SciencesPo.
OSC / CNRS
Observatoire sociologique du changement

Assets of longitudinal data in describing the immigrants' assimilation process: Analysing Ethnic Inequalities in the French Labour Market

Mirna Safi

Notes & Documents

n° 2008-01 Avril 2008

Résumé :

Le processus d'intégration des immigrés est un objet social éminemment temporel. Les données de type longitudinal, et notamment les données de panel récoltant des observations répétées sur un ensemble d'individus à plusieurs dates successives, devraient donc être privilégiées lorsqu'on étudie le parcours des immigrés dans une société d'accueil. Pourtant, leur usage en sciences sociales reste limité malgré un essor particulier d'une littérature méthodologique sur ce sujet depuis les années 1980.

L'objectif de ce travail est de présenter les apports de l'exploitation de données longitudinales pour la connaissance sociologique sur l'intégration des immigrés. L'accent sera mis sur deux atouts fondamentaux des méthodes appliquées à ce type de données : la capacité à prendre en compte des variables inobservées ou/et inobservables d'une part et à analyser des problématiques de causalité entre différentes dimensions de l'intégration d'autre part. Deux exemples empiriques seront développés ; l'un se fonde sur l'usage des méthodes d'économétrie des panels dans l'analyse des inégalités entre immigrés et natifs sur le marché du travail en France, et l'autre repose sur la mesure de l'effet causal de la naturalisation sur l'emploi grâce à l'estimation de modèles à équations simultanées.

Pour citer ce document :

Safi, Mirna (2008). « Assets of longitudinal data in describing the immigrants' assimilation process : analysing ethnic inequalities in the French labour market », Notes & Documents, 2008-01, Paris, OSC, Sciences Po/CNRS.

Pour une version électronique de ce document de travail et des autres numéros des Notes & Documents de l'OSC, voir le site web de l'OSC : http://osc.sciences-po.fr/publication/pub_n&d.htm

Abstract:

This work presents the main advantages of the use of longitudinal data for the sociological research on the assimilation process. Two fundamental issues will be developed: the advantages panel data offer in controlling for unobserved individual heterogeneities on one hand, and, on the other hand, their capacity to make some advances in the estimation of causal effects. Two empirical examples related to each of these advantages are provided by analysing inequalities between immigrants and natives in the French labour. These examples put the stress on some methodological issues linked to the use of such data.

Readers wishing to cite this document are asked to use the following form of words:

Safi, Mirna (2008). "Assets of longitudinal data in describing the immigrants' assimilation process : analysing ethnic inequalities in the French labour market", Notes & Documents, 2008-01, Paris, OSC, Sciences Po/CNRS.

For an on-line version of this working paper and others in the series, please visit the OSC website at: http://osc.sciences-po.fr/publication/pub_n&d.htm

1. Introduction

Assimilation of immigrants is fundamentally a temporal process. Longitudinal data, particularly those that follow the same individuals at several successive dates, should be therefore chosen whenever it is possible in order to analyze the immigrants' situation in a host society. Nevertheless, their use is still rather rare in empirical studies in sociology despite a rapid expansion of methodological literature on this subject since the 1980s.

The aim of this work is to present the main advantages of longitudinal data for research on the assimilation process. It starts by studying the reasons that explain the shortcomings of the sociological research in this field, analysing the American and the French experiences. It puts the stress next on the necessity of longitudinal data in immigration research. Two fundamental issues will be developed: the advantages panel data offer in controlling for unobserved individual heterogeneities on one hand, and, on the other hand, their capacity to achieve a major sociological concern, the estimation of causal effect.

Empirical examples are provided to illustrate each of these advantages. They are based on the use of a French longitudinal dataset and try to analyse inequalities between immigrants and natives in the French labour market. The first example describes the evolution of the differences between native and immigrant unemployment from 1968 to 1999. The second one tries to measure the causal effect of naturalization on immigrants' employment in the same period.

2. The temporal gap: comparing the longitudinal dimension of assimilation in the theoretical and empirical works

Ever since its emergence as a scientific paradigm in the works of the Chicago School, and whatever the precise words used or their ever changing meaning, sociologists have kept speaking about assimilation as a process. Park and Burgess provided a widely used definition that relies on this temporal apprehension of the concept: "assimilation is a process of interpenetration and fusion in which persons and groups acquire the memories, sentiment, and attitudes of other persons or groups, and, by sharing their experience and history, are incorporated with them in a common cultural life" (Park and Burgess 1921, p. 735).

From then on, assimilation has been regarded as an emblem of social change (Abramson 1994), and social change occurs by definition over time. The usefulness of the understanding of assimilation as a process has become more and more important as the "assimilationist" connotation of the term was loosing ground. While the "old" conception stresses assimilation as the natural end point of the immigrants' incorporation (Alba and Nee 2003) the new theories of assimilation point out the existence of several modes of incorporation that are not always "assimilationist", and which empirical observation requires a temporal follow-up over years and generations (Portes and Zhou 1993 ; Portes 1997). Longitudinal data, particularly those that allow us to follow the same individuals at several successive dates, that is to say panel data (Baltagi 2005), should be therefore chosen whenever it is possible in order to analyse the immigrants' situation in a host society[1].

Almost all the thinking about assimilation ensues from this temporal nature of the concept: the analysis of the duration of the process and its more or less rapidity, the description of its stages and directions, the decomposition of its different dimensions, etc. Obviously, authors did not agree on these points and developed different theoretical frames. Park and Burgess used the metaphor of the race-relations cycle when analysing social change, a theory that was rejected later basically because of its ethnocentric nature[2]. Milton Gordon (1964) defined seven stages of assimilation and separated for the first time acculturation from the structural assimilation, opening thus the way in front of a huge literature that criticized the assimilation paradigm. Authors have debated at length about the duration of the process: while Warner and Srole (1945) defined a "short" assimilation as lasting less than six generations, the most recent theoretical developments emphasize the importance of the study of the first and second generations (Portes 1996). H. Gans (1973). pointed out the "bumpy" form of the "assimilation curve" showing that some cultural characteristics may remain, and even re-emerge in the third generation All these concepts and assumptions require the observation of change during time, on the individual level or the generational one. Therefore, and in spite of disagreement among sociologists on the trajectory taken by immigrant incorporation in the host society, the theoretical literature on this subject relies all together on the temporal dimension of assimilation.

This theoretical common point doesn't hold out when we look and the empirical studies. Indeed, the overwhelming majority of empirical researches on immigration, particularly works that rely on quantitative investigation, are carried on regardless of the temporal dimension of assimilation. This "neglect of time" (Maines 1987) is linked to two different reasons.

The first reason is related to the shortage of the available data. The dynamics of the assimilation process can't be properly observed without longitudinal panel data, which allow a life course following, at least for a certain period. Such data are very rare. Even in the United-States where immigration studies has been very dynamic ever since sociology exists as a scientific field, the data used in the works on immigration suffer from the fact that they are not designed specifically to this aim (Levine, Hill, and Warren 1985 ; Jasso et al. 2000). The major databases used to study immigration are indeed the Census and the Current Population Surveys. Even if, comparing to the equivalent data in France, and more generally in Europe, the American census gives some important information about the immigrant population, its major pitfall lies in the fact that it is cross-sectional. The data is collected at fixed points in time and doesn't allow researchers to follow up the same individual over time. Using repeated cross sections from the Census certainly accounts for a longitudinal dimension especially when a cohort design is used. This method

---

[1] There are several types of longitudinal data, among which the more important are history data and panel data. The longitudinal attributive adjective is not always properly used in describing certain data. In this paper, it will be used only referring to panel data.

2 Alba and Nee argue in favour of a rehabilitation of the assimilation concept developed by Park and Burgess ascribing its ethnocentric apprehension to some later works that claim to be followers of the Chicago School's ones (Alba and Nee 1997) (Alba and Nee 2003).

became widespread, especially in works describing assimilation of immigrants in the labour market, since the pioneer work of G. Borjas (Borjas 1990 ; Myers and Cranford 1998). Nevertheless, using data from successive censuses doesn't really compensate the lack of panel data; the longitudinal description it allows is indeed aggregated and doesn't give precise information about the individual trajectory. Moreover, the cohort tracking method may lead to important bias in the results: the changes measured may be linked to changes in the cohort composition rather than to changes in the individual situation of immigrants. A major problem is related to emigration. Return migration substantially differs across countries of origin and is not an independent phenomenon. Immigrants who leave the host country after a certain period of time may be different from the others and their omission may lead to a significant selectivity bias in the results (Hu 2000). Some researchers tried to avoid these problems by using existent nationally representative longitudinal data; in the United States for instance, the Panel Study of Income Dynamics (PSID) and the National Longitudinal Surveys (NLS) are sometimes used in immigration research. But these databases suffer from the fact that they were not designed for immigration studies: the sizes of the samples available to study immigrants are problematic, many key migration variables are omitted, etc. These problems clearly point to the importance of designing a specific immigration longitudinal database. The New Immigrant Survey (NIS) was created in order to fulfill this purpose (Jasso et al. 2000). The availability of these data will certainly change the shape of the research on the assimilation process.

Such data do not exist in France yet. In this country, the problem of the availability of data on immigration has been more serious, because of the fact that census data have been unable to properly spot the immigrants and their descendents. Until now, many key variables in the study of the assimilation process are lacking in the census (arrival date, naturalisation date, French fluency, etc). However, in 1992, a survey especially dedicated to the analysis of immigration was carried out; "Mobilité Géographique et Insertion Social" (MGIS). A new specific immigration survey that should improve this first version shall be available soon. Although, these specific surveys were based on a very rich questionnaire and involve extensive information about immigrants' characteristics giving retrospective information about their life, they are cross-sectional. However, similarly to the case of the United-State, there are some longitudinal data which weren't initially designed to analyse immigration that can be used for this purpose. One of the most useful empirical dataset in France to study this issue will me presented later in this article.

There is a second reason for the gap between the theoretical and empirical importance given to the temporal dimension of assimilation. It is not related to the data but rather to their use. In fact, even in the rare works that manipulate the available longitudinal data, the individual temporal design is not properly incorporated in the analysis. Indeed, researchers in social science and especially in the sociological field remain very reluctant to the methods and models constructed by statisticians during the last three decades in order to analysis longitudinal data (Halaby 2004). Yet, analysing panel data grows out of a tradition that dates back to the 1970's and great names of empirical sociologists took actively part in its developments (Goodman 1973 ; Duncan 1980 ; Duncan 1981). From then on, methodological literature on panel data has been more and more flourishing and essays that recommend their use are more and more visible in prestigious sociological journals[3]. Nevertheless, applications of such methods are mostly found in the economic literature sociologists having been rather slow in absorbing this literature and in capitalizing on the advantages of panel data. This is particularly true in the sociological works on immigration.

