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

Yann Algan† Thierry Mayer‡ Mathias Thoenig§

July 17, 2012

Abstract

This paper estimates a structural model of cultural transmission by looking at the cultural type of the names given by parents to their newborn children. More specifically, we focus on the transmission of Arabic name versus Non-Arabic name in the French society. We disentangle the relative contribution of the vertical transmission from the parental culture, the horizontal transmission from the neighborhood culture and the economic cost associated to names that sound culturally distinctive in the French society. Our main contribution is to estimate the extent to which parents are willing to allow for a deterioration of expected economic outcomes in order to satisfy their desire to transmit their own trait, or to conform to the dominant trait in the surrounding area. We also make progress in the identification of the channels of cultural transmission by using the French Labor Force Survey that collects data on names and socio-economic characteristics of all the individuals living within the same blocks. To control for endogeneous residential sorting, we focus on the naming choice of parents who are randomly allocated across social housings. We find that parents do take into account the expected economic cost that they inflict to their child by choosing a culturally distinctive name. By transmitting an Arabic Name, the parents are ready to give up one year of expected average income for their child in order to maintain their cultural trait.

*We thank Esther Duflo for initial discussion on the motivation of this paper, and Laurent Gobillon for a very helpful discussion. We also thank participants to seminars and conferences in RSAI in Miami, the NBER Summer Institute in Boston, INED in Paris, UCD in Dublin, UAB and Pompeu Fabra in Barcelona for their helpful comments. The research leading to these results has received funding from the European Research Council for Yann Algan under the European Community Seventh Framework Programme (FP7/2007-2013) / ERC grant agreement 240923.

†Sciences Po and CEPR.

‡Sciences Po, CEPII and CEPR.

§University of Lausanne and CEPR.

1 Introduction

In this paper we estimate a model of inter-generational cultural transmission at the household level. The cultural trait under scrutiny corresponds to the type of first name given by parents to their newborn children. We focus on the desire of parents to transmit their own trait, potentially conflicting with peers' influence, and the adverse impact that a very distinct name can have on children future wage/employment status. Our main contribution is to estimate the extent to which parents are willing to allow for a deterioration of expected economic outcomes (for the child) in order to satisfy their desire to transmit their own trait, or to conform to the dominant trait in the surrounding area. We consider this analysis as a first attempt to identify the economic determinants of cultural transmission, while the previous literature has focused so far on the effect of culture on economic outcomes (see Fernandez, 2008).

The analysis of naming decisions is an ideal experiment to probe into the interplay between economic incentives and cultural transmission. The first name is a crucial marker of cultural identity. As stressed by sociologists (Lieberson, 2000) the choice of first names are available to all parents, without material constraints, and thus are a pure expression of cultural identity. Levitt and Fryer (2004) have provided additional evidence on the cultural component of first name by showing that the surge in distinctively Black names in the US since the seventies could be associated to a rise in the Black cultural identity. But the first name is also perceived as a signal for the employer about the cultural and socio-economic background of the individual and can have direct economic consequence. 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). We add to this literature by estimating to what extent parents do take into account this economic cost in their naming decision.

We start by providing a theoretical model of naming decision. We incorporate the two traditional vertical and horizontal channels analyzed in the literature on cultural transmission (see Bisin and Verdier, 2001), but include on top of that an economic channel. The vertical transmission channel results from the utility gain for *parents* of 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.

In a second stage we estimate our structural model of naming decision. Our dependent variable is the type/cultural category of the first names of babies born in France between 2003 and 2007. We focus in particular on the transmission of Arabic names as opposed to non-Arabic names in the French society. Non-Arabic Names in France are mainly associated with Saint Names that come from the French calendar of Saints. They are deeply associated with the French culture and are typically hold by natives whose both parents and grand-parents were born in France. By opposition the Arabic names are associated with the most important population of immigrants in France after World War Two, in particular in the aftermath of the decolonization initiated in the 1960s. Those names capture a cultural heritage that is potentially the most conflicting with that of the French culture. First 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. Second, they are associated with a hatred decolonization process in the French history, exemplified by the war of Algeria (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 Duget 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 their offsprings with what can be perceived as a lack of integration with the dominant culture.

