DOI: http://dx.doi.org /10.22185/24487147.2016.90.036

Modelling interstate migration in México: Static Markov Chains versus Dynamic Markov Chains with Moving Average

Carlos GARROCHO, Eduardo JIMÉNEZ-LÓPEZ
and José Antonio ÁLVAREZ-LOBATO

El Colegio Mexiquense, México

Resumen

Se compara la precisión / imprecisión de las Cadenas de Markov Estáticas (CME) *versus* las Cadenas de Markov Dinámicas con Medias Móviles (CMD) para reproducir los flujos migratorios interestatales de México. En México aún no se ha realizado una comparación detallada de ambos modelos. La principal aportación de este texto es la manera sistemática como se compara la precisión de los dos modelos markovianos. Se detallan las métricas de bondad de ajuste para comparar la exactitud / inexactitud de las CME y las CMD. Cada métrica se orienta a una *dimensión distinta* del sistema migratorio: i. La *Desigualdad migratoria* se analiza con índices de Gini y Curvas de Lorenz; ii. La *Competencia migratoria* con índices de Herfindahl-Hirschman; iii. La *Similitud Estadística* con Coeficientes de Correlación de Pearson y de Correlación de Rangos de Spearman; y, iv. El *Error Global* con el *error proporcional promedio.* El modelo derivado de las CMD resultó ligeramente más preciso que el modelo estático para replicar los flujos migratorios interestatales observados de México y seguramente será aún más exacto conforme se disponga de más información y se amplíe el periodo de estudio.

Palabras clave: Migración, México, Cadenas de Markov Estáticas *versus* Cadenas de Markov Dinámicas con Medias Móviles.

Abstract

Modelling interstate migration in México: Static Markov Chains versus Dynamic Markov Chains with Moving Average

This paper compares the accuracy / inaccuracy of Static Markov Chains (CME) *versus* Dynamic Markov Chains with Moving Averages (CMD) to reproduce interstate migration flows in Mexico. This comparison has not been done in Mexico. The main contribution of this paper is the systematic way the precision of the two Markov models is compared. We detailed the metrics of comparison. Each metric *is geared to a different dimension* of the migration system: i. *Inequality of immigration* is analyzed with Gini indexes and Lorenz curves; ii. *Competition of migration* is compared with the Herfindahl-Hirschman index; iii. *Statistical Similarity* is examined with Pearson correlation coefficients and Spearman rank correlation; and iv. The *Global Error* is estimated with the proportional average error index. The model derived from the CMD was slightly more accurate than the static model to replicate interstate migration flows observed in Mexico and will surely be even more accurate as more information becomes available and the study period is extended.

Key words: Migration, Mexico, Static Markov Chains *versus* Dynamic Markov Chains with Moving Averages.

Introducción

One of the most complex problems faced by contemporary geo-demography is to disclose and anticipate the population's spatial distribution (Garrocho, 2011). The reason is that out of the four basic components of demographic component: births, deceases, immigrations and emigrations, the two first have lost *relative* importance as explanatory factors and are more feasible to anticipate, while the latter have become relevant, volatile and difficult to anticipate (Goodkind and West, 2002).[1]

The topic is relevant, as acknowledging the distribution of land demand and over time is key for the decision making governments and various economic agents. Fails in information in this regard produce uncertainty and hinder the strategic and tactical planning in the public and private sectors, which eventually becomes governmental inefficacy and social and economic inefficiency. Even if the market produces instantaneous signals and reacts promptly, governments do not (Geys, 2006), and they can act erratically for long, locating and dimensioning in a wrong manner its investments in *spatial* terms (i.e., placing services in the wrong sites), *temporary* (i.e., opening or closing services at wrong times), *sectoral* (i.e., allotting resources to wrong or poorly interconnected sectors) or of the characteristics of the consumer market (i.e., estimating wrongly the dimensions and other peculiarities of the demand defined by market segments). Consequentially, it is fundamental both for the public and private sectors, not only to have a the *right picture* of the population's spatiotemporal distribution, but to *anticipate* its distribution patterns to make accurate, assertive and timely decisions (Harvir *et al.*, 2005).

To advance in the solution of the enigma of the population's spatiotemporal distribution, a key is to discover the migratory patterns that generate noticeable changes in the demand for services and development opportunities, both in space and time (Bell *et al*., 2015).[2] Revising the international literature, it is possible to establish four model categories that allow replicating migratory flows (a crucial element in the population's spatio-

1 An additional element: migratory movements directly affect the spatial and temporary patterns of births and deceases (Bustos, 2013).
2 Spatiotemporal demand of services and development opportunities is also influenced by the population's mobility. For a detailed discussion on migration, mobility and floating population, see Garrocho (2011).

temporal distribution): i) Markov models; ii) spatial interaction models; iii) econometric models; and, iv) models based on the master equation (Martínez-Faura and Gómez-García, 2001). This work focuses on Static Markov Chains v. Dynamic Markov Chains with moving average to reproduce interstate migratory flows in Mexico.

Markovian models are supported on Markov processes and are systemic (i.e., they simultaneously consider the origins and destinations of migratory flows) and predictive in nature, however they contribute very little in *explanatory* terms. They project the migratory balances of *all* the origins and destinations in a territory where they *simultaneously interact*. This implies that departures and arrivals compensate in each unit (e.g., federal states in this work,), which usually generates projections of migratory flows with acceptable accuracy (for Mexico see, for example, Bustos, 2013).

Objective and presentation strategy

The objective of this article is *methodological*: it consists in comparing the accuracy (or inaccuracy) of Static Markov Chains v. Dynamic Markov Chains with moving averages to reproduce interstate migratory flows in Mexico. It is not intended whatsoever to propose corrections to the Markovian methods used in this work, nor to propose a new method to model and project migratory flows. Our research question is much more modest: which Markovian method to model interstate migratory flows is more accurate (or inaccurate) for Mexico, Static Markov Chains or Dynamic Markov Chains with moving averages? Albeit simple, the question is relevant in methodological terms, as it allows orienting the analytical efforts of migration in Mexico to support on Markov chain's reasoning. Even if the literature reports that DMC model better the migratory systems than SMC (Guijarro and Hierro, 2005), there no systematic comparison of both models for Mexico yet. Perhaps the main methodologic contribution of this text is the systematic way to compare the Markovian models' accuracy.

The article structures into five sections. In the first, the estimation methods for Static Markov Chains (SMC) and Dynamic Markov Chains with moving averages (DMC) are explained; their main suppositions, advantages and weaknesses are included and it is suggested how some of their inconveniences can be overcome. In the second section, the measures of goodness of fit used in this work to compare the accuracy inaccuracy of SMC and DMC are explained. Using several measures of goodness of fit can lead to "the paradox of desynchronized watches": which one tells the right time? (see for example the classic articles by Plane and Mulligan, 1997, and Rogers and Sweeney, 1998). In order to avoid it, each measure we use is oriented toward a *different dimension* of the migratory system: i) migratory *inequality* is analyzed by means of Gini index and Lorenz curves; ii) *migratory competence* with Herfindahl-Hirschman indexes; iii) *statistical similarity* with Pearson Correlation Coefficient and Spearman's rank correlation coefficient; and iv) the *Global Error* with *average proportional error*.[3] In the third section the sources of information are disclosed (including a comment on their limitations) and the software utilized is mentioned. These three first sections are key due to the methodological nature of the article. In section four, we systematically compare SMC with DMC and compare the accuracy or inaccuracy with which they reproduce the Mexican migratory system. Finally, section five summarizes the findings, limitations and advantages of the utilized methods.

Migration and Markov's models

The Markovian method usually supposes the temporal homogeneity of the phenomenon it tries to model, so it estimates the transition possibilities (i.e., the *change* from one situation to another, or *move* from one place to another) with statistical techniques (Modica and Poggiolini, 2013). However, migration is a highly dynamical socio-spatial phenomenon that is influenced by a series of factors whose effect is not always well known and complicated to anticipate (Castles *et al*., 2013). Maybe because of these reasons, Markov chains, used as of the 1980's in Mexico (see the excellent pioneering contribution by Partida, 1989), have lost impulse in recent years as an instrument to *model and project migratory flows*, but not as instruments to project population (in this topic it is key to review the noteworthy contributions made by Bustos, 2013, for the Mexican case).