This article deals with two major advantages of using panel data giving direct applications in the sociology of immigration. Two fundamental issues will be developed: the advantages panel data offer in controlling for unobserved individual heterogeneities on one hand, and, on the other hand, their capacity to achieve a major sociological concern, the estimation of causal effects.

3. The advantages of panel data

There are various types of longitudinal data among which the most known are time series, event history data and panel data. The scope of this article is limited to the latter type. The central feature of a panel dataset is that it records at regular intervals the state each individual occupies at each time an observation is taken. Thus, for each observation of the dataset, variables available could be double-indexed using an individual and a temporal index: variable Xit gives information about the individual i on a date t. The strengths of panel data lie in this double dimension of the collected information.

Baltagi (2005) enumerates several advantages of panel data in the introduction of his book. This study won't develop all the uses of panel data. It chooses to stress upon two fundamental assets that are classically distinguished in the literature. The first one lies in their capacity of accounting for unobserved or non-observable variables and the second one resides in their leverage to estimate causal effect.

Let's consider a single-equation model using panel data where the dependent variable is denoted Yit and the explanatory variables Xit. β denotes the effect parameters pertaining to Xit and Uit denotes the error term. Estimating equation (1) as if the data were cross-sectional neglects the correlation between the error terms of the same individual.

$$Yit = Xit\,\beta + Uit \qquad (1)$$

The main difference between the cross-sectional and the longitudinal models lies in the fact that, in the latter one, the error can be decomposed into two terms, one that depends only on the individual and the other depends on both time and individual. Thus equation (1) can be written:

3 The Annual Journal of Sociology for instance, published three methodological articles on longitudinal data (Hannan and Tuma 1979 ; Petersen 1993 ; Halaby 2004).

$$Yit = Xit\ \beta + \alpha i + \varepsilon it \qquad (2)$$

$$\text{where } Uit = \alpha i + \varepsilon it \qquad (3)$$

αi is an individual specific error term that is time invariant. As Peterson points out, one appealing interpretation of this individual non temporal error term can be obtained by letting Zi denote all unobservable variables for individual i and δi denote their effect parameters (Petersen 1993). Hence, αi is no longer different from δiZi : it captures the effect of all unobserved time-invariant variables for individual i. These unobserved variables, also called unobserved heterogeneities, refer to variables that are either omitted in the dataset, or fundamentally non-measurable. This capacity of accounting for unobserved variables is of great use in the study of immigration data: as developed above, the major part of data used in immigration research aren't designed to this aim and suffer most likely from a lack in some very important immigration specific variables. If these variables are supposed to be invariant to time, the structure of panel data can account for their effect. Thus, variables such as the arrival date, the naturalisation date, the specific visa category that permitted entry, the pre-migration characteristics etc., all of which are usually omitted in mainstream social science surveys, can be regarded as unobserved variables if a panel architecture of the data exists. However, an important point ought to be stressed upon: the effects of the unobserved variables can't be estimated in the model. For instance, if the aim of the study is to estimate the effect of a specific pre-migration characteristic - let's say the father's occupation - it is obviously impossible to achieve it if the variable information isn't available in the data. In other words, unobserved heterogeneities can't be specified as interest variables in the model. However, whenever one is interested in the effect of observed variables Xit, panel data architecture enables him or her to control for the unobserved heterogeneities leading to consistent estimation ofβ. Moreover, as the empirical example below will show, the model gives some information about the share of these unobserved heterogeneities in the total variance of the error term asserting thus for the extent to which the model unexplained variables can be imputed to these unobserved characteristics.

The second asset of panel data, while linked to the first one, is more general. As Halaby (2004) asserted, the structure of panel data "provides the analytical leverage for rigorously achieving the central aim of quantitative research: the estimation of causal effects."

Indeed, when working on panel data, researcher can go beyond some descriptive measurements of the correlation between variables. This is due to the fact that panel architecture of data allows ones to deal with the problem of causality. Causality in social sciences has been a long subject of concern because while the importance of its detection is capital in order to apprehend the social world, the principles of its definition and its measurement are far from being consensual. Researchers in the econometric field defined a temporal concept of causality that can be applied to social science experience (Wold 1954 ; Granger 1969). Econometrics consider that, X is one of the causes of Y, if the expectancy of Y at time t, controlling for the past of Y and the present and past values of X, is different from the expectancy of Y controlling only for the past of Y (Lollivier and Verger 2005 ; Lollivier 2006). In other words, the knowledge of the past value of X significantly modifies the prediction one can make for Y (Sobel 2000). It is however important to underline that the interpretation of the causal relation between X and Y remains above all an intellectual act: it depends on the scheme the researcher fixes in order to analyse the data. Causal inference is not "a simple affair that can be reduced to a formula applied mechanically" (Duncan 1972). Nevertheless, when a specific causal scheme between two variables (or more) is argued by the researcher, working on panel data is certainly the best way to validate it empirically because panel data can determine a chronology between the temporal values of X and Y.

These advances in the study of causality can be very useful in the research on immigration and the assimilation process. Indeed, since the canonical work of Gordon (1964), works on the assimilation process have been trying to measure the causal relations that may exist between assimilation dimensions. Is acculturation, as the assimilation paradigm supposes, a clue dimension that can boost the others? Is there a causal link between cultural assimilation and social mobility? What factors can be identified as the causes of structural assimilation? When the available data are longitudinal, it becomes much more precise to deal with such questions making thus some advances in the inference of causal relations among assimilation dimensions.

4. Two empirical examples

4.1. The data

The empirical analyses developed in this article are based on the exploitation of a French longitudinal dataset: the *Échantillon Démographique Permanent* (EDP). EDP is a panel sample that allows us to follow almost 1% of the French population through information contained in the 1968, 1975, 1982, 1990 and 1999 French censuses.

This database was created in 1967 with the aim of accumulating, through the different censuses, some demographic characteristics of the French population measured on a representative sample. Individuals who go into EDP, whom can be called the EDP individuals, were included in the database according to a day of birth rule. If an individual is listed in the census and meet the date of birth criterion, one can follow him/her during the succeeding censuses if he/she is listed again. In addition to the characteristics contained in the censuses, EDP includes civil status forms registered in France and gives information about the major events in the EDP individual's life course: marriage(s), birth of children, death, etc. For each new census, individuals who meet the birth rule are further added to the existent sample. As far as immigrants are concerned, they appear in the EDP sample as soon as they are listed in the census or as soon as a registered civil status certificate involving them is collected. Immigrants can also disappear from the EDP base due to emigration or death, which is also the case for the entire population of the sample.

EDP is valuable for the study of immigration in France for two main reasons. First, it allows us to work on reasonably satisfactory number of immigrants belonging to about ten different groups representing thus over 90% of the French immigrant population in the 1968-1999 period. Second, EDP lends itself to a long term analysis of data and is convenient for describing the progress of the assimilation process.

Nevertheless, as it is the case for almost all national population longitudinal data, EDP, or more specifically the census data it was extracted from, was not conceived to analyse specifically immigration in France. It lacks thus a great amount of information that is crucial for the study of this population, as for example the length of stay or the language fluency. However, when compared to the existent statistical material in France, EDP can treat the dynamic of immigrants' assimilation in a rather acceptable way. Several research works on immigration were indeed based on the exploration of this data among which two PhD thesis: the first analysis the dynamic of the second generation integration and examines their possible return migration (Richard 2004) while the second used the EDP data to analyse ethnic inequalities in France dealing with several dimensions of the assimilation process (assimilation in the labour market, spatial segregation, citizenship, intermarriage) (Safi 2007).

In order to back up the assertion developed above on the usefulness of panel data, two empirical examples related to the immigrants' assimilation process in France will be presented below. Each of them illustrates one of the two advantages of longitudinal data developed in the former section.

4.2. Example I. A longitudinal study of ethnic inequalities in the French labour market

In France, it's well known that some groups experience, more than others, discrimination and prejudice in the labour market. However, there are very few studies that are able to measure the differences between the socioeconomic integration of immigrants and natives, and even fewer are the ones that can give some information about the evolution of these differences over time. This empirical study conducted on the EDP longitudinal data set, aims to describe the situation of immigrants in the labour market during the last 30 year period in France, and to provide some explanations for their durable socioeconomic inferiority.

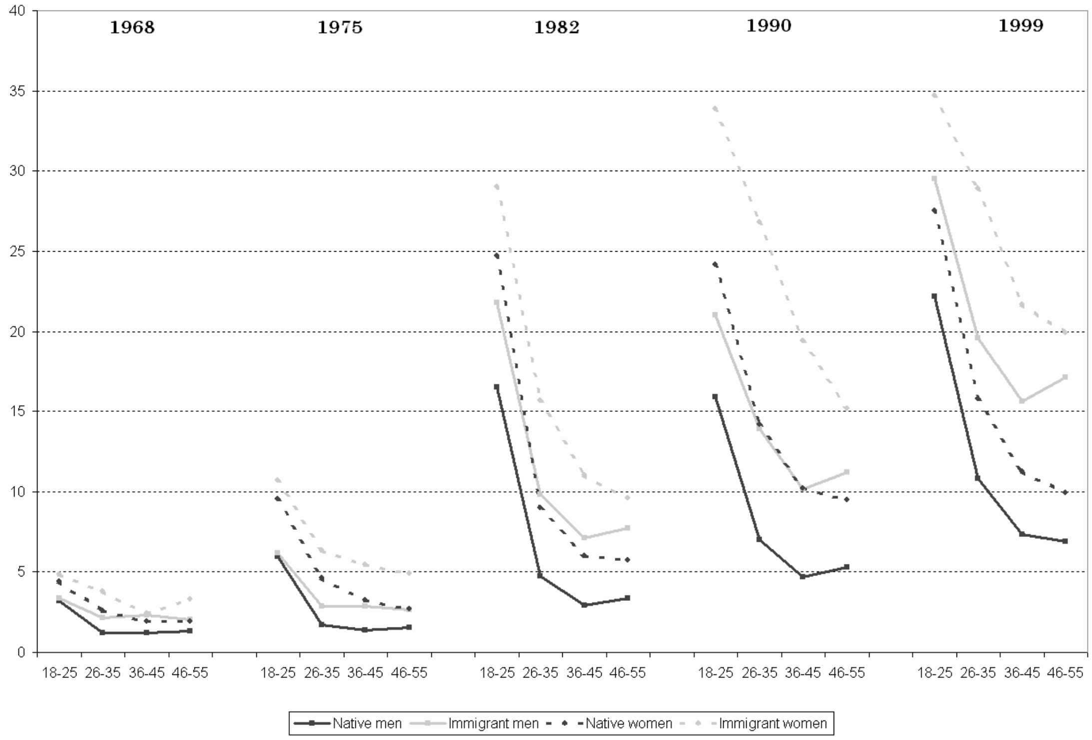

Figure 1.1. – Unemployment rates for immigrants and natives (1968-1999)

Using data of the EDP as cross-sections, figure 1.1 shows the evolution of unemployment rates over the period among immigrants taking all together and natives. Even if the general form of the curve is almost the same, immigrants' rates are always above the natives' ones and this becomes more and more salient during the period. The difference is also more important between the rate of unemployment for immigrant and native women. Figure 1.2 shows that these ever growing differences between natives' and immigrants' unemployment are not due to a different age composition of the two populations. Whatever the defined age class, immigrant women's unemployment curve is on the top, and the one for the native men at the bottom. An inversion in the position of immigrant man rate and native women rate is noteworthy since 1990.This figure also shows clearly that the gap between the curves of natives and immigrants is becoming more and more important.