From this perspective, our key variable of interest is the choice of the first name instead of the last name since the latter one is not manipulable by the parent. Obviously the last name is also a marker of cultural identity. Besides, the last name can be correlated with productive characteristics that affect the economic penalty associated with culturally distinctive first name. Thus in the robustness analysis, we control for the nationality at birth of the parents as a measure of the origin of the last name and of ethnic backgrounds of individuals (color of skin, language...). This allows us to isolate, for a given last name, the specific signal associated to the transmission of a culturally distinctive first names.

In the last part of our paper, we provide several quantification and simulation exercises. We try in particular to evaluate the effects of different counterfactual experiments, which might be relevant in policy terms. Based on our structural model, we build various scenarios to assess the dynamic interplay between labor market policies, public housing policies and cultural transmission. For instance, we look at the equilibrium impact of lowering the extent of discrimination on the labor market for arabic-named individuals when naming patterns are endogenous.

In terms of results, we show that economic factors deeply shape the individual decision of cultural transmission. If the vertical channel associated with the culture of origin of the parents plays a key role in the cultural transmission, in line with the previous literature (Fryer and Levitt, 2004), we also find that parents do take into account the economic costs or advantages associated with their cultural trait in the naming decision of their babies. In particular, the unemployment penalty associated to Arabic name holders (which on the French labor market is around 5 percentage points) 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 annual number of babies born with an Arabic name would increase by almost 60 percent, from the observed 201,895 births to a counterfactual 322,159 births.

Our empirical analysis is based on the French Labor Force Survey (LFS henceforth) from 2003 to 2007. The LFS combines three main features that provide a unique opportunity to estimate the various channels of cultural transmission. First, the LFS is a representative rolling panel of 6 quarters at the level of the households that reports the first names of *all* the household members, including children, and report detailed information on socio-economic characteristics. This unique information makes it possible to measure the vertical transmission channel. Secondly the LFS enables us to get various proxies for the economic penalty associated with Arabic name as perceived by the parents. Thirdly the LFS enables us to use residential neighbors as the source of information for the economic and horizontal channels since the data collection is based on close neighbors. The sampling unit in the LFS consists of groups (which we refer to as blocks) of adjacent households (on average 20 of them). In each wave, all the households and all members of the households, within the same block are interviewed. The LFS thus provides detailed information on both first names and socio economic characteristics of all the other individuals living in the block. We perceive this as a considerable data-related improvement upon the existing studies, which mostly look at neighborhood effects using spatial information from census data, and therefore have to rely on large areas (counties for instance) to define neighbors. Since close neighbors are likely to have a much stronger influence than distant neighbors on the individual decisions (a statement that we are able to quantify properly in our empirical analysis), our paper provides more accurate estimates of the horizontal and economic channels in the transmission of naming patterns.

A salient issue in the identification of the channels affecting the cultural transmission relates to the self-sorting of parents into different economic and neighborhood environments, leading to biased estimates of factors in the naming decision. In particular, two main biases arise from the identification of the economic and the horizontal channels. Regarding the estimation of the horizontal channel, a potential bias come from the self-sorting of parents into given areas. The spatial clustering of Arabic names could thus be potentially driven by unobserved characteristics of the neighborhood. The same issue is raised by the identification of the economic channel. Parents attached to the transmission of the Arabic type to their offspring should, everything else equal, prefer occupations where the economic cost is lower. This potentially leads to a negative correlation between the economic cost and the unobserved heterogeneity in taste for Arabic names. We can partly alleviate this concern by focusing on the information that parents retrieve from their neighbors rather than from their own case. But as for the horizontal channel, potential self-sorting of parents into specific neighborhoods will bias the identification of this source of information. Since the main source of information for the horizontal and economic channel is based on residential neighbors, we need to look at parents who are allocated across housing blocks in an exogeneous way with respect to their cultural type.