However, recently an improvement to SMC has been proposed (supported on the hypothesis of temporal homogeneity) by means of a dynamic

3 To prevent the "problem of desynchronized watches", in this text we neither estimate the Coefficient of Variation (CV) nor Hoover Index, which overlap Gini Index (GI) (Mulligan, 1991; Plane and Mulligan, 1997). Rogers and Sweeney (1998) engaged in polemics with Plane and Mulligan (1997) on which of the indicators was better to measure migration's spatial distribution. Time partially gave reason to Plane and Mulligan, nowadays GI is much more popular than CV. Additionally, López-Vega and Velarde (2011: 130) report: "Rogers and Sweeney (1998: 23) consider that CV is a much more transparent coefficient in terms of reading and interpretation. However, in the case of internal migration in Mexico, both CV and GI are sensible to the characteristics of the change in the recent migration flows…" (Text in brackets is ours). This suggests that the selection of indicators to explore the concentration /dispersion of migratory flows depends on the goals of each research.

method of moving averages, which arguably models the migratory systems better than the usual statistical method of linear tendency (Guijarro and Hierro, 2005). As far as we know, Markov chains have not been applied to migratory systems in Mexico or in Latin America yet, so their supposed advantages regarding SMC have not been systematically assessed in the region.[4]

The application of moving averages methods to Markov chains is revitalizing their use to model and project migratory systems (Guijarro and Hierro, 2005; Hierro and Guijarro, 2006), maybe because they are displaying their predictive power in various fields of knowledge: economy (Startz, 2008), finance and economic analyses (Cont and Wagalath, 2013; Kayacan *et al.*, 2010), genetics, biology or medicine (Lin *et al.*, 2012), to mention some examples.

Markov Chains

In this work by Static Markov Chains, we understand those that support on the hypothesis of temporary homogeneity. Their key element are the *probabilities of transition*: the *conditional probability to move* (i.e., *migrate*) from one space unit to another (i.e., states, municipalities, for instance) in *two points in time*.

The literature reports three main techniques to estimate Markov chains homogenous in time: the *static Markovian* procedure, *homogenous matrix*, and the *average matrix* method. The three procedures generate similar results, as they estimate a matrix with constant transition possibilities from the theoretical transition matrix that characterizes the chain. Then, the constant transition matrix calculated is used to project transition matrixes in consecutive points of time (Guijarro y Hierro, 2005).

As a hypothesis it can be advanced that static Markov chains (i.e., homogenous in time) may not be the best alternative to model migratory systems as dynamical as the Mexican, indeed owing to their *temporal homogeneity* (Romo *et al*., 2012). Then the alternative is to opt for a *dynamical* perspective that estimates the transition probabilities by means of a relation *linked to time*. Maybe, the most known is the linear tendency with a background in Rogerson (1979) and which has been applied with relative success in realities lose to the Mexican (Gómez-García *et al*., 1997; Hierro and Guijarro, 2007). If this is so, the hypothesis can be extended and state

4 In the international literature the term is moving average.

that it is very likely that the moving averages method is more accurate then the static Markovian to model the *migratory system* of Mexico.[5]

The advantage of the moving averages method apply to the modeling of migration is that it generates a softened estimation of the migratory flows by means of the repeated calculation of average values. Since the purpose is to *soften the tendency* of the migratory flows, controlling abrupt changes but maintaining the tendencies, the process is to calculate consecutively and iteratively the arithmetic averages of successive data over time. The number of successive data is called *order* of the arithmetic average: at the end the *most exact order according to the measures of goodness of fit* is selected and the values of the matrix of the flows calculated by the values of the arithmetic means that are estimated are replaced (*softened values*: Guijarro and Hierro 2005).

The selection of the order of the moving average (i.e., the dimension of the observation *time frame*) is not an easy aspect (Vandewalle et al., 1999). When the order is very low, the moving average is tightly linked to changes in the transition probabilities, which allows the method to adjust to face abrupt dynamic changes, but has the inconvenience of not being able to ring *false alarms* facing small dynamical changes, which might be adjusted later. In the opposite direction, a high order in the moving average is adequate when the dynamic of the chain (e.g., *the tendency*) is clear, as the corrections made by the moving average will be lower (Alessio *et al*., 2002). However, in this situation an earlier reaction of the method before an abrupt dynamic change is sacrificed (Chakraborti and Germano, 2010). Hence, the importance of the measures we use in this work to estimate the goodness of fit of the flows calculated in relation to those observed (see Introduction).

Static Markov chains: definition and calculation

Markov chains are one of the stochastic processes most useful to model phenomena from a probabilistic evolution in the future, only knowing the present state. A *stochastic process* registers a *non-determinist* behavior. This is to say, the trajectory or evolution of the process depends both on *inherent or causal* variables of the process and on *random or stochastic* variables. In mathematics, a stochastic process is a useful concept to disclose the *behavior* of random variables that evolve in function of time

5 Well now, if it is necessary to project population in Mexico, using Markov Chains, one of the best alternative (if not the best so far) is to support on Bustos' method (2013) that allows making projections by age, sex, region, explicitly incorporating the geographical dynamics of origins and destinations.

(however, on occasions it may be another variable: space, for example) (Ching *et al.*, 2013).

Migratory flows (and many other processes) observed over time are often modelled by stochastic processes, understood as any collection of random variables $\{X(t)\}$ dependent on time t (Cameron and Poot, 2011). Time can be discrete, for example $t = 0, 1, 2, \ldots$, or continuous: $t \geq 0$. At any moment, it describes the observation of a random variable we will call X_t or $X(t)$. Be $\{X_n\}_{n \geq 0}$ a discrete stochastic process with countable state spaces $E = \{i, j, k, \ldots\}$. If for all the integers $n \geq 0$ and all the states $i_0, i_1, \ldots, i_{n-1}$, i, j, then:

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \ldots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (1)$$

Where both sides of the equation are well-defined. Then, this stochastic process is called Markov chain or Markov process and it is called homogenous chain if the tight side of the chain (1) is independent from *n*. Equation (2) are Markov properties for all the *i,j* states (Yin and Zhang, 2010).

$$p_{ij} \geq 0, \sum_{j \in E} p_{ij} = 1 \quad (2)$$

The space of situations can be infinite, therefore the matrix is not generally studied by means of linear algebra. However, they meet the basic operations of addition and multiplication; this way, $p_i = (p_{ij})_{j Î E}$ is a probability vector for each *i*.

Dynamic Markov chains with moving averages: definition and calculation

While SMC try to model phenomena from the present situation, the method of DMC opens a "temporal observation frame" and considers *present and past* information. This is to say, it considers the *historic trajectory of the process*, which allows this method to obtain theoretical transition matrix for instants consecutive in time. For this article, for any temporal sequence of migratory movements, the estimation of the corresponding probability is the arithmetic average of the migratory flows between the different consecutive points in the observation period (Guijarro and Hierro, 2005).

When the method of moving averages is used, it is assumed that all the time-series observations are equally important to estimate the parameter to model and/or forecast (in this case the parameter is the arithmetic mean of the flows, according to the selected order *(n)*). This way, the average of the *n* values considered in the "temporal observation frame" is utilized to find the tendency or trajectory of the observed migration data.

In mathematical terms the moving average is calculated as follows:

The *moving* terms means that as observations appear in the time series (a *new temporal frame*), these replace the previous observations in the

$$Media\ Móvil = \frac{\sum(datos\ en\ "n"\ puntos\ en\ el\ tiempo)}{"n"} \quad (3)$$

equation and a new arithmetic average is calculated. Given any projection of the matrix of migratory flows with the help of equation (1), a procedure of moving averages of n order was carried out (n = 2, 3, …, k) for every, *i, j* Î E, as estimation of the probability of transition between instants *t* and *t+1*. This implies that multiple values of *n* are proved:

Using the technique of moving averages helps find and soften the tendency of the probabilities of the transition matrix, according to its trajec-

$$\overline{P_{ij}}(t, t+1) = \frac{1}{n}\sum_{u=1}^{n} P_{ij} \quad (4)$$

tory or historic behavior, which allows filtering nonsignificant variations in the migratory flows. This is an advantage on SMC which only consider the present.

Static Markov chains: advantage, limitations and solutions

It is important to clarify the main advantages and disadvantages of Markov chains in order to appreciate this work's results in its right dimension. The use of Markov chains to model migratory flows is supported on the following hypotheses, some of which become its main *disadvantages*:

i. The formulation of Markov chains implies that *population is homogenous*. This implies that in aggregated studies, in reality *heterogeneous* populations are modelled as if they were *homogenous*. By considering only one transition matrix and only one mobility rate, the population homogeneity considered by the Markov chains does not fully reflect the reality (this is common to al models: the only thing that faithfully reflects reality is reality itself). Usually Markov chains do not consider the breakdown of the population into groups distinguished by age, sex, income, among other relevant variables (Rabiner and Juang, 1986; Rabiner, 1989; Giménez *et al*., 2012). Maybe the best alternative to solve this for the case of Mexico is to follow Bustos' (2013) analysis strategy.

ii. The hypothesis of temporal homogeneity is conditioned by the guarantee that the chain is discrete, finite, and regular and that possesses a distribution in balance (Martínez and Gómez 2001). A strategy to avoid temporal homogeneity is to apply the method of moving averages (Le Gallo and Dall'erba, 2006).
iii. Changes of state (or situation) take place over a discrete time, this is to say, the transitions take place at regular time intervals. However, in some cases changes from one state to the other do not occur regularly, but they can be the result of other observations that have their own probability distributions. The method of moving averages helps reduce this problem as it considers tendencies over time.