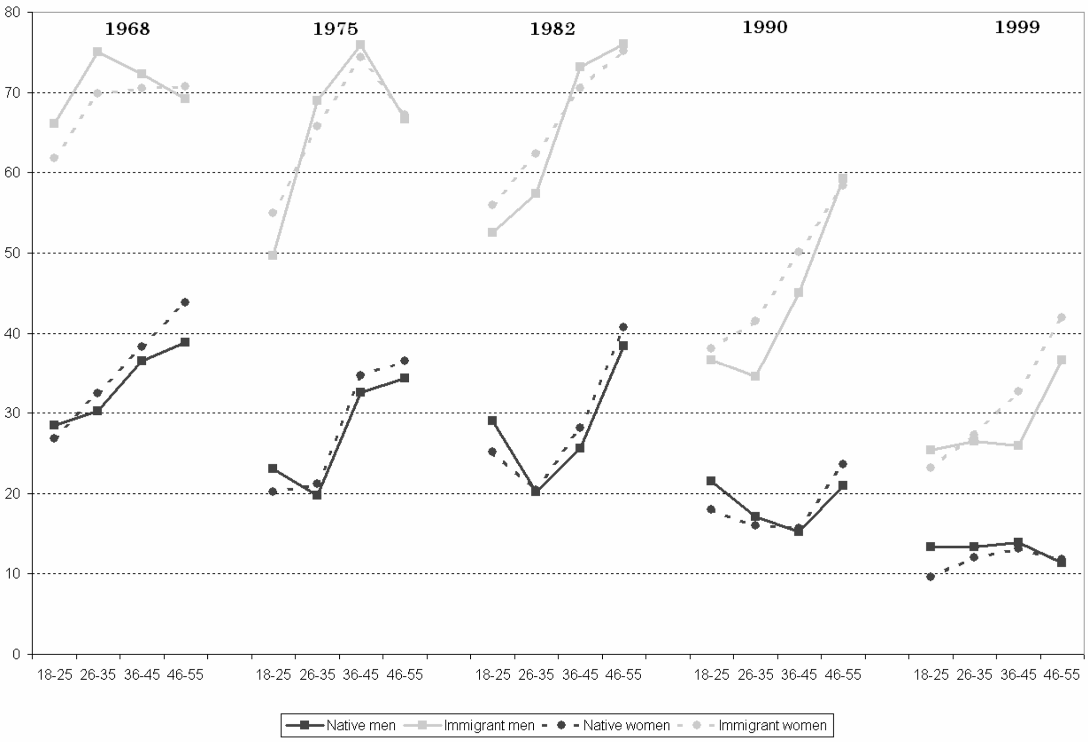

Figure 1.2. – The Evolution of unemployment rate for the native and immigrant population by age and sex

One can think that these observed differences in the labour market between immigrants and natives are due to individual characteristics that would be radically different in the two populations as for instance qualifications and occupations. In fact, a rapid examination of these characteristics reveals rather a convergence between the immigrants' and the natives' characteristics. This convergence can be represented briefly in the next two figures[4]. Figure 1.3 shows that, even though the number of diplomaless is still more important among immigrants, the education gap between immigrants and native has been rapidly diminishing. At the same time, the curves of post-graduated become almost superposed at the end of the period. The same conclusions can be inferred when one takes a look at the occupations evolution (figure 1.4). Immigrants are certainly much more frequently blue-collar than natives, but here also, the difference tends to diminish during the period. As for the executive and managers occupations the curves for immigrants and natives are almost superposed.

[4] Exploitation of the last census data shows very clearly this convergence between the socio-demographical characteristics of immigrants and natives (Borrel, 2006).

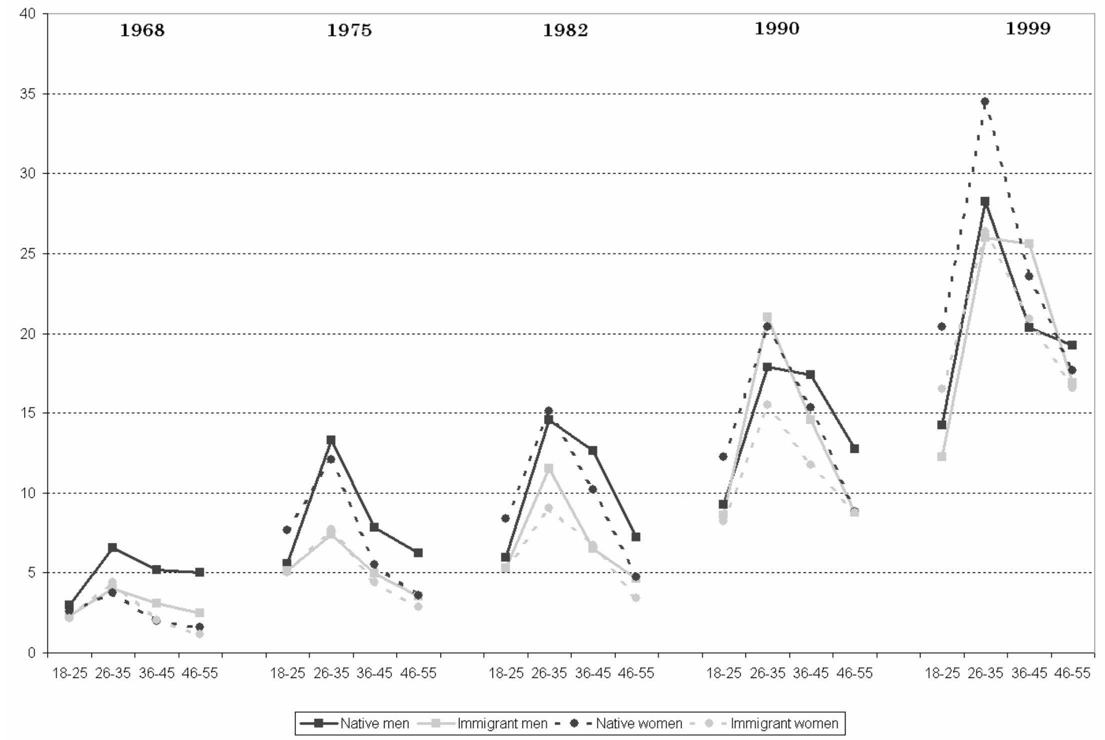

Figure 1.3. – Percent of holders of post-graduate diploma and persons without any diploma among immigrants and natives

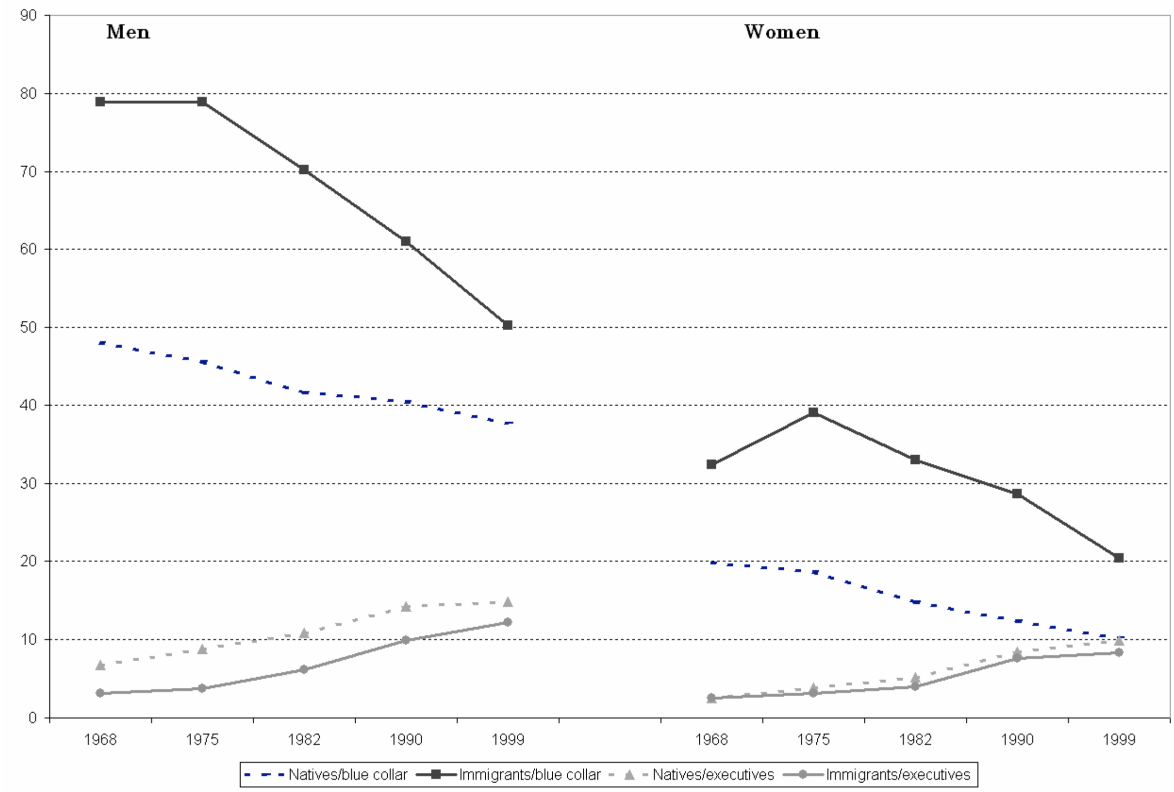

Figure 1.4. – Percent of executives and blue-collar workers among employed immigrants and natives

These general descriptive figures make it clear that, although the principle characteristics of immigrants tend to converge toward the ones of natives, the gap between their unemployment rate and the one the natives experience tends on the contrary to become more and more pronounced. What explanation can be provided to account for these differences? Are they due to individual characteristics that make immigrants less efficient in the labour market or rather to a difference of treatment in the recruitment procedure that tends to prefer natives?

4.2.1. The panel model

In order not to loose a great part of the dataset, and since we are working on a very long period, it seems better to choose only the last three dates 1982, 1990, 1999. To construct the panel architecture of the sample that shall be used in the estimation of the model, an entry date in the panel should be fixed: all individuals who are present in the 1982 census, who are 18-55 years old at the 1982 census date and who declared their employment status at this date are included. The same individual gives as many observations as he or she is listed in the following two censuses. This leads to several temporal trajectories of individuals that are reported in Table 1.1.