In the robustness check section, we make use of the French public housing policy which provides a natural experiment of an exogenous residential allocation of households. Due to a 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 inside the public housing market can be considered to a large extent as exogenous relative to cultural and ethnic characteristics. We then show the robustness of our results when estimations are restricted to the subsample of individuals living in public housing.

Our paper follows three main strands of related research. The first strand deals with the transmission of cultural values and the formation of identity (Akerlof and Kranton, 2000). Bisin and Verdier (2001) provides a seminal model of cultural transmission distinguishing vertical transmission inside the family from parents and oblique or horizontal transmission due to social interactions. Tabellini (2008) and Guiso et al. (2008) model the interactions between norms and economic incentives in the intergenerational transmission of values like trust. But those theoretical models have received little empirical support so far due to the lack of quantitative data on intergenerational transmission of culture. The closest paper to ours is Bisin and al. (2004) who use inter-religion marriages from the GSS to estimate a structural model of transmission of religious values. We distinguish from this literature in important dimensions. First, the naming decision provides an ideal experiment for estimating the determinants of cultural transmission. As emphasized by Lieberson (2000) the naming pattern is a "pure" expression of the parents' choice in the formation of cultural and social identity, since they are free and available to all parents without any material constraints. Alternative attributes of the cultural identity, such as wearing clothes with specific brands or religious values, are in contrast a costly investment that impose material constraints on the identity formation. This is obviously the case for cultural attributes associated with brands of consumption goods that could be manipulated by producers. But this is also true in the context of transmission of cultural values such as religious faith that could be constrained by the availability of churches, mosques or any other worship places in the neighborhood. Besides, naming practices provide an objective quantitative measure of cultural identity. Previous papers have used social surveys on attitudes and transmission of values (Bisin et al., 2010; Bisin et al., 2004), while the act of name-giving is a concrete cultural and social act. Second, our paper provides a structural model encompassing a new channel, associated to economic incentives, to the traditional horizontal and vertical channels discussed in the literature on cultural transmission. Third, previous papers on cultural transmission do not deal with endogenous spatial sorting to isolate the causal impact of the neighborhood (Bisin et al, 2010).

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) study unique first names created by African American parents, and show that if such names reflect well the African origin of the parents (vertical transmission; e.g. root of the name), they are also to a large extent influenced by American norms in naming patterns (oblique transmission; e.g. sounding of the name). More recently, economists have sought to explain naming pattern, controlling more specifically for parents socio-economic background. Fryer and Levitt (2004) provide evidence that names given to children are an expression of cultural identity: they argue that the sudden rise in the choice of distinctively Black names in the 1970's in the USA was a way to reinforce Black identity in the midst of the Black power movement, in line with the predictions of an identity model [1]. Our

[1] They follow the Akerlof and Kranton (2000) identity model, where individuals gain a utility benefit when they behave according to what is "prescribed" by their type.

paper differentiates from Fryer and Levitt (2004) by identifying the interplay, and potential conflict, between economic and cultural incentives in the inter-generational transmission of culture. Fryer and Levitt (2004) estimate the average economic costs of distinct African American names, but do not consider how economic factors affect the individual decision to transmit a naming pattern. Head and Mayer (2008) measure the respective influence of socio-economic distance and geographical distance for the differences in naming patterns that emerge across French regions. While they are able to analyze nearly half a century of naming decisions, the lack of individual-level data makes it impossible for them to investigate the vertical and horizontal transmission channels in a sufficiently detailed way. Goldin and Shim (2004), focus on women's choice to keep or change their surname after wedding or child birth. Here, social determinants are examined (e.g religion, in laws) along with economic determinants (all the characteristics affecting her position/career achievement, such as education or age). They find that a woman will be more likely to keep her maiden's name if she has "made a name" in her career, highlighting the importance of economic background. We distinguish from this paper by focusing on the interplay between norms and economic incentives in the intergenerational transmission of first names.

The third related literature focuses on the effect of culture on economic outcomes. There is growing evidence that cultural values could shape individual economic behavior and affect aggregate economic outcomes (see Fernandez for a recent survey, 2008). But we still have scant information on how the economic environment and economic incentives could shape cultural values and change their transmission pattern. Our contribution to this literature is that we are –to the best of our knowledge– the first to estimate how economic incentives affect individual decisions in the transmission of cultural values.

