10-2013

The queries to Google Search as predictors of migration flows from Latin America to Spain

Dawid Wladyka
The University of Texas Rio Grande Valley

The Queries to Google Search as Predictors of Migration Flows from Latin America to Spain

by

Dawid Wladyka

A Thesis Presented to the Faculty of the College of Liberal Arts in Partial Fulfillment

of the Requirements for the Degree of

Master of Arts in Interdisciplinary Studies

In the Field of Individualized Studies

Approved by

Dr. William Yaworsky
Thesis Director

Dr. Bernardo De La Garza

Dr. Leslie D. Meyer

Dr. Charles Lackey

Dean of Graduate Studies

Graduate School

University of Texas at Brownsville

The Queries to Google Search as Predictors of Migration Flows from Latin America to Spain

A Thesis Presented to the Faculty of the

College of Liberal Arts

The University of Texas at Brownsville

In Partial Fulfillment

of the Requirements for the Degree

Master of Arts in Interdisciplinary Studies

by

Dawid Wladyka

October 2013

Abstract

Recently, the development of global network and ITC technology provided new opportunities to improve the estimations and predictability of migration flows. The activity of users of e-mail and other web-based services was compared in time and space in order to track international human mobility. At the same time, the IP based geolocation linked to Google Search proved to be efficient in geographically tracking the outbreaks of several illnesses, and also in predicting changes in economic indicators and travel patterns. This research draws from both experiences. It compares the popularity of migration-to-Spain related queries introduced to Google Search in Argentina, Colombia and Peru, to changes in a quantity of residents' registrations in Spain, performed by immigrants proceeding from these countries between the years 2005 and 2010. Following the preliminary visual trend analysis, the time series are pre-whitened in order to formally test for a time-shifted correlation and predictability not-influenced by a general series trend. The analysis was performed on the datasets of queries popularity derived from Google Trends and anonymized micro-data of Residential Variation Statistics based on the Municipal Register of Spain. The predicted lags of one or more months that showed to be significantly correlated according to the Cross-Correlation Function have been further used to evaluate its predictability with regression analysis. The results show a significant correlation and weak to moderate predictability for the lags of several months depending on the particular country. The findings support the assumption that popularity of queries to Google Search provided by Google Trends might constitute a useful predictor of migration flows while at the same time it indicates further developments necessary in order to improve its analytical capacities.

Keywords: migration flows, IP geolocation, Google Trends, Latin America, Spain

Contents

Chapter I – Introduction to the Study

1.1. Statement of the Problem and Rationale

The aim of the study is to evaluate the relation between the changes in proportion of the migration-related queries made to Google Search and changes in volume of international migration flows between particular origin and destination countries. The purpose of this analysis is to determine serviceableness of the Google Trends application in predicting the international migration flows. Therefore, this study assesses if a cost-free and relatively easy to use Google Trends improves the tracking and prediction of the international migratory flows, in a sense that it could be proposed as a tool for organizations and policymakers.

The quality of statistical data on international migratory flows has been a constant challenge not only for researchers, but also for local and regional policymakers that have been eager to obtain a possibility of predicting the number and origin of immigrants that seek a new, at least temporary home in their country, region or city. Despite the attempts of normalization of international flows statistics between the countries (e.g., by the European Union[EU]) the data provided is still unreliable and not comparable (Kupiszewska, 2008; Moses, 2012; United Nations[UN], 1998; Zagheni & Weber, 2012).

In the terms of application of knowledge, the abovementioned shortcomings not only prevent local authorities from adjusting their policies in order to meet the social reality produced by an immediate future of immigration flows, but also results in lack of awareness about the current demographic image of the region or city in question. This lack of data results with limited possibility of adjusting the demand for use of local services (such as kindergartens, schools, medical facilities) or policies

related to housing, welfare state, and employment (UN, 1998). Furthermore, not only a raw number or estimated proportion of immigrants, but also their geographic origins that could indicate cultural traits constitute knowledge useful from a point of view of receiving country authorities. This information would allow undertaking such basic challenges as to train and employ an appropriate number of relevant translators and social-cultural mediators.

The data is not only valid information for receiving, but also to sending countries since the migrant's remittances directly influence the economy of origin. The impact of remittances can be crucial for sustaining the robustness of the country's economy, as, for example, during the (add) the global economic crisis (UN, 1998). The recent economic recession, during which a recent European Union member – Poland – did not formally enter the economic recession (lack of Gross Domestic Product decrease), is a perfect example of such an impact. The hundreds of thousands of Polish that migrated to the more developed European countries (mainly UK, Ireland, and Germany) not only strengthened the Polish economy with remittances, but also alleviated issues related to the unemployment on the Polish labor market noted during the European recession (Ministerstwo Spraw Zagranicznych, 2013). While this positive effect should be contrasted to the negative "brain drain" phenomenon (Krajowy Punkt Kontaktowy Europejskiej Sieci Migracyjnej, 2011), both demonstrate the economic significance of international migration flows.

Luckily enough, the recent two decades, altogether with the growing number of international migration flows brought also the development and dynamic dissemination of information technology. Therefore, the spatial mobility of individual has already begun to be researched with use of the IP [Internet Protocol] geolocation technology. The same technology is used by Google Search engine in order to

geographically track incoming queries, which can be further examined by the Google Trends application. Therefore, this research examines if, at the current stage of the technological development and data accessibility, the later mentioned tool can be a useful addition to other methods of estimating and predicting international migration flows.

1.2. Definitions of Basic Concepts

Four crucial concepts should be introduced at this point in order to allow the reader a seamless lecture and comprehension of the following chapters. These include: international migration phenomenon, IP address geolocation technology, and Google Search and Google Trends online services.

1.2.1. International migration.

Migration is a long-distance displacement undertaken by human beings, usually in large groups (Demuth, 2000, p. 26). This understanding is rooted in the first human migrations started with *Homo erectus* leaving the African continent towards Eurasia about 700,000 years ago.[1] After crossing the borders of Southern Europe, the progenitor of contemporary human began to spread towards the north and north-east of the continent. Among the possible reasons for this expulsion from Africa were the abundance of food sources and shelter places (like caves) in Europe, and competition for scarce resources in Africa resulting in exile of weaker groups (Schrover, 1995a; Schrover, 1995b). As the above example shows, in order to appreciate the

[1] The most recent paleoanthropological discoveries tentatively indicate that early *Homo* could leave Africa over a million years earlier. The analysis of fossils in Dmanisi, Republic of Georgia not only poses a question over the patterns of these migrations, but also bring new perspectives on the genes variations between the early *Homo* species (Gibbons, 2013; Lordkipanidze, 2013).

significance of the *migration* processes in the development of humanity one could take into account that migration (gene flow) and subsequent separation of human groups (technically the founder effect is a form of genetic drift) are among the four causative factors of evolution.[2]

On the other hand, in the contemporary societies, migration is one of the most important factors of population growth (or decline). Therefore, at the turn of the XX and XXI centuries, there were 125 million international migrants[3], and over one billion internal migrants[4]. The number of international migrants grew to 214 million in the year 2010 (see International Organization for Migration [IOM], 2010; IOM, 2013). This intensification of the migration phenomena drove scholars to use of the term "*The Age of Migration*" while talking about our times (Castles & Miller, 1993; Demuth, 2000, p. 22 - 26). Still, the migration processes of today differ significantly from the ones existing in the past. The beginning of the twentieth century has seen the international economic migrations that reached a number of three million people yearly. At the very same time, the idea of the nation-state has been spreading dynamically, and one of its bases was the introduction of the limitations in the free movement of people, but also creation of myths, that should mirror the importance of the migration processes in setting up particular nations, especially in the Americas (Collinson, 1994, pp. 35 - 45; Devoto, 200, pp. 45 - 91; Manning, 2005, pp. 132 - 162; Zolberg, 2005, pp. 119 - 123).

The popularization of travel documents (i.e. passports) and closure of the borders that last till today were visible especially from the ninetieth and beginning of

[2] Next to: natural selection, genetic drift, and mutations.
[3] International migrants– people crossing the country border during migration process. Depending on sources and statistic, there were about 125 to 190 million at the turn of the centuries (Demuth, 200, pp. 22 – 26)
[4] Internal migrants – people moving within the territory of one state/political entity (Demuth, 200, pp. 22 – 26)

the twentieth century. The borders reopened only in a very few cases, e.g., Schengen Treaty related to the internal European migrations. However, these closed borders and restricted immigration laws did not end the migration processes. This contrary effect is especially visible in the last three decades**.** Consider the example of the EU. Despite tightening of immigration policies, the inflow of immigrants in 2011 constituted 68% of the population growth in the entire EU.[5] The difference between planned and obtained effects of the immigration policies has several reasons, but the need for inexpensive manpower that guarantees the stability of the economy and welfare state should be highlighted. This significance of immigration for European demography is a fundamental change compared to previous centuries, when Europe was a sending territory, especially to the Americas and Australia. One of the particular examples of this shift of patterns is Spain. This country no longer constitutes neither a significant source of migration to the Americas, nor to wealthier European states. On the contrary, during the last twenty years it became a country receiving important numbers of immigrants from Latin America, Africa, and Eastern Europe, but also converted into one of the favorite destinations of Western European entrepreneurs, and especially retired sun-seekers from Germany and United Kingdom. While this trend slightly slowed down with the economic crisis, the emerging emigration of Spaniards, especially to the better-off European countries, has to be considered as still a limited phenomenon (Colectivo Ioé, 2008).

The worldwide introduction of borders and passports brought up a new phenomenon, and a new research field: *international migration*. The definition of what *international migration* is, and who is considered an *immigrant* from its perspective varies not only between countries, but also poses legal and statistical

---

5 http://epp.eurostat.ec.europa.eu/statistics_explained/index.php/Population_and_population_change_statistics

issues within the particular national legal systems (see e.g. Anderson & Blinder, 2013). The frequent common ground for describing this phenomenon is the United Nations indication that

> an *international migrant is defined as any person who changes his or her country of usual residence*. A person's country of usual residence is that in which the person lives, that is to say, the country in which the person has a place to live where he or she normally spends the daily period rest. Temporary travel abroad for purposes of recreation, holiday, business, medical treatment or religious pilgrimage does not entail a change in the country of usual residence." (UN, 1998, p.17)

The above cited definition alters the use of residence as an indicator of international migration. Therefore, the general rule of the thumb says that "the change of country of usual residence necessary to become an international migrant must involve a period of stay in the country of destination of at least a year" (UN, 1998, p.18), which implies that the destination country becomes a usual residence country for a migrant after a year of stay. Subsequently, the same United Nations document (1998) introduces two crucial types of international migration: long-term and short-term. The long-term migrant would be the one fulfilling the above explained requirement of 12 months residence. The short-term migrant, on the other hand, would "move to a country other than that of his or her usual residence for a period of at least 3 months but less than a year […] except in cases where the movement to that country is for purposes of recreation, holiday, visits to friends and relatives, business, medical treatment or religious pilgrimage." (UN, 1998, p. 18)

This research assumes a simplified, residence based understanding of the *international migration* event, and therefore does not make distinction between short-

and long-term immigrant. Subsequently, the official enrollment in the municipal register of residents is understood here as a threshold of migration event (see Section 3.1.2. for details). Still, it is important to mention at this point, that this effort of clarification and categorization undertaken by the United Nations has been driven by the important challenges of estimating the international migration flows. Since these issues constitute a rationale of this research, a section (2.1.) will introduce the several coexisting typologies of migration phenomena in the further part of the thesis.

1.2.2. IP address geolocation.

The notion of geolocation bears multiple meanings depending on particular discipline and service it describes. In this thesis however, the geolocation (also called geo-tagging) will refer to technological means (e.g. Internet IP address, cellular network, Wi-Fi network, GPS [Global Positioning System] device coordinates) that allow to determine the individual's location, and to track the change of this location at various geographical and administrative levels (e.g. international, cities, blocks or streets). The focus here will be on IP address geolocation which is one of the most sophisticated, but at the same time common, and quick ways to determine the location of the Internet service’s user (King, 2010).

In order to understand how the geolocation based on the IP address works, it is essential to explain the process of web-browsing from the perspective of an individual user and to shed light on the interaction of the user’s device with the server where the accessed website is hosted. Svantesson (2004), and building on his work, King (2010), provided accessible and straightforward descriptions of the process, thus the following explanation is founded on their works.

In the most basic form, the user-side process of browsing the Internet websites is composed of either introducing the website address called Uniform Resource Locator [URL] into the address bar of the Internet browser, or clicking a hyperlink containing the information about a desired URL. After completing either of these actions, the browser sends an access-request to the server hosting the desired content in order to provide it to the user. While there are some in-between technical steps that contain the mechanical translation of the human accessible URL to the numeric Internet Protocol [IP] address by the appropriate Domain Name Service [DNS], these will be omitted here since they are not essential to the studied feature. After receiving the access-request, the server hosting the content makes a location-request for comparison of user's IP address with the database of the provider of the geolocation service. The provider of geolocation service returns an educated guess regarding the user's geographic location, and sends it to the server hosting desired content. While the server hosting the content can disregard this information, it could also restrict the access to the website if its content is for example protected with the copyrights limited to the particular territory. For example, a Spanish user could be denied access to the movie that, for this particular website, is restricted to users only from the United States. Obviously, there are more complex applications of this technology, like providing profiled content, and especially targeting the advertisement.

It should also be observed that the described above IP address geolocation could be substituted by other forms of detecting user's location. The websites often take advantage of the user's browser settings regarding country, time zone, and language of user. Still this kind of location's guess is less accurate than the one based on the IP address location which reaches 99% accuracy at the country level, and over 90% accuracy for the cities worldwide (Svantesson, 2004; King, 2010).

1.2.3. Google Search and Google Trends.

The research presented in this thesis focuses on and uses two tools: Google Search and Google Trends. While the former is a standalone flag product of Google, the later is quasi-analytical tool that allows relatively simple analysis of the data gathered by the Google Search. In order to understand the relation between both tools, and to comprehend why they have been chosen for this study, both will be introduced in this section. The broader examination of their utilization in this and similar studies will be provided in the further chapters.

The Google Search is an Internet search engine that allows users to search the Internet resources following the introduction of the particular *query* into the search engine. The *query* could constitute one word, a set of loosely related (or not) words, as well as an entire sentence(s). The Google Search is a most popular Internet search engine worldwide. It is responsible for over 78% Internet searches in the US (FairSearch, 2013), about 90% in Latin America (ComScore, 2011) and 93% in Europe (FairSearch, 2013). Thus, the FairSearch organization claims that Google Search dominance threatens to gain a monopoly over the information, and in particular is undesirable for the development of the Internet business. Nevertheless, this monopoly of the Google Search also constitutes an advantage for the potential analytical purposes since the *queries* submitted to Google are gathered within the database together with e.g. geolocated IP address data. The examples of the direct use of Google Search database for analytical purposes are presented in the further chapters.

Google Trends is an online based and freely accessible application that provides the frequency with which the given *query* has been entered in the Google Search in relation to the total search volume. It has been primarily designed to

compare the popularity of the various keywords in the Google Search, and to find relation between real world events and outbreaks in keywords popularity. Importantly, besides the search volume, it also provides the geographic location of the *queries'* origin. As in the case of Google Search, the examples of analytical exploitation of the Google Trends application, and its' use in this research are provided in the further sections of this thesis.

1.3. Assumptions and Limitations

1.3.1. Time shifting of the information search stage of migratory process.

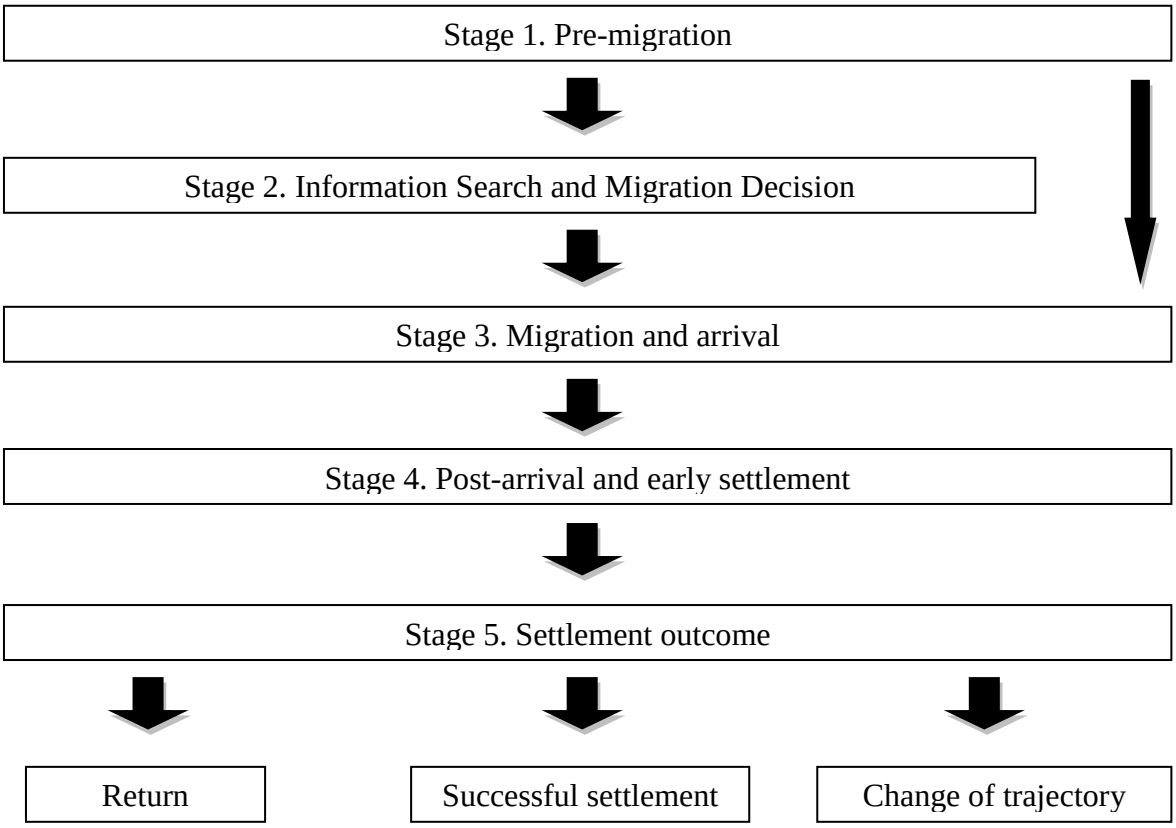
The main assumption of this study is that persons that plan any migratory activity would acquire knowledge about a potential destination through the media, and as a part of this process they would use the Internet to search certain information related to the geographical destination that they aim to reach (Desiderio, 2012; Hamel, 2009). That assumption is similar to the influenza-like illnesses detection study (Ginsberg et al., 2009) described in the section 2.4. of this thesis. Nevertheless there would be some important modifications in the construction of the relation between real-life occurrence of events and a change in query's popularity in Google Search. The occurrence of queries related to influenza study was assumed to emerge as a result of arising influenza-like-illness epidemic (Ginsberg, 2009). In other words, people or relatives of people who detected symptoms would relatively immediately search for the treatment advices and medications in Google Search what would prompt an outbreak of searches.