Table 1.1. – The temporal structure of the sample

Census dates	Immigrants		Total sample	
	Frequency	%	Frequency	%
82-90-99	7 127	30.23	129 973	49.67
82-90-.	5 900	25.02	59 329	22.67
82-.-.	9 797	41.54	65 043	24.86
82-.-99	757	3.21	7 325	2.8
Total	23 579	100	261 670	100

The total size of the sample used is 261,670 observations among which 23,579 concern immigrants. Almost 50% of these observations are listed in the 1982, 1990 and 1999 censuses. This percent falls to 30% for the immigrants' sample which is quite normal since, as mentioned before, immigrants are more mobile then natives.

Several models are estimated using these data. They aim at comparing the situation of immigrants and natives in the labour market on one hand, and the differences among the immigrants groups on the other hand. Models are estimated separately for men and women and for each model estimations are conducted twice: on the entire longitudinal sample and on the sub-sample where attrition is excluded (*i.e.* all individuals are present at the three dates of the samples). This approach measures thus the extent to which the results are altered by the attrition problem.

4.2.2. Some estimation considerations

The employment variable Y is a discrete dichotomous variable which value can change over time. The model estimated is thus a discrete panel model and equation 1 should be written as below:

$$Y_{it} = 1(X_{it}\beta + \alpha_i + \varepsilon_{it} > 0) \quad (4)$$

This paper won't develop all the methodological issues related to the estimation of such an equation. Some general considerations are though necessary to mention in order to give a better idea of both strengths and limits of such a temporal modelling.

A key issue is the specification one makes for the $\alpha_{i.}$ In general, two different specifications can be used. The first regards the α_i as an additional set of co-variables. It is what the literature calls the fixed effects model. The great advantage of such a model lies in the fact that there is no restriction on the possible correlation between the individual heterogeneities and the explanatory variables. Neither does it need a distribution hypothesis for the $\alpha_{i.}$. Nevertheless, such a model can only estimate variables that change over time whereas the coefficients of the time invariant variables can't be identified. In the case presented in this article, the major interest variable is the one that distinguish immigrants from natives; it is thus an invariant variable and its effect can't be estimated in a fixed effects model.

The second type of specification regards the α_i as random occurrences of a variable for which a distribution can be attributed, as for instance the normal one $N(0,\sigma^2_{\alpha})$. It becomes thus possible to estimate a probit model with random effects. In such a model, the error term ε_{it} is N(0,1) distributed (by definition in a probit model), and is supposed to be independent from the α_{i}. This latter hypothesis means that no correlation is allowed between the α_i and the other explanatory variables, which is a very restrictive one. This is due to the fact that the α_i are no more specified as co-variables but rather as random terms. In the empirical case this article is studying, we are yet obliged to choose this specification because the variable in which we are interested is time invariant[5].

In a random effect model, one can estimate not only the co-variables coefficients, but also the extent to which the residual variance can be explained by the individual heterogeneities. It can indeed be calculated as the ratio of the α_i variance and the total residual variance. The latter is nothing else then the sum of the ε_{it} variance (that is equal to 1) and the variance of the $\alpha_{i.}$. If we define ρ as below (equation 5), it gives us an estimation of the extent to which the total variance is explained by the individual heterogeneities.

$$\rho = \frac{\delta_{\alpha}^{\ 2}}{1+\delta_{\alpha}^{\ 2}} \quad (5)$$

5 For more details on the estimations methods and the specification of panel discrete models, see the (Arellano and Honoré 2001 ; Halaby 2004 ; Baltagi 2005 ; Lollivier 2006). The 14th Chapter of (Wooldridge 2005) is also very clear on this subject.

If ρ is equal to zero, the temporal architecture of the panel is not worthy; one can estimate unbiased parameters using a classical probit model, without specifying the individual heterogeneities.

A last characteristic of panel models should be mentioned. It is related to the problem of attrition: some individuals may leave the panel for several reasons. Should one take them into account or should they just be ignored in the model estimation? In the case studied in this article, individuals come into the panel at different dates and information about them may be lost if they are not listed in the successive censuses. Ideally, one should treat in a separate equation the outgoing of such individuals and analyse its determinants. Strictly speaking, this attrition problem shouldn't be stated as an estimation one because the methods used remain practically unchanged even when taking into account attrition. The lost of observations over time poses rather an interpretation problem especially when it refers to selective phenomena as return migration. To simplify matters, and to concentrate the concern of this article within the question of individual heterogeneities, the selection bias due to attrition is not treated. Nevertheless, the estimation of the model on several types of samples, some of them constructed without attrition and some others taking all the observations into account, gives some information about the extent to which selectivity bias are important.

4.2.3. Findings

The results are summarized in the figures below. Tables with the entire model estimation can be found in the appendix.

We begin by estimating a general model that compares the probability of being employed for native and immigrants. An additional model of labour market participation is estimated for women. An interaction term between the date of the census and the immigrant status variable is introduced to capture the evolution of inequalities before unemployment over time. Figure 1.5 compares the estimated parameters associated to being an immigrant rather than a native, for each of these models.

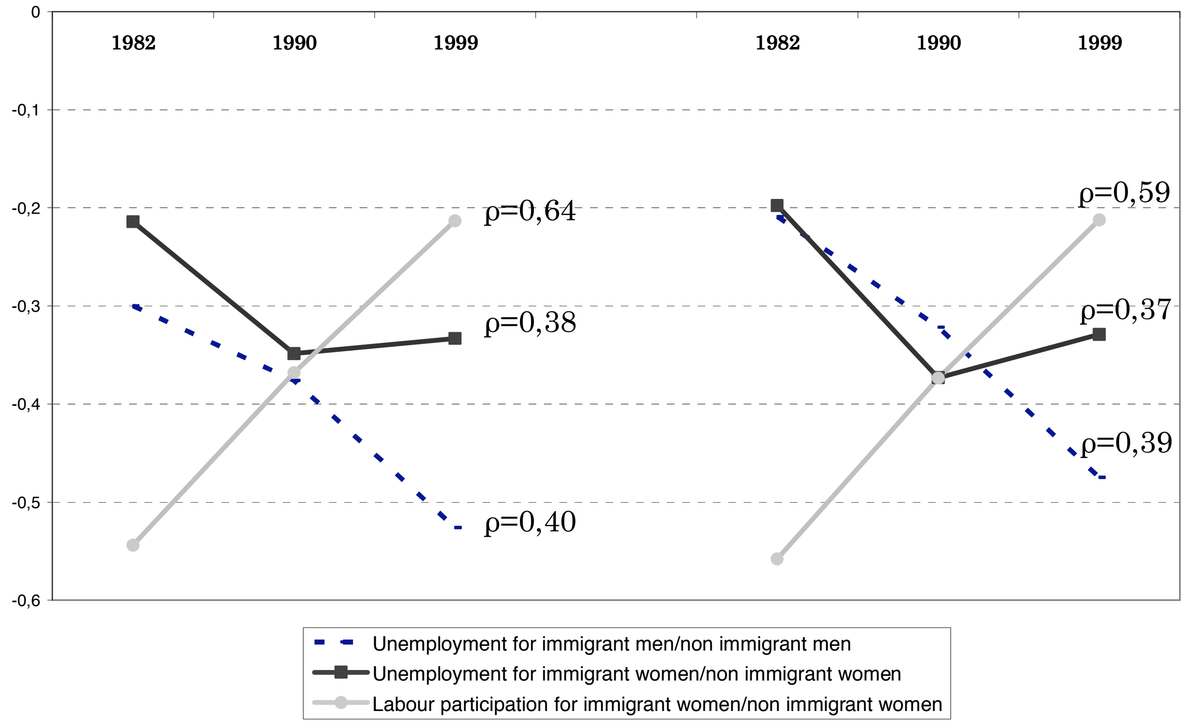

Figure 1.5. – Estimation of the effect of being an immigrant *versus* native interacted with time on the probability to be unemployed (random effect models)

The effect of being immigrant on the probability to be employed rather than unemployed remains negative and very significant in such a model both for men and women. Moreover, this effect tends to be more and more negative over the period, particularly for men. On the contrary, immigrant women participation is more and more important between 1982 and 1999, even if it remains lower that the one of native women. In addition to these quite known differences between immigrants and natives in the labour market, what these panel models tell us more is that individual heterogeneities (α_i) accounts for around 40% of the residuals of the unemployment models. Which mean that, even if we take into account the effect of the unobservable individual variables, we would still have 60% of the residuals variance that remain unexplained by the model. On the contrary, when the model is interested in the labour market participation, the share of the individual heterogeneities becomes more salient; 64% of the residual variance between immigrant and native women is due to individual differences that remain stable over time[6]. Hence, while the differences observed in the labour market participation between immigrant and native women can largely be attributed to individual characteristics, as far as unemployment is concerned, the large unexplained share of variance seems to be rather related to a different treatment, particularly on concrete mechanisms of durable discrimination in the labour market in France.

The same kind of model is estimated only within the immigrant population in order to compare the situation of different groups. Figures 1.6, 1.7 and 1.8 show the evolution of the magnitude of the estimated

6 One can here think about the number of children variable which was not introduced in the model or many other individual factors that have been identified in the literature on women's participation (partner's salary, social background, etc).

parameters for each ethnic group over the census dates. Only the coefficients estimated on the total sample are reported here; the ones estimated on the sub-sample without attrition are quite similar and can be checked for in the table B in the appendix. All ethnic groups are compared to Spanish immigrants. Here also the interpretation schemes are different for unemployment and labour market participation. The estimated coefficients are more and more negative when employment is the dependent variable; the differences among ethnic groups tend to be more and more pronounced during the period. On the contrary, the magnitude of the estimated parameters tends to shrink in the labour market participation model over time. Moreover, when the relative position of ethnic groups is compared, and as far as unemployment is concerned, the figures show that socioeconomic assimilation don't occur for all groups. The European groups are always the most advantaged ones while African and Turkish immigrants experience the most difficult situation in the labour market all over the period. These results control for observable and non observable variables. But here again, individual heterogeneities account only for a small part of the residual variance (30% for men and 38% for women). This figure is quite more important for the labour market participation model; 60% of the variance residual can be attributed to individual differences.