The remainder of the paper is as follows. Section 2 presents our theoretical model of naming decision, that will guide the estimated equation. Section 3 provides a detailed description of the data we use, and Section 4 presents the benchmark estimation results. Sections 5 and 6 provide robustness checks for the identification of the economic and horizontal channels of cultural transmission. Section 7 proposes several quantification and simulations exercises based on our model of naming decision, fed with structural parameters estimated from the econometric work.

2 Model and Identification of Naming Decision

2.1 A simple model of baby name choice

In this section we build a model of baby naming decision. We provide a framework meant to be 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 name. In most of our analysis, the cultural type corresponds to being from Arabic culture/origin. The utility for a household i, member of social network k, in year t derived from choosing a given name type for its baby is defined as $U_{ik,t}(1)$ if the name is Arabic and $U_{ik,t}(0)$ otherwise,

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

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 our empirical analysis we consider two types of social network k: the residential neighborhood and the professional occupation. Potential self-selection of households into their social network raises important identification issues that we address below.

In such a discrete choice setting, only differences in utility over alternatives can be identified in the data (for a recent survey of discrete choice models with social interactions, see Blume, Brock, Durlauf and Ioannides, 2010). 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

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

where $\Delta V_{ik,t}$ is specified as a three part linear function, which we label "Vertical", "Horizontal", and "Economic cost" channels of influence. $\text{Parents}_{ik,t}$ is a parental attribute equal to one when the name of one of the two parents is Arabic and zero otherwise, while $\text{Baby}_{jk,t}$ codes for choices of baby names by the $\mathcal{N}_{k,t}$ other parents of the social network k. 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 naturally much less salient (see Section 3.2). The second RHS component reflects social influence, i.e. the share of parents in social network k expected to make the same choice as i, with intensity α_2 expected to be positive. The third RHS component relates to economic incentives: $\mathcal{C}_{k,t}$ is the expected economic loss resulting from future discrimination toward Arabic name holders on the labor market. This loss is determined by the economic environment specific to the social network k and is thus not conditioned by individual characteristics. Presumably the higher it is, the less parents want to attach an Arabic cultural type to the name of their babies.

The coefficient α_3 relative to α_1 in particular reflects the parental tradeoff between their own attachment to a particular cultural type and their altruistic concern toward the future economic performance of their babies. There are a number of identification issues raised by the estimation of the horizontal channel and economic channel in equation (2), to which we now turn.

2.2 Estimation issues

2.2.1 Horizontal channel

The horizontal transmission channel causes several estimation issues well-known in the social interaction literature (see Blume et al. 2010 for a very complete survey). Indeed $\text{Horizontal}_{ik,t}$ depends on the expected realization of $\text{Baby}_{jk,t}$ of socially connected parents $j \in k \neq i$, that is on their utility differential $\Delta U_{jk,t} = \Delta V_{jk,t} + \varepsilon_{jk,t}$.

- *Reflection problem*: The $\Delta V_{jk,t}$ term includes the choice of household i (expected by j), inducing a *"reflection problem"*. If the social network k is small enough such that household i is not negligible, we have the classical Manski (1993) reflection problem. This reverse causality issue can be resolved using an IV strategy, where natural instruments of $\text{Horizontal}_{ik,t}$ are $\text{Parents}_{jk,t}$ (they should not influence the choice of household i *directly* but through the choice of surrounding parents j). Another way to mitigate the problem is to assume that parents i form their expectations on lagged decisions of neighbors such that

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

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 utility differential becomes

$$\Delta U_{ik,t} \;=\; \alpha_0 + \alpha_1 \underbrace{\text{Parents}_{ik,t}}_{\text{Vertical}_{ik,t}} + \alpha_2 \underbrace{\sum_{\tau=1}^{10} \sum_{j\in k\neq i} \frac{\text{Baby}_{jk,t-\tau}}{\mathcal{N}_{k,t-\tau}}}_{\text{Horizontal}_{ik,t}} + \alpha_3 \underbrace{\mathcal{C}_{k,t}}_{\text{Economic Cost}_{k,t}} + \varepsilon_{ik,t}. \quad (3)$$

