurity, and natural disasters (to name a few) have contributed to reversed migration from cities into the countryside, or from the capital city to other small cities, at least in three different points in time: within 1970s, after the earthquake in 1985, and within the 1990s (Izazola, 2004; Massey, Durand et al., 2009; Pelaez Herreros, 2013).

Moreover, there is a well-established, traditional corridor of international migration from Mexico (and other Latin American countries) to the United States, and Canada (Massey, 1990; Reuveny, 2007; Massey, Durand et al., 2009). Approximately ~12 million Mexicans lived abroad by 2010 (Zapponi, 2010). Ninety-eight percent of these emigrants lived in the United States (idem). Almost 11.6 million people have crossed the border between Mexico and the United States making it one of the busiest global migration corridors (idem).

In that sense, the Bracero Program (Mexican Farm Labor Agreement) may bring some insight to this phenomenon. Implemented in 1942[4], the Bracero Program was supposed to cover for the shortage of farm labor in agricultural lands in the United States during the Second World War. At the end of the war subsequent program extensions

[4] Previous bilateral agreements between Mexico and the United States started as early as 1909, between presidents Diaz and Taft (Durand, 2007).

provided temporary manual and farm labor opportunities to Mexican farmers until 1964. When the Bracero Program ended, social and economic impacts felt negatively within rural Mexico, and lasted for several years (Durand, 2007). In this sense, understanding the history behind the Bracero program provides sufficient grounds to understand illegal immigration, social networks, and labor rights today.

Another important event in the history of Mexico-United States immigration was the Agrarian Reform promulgated in 1992, which modified the Mexican Constitution to allow for the liberalization and free trade of communal agricultural lands in Mexico (Schwartz and Notini, 1994; Audley, Papademetriou et al., 2004). Part of the motivation for the Agrarian Reform was to attract foreign investment into Mexican lands at the onset of the North American Free-Trade Agreement (NAFTA) (Audley, Papademetriou et al., 2004). However, this move felt negatively among smallholders and subsistence farmers. Many smallholders became land-insecure; another portion ended leaving the countryside in search for better opportunities elsewhere (internally, among Mexican urban areas; or internationally, mainly the US) (idem).

4.4. Previous work on migration-environment nexus in Mexico.

Understanding migratory processes and the linkages with climatic and other hazardous events has significant policy implications. There is an interesting line of research led by the Population Program at the University of Colorado Boulder that has taken into understanding international migration between Mexico and the United States driven by both changes in rainfall and temperature. Significant connections between changes in rainfall patterns and increasing temperatures and international migration were

found repeatedly using various data sources, including: census data from the Mexican Statistics Office, INEGI; survey data from the Mexican Migration Project[5]; weather-station data; DesInventar dataset of natural disasters[6]; as well as satellite imagery. One of the main findings was that the lack of rains and increased temperatures intensified international migration (Nawrotzki, Riosmena et al., 2013; Riosmena, Nawrotzki et al., 2013; Runfola, Romero-Lankao et al., 2013; Nawrotzki, Hunter et al., 2015; Nawrotzki, Riosmena et al., 2015).

Qualitative work has also been conducted in this respect. Saldaña-Zorrilla investigated smallholders' coping mechanisms to natural disasters in small rural communities in Chiapas (Saldaña- Zorrilla, 2008). Similar anecdotal work was also conducted in rural communities from Tlaxcala and Chiapas (Alscher 2010); and in Oaxaca (Cohen 2004). International migration (mainly to the US) and rural-urban migration was found as one frequent coping mechanism after hazardous episodes in all the three studies.

Later on Saldaña-Zorrilla and Sandberg (2009) estimated that weather-related disasters have accounted for 80% of economic losses in the agricultural sector in Mexico since 1994. The authors argue this is a relevant finding since 25% of the population's livelihood depend on agriculture (Saldaña- Zorrilla and Sandberg, 2009). In that sense, Feng and others were able to quantify these losses and tie them up with estimates on international migration flow: for each 10% in expected crop losses, there is a 2% increase of migrants to the United States (Feng, Krueger et al., 2010). The socioeconomic impact that internal migration may have in cities or other rural areas is not known to the author.

[5] See http://mmp.opr.princeton.edu/home-en.aspx
[6] See http://www.desinventar.net/index_www.html

Though Runfola and others (2015) inspected the exposure to disastrous events considering a scenario of no migration, and then comparing it to the next quinquennial data source to estimate the exposure contribution. Overall, the level of exposure decreased with migration, which means that people moving out of hazardous-prone areas made areas less exposed, which was expected. However, an intensified movement toward cities increased exposure to natural disasters at hazardous-prone urban destinations. Therefore, the selection of urban destinations may not have been based on the probability of floods or droughts, but based on other (presumably) economic reasons (Runfola, Romero-Lankao et al., 2013). Rodriguez- Oreggia and others found that the presence of natural disasters in Mexico increases social vulnerability (Rodriguez-Oreggia, de la Fuente et al., 2013).

The literature shows that migration is a complex and multidimensional issue to model. However, there are some aspects that may help explain migration. In this study the main motivation is to test whether the frequency of natural disasters has influenced internal migration in Mexico. Previous work on this regard has proved a significant relationship between internal migration and natural disasters. This study will build upon previous work conducted by Saldaña-Zorrilla and Sandberg (2009), with two main differences: 1) the authors' methodology will be applied toward internal migration, instead of international migration; and 2) data from 2010-2015 will be used. The same methodological framework will be applied.

5. Model Specification

Spatial Lag and Spatial Error models[7] are proposed to best understand both rates of in and out migration at the municipal level. As Lee pointed out in his paper "Theory of Migration", distance is one of the intervening obstacles to migration (Lee, 1965). In this sense, it is possible to assume that Tobler's first law of geography[8], which is the underlying principle of spatial models through the presence of spatial dependency (i.e. autocorrelation), can apply to study migration rates and how it is affected by the incidence of natural hazards. The main assumption is that contextual characteristics among neighboring units will act as either push or pull factors, which may influence the different rates of migration across space.

A similar methodology was proposed by Saldaña-Zorrilla and Sandberg (2009) to study the effect of natural disasters on emigration in Mexico, between 1990 and 2000, from which this study builds upon.

5.1. Spatial Lag model

The spatial lag model assumes spatial dependency by adding a "spatially lag" dependent variable y on the right-hand side of the equation. This model is adequate when predicted values of y_i are influenced by the values of the neighbors of i (Ward and Gleditsch, 2008). Spatial lag models are appropriate for continuous dependent variables. Following the notation of Anselin (Anselin, 1998), the spatial lag model can be expressed as:

[7] Anselin (1998), Schabernberger and Gotway (2005), Viton (2010) among others, refer to the Spatial Lag model as Spatial Autoregressive model (SAR). But in this paper, it will be referred to as Spatial Lag.
[8] Waldo Tobler's First Law of Geography (1979): "Everything is related to everything else, but near things are more related than distant things" (Viton, 2010).

$$Y = \rho \mathbf{W}\mathbf{y} + \mathbf{X}\beta + \epsilon,$$

$$\epsilon \sim N(0, \sigma^2 \mathbf{I})$$

Where **I** represents an identity matrix, and the $N(0, \sigma^2 I)$ indicates that the errors follow a normal distribution with mean equal to zero and constant variance. When ρ is zero, the lag-dependent term is cancelled out, leaving the model under the Ordinary Least Squares (OLS) form. Though when ρ is not zero, it means that spatial dependency exists, and that non-random spatial observable interactions are present (Ward and Gleditsch, 2008).

5.2. Spatial Error model

In the case of spatial error models, the spatial dependency is accounted for through the error term. Instead of letting neighboring values to affect the values of y_i, the spatial error model assumes the errors of the model to be spatially correlated. The spatial error model can be modeled directly following geostatistical principles, or by utilizing a spatial autoregressive process for the error term (Anselin, 2002). In this paper, the former approach is used as defined in Ward and Gleditsch (2008), by dividing the error term into two pieces: one error term that assumes no spatial correlation and satisfies the normal assumption (ϵ), and another component (ξ) that includes the spatial dependency. The parameter λ indicates the level of correlation between error components, incorporating also the relationship defined by the contiguity (or neighbors) matrix. The spatial error model can be therefore expressed as

$$y = \mathbf{X}\beta + \lambda \mathbf{W}\xi + \epsilon,$$

$$\epsilon \sim N(0, \sigma^2 \mathbf{I})$$

If there is no spatial correlation between error components, then the parameter λ equals zero, and the model behaves as a regular OLS. Though, if the parameter λ is not zero, then a pattern of spatial dependence between the errors of connected observations is present. Moreover, having λ with a non-zero value could be the result of other processes happening such as misspecifications of the model, omitted variables that are spatially clustered too, or other unexplained effects (Ward and Gleditsch, 2008; Schabenberger and Gotway, 2005).

The main difference between Spatial Lag and Spatial Error models is that in the former case the spatial dependency has an effect on the predicted values; whereas in the spatial model the observations are related through unexplained factors (i.e. error term), which for some unknown reason, are correlated in space (Ward and Gleditsch, 2008). Since there is no effect on the coefficients, using OLS estimates in place of Spatial Error coefficients is inefficient, though not necessarily incorrect. However, using reported standard error terms from OLS is erroneous (Anselin, 2002; Ward and Gleditsch, 2008).

5.3. Spatial Neighbors

Both Spatial lag and Spatial error models take municipalities as the unit of analysis. Municipalities are areal units of unequal size with sharing borders. Choosing the neighboring criteria is the first step toward identifying spatial patterns (Bivand, et. al., 2008). Given the difference in sizes and non-constant number of neighboring units, a queen-style contiguity matrix was created to illustrate the association among observations. This style of contiguity is met when "at least one point on the boundary of one polygon is within the snap distance of at least one point of its neighbor" (idem). This

means that all potential points of contact for each unit will be considered as neighbors, regardless of the extent of that contact (i.e. if it is just a point, or the whole border-line). First-order contiguity neighbors (i.e. queen 1) matrix is used as the neighboring parameter for spatial modeling (see figure 1).

5.4. Spatial Weights

A row-standardized weighting scheme was preferred in this study. What this means is that the share of the weights depends on the number of neighboring units, per unit of analysis. Units with few neighbors will have higher standardized weights than units with many neighbors (Bivand et.al., 2008).

5.5. Spatial Autocorrelation

Spatial Autocorrelation can be defined as "the coincidence of value similarity with locational similarity" (Anselin and Bera, 1998 in Viton, 2010). Positive spatial autocorrelation means that values tend to cluster together, whereas negative spatial autocorrelation means that values are surrounded by significantly different values (idem). The assumption of spatial autocorrelation is the essence of spatial modeling. Therefore, suitable tests are needed in order to confirm these spatial patterns. Moran's I was the preferred statistic to test for spatial autocorrelation following Anselin (2003), Bivand, et. Al. (2008), Ward and Gleditsch (2008); Viton (2010) examples.

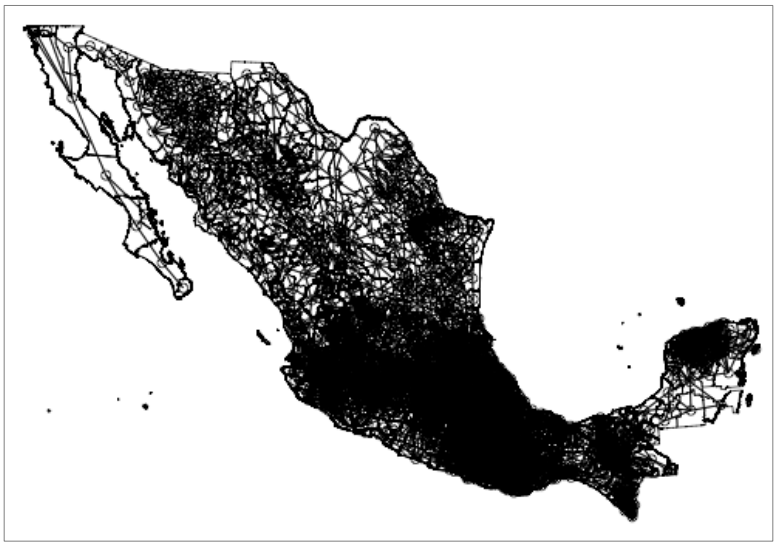


Figure 1 Queen- style, first-order municipal contiguity in Mexico.

