The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France

Yann Algan, Thierry Mayer, Mathias Thoenig

The Economic Incentives of Cultural Transmission: Spatial Evidence from Naming Patterns across France*

Yann Algan† Thierry Mayer‡ Mathias Thoenig§

March 26, 2013

Abstract

This paper aims at studying how economic incentives influence cultural transmission. We do so in the context of naming decisions, a crucial expression of cultural identity. Our focus is on Arabic versus Non-Arabic names given by parents to their newborn babies in France over the 2003-2007 period. Our model of cultural transmission disentangles between three determinants: (i) vertical transmission of parental culture; (ii) horizontal influence from the neighborhood; (iii) economic penalty associated with names that sound culturally distinctive. Our identification is based on the sample of households being exogenously allocated across public housings dwellings. We find that economic incentives largely influence naming choices: In the absence of economic penalty, the annual number of babies born with an Arabic name would have been more than 50 percent larger. Our theory-based estimates allow us to perform a welfare analysis where we gauge the strength of cultural attachment in monetary units. We find that the vertical transmission of an Arabic name provides the same shift in parents' utility as a 3% rise in lifetime income of the child.

*We thank Alberto Alesina, Alberto Bisin, David Laitin, Paola Giuliano, Laurent Gobillon, Camille Hémet, Yves Zenou, Fabrizio Zilibotti, for helpful comments. We also thank participants to seminars and conferences at RSAI in Miami, the NBER Summer Institute, CEPR conference on culture and economics in Rome, INED in Paris, UCD in Dublin, UAB and Pompeu Fabra in Barcelona, CIFAR IAST at the Toulouse School of Economics, Universities of Bern, ENSAI in Rennes, Lausanne, and Lugano, for their helpful comments. Yann Algan acknowledges financial support from the ERC Starting Grant 240923 under the European Community Seventh Framework Programme (FP7/2007-2013). Mathias Thoenig acknowledges financial support from the ERC Starting Grant GRIEVANCES-313327.

†Sciences Po and CEPR.

‡Sciences Po, CEPII and CEPR.

§University of Lausanne and CEPR.

1 Introduction

The aim of this paper is to study how economic incentives influence cultural choices made by individuals. While there is a growing literature looking at the effect of cultural background on economic outcomes, much less is known on the endogenous response of cultural traits to economic stimulus.[1] This is despite widespread interest and concern about economics-driven cultural change, such as the effect of global trade integration on cultural homogenization or the impact of technological progress on family values and gender roles to take only a few examples. This paper challenges the view that culture is exogenous with respect to economic factors.

Our approach to the question is to study whether there is room for economic determinants in the type of first name given by parents to their children. More precisely, we estimate whether the economic penalty associated with a name type on the labor market deters parents from transmitting their own cultural trait. The naming decision strikes us as a near-to-ideal object of inquiry in order to highlight the interplay between economic incentives and culture. The first name is a crucial marker of cultural identity. As stressed out by sociologists, the choice of first names is available to all parents, without material constraints, and thus is a pure expression of cultural identity (Lieberson, 2000). Fryer and Levitt (2004) have provided additional evidence on the cultural component of first names by showing that the surge in distinctively Black names in the US since the seventies could be associated to a rise in Black cultural identity. But the first name can also have direct economic consequences. In particular, various audit studies show that first names associated to a cultural minority are perceived negatively by employers (see the seminal study by Bertrand and Mullainathan, 2004). Quantifying how much this economic cost affects the cultural transmission decision is the main innovation of the paper.

While our paper is mostly empirical, we frame it around a random-utility discrete-choice model of naming decision. It incorporates the two traditional vertical and horizontal channels analyzed in the literature on cultural transmission (see Bisin and Verdier, 2001), to which an economic channel is added. The vertical transmission channel results from the utility gain for parents when transmitting their own cultural type; the horizontal transmission channel results from spatial externalities associated to the cultural type of peers and neighbors; and lastly the economic channel results from the utility gains/losses for children linked to the expected economic outcomes associated with their cultural type.

We estimate our theoretical model on the type/cultural category of first names given to babies born in France between 2003 and 2007. We specifically focus on the transmission of Arabic names as opposed to non-Arabic names in the French society. This type of name captures a combination

[1] We adopt here a popular definition of culture among economists, as being the belief, behavior or identity patterns that are transmitted from one generation to another (Fernandez, 2011). Cultural values, such as family values or trust have been found to affect a wide range of economic outcomes: consumption and savings (Carroll et al., 1994; Guiso et al., 2004), female labor force participation and living arrangements (Fernandez and Fogli, 2009; Giuliano, 2009), firm organization (La Porta et al., 1997; Bloom et al., 2012), trade (Guiso et al., 2009) economic development (Guiso et al., 2008; Tabellini, 2010; Algan and Cahuc, 2010), institutions and the demand for regulation (Alesina and Fuchs-Schundeln, 2007; Aghion et al., 2010, Luttmer and Singhal, 2011). Recent surveys include Guiso et al. (2006), Bisin and Verdier (2011), Fernandez (2011) or Algan and Cahuc (2013).

of elements that are associated with a culture potentially conflicting with the "traditional" (native) one: Religion, migration, political tensions, historical legacy with ex-colonies, or even consumption habits. A first major characteristic of Arabic names is to be associated with the most important population of immigrants (from Maghreb countries: Algeria, Morocco and Tunisia) in France after World War Two. Moreover, Arabic names in France are a signal of the Muslim religious affiliation since most of those names come from the Qu'ran, and the transmission of first names associated with the Qu'ran is a natural practice for religious people. By contrast, Non-Arabic Names in France are mainly associated with Saint Names that come from the French calendar of Christian Saints. Finally, Arabic names are associated with the most conflictual French decolonization episode–independence of Algeria–taking place after several years of violent war (1952-1964). Besides, it has been well documented that second generation immigrants from Maghreb face the highest penalty on the French labor market among the different immigrant groups (see Algan et al., 2010, and Duguet et al., 2010). Recent audit studies show that this labor market penalty is partly driven by pure cultural discrimination (Adida et al., 2010). We thus expect the transmission of Arabic names to raise an important trade-off between the desire to perpetuate one's own culture and the associated economic penalty inflicted to the offsprings due to what can be perceived as a lack of willingness to assimilate into the dominant culture.

We find that economic factors deeply shape the individual decision of cultural transmission. While the vertical channel plays a key role, we also find that parents do take into account the perceived economic cost attached to their cultural trait in the naming decision of their babies. The unemployment penalty associated with Arabic name holders (which we estimate to be around seven percentage points on the French labor market) significantly reduces the probability for parents to give such names. The magnitude of the effect is also quite sizeable. Our estimates imply that, in absence of any unemployment penalty for Arabic name holders, the number of babies born with an Arabic name would increase by more than 50 percent. By contrast, the horizontal channel is much less important. Finally, our theory-based estimates allow us to perform a welfare analysis. Focusing on the substitution rate between the vertical and the economic cost channels, we can express the strength of cultural attachment in monetary units. We find that allowing for the vertical transmission channel provides the same shift in parents' utility as a 3% rise in lifetime income of the child.

Our framework can also be used to assess the welfare effects of policies aiming at restraining cultural choice. We use the French regulation of first names as a natural example. Since Napoleon and the introduction of the Civil code in the early nineteen century, the choice of first names was restricted to Saint names, names from ancient Greece and Rome, and names from the Bible. This regulation was definitely removed in 1993, after several steps of gradual relaxation. We show that a counterfactual return to the strict version of such a policy ban on cultural expression would have very substantial welfare costs.

The empirical analysis is based on the French Labor Force Survey (LFS henceforth) that provides an unique opportunity to estimate the various channels of cultural transmission. First the LFS enables us to identify the vertical transmission channel by reporting the first names of *all* the household members, including children, in the sample. Second, the LFS also provides various proxies for the economic penalty associated with Arabic names. Third, the LFS data collection is based on a large representative set of 10541 sampling units spread all over the country, each unit consisting in a residential block of 20 *adjacent* households. This offers a considerable data-related improvement upon the existing empirical studies of peer/social influences, which rely on large areas (counties for instance) to define neighbors.

We use the spatial distribution of households for two purposes. First, as a measure for the horizontal transmission of naming choices from near neighbors. Second, we use the set of occupations held by neighbors, and the associated occupation-specific unemployment penalties, as a proxy for the *perceived* economic penalty. The idea is that parents form a belief about the future economic penalty associated with the name type of their child based on the information they can gather. Our hypothesis is that one of the main source of this information comes from people living in the same neighborhood. This is likely to be even more important for migrants, who arrive with a low initial knowledge on the local labor market. The use by parents of spatial information from neighbors' occupation rather than own's also mitigates the bias from self-sorting into occupation.

A salient issue with this identification assumption relates to the self-sorting of parents into residential neighborhoods, which might be leading to biased estimates of the horizontal and economic channels. We alleviate this concern by restricting our estimation to a sample of exogenously allocated households living in the French public housing sector. Due to an ideology deeply rooted in the French political system, and built into law, the government allocates state-planned moderate cost rental apartments to households without concern for their cultural background, mixing people indiscriminately. Furthermore, individuals rarely move, as the rents are much lower than market rates. We confirm with a variety of tests that spatial allocation *within* the public housing market can be considered to a large extent as quasi-random.

Our paper follows several strands of research. The first one deals with the transmission of cultural values and the formation of identity (Akerlof and Kranton, 2000). Bisin and Verdier (2001) provide a seminal model of cultural transmission distinguishing between vertical transmission by parents and oblique or horizontal transmission associated with social interactions. Tabellini (2008) and Guiso, Sapienza and Zingales (2008) model the interactions between norms and economic incentives in the intergenerational transmission of values like trust. The closest papers to ours is Bisin, Topa and Verdier (2004) and Bisin et al. (2010) who estimate structural models of transmission of religious values. Compared to this literature, our paper provides a new channel, associated with economic incentives.

The second strand of the literature focuses more precisely on the determinants of naming patterns. Sociology was the first social science to analyze these issues (Lieberson and Mikelson, 1995, is an early example). More recently, economists have sought to explain naming pattern, controlling more specifically for parents socio-economic background (Fryer and Levitt, 2004; Goldin and Shim, 2004; Head and Mayer 2008). Unlike in the influential paper by Fryer and Levitt (2004) looking at black names in the United States, the characteristics of the French labor market are such that the name-type of the minority (Arabic in our case) is associated with a significant economic penalty. Besides, we can identify economic and cultural variables at the individual and residential block level, while Fryer and Levitt (2004) have an aggregate proxy for economic background at the zip code level. We can therefore add to this strand of papers by looking at the trade-off between economic and cultural incentives in the inter-generational transmission of culture.

The remainder of the paper is as follows. Section 2 provides a detailed description of the data we use. Section 3 presents our theoretical model of naming decision, that will guide the estimated equation. This section also documents our identification strategy based on exogenous residential allocation of households in the public housing sector. Section 4 contains our estimation results. Section 5 quantifies the short-run and long-run contributions together with welfare effects of the vertical, horizontal and economic channels on cultural transmission. The long-run quantification looks at the dynamic and steady state effects of the channels.

2 Data

2.1 The French Labor Force Survey

Our empirical analysis is based on the French Labor Force Survey (LFS henceforth) from 2003 to 2007. The LFS is a representative survey of the French population, stratified at levels of around 3500 residential blocks per year, each block being defined as an average of 20 *adjacent* households. The LFS is a rolling panel of 6 quarters and all the households within a given block are interviewed every quarter. All the household members aged above 15 year old are interviewed and they report information on their first names and their socio-economic characteristics, including employment status (unemployed, inactive and employed), hourly wage and occupation. The occupation variable covers seven broad categories: farmer, craftsman, unskilled blue-collar, skilled blue-collar, clerk, intermediate, and executive. But the LFS also provides a more detailed classification of 29 occupations within those categories depending on the sector and infra-skill level of the occupation.

In addition, the survey records the first names of all the household members, including children below 15 years old. In particular, we observe the first-name given to any baby (i.e. less than 1 year old) born just before, or during, the 6 quarter period in which a household is surveyed. Moreover, the data collection being based on (very) close neighbors, the LFS provides a unique opportunity to understand the role of horizontal factors in the transmission of names. Given that the sampling unit in the LFS consists of small groups of adjacent households, and that all the members of the households within the same block are interviewed, we get detailed information on all the individuals living in the close neighborhood. Another important characteristic of the LFS is to distinguish between the public and the private housing sector. As discussed below, our identification strategy will be based on residential allocation of households within the public housing sector. We thus report henceforth information both on the total sample and on the sample of public housing residential blocks.

The time span of the rolling panel is too short (6 quarters) to exploit time variation in the socioeconomic composition within residential blocks. We thus keep one observation per member of the household.[2] Table 1 reports the main descriptive statistics of the full database when we use this selection criterion. Our total sample is made up of 10541 blocks, with 1535 blocks belonging to the state-owned housing market. The average block size is 18.31 adjacent households, each household being composed by around 3.31 members (babies, children and adults included). Overall, the total sample includes 425210 individuals, among whom 69458 are living in public housing.