In the study of migration flows the order is assumed to be the opposite, (i.e. the change in the popularity of searches would precede the actual change in migration flows). Another crucial difference to Ginsberg et al. (2009) study is a time lag

between change in query popularity in Google and the change in migration flow occurrence. In fact, in case of migration studies, this difference would be twofold. First, there is often occurring time lag between an immigrant's arrival at the destination and registering in the population register (or being registered by some system of population control). Second, there should be a time lag between the change in the popularity of search for migration related queries (e.g. job in the destination, flight tickets, apartment, etc.) and the change in migration flows itself. Since the studies on Internet use by the migrants focus mainly on the practices in destination countries, there are no estimations to a time distance that a prospective immigrant would search online for information before the event of migration (Chen, 2010; De Tona & Whelan, 2009; Fox & Livingston, 2007; Kissau, 2009; Orozco, Burgess, & Ascoli, 2010). However, for this study we will assume the *time shift* of several months between the events when summing up the both mentioned time lags. This assumption is concordant with migration models that include the pre-migration preparations, like the model presented by Benson-Rea and Rawlinson (2003) where the early settlement is not until fourth stage of a five stages migration process (see Figure 1).

Figure 1

Benson-Rea and Rawlinson (2003, p.66) model of migration stages.



1.3.2. The Spanish Residential Variations Statistic as an exemplary database.

The before mentioned assumption of time lag importantly impacts on the data sources accessible for the analysis. Most importantly, the predicted gap of several months would make irrelevant the use of annual migration flows data since these would not be able to precisely render the relation between the change of searched queries popularity and change of migratory flows. Since most of the statistics on migration flows are published on an annual basis, the relevant option is use of the anonymized micro-data of Residential Variation Statistics, based on Municipal

Register (i.e. *padrón*) provided by the National Statistics Institute of Spain. This dataset provides daily data on new registrations or address changes of Spanish residents with relation to particular country of previous residence.

Furthermore, this study is based, inter alia, on the previous research carried out by State et al. (2012) who underlined the role of language and former colonial links in the migratory flows based on the use of IP geo-tagging. Following this pattern, this study analyzes the migratory flows between Latin America and Spain. The selection of Spain as a destination country is also supported by the previously indicated growth of Spain's importance in receiving large number of immigrants. Furthermore, the Americas constitute the origin of the majority of newly registered residents with previous residence outside Spain in the 21th century. It is especially worthwhile to underline here that 34.19% of the new registrations of residents with previous foreign address regarded those that arrived from Americas, and that this number even slightly surpassed those with previous European residence (see Table 1).

Table 1

Enrollments of the inhabitants with previous foreign residence (Instituto Nacional de Estadística, 2012)

	2000 - 2012	**% of Total**
Total	7448095	
America	2546345	34,19
Europa	2544844	34,17
Africa	905359	12,16
Asia	351727	4,72
Oceania	8218	0,11

Still, it should also be stressed that the Google Trends does not cover reliably the search traffic from all Latin American countries or in certain cases the Google Search traffic volume may be too limited to be captured by the current version of the Google Trends. Consequently, the results from only Spanish speaking countries that are fully covered by Google Trends during a particular time span are analyzed. The very similar limitation addresses particular *queries* that could be used in this research, but are not be adequately covered for the whole time span of the study's interest, and therefore would be eliminated from the analysis. The detailed information on countries of origin analyzed, and queries used to analysis are provided in the *section 3.1* devoted to the description of the datasets used.

1.4. Hypotheses of the Study

Taking into account all the described assumptions and limitations, the hypotheses of the study are the following:

H_1: The time-shifted change in proportion of migration-related queries made to Google Search in sending country and further reported by Google Trends is significantly related to changes in volume of international migration flows from selected Latin American countries to Spain.

H_2: The time-shifted change in proportion of migration-related queries made to Google Search in sending country and further reported by Google Trends predicts changes in volume of international migration flows from selected Latin American countries to Spain.

Chapter II – Literature Review: Migration and Geolocation

2.1. Challenges in Estimations of International Migration Flows

As it has been previously mentioned, the phenomenon of international migration is directly linked to the formation of nation-states, and their sovereignty to grant or deny access to the state's territory. Therefore, the majority of data on international migration flows is collected and related to the systems of the control of territory access. Subsequently, national laws set-up various categories of people (e.g. based on the citizenship) that are expected to comply with distinct requirements for entering the state's territory. This situation furthermore results with various ways of classifying people in national migration statistics, and its further lack of compatibility at international level (UN, 1998, p. 9).

The UN (1998, p.9) also highlights the effect that the changing nature of the international migration has on the quality of the statistics describing it. Several factors, like demand for foreign workers influence how the receiving state treats the foreigners that enter its territory with an intention to stay there for a certain amount of time. In consequence, a number of international population movements fall into the "grey area" between non-migration, short-time, and long-time migration. Since the diversity and volatility of the phenomenon of international migration has such a significant impact on the quality of its statistics, it is a useful exercise to present some of the most important classifications of the migratory flows elaborated by the scholars.

There are four traditionally distinguished models of migration from the perspective of receiving countries (Giddens, 2005):

- *Classic:* regarding especially United States, Canada, and Australia, called "nations of immigrants". Their characteristic is that, at least at some stage of history, government activities were encouraging immigrants to arrive offering guarantees of the citizenship. Still, these were limited by various political instruments, like annual quotas, etc.
- *Colonial:* regarding countries with colonial past (e.g. United Kingdom or France), that introduce simplification of immigration procedures for the citizens of former colonies (e.g. Commonwealth citizens in the case of the United Kingdom)
- *Gust workers* (or *Gastarbeiters* from German original): the model especially popular in Germany, Switzerland, and Belgium. It consists of admitting immigrants for the particular and limited period as workers, with no assurances of citizenship acquisition regardless of the length of stay.
- *Illegal migration:* the phenomena resulting from the tightening migration laws, and borders' sealing. It basically consists in clandestinely entering the country, or staying on its territory with no appropriate document allowing the stay. It drives to several pathologies, starting with migration process itself (e.g. human trafficking), and concluding with the social cohesion of the receiving society (life outside the community).

Still, the demography of the contemporary migration processes let Castles and Miller (1993) distinguish its characteristic trends:

- *Acceleration:* Increase of the immigrants number comparing to the previous periods;

- *Differentiation:* previously prevailing forms of migration, like economic or exile, have been substituted with more diverse categories of migrants;
- *Globalization:* The global range of the migration phenomenon is reflected in an increasing number of sending and receiving countries;
- *Feminization:* The increase of the share of woman in the international migration rooted in: the changes of global market labor structure, demand for babysitters and housekeepers, expansion of sexual tourism and "order brides".

One of the basic criteria in the description of migration is its voluntariness. This feature influences directly on the migratory statistics and its interpretation. While the international law recognizes *migration* as voluntary movement from one point to another, the *exile* constitutes a constrained movement between the states' borders (Demuth, 2000).

Another important criterion is a migration range. There are *internal* and *cross-border migrations*. The first ones take place within the limits of the administrative or political unit, and the later means crossing that border. In the spatial sense, the migration processes can also be analyzed from the perspective of the *sending* and *receiving* society. These kind of opposing descriptions include *rural – urban*, *urban – rural*, *rural – rural*, and *urban – urban* migrations (Solomon, 2003). Yet another geographic point of view on migration processes is related to politics. The typology based on geopolitical criterions includes most commonly *East – West*, *North – South*, *South – North*, and *South – South* denominations of migrations. This last typology is obviously rooted in the Eurocentric perspective assuming that most migration movements focus on reaching the destination labeled as *West* or *North* in a sense of reaching the developed countries of Europe, but also Northern America, and

contradictory to geographical conventions, Australia and Oceania. Still, while denominating migration movement with these labels one should be extremely cautious to not to directly imply the migration reasons. Also, it should be acknowledged that during a long time, the Eurocentric perspective ignored the fact that the South – South migrations are more numerous, and that about 90% of migration worldwide starts and terminates in the developing, non-European countries (Agozino, 2005).

As it has been already mentioned, the factor that often distinguishes *migratory* from *non-migratory* movement in both academic research, and legislation is duration of migration/residence. In the simplest scheme, one can distinguish *temporary* (e.g. circular) and *permanent* migration. While the former includes temporary workers and students, the later includes people that permanently displaced their residence (Solomon, 2003).

The category of migration that catches a lot of mass-media attention is *undocumented migration*. The notions of *illegal* or *undocumented* migrants would refer always to the *cross-border migration*. The ideas behind this notions goes back again to the emergence of the idea of nation-state, and its sovereignty that implies the liberty of the authorities to make decision who, and on what conditions, can cross a border of the state. Obviously, over the years, these criteria change, depending on the economic situation and labor market structure of the receiving country, the political stances of the currently governing bodies, or simply immigration policy applied at the very moment. Therefore, while International Declaration of Human Rights grants the unconditional right to enter a country to the country citizens, none of the countries is obliged to allow the unconditional entry of foreign citizens. Still, there are some exceptions imposed by international agreement (like free movement of people within

the European Union) or the obligation to admit the refugees on the basis of the Geneva Convention (Demuth, 2000). Subsequently, the *legal alien* is the one whose arrival and stay in the receiving country do not violate internal regulations, and international agreements or conventions (Solomon, 2003). Determining then, who is an *undocumented migrant* is problematic (Demuth, 2000). The simplest and most similar criteria would be the following (Solomon, 2003):

- Trespassing the border outside of the official crossing;
- Staying in the country without the valid permit/visa;
- Developing activities that are not concordant with a permit/visa held
- Staying in the country after the permit/visa expired;
- Having a status of *persona non grata* while entering or staying in the country;

Therefore, the line between being a tourist and becoming an undocumented migrant in the perspective of local law can be as thin as the touristic visa expiration date. On the other hand, the entrance to the country on e.g. tourist visa, and subsequent overstay while undertaking activities (mostly employment) not permitted by the particular visa type, is one of the common ways of entering into the *undocumented migrant* status. Because the presence of *undocumented immigrants* has several drawbacks to the local economy and society (e.g. lack of tax income from the employment or lack of participation in the community) some governments attempt to provide massive regularizations of stay of the undocumented immigrants. Spain is an example of a country that since becoming an attractive immigration territory, every several years undertakes actions of massive “amnesty”. The critiques of this approach underline that apart from gaining new taxpayers, the massive regularizations could constitute bait for the next waves of undocumented immigrants that would enter the

country with an intention to await the next similar policy of that type (Moreno, 2004, pp. 18 - 22; Nowaczek, 2004, pp. 158 - 162).

As one can observe, the categories introduced in the previous paragraph brought a perspective focused more on the migrant as a person, rather than on the process of migration. Following this direction, it is worth to present here a typology that relates a motive for migration with the migration process voluntariness. Table 2 showcases the four most extreme cases. Still, it should be remembered that these categories usually mingle, and therefore it is difficult to find a pure extreme example in reality. For example, the economic and voluntary migration never appears in the political vacuum, as the access to the particular labor market for the particular immigrants is politically established. On the other hand, the migration of contract workers is frequently provoked by the need to overcome extremely difficult economic situations, though it has some element of forced migration (Collinson, 1994).

Table 2

The classification of migration by motive and voluntariness (Collinson, 1994, p. 2)

	Economic Migration	**Political Migration**
Voluntary Migration	1) Reason and motivation are rooted in economic circumstances, and the process is voluntary **(e.g. migration in search for better employment, salary, economic opportunities)**	2) Motivation is political (ideological), but the migration process is voluntary **(e.g. Jews migrating to Israel)**
Forced Migration	3) Economic reasons compel to leave the current residence **(e.g. refugees from hunger, environmental hazards, and catastrophes)**	4) Because of the political (ideological) reasons, person is forced to leave current residence **(classical example of refugees)**

However detailed this or similar a matrix would be, one should realize that there is no point in attempting to find a cell for every possible case. Just to give an example, the Table 2 could be easily enlarged with social motives for migration (e.g. family rejoining) or cultural motives for migration (e.g. similarity of languages between sending and receiving countries). That would obviously not terminate the possibilities since the highly skilled and disadvantaged migrants, legal and undocumented ones, or temporary and permanent could be added. Therefore, the common research perspective is examining the individual context and motives for migration. This approach, from the point of view of the particular study, could be generalized by describing the migration process with *pushing* and *pulling* factors (see Table 3). While the first ones are responsible for the "expulsion" from the *sending* country, the later explain why the migration finishes in the particular *destination*. This typology, gives a glimpse of the migrant's perspective on motives of migration while at the same time still provides some possibility for attempts of migrants' categorization in relation to other typologies. The elasticity is an important advantage here, since the additional factors can be easily added and analyzed (Demuth, 2000).

Table 3

Voluntary migration in perspective of the push and pull factors (Agozino, 2000, p. 36)

Push Factor	**Migration Type**	**Pull Factor**
Degradation of the life quality, poverty, overly bad situation in poor country	Economic migration	Stability of employment, the example of life in the affluent society
Unemployment, disadvantaged work conditions, low salaries	Economic migration	New Job, better conditions, better salary
Separation from family	Chain migration Migration networks	Family rejoining

Study and research, acquiring special qualifications abroad, foreign delegations related to employment	The population movements due to this reasons are usually not described as migration	Particular conditions or place for studies, research or development

Regardless the assumed categorization, the data on migration flows are often generated a posteriori by comparison on time series of migration stocks as extracted from census data. The comparison of statistics based on definition of immigrant (or migration process) is often biased because of the use of different definitions by particular national statistical authorities. Another problematic issue is a time lag in which migration statistics are delivered (State et al., 2012). In countries where statistical authorities build their detailed data only on population censuses, carried out in a certain threshold of time (usually in ten-year periods) the availability of data makes them virtually useless to everyday analysis and policymaking. In some cases (e.g. developing countries) the data on migration flows can be entirely unavailable because of the financial limitations and/or lack of infrastructure. During the economic recession financial cut backs in the statistical authorities of developed countries could result with deterioration of the population statistic's quality.

On the other hand, already existent systems of capturing the migration flows deal with a number of drawbacks inherent in the previously explained nature of migratory processes. These issues are primarily related to the undocumented immigrants' hesitation (or inability) to register because of the status related reasons. However, the delays between the displacement itself and registration apply also to the general immigrants' population. These individuals need time in order to familiarize themselves with appropriate procedures and pros/cons of fulfilling administrative formalities. Additionally, the economic calculations (i.e. international taxation rules)

would negatively impact on the decision of deregistering from the population censuses of countries of origin. Therefore, also the flows estimations based on the population registers of origin countries could provide misleading data (Zagheni & Weber, 2012).

2.2. Application of Geolocation in Human Mobility Tracking

During the first endeavors with geolocation, the scholars' efforts have been principally focused on the description of short distance displacement patterns at the city and regional level. The well-known study carried out by Ferrari, Rosi, Mamei & Zambonelli (2011) used Twitter data in order to shed light on New York's urban displacement patterns. Also, mobile phone data and applications, such as Google Latitude and Foursquare, allow sharing of location data with friends have been subsequently used for description of short distance mobility displacements. On the other hand, the attention of the tourism oriented researchers has been drawn to analysis of spatial data. Photos and online posts which included geo-references in web-based services like Flickr and CoachSurfing, all together with mobile network data have provided the new tools for the development tourism industry (State et al., 2012).

Also the geo-location data on displacements, "long" with both distance and duration, have begun to be tracked and analyzed. On the one hand, statistical information provided by websites like wheresgeorge.com, which tracks one-dollar bills, and geocaching.com, which tracks various items (e.g. books, teddy bears, etc.), worldwide were used to examine displacement patterns (Zagheni & Weber, 2012). However, recently the more complex databases have been applied in order to track and analyze human mobility on the basis of geolocation of IP addresses. The latter are

numerical names of every particular device connected to the Internet. The IP number usually corresponds to the geographic region of the Internet provider and can point not only to country, but also to smaller units like regions and cities. Although the geolocation by the IP number brings some shortcomings (e.g. use of proxy servers that are located in a different physical place), they have provided a reliable source of spatial data, especially at a country level (Gueye, Uhlig & Fdida, 2007; Hui, Mortier, Piorkowski, Henderson & Crowcoft, 2009).