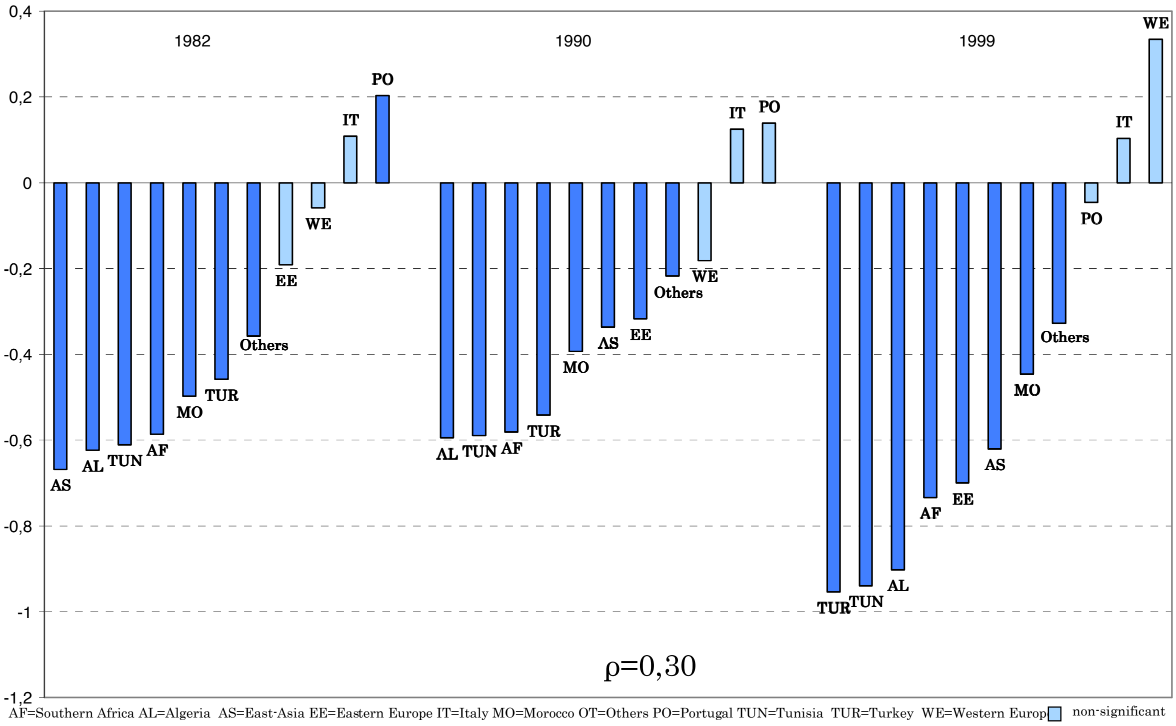

Figure 1.6. – Estimation of the probability to be employed for men: effect of the interaction between the census date and the ethnic group (random effects)

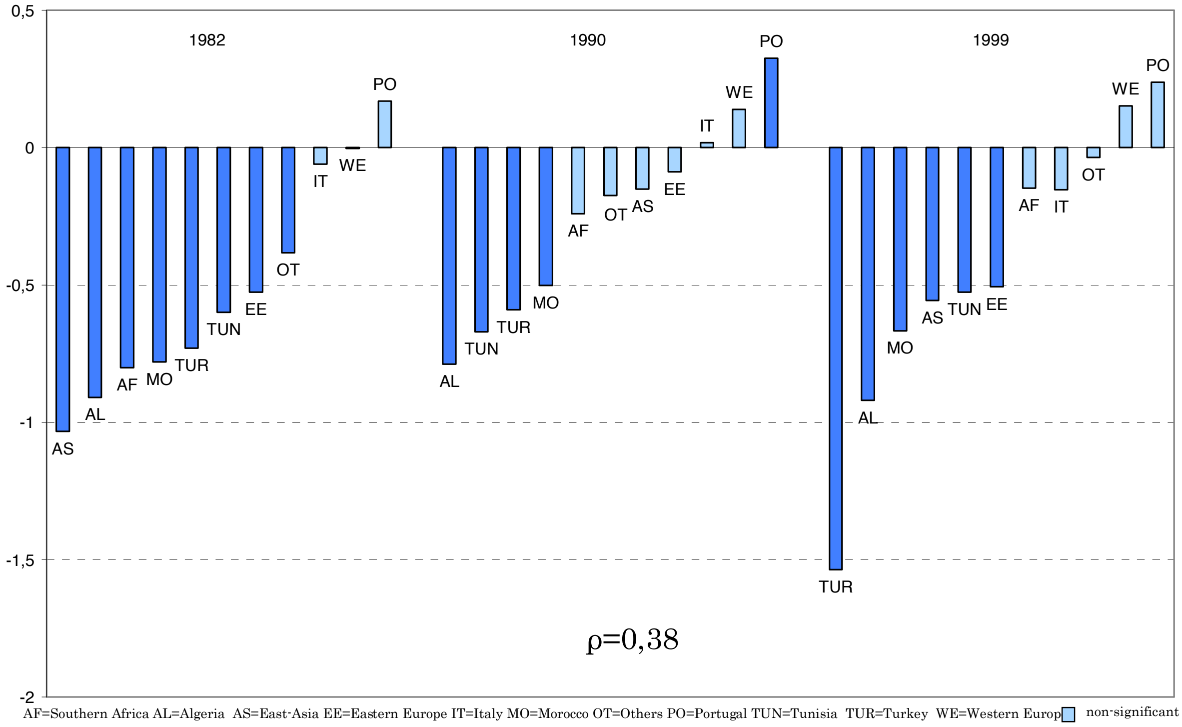

Figure 1.7. – Estimation of the probability to be employed for women: effect of the interaction between the census date and the ethnic group (random effects)

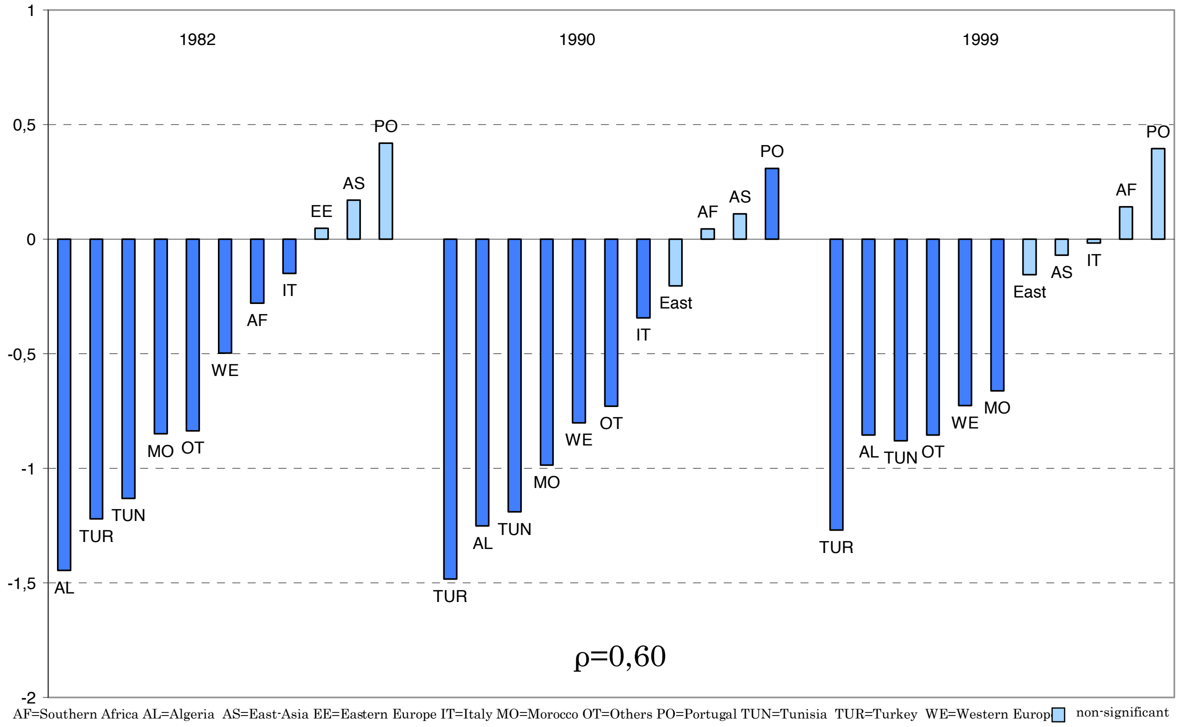

Figure 1.8. – Estimation of the probability to participate in the labour market for women: effect of the interaction between the census date and the ethnic group (random effects)

These models put forward the disadvantages of immigrants in the labour market in France. These employment penalties remain strong even after the control for individual observable and non observable variables made possible thanks to the panel dimension of the data. Moreover, the disadvantages are not gradually breaking up as the assimilation paradigm would predict. On the contrary, some immigrant groups seem to be durably underprivileged in the labour market (immigrants coming from the African continent or from Turkey) and their individual characteristics don't account for their disadvantaged position. They seem to suffer from a long term discrimination that affect their performance in the labour market and tend to make their mode of assimilation closer to the downward assimilation than to the straight line one(Safi 2006). In order to confirm these hypotheses, this work should be pursued and confronted to other results asserting for these discrimination mechanisms.

4.3. Example II: Causal inference: the effect of naturalisation on immigrants' employment

The second empirical case puts the stress on the causal inference facilitated by the use of panel data. It deals with the causal effect of naturalisation on immigrants' employment and is extracted from an article published by Fougère and Safi on this issue in 2005[7].

While gaining the nationality of the host country is often presented as the final step in the immigrants' assimilation process, questions can be raised as to whether it is not more an intermediate step, one that strengthens integration, in particular in its socio-economic aspect. Indeed, naturalization may affect immigrants' status in the labour market and, in particular, their ability to secure a job. This is particularly true in countries like France where the labour market discriminates legally between foreigners and citizens. When an immigrant gains French nationality, the range of jobs available to him opens up to include in particular all jobs requiring French nationality[8]. At the same time, naturalized immigrants can more easily circumvent discriminatory situations during the hiring process. Several empirical studies have revealed such forms of discrimination using among other methods, testing. These two reasons, legal and illegal discrimination in the labour market assert the hypothesis of a positive effect of naturalisation on immigrants' employment.

Many studies treated this subject using individual wages (Chiswick, 1978; Brastberg and al., 2002; DeVoretz and Pivnenko; 2004). These works show that naturalization brings about a greater wage increase and that the extent to which this so called naturalization "wage premium" is important depends on the country of origin: immigrants from developing countries see their job status improve to a greater extent after gaining U.S. nationality than other immigrants. Unfortunately, EDP provides no information about individual wage levels. The focus of this study will be oriented on the measure of the naturalization impact on immigrants' employment status.

7 For more details about this example see (Fougère and Safi 2005).
8 The set of this kind of jobs is rather large. Its scope covers not only the strict civil service and the public firms but also many liberal and entrepreneurial professions (Math and Spire 1999). All together, about 20% of the French labour market is in fact closed to foreigners (GELD 2000).