Markov chains also have *advantages* to analyze migratory systems, for example some important are:

a) They have a systemic approach that allows considering all the origins, destinations and flows simultaneously, as well as the dependence between flows. Therefore, Markov chains are capable of offering a vision of the most probable changes in very complex systems such as the migratory, with a longitudinal vision (Guijarro and Hierro, 2007).

b) Their calculation is not extremely complex much less now there is varied software: Java Modelling Tools (JMT),[6] MARCA (MARkov Chain Analyzer),[7] or MARCH,[8] and even more conventional computing programs such as MatLab,[9] among many other.

c) The Markovian approach has a predictive capability without the need to acknowledge the *causes* of the phenomena it models (as it considers them stochastic models). This way, Markov chains are a valuable instrument capable of shedding light on short-term future migration (five or ten years).

6 https://sites.google.com/site/jaracohe/softwareparaestudio
7 http://www4.ncsu.edu/~billy/MARCA/marca.html
8 http://www.andreberchtold.com/march.html
9 http://www.mathworks.com/products/matlab/

MEASURES OF GOODNESS OF FIT

In spite of the frequent references in demographic literature to the spatial distribution of migratory flows (spatial focus: the degree of concentration of migration between spatial units of origin and destination: O'Kelly and Horner, 2003), there has been little advancement in this regard. Contemporary studies (e.g., López-Vega and Velarde, 2011) still resort to the same indicators reported by Plane and Mulligan (1997) and Rogers and Sweeney (1998) almost two decades ago (De Castro, 2007).[10] Because of that here we also introduce Herfindahl-Hirschman Index, which has demonstrated its power in economics literature on market competence (in the case of this work: *migratory competence* between federal states) and which complements well GI (D'Amico and Di Biase, 2010). We also utilize Lorenz Curve (LC), as GI can generate the same value for different distributions (e.g., for different migratory landscapes: Giorgi, 1993). Following, the indicators of goodness of fit used in this work and the reasoning underlying its use is detailed.

Gini index and Lorenz curve

In this text Gini Index (I_G) is a measure of the inequality of distribution of the flows of immigrant and emigrants between federal states and it is estimated as follows:

$$I_G = \frac{\sum_{i=1}^{N}(p_i - q_i)}{\sum_{i=1}^{N} p_i} \qquad (5)$$

Where p_i and q_i represent the occurrence frequencies of two variables in a set of data (1, 2, …, i, …, N). Gini index cannot be negative and ranges between 0 and 1. If $p_i = q_i$, it means that the analyzed variable distributes equally over the elements (it implies *maximum equality* between the elements: $I_G = 0$). For its part, $I_G = 1$ represents maximum *inequality between* the elements $p_i \neq q_i$ (Raskall and Matehson, 1992; Spicker *et al*., 2007).[11]

10 In reality some of these indicators date back to the 1980's (Watkins, 1986), as reported by Plane and Mulligan (1997).

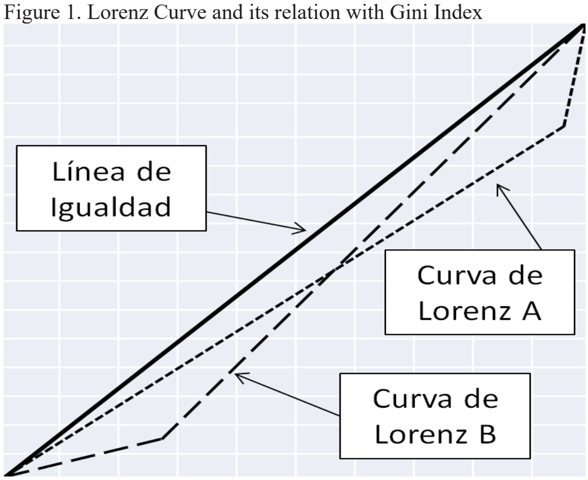
11 In the excellent work by López-Vega and Velarde (2011; p. 130) there is a mistake in their explanation when they say: "… the closer to 0 the value [of GI] is means there is a greater heavier oncentration of the flows at certain points, and the opposite case occurs when the value of closer to 1". The interpretation of GI is precisely contrary. The text in quotation marks is ours.

In addition to the traditional Gini Index, in this work we follow the example of López-Vega and Velarde (2011) and estimate the *Global* Gini index with Plane and Mulligan's (1997) methodology, with together with the average proportional error are the global indicators used in this work (see further in the text the explanation of average proportional error).[12]

Gini Index is directly linked to Lorenz curve, which is a graphic Cartesian representation of inequality. In this work LC is sketched placing on *X* axis the accumulated values of the *observed migrants* expressed as percentage regarding the total migrants and on *Y* axis the accumulated values of the *calculated migrants* also expressed as percentages regarding the total. Of both series of accumulated values are exactly the same, there is no LC but a diagonal line from the bottom left corner of the Cartesian plane to the top right corner. This so called Line of Equality and corresponds to perfectly equal distributions.

LC is a *necessary* complement of GI, as the index exhibits an important weakness: it can generate the same value for different situations of inequality (Giorgi, 1993). Since Gini index measures the *proportion* of the area under the Line of Equality limited by CL regarding the total area of the triangle made by the Cartesian axes and the Line of Equality, it can produce one same value for an obese curve in the top right part of the curve and for an obese curve in the bottom left part, which are utterly different scenarios of inequality (figure 1).

Figure 1. Lorenz Curve and its relation with Gini Index



Source: own elaboration.

12 The explanation of Global Gini Index is thoroughly described in López-Vega and Velarde (2011) and in Plane and Mulligan (1997). In the latter, there is even an example of the calculations "step by step".

In figure 1, LC “A” represents a highly *uneven* distribution of migrants between the federal states that record the *most* migrants. On it own, LC “B” also represents a high *uneven* distribution of migrants between the federal states, but in this case regarding the states with the *fewest* migrants. However, since the areas limited by LC are equal (CL “A” is the mirror of LC “B”), the value of GI is the same. This is to say: Gini Index *does not detect difference between both situations*. It is noticeable that LC is used much less than GI in migratory studies, as they are mutually necessary complements.

The *shape* CL is a good visual indicator of inequality. Albeit, visual inspection involves a number of inequalities (Metzger, 2006), hence some quantitative indicators that *summarize its profile* are needed. In this work use three:

Average: where the LC slope is equal to the Line of Inequality (p = 1.0). This is the point “A” in figure 2, which has as *F(ų)* as an abscise. The federal states to the right are *dominant* in terms of migration: they concentrate migratory movements (emigration or immigration) more than proportionally. The breakpoint *F(ų)* is important because it indicates us how convex LC is. The more convex the curve, the higher inequality.

Median: it indicates how many federal states concentrate 50 percent of the migratory movements. In figure 2, it is observed that such point (“B”) is located in the LC after abscise 0.50 and the angle *((Q0.5)/ ų)*.

Schutz index (also known as Pietra Index). It estimates the maximum deviation between the LC and the Line of Equality. The value of Schutz index is calculated as follows: (figure 2):

$F(\mu) - L(F(\mu))$.

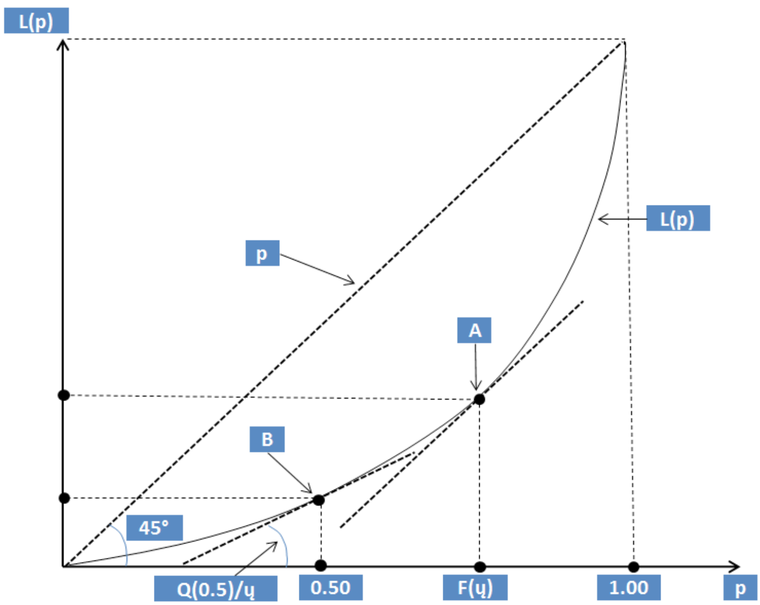
These three indicators adequately summarize the *shape* of LC, especially when differences between the curves are difficult to distinguish visually (Lubrano, 2015) (figure 2).

Herfindahl-Hirschman index

Owing to the weaknesses of GI, in this work we also use Herfindahl-Hirschman Index (HHI), which is a second-order concentration measure more powerful than GI (Masciandaro and Quintyn, 2009), albeit it is recommended using both indicators complementarily (D’Amico and Di Biase, 2010).