- *Social sorting*: The horizontal channel, $\text{Horizontal}_{ik,t}$, now depends on realizations of $\text{Baby}_{jk,t-\tau}$, and therefore on $\Delta U_{jk,t-\tau} = \Delta V_{jk,t-\tau} + \varepsilon_{jk,t-\tau}$. *Social sorting* might lead to a non-zero correlation between $\varepsilon_{ik,t}$ and $\varepsilon_{jk,t-\tau}$ for households i and j belonging to the same social network k. This would create a correlation between $\text{Horizontal}_{ik,t}$ and the error term in (3). Indeed $\varepsilon_{ik,t}$ captures unobservable taste shocks for the considered cultural type for 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 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 is also potentially partly driven by unobserved characteristics of the area. Our estimates could thus be biased by the endogenous spatial sorting of households. In the robustness check section, we will address this issue by exploiting the random allocation of households to their neighborhoods in the French social housing. We will restrict our analysis to the sub-sample of households living in social housing and therefore randomly allocated inside a given département. For those, the change in the composition (and therefore choices) of neighbors is considered exogeneous.

2.2.2 Economic cost channel

Let us now consider the issues raised by the expected economic loss of an Arabic name, $\mathcal{C}_{k,t}$. This cost can be decomposed into two components, a part observable to the researcher, $c_{k,t}$, and a random/unobserved part $u_{k,t}$. Typically, $u_{k,t}$ can arise because of spatial variation in the degree of discrimination on the local labor market, which is known to the parents,[2] but unknown to the econometrician:

$$\mathcal{C}_{k,t} = c_{k,t} + u_{k,t}. \tag{4}$$

When replacing in (3), utility becomes

$$\Delta U_{ik,t} = \alpha_0 + \alpha_1 \underbrace{\text{Parents}_{ik,t}}_{\text{Vertical}_{ik,t}} + \alpha_2 \underbrace{\sum_{\tau=1}^{10} \sum_{j \in k \neq i} \frac{\text{Baby}_{jk,t-\tau}}{\mathcal{N}_{k,t-\tau}}}_{\text{Horizontal}_{ik,t}} + \alpha_3 \underbrace{c_{k,t}}_{\text{Economic Cost}_{k,t}} + \delta_{ik,t}, \tag{5}$$

where $\delta_{ik,t} \equiv \alpha_3 u_{k,t} + \varepsilon_{ik,t}$ is the new error term.

- *Estimation of the economic cost*: The first issue is the identification of the economist cost $c_{k,t}$ in a first stage regression. A measurement error in the economic cost in the first stage will result in a bias when estimating its impact on naming pattern in the second stage equation in (5).

  The economic cost we are interested in is the economic cost perceived by the parents when they transmit an Arabic name to their offspring. The information on the economic cost $c_{k,t}$ is retrieved from their network k (defined as their professional occupation in our empirics). From the parents' perspective, their offspring can suffer from two types of discrimination. The first one is known as statistical discrimination. Presumably, the first name should have no direct impact on productivity of individuals. However, a name type is likely to be correlated with some unobservable characteristic that affect productivity (i.e. social capital, language skills...). Employers can use the name type to detect extract unobserved productivity information, applying statistical discrimination. Second, a first name can also be hit by "pure" discrimination (by taste). As an economic cost variable, we want to capture both effects, since both effects will be relevant for the economic well-being of the children being named. We shall stress that our objective is not to decompose those two sources of discrimination, but to estimate their overall effect. Yet, the main issue is that the econometrician observes a potentially different set of information about the productivity determinants than do the employers

[2] Assuming either that their children will stay in the same area when entering the labor market, or that the parents assume that on average the kid will face the local level of discrimination.

or the parents. The most likely situation is that the econometrician has less observables at hand. Suppose that those omitted variables are positively correlated with arabic names, then the economic cost of a name will be overestimated with respect to the real one (perceived by parents). An opposite correlation or a larger set of observables for the econometrician would result in a downward bias. Both first-stage issues would result in measurement error of economic costs, and therefore will result in an attenuation bias when estimating its impact on naming patterns.