4.3.1. Construction of the sample

Since each individual declares his or her nationality at each census, it is possible to identify immigrants who have gained French citizenship between two censuses. Thereby, a sample that includes all the individuals present in two successive censuses was built[9]. This sample is restricted to individuals having declared themselves as foreigners when they first appeared in the EDP. In order for an individual to give rise to an observation, he or she must be present (or more specifically, identified in the census) in two consecutive censuses. This means that an individual citizenship may evolve in one of three ways:

- Foreigner in t, foreigner in t+1
- Foreigner in t, French in t+1
- French in t, French in t+1

Individuals on whom the third observation is made are discarded, considering that what we wish to detect is the transition from foreigner to French citizen. If an individual is of a foreign nationality in 1968 and 1975, he or she gives rise to an observation with the variable "naturalization" taking 0 as a value and the variable "observation period" which takes value 1 (the first wave in the panel came between 1968 and 1975, the second between 1975 and 1982…). On the other hand, if the person is a foreigner in 1968 and became French in 1975, he or she gives rise to an observation with the variable "naturalization" taking 1 as a value, and the variable "observation period", which takes also value 1[10].

Lastly, as this work aims to analyze the interaction between naturalization and employment, the sample is limited to individuals between ages 18 and 55, who were neither student nor engaged in the military at the time. Given those restrictions, the sample was reduced to 36 685 observations (or 21 779 individuals). Table 2.1 gives a view of the sample size and the countries involved in this study.

9 For this reason, the same individual may give rise to different observations at several points in the sample. The number of observations he or she has is equal to the number of inter-census periods during which he or she was present in the EDP panel dataset.
10 As a result of the sample design principle described above, all naturalizations not specifically identifiable as occurring between two census periods are eliminated. For instance, a foreigner in t, absent in t+l and naturalized in t+2 cannot be taken into account in the analysis.

Table 2.1. – Countries of origin in the sample

Country of origin	Number of observations	Percent
Portugal	9670	26.36
Algeria	6577	17.93
Italy	6227	16.97
Spain	4571	12.46
Tunisia	2124	5.79
Western Europe	1949	5.31
Eastern Europe	1518	4.14
Turkey	1266	3.45
Morocco	1056	2.88
South-East Asia	888	2.42
Southern Africa	839	2.29
Total	**36 685**	**100**

Source: EDP, Insee(1968-1999)

4.3.2. The causal effect of naturalization on employment: why isn't a simple regression model efficient?

Measuring the causal effect of naturalization on employment can't be achieved by a simple regression model (a univariate probit model for instance) that would use the naturalization variable among other co-variables explaining employment probability. It is in fact difficult to pinpoint the direct effect of naturalization because immigrants who gain French citizenship are not a sample randomly drawn among immigrants living in France. They differ from others in observable characteristics (education), but also through other characteristics, which are not observable (fluency in French) or even non observable at all (intelligence). Yet those characteristics also affect their probability of finding a job, and this needs to be taken into account. Thus, the possible positive impact of naturalization on employment estimated by a standard regression model can be due to the unobserved individual characteristics that affect significantly both phenomena. The magnitude of the impact measured by using an univariate probit model may thereby be "distorted" by what statisticians call an "endogeneity bias". It can be corrected by simultaneously estimating the probability of being naturalized and having a job period with a bivariate probit model. Because the data used in this study are longitudinal, the two phenomena can be situated in a chronological way; the naturalization is observed between two dates of the census while employment status is observed in the end of the inter-census period that is to say after the naturalization possible occurrence.

4.3.3. Some estimation considerations: the bivariate probit model

When several variables are analyzed simultaneously, the model used is a simultaneous equation model. In the case of naturalization and employment it is a double equation model. Since both of the variables analyzed in this example are dichotomous, a bivariate probit model is used: the residuals of both of the equations are supposed to be *N(0,1)* distributed. In the model used here, an additional difficulty lies in the fact that one of the dependent variables (namely naturalization) is supposed to affect the other dependent variable (namely employment). In such a model, called a discrete bivariate model with endogeneity, the identification of the parameters requires that the causal variable, that is to say the naturalization variable, depends from at least one additional variable that is not included in the set of the explanatory variables that affect the result variable, namely the employment one (Maddala 1983). These variables, excluded from the employment equation and included in the naturalization one, are called instrumental variables. They assure the identification of the model. However, instrumental variables cannot be valid if the parameters associated to their estimation aren't statistically significant in the naturalization equation.

Thus, If X_1 denotes a set of explanatory variables that affect the probability $Nat_{t,t+1}$ to be naturalized between *t* and *t+1* and X_2 a set of explanatory variables that affect the probability Emp_{t+1} to be employed on *t+1* (X_1 and X_2 can be identical), Z a set of instrumental variables supposed to affect the naturalization probability and to be non correlated to the employment one, U_1 and U_2 the random terms of each equation, the bivariate probit model can be written as in equation (6).

$$\left.\begin{array}{l} Nat_{t,t+1} = X_1\beta + Z\gamma + U_1 \\ Emp_{t+1} = Nat_{t,t+1} + X_2\beta + U_2 \\ avec \\ t \in (1968{,}1975{,}1982{,}1990) \\ \begin{pmatrix} U_1 \\ U_2 \end{pmatrix} \propto N\left[\begin{pmatrix} 0 \\ 0 \end{pmatrix},\begin{pmatrix} 1\rho \\ \rho 1 \end{pmatrix}\right] \end{array}\right\} \qquad (6)$$

In addition to the explanatory variables parameters, such a model gives an estimation of the correlation coefficient between the two residuals terms. If ρ is equal to zero, the causal variable is in fact exogenous and the estimation of its causal effect can be correctly made only by estimating the second equation by an univariate probit model. If ρ is different from zero, it is only this simultaneous model that can guarantee the estimation of unbiased parameters for the causal variable and the other co-variables.

4.3.4. Findings

Parameters of equation (6) are reported in Table C (Appendix). Models are estimated for men and women.

The naturalization equation controls for age, education, time period, matrimonial status and the country of birth. The employment equation controls for age, education, matrimonial status, size of the residential unity and the date of the census. Since one may think that the magnitude of the causal effect of naturalization may be more or less important among immigrant groups, an interaction is allowed between the possible naturalization occurrence during the inter-census period and the immigrants' country of origin.

Identification of the bivariate probit model is insured by restriction exclusions: some contextual covariates are introduced in the naturalisation equation and excluded from the employment one.

The number of foreigners in the local residential zone, called "départment" in French, is used as a proxy for the length of the "waiting line" for those applying for nationality. Foreigners are indeed implicitly considered as potential candidates for naturalization. This variable is calculated in the "départment" because it is the local administrative unit where the demands for naturalization are to be deposed. It is thus expected that the impact of this variable on the probability of naturalization will be negative: the more foreigners in the "départment", the longer the waiting line in the naturalization office and the lower the probability of naturalization between two census dates.

Another contextual variable is also introduced in the model: the size of the community of origin in the residential region. Its possible effect is related to the concept of community network used in several studies on the naturalization procedure (Portes and Curtis 1987 ; Yang 1994). According to these research works, the existence of a dense community network can facilitate the circulation of administrative information regarding the naturalization procedure and enhance thus the probability of naturalization of members.

Thereby these two contextual variables, both calculated at the beginning of the inter-census period of observation, are introduced in the model as instrumental variables: they are supposed to affect the probability of naturalization and not the probability of employment at the end the period. This assumption seems realistic because these covariates represent the local context seven or nine years before the immigrant's employment status is observed. Their significant coefficients, by the way concordant with the hypothesis formulated a few lines above assure that the estimation of the causal effect of naturalisation is consistent.

The results put forward the selective nature of naturalization: the probability to be naturalized raises with education and occupation. On the other hand, the effect of gaining nationality on employment probability turns out to be rather important. A simple way to present the results of this causal effect is possible by calculating a "naturalization premium" (table 5). For each group of immigrants, an average probability to be employed is calculated for those who were naturalized and those who weren't. This calculation is based on the model estimated parameters. The difference, for each immigrant group, between the probability to be employed when naturalized and the probability to be employed when non-naturalized measures an average "naturalization premium". On average, gaining French nationality increases the probability of being employed at the end of the period by almost 24 probability points for men and 23 for women. (Table 2.2).

Table 2.2. – Estimated marginal effect of naturalization on the employment probability by the country of origin (bivariate probit model)

Country of origin	Men	Women
Southern Africa	0,404	0,170
Morocco	0,353	0,302
South-East Asia	0,343	0,290
Eastern Europe	0,336	0,214
Tunisia	0,323	0,250
Algeria	0,268	0,293
Turkey	0,260	0,424
Italy	0,231	0,228
Western Europe	0,221	0,219
Spain	0,217	0,224
Portugal	0,154	0,176
Total	0,238	0,231

As table 2.2 shows, the naturalization premium is the most important for immigrant men coming from Southern-Africa or Morocco, while Turkish and North-African women seem to take the most important advantage of naturalization. Men and women coming from Western and Southern Europe also enjoy a "premium", nonetheless relatively small. The negative value of the correlation coefficient suggests that, given equal observable factors, immigrants who gain access to employment least easily, due to unobservable or non-measurable characteristics are those who probably stand the most to gain from becoming French citizens.

Thanks to the longitudinal structure of data, it was possible to estimate a direct causal effect of naturalization on employment. This effect, which proved to be rather important, especially for the immigrant groups facing a difficult socioeconomic situation, shows that gaining French nationality can significantly offset the extent of the discrimination some immigrants suffer from in the labour market. This study makes thus some advances in the analysis of the causal relations between the dimensions of the assimilation process. From this point of view, it echoes the theoretical work of Gordon using an empirical approach. While most theories consider the civic dimension of integration to be a sort of a crowning achievement of assimilation, the results presented here suggest that it is rather an important stage of the process that can enhance its socioeconomic dimension. By recognizing full citizenship to the immigrant aspiring to French nationality, the State greatly facilitates his or her mainstreaming into the labour market and society as a whole. These results measured and validated empirically give support to many studies in political sociology and political sciences that emphasized the role of citizenship in the assimilation process (Castles 1992 ; Heisler 1992 ; Castles 1995)

5. Conclusion

This article puts the stress on the usefulness of longitudinal data for immigration studies. When it is possible, following the same individuals over time is the best way to analyse the dynamic of the assimilation process and to contribute thus to the theoretical debate about assimilation by providing empirical evidence.

This article exposed some methodological issues linked to the use of such data through the presentation of two concrete examples extracted from research works on a French longitudinal dataset: the EDP. The first example shows the advances made by the use of longitudinal data in the interpretation of ethnic inequalities in the labour market. The durable disadvantages of some immigrants groups before employment can be explained only for a small part by uncontrolled individual characteristics and seems to be rather related to discrimination mechanisms. The second example emphasizes the conveniences of longitudinal data when dealing with causal interrogations in social sciences. When a proper methodology is applied, the longitudinal architecture of the EDP makes the estimation of a causal effect of naturalisation on employment possible. This effect seems to be the most prevailing for immigrants that experience the most disadvantaged situation in the labour market.