HHI offers and image of the *competence in the migratory market* from detecting the *dominant* spatial units to a lesser or greater extent: for example: federal states that might dominate the attraction or ejection of migrants. As GI, HHI uses the proportion of flows that correspond to each federal state in terms of migrants, but *squared*. This way, the weight of the most important states in terms of migrants is underscores and the problem of producing values of HHI for different situations as it occurs in the case of GI (Giorgi, 1993).

Figure 2. Key points of the Lorenz Curve



Source: Duclos and Araar, 2006.

HHI is calculated adding the squares of the participations of each federal state in the migratory market (immigration, emigration and/or total). For example, if all immigration reached a single federal state, it would be said that such state concentrates 100 percent of the *migratory market* and there would be competence. HHI would be: 100 x 100 = ten thousand, which is the maximum value of HHI. The extreme opposite situation (also hypothetical) would be the inexistence of migration and then all the federal states had a participation in migratory market of 0.0. In this situation, HHI would be 0.0 (the addition of all the squares of zero). Therefore, the closer to 10 thousand the value of HII, *the lower the migratory competence*, and the closer to zero, the *higher the migratory competence* (values close to zero would indicate, in terms borrowed from economics, an *almost perfect migratory competence*).

HHI is calculated as follows:

$$IHH = \sum_{i=1}^{N} s_i^2 \tag{6}$$

Where

$$s_i = \frac{x_i}{\sum_{i=i}^{N} x_i}$$

Is the addition of the fraction of the squared migratory total by *federal state* of the data that compose it.

Additionally, the literature reports HHI *thresholds* to interpret competence: values under 1 500 indicate *high competence*; values between 1 500 and 2 500 are interpreted as *moderate competence*; and of 1800 and over, they are interpreted as *lack of competence* (presence of agents highly dominant or monopolistic) (Sys, 2009, among others).[13]

Person's and Spearman's correlation coefficients

Pearson's or Spearman's correlation measures are statistical indexes that measure the linear relation between two variables. Here, they are utilized as measures of *similitude of statistical behavior*. Pearson's correlation is estimated from the comparison of two data distributions (equation 7). For its part, Spearman's correlation coefficient is calculated from the comparison of the ranges of paired data (equation 8). Spearman's correlation can be calculated with Pearson's formula if data are previously grouped in ranges. Both coefficients vary from -1 to +1 and are interpreted in the same manner: equal or close to zero means that the distributions of the observed and calculated migratory flows are not similar; if they are close or equal to +1, it means that the variables' statistical behavior is very similar; and, if they are close or equal to -1, it means that the observations behave in perfectly opposite manner: they are absolutely *dissimilar*.

Pearson's (r) and Spearman's (r_s) coefficients are estimated as follows:

13 This does not occur with GI: When is value low, mid or high? The interpretation of GI is made in comparative terms for each study case.

$$r = \frac{S_{xy}}{S_x S_y} \tag{7}$$

Where S_{xy} is the covariance of the flow observed regarding the estimated flow, S_x and S_y are the standard deviations of each flow.

Where *d* is the difference between ranges (i.e., the observed flow minus the estimated flow); *k* is the number of data in the range.

$$r_s = 1 - \frac{6\sum d^2}{k(k^2-1)} \tag{8}$$

Average proportional error

The average proportional error allows estimating the global difference between the *observed* and *calculated* matrices, comparing all the migratory flow. Therefore, it allows two basic things in this text: i) to identify the order (*n*) of the method of moving averages that best adjusts to the observed flows. This is to say, the order (*n*) that minimizes the average proportional error between the migratory system calculated with the model and the *observed* (or *real*) *migratory* system; and, ii) to estimate the deviation or the average proportional fail of the matrices calculated by means of SMC or DMC regarding the matrix of the observed flows.

The average proportional error (*e*) is estimated as follows:

$$e = \frac{\sum |F_{ijo} - F_{ijc}|}{(2\sum F_{ijc})} \tag{9}$$

Where

e = average proportional error of the calculated migratory flows (proportion of migrants *wrongly calculated* by the model):

F_{ijo} = percentage structure observed in the migratory flows.

F_{ijc} = percentage structure calculated for the migratory flows.

It is worth mentioning that (*e*) estimates the average error of calculations regarding the reference parameter, which in this work is the matrix of flows observed in 2010 and expresses them as proportions. If they are multiplied by 100 they can be interpreted as *percentages of error*.[14] The minimum value of (*e*) is 0.0, which implies there is no difference between the

14 For example, (e) = 0.20 means that the model has registers an average proportional error of 20 percent regarding the observed data.

flows calculated the matrix of *observed* flows for the same year: the model perfectly replicates the observations. However, this situation is difficult to find, the most usual is that there is a certain average proportional error (*e*). The higher the value of (*e*), the greater the difference between calculations and observations: the model is less accurate.

The valuing of the importance of (*e*) shall be contrasted against the nature of each situation that is being modelled and with the number of data the model projects. If there are few data, a value of (*e*) = 0.10 can be very high, if it is a delicate process (related with some health issue, for example), it may be unacceptable (Garrocho *et al*. 2002; Webber, 1984).[15] In the case of this study on migration, a value of (*e*) = 0.10 would be very good, as it means that the model only misses its projection in ten percent on average, even if it is projecting 992 data (the flows between the 32 states, eliminating the diagonal of the matrix). We also have to bear in mind that (*e*) is an average *global* indicator (as the global Gini index mentioned earlier) and so, there would be cases (e.g., flows between federal states) with errors or smaller and greater deviations than the value of (*e*).

Information sources and software utilized

In this work we had a temporal sequence of transition matrices of the Mexican migratory system at the scale of federal states. The matrixes were built with support from the National Council of Population (Conapo) using data on recent migration from the 1990, 2000 and 2010 Censuses of Population and Housing and the 1995 and 2005 Population Counting.[16]

Recent migration only accounts for the usual place of residence of the population five years before being interviewed to build the information sources.[17] In the group from 0 to four years the place of birth is considered as the migratory origin. This approximation to the recording of recent migration implies that possible changes in the place of residence within five years are not accounted, which is a limitation of the sources of information on Mexican migration, as there are no continual records of changes of residence (López-Vega and Velarde, 2011).

15 The assessment of how acceptable the value of (e) is does not respond to pre-established rules, but to the characteristics of the study undertaken. Something similar occurs in other statistical indicators such as Gini index or Pearson's correlation coefficient (R): when is a coefficient high, mid or low? (Field, 2005).
16 As López-Vega and Velarde do in their splendid work of 2011.
17 See a discussion on the diffuse of the term habitual residence in Garrocho, 2011.

Additionally, the sources of information considered residence change in Mexico are the ones that imply crossing municipal or state borders. Therefore, in intra-state migration there is not a change of residence, even though it involves a deep change for the population (distance, cultural change); as in the case of a family that moves from the city of San Luis Potosí to the Huasteca Potosina. However, a change in residence that involves moving one block (which theoretically in Mexico is 50 meters), but which implied a change in federal state is recorded as interstate migration (for instance, a family that moves inside the same neighborhood, such as Lomas de Sotelo, which has a part in Mexico City and the other in the State of Mexico [they are adjoining states]). Indeed, more detailed and accurate data on migration are needed in our country (Garrocho, 2011).

These limitations are inherent to Mexico's sources of information. However, even so it is possible to reasonably summarize the data on the interstate migratory flows in origin-destination matrices (ODM) of 32 rows and 32 columns. ODM records the population interchange between spatial units (e.g., states, municipalities). It is read from right to left and from top to bottom. In each cell of the matrix the outflows are recorded (if read by row) or inflows (if read by column) of interstate migrants. The addition of each row is the total population that left *from* each state. The addition of each column is the total population that arrived *in* each state. The diagonal of the matrix represents the population that did not migrate. The cell of the matrix in the crossing of column and row of additions (e.g., that one that simultaneously registers the total by column and row) is the total population that migrated in the period considered. The interstate ODM for Mexico is composed of 1 024 cells (32 x 32 = 1 024). If the diagonal is subtracted (32 cells: the population that did not migrate) the result is 992 interstate migratory flows. With this information Static and Dynamic Markov Chains were calculated in this work. The calculation were mainly carried out in MatLab.

Static Markov Chains *versus* dynamic Markov Chains with Moving Averages

Comparison strategy for Static Markov Chains *versus* dynamic Markov Chains with Moving Averages

The strategy followed to compare the accuracy of estimations using SMC and DMC was:

a) The matrix of observed flows in 2010 (the most recent available) was the parameter that was used to assess the accuracy/inaccuracy of estimations of DMC and SMC. Especially the row and column of the matrix totals.

b) The *measures of goodness of fit* between the matrix of observed flows (reference parameter) and the values calculated by the static and dynamic Markov chains allowed estimating the four dimensions of the results: inequality, migratory competence statistical similitude and global error. In this work, it is considered that the first three dimensions are *equally important* (a similar strategy to that of López-Vega and Velarde, 2011) and that the average proportional error is one of the *most powerful* indicators of goodness of fit due to its global character (as considered from Webber, 1984, to Boutros *et al*., 2015; Chormanski *et al*., 2008 or Singh, 2001, among many more).

c) SMC for 2010 were estimated considering the migration matrixes for 2000 and 2005.

d) DMC were calculated considering the following time frames: *i.* Matrices for 1990 and 2000 ($n = 2$); *ii.* Matrices for 2000 and 2005 ($n = 2$); *iii.* Matrices for 1995, 2000, 2005 ($n = 3$); *iv.* Matrices for 1990, 1995, 2000 ($n = 3$); and, *v.* Matrices for 1990, 1995, 2000, 20005 ($n = 4$).