Table 1: Descriptive statistics of the residential blocks

	Total sample	Public housing
Number of blocks	10541	1535
Number of blocks by department	174.35	45.12
Average number of households per block	18.31	17.99
Average number of members per household	3.31	3.70
Average number of children per household (aged below 15 years old)	2.19	2.40
Total number of households	173154	26749
Total number of individuals	425210	69458

2.2 Sample of babies' names

Our main variable of interest is the individuals' name type and the cultural background that is associated with it. In particular, we focus on the transmission of Arabic names, as opposed to non-Arabic names, in the French society. As explained in the introduction, this focus is motivated on two main grounds. First, Arabic names are associated with the most important population of immigrants in France – Maghreb – and to a lesser extent with the Middle East (other Arabic countries and Turkey), in the aftermath of the decolonization initiated in the 1960s. According to INSEE (2012), people with Maghreb origins (i.e. Algeria, Morocco, Tunisia) represent almost 60 percent of non-OECD migrants from first and second generations in France in 2008; this corresponds to 3.5 million individuals (1.6 millions for the first generation and 1.9 millions for the second generation) out of a total French population of 64.3 millions. Second, the Arabic names capture a cultural heritage that is potentially the most distinctive from the "locals", sometimes called *Français de Souche*, that is native French whose parents were also born in France. They are to a large extent a signal of the Muslim religious affiliation since most of those names come from the Qu'ran, and the transmission of first names associated with the Qu'ran is a natural practice for religious people. They are also associated in the French history with a hatred decolonization process linked to the independence war in Algeria. In our data we identify the Arabic first names

[2] In general this observation corresponds to the first wave of interview of the block. If a baby is born in the subsequent waves of interview, we explain the naming decision by the socio-economic characteristic of the household and of the block that prevailed at the time of the first interview. We allow for a gap up to one year (4 quarters) between the explained outcome, e.g the choice of a baby's name, and the explanatory variables.

by using the classification of Jouniaux (2001).

Table 2 displays the descriptive statistics of the sample of newborn babies. In our econometric analysis, we relate naming decision to various observable characteristics of the residential block where the parents live. For newborn children, the current block is the relevant one. We observe 3451 babies for whom we have all the relevant information on the parents' and blocks' characteristics. 3216 babies (90.8%) receive a non-Arabic names. Among them, 1879 babies (58%) are given traditional names, that is names that were already given in France in the early twentieth century.[3] Naming patterns that "sound" *Français de Souche* are thus still the most popular in the French society. Those traditional names are generally associated with Christian saint names, or names deeply ingrained in the French culture like Leo for boys or Manon for girls.

Table 2: Transmission of name types

	Babies with:	
	non-Arabic name	Arabic name
New Born (full sample):		
Parents with non-Arabic name	2982 (489)	80 (28)
Parents with Arabic name	234 (95)	245(132)
0-3 years old (2nd/3rd generation):		
Parents with non-Arabic name	416 (183)	111 (47)
Parents with Arabic name	658 (317)	789(461)

Note: This table reports the full number of babies born with the two name types and allocates them according to the name type of their parents. The top panel gives figure for the whole sample of babies born within the year. The bottom panel considers babies aged 0 to 3 at time of survey, born from at least one parent or grandparent with Arabic origins. In parentheses, the sample of babies in the public housing sector.

Among parents with Arabic names, the naming decision is rather balanced since 51.1 percent of those parents give an Arabic name to their offspring. Around one half of the parents from an Arabic cultural background transmit to their children a first name that sounds more traditional or more neutral relative to the French culture. In this latter case, they rarely choose saint names, but choose instead names that are culturally less distinctive. In particular, the two non-Arabic first names that are the more frequently selected are Adam or Yanis for boys, and Ines or Sarah for girls, names that seem to be attached to different cultures and are also given by the *Français de Souche* group of parents.

In contrast, among parents with non-Arabic names, the adoption of Arabic names is marginal, with a frequency of adoption of 2.8%. The top Arabic names given by those parents are Louna for girls and Rayan for boys. Those first names are rather neutral, they are hardly selected by parents with Arabic names.

Since we observe very little adoption of Arabic names by parents with non-Arabic names, we also look in our analysis at the pure transmission decision of giving an Arabic name when it is part of the original culture. We focus on households where at least one parent or grandparent is

[3]To identify those, we use INSEE's national database called "fichier des prénoms".

a national from an Arabic country. The babies are thus born in France, but the parents (babies of second generation) or the grandparents (babies of third generation) were born in an Arabic country. Since this selection criterion would leave us with a too small sample of babies, specially in the public housing sector, we consider children between 0 and 3 years old instead of just newborn babies to carry out this analysis. While the use of this expanded sample of children could be a source of statistical noise in our econometric analysis, Section 4 shows that this problem is of little concern in practice.[4]

2.3 Employment penalty associated with Arabic names

The goal of this section is to document the extent of the economic penalty attached to Arabic first names across occupations. It should be clear that this investigation is only a first step for us, our final goal being in Section 4 to analyze how the information and perception about this economic penalty affect the individual decision to transmit a name type. We therefore keep our estimation method voluntarily simple in this first step. It is however interesting to note that our estimates of the penalty attached to Arabic first names are very much in line with the existing findings based on more elaborate econometric methods.[5] Another important point is that an accurate estimate of discrimination associated with an Arabic name—beyond its intrinsic complexity to obtain—is in fact not crucial for the purpose of this paper. Indeed , as will become clear below, our identification strategy is not based on the true penalty but rather exploits the *subjective* expected penalty that is perceived by parents after retrieving information from their neighbors (see equation 4 below).

Table 3 displays the unconditional unemployment rate associated with the 10 most popular non-Arabic and Arabic names. The data derives from the population aged 25-55 in the LFS 2003-2007. The unemployment rate of men with popular Arabic names is between four to eight times as high as the unemployment rate of men with popular non-Arabic names. A striking example is given by men named Abdelkader, whose unemployment rate reaches 37 percent against 5 percent for individuals named Philippe.[6] The unemployment gap is even more pronounced among the female population. Women named Fatma (ranked 6th in the list of the most popular Arabic names for women) have an average unemployment rate of 42 percent, against 10 percent for women named

[4]There will be an issue if the child changes neighborhood between age 1 and 3, a relatively rare event in our sample.

[5]In particular, Duguet et al. (2010) use pair auditing to test access to job interviews of individuals who share the same characteristics, except Arabic and non-Arabic names. They find that the probability to get interviewed is 7 percentage points lower for Arabic name holders in the French labor market, which is really close to our results. Adida et al. (2010) isolate the source of discrimination by identifying the effect of being Muslim on the French Labor Market. Using a large-scale survey on immigrants from Senegal, they are able to identify typical first names from the Muslim and the Christian parts of this population, which they report to be otherwise quite similar on all measurable aspects. The authors then run an audit survey with CVs identical in all dimensions, but with a different first name. The CVs would in particular have the same family name, for instance Diouf, a typical Senegalese one. But one CV would have a typical Muslim first name (for instance Khadija for women) and the other with a well-known Catholic first-name (Marie). Adida et al. (2010) find a statistically significant difference of 13 percentage points in the response's rates to job applications between the holders of Catholic first names and those with Muslim first names.

[6]Interestingly, emir Abdelkader was the military leader who led the struggle against the French colonial invasion of Algeria in the mid-19th century (we thank Nour Meddahi for pointing that out during a presentation of this paper).

Sandrine or Patricia (who have the highest unemployment rate among the most popular non-Arabic names). The cross-name heterogeneity is also much larger for Arabic names but this is probably driven by small sample issues.

Table 3: Unemployment rate by name

	Name type			
Rank of name	Non-Arabic		Arabic	
1	Philippe	0.05	Mohamed	0.19
2	Alain	0.05	Said	0.30
3	Christophe	0.07	Rachid	0.20
4	Frederic	0.07	Ali	0.20
5	Patrick	0.07	Abdelkader	0.37
6	Michel	0.05	Karim	0.17
7	Thierry	0.05	Ahmed	0.18
8	Pascal	0.05	Mustapha	0.18
9	Laurent	0.06	Kamel	0.20
10	Stephane	0.08	Farid	0.29
1	Nathalie	0.08	Nadia	0.18
2	Sylvie	0.07	Fatima	0.26
3	Isabelle	0.08	Malika	0.24
4	Catherine	0.07	Aicha	0.26
5	Christine	0.06	Naima	0.15
6	Martine	0.08	Fatma	0.42
7	Valerie	0.08	Khadija	0.17
8	Sandrine	0.10	Rachida	0.37
9	Veronique	0.08	Samira	0.31
10	Patricia	0.10	Yamina	0.36

Notes: The sample covers the 4 years of employment survey we have access to (2003-2007). The statistics are for adults between 25 and 55 years old.

Table 4 breaks down the unconditional unemployment penalty associated with Arabic names by occupation. The average rate is also reported. For the sake of clarity (in this table only), we group together the 29 different occupations listed by INSEE into 7 main categories: farmer, craftsman, unskilled blue-collar, skilled blue-collar, clerk, intermediate, and executive. On average, Arabic name holders have an unemployment rate of 20 percent, around three times as high as the unemployment rate of non-Arabic name holders (7 percent). But this average comparison hides a lot of variance across occupations. The unemployment rate of Arabic name holders among executives is only 7 percent and the unemployment gap with non-Arabic name holders falls to 3 points for this occupation. In contrast, the unemployment rate of Arabic name holders reaches 29 percent among the (unskilled) blue collars, which represents an unemployment gap of 14 percentage points with the non-Arabic name holders belonging to the same occupational category. This variation across occupations suggests to estimate the employment penalty associated with an Arabic name at a very detailed level among the 29 occupations registered in the French national statistics.

Table 5 documents the *conditional* employment penalty by running a standard Mincer-type

Table 4: Unemployment rates by Name type and Occupation

	Unemployment rate:	
	Arabic name	non-Arabic names
Executive	0.07	0.04
Intermediate	0.14	0.05
Clerk	0.20	0.09
Blue collar (skilled)	0.20	0.07
Blue collar (unskilled)	0.29	0.15
Craftman	0.15	0.04
Farmer	0.10	0.00
Total	0.20	0.07

Notes: The sample covers the 4 years of employment survey we have access to (2003-2007). The statistics are for adults between 25 and 55 years old.

equation estimated on the LFS subsample of active persons aged between 25 and 55. The left-hand-side variable is the employment status, equal to 1 if the respondent is employed, and 0 if unemployed. Logit regressions are performed with marginal effects reported in the table. We consider a set of standard controls, including nationality at birth of the respondent and parent's respondent, individual characteristics (age, age squared, gender, marital status and number of children), educational, occupational, spatial and year fixed effects. Our variable of interest is *arabic name*, a binary variable coding for a first name from Arabic origins. Column (1) reports the employment penalty associated with an Arabic name, without controlling for the nationality at birth of the respondent and of the parents' respondent. Holding an arabic name decreases the probability to be employed by 10 percentage points and the effect is statistically significant at the 1 percent level. However, most of Arabic name holders being first or second generation migrants, the previous correlation captures both the discriminating impact on the labor market of foreign origins and of foreign names; while closely related, the latter dimension is manipulable by parents but the former is not. To isolate the specific penalty from a name that sounds culturally distinctive, we also control for other attributes of the country of origin. Column (2) includes a dummy variable equal to 1 if the nationality at birth of the respondent or of the parents' respondent is from an Arabic country, and 0 otherwise. The estimated employment penalty associated with an Arabic name remains fairly high at 7 percentage points and remains highly statistically significant. The estimated employment penalty associated with an Arabic Name is of the same order of magnitude as the one associated with having an Arabic-related nationality, suggesting that a specific employment penalty is attached with the first name.[7] In the next two columns, the sample is restricted to individuals living in the public housing sector (on which our main econometric analysis will be based). Column (3) shows that the conditional employment penalty is identical for this subsample

[7]This might reflect the absence of clear morphological markers of ethnicity (e.g. skin color, size) for individuals with Arabic origins and living in France. Hence the first name conveys meaningful information on the ethnical background.

and is robust to the inclusion, in Column (4), of a variable coding for the number of children *with* Arabic name in the household. This last variable is likely to be correlated to a bundle of *unobservable* characteristics related to the degree of attachment of the individual to the Arabic culture (e.g. religiosity) that may simultaneously influence the penalty. Column (5) estimates the conditional employment penalty for each broad occupational category. The reference category is executives.

How big is the implied loss in lifetime expected income associated with an Arabic name? A simple "back of the envelope" calculation suggests that the loss is substantial. Breuil-Grenier's (2001) provides detailed estimates of the income variation induced by a transition from employment to unemployment on the French labor market (including in particular social benefits, that we do not observe in the LFS). She finds an average income loss of 50 percent. From Table 5 - Column (2), we know that every period, the conditional unemployment gap of Arabic name holders is 7 percentage points relative to non-Arabic name holder. This means that the total income loss of typical Arabic name holder during their active life is $0.07 \times 0.5 = 3.5\%$ of expected income. Since the average participation to the active population is 39 years in France, this is equivalent to $39 \times 0.035 = 1.365$ years, i.e. around 16 months of income.

Table 5: The penalty of an Arabic name

	(1)	(2)	(3)	(4)	(5)
Dep.Var:		emp. / unemp. status			
arabic name	-0.10^a	-0.07^a	-0.07^a	-0.06^a	0.00
man	0.01^a	0.01^a	0.03^a	0.03^a	0.01^a
age	0.01^a	0.01^a	0.01^a	0.01^a	0.01^a
age squared	-0.00^a	-0.00^a	-0.00^a	-0.00^a	-0.00^a
nationality from Maghreb/Middle-East		-0.06^a	-0.07^a	-0.06^a	-0.06^a
count of kids with Arabic name				-0.01^b	
arabic × intermediate					-0.06^a
arabic × clerk					-0.08^a
arabic × blue collar (skilled)					-0.09^a
arabic × blue collar (unskilled)					-0.09^a
arabic × craftsman					-0.08^a
arabic × farmer					-0.04
Observations	148582	148582	90693	90693	148582
R^2	0.041	0.042	0.047	0.047	0.049

Note: Column (5) has executives as the baseline occupation group. All regressions include dummies for education level, occupation group, *département* of residence, years, as well as number of children, and marital status. The sample includes active persons aged between 25 and 55. The unconditional unemployment rate in this sample is 8%. Standard errors are in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%

3 Model and Identification of Naming Decision

3.1 A simple model of baby name choice

In this section we build a random utility discrete choice model of baby naming decision. Our framework is rich enough to highlight the underlying estimation issues while remaining sufficiently tractable for structural estimation. The parental decision under scrutiny is binary and relates to the cultural type attached to the baby's first name. The utility for a household i, living in residential block k, derived from choosing a given name type for its baby born in year t is defined as $U_{ik,t}(1)$ if the name is Arabic and $U_{ik,t}(0)$ otherwise,

$$U_{ik,t}(\text{Baby}) \equiv V_{ik,t}(\text{Baby}) + \epsilon_{ik,t}(\text{Baby}), \tag{1}$$

where Baby $\in \{0, 1\}$ denotes alternatives, $V_{ik,t}(\text{Baby})$ is the observed part of utility and $\epsilon_{ik,t}(\text{Baby})$ is the unobserved parental-specific random shock across alternatives.