To address this identification issue, we exploit difference in the economic cost across occupation and within time. The economic cost $c_{k,t}$ is estimated in the within dimension from exogenous (at least to the individuals) temporal variation over t for each occupation k. We consider this strategy as a solution for identification for the following reasons. First, differences in the set of information about the observed productivity determinants presumably have more limited variability across occupation. Second, we have no reason to suspect that those differences in the set of observable characteristics should vary over time within each occupation. Therefore, identifying in this dimension should purge most of the bias between the real and estimated economic cost of a name.

- *Identification of the economic channel*: In the second stage estimate of equation (5), the coefficient α_3 associated with the economic cost of a name type, $c_{k,t}$, is also potentially ill-estimated due to self-selection into occupations by households.

  Again, endogenous social sorting of parents might create a selection bias. Parents attached to the transmission of the Arabic type to their offspring should, everything else equal, prefer social networks where the economic cost $c_{k,t}$ is lower. This potentially leads to a negative correlation between $c_{k,t}$ and the unobserved heterogeneity in taste for Arabic names $\epsilon_{ik,t}$ included in $\delta_{ik,t}$.

  We address this issue by controlling for occupation 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. The impact of the economic cost on naming decisions is therefore identified through time country-wide variation of economic penalty within a given occupation.

2.3 Estimation methods

The empirical section will estimate the utility function (5) while addressing the identification issues above mentioned. Before proceeding to the estimation, we need to go from the utility function (5) to an estimable discrete choice equation. To obtain a closed-form solution of the choice probability, a logistic distribution for the error term $\delta_{ik,t}$ is specified, with σ the scaling parameter and μ the location parameter of this distribution. One can then express the probability of choosing an Arabic name as a classical logit model

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

The observable utility differential $\Delta V_{ik,t}$ is retrieved from the coefficients in (5) that can be readily estimated with logit (up to the scaling parameter). This also enables to run counterfactuals without probabilities going out of bound.

3 Data

3.1 Database

Our 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 blocks (aires in French) 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. The LFS records the first names of all the household members alongside with their detailed individual economic characteristics. All the household members aged above 15 year old are interviewed and report their information on their socio-economic characteristics and first names. In addition, the survey records the first names of all the children in the household who are aged below 15 years old. This information makes it possible to estimate the economic and cultural factors in the naming pattern at a detailed individual level, contrary to Fryer and Levitt (2004) who look at county level economic outcomes associated to first names. Second, the LFS makes it possible to identify the determinants of the vertical transmission of first names, since we have information on the first names of both parents and children and we observe the choice of the first-name given to a child at birth. Third, the LFS provides a unique opportunity to understand the role of horizontal factors in the transmission of names since the data collection is based on (very) close neighbors. Given that the sampling unit in the LFS consists of groups of adjacent households, and that all the members of the households, within the same block are interviewed, we get detailed information on the first names of all the other individuals living in the close neighborhood.

The LFS also provides detailed information on the socio-economic characteristics of the household members, including the employment status (unemployed, inactive and employed), the hourly wage and the 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.

The time span of the rolling panel is too short (6 quarters) to get enough time variation in the socioeconomic composition within the blocks and to exploit the panel dimension of the LFS. We thus keep one observation per member of the household. 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.

Table 1 reports the main descriptive statistics of the whole database when we use the previous selection criterion. Our total sample is made up of 10,541 blocks, with 1,535 blocks belonging to the state-owned housing market. Each block consists on average of 18.31 adjacent households, each household being composed by around 3.31 members (babies, children and adults included). Overall, the total sample includes 425,210 individuals, among whom 69,458 are living in social housing.