The most promising data has been brought by research based on the assumption of the common use of Internet and Google Search in particular for travel planning. Choi and Varian (2009b) found with use of Honk Kong visitors' data, that Google Trends was successful in predicting visits to touristic destination. There are three crucial findings in the Choi and Varian (2009b) study that support the assumptions underlying this research. First, the successful prediction of travel data by Choi and Varian (2009b) paves the way to the assumption that Google Search can be useful in predicting migration flows. Second, the query search time series has been lagged which adds the predictive power to the analysis of travel to Hong Kong. Third, the study on Hong Kong visitors use the data directly from the Google Trends open access web-based application instead of the raw search data from the Google database like in the case of the pioneering influenza study (Ginsberg et al., 2009). This supports the assumption that the combination of appropriately selected IP geolocated Google search terms could provide an evaluation of interest in migration from particular location to particular destination.

2.3. Application of IP Address Geolocation in Estimation of Migration Flows

Following some initial successes of IP address geolocation in previously mentioned fields, Zagheni and Weber (2012) attempted to estimate international migration rates with the help of geolocated e-mail messages, using a large sample of IP geolocated messages sent between 2009 and 2011 from Yahoo! e-mail accounts. The country where most of the e-mail messages were sent from by a user was considered the country of residence, and this data has been compared to national statistics on immigration rates gathered in eleven European countries. According to authors' findings, their predictions of age and gender were consistent with the archival data. Zagheni and Weber (2012) therefore suggest that geolocation of e-mail messages can be an added value to the worldwide migration statistics.

Another relevant study, instead of tracking where the e-mail messages originated, examined the locations of where the users of Yahoo! services log in during a one year period. In particular, State et al. (2012) analyzed global country-to-country flows and attempt to build a reflection of current human flows worldwide. A number of important findings have been underlined in this study, among them the similarity in patterns of short and long-term displacements and relatively strong connections between some former colonies and metropolises like Spain or England.

Whether the findings of both above mentioned studies support and open new perspectives for IP geolocation as a tool in migration flow research, they are not without disadvantages. One of the primary concerns is directly related to the data source used, which is Yahoo! database. While a set of services related and branded as Yahoo! is a popular service in the United States, it is not as common in other parts of the world, and even within the United States it is not a prevailing one. This may provide a handful of biases, including the "surprising", as proposed by State et al.

(2012), migratory connections between the majority of countries and the United States. Secondly, while the delivery of data is extremely fast and relatively egalitarian in sense of worldwide access, the data itself does not differ in its nature from other sources of information on migration flows. This is to say that a geolocated service user needs to arrive at the destination (e.g. to Spain) and log into the particular Internet service to be geo-tagged for a first time. Subsequently, the user should use it several more times during a period of time in order to be considered short-time traveler or migrant.

2.4. Application of Google Search in Description and Prediction of Real Events

The interesting supplement, but also a conceptual foundation for some of the above mentioned approaches is derived from the study dedicated to detecting the epidemics of influenza with use of Google Search *queries* (Ginsberg et al., 2009). This study developed state-level model of estimation of influenza-like-illness outbreaks in United States through the use of a five-year database of hundreds of billions of individual searches in Google Search engine. While the IP geolocation was the technological foundation of the study, the Google users that searched for the influenza related terms were not expected to possess an account on Google website. Although the disadvantage is that no knowledge about those who searched for given terms could be acquired, the simple aim of constructing a model that would instantly warn about a geographical location of epidemic influenza outbreak could be achieved with searches analyzed, which is virtually impossible to achieve through any other research method. The study based on the assumption that a person with symptoms of influenza-like-illness would start to seek out information regarding drugs and

treatment methods provided the model that is able to predict in a reliable mode the outbreaks of influenza long before the national health authorities.

The successful predictive role of the queries to Google Search in this study has been followed by the further development and improvement of Google Trends application, and the predictions of outbreaks of other illnesses like chickenpox or salmonella (Mohebbi, Vanderkam, Kodysh, Schonberger, Choi, & Kumar, 2011). Other economy- related predictions were successful, specifically those focused on the retail sales of various goods. In the perspective of this research, the relation between Google searches and unemployment claims also appear relevant (Choi and Varian, 2009b; Mohebbi et al., 2011).

Chapter III – Method: Time Series Data Analysis

3.1. Identification and Description of Datasets

There are two sources of data used in this study: the time series of popularity of selected migration-related *queries* made to Google Search acquired from Google Trends, and the Spanish Residential Variation Statistics based on Spanish Municipal Register (i.e. *padrón*). The data for both is examined in monthly intervals and covers a time span of six years from January 2005 to December 2010.

3.1.1. The Google Trends dataset.

The Google Trends is a publicly accessible tool that needs to be efficiently operated and does not provide raw levels of *queries* searched, but reports an index with data categorized at national, regional and municipal level depending on the geographic region coverage of the tool. Also, the data accessible for Google Trends users is sampled from the entire Google Search database and limited to the queries with significant volume (Choi & Varian, 2009b). Choi and Varian (2011) offer precise description of the index construction:

> The query index is based on query share: the total query volume for the search term in question within a particular geographic region divided by the total number of queries in that region during the time period being examined. The maximum query share in the time period specified is normalized to be 100 and the query share at the initial date being examined is normalized to be zero.

The analyzed *queries* are intended to be related to the elements of international migration processes and their use depends on their coverage by Google Trends

database. In particular, the selection of *queries* studied here is based on the introduced literature on international migration, and follow the logic presented in the previously introduced example of Choi and Varian (2009b). Therefore, at the preliminary stage of this research several similar *queries* e.g. *"trabajo en España", "trabajo España", "empleo en España", "visado España", "visado para España", "embajada de España", "embajada España"* have been retrieved from Google Trends database for the countries of interest in order to select the queries with the most comprehensive geographic, and longitudinal coverage (see Table 34).

This preliminary review resulted with finding several limitations that are applied to this study. First of all, the vast majority of the *queries* tested, provided outlying result for the year 2004. This can be explained with the year 2004 as being the first year for which Google Trends recollects data, and therefore some lower volume *queries* might been suffered shortcomings in coverage. Furthermore, at the 1st January 2011, there was a significant change introduced in the standards of data exploitation by the Google Trends, which may result with the discontinuity with the previous longitudinal data.[6] Based on these assumptions, the actual analysis has been conducted on a data that covers a time span of six years from January 2005 to December 2010.

Additionally, the above mentioned preliminary exploitation of *queries* regarded the geographical coverage for particular Latin American countries. The list of countries that has been checked against Google Trends data availability has been constructed from the summed *padrón* statistics of the new registrations of Spanish inhabitants that have previously resided in Spanish speaking Latin American countries (see Table 4). The data covered the period of research interest between the years 2005

[6] https://support.google.com/trends/answer/1383240?hl=en

and 2010. The results of this preliminary analysis demonstrated that, as mentioned by other authors (e.g. Choi & Varian, 2009a), Google Trends still carries out significant shortcomings, especially in relation to data provided for developing (here Latin American) countries. In order to overcome these Google Trends drawbacks three *queries* based on criteria mentioned in the previous paragraphs that are simultaneously covered by Google Trends for three large migration sending countries have been chosen for analysis.

Table 4

Enrollments of the inhabitants with previous residence in Spanish speaking Latin America (INE, 2012).

	2005	**2006**	**2007**	**2008**	**2009**	**2010**	**2005 /10**	**% of (a)**
Total	719284	840844	958266	726009	498977	464443	4207823	
Americas	215047	285527	302058	234718	131589	107986	1276925**(a)**	
Bolivia	38654	69796	46323	8835	4265	2874	170747	13,4
Colombia	21351	28650	36434	34577	19469	12622	153103	12,0
Argentina	26874	25966	23651	17786	9424	7875	111576	8,7
Peru	17563	19384	25093	27788	13249	7520	110597	8,7
Ecuador	11830	14584	25008	29743	11436	5159	97760	7,7
Paraguay	11272	19951	22366	18247	10954	9324	92114	7,2
Venezuela	15071	14452	16761	13101	9261	10827	79473	6,2
Dominican Republic	11082	13036	16730	15730	9013	6651	72242	5,7
Cuba	6090	7761	8953	9609	6383	8851	47647	3,7
Chile	7731	8959	8844	5939	3539	3220	38232	3,0
Mexico	5660	6129	6357	6363	5345	5145	34999	2,7
Uruguay	8142	8941	8560	5425	2024	1757	34849	2,7
Honduras	2335	5785	7926	4510	3519	4493	28568	2,2
Nicaragua	925	2062	3887	2735	2271	2846	14726	1,2
El Salvador	601	1053	1564	1131	799	899	6047	0,5
Guatemala	636	964	1276	1181	829	863	5749	0,5
Panama	474	697	672	582	514	464	3403	0,3
Costa Rica	338	497	642	556	409	437	2879	0,2

In particular, the *queries* will be analyzed for Colombia, Argentina, and Peru that respectively were positioned as 2nd, 3rd, and 4th at the list of Spanish-speaking Latin American countries that constituted previous residence for newly registered Spanish inhabitants during the period of interest. Concretely, these countries are reported as previous residences of almost 30% of all newly registered inhabitants with previous Latin American address, and almost 9% of all newly registered inhabitants with previous foreign address for this period (see Table 4).

In conclusion, the selection is related to the particular elements of migration processes and it additionally depends on already mentioned shortcomings of the Google Trends database's coverage. The particular *queries* selected for analysis in this study are: "*trabajo en España"* as related to economic migration process, "embajada de España" as related to Spanish (and EU) visas and entry requirement and simply "España" based on the previous findings of the similarity between tourism and migration related global network activity (State et al., 2012).

Still, one could argue that the more appropriate *queries* would be ones related to search for employment in a particular municipality. This is a legitimate request taking into account the geographic precision in the practice of employment seeking. In this case, the particular examples could be related to the Spanish gateway cities for immigrants (Colectivo Ioé., 2008), i.e. "*trabajo en Barcelona" or* "*trabajo en Madrid".* Unfortunately, Google Trends does not provide the longitudinal data for these *queries* for satisfying time range in order to carry out the comparison. This however, does not invalidate the use of the *query* "trabajo en España". There are two premises significant for this assumption. First, based on logic, says that the more general query might be used at the beginning of job-seeking process in order localize general resources (e.g. internet job boards or classified) provided at national level. In

this case, the process of geographic narrowing would be next in order. This logic is supported by statistical premise provided by Google Trends itself. The mere fact that coverage is provided for the more general *query* vs. more locally narrowed one indicates its' higher popularity in Google Search, and therefore relevance for this study.

3.1.2. Spanish Municipal Register (i.e. padrón).

According to the Spanish Law[7] "The *pardón* is an administrative register which consist of the neighbors of the municipality…" The regulation indicates that each and every individual that resides in the municipality is required to enroll under the appropriate penalties if this obligation is not fulfilled. The first and basic function of the *pardón* is to recount the number of inhabitants of each and every Spanish municipality (García-Pérez, 2007). Because of the information gathered at the moment of enrollment it also provides a basic demographic profile of the municipalities, and subsequently, *comarcas* (counties), provinces, autonomous communities, and the state. The following information is required to be provided by inhabitant during the enrollment in the *padrón* (Suero-Salamanca, 1999):

a) Name and lastname(s);
b) Sex;
c) Habitual residence;
d) Nationality;
e) Locality and date of birth;
f) National Identity Number or, in case of foreigners, a document which substitutes for it;

[7] el art. 16 de la Ley 7/1985 de 02 de abril, reguladora de las bases del Régimen Local, modificada por la Ley 4/1996 de 10 de Enero11

g) Certificate or title granted by school or higher education institution held;

h) Any other data that might be necessary in elaboration of the *padrón* (only if requesting these data complies with the Constitution);

The possibility of demographic description of each municipality has more complex implications than being solely an example of statistical inquiry. The basic economic, political, social, and administrative consequences are listed below (García-Pérez, 2007):

a) Local treasury receives an amount of state's taxes proportional to the number of inhabitants;

b) The number of municipality's councilors depends on its population;

c) The minimum obligatory services (e.g. schooling) that local authorities are required to provide to the municipality depends on its population;

d) Exceeding certain threshold of inhabitants, the administrative organization of municipality increases its complexity.

For this study in particular, it is crucial that each and every person that enrolls in the *padrón* is required to provide a location of his or her previous municipality or country of residence (Suero-Salamanca, 1999). The later allows tracking international migration flows on the basis of the *padrón* data.

Table 5

Some basic differences between the Population Census and the Municipal Register (García-Pérez, 2007)

	Municipal Register (*padrón*)	**Population Census**
Function	Administrative register	Statistic
Relevance	Continuously updated	Elaborated every ten years since 1981

Authority	Municipal Council (yearly data corrected and approved by INE)	National Statistics Institute [Instituto Nacional de Estadística: INE]
Participation	Obligatory	Voluntary

The second function of the *padrón* is for the individual to be able to demonstrate the residence in the municipality if needed. This function is met through the certifications elaborated from the *padrón* database. These are public documents with an administrative power (García-Pérez, 2007). In the perspective of the individual, being enrolled to *padrón*, gives also some particular privileges like the right to use public services present in the municipality (Suero-Salamanca, 1999).

It is important to recapitulate some of the characteristics of *padrón* that makes it a valid source of information on migration flows that is suitable for this research. In contrast to many registration systems, *padrón* rather encourages the immigrants to undergo the enrollment procedure. First of all, it is accessible for all the inhabitants with no prejudice or status check for irregular immigrants. Second, before the year 2012, the enrollment gave immediate access to free public health care regardless to immigration status. That has been currently limited in some autonomous communities due to the economic recession. Third, the enrollment might be used as further documentation of seniority of residence in order to undergo processes related to residence, nationalization or family rejoining. What an immigrant needs in order to register is a valid document (e.g. passport) and a relatively easily accessible apartment lease contract.

This research makes use of the anonymized micro-data of Residential Variation Statistics based on *padrón*, and provided by National Statistics Institute of

Spain. This data set provides information about each particular event of an individual's registration in each and every community over 10.000 inhabitants in Spain. The data for the towns and villages below 10.000 are additionally anonymized by not providing the community name but only the region name in order to ensure data anonymity.

The original daily data of individual enrollments present in the Residential Variation Statistics has been aggregated into the monthly intervals in order to facilitate comparison with the data retrieved from Google Trends web-based application. The original data set of 15,684,892 data points that represent each enrollment event that occurred in the given time span in Spain (2005: *2357656*; 2006: *2715449*; 2007: *2980684*; 2008: *2635679*; 2009: *2475632*; 2010: *2519792*) has been narrowed down to the data points representing individuals with direct previous residence in Argentina, Columbia or Peru. These data was subsequently aggregated into the monthly data for each country. The aggregate data includes 72 monthly intervals (aggregated data points) for each analyzed country.

3.2. Time Series Data Analysis Applied in this Study

In the study that analyzes the time series data, pre-whitening process should be applied before the formal statistical analysis. The Auto-Corelation Function (ACF) and Partial Auto Correlation Function (PACF) should be performed in order to assess stationarity. The first order differentiation or log transformation as appropriate should be performed in order to stationarize the time series, if necessary. This procedure is generally used if the time series data are non-stationary, thus stationarizing is done before the formal statistical test can be performed upon them. In this understanding, stationary is "random," and non-stationary is the opposite. In the case of truly random

time series, its posterior value does not depend on the precedent ones, i.e. there is no autocorrelation of the series. The most basic example is a time series that maintains a similar trend throughout the entire data set. The virtually continuous downward trend of the time series makes it non-stationary and concomitantly the general downward trend of the time series analyzed could potentially influence on a vast part of the correlation coefficient value found if the statistical test is performed upon non-stationarized times series data. This is to say that although the trend of both time series is similar, there might be no true correlation between two time series (McCleary & Hay, 1985; Tabachnick & Fidell, 2010).

In fact, the vast amount of researchers suggests a comparison of ARIMA (auto regressive, integrated, moving average) models in order to evaluate the relation and predictive potential of one time series on another (Choi & Varian 2009a; Tabachnik & Fidel, 2010). However, due to the limited scope of this study, the performed analysis followed an alternative procedure that gives more straightforwardly interpretable results although it is similar to ARIMA models in nature (McCleary & Hay, 1985; National Cooperative Highway Research Program Report, 1997). The details of the procedure are exposed in the following section.

Additionally, the aims of this particular research carry a conflict between the formal statistical testing of correlation between time series and simple indication of relation between the general trends of both time series that could be useful enough in prediction of increased interest in migration from a particular country to another. Still, the limited time span of accessible Google Trends data might result with a constant trend of both time series, which would require extreme caution with the interpretation of the non-stationary time series correlation coefficient and appeal for a construction of a more complex statistical model.

3.3. Statistical Procedure Applied in the Study

Following the indications of the previous research and literature, this study applies the Cross-Correlation Function on the pre-whitened time series data in order to assess the relation between Google Search and Residential Variation Statistics at various monthly lags of the predictor. Furthermore, regression models are examined in order to assess the predictability at the particular lags that result to be significantly correlated (Choi & Varian, 2009a; Ruiz et al., 2012; Tabachnick & Fidell, 2010).