5.6. Caveats and limitations

Spatial regression models inherit the same underlying assumptions as linear regression models. For instance, the assumption of no collinearity needs to be observed closely. Predictors need to be independent of each other, without inherent linear relationship among factors (e.g. regressing age against year of birth), or exceeding the convened number of dummy variables as predictors (Berry, 1993). In this sense, the variance inflation factor (VIF) will be calculated at a threshold value = 5 among all potential variables to test for collinearity (Agresti and Finlay, 2009).

In terms of the assumption of no correlation between the error term and each independent variable, there is one situation that Berry (1993) refers to as a potential factor that violates this assumption: reciprocal causation. Reciprocal causation happens when the dependent variable influences one or more of the independent variables (Berry, 1993). This assumption could be violated in this study. As stated before, migration is a complex, multi-dimensional process. In this sense, it is possible that the presence or absence of internal migrants may have an influence among contextual processes –for

example, highly educated in-migrants may contribute with the average income level of a given unit; at the same time, average income levels, as a proxy of expected average income, usually act as a pull factor (Harris and Todaro, 1970). Then the relationship between in-migration and average income can be reciprocal. In this paper, this issue is acknowledged and assumed to be minimal by studying migration as an aggregate phenomenon (i.e. at the municipal level) and by modeling *rates* of in and out migration, rather than migration volumes.

A third assumption is the fact that the selection of independent variables is correctly defined, and based on a sound theoretical framework (Berry, 1993). In this respect, the theory was followed to highlight the *main* social and economic dimensions to test against rates of in and out-migration. These dimensions were represented by proxy variables, available to the researcher by the time of this study. It is possible that other, more appropriate metrics exist and were either overlooked or not available for the time frame and/or aggregation level of the study. In this sense, this assumption could be violated, and could be a limitation of the study findings.

Another assumption of the linear regression is that the variance of the error term ε_i is constant among all observations. When the error terms present clustering or identifiable patterns, the assumption is violated; this is known as heteroskedasticity. Kaufman (2013) mentions three common situations where heteroskedasticity is commonly present: 1) when analyzing an aggregate dependent variable; 2) when making comparisons among social groups; 3) when the distribution of the dependent variable is far from the symmetric and bell-shaped distribution (Kaufman, 2013). This study uses aggregates as dependent variables; therefore, heteroskedasticity is possible, and is an

acknowledged limitation in this regard. To correct for #3, both in and out-migration rates (i.e. the dependent variables) have been log-transformed to *normalize* their distributions, though they do not resemble the normal curve, strictly speaking. Kaufman's point #2 is not applicable in this study.

Stationarity is an underlying assumption among spatial autoregressive models. Stationarity means that the statistical processes are global and do not depend on individual locations (Schabenberger and Gotway, 2005). In that sense, under stationarity, it is expected that the mean and variance of a variable remains constant as locations change within the overall area of study. It also assumes that the correlation between any two locations depends only on the vector that separates them, and not on their exact locations in space (Krivoruchko, n.d.). When this assumption is violated, non-stationarity happens. Non-stationarity is a common feature of many spatial processes (Schabenberger and Gotway, 2005), and may be indicative of other, more efficient ways to model data – for example, using geographic weighted regressions (GWR), or other probabilistic models. In this study, the use of GWR was not explored, though will be proposed as an additional line of research.

Finally, the difference in sizes and the nature of boundaries among municipalities may constitute a limitation in regards to the Modifiable Areal Unit Problem (MAUP). The MAUP happens by the "imposition of artificial units of spatial reporting on continuous geographical phenomena resulting in the generation of artificial spatial patterns" (Heywood, 1988, in Ervin, 2016). MAUP constitutes another limitation when individual-level data is scaled-up toward municipal level aggregation.

6. Data sources and variable construction

A description of each variable, its data source, the acronym used during analysis, and its corresponding transformations are explained below.

2) **ln_INRATE/ ln_OUTRATE**- these two variables represent the dependent variables, and refer to log-transformations of both in-migration and out-migration rates in Mexico, for the period 2010- 2015. Both migration rates were calculated using the question of place of residence 5-years ago from INEGI's Encuesta Intercensal extended questionnaire in Mexico, in 2015[9]. Using weighted factor values from micro-data, it was possible to calculate the internal migration matrix. This study followed the approach and methodology from the Centro Latinoamericano y Caribeño de Demografía CELADE's Internal Migration in Latin America and the Caribbean (MIALC) database[10] to calculate the internal migration matrix and internal migration rates (CELADE, 2016). This methodology excludes people living abroad five years ago and currently living in Mexico (i.e. immigrants). Likewise, this methodology uses residents calculated from the matrix for both years as the denominator, instead of total populations figures. Equations to calculate number of residents in both years as well as those used to calculate in-migration and out-migration rates are presented below.

$$\text{Residents}_{2010} = \text{not migrants} + \text{out} - \text{migrants}_{2010-2015}$$

[9] This question excludes children under five years of age.
[10] For more information, see http://www.cepal.org/celade/migracion/migracion_interna/

$$\text{Residents}_{2015} = \text{not migrants} + \text{in} - \text{migrants}_{2010-2015}$$

$$\text{In} - \text{migration rate} = \frac{\text{in} - \text{migrants}_{2010-2015}/5}{\left(\text{residents}_{2010} + \text{residents}_{2015}/2\right)} * 1000$$

$$\text{Uut} - \text{migration rate} = \frac{\text{out} - \text{migrants}_{2010-2015}/5}{\left(\text{residents}_{2010} + \text{residents}_{2015}/2\right)} * 1000$$

Not migrants represent the population who lived in the same place of residence 5 years ago during enumeration. Not migrants are the individuals placed at the 'diagonal' within the migration matrix. *In-migrants* $_{2010\text{-}2015}$ refer to the count of individuals who currently live at the place of enumeration, and did not live there 5 years ago. *Out-migrants* $_{2010\text{-}2015}$ refer to the count of individuals who left their residences 5 years ago, by place of origin. The calculation of out-migrants is possible because individuals who relocated within this 5-year period also replied to the question "in which municipality did you live 5 years ago?".

The reason why the volume of in/ out-migrants is divided by five in the numerator is because both migration rates are calculated in an annual basis, which assumes a constant distribution of migrants among the five-year period. Likewise, the reason why the volume of residents in 2010 and 2015 is divided by two in the denominator is because it estimates the resident population as a middle point between both years.

Internal migration is modeled separately (as in and out-migration) because of theoretical purposes. The same contextual factors may affect both groups differently –for example, the incidence of flooding may act as a push factor for out-migrants whereas the

same phenomena may not have any effect on in-migration. In order to capture these differences, these groups need to be differentiated and modeled separately.

2) ln_ECONEST- this variable refers to the log-transformation of the total number of economic establishments registered under INEGI's National Inventory of Economic Units (DENUE), 2011-2015. Economic establishments comprise primary, secondary and tertiary economic activities between the period 2011-2015. This variable was included following Lee's hypothesis relating volumes of migration with the size of the economy (Lee, 1965). A more appropriate measure could have been average GDP, however this metric was not available at the municipal level.

3) ln_INCOME- this variable refers to the log-transformation of the average income in 2010, at the municipal level. This variable was downloaded from the National Institute toward Federalism and Municipal Development (INAFED)'s website in March 2016, though the data was authored (probably because of some processing) by the United Nations Development Programme (UNDP) to calculate the Human Development Index (HDI) at the municipal level. This variable was included in this study as part of Harris and Todaro's assumption of *expected income* as the pull factor for in-migrants (Harris and Todaro, 1970).

4) EXPSCHOOLI- this variable pertains to the *expected* years of education in a municipality, given the average age of the population in 2010. Similar to *ln_INCOME*, this variable is used as input for the HDI. The data was downloaded from INAFED's website too. This variable was not log-transformed because the distribution of the original data was better than the log-transformed one.

5) ln_FLOODS, ln_FROST, ln_EQK, ln_RAIN, ln_HURR, ln_ALLHAZARDS- these variables constitute the log-transformation of the count of disastrous events classified as emergencies from the Sistema Nacional de Protección Civil from 2011 to 2015. The data was obtained from official records declaring state of emergency from various disasters. The records are available at the municipal level, and can be accessed on the Fondo Nacional de Desastres (FONDEN)'s website. These records were manually entered in a spreadsheet for further processing and analysis.

- ln_FLOODS: log-transformation of flooding events
- ln_FROST: log-transformation of frost, hail and snow storm events
- ln_EQK: log-transformation of earthquakes and landslides events
- ln_RAIN: log-transformation of heavy rains, tropical storm events
- ln_HURR: log-transformation of hurricanes, strong wind events
- ln_ALLHAZARDS: log-transformation of the sum of all of the above

6) PDROUGHT, PFROST, PHURR, PEQK, PFLOOD- these variables constitute a separate set of disaster-related variables, and refer to the percentage or rural localities that presented any damages due to drought, frost, hurricanes, earthquakes, and flooding (as listed respectively in the header) for any given municipality between the period 2010 and 2015. This data was obtained as part of INEGI's preliminary fieldwork toward the 2015 Encuesta Intercensal in Mexico. A potential source of bias with this data is that it represents the *views or perception* of key informants within rural localities. Similarly, the fact that it only illustrates potential damages at rural locations limits the analysis within rural municipalities only (n= 840 municipalities, 34.35%)

- PDROUGHT: % rural localities with damages from droughts between 2010-2015

- PFROST: % rural localities with damages from frost, hair or snow storms between 2010-2015
- PHURR: % rural localities with damages hurricanes between 2010-2015
- PEQK: % rural localities with damages from earthquakes between 2010-2015
- PFLOOD: % rural localities with damages from floods between 2010-2015

There were other variables considered during the analysis but eventually dismissed because they were not significant and affected the model negatively:

ln_PINV- log-transformation of the average public investment, in USD, in 2013. This data was obtained from Finanzas Publicas Municipales, INAFED, 2013.

Ln_VDEATH- log-transformation of the percentage of violent deaths respect to all deaths in 2013. Theoretically speaking, this variable was relevant because of the security issues that prevail in Mexico currently. Surprisingly, it had to be dismissed during the analyses due to its lack of significance.

AVGSCHOOLI- average number of years at school in the municipality. This variable was downloaded from INAFED's website in March 2016, though the data was authored (probably because of some processing) by the United Nations Development Programme (UNDP) to calculate the Human Development Index (HDI) at the municipal level.

PCURBAN- percentage of the population living in urban areas. This data was obtained from INEGI's Encuesta Intercensal extended questionnaire in Mexico (2015), was and calculated using tables depicting the distribution of the population by size of the settlement. INEGI's definition of urban (settlements with population equal to or higher

than 2500 are considered urban) was used during this calculation. However, when integrated into the model it was not significant, and was therefore dismissed.

AREA- Using INEGI's latest geo-statistical framework (2014) as shapefile, the areal extent in km2 was calculated using ArcMap 10.3. The inclusion of this variable was assumed to control for potential MAUP problems, or account for the differences in units' sizes. Contrary to what was expected, this variable hindered the overall performance of the model; therefore, it was dismissed.