Nonetheless, even if EDP is one of the best available data in France to treat the immigrant assimilation process, its drawbacks shouldn't be underestimated. In addition to the lack of adaptation to the immigrant population, its principal shortcoming is related to the time period separating two observations for the same individual. Indeed, EDP is rather a "weak" example of longitudinal data since the observations are not repeated closely enough to properly track the dynamic of assimilation. This is all the more true when working on labour market problematic; it would have been much more convenient to observe the situation of employment each year or every other year for instance. However, the main purpose of this article is to show that longitudinal data is of great use to sociologists dealing with immigration issues, and hopefully, some new longitudinal data much more appropriate than EDP will soon change the shape of research on assimilation process. It is nevertheless a fact that the only way to make the most of the advantages this kind of data provides is by opening up to the methodological literature on panel data.

BIBLIOGRAPHIE

Abramson, Harold J. 1994. "Assimilation and Pluralism", p. 150-160 in Stephan. Thernstorn. *Harvard Encyclopedia of American Ethnic Groups*. Cambridge: Belknap Press of Harvard University Press.

Alba, Richard and Victor Nee. 1997. "Rethinking assimilation theory for a new era of immigration". *International Migration Review* 31:826-874.

—. 2003. *Remaking the American mainstream. Assimilation and Contemporary Immigration*. Cambridge, Massachussetts and London, England: Harvard University Press.

Arellano, Manuel and Bo Honoré. 2001. "Panel data models: some recent developments", p. 3265-3295 in James J. Heckman and Edward Leamer. *Handbook of Econometrics*: Elsevier Sciences B.V.

Baltagi, Badi H. 2005. *Econometric analysis of panel data*: Chichester.

Borjas, George J. 1990. *Friends or strangers: the impact of immigrants on the US. economy*. New York: Basic Books.

Castles, Stephen. 1992. "The Australian model of immigration and multiculuralism: Is it applicable to Europe?" *International Migration Review* 26:549-567.

—. 1995. "How nation-states respond to immigration and ethnic diversity". *New Community* 21:293-308.

Duncan, Otis Dudley. 1972. "Unmeasured Variables in Linear Models for Panel Analysis". *Sociological Methodology* 4:36-82.

—. 1980. "Testing Key Hypotheses in Panel Analysis". *Sociological Methodology* 11:279-289.

—. 1981. "Two faces of Panel Analysis : Parallels with Comparative Cross-Sectional Analysis". *Sociological Methodology* 12:281-318.

Fougère, Denis and Mirna Safi. 2005. "L'acquisition de la nationalité française : quels effets sur l'accès à l'emploi des immigrés?" *France, Portrait Social* Edition 2005-2006:163-184.

Gans, Herbert J. 1973. *Ethnic identity and assimilation : the Polish community*. New York: Praeger.

GELD, Groupe d études et de lutte contre les discriminations. 2000. "Une forme méconnue de discrimination : les emplois fermés aux étrangers".

Goodman, Leo A. 1973. "Causal Analysis of Data from Panel Studies and Other Kinds of Surveys". *The American Journal of Sociology* 78:1135-1191.

Gordon, Milton M. 1964. *Assimilation in American life: the role of race, religion, and national origins*. New York: Oxford University Press.

Granger, C. W. J. 1969. "Investigating Causal Relations by Econometric Models and Cross-spectral Methods". *Econometrica* 37:424-438.

Halaby, Charles N. 2004. "Panel models in sociological research". *Annual Review of Sociology* 30:507-44.

Hannan, Michael T. and Nancy Brandon Tuma. 1979. "Methods for Temporal Analysis". *Annual Review of Sociology* 5:303-328.

Heisler, Barbara Schmitter. 1992. "The future of immigrant incorporation : which models? which concepts?" *International Migration Review* 26:623-645.

Hu, Wei-Yin. 2000. "Immigrant Earnings Assimilation: Estimates from Longitudinal Data". *The American Economic Review* 90:368-372.

Jasso, Guillermina, Douglas S. Massey, Mark R. Rosenzweig, and James P. Smith. 2000. "The New Immigrant Survey Pilot (NIS-P): Overview and New Findings about U.S. Legal Immigrants at Admission". *Demography* 37:127-138.

Levine, Daniel B., Kenneth Hill, and Robert Warren. 1985. *Immigration Statistics: a Story of Neglect*. Washington DC: National Academy Press.

Lollivier, Stéphane. 2006. *Econométrie avancée des variables qualitatives*. Paris: Economica.

Lollivier, Stéphane and Daniel Verger. 2005. "Trois apports des données longitudinales à l?analyse de la pauvreté". *Economie et statistique* 383-384-385:245-281.

Maddala, G. S. 1983. *Limited Dependent and Qualitative Variables in Econometrics*. New York: Cambridge University Press.

Maines, David R. 1987. "The Significance of temporality for the development of sociological theory". *Sociological Quarterly* 28:303-311.

Math, Antoine and Alexis Spire. 1999. "Des emplois réservés aux nationaux ? Dispositions légales et discriminations dans l'accès à l'emploi". *Informations sociales, Droits des étrangers* 78:50-57.

Myers, Dowell and Cynthia J. Cranford. 1998. "Temporal Differentiation in the Occupational Mobility of Immigrant and Native-Born Latina Workers". *American Sociological Review* 63:68-93.

Park, Robert E and Ernest W. Burgess. 1921. *Introduction to the science of sociology*. Chicago: University of Chicago Press.

Petersen, Trond. 1993. "Recent Advances in Longitudinal Methodology". *Annual Review of Sociology* 19:425-454.

Portes, Alejandro. 1996. *The New Second Generation*. New York: Russell Sage Foundation.

—. 1997. "Immigration theory for a new century: Some problems and opportunities". *International Migration Review* 31:799-825.

Portes, Alejandro and John W. Curtis. 1987. "Changing Flags: Naturalization and Its Determinants among Mexican Immigrants". *International Migration Review* 21:352-371.

Portes, Alejandro and Min Zhou. 1993. "The New Second Generation: Segmented Assimilation and Its Variants". *Annals of the American Academy of Political and Social Science* 530:74-96.

Richard, Jean-Luc. 2004. *Partir ou rester?* Paris: Presses Universitaires de France.

Safi, Mirna. 2006. "Le processus d'intégration des immigrés en France : inégalités et segmentation". *Revue française de sociologie* 47:3-48.

—. 2007. "Le devenir des immigrés en France. Barrières et inégalités". EHESS.

Sobel, Micheal E. 2000. "Causal inference in the social sciences". *Journal of the American Statistical Association* 95:647-651.

Warner, W. Lloyd and Leo Srole. 1945. *The social system of American ethnic groups*. New Haven: Yale University Press.

Wold, Herman O. A. 1954. "Causality and Econometrics". *Econometrica* 22:162-177.

Wooldridge, Jeffrey M. 2005. *Introductory econometrics: a modern approach*: South-Western publishing.

Yang, Philip Q. 1994. "Explaining Immigrant Naturalisation". *International Migration review* 28:449-477.

ANNEXE

Table A – Longitudinal estimation of employment inequalities between immigrants and natives (random effect model)

		Men		Women			
		Employed/ Unemployed		Employed/ Unemployed		Participate/non-workii	
		Total sample	82-90-99 sample	Total sample	82-90-99 sample	Total sample	82-90-99 sample
Individual national status * census date	*Native*census date*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	Immigrant*1982	**-0,300*****	**-0,209*****	**-0,214*****	-0,198***	**-0,544*****	**-0,558*****
	Immigrant*1990	**-0,376*****	**-0,322*****	**-0,349*****	-0,373***	**-0,368*****	**-0,374*****
	Immigrant*1999	**-0,526*****	**-0,475*****	**-0,333*****	-0,329***	**-0,213*****	**-0,212*****
Education	*no diploma*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	Primary leaving certificate	0,155***	0,087***	0,066***	0,052***	0,364***	0,322***
	Exam taken at the age of 16	0,342***	0,319***	0,291***	0,333***	0,618***	0,607***
	Technical school certificate	0,350***	0,327***	0,285***	0315***	0,730***	0,728***
	High school diploma	0,484***	0,486***	0,612***	0,652***	0,942***	0,973***
	University diploma	0,735***	0,713***	0,926***	0,945***	1,217***	1,196***
Age		0,014***	0,021***	0,035***	0,044***	-0,022***	-0,015***
Matrimonial Status	*Single*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	Married	0,816***	0,787***	0,238***	0,216***	-1,137***	-1,114***
	Widow or divorced	0,129***	0,117***	-0,038**	-0,049**	-0,299***	-0,364***
Size of the residential area	*less than 20 000 inhabitants*	*ref.*	*ref.*	*ref.*	*ref.*	ref.	ref.
	between 20 000 and 100 000 inhabitants	-0,195**	=-0,191***	-0,104***	-0,093***	0,039***	0,061***
	More than 100 000 inhabitants	-0,224***	-0,195***	0,026**	0,042***	0,152***	0,154***
Census date	*1982*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*	*ref.*
	1990	-0,103***	-0,073	-0,274***	-0,289***	0,259***	0,331***
	1999	-0,378***	-0,492	-0,446***	-0,580***	0,711***	0,599***
Intercept		0,948***	0,817***	-0,038 ns	-0,291***	0,167***	1,706***
Standard deviation of the individual heterogeneities		0,832	0,814	0,791	0,771	1,329	1,218
Share of the individual heterogeneities in the total residual variation		**0,409**	**0,398**	**0,385**	**0,373**	**0,638**	**0,597**
Number of observations (i et t)		276701	180643	215945	156566	299300	203580
Number of individuals		126420	61731	103354	61908	131398	67860

Source: EDP, * 10% significance; ** 5% significance; *** 1% significance

Table B – Longitudinal estimation of employment inequalities among immigrants groups