In such a discrete choice setting, only differences in utility over alternatives can be identified from the data. The econometrician observes a parental choice $\text{Baby}_{ik,t} = 1$ if and only if $\Delta U_{ik,t} \equiv U_{ik,t}(1) - U_{ik,t}(0) \geq 0$. Let us denote the difference in the observed part of utility as $\Delta V_{ik,t} \equiv V_{ik,t}(1) - V_{ik,t}(0)$, and the difference in unobserved utility as $\varepsilon_{ik,t} \equiv \epsilon_{ik,t}(1) - \epsilon_{ik,t}(0)$, such that

$$\begin{aligned}
\Delta U_{ik,t} &= \Delta V_{ik,t} + \varepsilon_{ik,t} \\
&= \alpha_0 + \alpha_1 \underbrace{\text{Parents}_i}_{\text{Vertical}_i} + \alpha_2 \underbrace{\mathbb{E}\left[\frac{1}{\mathcal{N}_{k,t}} \sum_{j \in k, j \neq i} \text{Baby}_{jk,t}\right]}_{\text{Horizontal}_{k,t}} + \alpha_3 \underbrace{\mathbb{E}[\mathcal{C}_{ik}]}_{\text{Economic Cost}_{ik}} + \varepsilon_{ik,t},
\end{aligned} \tag{2}$$

where $\Delta V_{ik,t}$ is specified as a three part linear function, which we label "Vertical", "Horizontal", and "Economic cost" channels of influence. Parents_i is a parental characteristic equal to one when the name of one of the two parents is Arabic and zero otherwise (with alternative definitions investigated in Section 4), $\text{Baby}_{jk,t}$ codes for choices of names among the $\mathcal{N}_{k,t}$ other babies born in t and living in residential block k and $\mathbb{E}[\mathcal{C}_{ik}]$ is the *perceived* economic penalty that parents i expect to be attached to their baby if they choose an Arabic name.

The first RHS component therefore corresponds to the desire by parents to transmit their own cultural type (measured by the coefficient α_1). Our specification of utility is flexible as it allows both for cultural transmission and cultural adoption. Transmission is the case where the names of parents and babies belong to the same cultural type. Adoption corresponds to the two other cases: e.g. parents with Arabic names that do not transmit their cultural type to their baby or parents with non-Arabic names adopting an Arabic name for their baby. Both patterns are observed in the data although the latter is less salient (see Table 2).

The second RHS component reflects social influence, i.e. the share of parents of newborn babies in residential block k expected to make the same choice as i, with intensity α_2 expected to be positive. In our data, the block k is small enough that household i is not negligible and this creates the classical Manski (1993) reflection problem. We assume that parents i form their expectations on lagged decisions of neighbors[8]:

$$\mathbb{E}\left[\frac{1}{\mathcal{N}_{k,t}}\sum_{j\in k, j\neq i}\text{Baby}_{jk,t}\right] \equiv \frac{\sum_{\tau=1}^{\Upsilon}\sum_{j\in k,j\neq i}\text{Baby}_{jk,t-\tau}}{\sum_{\tau=1}^{\Upsilon}\mathcal{N}_{k,t-\tau}}, \tag{3}$$

that is they expect the current choices of neighbors to be on average similar to the ones taken since year $t-\Upsilon$ (we will take $\Upsilon = 10$ in our application).

The third RHS component relates to economic incentives: Presumably, the higher the expected penalty is, the less parents want to attach an Arabic cultural type to the name of their babies. The comparison of the coefficients α_1 and α_3 reflects the parental tradeoff between their own attachment to a particular cultural type and their altruistic concern towards the future economic performance of their babies. The *perceived* expected penalty, $\mathbb{E}[\mathcal{C}_{ik}]$, is sensitive to the parental information set and to a wide set of observed and unobserved parental characteristics influencing the future spatial and social mobility of the baby. Our identification strategy exploits the fact that part of this information set is based on information on the labor market that households retrieve from social interactions and communication with their neighbors. The idea is that parents surrounded by neighbors working in occupations with high levels of penalty tend to update upwards their beliefs on the extent of the penalty. Formally, the *perceived* expected penalty is broken down into a block-specific informational component and an unobserved parent-specific residual component:

$$\mathbb{E}[\mathcal{C}_{ik}] = \sum_{l\in\mathcal{O}}\omega_{lk}\times\hat{\gamma}_l + u_i, \tag{4}$$

where $\mathcal{O} =$ is the set of occupations, ω_{lk} is the share of neighbors in block k working in occupation l, $\hat{\gamma}_l$ is a proxy for the labor market penalty observed in occupation l and u_i is the unobserved residual parent-specific part. In the remainder of the paper, $\sum_{l\in\mathcal{O}}\omega_{lk}\times\hat{\gamma}_l$ is labeled as the *local information on penalty.*

Combining (2), (3) and (4), utility becomes

$$\begin{aligned}\Delta U_{ik,t} &= \Delta\mathcal{V}_{ik,t} + \delta_{ik,t} \\ &\equiv \alpha_0 + \alpha_1\text{Parents}_i + \alpha_2\frac{\sum_{\tau=1}^{\Upsilon}\sum_{j\in k,j\neq i}\text{Baby}_{jk,t-\tau}}{\sum_{\tau=1}^{\Upsilon}\mathcal{N}_{k,t-\tau}} + \alpha_3\sum_{l\in\mathcal{O}}\omega_{lk,t}\times\hat{\gamma}_l + \delta_{ik,t}, \end{aligned} \tag{5}$$

where $\Delta\mathcal{V}_{ik,t}$ is the observable utility and $\delta_{ik,t} \equiv \alpha_3 u_i + \varepsilon_{ik,t}$ is the new error term.

In this equation, our source of identification is based on neighbors from the residential block. Neighbors serve both as: i) A source of peer-pressure for the horizontal transmission channel,

[8] An alternative way to mitigate the problem would be through IV strategy, where natural instruments of parental expectations are some observable characteristics of the neighbors. Those should not influence the choice of household i *directly* but through the choice of surrounding parents j (Glaeser and Scheinkman, 2001).

and ii) A source of information for the perceived expected penalty associated with a name. Since individuals tend to self-segregate, e.g most households choose their location, our estimates of equation (5) could be biased by endogenous residential sorting. To address this bias, we restrict our estimations to a subsample of households allocated exogeneously, namely those living in the public housing sector. In the next section, we document step by step the issues raised by the identification of equation (5). In Section 3.3 we document and test the fact that the residential allocation within the French public housing sector is quasi-random and exogenous with respect to ethnic characteristics.

3.2 Estimation issues

The horizontal transmission channel raises several estimation issues well-known in the social interaction literature (for a recent survey of discrete choice models with social interactions, see Blume, Brock, Durlauf and Ioannides, 2011). Indeed, in equation (5), the realizations of $\text{Baby}_{jk,t-\tau}$ depend on $\Delta U_{jk,t-\tau}$. *Spatial sorting* might lead to a non-zero correlation between $\delta_{ik,t}$ and $\delta_{jk,t-\tau}$ for households i and j belonging to the same residential block k. This would create a correlation between $\text{Baby}_{jk,t-\tau}$ and the error term in (5), $\delta_{ik,t}$, potentially capturing unobservable taste shocks for the considered cultural type common to households i and j. For example, it is clear that the degree of religiosity of the household, which is unobserved by the econometrician, affects positively the choice of an Arabic Name for the baby; moreover religious people tend to live in the same residential areas (e.g. close to a Mosque or to Halal shops). This example makes it clear that spatial clustering of Arabic names is not only driven by horizontal transmission but might also be partly driven by unobserved characteristics of the area. Our estimates could thus be biased by the endogenous spatial sorting of households. To get rid of this bias, we identify the coefficient α_2 on the subsample of households that are exogenously allocated across the different public housing blocks within *départements*.[9] This *within* identification strategy calls for including *département* fixed effects in all specifications.

The coefficient α_3 associated with the economic cost of a name type is also potentially ill-estimated due to self-selection into occupations and locations by parents. The concern is that religious (muslim) parents, attached to the transmission of the Arabic type to their offspring, tend to work in occupations with low discrimination, and are located in residential blocks with religious neighbors working in non-discriminating occupations. We address this issue first by controlling for parental occupation and education fixed effects. Although parental occupation is not a random choice, the inclusion of fixed effects for the parental occupation captures all time-invariant co-determinants of the parental occupation choice and the naming pattern. Second, rather than using the parental occupation as a source of information on the perceived expected penalty, we use the block-specific informational component of the penalty. Thus the key issue is now that of the exogeneity of the composition of occupations within the residential block. We identify the coefficient α_3 by restricting once again our estimates to the subsample of exogenously allocated

[9] Metropolitan France is divided into 95 administrative areas, called *départements*.

households living in the public housing sector within a given *département*.

As discussed earlier, our goal is not to identify accurately the discrimination associated with an Arabic name. Our model is based on the *subjective* expected penalty of an Arabic name that is perceived by parents from the observation of their neighbors who live in the very same housing block. In our empirical application we proxy $\hat{\gamma}_l$, the *subjective* labor market penalty perceived in occupation l, as the observed unconditional unemployment gap between Arabic and non-Arabic name holders. Parents just observe among their Arabic and non-Arabic neighbors those who work and those who do not. Yet, the unconditional unemployment gap might be an imperfect measure of the information truly used by the parents to assess the labor market penalty, leading to an attenuation bias in the estimation of α_3. In particular, parents might use additional information from the observed characteristics of their neighbors, such as the education or the country of origin, to assess the specific penalty associated with an Arabic name. We will consequently consider an alternative measure of $\hat{\gamma}_l$ based on the *conditional* unemployment gap as computed in Table 5, taking into account additional observable characteristics. The comparison of the estimates obtained with each of the measure (Columns 3 and 6 in Tables 8 and 9) confirms a very light, but statistically insignificant, increase in α_3 when estimated with the second measure, showing that this attenuation bias is unlikely to be a first-order issue.

3.3 Identification with exogenous spatial residential allocation in public housing

As discussed above, the estimation of the horizontal and economic channels of cultural transmission raises the issue of endogenous residential sorting of households. Individuals might tend to self-segregate if they prefer to live close to neighborhoods with whom they share common characteristics, in particular people from the same ethnic and socio-economic backgrounds. We draw on the previous literature using public housing as a source of exogenous spatial variation to identify the neighborhood effects on socio-economic outcomes (see, among others, Oreopoulos in the Canadian context, 2003; and Goux and Maurin in the French context, 2007). We document that the French public housing provides a source of exogenous residential allocation to mitigate this bias. Since public housing is administrated at the *département* level [10], we investigate whether households are allocated exogenously across the different public housings within a given *département*.[11]

3.3.1 Formal allocation process in public housing

We start by documenting the actual process of allocation of households across public housing dwellings. More details on the institutional and legal aspects are provided in Appendix C.1. The

[10] Metropolitan France is divided into 22 large administrative areas, called *régions*, and into 96 smaller administrative areas, called *départements*. Each *département* is hence a subdivision of a region, and several *départements* can belong to the same region. Each *département* is administered by an elected General Council (*Conseil Général*) and its President, whose main areas of responsibility include the management of a number of social and welfare programs, junior high schools, buildings and technical staff, local roads, schools, rural buses, and municipal infrastructure.

[11] See Algan, Hémet and Laitin (2012) for an in-depth analysis of the exogeneous allocation of households, with respect to their ethnic characteristics, across the public housing dwellings in France

main eligibility requirements for admittance into the public housing sector are to be legally living in France (as a French citizen or migrant with a valid residence permit) and to be living under a certain threshold of income per unit of consumption. This income ceiling is usually rather high: in 2009, this threshold was between 36748 and 50999 Euros per year for a four-person family, depending on the *département* of residence. Jacquot (2007) estimates that around two thirds of households living in Metropolitan France could apply for a public housing unit. Moreover the rents are considerably lower in public housing than in private housing. As a result, there is a strong excess demand for public housing. Just for the case of Paris, there were 121937 ongoing applications, to be compared to 12500 public housing units allocated over the year 2010. Due to those stringent constraints, other eligibility criteria are taken into account : the family situation and household size (to ensure a suitable match with the characteristics of vacant dwellings), as well as the emergency of the application.[12]

The selection committees in charge of allocating households to vacant public housing dwellings are held at the *département* level. For each vacant housing unit, at least three households must be considered by the committee members, who finally decide which household will be allocated to which housing unit according to the eligibility and priority criteria detailed above. The application form contains very limited information about the ethnicity of the applicant: he or she only needs to inform about his or her nationality, which is limited to three possible categories (French, European Union, or non European Union). Legally, applicants can refuse up to three offers but in practice they rarely do, given the large opportunity cost of declining an offer. This makes it unlikely that the selected households could be really picky about the characteristics of their neighborhood. It is also formally possible to indicate a precise neighborhood within the *département* in the application form, but in practice, very few applicants (6.6 percent) do provide this information probably due to the fear of being rejected on this ground. Residential mobility within the public housing sector is very low, due to the current strong shortage in the supply of public housing dwellings. People who move are seeking larger space following an increase in their household size (only 12 percent of the public housing dwellings have more than three rooms).

In short, residential self-sorting on ethnic or socio-economic characteristics is not a common practice within public housing (Simon, 2003). The public housing market is very tight, and highly regulated. This implies that households have very limited control over the time when they will be assigned to a public housing dwelling and the precise place within a *département* where they will be located.