In particular the change of the proportion in each of the Google Search *queries* has been compared to change in proportion of the number of residents enrolled in *padrón* for which the previous residence was the particular country of origin for which the *query* data were retrieved from Google Trends. That resulted with nine comparison pairs. Four stages of the analysis have been conducted on each pair of the time series:

1) The visual analysis of the Cross-Correlation Function (CCF) and sequence charts of the original non-stationary data that show the correlation between the general trends of the time series while taking into account possible lags in reaction of *padrón* to the Google Search *queries* changes. As mentioned before, their statistical significance is not formally assessed due to the possible lack of data randomness;

2) The analysis of time series' stationarity based on the visual inspection of Auto-Corelation Function (ACF) and Partial Auto Correlation Function (PACF) has been performed. If series is found to be non-stationary, it was stationarized with use of first order differentiation (that occurred enough in order to stationarize these particular time series according to the repeated PACF and ACF results). Therefore instead of the direct comparison of equivalent months**,** a comparison of the difference in the number

of immigrants enrolled in *padrón* to Google Search *queries* for the particular month (t) and the previous month (t-1) have been assessed. The transformed time series has been used in further analysis;

3) The CCF has been again performed on the stationarized time series in order to evaluate actual correlation between two time series and find the most highly correlating months' lag;

4) The lags that have been showed to be significantly correlated by the CCF and that were in accordance to the previous theoretical assumption have been assessed for the predictability with the Linear Regression model. The Durbin-Watson statistics has been additionally assessed to examine if autocorrelation of the time series has not influenced the regression results.

Chapter IV - Results

4.1. Individual Country Results: Argentina

In the case of Argentina the visual inspection of the time sequence chart (see Figure 20) showed the strong downward trend in all the time series analyzed. This indicates the similarity of the general trend over time between the predictors and the outcome variable, but also the strong non-stationarity of the time-series. CCF performed on the non-stationarized time series confirm the former that the series are highly correlated. The "*trabajo en España"* (see Figure 2 and Table 6) *and* "*embajada de España"* (see Figure 3 and Table 7) showed the strongest correlation with the Residential Variation at lag 2 of both predictors. The query "*España"* (see Figure 4 and Table 8) showed the strongest correlation with the Residential Variation at lag 0 of both predictors. Since, the ACF and PACF confirmed the series to be non-stationary, no conclusion could be inferred on the significance of the correlation coefficient for these time series before the first order differentiation that was further applied and evaluated with repeated ACF and PACF. Afterwards, the CCF performed on the stationarized time series indicated the significant correlation with the Residential Variation data on lag 4 for the "*trabajo en España"* (see Figure 5 and Table 9) and on the lag 8 for the "*embajada de España"* (see Figure 6 and Table 10). The query "*España"* did not show significant correlation with the Residential Variation data (see Figure 7 and Table 11).

4.2. Individual Country Results: Colombia

In the case of Columbia the visual inspection of the time sequence chart (see Figure 21) showed the strong downward trend in all the predictors analyzed. The

DV's trend was more diverse, but would follow the same downward trend if the lag was applied to the predictors. This again indicates strong non-stationarity of the time-series in question. CCF performed on the non-stationary time series showed that "*trabajo en España"* (see Figure 8 and Table 16), "*embajada de España"* (see Figure 9 and Table 17), and "*España"* (see Figure 10 and Table 18) *show the strongest* correlation with the DV at a lag of nine months applied to the predictors. Since, the ACF and PACF confirmed the series to be non-stationary, no conclusion could be inferred on the significance of the correlation coefficient for these time series before the first order differentiation that was further applied and evaluated with repeated ACF and PACF. Afterwards, the CCF performed on the stationarized time series again indicated the significant correlation with the Residential Variation data on lag nine for the "*trabajo en España"*(see Figure 11 and Table 19), *embajada de España"*(see Figure 12 and Table 20), and "*España"* (see Figure 13 and Table 21).

4.3. Individual Country Results: Peru

In the case of Peru the visual inspection of the time sequence chart (see Figure 22) again showed the strong downward trend in all the predictors analyzed and more diverse trend of DV that would follow the predictors' downward trend if the lag was applied to the predictors. This again indicates strong non-stationarity of the time-series in question. The CCF performed on the non-stationary predictors showed the strongest correlation with the DV at lag eight for "*trabajo en España"* (see Figure 14 and Table 28), lag nine for "*embajada de España"* (see Figure 15 and Table 29), and lag 12 for "*España"* (see Figure 16 and Table 30). Since, the ACF and PACF confirmed the series to be non-stationary, no conclusion could be inferred on the significance of the correlation coefficient for these time series before the first order

differentiation that was further applied and evaluated with repeated ACF and PACF. Afterwards, the CCF performed on the stationarized time series indicated no significant correlation for queries "*trabajo en España"* (see Figure 17) and "*España"* (see Figure 19). The significant correlation with the Residential Variation data on lag nine for the "*embajada de España"*(see Figure 18 and Table 31) has been found.

4.4. Comparative Results: Hypothesis 1

The above explained analysis of the series of CCF performed support the H_1 that the time-shifted change in proportion of migration-related queries made to Google Search in sending country and further reported by Google Trends is significantly related to changes in volume of international migration flows from selected Latin American countries to Spain. The high correlation coefficient for the non-stationary time series cannot be formally interpreted since the general trend of the time series analyzed have potentially very strong input into the correlation found. Still, in the terms of migration flows prediction, the observation of changes in general trends could become a satisfactory predictor. Nevertheless, the stationarized time series also showed significant correlations between the some of the predictors and the DV.

4.5. Comparative Results: Hypothesis 2

In order to statistically asseses the predictive power of the queries made to Google Search on the international migration flows, the linear regression has been performed on the lags of the stationarized IVs that showed to be significantly correlated according to the CCF results. In the case of Argentina, the query "*trabajo en España"* lagged by four months [R= .360, R2 = .130, adj.R2 = .116, F(1, 65) =

9.696, p = .003] (see Table 12 and Table 13 for details), and the query "*embajada de España"* lagged by eight months [R= .389, R2 = .151, adj.R2 = .137, F(1, 61) = 10.869, p = .002] (see Table 14 and Table 15 for details) appeared to significantly explain small to moderate proportion of variance in the Residential Variation. In the case of Colombia all three queries have been tested for its nine months lags have predictive power. The "*trabajo en España"* [R= .292, R2 = .085, adj.R2 = .070, F(1, 60) = 5.597, p = .021] (see Table 22 and Table 23 for details) and "*embajada de España"* [R= .348, R2 = .121, adj.R2 = .107, F(1, 60) = 8.276, p = .006] (see Table 24 and Table 25 for details) occurred to significantly explain small proportion of variance in the Residential Variation, while the "*España"* [R= .459, R2 = .211, adj.R2 = .198, F(1, 60) = 16.037, p = .000] (see Table 26 and Table 27 for details) occurred to significantly explain small to moderate proportion of variance in the Residential Variation. In the case of Peru the only tested query was "*embajada de España"* lagged by nine months in respect to the Residential Variation. The regression results showed this query to significantly explain small proportion of variance in the Residential Variation [R= .281, R2 = .079, adj.R2 = .064, F(1, 60) = 5.147, p = .027] (see Table 32 and Table 33 for details).

The above mentioned results of the regression analysis only partially support the H_2 that the time-shifted change in proportion of migration-related queries made to Google Search in sending country and further reported by Google Trends predicts changes in volume of international migration flows from selected Latin American countries to Spain. The differences of the result between country shows that a number of particular migration flow characteristics should be taken into account while building the Residential Variation prediction models based on the queries to Google Search.

Chapter V – Conclusions and Discussion

The study findings partially supported the first and second hypotheses and therefore positively evaluated the relation between the changes in Google Searches patterns and international migration flows, and additionally provided support for the assumption that Google Trends database can be used as a predictor of international migration flows at a country level.

While the data for Columbia and Peru showed the correlation of stationarized time series at lags nine and eight months respectively, the data for Argentina showed more diverse results. The further exploratory analysis of the Residential Variation datasets showed an important difference between the structure of the inhabitants that arrived from Argentina, and these that arrived from Peru and Columbia. The important proportion (up to 40 – 50%) of those that enrolled after arriving from Argentina has been shown to be in a possession of European nationality. That fact might importantly influence not only the travel and employment requirements applied to them, but also the requirements applied to the families that would further rejoin them. Concomitantly, the time patterns of lags between search for particular queries (e.g. search for "*embajada de España")* would be irrelevant to these immigrants. This condition in the case of Argentinean immigrants has already been found in previous research carried out in Spain (e.g. Moren, Mas & Wladyka, 2012). The suggestion provided here is to control for this and similar factors in further studies.

The relatively weak or at best moderate predictability of the changes in the Residential Variation could be related to the average rate of the internet penetration during the researched period. Because of the scarcity and inconsistency of data on Internet penetration from developing countries, it is difficult to test statistically this assumption. Still, some preliminary indications may be provided here as the starting

point for the future research. The predictability of results given for Peru is shown to be less powerful than these for Columbia and Argentina. Accordingly, the average Internet penetration rate for the researched period between 2005 and 2010 was the following: Argentina – 26.09%, Columbia – 21.33%, and Peru – 22.13% (The World Bank, 2013). It is worth to underline, that while there is no visible relation between the predictability found and the Internet penetration rate, the Internet penetration rates are in general concordant with the predictability found for Argentina and Columbia.

This suggestion would support the idea that the applicability of the Google Search queries for predicting the Residential Variation depends highly on Internet expansion in developing countries. On the other hand, this study shows the fragility of Google Search predictability interpretation for the extraneous factors (e.g. citizenship). Also other variables like language barriers and migration type should be considered in the further studies as stated in the previous chapters of this thesis.

As explained through the previous pages, there has been no research done on the pre-migratory web-searching habits by individuals that undertake migration processes. Therefore an interesting step in the future studies, could be an introduction of the qualitative stage, based on the semi-structured interviews, that would provide the information on what kind of information and in what ways the immigrants search on-line about their destination country. The information obtained could be used to generate potential queries that immigrants' make to Google Search in order to find information about the destination country and city.

Although many challenges in construction of international migration flows prediction model based on Google Search queries has been brought up in the previous paragraphs, the results of the current study support the assumption that data on changes in Google Search queries popularity obtained from Google Trends might

constitute a useful and at least additional tool in predicting the changes in migration flows with several months of advance. Furthermore, the development of the Google Trends, especially in the sense of coverage for developing countries and less popular queries (e.g. focused on particular municipalities), and the growth of Internet penetration rate therein, are the factors that could convert Google Trends into a multipurpose tool in estimating and predicting international migration flows.

References

Agozino, B. (2000) Methods of Data Recepcion, In. B. Agozino (Ed.), Theoretical and Methodological Issues In Migration Research (145 – 164), Aldershot

Anderson, B., & Blinder, S. (2013) Who Counts as a Migrant? Definitions and their Consequences. London: The Migration Observatory at the Univerisity of Oxford. Retrieved from: http://www.migrationobservatory.ox.ac.uk/sites/files/migobs/Briefing%20-%20Who%20Counts%20as%20a%20Migrant_0.pdf

Benson-Rea, M., & Rawlinson, S. (2003). Highly Skilled and Business Migrants: Information Processes and Settlement Outcomes. *International Migration*, 41 (2), 59-79. doi: 10.1111/1468-2435.00235

Castles S., & Miller M. J. (1993) The Age of Migration: International Population Movements in the Modern World. New York: The Guilford Press

Chen, W. (2010). Internet-Usage Patterns of Immigrants in the Process of Intercultural Adaptation. *Cyberpsychology, Behavior, And Social Networking*, 13 (4). doi: 10.1089=cyber.2009.0249

Choi H. & Varian H. (2009b). *Predicting the Present with Google Trends*. Technical report. Retrieved from Google website: http://google.com/googleblogs/pdfs/google_predicting_the_present.pdf

Choi H. & Varian H. (2011). *Predicting the Present with Google Trends*. Technical report. Retrieved from: http://people.ischool.berkeley.edu/~hal/Papers/2011/ptp.pdf

Choi H., & Varian H. (2009a). *Predicting initial claims for unemployment benefits*. Technical report. Retrieved from Google website: http://research.google.com/archive/papers/initialclaimsUS.pdf

Colectivo Ioé. (2008). *Inmigrantes, nuevos ciudadanos: ¿Hacia una España plural e intercultural?* Madrid : Fundación de las Cajas de Ahorros (FUNCAS).

Collinson, S. (1994) Europe and International Migration, London

ComScore (2011) Google Sites Accounts for 9 of 10 Searches Conducted in Latin America. Retrieved from: http://www.comscore.com/Insights/Press_Releases/2011/5/Google_Sites_Accounts_for_9_of_10_Searches_Conducted_in_Latin_America

De Tona, C., & Whelan, A. (2009). Re-mediating' the ruptures of migration: the use of internet and mobile phones in migrant women's organisations in Ireland. *Translocations: Migration and Social Change.* Retrieved from: http://ec.europa.eu/ewsi/UDRW/images/items/docl_10859_479879574.pdf

Demuth A. (2000), Some Conceptual Thoughts on Migration Research. In: B. Agozino (Ed.), Theoretical and Methodological Issues In Migration Research (22 – 26), Aldershot

Desiderio, M.V. (2012). Practices of Access to Labour Market Information for Migrants and Employers: an overview LINET 2012 Research Findings. Independent Network of Labour Migration and Integration Experts. *Retrieved from* www.labourmigration.eu

Devoto, F. (2003) *Historia de la inmigración en la Argentina*, Buenos Aires

FairSearch (2013) FairSearch Fact Sheet. Retrieved from: http://www.fairsearch.org/wp-content/uploads/2011/06/Draft-Core-FairSearch-Fact-Sheet-051812.pdf

Ferrari, L., Rosi, A., Mamei, M., & Zambonelli, F. (2011). Extracting urban patterns from location-based social networks, *Proceedings of the 3rd ACM SIGSPATIAL International Workshop on Location-Based Social Networks.* pages 9–16.

Fox, S., & Livingston, G. (2007). Latinos online: Hispanics with lower levels of education and English proficiency remain largely disconnected from the internet. *Washington, DC: Pew Hispanic Center and Pew Internet Project.* Retrieved from: http://www.pewinternet.org/pdfs/Latinos_Online_March_14_2007.pdf.

García-Pérez, Mª. S. (2007) El padrón municipal de habitantes: origen, evolución y significado. Hispania Nova. Revista de Historia Contemporánea. No 7. Retrieved from: http://hispanianova.rediris.es

Gibbons, A. (2013) Stunning Skull Gives a Fresh Portrait of Early Humans. Science 342 (6156), 297-298. doi:10.1126/science.342.6156.297

Giddens, A. (2005) *Socjologia*, Warszawa

Ginsberg, J., Mohebbi, M. H., Patel, R.S., Brammer, L., Smolinski, M.S., & Brilliant, L. (2009). Detecting influenza epidemics using search engine query data. Nature 457, 1012-1014. doi:10.1038/nature07634

Google Trends. (2012). *Analyzing Data*. [Data file]. Retrieved from Google website: http://support.google.com/trends/topic/13975?hl=en&ref_topic=13762

Gueye, B., Uhlig, S., & Fdida, S. (2007). Investigating the imprecision of IP block-based geolocation. *Proceedings of PAM 2007, Louvain-la-neuve, April, 2007* (pp. 237–240). Retrieved from http://www.nas.ewi.tudelft.nl/people/Steve/papers/Geolocation-pam07.pdf

Hamel, J.Y. (2009). Information and Communication Technologies and Migration. *Human Development Research Paper 39*. New York: United Nations Development Programme, Human Development Report Office. Retrieved from: http://hdr.undp.org/en/reports/global/hdr2009/papers/HDRP_2009_39.pdf

Hui, P., Mortier, R., Piorkowski, M., Henderson, T., & Crowcoft, J. (2010). Planet-scale human mobility measurement. *Proceedings of the Second ACM International Workshop on Hot Topics in Planet-Scale Measurement (HotPlanet).* New York: ACM, doi: http://dx.doi.org/10.1145/1834616.1834618

Instituto Nacional de Estadística. (2012). Data extracted from the census database for 1st of January of 2012. Retrieved from: http://www.ine.es

International Organization for Migration (2010) World Migration Report 2010. The Future of Migration: Building Capacities for Change. Retrieved from: http://publications.iom.int/bookstore/index.php?main_page=redirect&action=url&goto=publications.iom.int%2Fbookstore%2Ffree%2FWMR_2010_ENGLISH.pdf

International Organization for Migration (2013) World Migration Report 2013 – Migrant Well-being and Development. Retrieved from: http://publications.iom.int/bookstore/index.php?main_page=redirect&action=url&goto=publications.iom.int%2Fbookstore%2Ffree%2FWMR2013_EN.pdf

King, K. F. (2010) Geolocation and Federalism on the Internet: Cutting Internet Gambling's Gordian Knot. The Columbia Science and Technology Law Review 11(41). Retrieved from: http://www.stlr.org/cite.cgi?volume=11&article=2.