6.1. Data descriptives

The spatial distribution of in-migrants and out-migrants in Mexico depict identifiable patterns. In terms of in-migration, the areas concentrated on the South West and some on the central-north show less than average rates (<-.50 standard deviations); whereas municipalities concentrated along the border with the United Sates, along with Los Cabos, Cancun, and areas close to the main metropolitan areas: Mexico City, Monterrey and Guadalajara, depict higher than average in-migration rates. With respect to out-migration, the above-average rates are observed among the northern municipalities, and some along the east coast, including again areas close to the main metro areas. Below-average out-migration is mostly concentrated on the central areas, and within Tehuantepec's isthmus. See figure 2.

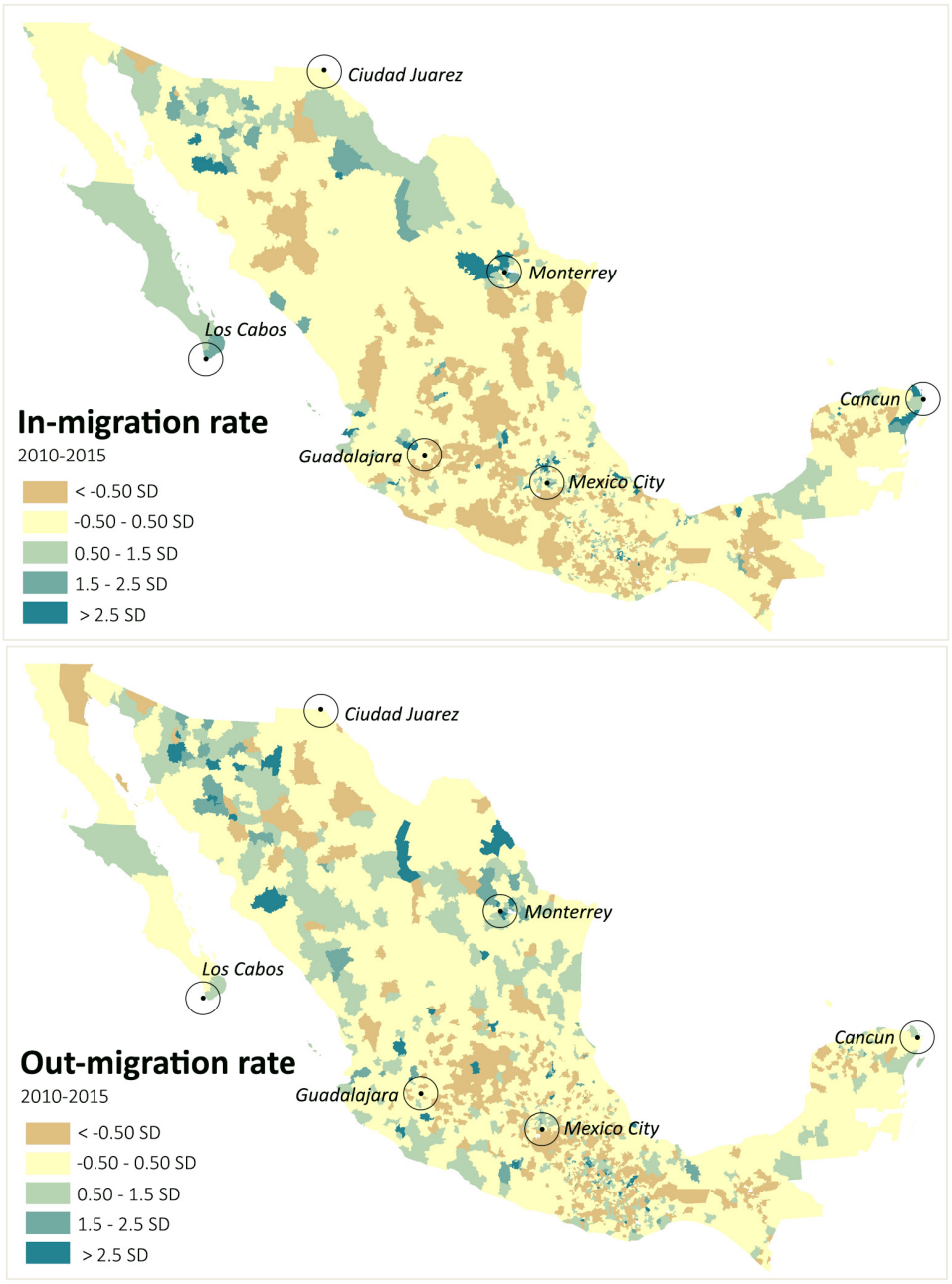


Figure 2 In-migration rate (above) and out-migration rate (below) in Mexico, annual average between 2010 and 2015.

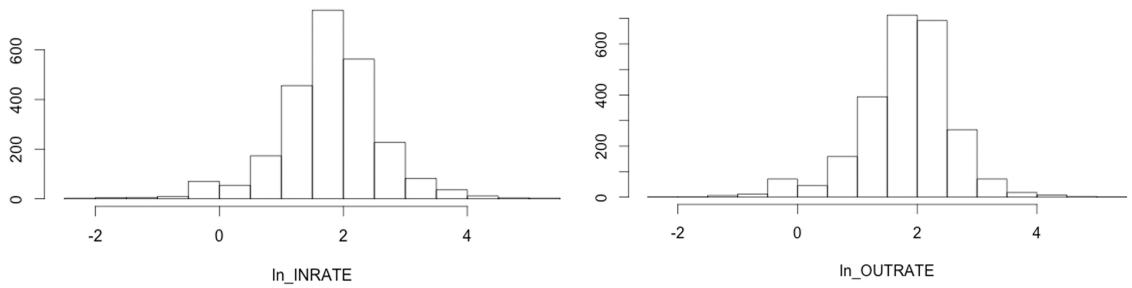


Figure 3 Histograms of log-transformed variables: in-migration rate (left) and out-migration rate (right) in Mexico.

After logarithmic transformations, the variables ln_OUTRATE, ln_INRATE, and ln_ECONEST depict 'close-to' normal bell-shaped distributions; which is what was expected and needed to run the model (see figure 3). EXPSCHOOLI's histogram shows a some of the observations with values close to zero. The values not close to zero follow a bell- shape curve. In terms of ln_INCOME, some outliers with values close to zero can be identified, though most of the data observations lie between ln values 8 and 10. The least normal-shaped is the variable ln_ALLHAZARDS, with the majority of the observations equal to zero (see figure 4).

Logarithmic transformations of values equal to zero were not log-transformed; instead, were instantly assumed 'zero'. This strategy was made in order to avoid transformed values equal to infinity, similar to what Saldaña-Zorrilla and Sandberg did (2009).

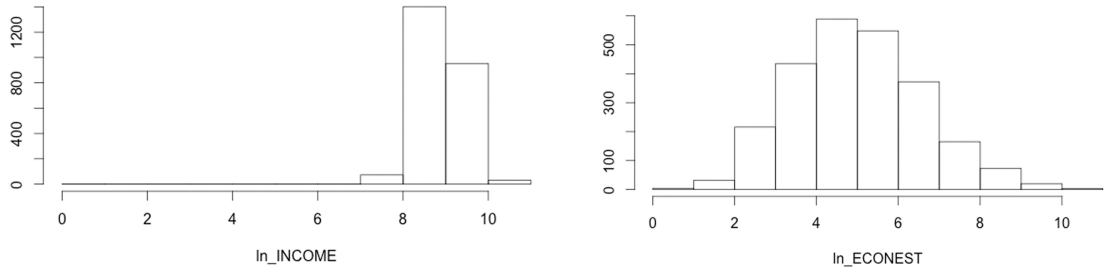


Figure 4 Histograms of log-transformed variables: income, 2010 (left) and economic establishments 2010-2014 (right).

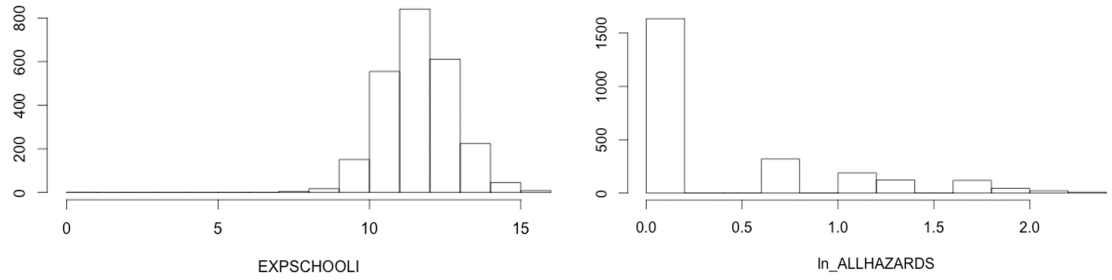


Figure 5 Histograms of log-transformed variables: expected average schooling 2010 (left); and frequency of total hazards 2011-2015 (right).

There is an interesting difference to observe in terms of the spatial distribution of hazardous events. Municipalities colored in red represent actual hazardous instances classified as state of emergency by Mexican authorities between 2011 and 2015. Municipalities colored in green refer to the percentage of rural localities within municipalities, where damages from hazardous events were reported by key informants between 2010 and 2014. Even though the time frames are shifted for a year, the differences are comparable, and still significant. Interviews may be capturing all levels of the 'damages' spectrum, whereas data from Mexican authorities is focusing only on the events with the greatest impact, or highest intensity. Therefore, it would be interesting to test the same model with both types of variables, to see whether results can be validated. One important difference is the data based on interviews are only applicable to rural localities, potentially biasing results at municipalities within a mixture of urban and rural localities. In that sense, the interview data has an important limitation in that respect. See figures 6, 7, 8, 9 and 10 for all hazard-related maps.

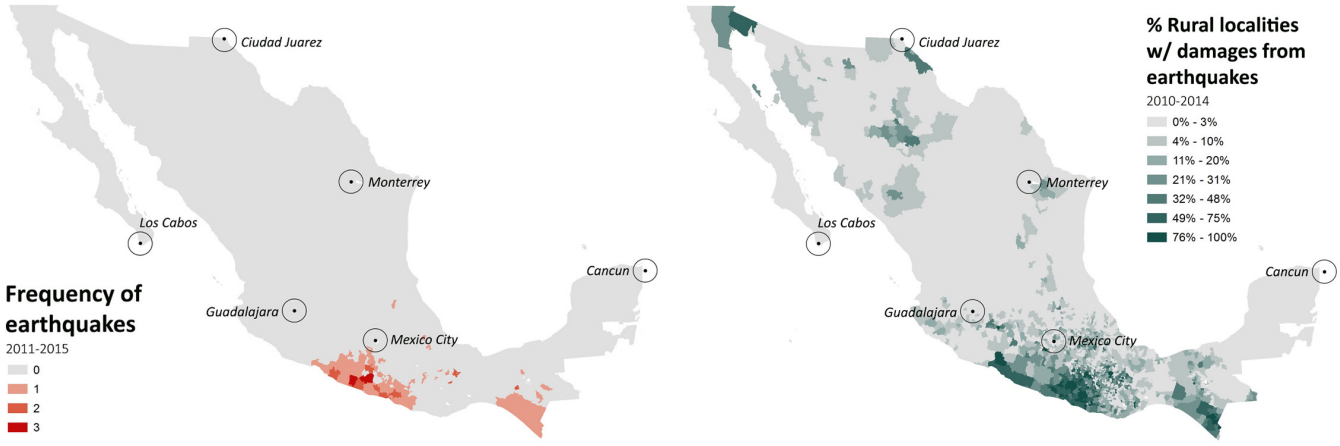


Figure 6 Frequency of earthquakes (left) from National System of Civil Protection 2011-2015; and percentage of localities with reported damages from earthquakes (right) from INEGI's 2015 Inter-census survey 2010-2014.