Goodness of fit in the four dimensions considered: inequality, migratory competence, statistical similitude, and average proportional error

Dimension 1: Inequality

Emigration. In 2000, the GI for total observed migration by federal state had a value of barely 0.10, which means low interstate inequality in terms of population ejection. This situation remains virtually unchanged in 2010: GI is 0.11 (table 1). SMC replicate well the inequality of emigration observed in 2000 (with a GI of 0.08) and they also show a scenario with no important changes in inequality for 2010 (as it is the case of the observed situation). Conversely, DMC generate scenarios far from the observed: GI for 2000 is 0.02 and for 2010 is 0.01, much less uneven than those observed and estimated SMC. Although SMC and DMC of low emigratory inequality and stable over the period (which is the same behavior as the observed flows), SMC were close to the observed situation. Point for SMC.

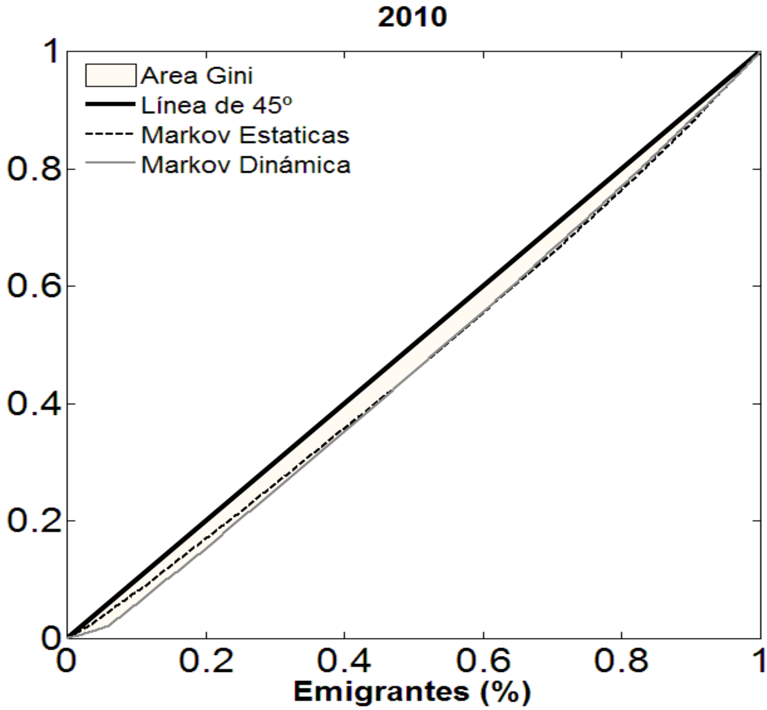
For their part, the indicators of the shape of Lorenz Curves (LC) show that the curves generated by both Markovian models are very similar that that which comes from the observed values, which is consistent with the visual analysis that does not allow detecting at first sight the difference between them (figure 3; see columns *Differences* in table 1). The Euclidian distance between the point of *Mean emigration* of the observed LC and that produced by SMC for 2000 is barely 0.02 and 0.04 for DMC; differences in the *Median* of LC of SMC and DMC regarding the *Median* of the observed LC are 0.0 and 0.01, respectively; and the difference between the Schutz indexes are also marginal: 0.09 and 0.05.

Also in 2010, differences in the shape of the observed LC and the LC calculated were low: the *Mean* of the LC derived from the calculations of SMC (curve we will shorten to SLC) deviated three hundredths from the mean of the observed LC and the mean of the LC profiled with calculations from DMC (shortened to DLC) failed only by fur hundredths. The *Median* of SLC deviated nine hundredths from the location of the Median of the observed LC and DLC only erred in two hundredths (table 2).

There is also much similitude in the shape of LC when the points of maximum deviation between LC and the Line of Equality are analyzed; this is estimated by means of Schutz Index. The difference between the observed LC and SLC is only two hundredths and seven hundredths regarding DLC (table 2). It is noticeable the great capacity of Markovian models to replicate the shape of LC profiled with the observed data. For practical purposes, the calculated and observed CL are the same. Therefore, in terms of the shape of LC we call it a *draw*.

Immigration. In 2000, the inequality of the total observed immigration is low: it records a GI of 0.10, which slightly descends by 2010, when GI reduces to 0.06, which indicates a reduction in the immigration inequality between the federal states (table 1). SMC replicate well the situation observed in 2000 as they accounted for a GI of the state immigration totals of 0.06. For 2010, their accuracy increases: they keep their GI in 0.06 with which the inequality they replicate in the migratory system is the same as the observed.

Figure 3. Impossibility to visually compare Lorenz Curves when they are very similar



Source: own elaboration.

Conversely, DMC fail in 2000 with a GI of 0.01, lower than the observed and the estimated by SMC, and in spite that in 2010 the noticeably approach the observed situation with a GI of 0.04 they do not accomplish the perfect exactitude of SMC. Another point for SMC.

As in the case of emigration, the shape of the observed immigration LC is replicated well with data calculated by the Markov models. In 2000, the point of *mean* immigration (that allows identifying the states that concentrate more than proportionally the immigration movements) of SLC deviates only eight hundredths from the Mean of the observed LC (let us remember that the LC scale ranges from 0.0 to 1.0) and the Mean of DLC only misses two hundredths (almost perfectly matching the Mean of the observed LC). The differences in the localization of the *Median* (that indicates how many federal states concentrate 50 percent of the migratory movements) of SLC only missed by a hundredth regarding the Median of the observed LC and the DLC only deviated three hundredths.

For its part, the point of maximum deviation between LC and the Line of Equality (calcualetd by means of Schutz index) also records high coincidences between the observed LC and those calculated: in the case of SLC, it perfectly fits (deviation is 0.00), while DLC fails in three hundredths regarding the observed LC. In 2010, the adjustments are also high. The *Mean* of SLC fails in one hundredth the observed LC and DLC deviates four hundredths. The *Median* and the point of maximum deviation between LC and the Line of Equality of LC show a very similar situation: deviation of one and six hundredths for the case of the Mean and eight for a perfect fit (that of DLC) for the case of Schutz index. Both Markovian models replicate as well the shape of the observed LC: a draw.

Global migratory inequality. The global inequality observed in the migratory system in 2000 (estimated with the method of Plane and Mulligan, 1997) can be considered high as the GI is 0.24 and remains the same in 2010 (table 1). SMC reflect well this great inequality as its GI for 2000 is 0.22 and of 0.24 for 2010, which almost exactly replicates the observed situation. DMC show a performance very similar to that of SMC: GI = 0.21 in 2000 and 0.23 in 2010, so we can declare a draw in this concept. Migratory inequality with the method of Plane and Mulligan (1997), does not allow deriving Lorenz Curves.

Dimension 2: migratory competence

Emigration. By and large, Markovian models replicate relatively well the interstate emigration competence observed for 2000 and 2010 (let us remember that HHI scale ranges from 0 to 10,000) (table 1). The HHI observed in 2000 is 492.8, that of SMC is 714.7 and for DMC of 774.8. For 2010, these values remain stable (455.2, 413.0 and 442.7, respectively). Values close to zero indicate that the observed and calculated emigration competence is high, which implies there are no dominant federal states as population ejectors. This concurs with the evidence supplied by GI. As regards, emigration measured with HHI, we can declare a draw between SMC and DMC: the differences between their HHI regarding the observed migration are marginal, since the measuring scale goes from 0.0 to 10 thousand.

Immigration. The situation in this concept is very similar to that of emigration: the Markov models reproduce with accuracy the immigration competence observed at interstate scale both for 2000 and 2010 (table 1). The values of HHI in 2000 range from 455.2 to 754.7 and on 2010 the range

Table 1. Goodness of fit for the observed and calculated migratory interstate flows by emigration, immigration and global, 2000-2010

	2000			2010		
Indicators of goodness of fit	Observed	Estimated		Observed	Estimated	
		Static MC	Dynamic MC		Static MC	Dynamic MC
Gini index						
Emigration	0.10	0.08	0.02	0.11	0.07	0.01
Immigration	0.10	0.06	0.01	0.06	0.06	0.04
Global	0.24	0.22	0.21	0.24	0.24	0.23
Herfindhal-Hirschman index						
Emigration	492.8	714.7	774.8	455.2	413.0	442.7
Immigration	455.2	621.7	754.7	485.2	779.8	639.9
Global	461.2	741.1	787.2	714.7	731.2	954.7
Pearson correlation coefficient						
Emigration	N.A.	0.35	0.37	N.A.	0.36	0.39
Immigration	N.A.	0.45	0.33	N.A.	0.35	0.35
Global	N.A.	0.46	0.41	N.A.	0.37	0.45
Spearman correlation coefficient						
Emigration	N.A.	0.25	0.40	N.A.	0.30	0.41
Immigration	N.A.	0.20	0.45	N.A.	0.28	0.38
Average proportional error	N.A.	0.06	0.04	N.A.	0.04	0.02

Source: own elaboration.