Kissau, K. (2009). Online spheres of migrants in Germany: How and why migrants use the Internet. In C. Dietz, & P. Stammen. *Media on the move: Migrants and minorities and the media* (pp. 80-84). 4th Symposioum on Forum Media and Development, Bonn 2008, Aachen Germany, Catholic Media Council (CAMECO). Retrieved from: http://www.cameco.org/files/mediaonthemove-kissau.pdf

Krajowy Punkt Kontaktowy Europejskiej Sieci Migracyjnej (2011) Migracja tymczasowa i cyrkulacyjna w Polsce: dotychczasowe doświadczenia, uregulowania prawne i opcje na przyszłość Lata 2004-2009. Retrieved from: http://ec.europa.eu/dgs/home-affairs/what-we-do/networks/european_migration_network/reports/docs/emn-studies/circular-migration/pl_study_on_temporary_and_circular_migration_pl_version_pl.pdf

Kupiszewska, D. & Nowok, B. (2008). Comparability of statistics on international migration flows in the European Union. In: J. Raymer, F. Willekens (Eds.), *International Migration in Europe: Data, Models and Estimates* (41-71). Chichester: John Wiley and Sons.

Lordkipanidze, D., Ponce de León, M. S., Margvelashvili A., Rak, Y., Rightmire, G. P., Vekua, A., & Zollikofer, C. P. E. (2013) A Complete Skull from Dmanisi, Georgia, and the Evolutionary Biology of Early Homo. Science 342 (6156), 326-331. dos:10.1126/science.1238484

Manning, P. (2005) *Migration in World History*, New York

McCleary, R. & Hay, R. (1980). Applied Time Series Analysis for the Social Sciences. London: Sage Publications.

Ministerstwo Spraw Zagranicznych (30 April 2013) Społeczno-gospodarcze efekty członkostwa Polski w Unii Europejskiej (1 maja 2004 – 1 maja 2013). Główne wnioski w związku z dziewiątą rocznicą przystąpienia Polski do UE. Retrieved from: http://www.msz.gov.pl/resource/94e4e616-9554-4c76-af2e-6615e3196f83:JCR

Mohebbi, M., Vanderkam, D., Kodysh, J., Schonberger, R., Choi, H., & Kumar, S. (2011). *Google Correlate Whitepaper.* Technical report. Retrieved from Google website: https://www.google.com/trends/correlate/whitepaper.pdf.

Morén-Alegret, R., Mas, A., Wladyka, D. (2012). Concordia Discors Understanding Conflict and Integration Outcomes of Inter-group Relations and Integration Policies in Selected Neighbourhoods of Five European Cities. Barcelona Final Report. Retrieved from European Commision website: http://www.concordiadiscors.eu/assets/Barcelona-final-report.pdf

Moreno, F.J (2004) The Evolution of immigration policies in Spain. Between external constraints and domestic demand for unskilled Labour, Barcelona

Moses J. W. (2012). EMIG 1.2: A Global Time Series of Annual Emigration Flows. *International Migration*.doi: 10.1111/imig.12026

National Cooperative Highway Research Program Report. (1997). A guidebook for forecasting Freight Transportation Demand (Report No. 388) National Cooperative Highway Research Program Report) Washington, D.C: Transportation Research Board

Nowaczek, K. (2004) Polityka Unii Europejskiej wobec procesów imigracyjnych, Toruń

Orozco, M., Burgess, E., & Ascoli, N. (2010). Is there a match among migrants, remittances and technology*? Inter-American Dialogue,* September 2010. Retrieved from http://www.thedialogue.org/PublicationFiles/a%20match%20in%20migrants%20remittances%20and%20technology%20MO_FINAL_11.4.101.pdf

Ruiz, E., Hristidis, V., Castillo, C., Gionis, A., Jaimes, A. (2012). Correlating Financial Time Series with Micro-Blogging Activity. *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (pp. 513-522). doi:10.1145/2124295.2124358

Schrover, M. (1995a) *Farmers in Europe: 8.000 B.C. - 800 B.C.*, History of International Migration, Universiteit Leiden. Retrieved from: http://www.let.leidenuniv.nl/history/migration/chapter112.html

Schrover, M. (1995b) *First people in Europe: 700.000 B.C. - 8.000 B.C.*, History of International Migration, Universiteit Leiden. Retrieved from: http://www.let.leidenuniv.nl/history/migration/chapter111.html

Solomon, H. (2003) *Of myths and migration*, Unisa

State, B., Weber I., & Zagheni E. (2012). Using IP Address Geolocation to Analyze International Migration and Mobility Patterns, *Proceedings of WSDM 2013* (pp. 265-274*)*. doi: 10.1145/2433396.2433432

Suero-Salamanca, J. A. (1999) Estudio sobre el padrón municipal de habitantes. Actualidad Administrativa. No 15 (12). Retrieved from: http://hispanianova.rediris.es/7/articulos/7a005.pdf

Svantesson, D. J. B. (2004) Geo-location technologies and other means of placing borders on the 'borderless' Internet. Retrieved from: http://epublications.bond.edu.au/law_pubs/63

Tabachnick, B. & Fidell, L. (2010). *Using Multivariate Statistics* (5th ed.). Boston: Pearson Education.

The World Bank (2013) *Internet users (per 100 people).* International Telecommunication Union, World Telecommunication/ICT Development Report and database, and World Bank estimates. Retrieved from: http://data.worldbank.org/indicator/IT.NET.USER.P2?page=1

Zagheni, E., & Weber, I. (2012). You are where you E-mail: Using E-mail Data to Estimate International Migration Rates. *Proceedings of the 3rd Annual ACM Web Science Conference, New York, NY (USA)* (pp. 348-351). doi: 10.1145/2380718.2380764

Zolberg, A.R. (2005) Patterns of International Migration Policy: A Diachronic Comparision. In: A.M. Messina, G. Lahav (Eds.), The Migration Reader: Exploring Politics and Policies (110-118). London

Index of Tables

Index of Figures

Appendixes

Table 6

Cross Correlations between the query "trabajo en España" and Residential Variation for Argentina.

Cross Correlations

Series Pair:ARG_TeE with ARG_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	.406	.129
-11	.460	.128
-10	.500	.127
-9	.521	.126
-8	.519	.125
-7	.521	.124
-6	.519	.123
-5	.568	.122
-4	.600	.121
-3	.618	.120
-2	.648	.120
-1	.708	.119
0	.795	.118
1	.834	.119
2	.849	.120
3	.816	.120
4	.769	.121
5	.646	.122
6	.569	.123
7	.572	.124
8	.516	.125
9	.483	.126
10	.446	.127
11	.450	.128
12	.466	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 7

Cross Correlations between the query "embajada de España" and Residential Variation for Argentina.

Cross Correlations

Series Pair:ARG_EdE with ARG_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	.330	.129
-11	.353	.128
-10	.379	.127
-9	.410	.126
-8	.440	.125
-7	.452	.124
-6	.453	.123
-5	.495	.122
-4	.519	.121
-3	.542	.120
-2	.595	.120
-1	.670	.119
0	.732	.118
1	.759	.119
2	.771	.120
3	.758	.120
4	.700	.121
5	.602	.122
6	.534	.123
7	.520	.124
8	.509	.125
9	.453	.126
10	.436	.127
11	.464	.128
12	.492	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 8

Cross Correlations between the query "España" and Residential Variation for Argentina.

Cross Correlations

Series Pair:ARG_Esp with ARG_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	.254	.129
-11	.219	.128
-10	.248	.127
-9	.223	.126
-8	.215	.125
-7	.197	.124
-6	.217	.123
-5	.286	.122
-4	.355	.121
-3	.439	.120
-2	.488	.120
-1	.610	.119
0	.661	.118
1	.617	.119
2	.594	.120
3	.535	.120
4	.524	.121
5	.423	.122
6	.425	.123
7	.509	.124
8	.506	.125
9	.518	.126
10	.528	.127
11	.580	.128
12	.588	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 9

Cross Correlations between stationarized time series of the query "trabajo en España" and Residential Variation for Argentina.

Cross Correlations

Series Pair:ARG_TeE with ARG_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.062	.130
-11	.079	.129
-10	.074	.128
-9	.135	.127
-8	-.019	.126
-7	-.001	.125
-6	-.243	.124
-5	.071	.123
-4	.094	.122
-3	-.040	.121
-2	-.121	.120
-1	-.175	.120
0	-.018	.119
1	.033	.120
2	.235	.120
3	-.013	.121
4	.341	.122
5	-.218	.123
6	-.213	.124
7	.217	.125
8	-.032	.126
9	-.002	.127
10	-.094	.128
11	-.045	.129
12	-.038	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 10

Cross Correlations between stationarized time series of the query "embajada de España" and Residential Variation for Argentina.

Cross Correlations

Series Pair:ARG_EdE with ARG_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	.150	.130
-11	-.014	.129
-10	-.039	.128
-9	.023	.127
-8	.025	.126
-7	.091	.125
-6	-.185	.124
-5	.060	.123
-4	.022	.122
-3	-.118	.121
-2	-.059	.120
-1	.040	.120
0	-.084	.119
1	-.003	.120
2	.015	.120
3	.146	.121
4	.118	.122
5	-.096	.123
6	-.049	.124
7	-.101	.125
8	.327	.126
9	-.177	.127
10	-.145	.128
11	.081	.129
12	-.094	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 11

Cross Correlations between stationarized time series of the query "España" and Residential Variation for Argentina.

Cross Correlations

Series Pair:ARG_Esp with ARG_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	.236	.130
-11	-.214	.129
-10	.171	.128
-9	-.048	.127
-8	.014	.126
-7	-.135	.125
-6	-.140	.124
-5	-.001	.123
-4	-.046	.122
-3	.118	.121
-2	-.232	.120
-1	.242	.120
0	.153	.119
1	-.089	.120
2	.081	.120
3	-.087	.121
4	.167	.122
5	-.100	.123
6	-.139	.124
7	.031	.125
8	-.024	.126
9	-.010	.127
10	-.125	.128
11	.151	.129
12	.067	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 12

Regression Model Summary for stationarized time series of the query "trabajo en España" lagged by four months and Residential Variation for Argentina.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.360[a]	.130	.116	269.00857	2.026

a. Predictors: (Constant), LAGS(ARG_TeE_1_diff,4)

b. Dependent Variable: DIFF(ARG_RAW_PV,1)

Table 13

Regression Coefficients for stationarized time series of the query "trabajo en España" lagged by four months and Residential Variation for Argentina.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-8.086	33.076		-.244	.808
	LAGS(ARG_TeE_1_diff,4)	9.980	3.205	.360	3.114	.003

a. Dependent Variable: DIFF(ARG_RAW_PV,1)

Table 14

Regression Model Summary for stationarized time series of the query "embajada de España" lagged by eight months and Residential Variation for Argentina.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.389[a]	.151	.137	265.80852	1.972

a. Predictors: (Constant), LAGS(ARG_EdE_diff,8)

b. Dependent Variable: DIFF(ARG_RAW_PV,1)

Table 15

Regression Coefficients for stationarized time series of the query "embajada de España" lagged by eight months and Residential Variation for Argentina.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	-8.609	33.657		-.256	.799
	LAGS(ARG_EdE_diff,8)	11.459	3.476	.389	3.297	.002

a. Dependent Variable: DIFF(ARG_RAW_PV,1)

Table 16

Cross Correlations between the query "trabajo en España" and Residential Variation for Colombia.

Cross Correlations

Series Pair:COL_TeE with COL_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.173	.129
-11	-.098	.128
-10	-.038	.127
-9	-.023	.126
-8	.002	.125
-7	.016	.124
-6	.033	.123
-5	.116	.122
-4	.170	.121
-3	.256	.120
-2	.265	.120
-1	.273	.119
0	.337	.118
1	.378	.119
2	.456	.120
3	.431	.120
4	.456	.121
5	.430	.122
6	.416	.123
7	.488	.124
8	.486	.125
9	.581	.126
10	.546	.127
11	.452	.128
12	.484	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 17

Cross Correlations between the query "embajada de España" and Residential Variation for Colombia.

Cross Correlations

Series Pair:COL_EdE with COL_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.222	.129
-11	-.169	.128
-10	-.122	.127
-9	-.106	.126
-8	-.081	.125
-7	-.069	.124
-6	-.066	.123
-5	-.016	.122
-4	.037	.121
-3	.101	.120
-2	.099	.120
-1	.080	.119
0	.103	.118
1	.171	.119
2	.220	.120
3	.227	.120
4	.244	.121
5	.226	.122
6	.215	.123
7	.259	.124
8	.308	.125
9	.412	.126
10	.402	.127
11	.353	.128
12	.396	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 18

Cross Correlations between the query "España" and Residential Variation for Colombia.

Cross Correlations

Series Pair:COL_Esp with COL_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.265	.129
-11	-.216	.128
-10	-.178	.127
-9	-.158	.126
-8	-.143	.125
-7	-.138	.124
-6	-.125	.123
-5	-.073	.122
-4	-.006	.121
-3	.052	.120
-2	.043	.120
-1	.046	.119
0	.102	.118
1	.168	.119
2	.210	.120
3	.219	.120
4	.247	.121
5	.239	.122
6	.283	.123
7	.342	.124
8	.398	.125
9	.489	.126
10	.471	.127
11	.442	.128
12	.484	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 19

Cross Correlations between stationarized time series of the query "trabajo en España" and Residential Variation for Colombia.

Cross Correlations

Series Pair:COL_TeE with COL_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.071	.130
-11	.031	.129
-10	.105	.128
-9	-.014	.127
-8	.022	.126
-7	-.003	.125
-6	-.160	.124
-5	.078	.123
-4	-.097	.122
-3	.161	.121
-2	.025	.120
-1	-.136	.120
0	.019	.119
1	-.108	.120
2	.240	.120
3	-.110	.121
4	.127	.122
5	-.021	.123
6	-.192	.124
7	.159	.125
8	-.223	.126
9	.288	.127
10	.142	.128
11	-.268	.129
12	.109	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 20

Cross Correlations between stationarized time series of the query "embajada de España" and Residential Variation for Colombia.

Cross Correlations

Series Pair:COL_EdE with COL_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.092	.130
-11	.027	.129
-10	.112	.128
-9	-.032	.127
-8	.053	.126
-7	.003	.125
-6	-.168	.124
-5	.015	.123
-4	-.047	.122
-3	.198	.121
-2	.090	.120
-1	-.232	.120
0	-.118	.119
1	-.022	.120
2	.089	.120
3	.016	.121
4	.154	.122
5	.027	.123
6	-.151	.124
7	-.074	.125
8	-.203	.126
9	.342	.127
10	.168	.128
11	-.252	.129
12	-.040	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 21

Cross Correlations between stationarized time series of the query "España" and Residential Variation for Colombia.

Cross Correlations

Series Pair:COL_Esp with COL_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.011	.130
-11	.043	.129
-10	.069	.128
-9	.031	.127
-8	.042	.126
-7	-.045	.125
-6	-.170	.124
-5	-.042	.123
-4	.036	.122
-3	.235	.121
-2	-.024	.120
-1	-.251	.120
0	-.053	.119
1	.035	.120
2	.084	.120
3	.007	.121
4	.130	.122
5	-.033	.123
6	-.099	.124
7	-.096	.125
8	-.135	.126
9	.368	.127
10	.086	.128
11	-.242	.129
12	-.037	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 22

Regression Model Summary for stationarized time series of the query "trabajo en España" lagged by nine months and Residential Variation for Colombia.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.292[a]	.085	.070	469.37008	2.570

a. Predictors: (Constant), LAGS(COL_TeE_diff,9)

b. Dependent Variable: DIFF(COL_RAW_PV,1)

Table 23

Regression Coefficients for stationarized time series of the query "trabajo en España" lagged by nine months and Residential Variation for Colombia.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	-3.797	59.681		-.064	.949
	LAGS(COL_TeE_diff,9)	9.693	4.097	.292	2.366	.021

a. Dependent Variable: DIFF(COL_RAW_PV,1)

Table 24

Regression Model Summary for stationarized time series of the query "embajada de España" lagged by nine months and Residential Variation for Colombia.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.348[a]	.121	.107	460.07106	2.557

a. Predictors: (Constant), LAGS(COL_EdE_diff,9)

b. Dependent Variable: DIFF(COL_RAW_PV,1)

Table 25

Regression Coefficients for stationarized time series of the query "embajada de España" lagged by nine months and Residential Variation for Colombia.

Coefficients[a]

Model		Unstandardized Coefficients B	Unstandardized Coefficients Std. Error	Standardized Coefficients Beta	t	Sig.
1	(Constant)	16.650	59.196		.281	.779
	LAGS(COL_EdE_diff,9)	18.415	6.401	.348	2.877	.006

a. Dependent Variable: DIFF(COL_RAW_PV,1)

Table 26

Regression Model Summary for stationarized time series of the query "España" lagged by nine months and Residential Variation for Colombia.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.459[a]	.211	.198	435.95864	2.657

a. Predictors: (Constant), LAGS(COL_Esp_diff,9)

b. Dependent Variable: DIFF(COL_RAW_PV,1)

Table 27

Regression Coefficients for stationarized time series of the query "España" lagged by nine months and Residential Variation for Colombia.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	26.029	56.120		.464	.644
	LAGS(COL_Esp_diff,9)	33.467	8.357	.459	4.005	.000

a. Dependent Variable: DIFF(COL_RAW_PV,1)

Table 28

Cross Correlations between the query "trabajo en España" and Residential Variation for Peru.

Cross Correlations

Series Pair:PER_TeE with PER_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.148	.129
-11	-.130	.128
-10	-.091	.127
-9	-.031	.126
-8	.009	.125
-7	.027	.124
-6	.051	.123
-5	.141	.122
-4	.174	.121
-3	.212	.120
-2	.233	.120
-1	.264	.119
0	.291	.118
1	.322	.119
2	.410	.120
3	.458	.120
4	.468	.121
5	.416	.122
6	.431	.123
7	.510	.124
8	.537	.125
9	.535	.126
10	.511	.127
11	.504	.128
12	.480	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 29

Cross Correlations between the query "embajada de España" and Residential Variation for Peru.