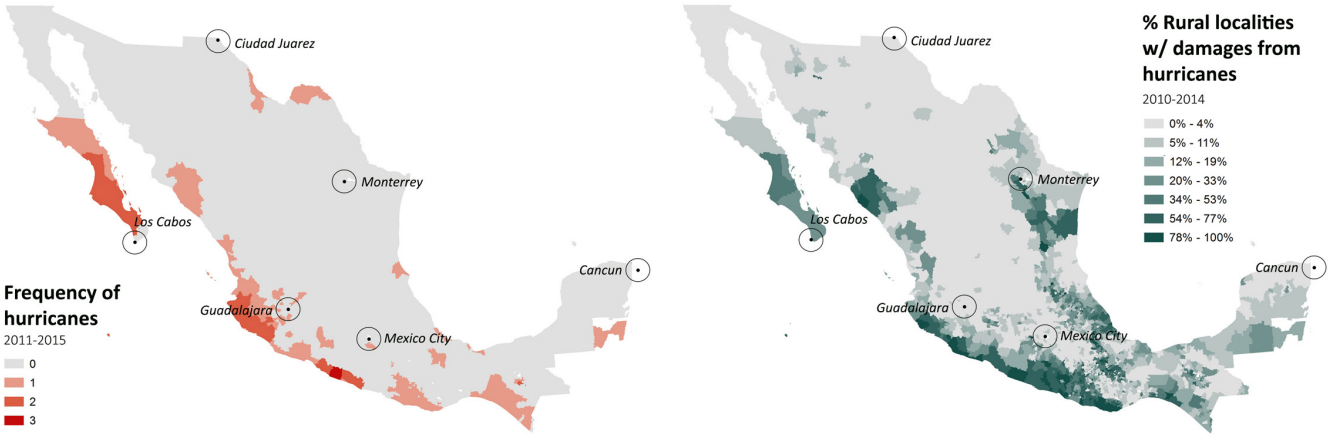


Figure 7 Frequency of hurricanes (left) from National System of Civil Protection 2011-2015; and percentage of localities with reported damages from hurricanes (right) from INEGI's 2015 Inter-census survey 2010-2014.

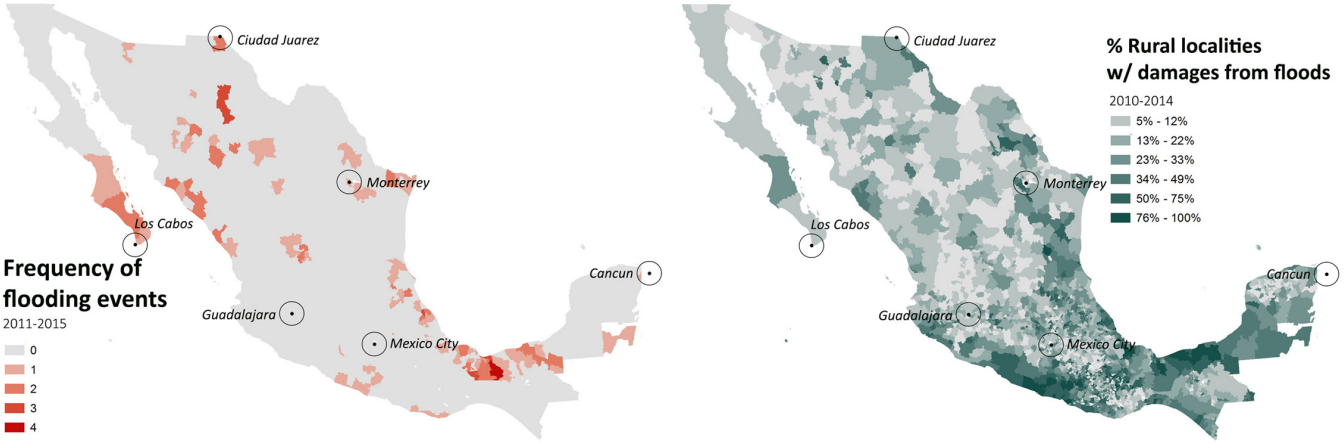


Figure 8 Frequency of flooding events (left) from National System of Civil Protection 2011-2015; and percentage of localities with reported damages from flooding events (right), from INEGI's 2015 Inter-census survey 2010-2014.

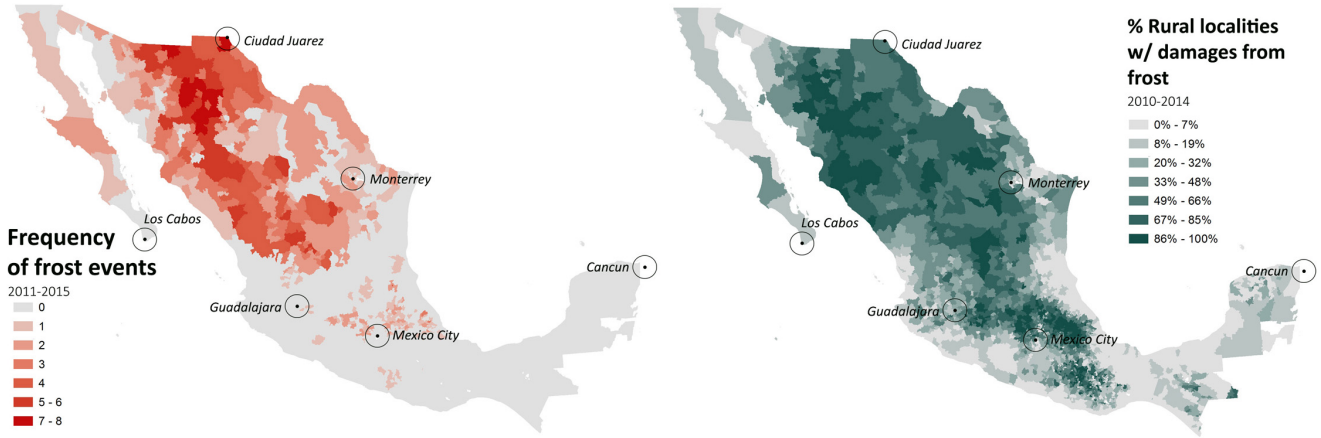


Figure 9 Frequency of frost, hail and snow storm events (left) from National System of Civil Protection 2011-2015; and percentage of localities with reported damages from frost, hail and snow storm events (right), from INEGI's 2015 Inter-census survey 2010-2014.

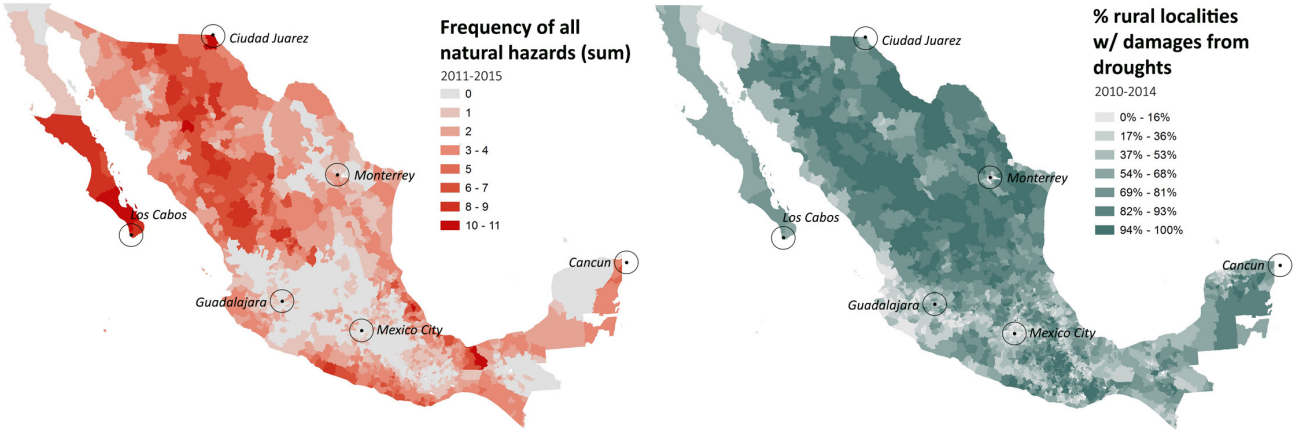


Figure 10 Frequency of all natural hazards catalogued as state of emergency (left) from National System of Civil Protection 2011-2015. Types of events included are: flooding, frost, hail and snow, hurricanes, heat waves, and earthquakes. Percentage of localities with reported damages from drougths (right) from INEGI's 2015 Inter-census survey 2010-2014.

7. Methodology

The analysis starts with the log-transformation of the variables of interest, depending on the overall distribution. Subsequently, the Variance Inflation Factor (VIF) is calculated using Marcus Beck's algorithm to estimate the VIF incorporating the stepwise function[11]. This functionality allows re-calculating the VIF as variables are dropped out due to collinearity. A threshold value of 5 was used, and three different scenarios were tested:

1) Dependent variables + socio- economic variables + set of disastrous events, entered separately.
2) Dependent variables + socio-economic variables + sum of disastrous events.
3) Dependent variables + socio-economic variables + set of damages from disasters, based on perception.

Once the definitive, independent variables were identified --which are the same explained in the "Data sources and variable construction" section--, Ordinary Least Square (OLS) linear regressions were performed. The stepwise function for linear

[11] For more information, see https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/

regression models in r was run to identify the variables that constituted best possible fit. In most cases, the stepwise function did not include the variables of interest (i.e. variables of disastrous events) as part of the final model. Since these are the variables to test in the analysis, these were manually included later on, so that the final models accounted for these variables, even if they did not contribute to the overall fitness –which in this case is treated as a result of the study. In this sense, six linear models resulted for further testing.

In-migration OLS models:

IN1 →

$$\ln_{INRATE} = \ln_{INCOME} + \ln_{ECONEST} + EXPSCHOOLI + \ln_{FLOOD} + \ln_{FROST} + \ln_{HURR} + \ln_{RAIN} + \ln \_EQK$$

IN2 →

$$\ln_{INRATE} = \ln_{INCOME} + \ln_{ECONEST} + EXPSCHOOLI + \ln \_ALLHAZARDS$$

IN3* →

$$\ln_{INRATE} = \ln_{INCOME} + \ln_{ECONEST} + EXPSCHOOLI + PDROUGHT + PFROST + PHURR + PEQK + PFLOOD$$

* This model is only tested among 840 municipalities (34.35% of Mexican municipalities)

Out-migration OLS models:

OUT1 →

$$\ln_{OUTRATE} = \ln_{INCOME} + \ln_{ECONEST} + EXPSCHOOLI + \ln_{RAIN} + \ln_{FROST}$$

OUT2 →

$$\ln_{OUTRATE} = \ln_{INCOME} + \ln_{ECONEST} + EXPSCHOOLI + \ln \_ALLHAZARDS$$

OUT3** →

$$\ln_{OUTRATE} = \ln_{INCOME} + \ln_{ECONEST} + EXPSCHOOLI + PDROUGHT + PFROST + PHURR + PEQK + PFLOOD$$

** This model is only tested among 840 municipalities (34.35% of Mexican municipalities)

Residual plots were conducted in order to identify potential patterns among residuals, as well as influential observations that could be biasing the results. Bonferonni's outlier test was also conducted in order to identify potential outliers. As a result of this screening, the variable ln_PINV (originally considered) was excluded, and a total of 12 outlier observations were dismissed from the analysis. OLS modeling was run again under these new considerations (see final residual plots in Annex 1).

Studentized Breusch-Pagan (BP) tests were conducted in all six instances to test for heteroskedasticity. As it can be seen in tables 1 and 2, as well as with the residual plots in Annex 1, the BP test is highly significant, which confirms the strong presence of heteroskedasticity in all six models. The literature proposes various alternatives to correct for this, and running linear models with robust standard errors is one of them (Berry, 1993; Kaufman, 2013). However, part of the spatial modeling process wants to incorporate heteroskedastic residuals to see whether the spatial lag component can explain (some) the excess variance in the residuals; or to model these errors spatially. Therefore, heteroskedasticity was not corrected in OLS models.