3.3.2 Statistical tests of exogenous residential allocation in public housings

We now provide more formal statistical tests for the exogenous spatial allocation of households across public housing residential blocks within a *département*.

Table 6 provides a first approach where, for various observable household characteristics, we

[12] Five priority criteria—none related to nationality—are defined by law to make sure that vacant housing will first be distributed to households with obvious social difficulties (see Appendix C).

test for the difference in means between residential blocks. We regress, for each *département* taken separately, each observable on a battery of fixed effects associated with the different residential blocks located in this *département*. Those regressions are performed on the subsample of household heads who are living in public housing. The exogenous allocation test consists in performing a standard F-test on the null hypothesis that the fixed effects are jointly not statistically different from zero. In the case of endogenous residential sorting in some public housing blocks, the fixed effects associated with those blocks should be statistically significantly correlated with the household characteristics, and the F-test will be rejected. We consider the two main characteristics used in our model of spatial information, e.g the ethnic and occupation composition of the residential block. We focus on the Arabic origin of the respondent's name, the nationality at birth (set to one if the nationality of the respondent or of the respondent's parent is from France, and zero otherwise) and the occupation of the respondent (coded as a binary variable equal to one for blue collars). Column (1) of Table 6 reports, for each observable, the share of *départements* for which the F-test is not rejected at the 10 percent level (a more conservative criterion in our context than the standard 5 percent level). For the sake of comparison, we run, in Column (2), the same F-test on the full sample of household heads, including both those who live in the public and private housing sectors. In this case, as expected, endogenous residential sorting is much more salient.

Table 6: F-Test of Residential Sorting

% departments without residential sorting relative to households' characteristics

	Public Housing	Total Sample
Household's characteristics:		
Nationality from Maghreb/Middle East	77.35	55.55
French Nationality at Birth	81.88	59.52
Occupation: blue collar	80.32	54.22

Note: The table reports the share of *département* for which F-tests (at the 10 percent level) do not reject the null-hypothesis of a null correlation between observable characteristics and residential block fixed effects. The F-test are based on a logistic regression of household characteristic on public housing fixed effect within each *département*. The sample includes household heads aged over 15 years old.

Our previous approach to testing exogenous spatial allocation has the advantage of simplicity but might not be ideally suited to our empirical context where the spatial units under consideration are small. Indeed, in the LFS, the average residential block is composed of only 18.31 household heads. In this type of context, as firstly pointed out by Ellison and Glaeser (1997), parametric test of spatial allocation/concentration (that assume independent location choices) might be ill-defined to test the null hypothesis of exogenous or random allocation.[13]

We thus propose an alternative approach. We perform a Monte Carlo simulation generating

[13] As mentioned previously, the allocation of households across public housing blocks takes place at the *département* level. If the members of the public housing committee strictly follow the legal criteria and do not take into account the ethnic characteristics in the allocation process, we should find an uniform distribution of households of a given

artificial random allocations that we later compare to the observed allocation. More precisely, for each *département*, we pool the public housing population and reallocate it randomly, without replacement, across the different residential blocks of the corresponding *département*, maintaining unchanged the actual size of each block. We get a simulated random distribution of individuals with a given characteristic across blocks. We then run a Kolmogorov-Smirnov (KS) test of equality of distribution with the actual spatial distribution. The final step calculates the percentage of *départements* for which the actual and simulated distributions across housing blocks are similar, i.e. those for which we cannot reject the null hypothesis of equality of the distribution at the 10 percent level. We run 100 draws of the Monte Carlo simulation. For each draw we compute the tests for the equality of distributions.

Table 7 shows the values of those tests averaging over 100 Monte Carlo draws. The labels in the first column indicate the characteristic under consideration: ethnic background (Arabic origins or French nationality at birth) and occupation (blue-collar worker or not). The second column reports the results of the Kolmogorov-Smirnov test within the public housing sector. For the sake of comparison, we run in Column (3) the same KS-test on the full sample of household heads, including both those who live in the public and private housing sectors. The equality of spatial distribution between the random simulated distribution and the real observed one is accepted in most departments in the public housing sector. In particular, the equality of distribution with respect to Arabic origin (French origin) is not rejected in 80.08 percent (70.23 percent) of the *départements* in the public sector. In contrast, Column (3) shows that in the full sample, the equality of distribution is not rejected in 54.3 percent of the departments for the characteristic Arabic origin and in 24.8 percent of the department for the characteristic French origins. Thus in the whole sample, French households do self-segregate a lot, and even more than Arabic households thanks probably to their social and economic capital. But this strong spatial self-sorting no longer holds in the public housing sector. Similarly, Column (2) shows that the distribution of individuals according to their occupation across the different housing blocks is close to the random simulated distribution. All in all, those tests are supportive of our identifying assumption that the allocation of households across the public housing blocks can be considered as exogenous with respect to their ethnic and occupational backgrounds.

Appendix C.2 provides descriptive statistics of the public housing sector. In Appendix C.3, we perform a variety of additional tests. First, we focus on movers and show that, in the public housing sector, movers do not select new neighborhoods where their ethnic and socio-economic characteristics (education and occupation) is over-represented. Since self-selection could occur prior to the move, we also look at the characteristics of households that have refused a public housing dwelling offer. We show that they display the same characteristics as those who accepted their

nationality across the various public housing residential blocks *only if the size of each block is large enough.* For the sake of illustration, let us assume that 10 percent of individuals with Arabic origins live in the public housing sector in Paris. We should find the same share of 10 percent within each Parisian housing block if the allocation was truly exogenous with respect to ethnic characteristics only if we have a sufficiently large number of individuals within each housing block; otherwise we will observe patterns of spatial concentration in some blocks.

Table 7: Monte-Carlo Test of Random Allocation

% departments without residential sorting relative to households' characteristics

	Public Housing	Total Sample
Household's characteristics		
Nationality from Maghreb/Middle East	80.08	54.36
French Nationality at Birth	70.23	24.89
Occupation: blue collar	97.02	60.95

Note: Comparison between the actual and simulated distributions by ethnic groups shares, education and occupations across public housing blocks (Col. 1) and across the whole sample of housing blocks (Col. 2). Percentage of *départements* where equality is not rejected at the 10 percent level using a Kolmogorov-Smirnov test.

first offer. Thus even if households try to be choosy with respect to the ethnic and socio-economic composition of their neighborhoods, they cannot self-segregate in the public housing sector due to the allocation process and the tight supply constraints of dwellings.

4 Econometric Results

In this empirical section we estimate the utility function (5) while addressing the identification issues discussed in Section 3.2. To proceed, we need to go from the utility function to an estimable discrete choice econometric model. A standard logistic distribution for the error term $\delta_{ik,t}$ is specified, with σ the scaling parameter. One can then express in closed-form the probability of choosing an Arabic name—a formula that enables, in Section 5, to run counterfactuals without probabilities going out of bound:

$$\mathbb{P}(\text{Baby}_{ik,t} = 1) = 1/[1 + \exp(-\Delta\mathcal{V}_{ik,t}/\sigma)], \tag{6}$$

The observable utility differential $\Delta\mathcal{V}_{ik,t}/\sigma$ is retrieved from the coefficients in (5) that can be readily estimated through standard logit. The dependent variable is a binary variable coding for the Arabic origins of a baby's name; explanatory variables relate to our set of parental and neighbor characteristics described above. All specifications add parental occupation fixed effects, parental education fixed effects, and spatial fixed effects at the *département* level (see Section 3.2). Standard errors are clustered at the residential block level. Average marginal effects are reported in all regression tables.

4.1 Babies from all origins

Table 8 reports the benchmark estimates for the sample of babies born over the 2003-2007 period and from all the potential countries of origins. The vertical transmission channel is represented by two binary variables coding for the type of parental first names and for the nationality of origins of the parents/grandparents. The former captures cultural transmission/adoption (as discussed in Section 3.1) and the latter isolates the specific effect attached to 2nd/3rd generation of Arabic migrants. Horizontal transmission is measured by the *share of Arabic names in the block* (defined by equation (3)). Our proxy for the perceived economic cost of a name corresponds to the *local information on penalty* that is based on the unconditional unemployment rate differentials by occupation between Arabic and non-Arabic name holders, weighted by the shares of occupations at the block level (see equation (4) and our discussion at the end of Section 3.2).

Column (1) reports the result for newborn babies in the full sample of private and public housings. All four coefficients of interest are statistically significant and have the expected sign. The t-stats of the two coefficients on vertical transmission are particularly high; this reflects the very stark contrast in the pattern of cultural transmission between parents of Arabic origins and parents without such origins, a feature of the data which has already been discussed in Section 2.2. The horizontal transmission channel is also significant at the one percent threshold. The coefficient associated with the local information on penalty is negative and statistically significant at the five percent level confirming that parents do take into account the information about the employment penalty of Arabic names when they transmit a first name: they are deterred from the fear to inflict an economic cost to their offspring. We devote Section 5 to the quantification of the different effects, focusing in particular on the economic penalty.

In Column (2) we address the endogenous residential sorting issue by restricting the sample to newborn babies of households living in public housing. With one fifth of the original sample left, the number of observations drops dramatically. The coefficients keep the expected sign but they are now much more imprecisely estimated. To overcome this lack of data (see the discussion of Table 2), we expand in Column (3) the sample to all children aged between 0 and 3 of households living in public housing, instead of restricting to newborn. To avoid any overlap, we adjust the age definition to measure the horizontal channel by considering babies aged between 4 and 10 in the housing block. The sample size in the public housing sector is now comparable to the one with the original full sample in Column (1). Column (3) shows that the coefficients keep the same magnitude than the ones in Column (2) with a statistical significance now restored at the 1 percent level.

Columns (4) and (5) document interesting features of the horizontal channel. In column (4) we add the share of arabic names for kids under 10 in a larger area, either at the *sector* level, which consists of 6 adjacent housing blocks, or at the *département* level. Those two variables measure the spatial decay of the horizontal transmission channel by looking at wider geographic unit. None of those two variables exhibit any influence, and the block-based horizontal estimate is unchanged. This points to the importance of studying those channels of transmission at a very fine-grained geographical level (namely at a block housing level), which is impossible with data usually at hand in the social interaction literature. Column (5) includes the share of arabic names for older cohorts to identify the groups to which the parents refer to in their naming decision. Only the coefficient of the cohort of kids under 10 years old has a statistically significant impact in their naming decision. It make us confident when concluding that this variable captures horizontal influence and that, in our sample of public housing, our estimates are unlikely to be pervasively

Table 8: The choice of an Arabic name - All Origins

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var:	Arabic name for baby					
one of grandparents/parents has Arabic-related nationality	0.07^a	0.08^b	0.09^a	0.09^a	0.09^a	0.09^a
	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
at least one parent with Arabic name	0.12^a	0.27^a	0.23^a	0.23^a	0.23^a	0.23^a
	(0.01)	(0.03)	(0.01)	(0.01)	(0.01)	(0.01)
share of Arabic names in block (aged 1-10)	0.05^a	0.06				
	(0.02)	(0.06)				
local information on penalty	-0.43^b	-1.14	-0.86^a	-0.92^a	-0.97^a	
	(0.21)	(0.79)	(0.33)	(0.33)	(0.34)	
share of Arabic names in block (aged 4-10)			0.09^a	0.10^a	0.07^a	0.07^a
			(0.02)	(0.02)	(0.02)	(0.02)
share of Arabic names in sector (aged 4-10)				0.01		
				(0.03)		
share of Arabic names in dept (aged 4-10)				-0.16		
				(0.13)		
share of Arabic names in block (aged 11-25)					0.02	0.03
					(0.04)	(0.04)
share of Arabic names in block (aged 26-49)					0.06	0.06^c
					(0.04)	(0.04)
share of Arabic names in block (aged 50+)					-0.04	-0.04
					(0.03)	(0.03)
local info. on penalty (Mincer-based)						-1.30^a
						(0.36)
Only public housing	No	Yes	Yes	Yes	Yes	Yes
Age of babies	0	0	0-3	0-3	0-3	0-3
Observations	3541	660	3829	3811	3777	3777
Pseudo R^2	0.446	0.443	0.399	0.402	0.402	0.403
Average prob.	0.09	0.21	0.19	0.19	0.19	0.19

Note: logit estimates (average marginal effects). Standard errors, clustered at the residential block level, are in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%. All regressions include dummies for parental education level, parental occupation group, *département* of residence, and years.

contaminated by endogenous residential sorting (otherwise the coefficients for older cohorts should also be non-zero).

In Column (6) the *local information on penalty* is now based on the *conditional* unemployment rate differentials by occupation between Arabic and non-Arabic name holders as computed in the Mincer-like specifications of Table 5. The coefficient associated with the economic penalty in the naming decision remains statistically significant at the one percent level and of the same order of magnitude. The stability of this coefficient in spite of conditioning the penalty by a large set of observables confirms that the attenuation bias due to measurement errors on the true level of the unemployment penalty is unlikely to be a first-order issue in our estimates.

4.2 2nd/3rd generations babies of immigrants from Arabic countries

The previous estimates considered indifferently the decisions of cultural transmission or adoption by parents from various cultural background. We now focus on pure transmission, that is we look at determinants of naming decisions for babies who are second or third generation of migrants from Arabic countries, e.g who are born in France while their parents or grand-parents are born with a nationality from an Arabic country. Table 9 replicates the specification of Table 8 on this new sample. We first report the results for the sample of new born babies, in both the private and public housing sector (Column (1)). In order to eliminate the residential sorting bias the next specifications are based on the public housing sample. The sample of newborn babies from Arabic origins in the public housing sector (Column (2)) is very small (and even lower than in Column (2) of Table 8) and coefficients are imprecisely estimated. Column (3) displays the result for 2nd/3rd generations of children aged between 0 and 3, instead of restricting to newborn, increasing by a factor of five our sample (992 observations). Remarkably, the effect of the local information on penalty is around 3.5 times as large for parents from Arabic origins as for the full sample of parents (Column (3) of Table 8)). In contrast the horizontal channel does not matter for this sample of parents, suggesting a polarization of the trade-off between the vertical and economic channels in the naming decisions. Columns (4) and (5) provide alternative specifications for measuring the effect of the horizontal channel by looking at the peer effect of older children or from larger geographic localities, without any substantial change. Column (6) shows that the coefficient of the economic channel remains stable and statistically highly significant when we use the conditional unemployment penalty instead of the unconditional penalty for building the local information on penalty.