Cross Correlations

Series Pair:PER_EdE with PER_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.157	.129
-11	-.112	.128
-10	-.081	.127
-9	-.046	.126
-8	-.004	.125
-7	.012	.124
-6	.027	.123
-5	.065	.122
-4	.116	.121
-3	.179	.120
-2	.183	.120
-1	.206	.119
0	.265	.118
1	.312	.119
2	.357	.120
3	.396	.120
4	.423	.121
5	.410	.122
6	.379	.123
7	.384	.124
8	.429	.125
9	.455	.126
10	.413	.127
11	.427	.128
12	.461	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 30

Cross Correlations between the query "España" and Residential Variation for Peru.

Cross Correlations

Series Pair:PER_Esp with PER_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	-.172	.129
-11	-.147	.128
-10	-.133	.127
-9	-.109	.126
-8	-.080	.125
-7	-.074	.124
-6	-.070	.123
-5	-.025	.122
-4	.048	.121
-3	.082	.120
-2	.089	.120
-1	.129	.119
0	.175	.118
1	.223	.119
2	.258	.120
3	.259	.120
4	.292	.121
5	.278	.122
6	.293	.123
7	.338	.124
8	.400	.125
9	.435	.126
10	.412	.127
11	.429	.128
12	.451	.129

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 31

Cross Correlations between stationarized time series of the query "embajada de España" and Residential Variation for Peru.

Cross Correlations

Series Pair:PER_EdE with PER_RAW_PV

Lag	Cross Correlation	Std. Error[a]
-12	.060	.130
-11	.032	.129
-10	.007	.128
-9	-.018	.127
-8	.106	.126
-7	-.004	.125
-6	-.111	.124
-5	-.045	.123
-4	-.085	.122
-3	.240	.121
-2	-.110	.120
-1	-.161	.120
0	.025	.119
1	-.030	.120
2	-.073	.120
3	.099	.121
4	.131	.122
5	.105	.123
6	-.161	.124
7	-.153	.125
8	.096	.126
9	.275	.127
10	-.190	.128
11	-.044	.129
12	.093	.130

a. Based on the assumption that the series are not cross correlated and that one of the series is white noise.

Table 32

Regression Model Summary for stationarized time series of the query "trabajo en España" lagged by nine months and Residential Variation for Peru.

Model Summary[b]

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin-Watson
1	.281[a]	.079	.064	409.49904	2.111

a. Predictors: (Constant), LAGS(PER_EdE_diff,9)

b. Dependent Variable: DIFF(PER_RAW_PV,1)

Table 33

Regression Coefficients for stationarized time series of the query "trabajo en España" lagged by nine months and Residential Variation for Peru.

Coefficients[a]

Model		Unstandardized Coefficients		Standardized Coefficients		
		B	Std. Error	Beta	t	Sig.
1	(Constant)	9.626	52.688		.183	.856
	LAGS(PER_EdE_diff,9)	16.274	7.173	.281	2.269	.027

a. Dependent Variable: DIFF(PER_RAW_PV,1)

Table 34

Queries preliminarily tested in Google Trends for longitudinal and geographical data accessibility.

Original query in Spanish	**Literal translation into English**
“apartamento Barcelona”	“apartment Barcelona”
“apartamento en Barcelona”	“apartment in Barcelona”
“apartamento en España”	“apartment in Spain”
“apartamento en Madrid”	“apartment in Madrid”
“apartamento España”	“apartment Spain”
“apartamento Madrid”	“apartment Madrid”
“apartamentos Barcelona”	“apartments Barcelona”
“apartamentos en Barcelona”	“apartments in Barcelona”
“apartamentos en España”	“apartments in Spain”
“apartamentos en Madrid”	“apartments in Madrid”
“apartamentos España”	“apartments Spain”
“apartamentos Madrid”	“apartments Madrid”
“bolsa de empleo Barcelona”	“employment exchange Barcelona”
“bolsa de empleo España”	“employment exchange Spain”
“bolsa de empleo Madrid”	“employment exchange Madrid”
“bolsa de trabajo Barcelona”	“labor exchange Barcelona”
“bolsa de trabajo España”	“labor exchange Spain”
“bolsa de trabajo Madrid”	“labor exchange Madrid”
“bolsa empleo en Barcelona”	“employment exchange in Barcelona”
“bolsa empleo en España”	“employment exchange in Spain”
“bolsa empleo en Madrid”	“employment exchange in Madrid”
“bolsa trabajo en Barcelona”	“labor exchange in Barcelona”
“bolsa trabajo en España”	“labor exchange in Spain”
“bolsa trabajo en Madrid”	“labor exchange in Madrid”
“consulado de España”	“consulate of Spain”
“consulado España”	“consulate Spain”
“cuarto Barcelona”	“room Barcelona”

“cuarto en Barcelona”	“room in Barcelona”
“cuarto en España”	“room in Spain”
“cuarto en Madrid”	“room in Madrid”
“cuarto España”	“room Spain”
“cuarto Madrid”	“room Madrid”
“cuartos Barcelona”	“rooms Barcelona”
“cuartos en Barcelona”	“rooms in Barcelona”
“cuartos en España”	“rooms in Spain”
“cuartos en Madrid”	“rooms in Madrid”
“cuartos España”	“rooms Spain”
“cuartos Madrid”	“rooms Madrid”
“departamento Barcelona”	“apartment Barcelona”
“departamento en Barcelona”	“apartment in Barcelona”
“departamento en España”	“apartment in Spain”
“departamento en Madrid”	“apartment in Madrid”
“departamento España”	“apartment Spain”
“departamento Madrid”	“apartment Madrid”
“departamentos Barcelona”	“apartments Barcelona”
“departamentos en Barcelona”	“apartments in Barcelona”
“departamentos en España”	“apartments in Spain”
“departamentos en Madrid”	“apartments in Madrid”
“departamentos España”	“apartments Spain”
“departamentos Madrid”	“apartments Madrid”
“embajada de España”	“embassy of Spain”
“embajada España”	“embassy Spain”
“empleo Barcelona”	“employment Barcelona”
“empleo en Barcelona”	“employment in Barcelona”
“empleo en España”	“employment in Spain”
“empleo en Madrid”	“employment in Madrid”
“empleo España”	“employment Spain”
“empleo Madrid”	“employment Madrid”
“España”	“Spain”
“habitación Barcelona”	“room Barcelona”
“habitación en Barcelona”	“room in Barcelona”

“habitación en España”	“room in Spain”
“habitación en Madrid”	“room in Madrid”
“habitación España”	“room Spain”
“habitación Madrid”	“room Madrid”
“habitaciones Barcelona”	“rooms Barcelona”
“habitaciones en Barcelona”	“rooms in Barcelona”
“habitaciones en España”	“rooms in Spain”
“habitaciones en Madrid”	“rooms in Madrid”
“habitaciones España”	“rooms Spain”
“habitaciones Madrid”	“rooms Madrid”
“piso Barcelona”	“apartment Barcelona”
“piso en Barcelona”	“apartment in Barcelona”
“piso en España”	“apartment in Spain”
“piso en Madrid”	“apartment in Madrid”
“piso España”	“apartment Spain”
“piso Madrid”	“apartment Madrid”
“pisos Barcelona”	“apartments Barcelona”
“pisos en España”	“apartments in Spain”
“pisos en Madrid”	“apartments in Madrid”
“pisos en Madrid”	“apartments in Madrid”
“pisos España”	“apartments Spain”
“pisos Madrid”	“apartments Madrid”
“trabajo Barcelona”	“job Barcelona”
“trabajo en Barcelona”	“job in Barcelona”
“trabajo en España”	“job in Spain”
“trabajo en Madrid”	“job in Madrid”
“trabajo España”	“job Spain”
“trabajo Madrid”	“job Madrid”
“visa España”	“visa Spain”
“visa para España”	“visa for Spain”
“visado España”	“visa Spain”
“visado para España”	“visa for Spain”

Figure 2

Cross Correlations between the query "trabajo en España" and Residential Variation for Argentina.

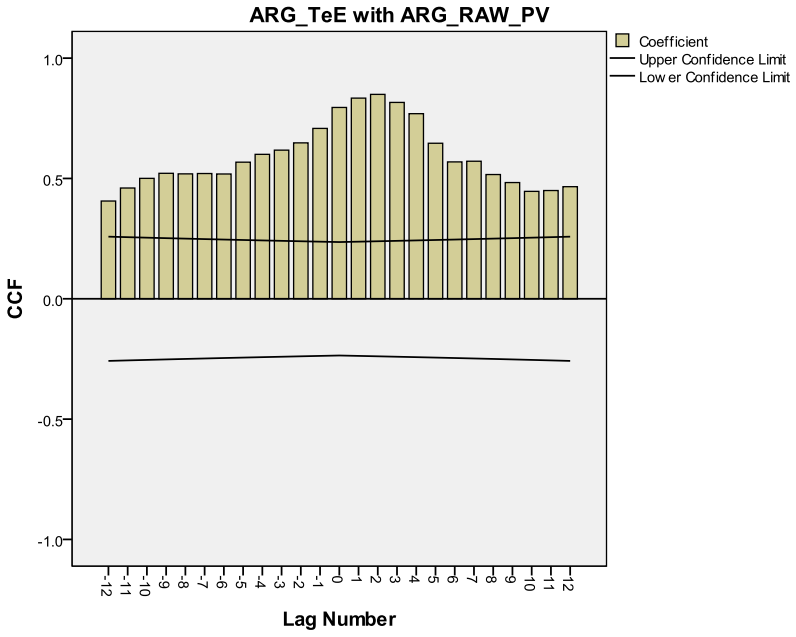


Figure 3

Cross Correlations between the query "embajada de España" and Residential Variation for Argentina.

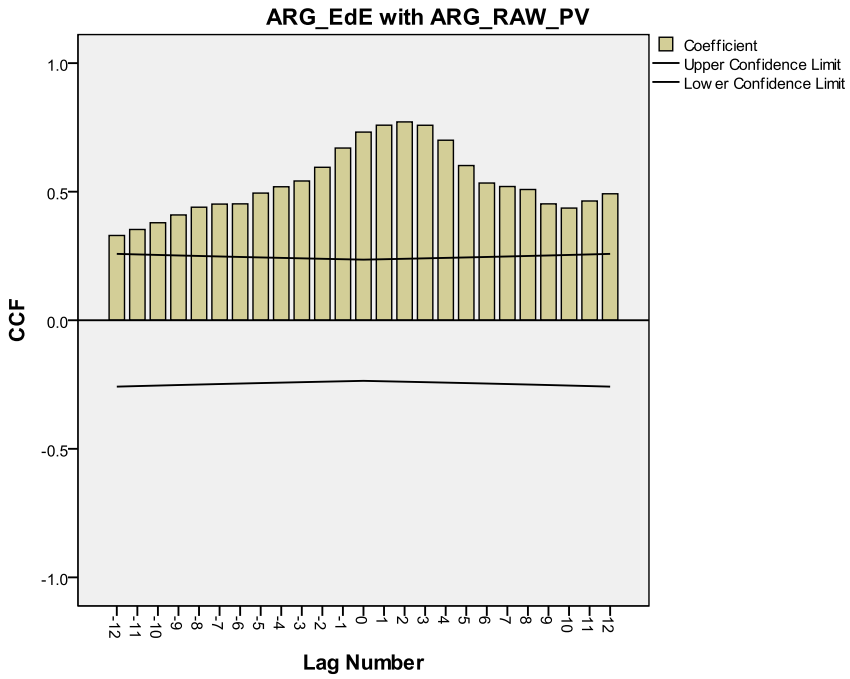


Figure 4

Cross Correlations between the query "España" and Residential Variation for Argentina.

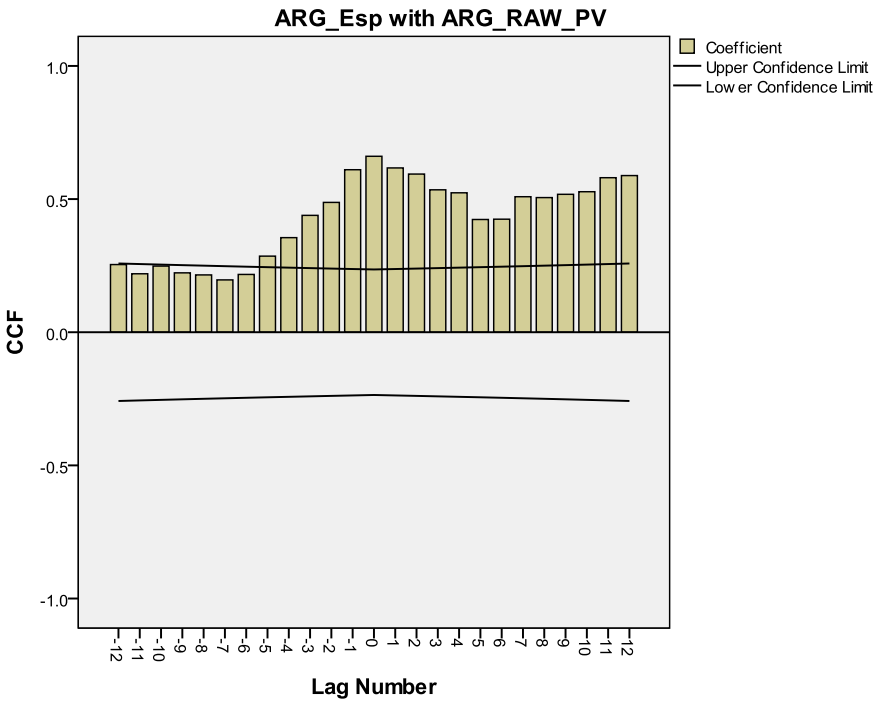


Figure 5

Cross Correlations between the stationarized time series of the query "trabajo en España" and Residential Variation for Argentina.

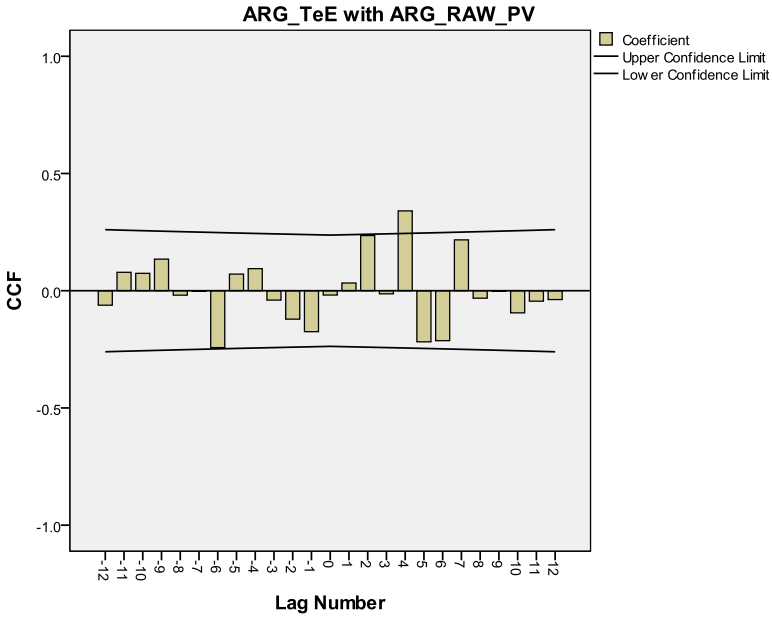


Figure 6

Cross Correlations between the stationarized time series of the query "embajada de España" and Residential Variation for Argentina.

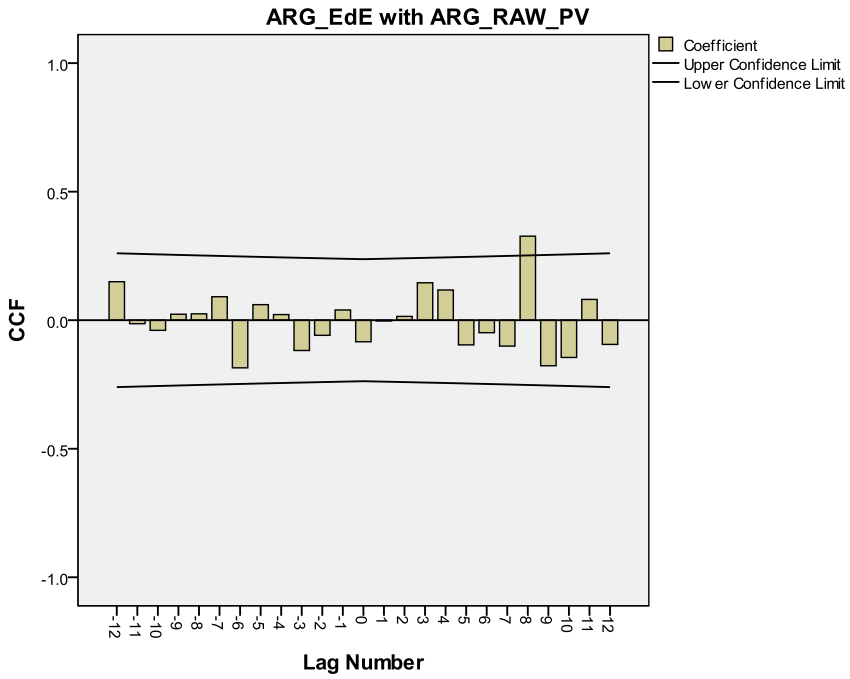


Figure 7

Cross Correlations between the stationarized time series of the query "España" and Residential Variation for Argentina.