The next step is to test for spatial autocorrelation through the Moran's I statistic. As it can be seen in tables 1 and 2, models IN1, IN2, OUT1, and OUT2 present positive and significant spatial autocorrelation—and this what was expected. Unfortunately, models IN3 and OUT3 could not go further down the analysis given that a significant portion of the 840 rural municipalities do not share borders with any other unit –they are isolated units in space. Therefore, the contiguity matrix cannot be built, so spatial analysis cannot be conducted in these two cases.

After spatial autocorrelation is confirmed, all four Lagrange Multiplier tests (LM tests) were conducted, in order to select the adequate spatial model (either Spatial lag or Spatial error models), following Anselin's (2003) methodology. Tables 1 and 2 provides the results in all four cases, as well as coefficient estimates, AIC values and the Studentized Breusch-Pagan tests values with residuals from spatial modeling.

8. Results

In-migration rates

When comparing all three OLS models, ln_INCOME, ln_ECONEST, and EXPSCHOOLI are all highly significant. The order of magnitude of these coefficients among all three models does not change drastically. Only ln_RAIN and ln_FROST were significant in IN1-OLS, and no *disastrous* variables were significant for IN2-OLS and IN3-OLS. All three OLS models are highly heteroskedastic (see table 1). IN1 and IN2 present significantly positive spatial autocorrelations with Moran's I statistics of ~0.33 in both cases.

Moving on to the spatial models, IN1-spatial provides a better fit than IN1-OLS with an AIC = 4246 and log likelihood = -2112, compared to AIC = 4796; log likelihood = -2388 from IN1-OLS. Likewise, IN2-spatial provides a better fit than IN2-OLS with an AIC = 4240 and log likelihood = -2113, compared to an AIC = 4806 and log likelihood = -2397 from IN2-OLS. In both cases, the spatial model provided a better fit for the data than OLS.

Following the methodology from Anselin (2002) and Anselin (2003) in regards to model selection from LM testing, a spatial error model for IN1 and a spatial lag model for IN2 were conducted. In the case of IN1-spatial, the significant effect of lambda

responds to the clustering effect of errors in the model. However, the BP test conducted with residuals from IN1-spatial indicates the presence of other unexplained error patterns happening as well inside the model (BP test = 18.28).

In terms of the effects of the coefficients, the signs and levels of association were as expected in regards to the variables ln_INCOME and EXPSCHOOLI. For each logarithmic unit increase in average income as well as expected schooling years, the rate of in-migration increases as well[12]. The decreasing effect of ln_ECONEST is contrary to what was expected. This variable was introduced as a proxy for size of local economy; therefore the expectation was a positive relationship. In this sense, it is possible that this metric is measuring other economic dimension related to the sectorization of economic establishments rather than the actual size of the economy in the municipality. In terms of hazardous variables, only ln_FROST was significant and with a negative sign, which is in agreement to what is expected in a risky scenario: for each logarithmic unit increase in frost, hail or snow storm event, the rate of in-migration decreases. The rest of the *hazardous* variables did not present significant effects.

In regards to IN2-spatial (spatial lag model), the significant positive effect of rho is indicative of a spatial dependence inherent in the data. A portion of the error term remains unexplained, as shown in the result from the BP test (BP test = 13.33). IN2-spatial provides a better explanatory fit than IN1-spatial, as shown in both AIC and log likelihood scores (see table 1). The effect of variables ln_INCOME, EXPSCHOOLI and ln_ECONEST are very similar to the ones from IN1-spatial. In addition, the variable ln_ALLHAZARDS presents a significant coefficient with a negative sign, which

[12] In this study I will not make reference of exact coefficients in the results because the interpretation is too convoluted. The interest is mostly to show association among proxy variables to relevant socio-economic dimensions related to in and out-migration, as well as with the presence of hazardous events.

corresponds to the expectation: for each logarithmic unit increase in *any* hazardous event classified as emergency, the rate of in-migration to that place decreases.

Table 1 Summary results for in-migration rate modeling

Parameter	OLS			Spatial	
	IN1	IN2	IN3	IN1	IN2
Rho	-	-	-	-	0.58449*
Lambda	-	-	-	0.57932*	-
Intercept	-6.520***	-6.313***	-6.113***	-6.2594***	-6.2476***
ln_INCOME	0.8662***	0.8269***	0.7979***	0.92900***	0.92650***
ln_ECONEST	-0.116***	-0.109***	-0.114***	-0.1074***	-0.1062***
EXPSCHOOLI	0.1039***	0.1132***	0.1178***	0.03112+	0.032381*
ln_FLOOD	-0.1070	-	-	-0.09933	-
Ln_FROST	-0.0847*	-	-	-0.09927+	-
Ln_RAIN	0.1202***	-	-	0.00946	-
Ln_HURR	0.1403	-	-	0.05918	-
Ln_EQK	0.2269	-	-	0.17555	-
Ln_ALLHAZARDS	-	0.02186	-	-	-0.05272+
PDROUGHT	-	-	0.00070	-	-
PFLOOD	-	-	0.00070	-	-
PFROST	-	-	0.00037	-	-
PHURR	-	-	-0.0015	-	-
PEQK	-	-	-0.00021	-	-
Adj R^2	0.3111	0.3069	0.2949	-	-
F-statistic	139***	271.6***	44.86***	-	-
LR test	-	-	-	551.92***	568***
AIC	4796.1	4806.9	1682.747	4246.2	4240.9
Log likelihood	-2388.061	-2397.466	-831.3	-2112.1	-2113.465
Heteroskedasticity Breusch-Pagan test	23.683**	20.638***	13.789+	18.828*	13.33**
Moran's I	0.3318***	0.3391***	-	-0.02592*	-0.02686*
Lagrange Multiplier (LM) error	703.08***	725.99***	-	-	-
LM-lag	650.01***	656.73***	-	-	-
LM- Robust error	70.662***	81.966***	-	-	-
LM- Robust lag	17.6***	12.705***	-	-	-

+ P-value < 0.10, * P-value < 0.05, ** P-value < 0.01, *** P-value < 0.001

Table 2. Summary results for out-migration rate modeling

Parameter	OLS			Spatial	
	OUT1	OUT2	OUT3	OUT1	OUT2
Rho	-	-	-	-	0.40483*
Lambda	-	-	-	0.43176*	-
Intercept	-2.488***	-2.491***	-3.50***	-2.422***	-2.1008***
ln_INCOME	0.3967***	0.3950***	0.561***	0.3925***	0.27123***
ln_ECONEST	-0.045***	-0.048***	-0.055	-0.02426+	-0.0262*
EXPSCHOOLI	0.0841***	0.0847***	0.051	0.0721***	0.07353***
ln_FLOOD	-	-	-	-	-
Ln_FROST	0.1710***	-	-	0.14580**	-
Ln_RAIN	0.1658***	-	-	0.09520*	-
Ln_HURR	-	-	-	-	-
Ln_EQK	-	-	-	-	-
Ln_ALLHAZARDS	-	0.1917***	-	-	0.12450***
PDROUGHT	-	-	0.0032**	-	-
PFLOOD	-	-	-0.005***	-	-
PFROST	-	-	-0.0018+	-	-
PHURR	-	-	0.0038**	-	-
PEQK	-	-	-0.0009	-	-
Adj R^2	0.1159	0.1258	0.1036		
F-statistic	65.07***	88.96***	13.12***		
LR test	-	-	-	260.16***	243.81***
AIC	5262	5233.3	5237.1	5003.8	4991.5
Log likelihood	-2623.978	-2610.636	-1069.91	-2493.896	-2488.733
Heteroskedasticity Breusch-Pagan test	77.219***	83.218***	24.177**	93.054***	104.54***
Moran's I	0.2284***	0.2284***	-	-0.0067 +	-0.0184 +
Lagrange Multiplier (LM) error	329.28***	304.55***	-	-	-
LM-lag	327.96***	306.73***	-	-	-
LM- Robust error	6.7621**	5.1132*	-	-	-
LM- Robust lag	5.4447*	7.2912**	-	-	-

+ P-value < 0.10, * P-value < 0.05, ** P-value < 0.01, *** P-value < 0.001

Out-migration rates

When looking at all three OLS models, OUT1 and OUT2 present similar scores for ln_INCOME, ln_ECONEST, and EXPSCHOOLI. Interestingly, only ln_INCOME is significant in OUT3. Similar to IN1-OLS, only ln_RAIN and ln_FROST are significant in OUT1. However, the ln_ALLHAZARDS coefficient is highly significant in OUT2, and all variables related to potential (or perceived) damages from hazards are significant

in OUT3, except for PEQK. Similar to OLS models for in-migration, all three OLS out-migration models are highly heteroskedastic (see table 2). OUT1 and OUT2 present significantly positive spatial autocorrelations with Moran's I statistics of ~0.22 in both cases.

Moving on to the spatial models, the OUT1-spatial model provides a better fit than OUT1-OLS with better AIC and log likelihood scores. A similar situation can be depicted in OUT2-spatial compared to OUT2-OLS (see table 2). In both cases, the spatial modeling provided a better fit for the data, accounting for clustering patterns in the error term.

Following the same methodology mentioned for in-migration modeling (Anselin, 2002; Anselin, 2003), OUT1-spatial resulted in a spatial error model, and OUT2-spatial resulted in a spatial lag model.

Similar to coefficient results for in-migration modeling, the signs for ln_INCOME and EXPSCHOOLI are positive, as expected. And in this case, the decreasing effect of ln_ECONEST resulted as expected: for each logarithmic unit increase in the number of economic establishments, out-migration rate decreases. However, the effect –although not really clear given all the data transformations performed to fit in the model—is quite low. In terms of hazardous variables, the same ln_FROST and ln_RAIN are significant and with the expected signs (i.e. positive).

In regards to OUT2-spatial (spatial lag model), the significant positive effect of rho is indicative of a spatial clustering; this result is similar for IN2-spatial. However, the model is still heteroskedastic with a significant P-value in the BP-test (BP-test = 104.54). Similar effects as in OUT1-spatial for ln_INCOME, EXPSCHOOLI, and ln_ECONEST

are also found in OUT2-spatial. Moreover, the variable ln_ALLHAZARDS is also significant for OUT2-spatial.

The spatial lag model (i.e. IN2-spatial for in-migration, and OUT2-spatial for out-migration) with the ln_ALLHAZARDS variable proved to have better fit, overall.

Now, rigorously speaking, in the presence of significant P-values on BP tests (i.e. that is, in the presence of unexplained variances in the error term that are not accounted for in the spatial lag or spatial error term), one suggestion is to conduct alternative modeling such as Geographic Weighted Regressions (GWR) (Bivand, et al., 2008). However, as Bivand et al. (2008), Schabenberger, O. and Gotway (2005), Krivoruchko, K. (n.d.) and others have pointed out, GWR is an exploratory data analysis technique, not a regression model. Explorations on the application of GWR on this topic, although interesting, are outside of the scope of this study. In that sense, even though the resulted spatial models in all four cases present significant heteroskedasticity (and potentially non-stationarity), no further modeling will be conducted.

8.1. Comparing both in-migration and out-migration models

Table 3 shows a summary of the results from the best fitting models for in-migration and out-migration rates. It is interesting to see how income (ln_INCOME) has an important effect on in-migration, compared to estimates from out-migration model. This result is in line with what Harris and Todaro (1970), Lee (1965), Massey et al. (1993) refer to in terms of *expected* income as one of the main drivers for migration. Expected income acts as a pull factor for in-migrants in Mexico.

Table 3 Summary table of resulting spatial models, comparing the effects on in-migration and out-migration rates

Parameter	**IN2-spatial**	**OUT2-spatial**
Rho	0.58449*	0.40483*
Intercept	-6.2476***	-2.1008***
ln_INCOME	0.92650***	0.27123***
ln_ECONEST	-0.1062***	-0.0262*
EXPSCHOOLI	0.032381*	0.07353***
Ln_ALLHAZARDS	-0.05272+	0.12450***
RMSE (residuals)	0.5532589	0.6587318

+ P-value < 0.10, * P-value < 0.05, ** P-value < 0.01, *** P-value < 0.001

The effect of ln_ECONEST is inconclusive. The negative effect for out-migration is as expected, but the level of magnitude is less than with in-migration. Hence, this variable may be measuring an unexplained dimension other than size of the local economy, in the case of in-migration.