Table 2. Indicators of the shape of Lorenz Curves, 2000-2010

2000 — Emigration

Indicators	Static MC				Dynamic MC				Observed				Difference	
	Coordinates		Slope	Angle (º)	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Observed Static Dis-tance	Observed Dynamic distance
	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ		
Mean emigration	0.43	0.39	1.00	45.00	0.38	0.37	1.00	45.10	0.41	0.39	1.00	45.0	0.02	0.04
50% concentration point	0.50	0.46	2.61	69.0	0.50	0.45	2.80	70.30	0.50	0.46	0.90	42.4	0.00	0.01
Point of maximum inequality	0.46	0.45	3.06	71.90	0.50	0.45	2.28	70.30	0.55	0.47	2.61	69.1	0.09	0.05

2000 — Immigration

	Static MC				Dynamic MC				Observed				Difference	
	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Observed static dis-tance	Observed Dynamic distance
	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ		
Mean emigration	0.13	0.11	1.01	45.20	0.09	0.07	1.01	45.30	0.07	0.06	1.00	45.00	0.08	0.02
50% concentration point	0.50	0.47	0.69	34.80	0.50	0.49	1.32	52.90	0.50	0.46	2.41	67.50	0.01	0.03
Point of maximum inequality	0.36	0.34	1.61	58.10	0.34	0.32	1.42	55.00	0.36	0.34	2.74	69.90	0.00	0.03

2010 — Emigration

Indicators	Static MC				Dynamic MC				Observed				difference	
	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Observed static distance	Observed Dynamics distance
	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ		
Mean emigration	0.13	0.11	0.99	44.90	0.07	0.08	1.00	45.10	0.11	0.09	1.00	45.00	0.03	0.04
50% concentration point	0.50	0.40	0.58	30.20	0.50	0.47	1.25	60.30	0.50	0.49	0.66	33.63	0.09	0.02
Point of maximum inequality	0.41	0.38	2.29	66.40	0.41	0.47	2.14	64.90	0.41	0.40	2.21	65.66	0.02	0.07

2010 — Immigration

	Static MC				Dynamic MC				Observed				difference	
	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Coordinates		Slope	Angle(º)	Observed static distance	Observed dynamic distance
	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ	x	y	L(p)	Q(x)/μ		
Mean emigration	0.36	0.35	1.00	45.00	0.39	0.38	1.00	45.00	0.37	0.35	1.00	45.00	0.01	0.04
50% concentration point	0.51	0.49	0.62	32.00	0.50	0.43	1.11	48.10	0.50	0.49	2.95	71.31	0.01	0.06
Point of maximum inequality	0.28	0.27	3.03	71.70	0.24	0.20	5.09	78.90	0.24	0.20	5.09	78.90	0.08	0.00

Source: own elaboration.

Table 2. Indicators of the shape of Lorenz Curves, 2000-2010 (continuation)

Indicators	Coordinates x	Coordinates y	Distance Slope L(p)	Angle(º) Q(x)/μ	Difference Observed Static Distance	Difference Observed Dynamic Distance
2000						
Emigration						
Mean emigration	0.41	0.39	1.00	45.0	0.02	0.04
50% concentration point	0.50	0.46	0.90	42.4	0.00	0.01
Point of maximum inequality	0.55	0.47	2.61	69.1	0.09	0.05
Immigration						
Mean emigration	0.07	0.06	1.00	45.00	0.08	0.02
50% concentration point	0.50	0.46	2.41	67.50	0.01	0.03
Point of maximum inequality	0.36	0.34	2.74	69.90	0.00	0.03
2010						
Emigration						
Mean emigration	0.11	0.09	1.00	45.00	0.03	0.04
50% concentration point	0.50	0.49	0.66	33.63	0.09	0.02
Point of maximum inequality	0.41	0.40	2.21	65.66	0.02	0.07
Immigration						
Mean emigration	0.37	0.35	1.00	45.00	0.01	0.04
50% concentration point	0.50	0.49	2.95	71.31	0.01	0.06
Point of maximum inequality	0.24	0.20	5.09	78.90	0.08	0.00

Source: own elaboration.

is from 485.2 to 779.8. This is to say, immigration competence remained stable. The values close to zero indicate there are no dominant regarding immigration, but immigration competence is high. The stable behavior of immigration competence in the studied years is also consistent with GI. Since differences in HHI of SMC and DMC regarding the observed migration are marginal, we can declare a draw between the Markovian models.

Global migratory competence. As expected (if the previous results were revised), global migratory competence is very high, both in the observations and in the Markov models' results: HHI vary in narrow ranges from 461.2 to 787.2 in 2000 and from 714.7 to 954.7 in 2010 (table 1). HHI of SMC are closer to the value of HHI for the observed situation than those of DMC, but these differences are so small (in a scale from 0.0 to 10 thousand) that we can state that for practical purposes the accuracy of both models is the same and it can be called a draw. In this concept, results are not directly comparable with those of GI, which showed growing global inequality, as global inequality was estimated with the method of Plane and Mulligan (1997).

Dimension 3: similitude of statistical behavior

Emigration. Pearson's correlation coefficients between the calculated and observed values are ata all cases positive and very similar in magnitude, both for 2000 and 2010. In the first year, they record 0.35 (SMC) and 0.37 (DMC) and in the second 0.36 (SMC) and 0.39 (table 1). If we consider Field's (2005) thresholds, it can be said that the two Markovian models reproduce with intermediate accuracy the statistical behavior of the observed interstate emigration.[18] In both years, DMC record correlation indexes slightly higher than those of SMC, especially in 2010. Owing to the sample size (32 state totals), differences are not marginal. This is verified if we revise the Spearman's rank correlation coefficient: DMC generate higher coefficients than SMC. In 2000 and 2010, Spearman correlation coefficients between the observed totals and those of DMC were 0.40 and 0.41, while those of SMC reached .025 in 2000 and 0.30 in 2010. Point for DMC.

Immigration. In 2000, SMC record a Pearson correlation coefficient with the observed pattern much higher than DMC: 0.45 versus 0.33, however, in 2010 both models produce the same coefficient: 0.35 (table 1). Albeit, it is interesting to that Spearman coefficient is much higher for DMC: in 2000 it is 0.45, while for SMC it is 0.20; in 2010 the values are 0.38 and 0.28, respectively. This is to say, the tendency of individual values is better expressed by SMC, but the importance (i.e., *hierarchy*) of federal states as population attractors is better expressed by DMC. In this situation, the

18 Field (2005: 112) makes the same question: how much is the correlation between two variables? And after a broad review of the literature he proposes three *guide thresholds* (all with a positive and a negative sign): a coefficient around 0.1 is *low*; around 0.30, *mid*; and, 0.50, *high*.

purpose of each study will define which Markovian model is the most adequate; this way, this concept is a draw.

Global statistical similitude. The first to underscore is that Pearson correlation coefficients show intermediate and high values, which indicates that both models replicate well the observed values (table 1). In 2000, the coefficients of the two models are similar (SMC= 0.46 and DMC = 0.41); however, in 2010 DMC clearly surpass the SMC: the values are 0.45 and 0.37, respectively. In terms of statistical similitude it seems as if while there is more information on the tendency (i.e., trajectory) better the performance of DMC. Point for DMC. For the case of global statistical similitude, it is not possible to calculate Spearman correlation coefficient.

Dimension 4: Average proportional error

The two Markovian models show a noticeable capacity to model interstate migratory flows: considering the 992 cells of the migratory flow matrix record a average proportional error (*e*) of maximum 6 percent and minimum of two percent (table 1). This is to say, the models are correct in their replications of the observed pattern in 94 percent on average, as a minimum, but reach an average accuracy of 98 percent. These surprising results were verified a number of times: they are right. For both years of the study DMC record lower (*e*) and despite the difference is not noticeable, we give the point to DMC owing to their better accuracy in this *key indicator* of global goodness of fit.

Conclusions and contributions

In this work a systematically *multidimensional* comparison of the static Markov models and dynamic with moving averages in view of assessing their accuracy/inaccuracy to replicate the interstate migratory flows observed in Mexico in 2000 and 2010. Four dimensions of the results were assessed: i) interstate *inequality* in the distribution of the migratory flows; ii) *migratory competence* between federal states; iii) *statistical similitude* of the behavior of flows between states; and, iv) *average proportional error* of the estimated flows regarding those observed.