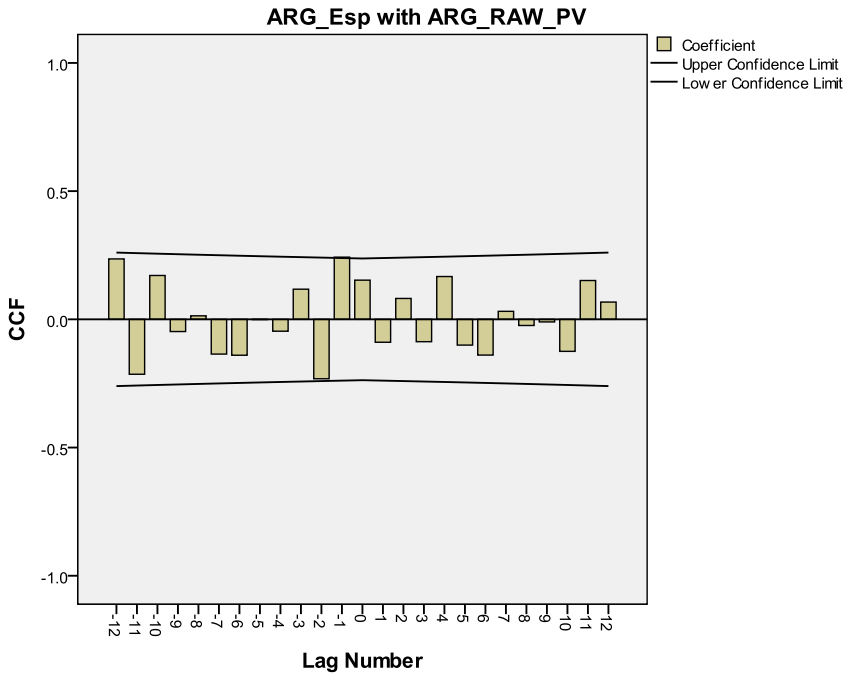


Figure 8

Cross Correlations between the query "trabajo en España" and Residential Variation for Colombia.

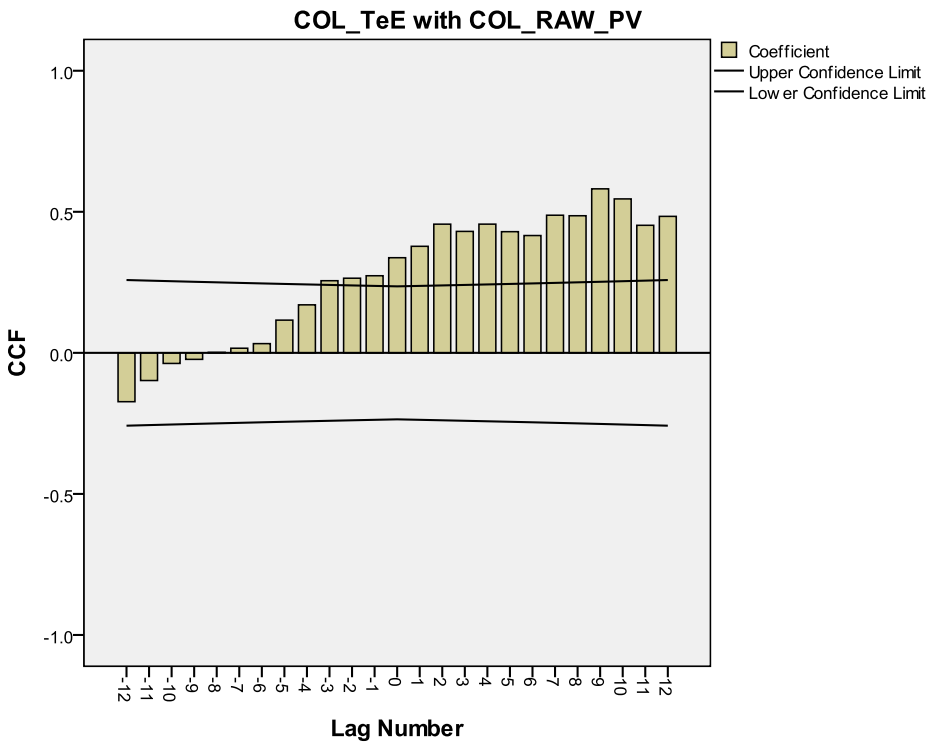


Figure 9

Cross Correlations between the query "embajada de España" and Residential Variation for Colombia.

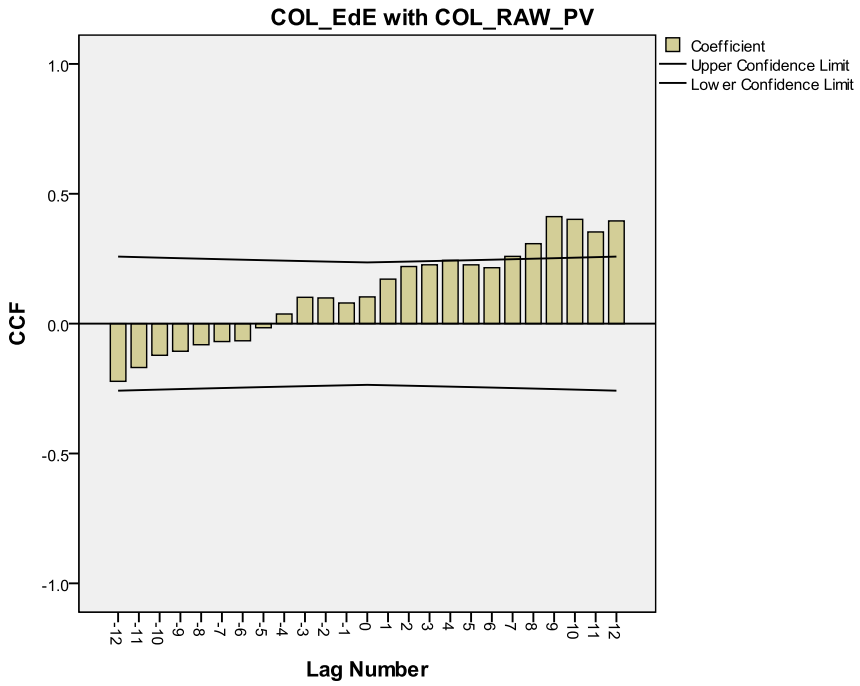


Figure 10

Cross Correlations between the query "España" and Residential Variation for Colombia.

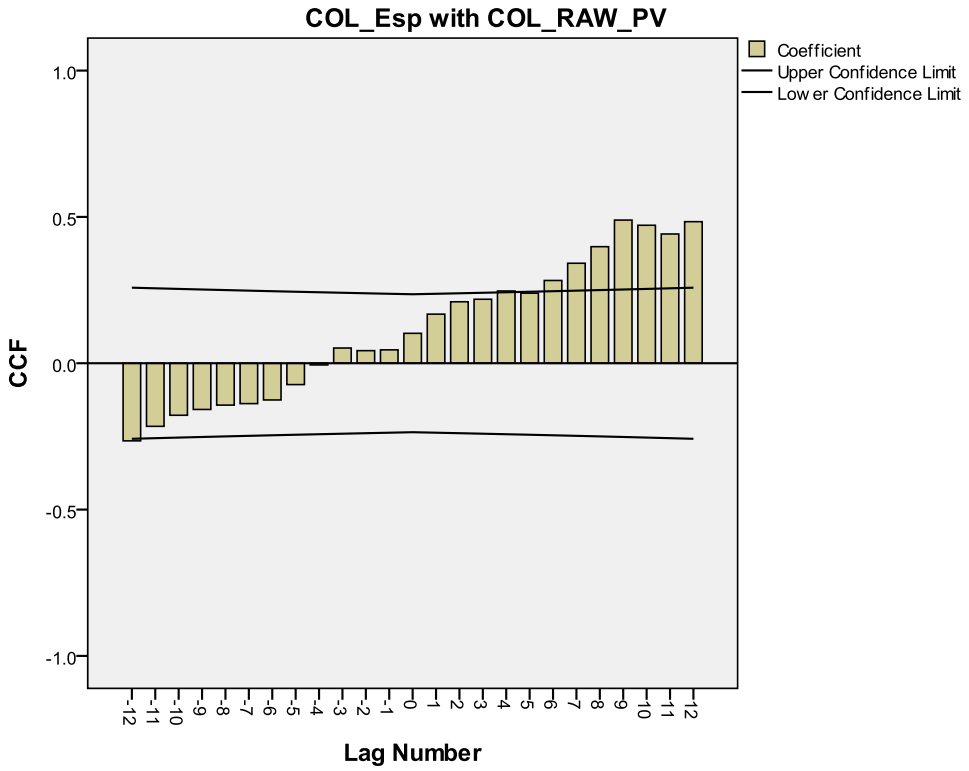


Figure 11

Cross Correlations between the stationarized time series of the query "trabajo en España" and Residential Variation for Colombia.

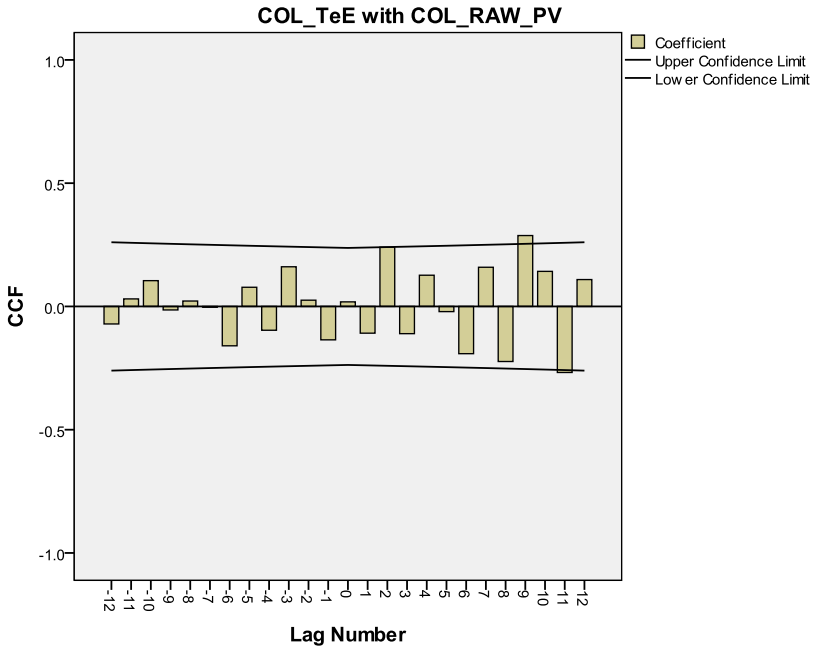


Figure 12

Cross Correlations between the stationarized time series of the query "embajada de España" and Residential Variation for Colombia.

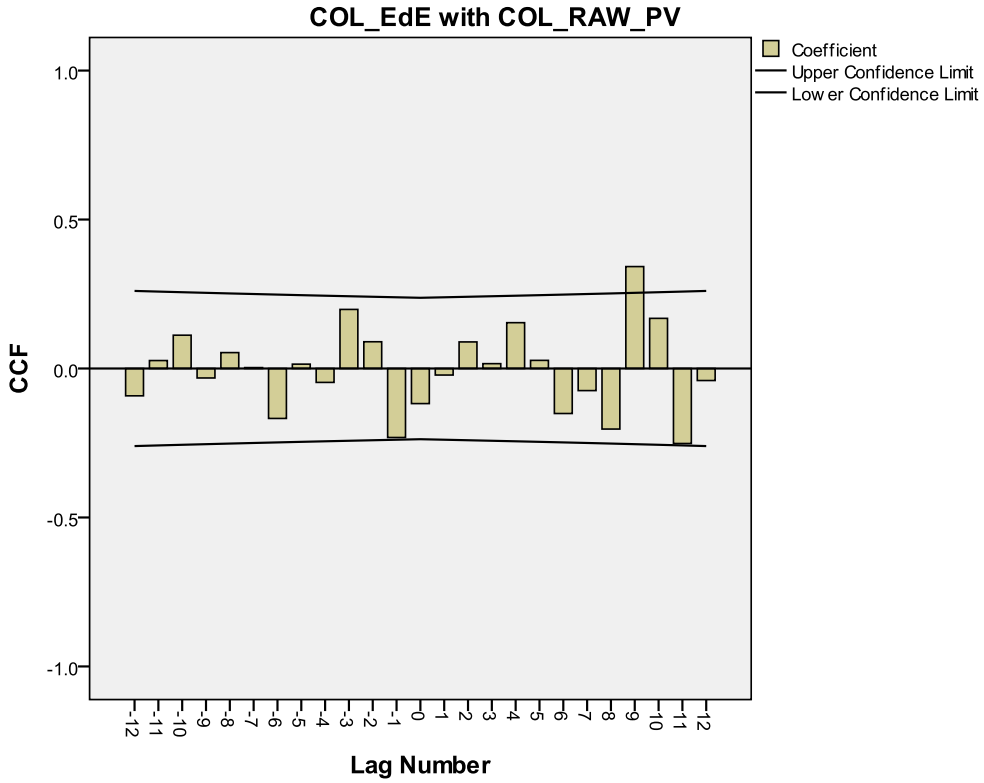


Figure 13

Cross Correlations between the stationarized time series of the query "España" and Residential Variation for Colombia.

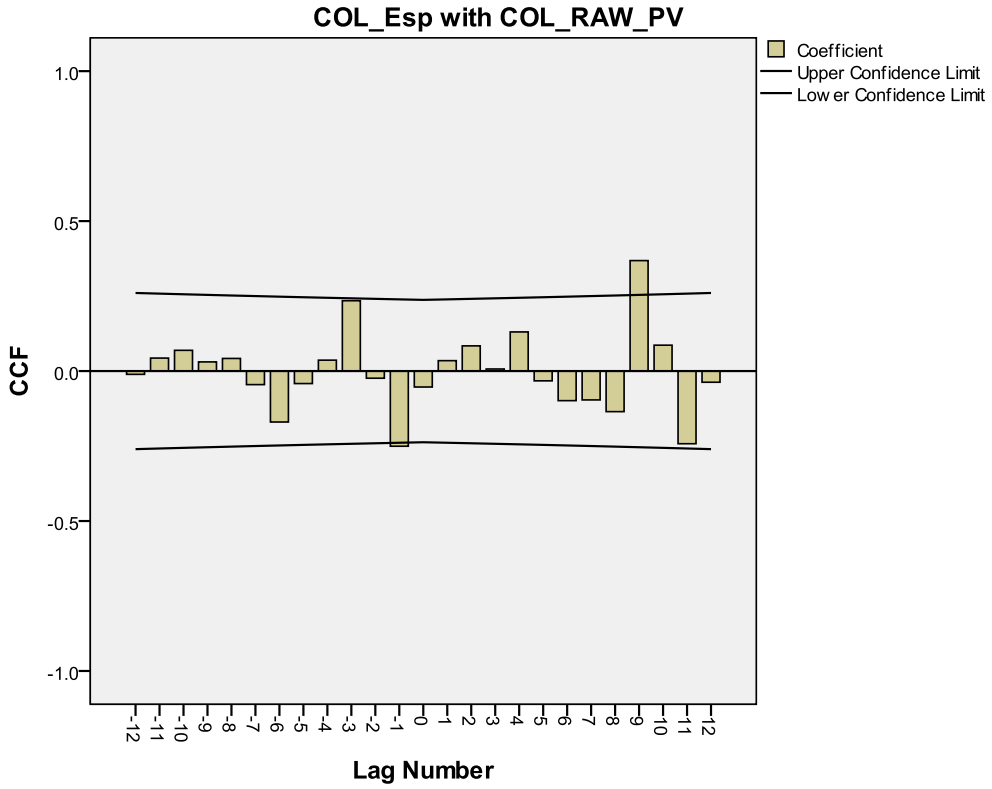


Figure 14

Cross Correlations between the query "trabajo en España" and Residential Variation for Peru.

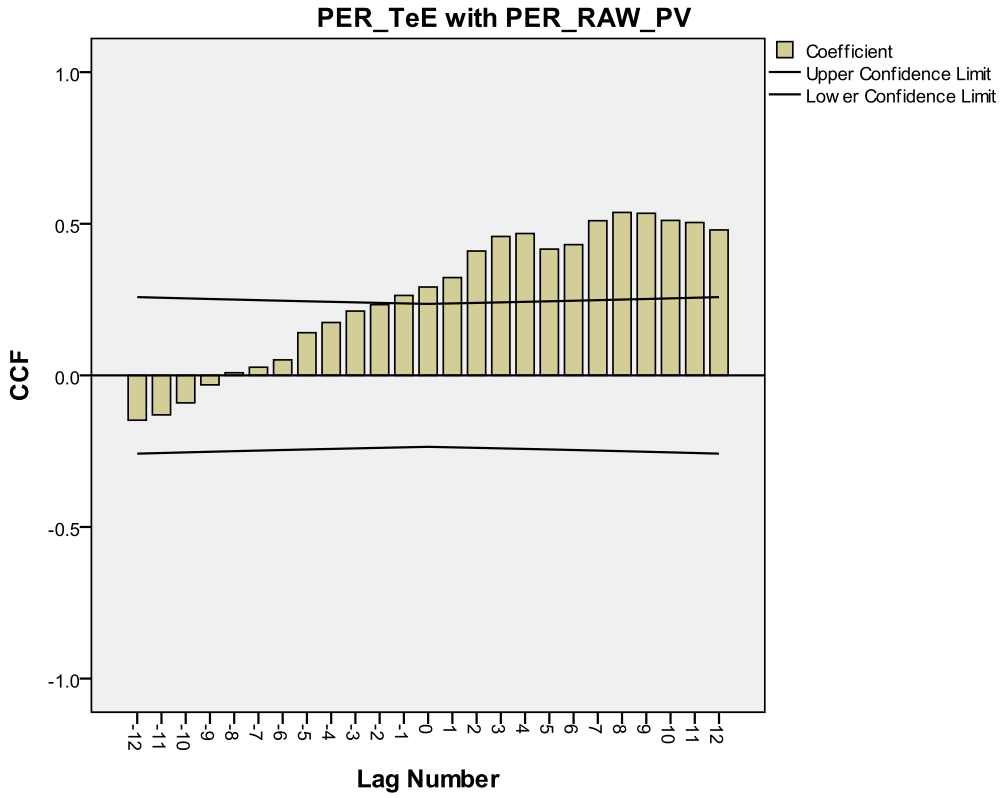


Figure 15

Cross Correlations between the query "embajada de España" and Residential Variation for Peru.