4.3 Additional Robustness checks

All robustness checks are provided for the subsample of children in public housing, aged between 0 and 3 and from Arabic origins (2nd/3rd generations). Due to the drop in sample size, those are the most demanding specifications.

In Table 10 the first column displays the benchmark estimates of column (3) in Table 9. In Column (2) we report the results for a weighted logit, where the individual representativeness

Table 9: The choice of an Arabic name - Migrants from Arabic countries

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var:	Arabic name for baby					
at least one parent with Arabic name	0.33[a]	0.39[a]	0.36[a]	0.36[a]	0.35[a]	0.35[a]
	(0.05)	(0.09)	(0.04)	(0.04)	(0.04)	(0.04)
share of Arabic names in block (aged 1-10)	0.05	0.04				
	(0.08)	(0.14)				
local information on penalty	-1.57	-5.01[c]	-2.95[a]	-3.17[a]	-3.32[a]	
	(1.25)	(2.74)	(1.07)	(1.06)	(1.03)	
share of Arabic names in block (aged 4-10)			0.03	0.05	-0.01	-0.02
			(0.05)	(0.06)	(0.06)	(0.06)
share of Arabic names in sector (aged 4-10)				-0.12		
				(0.11)		
share of Arabic names in dept (aged 4-10)				0.10		
				(0.36)		
share of Arabic names in block (aged 11-25)					0.05	0.07
					(0.10)	(0.10)
share of Arabic names in block (aged 26-49)					0.09	0.11
					(0.10)	(0.10)
share of Arabic names in block (aged 50+)					-0.17[b]	-0.16[b]
					(0.08)	(0.08)
local info. on penalty (Mincer-based)						-3.58[a]
						(1.15)
Only public housing	No	Yes	Yes	Yes	Yes	Yes
Age of babies	0	0	0-3	0-3	0-3	0-3
Observations	482	197	992	987	973	973
Pseudo R^2	0.225	0.348	0.160	0.164	0.169	0.169
Average prob.	0.43	0.50	0.50	0.50	0.51	0.51

Note: logit estimates (average marginal effects). Standard errors, clustered at the residential block level, are in parentheses with [a], [b] and [c] respectively denoting significance at the 1%, 5% and 10%. All regressions include dummies for parental education level, parental occupation group, *département* of residence, and years.

weights reported in the LFS are applied. As discussed in Section 2, the labor force survey is stratified at the *département* level and representativeness is thus not guaranteed at the residential block level, our level of analysis. Our 992 children in column (1) therefore represent 618314 children at the national level. Coefficients are only slightly affected.

In the next two columns we run placebo tests to rule out the possibility that our estimate of the economic cost could be driven by some residual statistical bias attached to endogenous residential sorting. We replicate our benchmark specification on a fake sample of parents/neighbors artificially reallocated to random occupations in Column (3) and to random residential blocks in Column (4). We see that in both cases the local information on penalty, that is based on neighbors occupations (see definition 4), drops and also looses its statistical significance. This makes us confident that our identification strategy, based on the sample of children living in public housing, gets rid of endogenous residential sorting in an efficient way. In columns (5) and (6), we test whether our effects depend on the gender of the children by splitting the sample in girls and boys respectively. We see that the horizontal transmission channel matters more for girls while the economic cost channel is more important for boys.

Table 10: The choice of an Arabic name - robustness 1

	(1)	(2)	(3)	(4)	(5)	(6)
Dep. Var:			Arabic name for baby			
one parent with Arabic name	0.36^a	0.35^a	0.36^a	0.36^a	0.41^a	0.34^a
	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)
share of Arabic names in block (aged 4-10)	0.03	0.06	0.04	0.04	0.15^c	-0.02
	(0.05)	(0.06)	(0.05)	(0.05)	(0.08)	(0.08)
local info. on penalty	-2.95^a	-2.23^b	1.32	0.13	-2.29	-3.98^a
	(1.07)	(1.13)	(1.10)	(0.36)	(1.64)	(1.54)
2nd/3rd generation, public housing	yes	yes	yes	yes	yes	yes
Specifications	Bench.	Weighted	Placebo CSP	Placebo Area	Baby girls	Baby boys
Observations	992	618374	992	986	464	470
Pseudo R^2	0.160	0.178	0.155	0.153	0.175	0.222
Mean pred. prob	0.50	0.51	0.50	0.51	0.44	0.57

Note: logit estimates (average marginal effects). Standard errors, clustered at the residential block level, are in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%. All regressions include dummies for parental education level, parental occupation group, *département* of residence, and years.

Columns (1) and (2) of Table 11 document the effects for couples where both parents have Arabic origins and for mixed couples, respectively. The magnitude of the economic cost channel (though less significant) is larger for the latter, while the vertical transmission motive is stronger when both parents have Arabic origins. Columns (3) and (4) display separately the results for 2nd generation children only (i.e. babies with parents who migrated to France) and for 3rd generation children only (i.e. babies with parents who are born in France). Clearly, the economic cost channel is much larger for the latter. Our interpretation is that parents from Arabic origins who are born in France are more exposed to information on discrimination.

Table 11: The choice of an Arabic name - robustness 2

	(1)	(2)	(3)	(4)
Dep. Var:	Arabic name for baby			
one parent with Arabic name	0.37^a	0.22^c	0.38^a	0.33^a
	(0.04)	(0.13)	(0.05)	(0.06)
share of Arabic names in block (aged 4-10)	0.06	0.17	0.09	-0.03
	(0.06)	(0.17)	(0.07)	(0.09)
local info. on penalty	-2.55^b	-3.72	-1.47	-3.69^b
	(1.24)	(2.66)	(1.47)	(1.62)
2nd/3rd generation, public housing	yes	yes	yes	yes
Specifications	Non Mixed	Mixed	2nd gen.	3rd gen.
Observations	782	143	517	432
Pseudo R^2	0.169	0.227	0.220	0.173
Mean pred. prob	0.52	0.50	0.49	0.52

Note: logit estimates (average marginal effects). Standard errors, clustered at the residential block level, are in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%. All regressions include dummies for parental education level, parental occupation group, *département* of residence, and years.

5 Quantification of the Channels and Welfare Analysis

Up to this point, we have mostly analyzed the signs and statistical significance of coefficients. Some quantification of the effect of the vertical, horizontal and economic channels in the naming decision is now provided. We first analyze the short-run contributions of each channel. Then we perform a welfare analysis. Finally we quantify the long-run effects taking into account the dynamics of intergenerational cultural transmission.

5.1 Short run effects

Column (3) of Tables 8 and 9, based on the public housing sector, is our preferred specification and is used for quantification. All coefficients are reported as average marginal effects over choices in our sample, and are therefore easy to interpret. From column (3) of Table 8, a ten percentage point increase in the estimated local information on penalty of Arabic names translates into a 8.6 percentage points drop in the propensity to give an Arabic name. The last line of the table gives the mean predicted probability of giving an Arabic name in the sample, around 19%, suggesting a quite large effect of labor market penalty. The effect is even larger for the vertical transmission motive, since in couples where at least one parent bears an Arabic name, the propensity to give the same name type jumps by 23 percentage points, when a 10% increase in the share of arabic names in births of the immediate neighborhood increases probability by "only" 0.9 points. In Table 9 we see that the economic cost channel matters even more when we focus on the pure cultural transmission mechanism of migrants. Parents from a specific culture, especially a culture that conflicts with the dominant one, do matter a lot about the economic penalty inflicted to their children channel in their cultural transmission decision. A 10 percentage point increase in the local information on penalty reduces the probability for Arabic parents to give their baby an Arabic name by 29.5

percentage points. Conversely, the fact to have at least one of the parents with an Arabic name increases the probability of transmitting an Arabic name by 36 percentage points.

A way to quantify those effects relative to each other is to look at the model's predicted numbers of babies born with an arabic name when we shut down each of the three channels in turn. In order to calculate such counterfactuals, we adopt the following strategy: We start by running our benchmark regression, to estimate the coefficients of interest which give us the benchmark probability of transmitting an Arabic first name in the sample. Then we run the counterfactual by changing the values for one or more explanatory variables. For instance we shut down the economic cost channel by forcing the local information on penalty variable to be zero for the whole sample. The logit formula (6), provides the counterfactual naming probability for each observation. Summing those over the sample gives the counterfactual number of babies born with an Arabic name in each of the experiment. This procedure ensures that the probability remains within the admissible range, while doing a "what if" experiment.

Results are reported in Table 12, where different lines present different scenarios. The first line reports the true number of arabic names births in the sample of parents originating from Arabic countries and living in public housing (501, representing 320851 babies nationally over the five years considered). The second line is the predicted number of babies born with arabic names in our benchmark regression of this sample (col 3 of Table 9). We then remove the vertical channel associated with parental name. The predicted number of arabic naming decisions falls to 221 in that case, that is 44% of true births. This is a quite drastic cut, specially when compared to the horizontal channel, where a similar thought experiment removes only 2% of arabic naming decisions from the benchmark. The economic channel has a much stronger effect than the horizontal one: removing the economic penalty increases the number of babies receiving an Arabic name by 56%. The line "no ghetto" shows the results from a slightly different experiment. We make as if all blocks in the country had the same neighborhood composition and the same information on unemployment penalty. That amounts to consider the predicted number of babies when averaging the horizontal and penalty variables, which induces effects that almost cancel out in naming choices on average.

5.2 Welfare analysis

Our random-utility discrete-choice model of naming choice allows for a quite simple characterization of welfare changes associated with the different thought experiments of Table 12. A natural metric for welfare in this model is the expected value of maximum utility between choices (the theoretical analysis is relegated to Appendix A). This varies across households, and we average this welfare over our sample (a simple average in the third column of Table 12, and a weighted one in the fourth). The absolute level of welfare has no meaningful unit in the logit model—as noted by Anderson et al. (1992) and Train (2003)—and a natural way to quantify welfare changes is to first take the difference between welfare in each scenario and the benchmark case and then compare across scenarios. Looking at the last two columns of Table 12, we see for instance that the negative impact of removing the vertical transmission motive would be more than twenty times larger than

Table 12: Quantification of the 3 channels

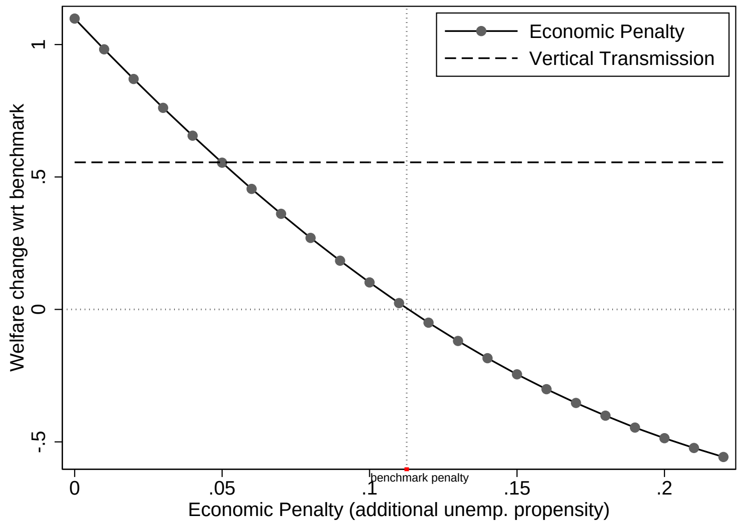
	# babies with arabic name		Mean Δ Welfare wrt benchmark	
Scenario:	count	weighted count	change	weighted change
true figure	501	320851	.	.
benchmark	501	316292	.	.
no vertical (parental name)	221	139636	-.555	-.567
no horizontal	491	310516	-.026	-.024
no penalty	783	490945	1.098	1.109
no ghetto	500	316275	.002	.005
no foreign names	0	0	-.822	-.838

Note: This table uses logit estimates (col 3 of Table 9). Each line presents a scenario, removing in turn one of the channels of influence in the regression. Second column shows predicted number of babies born with Arabic name, in the sample of 3829 children 0-3 years old in public housing (representing 2,425,238 nationally).

the effect of removing the horizontal one. Considering economic penalty, the effect varies according to the cut in the additional unemployment rate associated with Arabic names naturally. If this penalty was brought to zero, the gain in welfare would be about twice as large as the one arising from the utility boost linked to vertical transmission. Figure 1 spans over a wider set of changes in economic penalty, and compares it to the welfare changes associated with vertical transmission. The horizontal axis reports the counterfactual economic penalty (percentage point differences in unemployment rates). The y-axis measures welfare changes difference with the benchmark level (with zero change occurring at the sample average of economic penalty, around 11 percentage point difference). We also represent the utility gain (with respect to benchmark) associated with vertical transmission. An interesting conclusion from this figure is that we are now able to gauge the strength of cultural attachment in monetary units. Indeed the vertical transmission motive of one's cultural trait is equivalent in terms of welfare gains to a cut by around half of the perceived economic penalty associated with that trait. Since, in this sample, this amounts to cutting the penalty by around 6 percentage points, using the same estimates of unemployment related income loss as in section 2.3, the vertical transmission channel is on the same indifference curve as a 3% upward shift in lifetime income of one's child.