In terms of EXPSCHOOLI, the effect of the coefficient may be tricky. The variable refers to the average of *expected years at school*, which refers to a probable scenario, given the current age structure in the municipality. It does not refer to the actual average years at school[13]. A possible interpretation on model estimates is that the variable somehow depicts a proxy for age-structure, where places with prevailing youth population may be associated with higher levels of out-migration, and to a lesser extent, with in-migration.

Finally, the variable of interest, ln_ALLHAZARDS, presents significant and consistent effects for both in-migration and out-migration rates modeling. The order of magnitude is higher for out-migration than it is for in-migration, which is expected.

[13] This variable AVGSCHOOLI was included in the initial models testing, and was not significant.

9. Discussion

So what can we say about the incidence of natural hazards in relation to internal migration in Mexico? Internal migration in Mexico is a spatial problem. There is a significant, spatial association between the frequency of natural hazards and internal migration in Mexico. The models with the best fit for both in-migration and out-migration considered hazardous events as aggregates, rather than individual-type events. This finding is similar to Saldaña-Zorrilla and Sandberg's for emigration rates in Mexico between 1990 and 2000 (Saldaña-Zorrilla et al., 2009). The effect of natural hazards is more significant for out-migration than for in-migration, which means that the frequent occurrence of state of emergency-type events has more significance as a push factor than as a *negative* pull factor. In this sense, these two findings answer the initial research questions #1 and #3.

In regards to the research question #2, the fact that the data from reported damages in rural areas could not be modeled using Spatial Lag or Spatial Error models does not mean that there is something wrong with the data. On the contrary, perception data on natural disasters brings in the important social component which tacit among official records, though not always identifiable. In that sense, as a future line of research, internal migration could be modeled from a social vulnerability perspective, using individual-level data. INEGI's micro data for 2015 allows to tag the code of each rural locality with the individual. Therefore, it is possible to see if the perceived damages have a significant effect on internal migration, and to what extent. Modeling migration using micro data may also solve the issue of heteroskedasticity. However, it assumes that social groups may have homogenous error terms' means and variances—and I suspect they may

not due to intrinsic differences on income and educational among in and out-migrants… but this may be part of another study. Likewise, other more appropriate techniques may be used, like for example GWR or probabilistic spatial methods.

10. References

Adamo, S. and de Sherbinin, A. (2011). "The impact of Climate Change on the spatial distribution of the population and on migration". Population Distribution, Urbanization, Internal Migration and Development: An International Perspective. United Nations Department of Economic and Social Affairs, Population Division. New York: 161-195

Adamo, S. and H. Izazola (2010). "Human migration and the environment." Population and Environment 32(2): 105-108.

Agresti, A. and Finlay, B. (2009) Statistical Methods for the Social Sciences, 4th Ed. Essex: Pearson Education Limited.

Alscher, S. (2010). Environmental Factors in Mexican Migration: The Cases of Chiapas and Tlaxcala. Environment, Forced Migration and Social Vulnerability. T. Afifi and J. Jäger. Berlin, Heidelberg, Springer Berlin Heidelberg: 171-185.

Anselin, L. (1988) Model Validation in Spatial Econometrics: A Review and Evaluation of Alternative Approaches. International Regional Science Review, Vol. 11, No. 3 279-316.

Anselin, L. (1998) Spatial Econometrics: Methods and models. Dordrecht, The Netherlands: Kluwer.

Anselin, L. (2002) Under the hood: Issues in the specification and interpretation of spatial regression models. Agricultural Economics, 27 (2002) 247-267.

Anselin, L. (2003) An Introduction to Spatial Regression Analysis in R. University of Illinois, Urbana-Champaign http://sal.agecon.uiuc.edu

Audley, J. J., et al. (2004). NAFTA's Promise and Reality. Lessons from Mexico for the Hemisphere. Washington, DC, Carnegie Endowment for International Peace.

Beck, M. (2013) Collinearity and Stepwise VIF Selection. https://beckmw.wordpress.com/2013/02/05/collinearity-and-stepwise-vif-selection/ Accessed on April 2, 2016.

Berry, W. D. (1993). Understanding Regression Assumptions. Series: Quantitative Applications in the Social Sciences, Thousand Oaks, California: SAGE Publications.

Bivand, R.S., Pebesma, E.J., Gómez-Rubio, V. (2008) Applied Spatial Data Analysis with R. Series: Use R! New York, NY: Springer

Black, R., et al. (2011). "The Effect of Environmental Change on Human Migration." Global Environmental Change 21S (2011): S3- S11.

Black, R., et al. (2013). "Migration, immobility and displacement outcomes following extreme events." Environmental Science & Policy 27, Supplement 1: S32-S43.

Centro Latinoamericano y Caribeño de Demografía CELADE (2016). MIALC: Migración Interna en Latino América y el Caribe http://www.cepal.org/celade/migracion/migracion_interna/ Accessed on April 2, 2016.

Cerrutti, M. and D. S. Massey (2001). "On the auspices of female migration from Mexico to the United States." Demography 38(2): 187-200.

Cohen, J. H. (2004). The Culture of Migration in Southern Mexico. Austin, Texas, University of Texas Press.

Chi, J., Hayatdavoudi, L., Kruszona, S., Rowe, B. (2013) Reducing Drug Violence in Mexico: Options for Implementing Targeted Enforcement. University of

California, Los Angeles, Luskin School of Public Affairs (Applied Policy Project). Advisor: Mark A.R. Kleiman.

de Janvry, A., et al. (1997). Determinants of Mexico-US migration: the role of household assets and environmental factors. CUDARE Working Papers. California, University of California at Berkeley.

Durand, J. (2007). Braceros. Las miradas mexicana y estadounidense. Antologia (1945-1964). Mexico DF, Miguel Angel Porrua.

Ervin, D. MAUP. An Introduction to the Modifiable Areal Unit Problem. Advanced Spatial Analysis in the Population Sciences and Spatial Demography (GISpopsci.org) http://gispopsci.org/maup/ Accessed on April 12, 2016.

Feng, S., et al. (2010). "Linkages among climate change, crop yields and Mexico–US cross-border migration." Proceedings of the National Academy of Sciences 107(32): 14257-14262.

Fausset, R. (February 28, 2013) Mexican Towns, Once Frozen With Fear, Now Frozen in Time. Los Angeles Times.

Harris, J. R. and Todaro, M. P. (1970). Migration, Unemployment and Development: A Two-Sector Analysis. The American Economic Review, Vol. 60, No. 1, 126-142.

Hunter, L. M., Luna, J., Norton, R. (2015). "Environmental Dimensions of Migration." Annual Review of Sociology 41(1): 377-397.

Hunter, L. M., Murray, S., Riosmena, F. (2013). "Rainfall Patterns and U.S. Migration from Rural Mexico." International Migration Review 47(4): 874-909.

Instituto Nacional de Estadística y Geografía INEGI (2015). Directorio Estadístico Nacional de Unidades Económicas (DENUE). México, D.F. http://www.beta.inegi.org.mx/app/descarga/?ti=1 Accessed on March 26, 2016.

Instituto Nacional de Estadística y Geografía INEGI (2015). Encuesta Intercensal 2015. México, D.F. http://www.inegi.org.mx/est/contenidos/Proyectos/encuestas/hogares/especiales/ei2015/ Accessed on November 11, 2015.

IPCC, 2013: Summary for Policymakers. In: Climate Change 2013: The Physical Science Basis. Contribution of Working Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change [Stocker, T.F., D. Qin, G.-K. Plattner, M. Tignor, S.K. Allen, J. Boschung, A. Nauels, Y. Xia, V. Bex and P.M. Midgley (eds.)]. Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

Izazola, H. (2004). "Migration to and from Mexico City, 1995- 2000." Environment and Urbanization 16(1): 211-230.

Kaufman, R. L. (2013) Heteroskedasticity in Regression. Detection and Correction. Series: Quantitative Applications in the Social Sciences, Thousand Oaks, California: SAGE Publications.

Krivoruchko, K. (n.d.) Introduction to modeling spatial processes using Geostatistical Analyst, ESRI. http://www.esri.com/library/whitepapers/pdfs/intro-modeling.pdf

Laczko, F. (2010). Migration, Environment and Climate Change: Assessing the Evidence. Marseille, International Organization for Migration.

Lee, E. (1965) A theory of Migration. Paper presented at the Annual Meeting of the Mississippi Valley Historical Association, Kansas City, April 23 1965.

Levy, M.B. and Wadycki, W.J. (1974) “Education and the Decision to Migrate: An Econometric Analysis of Migration in Venezuela”. Econometrica, Vol. 42, No. 2, pp. 377-388.

Massey, D. S. (1990). Return to Aztlan. The Social Process of International Migration from Western Mexico. California, University of California Press.

Massey, D. S., Arango, J., Hugo, G., Kouaouci, A., Pellegrino, A., Taylor, E. (1993) Theories of International Migration: a Review and Appraisal. Population and Development Review, Vol. 19, No. 3 (Sep 1993) 431-466.

Massey, D. S., Axinn, G., Ghimire, J. (2010). "Environmental change and out-migration: evidence from Nepal." Population and Environment 32(2/3): 109-136.

Massey, D. S., Durand, J., Malone, J. (2009). Detras de la trama. Politicas migratorias entre Mexico y Estados Unidos. Mexico DF, Miguel Angel Porrua.

Moriniere, L. C. E. and M. Hamza (2012). "Environment and Mobility: A View from Four Discourses." Royal Swedish Academy of Sciences 2012(41): 795- 807.

Munshi, K. (2003). "Networks in the modern economy: Mexican migrants in the US labor market." Quarterly Journal of Economics(May 2003): 549-599.

Nawrotzki, R. J., L. M. Hunter, D. M. Runfola and F. Riosmena (2015). "Climate Change as Migration Driver from Rural and Urban Mexico." Environmental research letters : ERL [Web site] **10**(11).

Nawrotzki, R. J., F. Riosmena and L. M. Hunter (2013). "Do Rainfall Deficits Predict U.S.-bound Migration from Rural Mexico? Evidence from the Mexican Census." Population research and policy review **32**(1): 129-158.

Nawrotzki, R. J., F. Riosmena, L. M. Hunter and D. M. Runfola (2015). "Amplification or suppression: Social networks and the climate change—migration association in rural Mexico." Global Environmental Change **35**: 463-474.

Partida Bush, V. (2001) La Migración Interna, in La Población de México. Tendencias y Perspectivas Sociodemográficas hacia el siglo XIX. Gómez de León Cruces, J. and Rabell Romero, C. (cords). México DF: Consejo Nacional de Población, Fondo de Cultura Económica.

Pelaez Herreros, O. (2013) Desarrollo, transicion demografica y saldos migratorios: evidencia para los municipios de Mexico, 2000- 2010. Revista Iberoamericana de Estudios de Desarrollo 2, 38-62

Piguet, E. (2010). "Linking climate change, environmental degradation, and migration: a methodological overview." Wiley Interdisciplinary Reviews: Climate Change 1(4): 517-524.

Renaud, F. G., J. Bogardi, O. Dun and K. Warner (2007). Control, Adapt or Flee. How to Face Environmental Migration? Interdisciplinary Security Connections (InterSecTions). Bonn, United Nations University Institute for Environment and Human Security (UNU- EHS)**:** 48 p.

Renaud, F. G., O. Dun, K. Warner and J. Bogardi (2011). "A Decision Framework for Environmentally Induced Migration." International Migration **49**(S1): 5-29.

Reuveny, R. (2007). "Climate change-induced migration and violent conflict." Political Geography 26(6): 656-673.