This strategy of comparative analysis has at least three methodological particularities worth underscoring. *First*, it analyzes different dimensions of the results to accomplish two objectives: assess in the most complete manner the accuracy/inaccuracy of Markovian models and avoid the "paradox of two desynchronized watches" (which tells the right time?). This is to say, in this work the same dimension of the migratory flows is not measured differently, but different aspects of the same migratory network are analyzed. *Second*, migratory inequality was analyzed not only by means of GI as it is commonly carried out in studies on the spatial distribution of migratory flows (spatial focus), but we resorted to a deep study of the shape of Lorenz Curves using key indicators. Since Gini index can generate the same value for utterly different inequality scenarios, the use of Lorenz curves is fundamental, even though (surprisingly) seldom it is made on analyses of migratory flows. *Third*, the analysis of migratory competence was supported on Herfindahl-Hirschman Indexes, which are widely used in economics, but which inexplicably, are scantly applied in studies on the degree of concentration of migration between spatial units of origin and destination.

In analytical terms, it is noticeable the great capacity of Markovian models to not only replicate the network of migratory flows in their different dimensions, but even in the shape of Lorenz curves (in the case of interstate migratory inequality). More so, the results of the static and dynamic models always point in the same direction, they avoid the "paradox of the desynchronized watches" and are consistent in all the dimensions of their results. The only variation is the degree of accuracy (always high) in any of the dimensions. A solid datum: the global indicator of goodness of fit between calculations and observations shows that the models are right in their replications of the observed migratory network in 94 percent on average at least (when the statistic model) and can reach an average accuracy of 98 percent d the dynamic model is applied.

In the end, the dynamic Markov model with moving averages was slightly more accurate than the statistic model to replicate the interstate migratory movements observed in Mexico (table 3). Three arguments justify this conclusion: i) greater global accuracy (smaller average proportional error); ii) noticeably superior accuracy to replicate the statistical behavior of the migratory flows between states; and, iii) their capacity to integrate more information into the modelling of migratory flows (incorporates present and past: tendencies or trajectories), which implies that the broader the temporality of analysis is more information is incorporated into the migratory tendency and the greater superiority of the dynamic model over the static. This is not a theoretical speculation: over this work we noticed how the dynamic model noticeably improved its accuracy in 2010 regarding 2000.[19]

19 The superiority of the dynamic models is in line with the findings of Guijarro and Hierro (2005) for the Spanish migratory system.

However, static and dynamic Markovian models are vulnerable to their main criticism, they are good tools to model migratory flows, but they do not explain the phenomenon they model. In this work they noticeably replicate the interstate migratory flows in Mexico, but they do not explain *why* these flows generate. Explain them would be the subject of other complementary demographic approaches. Asking the Markovian models an explanation would be too much.[20]

Table 3. Summary: Static Markov Chains (SMC) *versus* Dynamic Markov Chains (DMC)

Dimensions to assess	Markov Chains	
	Static	Dynamic
Inequality		
Gini index		
Emigration	+	
Immigration	+	
Global	=	=
Shape of Lorenz Curve		
Emigration	=	=
Immigration	=	=
Migratory competence		
Emigration	=	=
Immigration	=	=
Global	=	=
Statistical similitude		
Emigration		+
Immigration	=	=
Global		+
Mean proportional error		+

Source: own elaboration.

20 The topic of explanation in social sciences implies an epistemological debate or wide proportions. There are even a number of *theories of explanation*. In social sciences it is accepted that not always it is possible to produce causal explanations as in the fields of material sciences (such as physics or chemistry) or in life sciences (such as biology), which is an epistemological deficit inherent to social sciences (Rappaport, 1995).

Bibliografía

ALESSIO, E., Carbone; A. G. CASTELLI and V. FRAPPIETRO, 2002, "Second-order moving average and scaling of stochastic time series", in *The European Physical Journal B-Condensed Matter and Complex Systems*, vol. 27, no. 2, pp. 197-200.

ARAAR, Abdelkrim y Jean-Yves DUCLOS, 2006, *Poverty and equity: measurement, policy and estimation with DAD,* Springer-Verlag, NY.

BELL, M., Charles-Edwards; E., KUPISZEWSKA, D., KUPISZEWSKI, M., STILLWELL, J. and Y. ZHU, 2015, "Internal migration data around the world: Assessing contemporary practice", in *Population, Space and Place*, vol. 21, no. 1, pp. 1-17.

BOUTROS, N.; M.R. SHORTIS and E.S. HARVEY, 2015, "A comparison of calibration methods and system configurations of underwater stereo video systems for applications in marine ecology", in *Limnology and Oceanography: Methods*, vol. 13, no. 5, pp. 224-236.

BUSTOS, A., 2013, "Estimación markoviana del tamaño y de las estructuras de una población", in *Realidad, Datos y Espacio Revista Internacional de Estadística y Geografía*, vol. 4, no. 1, pp. 122-159.

CASTLES, S.; H. DE HAAS, and M.J. MILLER, 2013, *The age of migration: international population movements in the modern world*, Guildford Press. New York–London.

CHAKRABORTI, A. and G. GERMANO, 2010, "Agent-based models of economic interactions", in G. Naldi, L. Pareschi y G. Toscani, *Mathematical modeling of collective behavior in socio-economic and life sciences*, Springer Science & Business Media, pp. 3-29. NY.

CHORMANSKI, Jaroslaw; Tim VAN DE VOORDE, Tim DE ROECK, Okke BATELAAN and Frank CANTERS, 2008, "Improving distributed runoff prediction in urbanized catchments with remote sensing based estimates of impervious surface cover", in *Sensors*, vol. 8, no. 2, pp. 910-932.

CHING, W.K., X. HUANG, M. NG, and T.K. SIU, 2013, *Markov Chains Models, Algorithms and Applications*, Springer. NY.

CONT, R. and L. WAGALATH, 2013, "Running for the exit: distressed selling and endogenous correlation in financial markets", in *Mathematical Finance*, vol. 23, no. 4, pp. 718-741.

D'AMICO, Guglielmo and Giuseppe DI BIASE, 2010, "Generalized concentration/inequality indices of economic systems evolving in time", in *Wseas Transactions on Mathematics*, vol. 9, no. 2, pp. 140-149.

DE CASTRO, M.C., 2007, S"patial demography: An opportunity to improve policy making at diverse decision levels", in *Population research and policy review*, vol. 26, no. 5-6, pp. 477-509.

FIELD, A., 2005, *Discovering Statistics Using SPSS*, Sage Publications. London.

GARCÍA-ABAD, R., 2003, "Un estado de la cuestión de las teorías de las migraciones", in *Historia Contemporánea*, no. 26, pp. 329-351.

GARROCHO, C., 2011, *Población flotante, población en movimiento: conceptos clave y métodos de análisis exitosos*, United Nations Population Fund-El Colegio Mexiquense-Conapo. Mexico.

GARROCHO, C., T. CHÁVEZ, and J.A. ÁLVAREZ-LOBATO, 2002, *La dimensión espacial de la competencia comercial*, El Colegio Mexiquense. Mexico.

GEYS, B., 2006, "Looking across borders: A test of spatial policy interdependence using local government efficiency ratings*",* in *Journal of Urban Economics*, vol. 60, no. 3, pp. 443-462.

GIMENEZ, O., LEBRETON, J. D., GAILLARD, J. M., CHOQUET, R., and R. PRADEL, 2012, "Estimating demographic parameters using hidden process dynamic models", in *Theoretical population biology*, vol. 82, n*o.* 4, pp. 307-316.

GIORGI, G.M., 1993, "A fresh look at the topical interest of the Gini concentration ratio", in *Metron*, vol. 51, no. 1-2, pp. 83-98.

GÓMEZ-GARCÍA, J., M.A. PALACIOS and J. GARCÍA-PÉREZ, 1997, "Movimientos migratorios intermunicipales en la Comunidad Autónoma de Murcia: un enfoque markoviano", in *Cuadernos de Economía Murciana*, vol. 21, no. 1, pp. 61-73.

GOODKIND, D. and L.A. WEST, 2002, "China's floating population: definitions, data and recent findings", in *Urban Studies*, vol. 39, no. 12, pp. 2237-2250.

GUIJARRO, M. and M. HIERRO, 2007, "Un análisis global y desagregado de la dispersión espacial de las migraciones interiores en España (1986-2003)", in *Cuadernos Aragoneses de Economía,* vol. 17, no. 1, pp. 163-180.

GUIJARRO, M. and M. HIERRO, 2005, "Un análisis de la dinámica de los movimientos migratorios interregionales en España (1986-2001): una explotación del método MCC", in *Investigaciones Regionales*, no. 6, pp. 125-140.

HARVIR, S.B., S.F. TAYLOR and Y. ST. JAMES, 2005, "Migrating to new service providers: toward a unifying framework of consumers' switching behaviors", in *Journal of the Academy of Marketing Science*, vol. 33, no. 1, pp. 196-115.

HIERRO, M. and M. GUIJARRO, 2006, "Un estudio mediante cadenas de Markov de la dinámica de los movimientos migratorios interterritoriales en España (1990-2003) desde un planteamiento de estimación dinámico"*,* in *Revista Asturiana de Economía*, no. 35, pp. 145-161.