It is possible to look for differences in the monetary valuation of cultural attachment across households. Probably the most meaningful source of heterogeneity is related to whether the parents choosing a name for their baby are themselves first generation migrants from Arabic countries versus being born in France from migrating Arabic parents. This distinction is done in the last two columns of Table 11, and we reproduce in Appendix B the equivalent of figure 1 for the samples of 2nd and 3rd generation babies (from, respectively, 1rst and 2nd generation parents). Interestingly, we find a much stronger monetary equivalent of the vertical transmission motive for first generation migrants. For those, the vertical transmission channel is on the same indifference curve as a 6.2% upward shift in lifetime income of one's child while for second generation parents this figure amounts to a 2.3% upward shift in lifetime income.

Figure 1: Welfare in the short run, economic penalty and vertical transmission



Finally, we consider an experiment where France would return to naming regulations from the early nineteen century. Between 1803 and 1993, the choice of first names was essentially restricted to Saint names, names from ancient Greece and Rome, and names from the Bible. The legal procedure was that a civil officer had to state whether the name proposed by the parents respected the 1803 Napoleonic law. If the answer was negative, the parents had to challenge the decision in court. Foreign names were hardly tolerated at all before a 1987 revision explicitly asked civil officers to be more liberal with names coming from a "foreign or French tradition, whether national or local". Note that the computation of this scenario involves shutting down all three channels emphasized in our paper, but also the occupation, regional and educational controls we have in the regression, in order to generate a predicted number of babies with Arabic names of 0. The mean welfare loss from this return to a strict ban on foreign—and therefore Arabic—names would be substantial, around 50% larger than the cut of the vertical channel alone.

5.3 Long run effects

5.3.1 The dynamic version of the model

In this section we assess the long run implications of our structural model on the naming patterns. We run various counterfactual experiments based on our estimates of the naming decision of descendants of migrants from Arabic countries (Table 9).[14] To this purpose we consider a simple dynamic extension of the static model described in Section 3. We restrict our focus to a partial equilibrium analysis where any potential feedback effect of naming patterns on the economic penalty is ignored –obviously an important simplification in a long run perspective. Let us consider that blocks are

[14]Simulations for the population from all origins are available upon request from the authors

populated by a large number of agents $\mathcal{N}$. To keep the model tractable, we consider that agents differ only in their name type. We denote m_t the share of Arabic name holders at date t. Time is discrete. Abstracting from demographic and fertility issues, we impose a constant $\mathcal{N}$ by assuming that just before death each agent gives birth to a unique child whose name is chosen by his parent. Mortality is ruled by a Poisson process with parameter θ. The naming decision follows the model described in Section 3. We denote $(\mathbb{P}_{0,t}, \mathbb{P}_{1,t})$ the probability of giving an Arabic name for, respectively, a non-Arabic parent and an Arabic parent. Those probabilities potentially differ because of the vertical transmission channel. The law of motion of the share of Arabic name holders is given by

$$m_{t+1} = (1-\theta) \times m_t + \theta[(1-m_t) \times \mathbb{P}_{0,t} + m_t \times \mathbb{P}_{1,t}]. \tag{7}$$

Labeling μ the steady state value of m_t, we have

$$\mu = (1-\mu) \times \mathbb{P}_0 + \mu \times \mathbb{P}_1, \tag{8}$$

where the steady state probabilities of transmission, $(\mathbb{P}_0, \mathbb{P}_1)$, are characterized by equation (6). Those can be conveniently rewritten as

$$\mathbb{P}_A = \left[1 + \tanh(\Delta\mathcal{V}_A/2\sigma)\right]/2, \tag{9}$$

where $\tanh(x) \equiv (e^x - e^{-x})/(e^x + e^{-x})$ and $\Delta\mathcal{V}_A$ is the observable utility differential (with $A \in \{0,1\}$). In $\Delta\mathcal{V}_A/\sigma$, the steady state value of the horizontal component is equal to $\mathbb{E}(m_t) = \mu$ and the parameters $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$ correspond to the point estimates retrieved from our empirical analysis. We thus have

$$\Delta\mathcal{V}_A/\sigma = \hat{\alpha}_0 + \hat{\alpha}_1 A + \hat{\alpha}_2\mu + \hat{\alpha}_3\mathcal{C} \text{ with } A \in \{0,1\}, \tag{10}$$

where $\mathcal{C}$ corresponds to an exogenous economic penalty attached to Arabic name holders.

Combining (8), (9) and (10), we obtain μ as a solution to the following fixed point equation

$$\mu = \frac{1}{2} + \frac{1-\mu}{2} \times \tanh\left(\frac{\hat{\alpha}_0 + \hat{\alpha}_2\mu + \hat{\alpha}_3\mathcal{C}}{2}\right) + \frac{\mu}{2} \times \tanh\left(\frac{\hat{\alpha}_0 + \hat{\alpha}_1 + \hat{\alpha}_2\mu + \hat{\alpha}_3\mathcal{C}}{2}\right), \tag{11}$$

A first noticeable point is that this equation does not depend on the value of the Poisson parameter θ. This makes us confident on the innocuity of our dynamic, albeit simple, demographic structure as long as we focus strictly our analysis on the steady-state, abstracting from any consideration on the transition dynamics. Second, while existence of μ follows directly from the Brouwer fixed-point theorem, uniqueness is not guaranteed and the previous equation may have multiple solutions. Contrary to Brock and Durlauf (2001, proposition 2) our dynamic setting with a non-homogenous population of agents forbids us to simply characterize the presence of multiplicity as a function of the parameter values.[15] We consequently rely on numerical computations of (11) to

[15] Without vertical transmission, i.e. with $\hat{\alpha}_1 = 0$, our model would be included in the class of model analyzed by

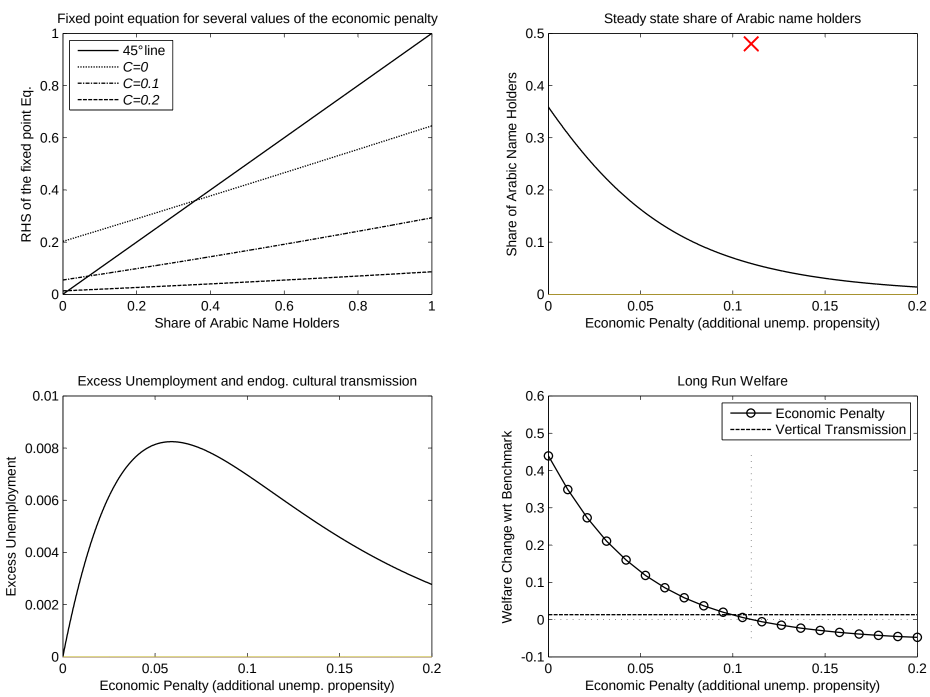
characterize the set of solutions.

5.3.2 Long run steady states and counterfactual experiments

The parameters $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$ correspond to the point estimates of a specification similar to our benchmark where the sample is restricted to descendants from Arabic migrants in public housing (Column 3, Table 9). A difference is that we remove all added covariates and fixed-effects which are not directly related to the three channels of our model in equation (5). The resulting values for our parameters are $\hat{\alpha}_0 = -1.37, \hat{\alpha}_1 = 1.8, \hat{\alpha}_2 = 0.17$, and $\hat{\alpha}_3 = 14.8$. We solve numerically the fixed-point equation (11) for values of $\mathcal{C}$ spanning the range $[0, 0.2]$; for each value of $\mathcal{C}$ this gives us the steady-state share of Arabic name holders $\mu(\mathcal{C})$. We then compute our second variable of interest, $\mathcal{U}$, which corresponds to the steady-state value of excess-unemployment due to discrimination towards Arabic name holders

$$\mathcal{U}(\mathcal{C}) \equiv \mu(\mathcal{C}) \times \mathcal{C}, \tag{12}$$

Figure 2: Long-Run Effects - Descendants from Arabic Migrants in Public Housing



The results are displayed on figure 2. The upper left panel depicts the fixed-point equation (11)

Brock and Durlauf (2001). Indeed, in that case, the population of babies has homogenous characteristics with respect to the naming process and our equation (11) is equivalent to their main equation (12).

for various values of economic penalty $\mathcal{C}$ within the range of 0 to 20 percent. We can check visually in this panel that the equilibrium is unique confirming that the horizontal channel is not large enough to generate multiple social equilibria. The upper right panel reports the steady-state value of $\mu(\mathcal{C})$. The red cross represents the actual values of the unemployment penalty (equal to 0.11) and the actual share of Arabic name holders (equal to 0.48) observed in our sample of first and second generations of migrants from Arabic countries living in public housing over the 2003-2007 period. We observe that for $\mathcal{C} = 0.11$, the steady-state share of Arabic name-holders predicted by our structural model is $\mu = 0.086$ which is much smaller than the actual one. This feature might be explained by the fact that migration from Arabic countries is still a quite recent phenomenon in France and that most babies born in the 2003-2007 period belong to the third generation of migrants only. Hence the actual share is still far from its steady-state value and transitory dynamics are expected to bring it down in the future. We also see that in absence of discrimination ($\mathcal{C} = 0$), this steady-state share is predicted to be much larger, at 0.36. This confirms that the economic cost channel is also a key driver of cultural transmission in the long run.

On the bottom left panel we report the predicted excess-unemployment $\mathcal{U}(\mathcal{C})$. We observe a non-monotonous relationship. This shows how a change in the degree of penalty (potentially resulting from public policy) may be partially counteracted by endogenous naming choices. Indeed, when the perceived penalty intensity ($\mathcal{C}$) falls, parents tend to raise their propensity to give Arabic names, everything else equal. This counteracting effect results in an ambiguous effect on the overall level of discrimination in the economy (total number of unemployed Arabic name holders because of the estimated penalty). Our simulation shows the interesting result that the overall discrimination starts by rising when the underlying penalty (the intensity of discrimination) decreases from a high initial level.

The bottom right panel replicates, for the long-run, the welfare analysis of figure 1. Considering the same benchmark and counterfactual conditions, we compute, for each, the steady-state value of aggregate welfare, taking into account the full dynamic impact of the counterfactual change in parameters. With respect to the short run, we see that the welfare impact of the vertical channel experiences a fivefold reduction: This stems from the long-run predicted fall of the share of Arabic name holders in the population (of comparable magnitude), which are the only individuals for which the vertical channel is relevant. The vertical transmission motive is now equivalent in terms of welfare gains to a cut by 1.2 percentage points of the penalty, namely a 0.6% upward shift in lifetime income of one's child. Naturally, the same figure calculated on the subsample of the population that keeps an Arabic name in the long run (8.6% of the considered population) would be much higher, at 6.9% of lifetime income.

6 Conclusion

While it might seem natural to consider culture as a deep characteristic of individuals, our paper shows that the cultural choices made by a person cannot be completely insulated from the economic context in which he/she operates. We focus on one cultural trait that has the advantage of being easily measurable, identified by social sciences as a key marker of cultural identity, and has economic consequences: first names. We use the propensity to transmit Arabic versus Non-Arabic names to newborn babies in France over the 2003-2007 period to capture the effect of economic incentives on cultural transmission. Our random-utility discrete choice model disentangles between three channels for the transmission of first names: a vertical channel from parental culture, a horizontal channel from the neighborhood's culture, and an economic penalty associated with names that sound culturally distinctive. We estimate those channels using the French Labor Force Survey, taking advantage of the exogenous allocation of parents in public housing to analyze the spatial interactions in naming decisions within blocks.

Our results show that economic factors deeply shape the individual decision of cultural transmission. While the vertical channel plays a key role in the process of cultural transmission, we also find that parents do take into account the perceived economic cost of their cultural trait in the naming decision of their babies. The horizontal channel, which has been the focus of much attention in the social interaction literature, is found to be much less important in our specific case. Our theory-based estimates allow us to perform a welfare analysis where we gauge the strength of cultural attachment in monetary units. We find that allowing for the vertical transmission channel provides the same shift to parents' utility as a 3% raise in lifetime income of the child. Our paper is also a first attempt to quantify the welfare effects of public policy affecting cultural choice. We show in particular that a return to an old regulation banning choice of names of foreign origins would cause very important losses to the well-being of parents.

While we have focused on naming decision, our paper therefore opens new questions on the welfare effects of public policies aiming at promoting or restraining expressions of cultural identity, such as wearing religious signs in public area. It also raises questions about the use in academic papers of cultural traits as determinants of economic outcomes. It seems clear from our results that at least some aspects of culture cannot be considered as exogenous to what happens in the economic sphere.

References

Adida, C., Laitin, D. and Valfort, M.A, (2010), "Identifying Barriers to Muslim Integration in France," *Proceedings of the National Academy of Sciences of the United States of America*, 107(52), pp. 384-390.

Aghion, P., Algan, Y., Cahuc, P., Shleifer, A., (2010), "Regulation and Distrust," *Quarterly Journal of Economics*, 125(3), pp.1015–1049.

Akerlof, G and Kranton, R., (2000), "Economics and Identity," *Quarterly Journal of Economics*, 115, pp. 715-753.