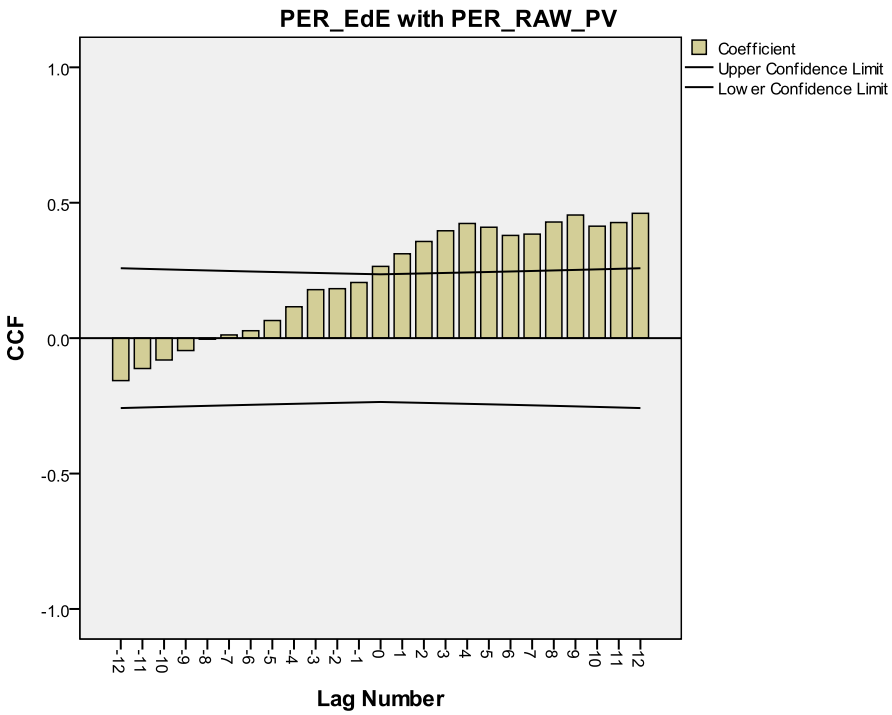


Figure 16

Cross Correlations between the query "España" and Residential Variation for Peru.

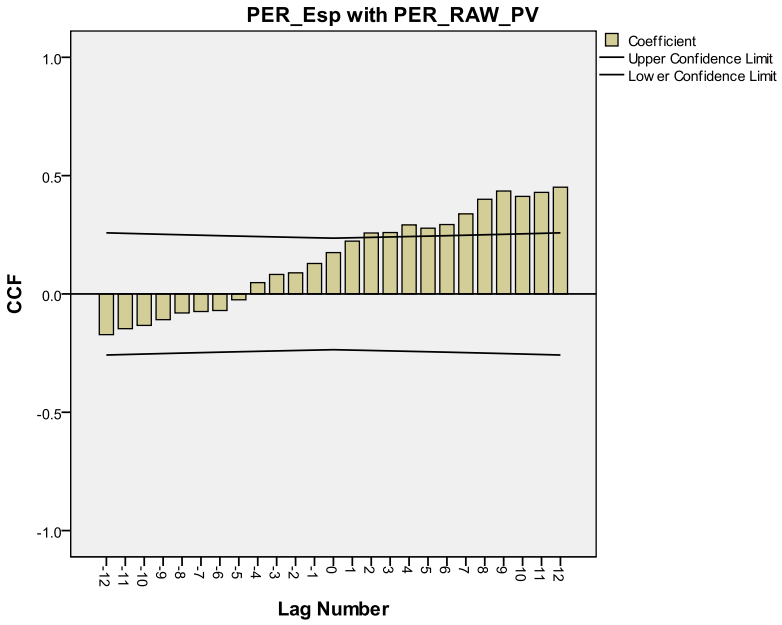


Figure 17

Cross Correlations between the stationarized time series of the query "trabajo en España" and Residential Variation for Peru.

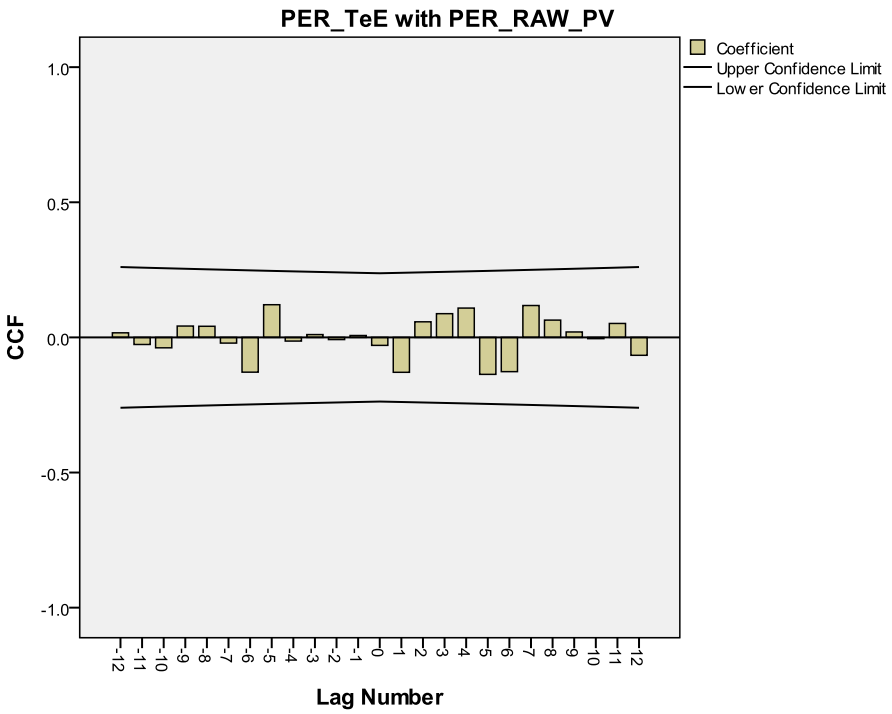


Figure 18

Cross Correlations between the stationarized time series of the query "embajada de España" and Residential Variation for Peru.

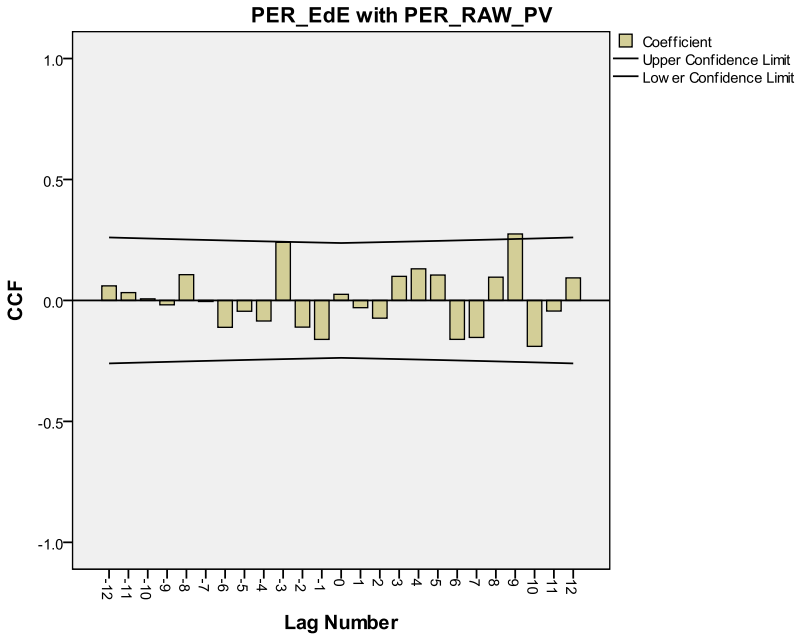


Figure 19

Cross Correlations between the stationarized time series of the query "España" and Residential Variation for Peru.

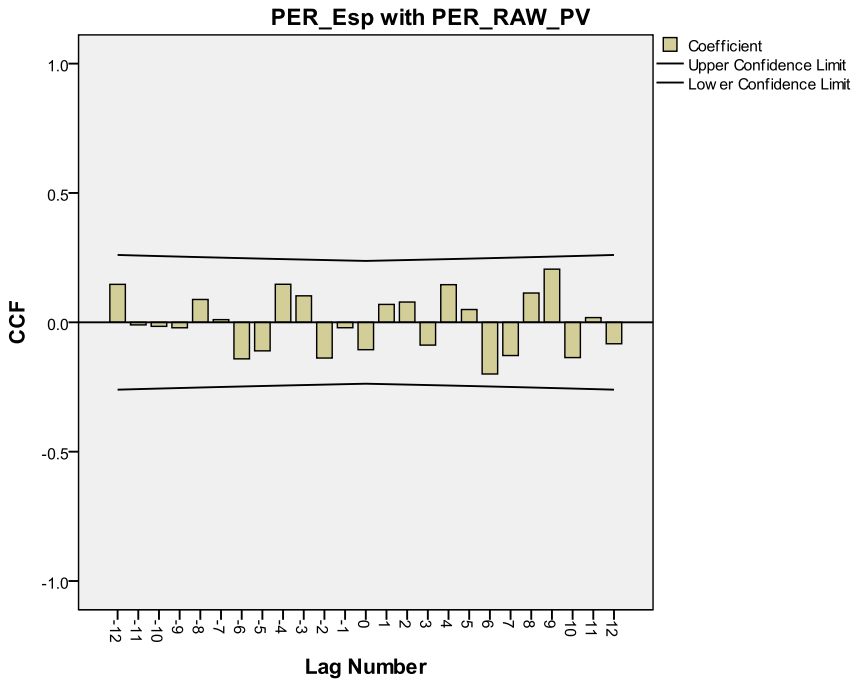


Figure 20

The sequence chart of non-stationary z-scores of the queries: "trabajo en España", "embajada de España","España", and Residential Variation for Argentina.

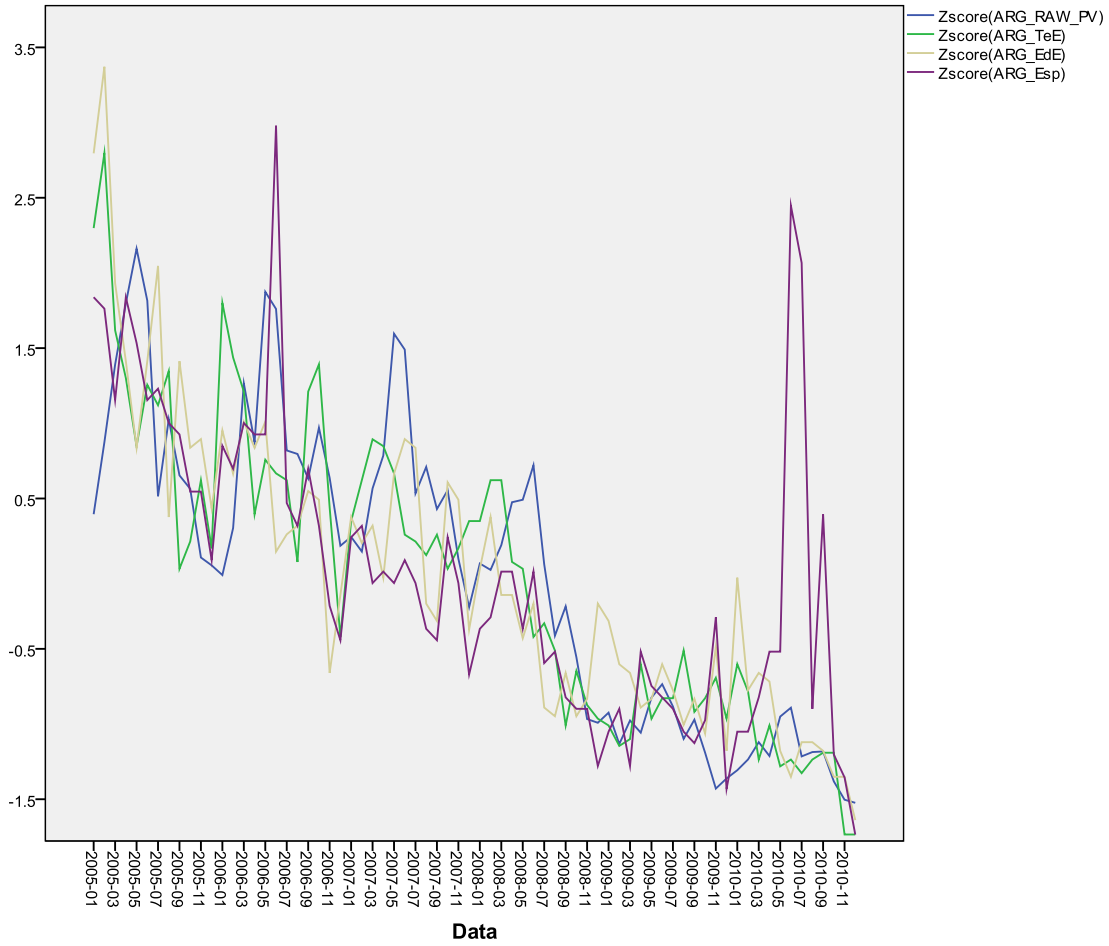


Figure 21

The sequence chart of non-stationary z-scores of the queries: "trabajo en España", "embajada de España","España", and Residential Variation for Colombia.

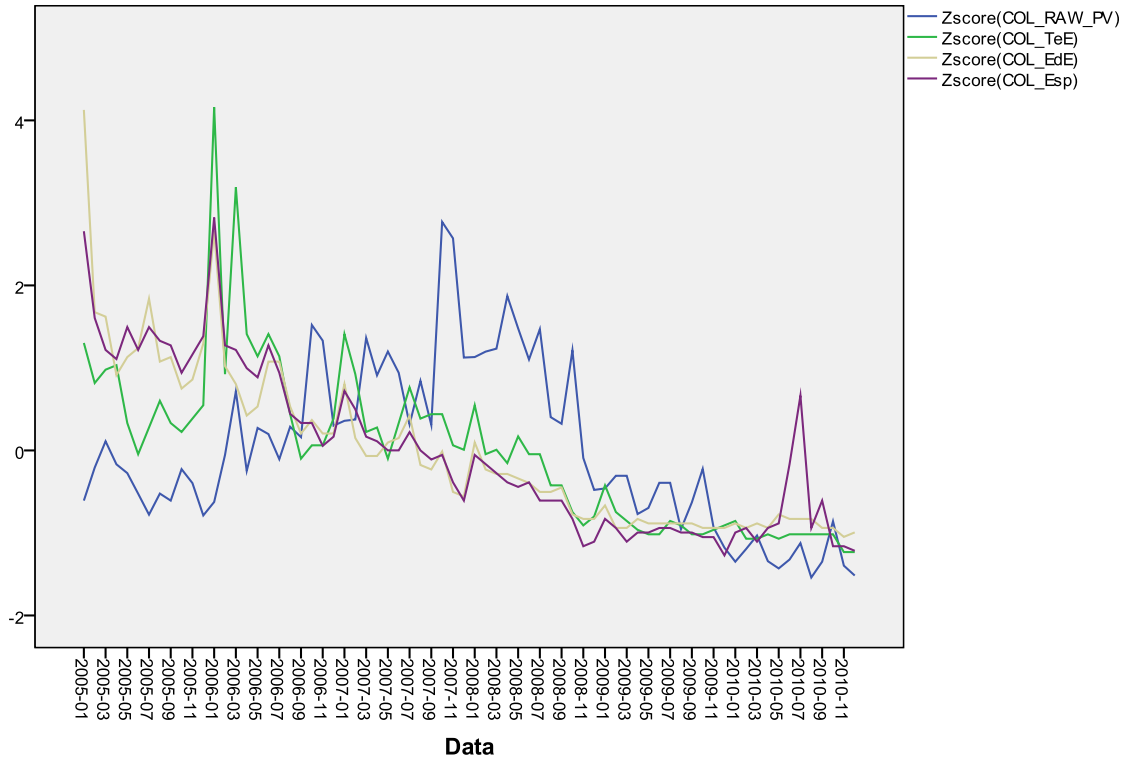


Figure 22

The sequence chart of non-stationary z-scores of the queries: "trabajo en España", "embajada de España","España", and Residential Variation for Peru.