Riosmena, F., R. Nawrotzki and L. M. Hunter (2013) "Rainfall trends, variability and US migration from rural Mexico: evidence from the 2010 Mexican census." IBS Population Program Working Paper.

Rodriguez-Oreggia, E., A. De La Fuente, R. De La Torre and H. A. Moreno (2013). "Natural Disasters, Human Development and Poverty at the Municipal Level in Mexico." Journal of Development Studies **49**(3): 442-455.

Runfola, D. M., P. Romero-Lankao, L. Jiang, L. M. Hunter, R. Nawrotzki and L. Sanchez (2013) "The Influence of Internal Migration on Exposure to Extreme Weather Events in Mexico." IBS Population Program Working Paper, 9 p.

Saldaña-Zorrilla, S. O. (2008). "Stakeholders' views in reducing rural vulnerability to natural disasters in Southern Mexico: Hazard exposure and coping and adaptive capacity." Global Environmental Change 18(4): 583-597.

Saldaña-Zorrilla, S. O. and K. Sandberg (2009). "Impact of climate-related disasters on human migration in Mexico: a spatial model." Climatic Change 96(1-2): 97-118.

Schabenberger, O. and Gotway, C. A. (2005) Statistical Methods for Spatial Data Analysis. Texts in Statistical Science. Boca Raton, Fl: Chapman & Hall.

Schwartz, M. L. and J. Notini (1994). Desertification and Migration: Mexico and the United States. San Francisco, California, Natural Heritage Institute.

Todaro, M. P. (1969) "A Model of Labor Migration and Urban Unemployment in Less Developed Countries". The American Economic Review, Vol. 59, No. 1 (1969) 138- 148.

United Nations Development Programme (UNDP) 2015. Average income in 2010. Obtained from Finanzas Públicas Municipales http://www.inafed.gob.mx/es/inafed/ Accessed on March 26, 2016.

United Nations Office for the Coordination of Humanitarian Affairs (UN-OCHA), (2011). Haiti: One Year Later. Accessed on April 12, 2016. http://www.unocha.org/issues-in-depth/haiti-one-year-later

Viton, P. (2010). Notes on Spatial Econometric Models. City and Regional Planning 870.03, unpublished manuscript.

Ward, M. and Gleditsch, K. S. (2008). Spatial Regression Models. Series: Quantitative Applications in the Social Sciences, Thousand Oaks, California: SAGE Publications.

Warner, K. (2010). "Global Environmental Change and Migration: Governance Challenges." Global Environmental Change 20(2010): 402- 413.

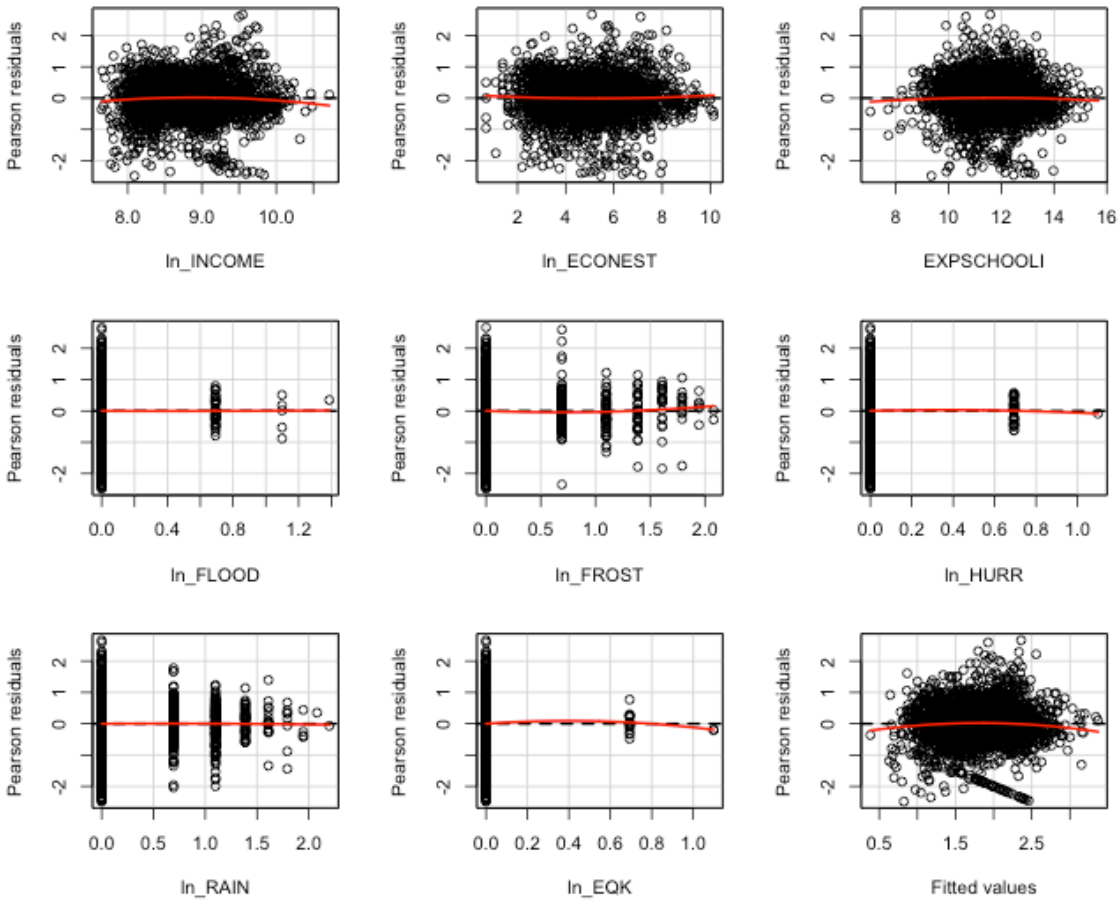
Warner, K., C. Ehrhart, A. de Sherbinin, S. Adamo and T. Chai-Onn (2009). In Search of Shelter. Mapping the Effects of Climate Change on Human Migration and Displacement. New York, NY, United Nations University Institute for Environment and Human Security; CARE International; Center for International Earth Science Information Network at Columbia University.: 36.

Winsemius, H. C., B. Jongman, T. I. E. Veldkamp, S. Hallegatte, M. Bangalore and P. J. Ward (2015). Disaster Risk, Climate Change and Poverty. Assessing the Global Exposure of Poor People to Floods and Droughts. Policy Research Working Paper. Washington, DC, World Bank Group CC Cross-Cutting Solutions Area: 36 p

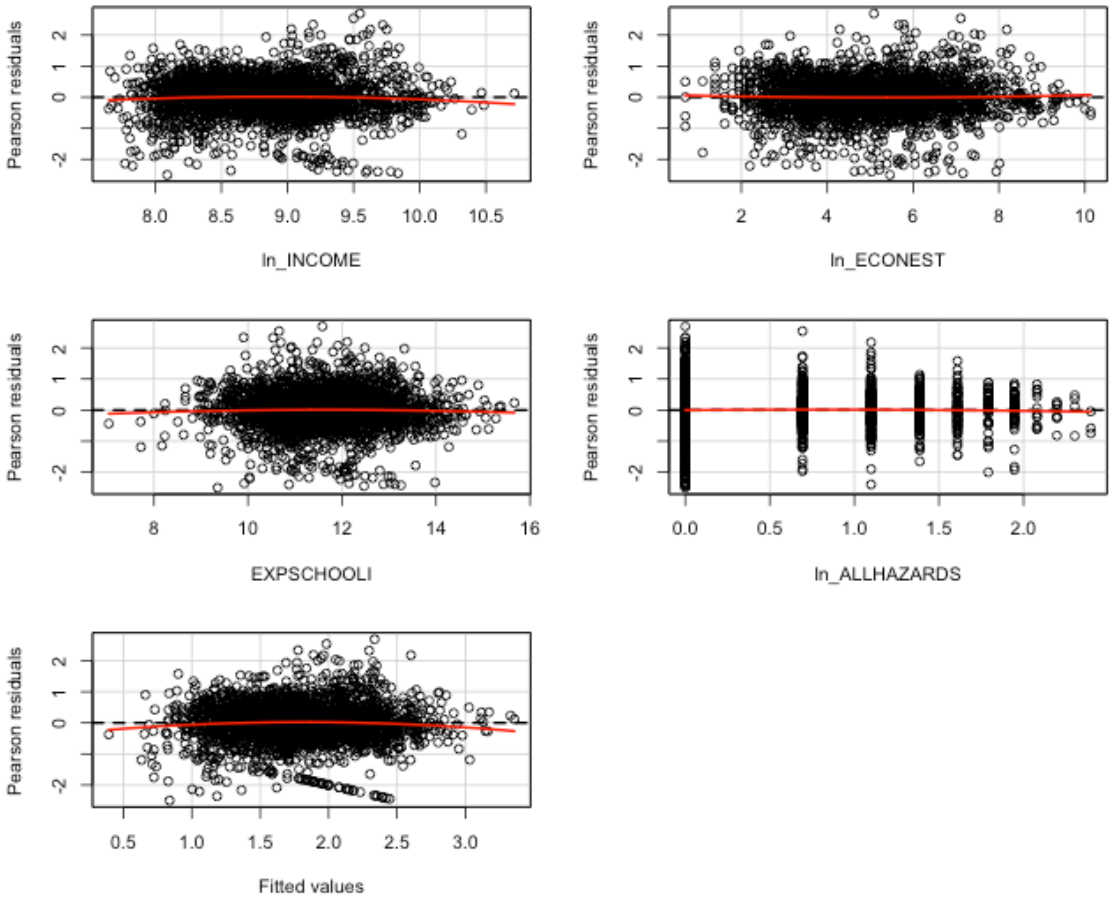
Zapponi, C. (2010) Peoplemovin http://www.peoplemov.in/ Accessed on January 25, 2016.