		Men		Women			
		Employed/ Unemployed		Employed/ Unemployed		Participate/non-working	
		Total sample	82-90-99 sample	Total sample	82-90-99 sample	Total sample	82-90-99 sample
Country of origin* census date							
	*Spain *census date*						
Western Europe	1982	-0,058 ns	0,051 ns	-0,003 ns	0,214 ns	-0,497 ***	-0,595 ***
	1990	-0,181 ns	0,713 ns	0,139 ns	-0,052 ns	-0,801 ***	-0,681 ***
	1999	0,335 ns	0,351 ns	0,152 ns	0,047 ns	-0,726 ***	-0,716 ***
Eastern Europe	1982	-0,191 ns	0,010 ns	-0,526 ***	-0,473 **	0,048 ns	0,110 ns
	1990	-0,317 **	-0,239 ns	-0,088 ns	-0,064 ns	-0,204 ns	-0,026 ns
	1999	-0,700 ***	-0,742 ***	-0,507 ***	-0,516 ***	-0,155 ns	-0,120 ns
Italy	1982	0,108 ns	0,101 ns	-0,061 ns	0,138 ns	-0,148 **	-0,080 ns
	1990	0,125 ns	0,047 ns	0,018 ns	0,113 ns	-0,343 ***	-0,101 ns
	1999	0,103 ns	0,161 ns	-0,153 ns	-0,130 ns	-0,017 ns	0,076 ns
Portugal	1982	0,203 **	0,221 ns	0,169 *	0,174 ns	0,419 ***	0,416 ***
	1990	0,139 ns	0,289 *	0,325 ***	0,365 ***	0,310 ***	0,360 ***
	1999	-0,046 ns	0,008 ns	0,238 ns	0,275 **	0,395 ***	0,419 ***
Eastern-Asia	1982	-0,668 ***	-0,604 ***	-1,033 ***	-0,776 ***	0,170 ns	0,217 ns
	1990	-0,336 **	-0,492 **	-0,152 ns	-0,047 ns	0,110 ns	0,207 ns
	1999	-0,621 ***	-0,596 ***	-0,557 ***	-0,497 ***	-0,070 ns	-0,092 ns
Southern-Africa	1982	-0,586 ***	-0,420 *	-0,801 ***	-0,698 ***	-0,279 **	-0,153 ns
	1990	-0,581 ***	-0,813 ***	-0,240 ns	-0,043 ns	0,046 ns	0,028 ns
	1999	-0,734 ***	-0,575 ***	-0,148 ns	-0,125 ns	0,141 ns	0,162 ns
Tunisia	1982	-0,611 ***	-0,725 ***	-0,599 ***	-0,506 **	-1,131 ***	-1,069 ***
	1990	-0,589 ***	-0,899 ***	-0,670 ***	-0,643 ***	-1,190 ***	-1,042 ***
	1999	-0,939 ***	-0,932 ***	-0,526 ***	-0,465 **	-0,880 ***	-0,730 ***
Algeria	1982	-0,623 ***	-0,652 ***	-0,909 ***	-0,913 ***	-1,445 ***	-1,426 ***
	1990	-0,595 ***	-0,642 ***	-0,788 ***	-0,787 ***	-1,251 ***	-1,115 ***
Source: EDP, * 10% sign	1999	-0,902 ***	-0,800 ***	-0,920 ***	-0,896 ***	-0,854 ***	-0,774 ***
Morocco	1982	-0,497 ***	-0,711 ***	-0,780 ***	-0,754 ***	-0,849 ***	-0,860 ***
	1990	-0,393 **	-0,343 ns	-0,502 ***	-0,426 **	-0,985 ***	-0,908 ***
	1999	-0,446 **	-0,051 ns	-0,667 ***	-0,660 ***	-0,661 ***	-0,664 ***
Turkey	1982	-0,458 ***	-0,592 ***	-0,730 ***	-1,073 ***	-1,221 ***	-1,321 ***
	1990	-0,542 ***	-0,724 ***	-0,590 **	-0,759 **	-1,483 ***	-1,318 ***
	1999	-0,954 ***	-0,930 ***	-1,536 ***	-1,415 ***	-1,269 ***	-1,178 ***
Other countries	1982	-0,357 ***	-0,414 *	-0,382 **	-0,010 ns	-0,836 ***	-0,558 ***
	1990	-0,217 ns	-0,475 **	-0,174 ns	-0,202 ns	-0,729 ***	-0,438 **
	1999	-0,328 ns	-0,402 *	-0,036 ns	-0,149 ns	-0,854 ***	-0,693 ***
Education	*no diploma*						
	Primary leaving certificate	0,078 ns	-0,012 ns	0,016 ns	0,005 ns	0,384 ***	0,286 ***
	Exam taken at the age of 16	0,351 ***	0,330 **	-0,046 ns	0,110 ns	0,515 ***	0,551 ***
	Technical school certificate	0,174 ***	0,134 **	0,054 ns	0,084 ns	0,741 ***	0,722 ***
	High school diploma	0,262 ***	0,288 ***	0,188 **	0,349 ***	0,750 ***	0,785 ***
	University diploma	0,520 ***	0,595 ***	0,446 ***	0,536 ***	1,119 ***	1,223 ***
Age		0,007 ***	0,010 **	0,024 ***	0,029 ***	-0,017 ***	-0,009 *
Matrimonial Status	*Single*						
	Married	0,590 ***	0,629 ***	0,079 ns	0,100 ns	-1,531 ***	-1,373 ***
	Widow or divorced	0,271 ***	0,266 **	-0,044 ns	-0,016 ns	-0,694 ***	-0,590 ***
Size of the residential area	*less than 20 000 inhabitants*						
	between 20 000 and 100 000 inhabitants	-0,191 ***	-0,143 *	-0,050 ns	-0,051 ns	0,040 ns	0,037 ns
	More than 100 000 inhabitants	-0,219 **	-0,157 **	0,081 *	0,026 ns	0,303 ***	0,249 ***
Census date	*1982*						
	1990	-0,136 ns	-0,041 ns	-0,479 ***	-0,443 ***	0,609 ***	0,477 ***
	1999	-0,284 ***	-0,401 **	-0,390 ***	-0,406 ***	0,943 ***	0,752 ***
Intercept		1,227 ***	1,192 ***	0,526 ***	.281017 ns	1,772 ***	1,432 ***
Standard deviation of the individual heterogeneities		0,661	0,670	0,783	0,745	1,223	1,080
Share of the individual heterogeneities in the total residual variation		0,304	0,310	0,380	0,357	0,599	0,538
Number of observations (i et t)		22180	9933	11936	6958	20958	11067
Number of individuals		12294	3427	6761	3018	10698	3689

Source EDP, * 10% significance; ** 5% significance; *** 1% significance

Table C – Effect of naturalization on the employment probability

The naturalization equation

Variables	Modalities	Men		Women	
Intercept		-1,190	***	-0,791	***
Country of origin					
	Morroco	*Ref.*		*Ref.*	
	Southern Africa	0,021	ns	0,405	***
	Algeria	-0,822	***	-0,375	***
	South-East Asia	0,410	***	0,869	***
	West Europe	-0,764	***	-0,681	***
	Spain	-0,394	***	-0,290	***
	East Europe	-0,012	ns	0,121	ns
	Italy	-0,514	***	-0,489	***
	Portugal	-0,838	***	-0,642	***
	Tunisia	-0,346	***	-0,083	ns
	Turkey	-0,817	***	-0,572	***
Occupation					
	Blue collar	*Ref.*		*Ref.*	
	Farmers	0,278	***	0,414	**
	Craftsman randail trader	0,172	***	-0,197	ns
	Managers, executives	0,297	***	0,241	*
	Intermediate professions	0,292	***	0,186	*
	Office workers	0,168	***	0,206	***
	Non working	-0,013	ns	-0,011	ns
	Unemployed	0,155	***	0,147	**
Education	*No diploma*	*Ref.*		*Ref.*	
	Junior high school	0,241	***	0,282	***
	Vocational high school	0,338	***	0,511	***
	High School	0,302	***	0,537	***
	Post-secondery diploma	0,470	***	0,545	***
Time period	68-75	*Ref.*		*Ref.*	***
	75-82	-0,004	ns	-0,193	***
	82-90	-0,057	ns	-0,326	***
	90-99	0,138	***	-0,242	***
Age	between18 and 25 years old	*Ref.*		*Ref.*	
	between26 and 35 years old	0,087	**	0,025	ns
	between36 and 45 years old	0,179	***	0,084	*
	more than 46 years old	-0,262	***	-0,214	***
Matrimonial status	Single	*Ref.*		*Ref.*	
	Married	0,360	***	0,027	ns
	Widowed or divorced	0,333	***	0,162	**
Number of foreigners in the département		-0,060	***	-0,079	***
Relative size of the community in the region of residence		3,178	***	2,897	***

The employment Equation

Variables	Modalities	Men		Women	
Intercept		0,979	***	-0,397	**
Country of origin*naturalization					
	Morroco*non naturalized	*Ref.*		*Ref.*	
	Morroco*naturalized	1,549	***	0,905	***
	Afrique*non naturalized	-0,014	ns	0,407	***
	Afrique* naturalized	1,733	***	0,983	***
	Algeria* non naturalized	0,375	***	-0,099	ns
	Algeria* naturalized	1,931	***	0,902	***
	Asie* non naturalized	0,057	ns	-0,038	ns
	Asie*naturalized	1,467	***	0,827	***
	Europe de l'Ouest*non naturalized	0,487	***	0,343	***
	Europe de l'Ouest* naturalized	2,010	***	1,057	***
	Spain*non naturalized	0,396	***	0,454	***
	Spain*naturalized	1,947	***	1,094	***
	East Europe*non naturalized	0,025	ns	0,319	***
	East Europe* naturalized	1,567	***	0,940	***
	Italy*non naturalized	0,368	***	0,263	***
	Italy*naturalized	2,036	***	0,973	***
	Portugal*non naturalized	0,761	***	0,715	***
	Portugal*naturalized	2,422	***	1,166	***
	Tunisia*non naturalized	0,191	***	-0,008	ns
	Tunisia *naturalized	1,808	***	0,854	***
	Turkey*non naturalized	0,433	***	-0,160	ns
	Turkey*naturalized	1,774	***	1,251	***
Previous Employment status					
(at *t*)	En emploi	*Ref.*		*Ref.*	
	Chômeurs	-0,557	***	-0,746	***
	Inactifs	-0,806	***	-1,173	***
Census date at which employment					
is observed	1975	*Ref.*		*Ref.*	
(at *t+1*)	1982	-0,280	***	0,115	***
	1990	-0,474	***	0,274	***
	1999	-0,662	***	0,204	***
Education	*No diploma*	*Ref.*		*Ref.*	
	Junior high school	-0,008	ns	0,157	***
	Vocational high school	0,024	ns	0,221	***
	High School	0,000	ns	0,259	***
	Post-secondery diploma	0,106	*	0,311	***
Size of the residential unity	Less than 20 000 hab	*Ref.*		*Ref.*	
	between 20 000 and 100 000 hab	-0,054	ns	0,081	*
	More than 100 000 hab	-0,089	***	0,165	***
Age	between18 and 25 years old	*Ref.*		*Ref.*	
	between26 and 35 years old	-0,008	ns	0,152	ns
	between36 and 45 years old	-0,090	ns	0,349	**
	plus de 46 years old	-0,489	***	0,171	ns
Matrimonial status	Single	*Ref.*		*Ref.*	
	Married	-0,013	ns	-0,147	***
	Widowed or divorced	-0,156	**	-0,046	ns
Correlation coefficient		-0,888	***	-0,323	***