Table 1: Descriptive statistics of the blocks

	Total sample	Public housing
Number of blocks	10.541	1.535
Number of blocks by department	174.35	45.12
Average number of households per block	18.13	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	173,154	26,749
Total number of individuals	425,210	69,458

3.2 Sample of babies' names

Our main variable of interest is the individuals' name type and the cultural component that is associated to it. In particular, we focus on the transmission of Arabic names, as opposed to non-Arabic names, in the French society. This focus is motivated on two main grounds. First Arabic names are associated with the most important population of immigrants in France from Maghreb and Sub-Saharan Africa and to a lesser extent from the Middle East (mostly from Turkey), in the aftermath of the decolonization initiated in the 1960s. We identify the Arabic names by using the classification of Jouniaux (2001). 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 such as the independence war in Algeria. In addition to the name type, we have information on the country of birth and the nationality at birth of the respondent and of the parents' respondents. The survey distinguishes between 29 countries of origin. Since we focus on the broad category of Arabic names, we group together the country of origin and the nationality at birth in an corresponding category of Arabic origin, made up of immigrants from Maghreb, Sub-Saharan Africa and Middle-East.

Table 2 shows the descriptive statistics of the sample of babies. We observe 3,387 babies for whom we have all the relevant information on the parents' and blocks' characteristics. 3,072 babies (90.7%) receive a non-Arabic names, and 315 babies are given an Arabic name (9.3%). Among the group of babies with non-Arabic names, 59.8 percent 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 Saint Names, or names deeply ingrained in the French culture like Leo for boys or Manon for girls. The majority of babies who are given an Arabic name have at least one parent with also an Arabic name (76.20%). The choice of an Arabic name is thus mostly, but not exclusively, attached to parents from an Arabic culture. This does not mean nonetheless that those parents systematically chose to transmit to their children a distinctive Arabic name. Their naming decision is rather balanced since 52.1 percent of those parents give an Arabic name to their offspring, the other 47.9 percent of this group choosing a non-Arabic name. Thus 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.

Table 2: Characteristics of the sample of babies

	Babies with:	
	Arabic name	non-Arabic name
Number of babies	315 (9.3%)	3,072 (90.7%)
of which:		
Parents with non-Arabic name	73 (2.1%)	2,875 (84.9%)
Parents with Arabic name	223 (6.6%)	216(6.4%)

Note: This table shows the number of babies born with the two name types (percentage) and allocates them according to the name type of their parents. The sample consists of the 3387 babies used in the estimations of the determinants of names.

4 Benchmark Results

This section documents the benchmark estimates of the econometric choice model characterized by equation (5). We document the contribution of the economic channel relative to the vertical channel arising from parental culture transmission and to the horizontal channel of influence of cultural traits from the neighborhood in the naming decision process. In the benchmark estimates, we start with a simple measure of the economic cost of a name, defined as the unconditional unemployment rate differential between Arabic and non-Arabic name holders.

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

4.1 Assessing the cost of a name

In the baseline regressions, we start with a basic measure of the observable part of the labor market penalty $c_{k,t}$ associated with an Arabic name that could be subjectively assessed by the parents.[4]

We consider two sources of information that the parents could use in the evaluation of the economic penalty. The main one is the economic penalty that they observe in their own occupation. The main indicator is thus the economic penalty by parental occupation. We also consider the possibility that the parents draw information about the economic penalty associated with an Arabic name from the block they live in. We therefore calculate a block-specific index of the economic cost associated with an arabic name, weighting the occupational economic penalty by the composition of occupations of the households living in the same block k.

As discussed in the theoretical model, we are interested in the overall economic penalty associated with both statistical discrimination and discrimination by taste. But a potential bias in the estimation of this economic penalty comes from the difference in the information set about the productivity that we have at hand and the one used by the parents in their naming decision. The most likely situation is that the econometrician has less observables at hand. Suppose that those omitted variables are positively correlated with arabic names, then the economic cost of a name will be overestimated with respect to the real one (perceived by parents). An opposite correlation or a larger set of observables for the econometrician would result in a downward bias. Both first-stage issues would result in measurement error of economic costs, and therefore will result in an attenuation bias when estimating its impact on naming patterns. We purge this bias by exploiting the variation in the economic cost across occupation and within time.