Alesina A., R. Baqir and Easterly, W., (1999), "Public Goods and Ethnic Divisions", *Quarterly*

Journal of Economics, Vol. 114(4), pp. 1243-1284

Alesina, A. and Fuchs-Schundeln, N., (2007), "Goodbye Lenin (or Not?): The Effect of Communism on People's Preferences," *American Economic Review*, 97, pp. 1507–1528.

Algan, Y. and Cahuc P., (2010), "Inherited Trust and Growth", *American Economic Review*, 100(5), pp. 2060-2092.

Algan, Y. and Cahuc P., (2013), "Trust and Human Development: Overview and Policy," in P. Aghion and S. Durlauf, editors, *Handbook of Economic Growth*, Vol. 1A, North-Holland, forthcoming.

Algan, Y., Dustman, C., Glitz, A. and Manning, A., (2010), "The Economic Situation of First and Second Generation Immigrants in France, Germany and UK," *Economic Journal*, 120, pp. 4-30.

Algan, Y., Hémet, C and Laitin, D., (2012), "The Social Effects of Diversity at the Local Level: A Natural Experiment with Exogeneous Residential allocation," Sciences-Po Working Paper.

Anderson S., de Palma, A., and Thisse, J.-F., (1992), *Discrete Choice Theory of Product Differentiation*, MIT Press.

Bernardot, M., (2008), "Loger les immigrés", *Bellecombe-en-Bauges, Rhone-Alpes: Editions du Croquant.*

Bertrand, M. and Mullainathan, S., (2004), "Are Emily and Greg More Employable Than Lakisha and Jamal? A Field Experiment Evidence on Labor Market Discrimination," *American Economic Review*, 94, 991-1013.

Bisin, A. and Verdier, T., (2001), "The Economics of Cultural Transmission and the Dynamics of Preferences," *Journal of Economic Theory*, 97, 298–319.

Bisin A., Topa, G. and Verdier, T., (2004), "Religious Intermarriage and Socialization in the United States," *Journal of Political Economy*, 112(3), pp. 615–664.

Bisin A., Patacchini, E., Verdier, T. and Zenou, Y., (2010), "Bend it like Beckham: Ethnic identity and Integration", NBER Working paper 16465.

Bisin A. and Verdier, T., (2011), "The economics of cultural transmission and socialization", in, J. Benhabib, A. Bisin and M., O. Jackson, editors, *Handbook of Social Economics*, Vol. 1A, North-Holland.

Bloom N., Sadun, R., Van Reenen, J., (2012), "The organization of firms across countries", *Quarterly Journal of Economics*, 127(4), pp. 1163-1705.

Blume, L. Brock, W., Durlauf, S. and Ioannides, Y., (2011), "Identification of Social Interactions," in, J. Benhabib, M. O. Jackson and A. Bisin, editors, *Handbook of Social Economics*, Vol. 1A, North-Holland.

Breuil-Genier, P., (2001), "Les variations des revenus individuels entre périodes de chômage et d'emploi", *Economie et Statistiques*, pp. 61-79.

Carroll, C.D., Rhee, B.K. and Rhee, C., (1994), "Are there cultural effects on saving? Some cross-sectional evidence," *Quarterly Journal of Economics*, 109, pp. 685-699.

Duguet, E., Landri, N., L'Horty, Y. and Petit, P., (2010), "Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area", *Annales dÉconomie et de Statistique*, pp. 99-110.

Ellison, G. and E. Glaeser, (1997), "Geographic Concentration in U.S. Manufacturing Industries: A Dartboard Approach", *Journal of Political Economy* 105(5), pp. 889–927.

Fernandez, R. and Fogli, S., (2009), "Culture: An Empirical Investigation of Beliefs, Work and Fertility," *American Economic Journal: Macroeconomics*, 1(1): pp. 146-177.

Fernandez, R., (2011), "Does culture matter ?," in, J. Benhabib, M. O. Jackson and A. Bisin, editors, *Handbook of Social Economics*, Vol. 1A, North-Holland.

Fryer, R. G. and Levitt, S. D., (2004), "The Causes and Consequences of Distinctively Black Names," *The Quarterly Journal of Economics*, 119(3), pp. 767-805.

Glaeser, Edward, L. and José A. Scheinkman, (2001), "Measuring Social Interactions" in S.N. Durlauf and H.P. Young (eds.), Social Dynamics . Washington, DC: Brookings Institution Press: 83132.

Goldin, C. and Shim, M., (2004), "Making a Name: Women's Surnames at Marriage and Beyond," *Journal of Economic Perspectives*, 18(2), pp. 143-160.

Goux, D. and Maurin, E., (2007), "Close neighbours matter: neighbourhood effects on early performance at school", *Economic Journal*, Vol. 117, pp. 1-24.

Giuliano, P., (2007), "Living Arrangements in Western Europe: Does Cultural Origin Matter?," *Journal of the European Economic Association*, 5(5), pp. 927-952.

Guiso L., Sapienza, P. and Zingales, L., (2004), "The Role of Social Capital in Financial Development," *American Economic Review*, 94(3), pp. 526-56.

Guiso L., Sapienza, P., Zingales, L., (2006), "Does Culture Affect Economic Outcomes?," *Journal of Economic Perspectives*, 20(2), pp. 23-48.

Guiso, L, P. Sapienza, and L. Zingales, (2008), "Social Capital as Good Culture," *Journal of the European Economic Association Papers and Proceedings*, 6(2), pp. 295-320.

Guiso, L, Sapienza, P. and Zingales, L., (2009), "Cultural Biases in Economic Exchanges," *Quarterly Journal of Economics*, 124(3), pp. 1095-1131.

Head, K. and Mayer, T. (2008), "Detection of local interactions from the spatial pattern of names in France," *Journal of Regional Science*, 48(1), pp. 67-95.

INSEE, (2012), Fiches thématiques - Population immigrée - Immigrés - Insee Références - édition 2012 [archive], Insee 2012

Jouniaux, L. (2001), *Les 20,000 plus beaux prénoms du monde.* Hachette eds.

La Porta, R., Lopez-de-Silanes, F., Shleifer, A., and Vishny, R., (1997), "Trust in Large Organizations," *American Economic Review Papers and Proceedings*, 87(2), pp. 333-338.

Lieberson, S., (2000), *A Matter of Taste: How Names, Fashions and Culture Change*, Yale University Press.

Lieberson, S. and Mikelson, K.S, (1995), "Distinctive African American Names: an Experimental, Historical, and Linguistic Analysis of Innovation," *American Sociological Review*, 60, pp. 928-946

Luttmer, E. and Singhal, M., (2011), "Culture, Context, and the Taste for Redistribution," *American Economic Journal: Economic Policy*, 3(1), pp. 157-179.

Manski, C., (1993), "Identification of Endogeneous Social Effects: The Reflection Problem," *Review of Economic Studies*, 60, pp. 531-542.

Oreopoulos, P., (2003), "The long run consequences in working in a poor neighborhood", *Quarterly Journal of Economics*, Vol. 118(4), pp. 1533-1575

Simon, P., (2003), "Le logement social en France et la gestion des populations a risques", *Hommes et Migrations*, No. 1246, pp. 76-91.

Small, K., and Rosen, H., (1981), "Applied welfare economics of discrete choice models," *Econometrica*, 49, 105-130.

Tabellini, G., (2008), "The Scope of Cooperation: Norms and Incentives," *Quarterly Journal of Economics*, pp. 905-950.

Tabellini, G., (2010), "Culture and Institutions: Economic Development in the Regions of Europe," *Journal of the European Economic Association*, 8(4), pp. 677-716.

Train, K., (2003), *Discrete Choice Methods with Simulation*, Cambridge University Press.

Appendix

A Welfare Analysis

The absolute level of utility of a discrete choice model such as (1) cannot be identified. Hence, without loss of generality, we can express individual i's welfare in our binary model specified in terms of cross-alternative utility differentials (see equation (5))

$$W_i \equiv \mathbb{E}\{\max[0; \Delta\mathcal{V}_i + \delta_i]\} \tag{13}$$

With δ_i logistically distributed and σ its scaling parameter, this expectation becomes (Small and Rosen 1981, Anderson et al. 1992)

$$W_i = \sigma \ln[1 + \exp(\Delta\mathcal{V}_i/\sigma)] + \lambda \tag{14}$$

where λ is a term that is constant across individuals and alternatives. Note that both λ and σ cannot be identified from the data and are unknown. This reflects the fact that the absolute value of utility cannot be interpreted in the logit, and that we can only put meaningful numbers on *relative* welfare.

The change in individual welfare between the benchmark condition `B` and a counterfactual condition `A1` is given by

$$\Delta W_i^{\mathtt{A1}} = \sigma \times \left(\ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{A1}}/\sigma)\right] - \ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{B}}/\sigma)\right]\right) \tag{15}$$

where $\Delta\mathcal{V}_i^{\mathtt{A1}}$ and $\Delta\mathcal{V}_i^{\mathtt{B}}$ correspond to utility differentials (of choosing Arabic vs non-Arabic names) in the benchmark and counterfactual conditions respectively.

The change in aggregate welfare is obtained by averaging (15) across the sampled population

$$\Delta\mathcal{W}^{\mathtt{A1}} = \frac{\sigma}{\mathcal{N}} \left(\sum_i \ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{A1}}/\sigma)\right] - \ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{B}}/\sigma)\right] \right), \tag{16}$$

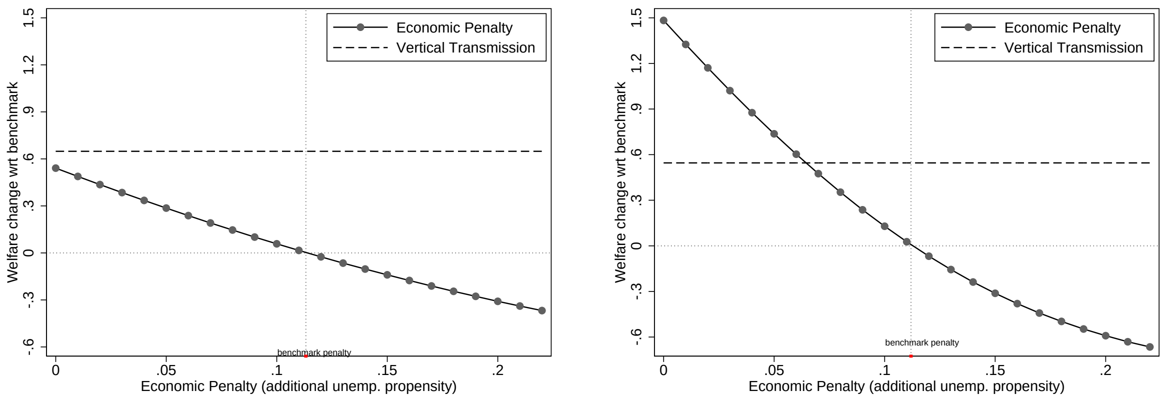
which still has no meaningful unit because of the unknown scaling parameter σ. The solution consists in computing the relative change in aggregate welfare between two different counterfactuals `A1` and `A2`

$$\frac{\Delta\mathcal{W}^{\mathtt{A1}}}{\Delta\mathcal{W}^{\mathtt{A2}}} = \frac{\sum_i \left(\ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{A1}}/\sigma)\right] - \ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{B}}/\sigma)\right]\right)}{\sum_i \left(\ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{A2}}/\sigma)\right] - \ln\left[1 + \exp(\Delta\mathcal{V}_i^{\mathtt{B}}/\sigma)\right]\right)}. \tag{17}$$

Equation (17) involves a double comparison (i.e. a ratio of welfare change) and its interpretation is subtle. For the sake of exposition, in the main text, Table 12 and figure 1 report, for each counterfactual `A`, the welfare change $\Delta\mathcal{W}^{\mathtt{A}}$ under the standard scaling assumption $\sigma = 1$. However, when discussing the effects, we pay attention to analyzing only the ratio of welfare change between counterfactuals.

Figure 3: Welfare in the short run, economic penalty and vertical transmission

(a) 2nd generation babies, 2006 (b) 3rd generation babies



B Welfare effects for different generations of migrants

Figure 3 reproduces Figure 1 for two samples of baby naming decisions. The (a) panel takes all babies born from parents living in French public housing themselves being born in an Arabic country. The (b) panel constrains the set of parents to have Arabic origins, but being born on French territory (the coefficients used come from the last two columns of Table 11). The x-axis axis reports the counterfactual economic penalty (percentage point differences in unemployment rates). The y-axis measures welfare changes (with respect to the benchmark level). The horizontal bar represents the utility gain (wrt benchmark) associated with vertical transmission.

C Public Housing

C.1 Formal Allocation Process in Public Housings

Due to a strong "Republican ideal", the French public housing system allocates state-planned moderate cost rental apartments (HLMs - *Habitations à Loyer Modéré*) to natives and immigrants without concern for their cultural and ethnic background, mixing people indiscriminately. The process of allocation across public housing blocks is mainly inspired by theories from the famous architect Le Corbusier (1887-1965). Le Corbusier insisted that France must avoid the homogeneous ghettoes of the urban landscapes elsewhere, and should therefore allocate housing blind to ethnicity, not permitting family networks to grow within housing establishments. These ideas were later translated into state regulation (Bernardot, 2008). This Appendix documents the legal framework for the residential allocation in the public housing sector. We show that the exogeneity of the allocation with respect to salient characteristics, such as ethnicity, is built into the law.