HIERRO, M. and M. GUIJARRO, 2007, "Una revisión de la aplicación de las cadenas de Markov discretas al estudio de la movilidad geográfica", in *Estadística española*, vol. 49, n*o.* 166, pp. 473-499.

KAYACAN, E., B. ULUTAS, and O. KAYNAK, 2010, "Grey system theory-based models in time series prediction", in *Expert Systems with Applications*, vol. 37, no. 2, pp. 1784-1789.

LE GALLO, Julie and Sandy DALL'ERBA, 2006, "Evaluating the temporal and spatial heterogeneity of the European convergence process, 1980-1999", in *Journal of Regional Science*, vol. 46, no. 2, pp. 269-288.

LIN, S.N., C.Y. CHOU, S.L. WANG and L. HUI-RONG, 2012, "Economic design of autoregressive moving average control chart using genetic algorithms", in *Expert systems with applications*, vol. 39, no. 2, pp. 1793-1798.

LÓPEZ-VEGA, R. and S. VELARDE, 2011, "Aplicación de medidas de concentración para el análisis demográfico de la migración interna de México", in *La situación demográfica de México 2011*, Consejo Nacional de Población, CDMX, pp. 123-139.

LUBRANO, M., 2015, *The econometrics of inequality and poverty: Lorenz curves, the Gini Coefficient and parametric distributions*, Le Groupement de Recherche in Économie Quantitative d'Aix-Marseille (GREQAM), Marseille, France.

MASCIANDARO, D. and M. QUINTYN, 2009, "Measuring financial regulation architectures and the role of the central banks: the financial supervision Herfindahl-Hirschman Index", in *Paolo Baffi Centre Research Paper*, no. 2009-55.

MARTÍNEZ-FAURA, Ú. and J. GÓMEZ-GARCÍA, 2001, "Modelos Migratorios: una revisión", in *Revista Asturiana de Economía*, no. 21, pp. 209-235.

METZGER, W., 2006, *Laws of seeing*, The MIT Press, Cambridge, Mass. USA.

MODICA, G. and L. POGGIOLINI, 2013, *A first course in probability and Markov Chains*, John Wiley & Sons. Chichester, West Sussex, England.

MULLIGAN, G.F., 1991, Equality measures and facility location, in *Papers in Regional Science*, vol. 70, no. 4, pp. 345-365.

O'KELLY, M.E. and M.W. HORNER, 2003, "Aggregate accessibility to population at the county level: US 1940–2000", in *Journal of Geographical Systems*, vol. 5, no. 1, pp. 5-23.

PARTIDA, V., 1989, "Aplicación de cadenas de Markov para proyecciones demográficas en áreas geopolíticas menores", in *Estudios Demográficos y Urbanos*, vol. 4, no. 3, pp. 549-571.

PLANE, D.A. and G.F. Mulligan, 1997, "Measuring spatial focusing in a migration system", in *Demography*, vol. 34, no. 2, pp. 251-262.

RABINER, L., 1989, "A tutorial on hidden Markov models and selected applications in speech recognition", in *Proceedings of the IEEE*, vol. 77, no. 2, pp. 257-286.

RABINER, L., and B.H. JUANG, 1986, "An introduction to hidden Markov models", in *ASSP Magazine IEEE*, vol. 3, no. 1, pp. 4-16.

RAPPAPORT, S., 1995, "Economic models and historical explanation", in *Philosophy of the Social Sciences*, vol. 25, no. 4, pp. 421-441.

RASKALL, P. and G. MATHESON, 1992, Understanding the Gini Coefficient, in *Newsletter: Social Policy Research Center,* University of New South Wales, vol. 6, Sidney, Australia.

ROGERS, Andrei and Stuart SWEENEY, 1998, "Measuring the spatial focus of migration patterns", in *The Professional Geographer*, vol. 50, no. 2, pp. 232-242.

ROGERSON, P.A., 1979, "Prediction: a modified Markov Chain approach", in *Journal of Regional Science*, vol. 19, no. 4, pp. 469-478.

ROMO-VIRAMONTES, R., Y. TÉLLEZ-VÁZQUEZ and J. LÓPEZ-RAMÍREZ, 2012, "Tendencias de la migración interna en México en el periodo reciente", in Conapo, *La situación demográfica de México 2013*, Conapo, pp. 83-106. Mexico.

SINGH, S., 2001, "Multiple forecasting using local approximation", in *Pattern recognition*, vol. 34, no. 2, pp. 443-455.

SPICKER, P.; S. LEGUIZAMON and D. GORDON, 2007, *Poverty: an international glossary*, Zed Books, NY, pp. 87.

SYS, C., 2009, "Is the container liner shipping industry an oligopoly?", in *Transport Policy*, vol. 16, no. 5, pp. 259-270.

STARTZ, R., 2008, "Binomial autoregressive moving average models with an application to US recessions", in *Journal of Business & Economic Statistics*, vol. 26, no. 1, pp. 1-8.

VANDEWALLE, N., M. AUSLOOS and Ph. BOVEROUX, 1999, "The moving averages demystified", in *Physica A: Statistical Mechanics and its Applications*, vol. 269, no. 1, pp. 170-176.

WATKINS, J.F., 1986, *Migration of the elderly in the United States: a multiregional analysis*, Tesis de Doctorado, University of Colorado, Boulder, Colorado, USA.

WEBBER, M.J., 1984, *Explanation, prediction and planning: the Lowry model*, Taylor & Francis. London.

YIN, G.G. and Q. ZHANG, 2010, *Continuous-Time Markov Chains and applications: a two-time-scale approach*, Springer. NY.

About the Authors

Carlos Garrocho
Master in Urban Development from El Colegio de México (1987) and Doctor in Socioeconomic Geography from the University of Exeter (England: 1992). National researcher level 3. Member of the Commission for the Assessment of Social Sciences of the National System of Researchers (2014-2016). In El Colegio Mexiquense: researcher since 1986; Member of the Government Board (2016); Member of the Academic Council (since 2008); Academic Coordinator (1994-1998); Founder of the Journal *Economía, Sociedad y Territorio* (1994) and director of this journal (1994-1998 and as of 2015). He has published 22 scientific books and 84 chapters and national and international scientific articles. Magistral lecturer on behalf of the Mexican Delegation in the Population Summits, 2014 and 2016 in UN at NY. Appointed by the United Nations Population Fund: Coordinator of an expert group on the tropic of sustainable cities (as of 2014). Appointed UN as an international expert on cities and Mexican representative in the XXIII Cumbre Iberoamericana de Panamá (2013). Awarded with the Premio Estatal de Ciencia del Estado de México (2011). Secretary of Planning of Development of the government of the State of San Luis Potosi (1997-2003). His most recent book is: *Segregación socioespacial de la población mayor en la Ciudad de México: la dimensión desconocida del envejecimiento,* El Colegio Mexiquense, CDMX, Mexico. In collaboration with Juan Campos.
Email address: cfgarrocho@gmail.com

Eduardo Jiménez López
Doctor in Applied Sciences from the Autonomous University of San Luis Potosí. Member of the national system of researchers (candidate since 2014). Researcher-professor in El Colegio Mexiquense A.C. His most recent publication is “Estructura profunda de los flujos migratorios en México, 1990-2010”, in Conapo, *La situación demográfica de México 2014*, Conapo, México, in collaboration with Carlos Garrocho and José Antonio Álvarez Lobato. His main research line is modelling of internal migration in Mexico.
Email address: eduardo.jimenez@ipicyt.edu.mx

José Antonio Álvarez Lobato
Doctor in Geografía, Master in Engineering and Computing Engineering. Member of the national system of researchers. Secretary General of El Colegio Mexiquense, A.C., Coordinator of the Laboratory of Socio-spatial Analysis and member of the seminar of Strategic Studies of the State of Mexico in the same institution. Member of the Academic Technical Committee of the Conacyt Thematic Network of Poverty and Human Development up to 2015. Author of a number of book chapters and scientific articles on topic of urban observation, spatial analysis and urban economy, as well a various dissemination articles on ICT. Among his publications one finds: *La dimensión espacial de la competencia comercial*; *Observatorios urbanos en México: lecciones, propuestas y desafíos*; "Acceso a oportunidades: el principal desafío" in Brambila, Carlos (Ed.) *Prioridades de Investigación sobre Pobreza y Desarrollo Territorial*, *México: 2020*, ITESM, 2015; "Spatial organization of banking branches in the intra-metropolitan space in Toluca, Mexico" (coauthored with Carlos Garrocho) in Garrocho, C. (ed.). *Advances in commercial geography*, Mexico, El Colegio Mexiquense, 2013. "Calculation intraurban agglomeration of economic units with planar and network K-functions: a comparative analysis, coauthored with Carlos Garrocho and Tania Chávez, in *Urban Geography*, 34(2), Abril, 2013. His areas of interest focus on urban geography, spatial analysis, accessibility and urban mobility, online GIS and development of indicators for urban monitoring.
Email address: jalvar@cmq.edu.mx

Article received on May 8th, 2015; approved on February 1st, 2016.