Annex 1. Residual Plots

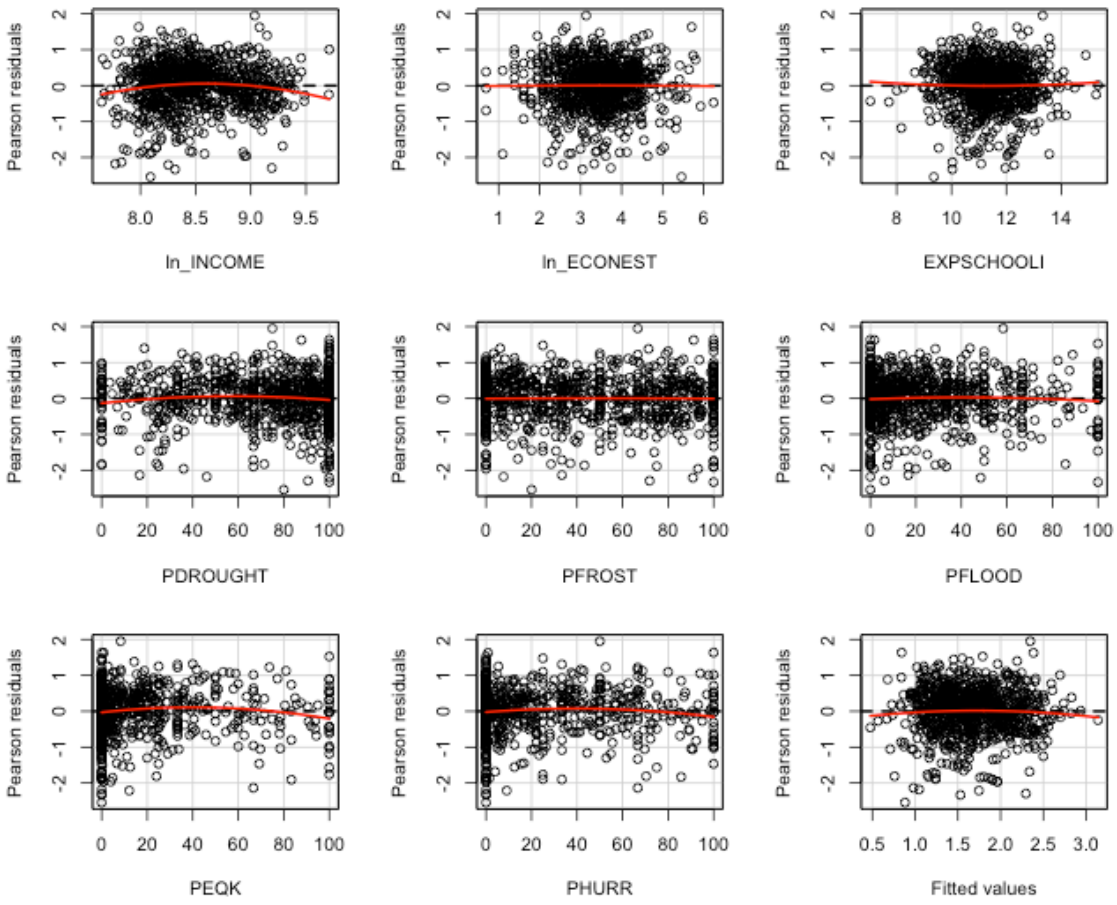
Model IN1



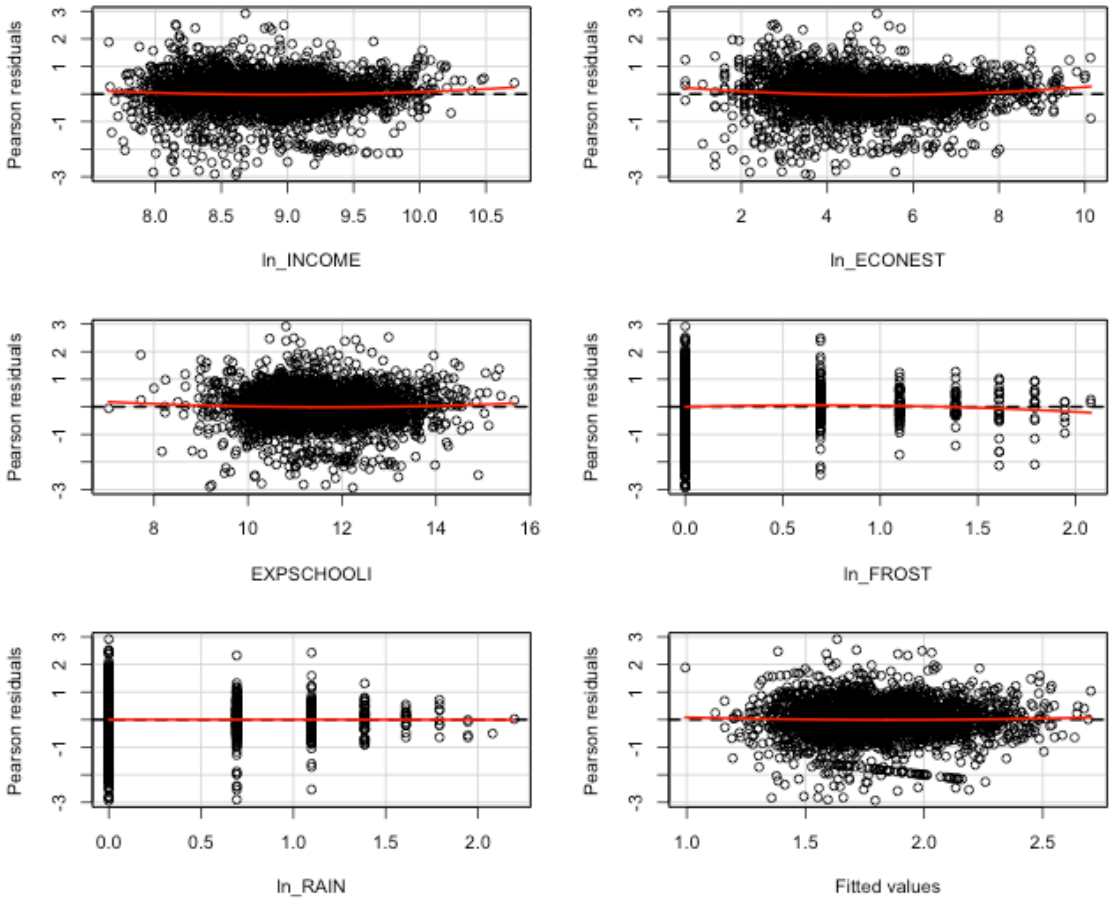
Model IN2


Pearson residuals
-2 0 1 2
8.0 8.5 9.0 9.5 10.0 10.5
ln_INCOME
Pearson residuals
-2 0 1 2
2 4 6 8 10
ln_ECONEST
Pearson residuals
-2 0 1 2
8 10 12 14 16
EXPSCHOOLI
Pearson residuals
-2 0 1 2
0.0 0.5 1.0 1.5 2.0
ln_ALLHAZARDS
Pearson residuals
-2 0 1 2
0.5 1.0 1.5 2.0 2.5 3.0
Fitted values

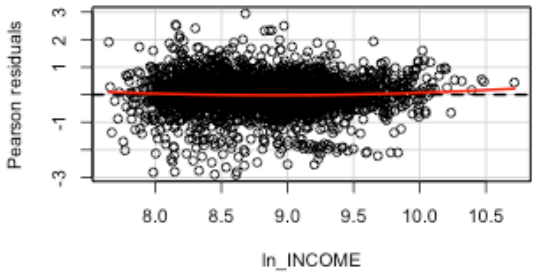
Model IN3

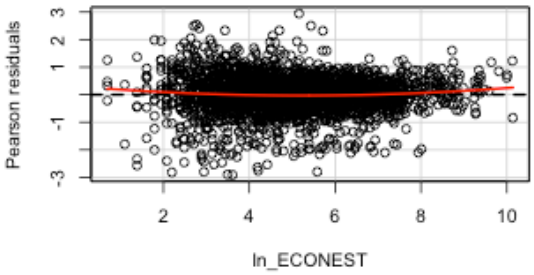

Pearson residuals
ln_INCOME
ln_ECONEST
EXPSCHOOLI
PDROUGHT
PFROST
PFLOOD
PEQK
PHURR
Fitted values

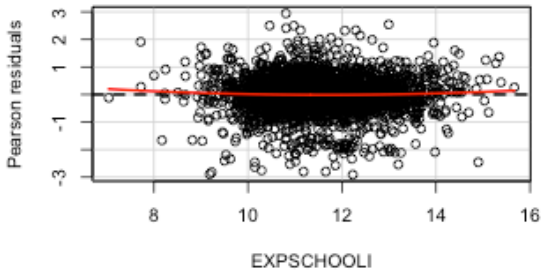
Model OUT1

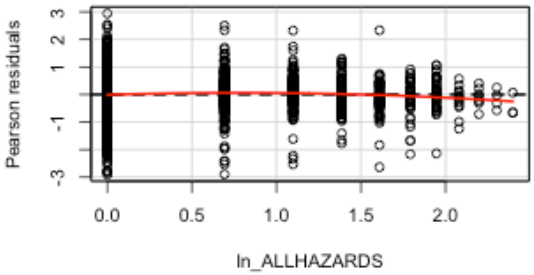

Pearson residuals
ln_INCOME
ln_ECONEST
EXPSCHOOLI
ln_FROST
ln_RAIN
Fitted values

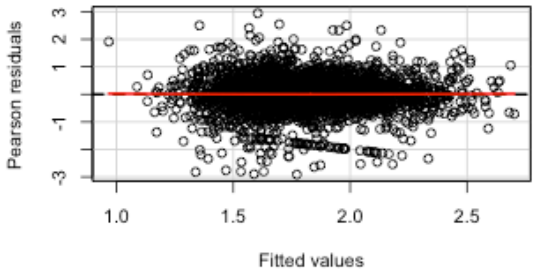
Model OUT2


Pearson residuals
-3 -1 1 2 3
8.0 8.5 9.0 9.5 10.0 10.5
ln_INCOME

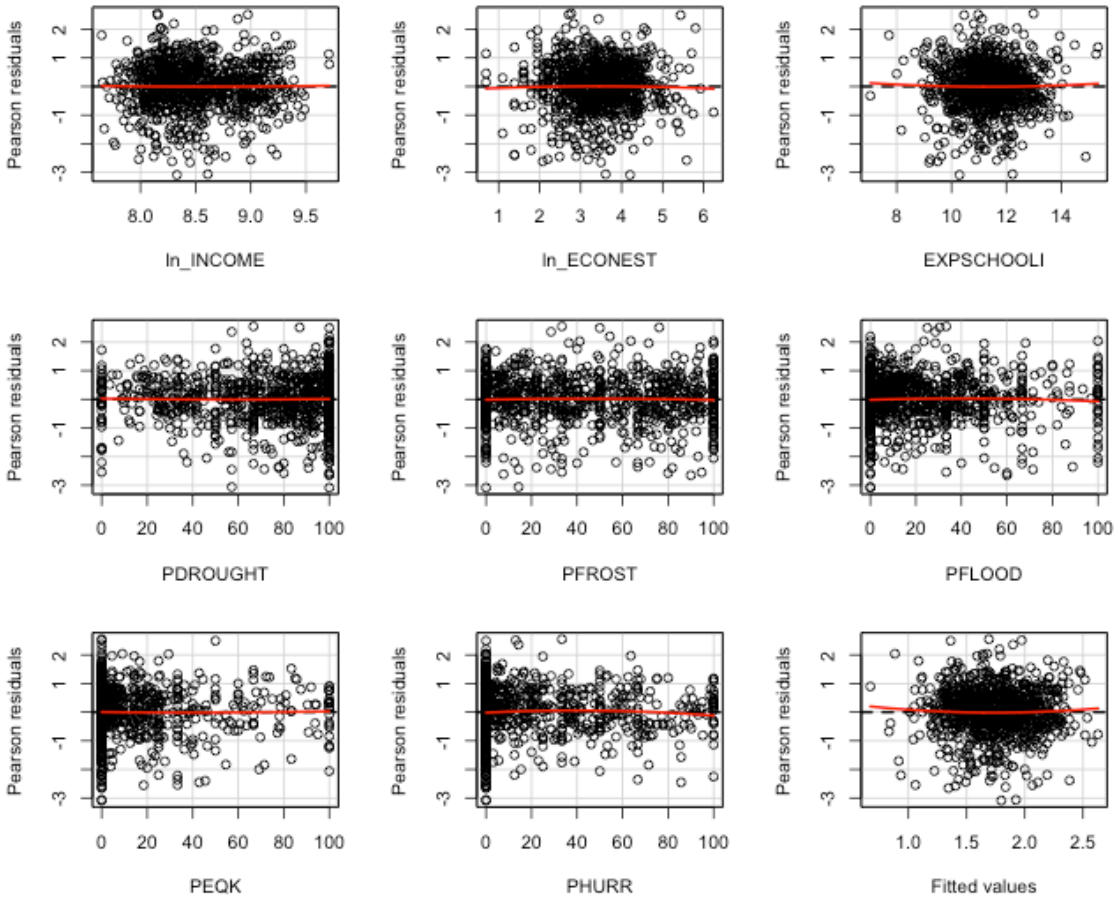

Pearson residuals
-3 -1 1 2 3
2 4 6 8 10
ln_ECONEST


Pearson residuals
-3 -1 1 2 3
8 10 12 14 16
EXPSCHOOLI


Pearson residuals
-3 -1 1 2 3
0.0 0.5 1.0 1.5 2.0
ln_ALLHAZARDS


Pearson residuals
-3 -1 1 2 3
1.0 1.5 2.0 2.5
Fitted values

Model OUT3


Pearson residuals
ln_INCOME
ln_ECONEST
EXPSCHOOLI
PDROUGHT
PFROST
PFLOOD
PEQK
PHURR
Fitted values