We first describe the eligibility criteria and the formal selection process. The only eligibility requirements for admittance into the public housing sector are to be legally living in France (as a

French citizen or migrant with a valid residence permit) and to be living under a certain threshold of income per unit of consumption. This income ceiling is usually rather high: in 2009, this threshold was between 36748 and 50999 Euros per year for a four-person family, depending on the region of residence. As a consequence, the population eligible for public housing is on average three to four times as large as the available space in vacant dwellings. But the situation is even much tighter in the most crowded areas. Take the example of Paris. According to the *Observatoire du Logement et de l'Habitat de Paris* (2011), as of January 2010, there were 186017 public housing dwellings in Paris. Public housing buildings are scattered across all Parisian areas, with a high concentration (69 percent) in six districts (the 13th, 14th, 15th, 18th, 19th and 20th *arrondissements*). Within Paris, 48.7 percent of households are under the income ceiling and could be theoretically eligible. In practice, only households with very modest incomes apply (71 percent have an income lower than the minimum ceiling for all France, equivalent to 2345 euros per month for a household with two children). On the 31st of December 2010, there were 121937 ongoing applications, to be compared to 12500 public housing units allocated over the year 2010. Due to those stringent housing supply constraints, other eligibility criteria are taken into account.[16] In addition to household income, the family situation and household size are taken into account to ensure a suitable match with the characteristics of vacant dwellings, as well as the emergency of the application. Those latter criteria have recently become the main criteria used by the commission due to the boom in housing prices in the private sector during the mid-90s and the 2000s. In particular, five priority criteria are defined by law (Article L441-1 of law relative to construction and housing - *Code pour la Construction et l'Habitat*) at the national level to ensure that vacant housing will first be distributed to households with obvious social difficulties. Households satisfying these priority criteria are those in which there is a (mentally or physically) disabled person, those living in precarious or hazardous shelter due to financial constraints, those living in a temporary accommodation, individuals living in a precarious shelter who recently found a job after a long unemployment spell, and spouse-abused individuals.

Regarding the selection process, the commissions of selection in charge of allocating households to vacant public housing dwellings are held at the *département* level (or at the city level in the case of Paris which is both a city and a *département* due to its size). The composition of the commissions is regulated by law: it includes six members of the public housing offices board, a representative of associations for social and economic insertion (appointed by the head of the *département*), mayors of the cities (or districts) in which vacant housings are to be attributed, as well as a representative of any association defending tenant rights. In addition, another *département* representative may attend the commission. For each vacant housing unit, at least three households must be considered by the commissioners, who finally decide which household will be allocated to which housing unit,

[16] To apply for an apartment in a public housing sector, one has to submit a form showing the identity, the family situation, the employment status and the resources of the household, the reasons for applying to the public housing sector (currently or soon to be homeless, or reasons related to health situation, family situation, job situation, inappropriate current housing, unpleasant environment), the type of housing looked for, whether the applicant is disabled and whether this is the first application. It is important to stress the fact that the application form contains very limited information about the ethnicity of the applicant: he or she only needs to inform about his or her nationality, which is limited to three possible categories (French, European Union, or non European Union).

according to the eligibility and priority criteria detailed above. Other criteria such as the number of children in the household are also taken into account in order to allocate suitable dwellings.

Despite this legal process of allocation, one might still be worried about the possibility of self-sorting of households that refuse the residential allocation proposed by the commission. In theory, households can refuse up to three offers. However, self-sorting, especially on ethnic characteristics, seems unlikely to be a common practice. Residential mobility within the public housing sector is very low, due to the current strong shortage of supply of public housing dwellings. This makes it unlikely that the selected households could be really picky about the diversity of their neighborhood (see the study by Simon, 2003). In addition, rents are considerably lower in public housing than in private housing, increasing the opportunity cost of moving, so that the turnover is very low. More specifically, the mobility rate in the public housing sector is even lower than for recent owners. Public housing allocation in Paris serves as a useful concrete example. The mobility rate (defined as the ratio of new entrants over the total number of public housing dwellings) is particularly low: it reaches 5.5 percent in 2010. It is formally possible to indicate a precise neighborhood in the application form, but in practice, very few applicants (6.6 percent) do provide this information. More than half of the 121937 applicants (52.9 percent) did not mention any particular area at all, probably due to the fear of being rejected on this ground. Among those who indicated an area of preference, 91.2 percent mentioned the area where they were already living. People who move within the public housing sector are people who moved for larger space following an increase in their household size (only 12 percent of the public housing dwellings have more than three rooms).

C.2 Descriptive statistics on public housing

Table 1 shows the summary statistics of the public housing units in the Labor Force Survey. The survey comprises 1535 public housings, with an average of 45.12 public housing blocks by *département*. The average number of households within a block is 17.99, which is similar to the whole sample.

Table 13 reports the descriptive statistics of the sample of individuals aged over 15 years and who are consequently interviewed in the LFS. There is an over-representation of Arabic Name holders in public housings: the share reaches 14 percent in this type of housing against 4 percent in the total sample. Similarly, the proportion of individuals who have Arabic origins is 16.8 percent in public housings. This proportion is three times as high as in the total sample (5.61 percent). Individuals who enter the public housing sector have also lower socio-economic backgrounds than the rest of the population. The share of unemployed is almost twice as high in public housings (9.6 percent) than in the total sample (5 percent). There is an over-representation of blue-collars and clerks (78 percent in the public housing, and 55 percent in the whole sample) and an under-representation of executives (3 percent in public housing, 12 percent in the whole sample). This table shows that there is a clear selection of individuals into the public housings, since the eligibility is based on the socio-economic characteristics. However, our key identification strategy relies on the exogeneous spatial allocation of individuals within the public housing sector.

Table 13: Descriptive statistics of individuals

	Total sample Mean (std)	Public housing Mean (std)
Age	45.22 (19.22)	40.93 (18.19)
Gender (Male)	0.48 (0.49)	0.46 (0.49)
Married	0.61 (0.48)	0.49 (0.50)
Arabic names	0.04 (0.21)	0.14 (0.35)
Arabic origins	0.05 (0.23)	0.16 (0.37)
Employed	0.52 (0.49)	0.50 (0.49)
Unemployed	0.05 (0.20)	0.09 (0.27)
Inactive	0.43 (0.49)	0.41 (0.49)
Hourly wage (euros)	9.70 (4.29)	8.24 (3.00)
Occupation: executive	0.12 (0.33)	0.03 (0.18)
Occupation: intermediate	0.20 (0.40)	0.14 (0.34)
Occupation: clerk	0.30 (0.46)	0.40 (0.49)
Occupation: blue collar (skilled)	0.15 (0.36)	0.20 (0.40)
Occupation: blue collar (unskilled)	0.10 (0.30)	0.18 (0.39)
Occupation: craftman	0.06 (0.24)	0.02 (0.14)
No education	0.21 (0.41)	0.36 (0.48)
Elementary school	0.44 (0.49)	0.42 (0.49)
High school	0.14 (0.34)	0.11 (0.31)
College	0.09 (0.28)	0.05 (0.22)
Graduate	0.10 (0.30)	0.04 (0.20)

C.3 Additional tests on the Exogeneous Spatial Allocation Process in Public Housing

C.3.1 Absence of self-sorting on ethnic backgrounds

Our first set of alternative tests consists in showing that while households tend to self-segregate in the unconstrained private housing market, there is no such evidence in the public housing market. We test this using the LFS and focusing on individuals who recently moved into an area (within the previous year).

We first estimate the correlation between the origin (nationality) of individuals moving into a new area and the share of the area's "long term" population of the same origin.[17] We expect a significant relationship in the private housing market where location choice is relatively unconstrained but not in the public housing sector. Table 14 reports the results from an OLS regression of the share of neighbors from the same origin as new movers on new movers' characteristics: nationality group, public housing dummy, quadratic function of age, hourly wage (log) education, socio-economic category, *département* fixed effects, and interaction of individual characteristics with the public housing dummy. We consider seven different nationality groups: native French, naturalized French, Europeans, Arabic, other Africans, Asians, and other nationalities, which is taken as the reference group.

Three facts are worth noting here. First, there is indeed evidence that on average native French are significantly more likely to move in neighborhoods where the share of natives is higher, compared to households from other nationalities. This is not surprising given the fact that natives make up a large majority in the French population. The second interesting point is that the coefficient for living in the public housing sector is negative and significant at the 5% level. More precisely, it reveals that HLM households move in areas where the share of individuals from the same origin is on average 18.4% lower than for households in the private housing sector. This result strengthens the idea that the extent to which households in the public housing sector live close to their co-ethnics is lower than in the private sector. Finally, when we interact nationalities with the public housing dummy, none of the coefficients are statistically significant. This comforts us with the idea that there is no particular self-segregation along ethnic lines in the public housing sector. We have run the same kind of test on other individual characteristics, and reach similar conclusions. We find that in the private sector, highly educated or low skilled workers are very likely to move into neighborhoods with higher levels of highly educated (respectively low skilled) people. This is not surprising and illustrates self segregation along education level in the private sector. On the contrary, such segregation does not appear in the public housing sector.

[17] A similar test was proposed by Goux and Maurin (2007) to show that the educational achievement of the children of newcomers in public housing is uncorrelated with that of the current residents. Individuals do not self-select in public housing neighborhoods according to the educational achievement of the neighbors' children. By contrast, the authors find a strong self-selection on the educational characteristics in the private housing sector.

Table 14: Correlation between new movers' nationality and the residential share of neighbors from the same nationality

Dep Var: % of neighbors from the same nationality	
Nationality (ref.: Other nat.)	
Native	0.067^b
	(0.030)
Naturalized French	-0.037^b
	(0.012)
European	-0.007
	(0.011)
Arabic	0.007
	(0.013)
Asian	-0.019
	(0.047)
Public Housing (HLM)	-0.184^b
	(0.066)
Nationality × HLM	
HLM × Native	0.040
	(0.033)
HLM × Naturalized	0.051
	(0.036)
HLM × European	0.010
	(0.037)
HLM × Arabic	0.024
	(0.036)
HLM × Asian	0.027
	(0.078)
Intercept	0.070**
	(0.035)
R-squared	0.864
N	11519

Note:The dependent variable is the share of neighbors from the same ethnic group as new movers in a given housing block. Additional controls are a quadratic function of age, gender, hourly wage (in log), education, occupation, housing block socio-economic characteristics and department fixed effects. Standard errors, clustered at the residential block level, are in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%.

C.3.2 Tests on the refusal rate of public housing offers

The previous tests point out the absence of self-selection along ethnic lines among the movers. But self-selection could occur prior to the move. In this case the sample of movers that we observe in the database would be biased. We address this issue by looking at households that have refused a public housing dwelling offer. We show that even if households declined at least one offer, possibly due to the ethnic diversity or the socio-economic composition of the neighborhood, they were still unable to choose the ethnic and socio-economic composition of the housing block in which they end up living. In this section, we report the results for the refusal rate of public housing offers depending on the ethnic composition of the neighborhood (measured with a standard fractionalization index; see Alesina et al., 1999). Similar (unreported) results are obtained when looking at the composition of the neighborhood by educational levels or occupations.

We run this analysis on an alternative database, the 2002 Housing Survey, to bring additional evidence on the absence of sorting. The Housing Survey (HS) shares exactly the same structure as the Labor Force survey, with information collected at the housing block level with adjacent neighbors (an average of 18.2 in the public housing sector). But the HS also reports specific questions on household satisfaction with housing quality and if they have turned down public housing offers.

If there were self-selection upon diversity, we should expect households that turned down proposals before being allocated to their current public housing dwelling to end up living in less diverse neighborhoods. To test this conjecture, we run OLS regressions of a variable indicating whether the household declined at least one offer (during the latest application process) on the level of ethnic diversity of the neighborhood in which it now lives. Panel A-I of Table 15 shows various estimates of the effect of ethnic diversity on the probability of having turned down offers. Column 1 shows the correlation without any additional control variables. In Column 2, we control for household characteristics. We add up the characteristics of the housing project in Column 3. Column 4 finally includes neighborhood characteristics and *département* fixed effects since the allocation of a public housing dwelling takes place at the *département* level. In each specification, the coefficient on ethnic diversity is not significantly different from zero, showing that households having declined offers during their past allocation process do not end up living in neighborhoods with significantly different levels of diversity.

Table 15: Rejection of Public housing offers and Ethnic diversity

Rows: Dependent Variables	Coeff. associated with Ethnic Diversity (1)	(2)	(3)	(4)
Panel A: Sample of households who currently live in public housing:				
I. Probability of having declined at least one public housing offer during the previous application process	0.058	0.069	0.017	0.123
	(0.058)	(0.063)	(0.067)	(0.0886)
N	1,779	1,779	1,748	1,744
R^2	0.001	0.021	0.023	0.089
II. Probability that the reason for having declined a public housing offer during the previous application was “unpleasant environment”	0.162	0.061	0.017	-0.0310
	(0.144)	(0.158)	(0.171)	(0.258)
N	417	417	415	414
R^2	0.003	0.035	0.050	0.308
Panel B: Sample of households who are currently applying to public housing:				
I. Probability of having declined at least one public housing offer during the current application process	-0.063	-0.043	-0.088	-0.116
	(0.057)	(0.064)	(0.071)	(0.103)
N	1,192	1,192	1,173	1,171
R^2	0.001	0.014	0.024	0.121
II. Probability that the reason for having declined a public housing offer during the current application was “unpleasant environment”	0.004	-0.007	-0.104	-0.122
	(0.194)	(0.237)	(0.250)	(0.506)
N	198	198	195	194
R^2	0.000	0.083	0.115	0.590

Note:Each of the coefficients is estimated from a separate regression of each of the four dependent variables described in the first column on an ethnic fractionalization index at the housing block-level. Column 1 does not include any control. Column 2 includes households characteristics (gender, age, education, employment status and nationality of the head of household, total income (in log) of the household per unit of consumption, and household size). Column 3 adds up the characteristics of the building (number of apartments (in log) and construction date). On top of that, column 4 includes neighborhood characteristics (socio-economic background (Tabard index), and local unemployment rate), as well as *département* fixed effects. In addition, a dummy variable indicating whether the household already lives in the public housing sector is included in specifications 2 to 4 of Panel B. Standard errors, clustered at the residential block level, are in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%.