Our identification assumption is that the bias is the same for each occupation and does not vary over time within a given occupation. Indeed, differences in the set of observable productivity determinants have all chances to have limited variability across occupation. For instance, an employer in a client-intensive sector might be very sensitive to the fact that the candidate speaks French fluently. The econometrician does not observe this language skill, but *at the CV stage*, neither does the prospective employer. We therefore will estimate our penalty at the earliest stage of the labor market relationship we observe (employed / unemployed).[5]

In the baseline regressions, we measure the employment penalty as the unconditional differential

[4] Ultimately the expected economic penalty corresponds to the current value of the differential in intertemporal flows of income between an Arabic name holder and a non-Arabic name holder, other individual characteristics being equal. It is assessed *subjectively* by parents and depends on their information set and on the future social and spatial mobility of their baby. A full-fledged estimation procedure of this cost would require to estimate a structural Bayesian model of intergenerational social mobility and labor outcomes, which goes far beyond the scope of this paper.

[5] Labor force participation could be partially explained by extra-economic characteristics. In particular, as shown by Fernandez (2008), labor market participation decisions are also related to cultural values like the gender division of roles. The wage penalty does not suffer from this criticism and could be considered as a true economic cost by parents. Yet from the perspective of the parents, it seems to be more difficult to observe the wage gap between Arabic and non-Arabic name holders in the neighborhood than observing the employment/unemployment status of the neighbors. As such, the wage gap might be an information that is less relevant than the employment status in the choice of a Name. Note also that once the worker has been hired, a lot of unobservable productivity determinants will become observable, and therefore we should expect a lower estimate of the impact of an Arabic name for wages than for employment/unemployment.

of unemployment between Arabic and non-Arabic name holders. In the robustness checks section, we will consider the conditional unemployment penalty by controlling for parental characteristics. The estimates are based on individuals aged between 25 and 55 years in the LFS over the 2003-2007 period.

Table 3 reports simple descriptive statistics on the unconditional unemployment rate among Arabic Names and non-Arabic names. We document both the average rates and the rates by occupational categories. For the sake of clarity, we group together all the different occupations listed by INSEE into 7 main categories: farmer, craftman, unskilled blue-collar, skilled blue-collar, clerk, intermediate,and executive. On average, Arabic name holders have an unemployment rate of 19.2 percent, more than three times as high as the unemployment rate of non-Arabic name holders (6 percent). But this average comparison hides a lot of variation across occupations. The unemployment rate of Arabic name holders among executives is only 6 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 18 percent and 25 percent among the clerks and the blue collars (unskilled), which represent an unemployment gap of 12 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 various occupations registered in the French national statistics. This variation of employment penalty across occupations also makes it possible to identify the impact of the economic cost by exploiting composition changes across blocks in the occupations of households.

Table 3: Unemployment rates by Name type and Occupation

	Unemployment rate:	
	Arabic name	non-Arabic names
Executive	0.06	0.03
Intermediate	0.11	0.03
Clerk	0.18	0.06
Blue collar (skilled)	0.16	0.05
Blue collar (unskilled)	0.25	0.13
Craftman	0.13	0.03
Farmer	0.25	0.00
Total	0.19	0.06

4.2 Determinants of the naming decision

We now turn to the estimate of our econometric choice model characterized by equation (5). Table 4 presents basic estimates on the full sample. The dependent variable is a binary variable coding for the Arabic origins of a baby's name; explanatory variables relate to various parental and spatial characteristics.

As discussed in Section 2, we identify the effect of the economic incentives in the parental naming decision within parental occupation by including in all specifications fixed effects for the parental occupation. Self-sorting of parents within occupation could raise a selection bias. The occupation fixed effects should address this bias by capturing all the time-invariant co-determinants of the parental occupation choice and the naming pattern. The impact of the economic cost on naming decisions is therefore identified through time variation of economic penalty within a given occupation. The precise causes to those changes in economic penalty attached to Arabic names goes beyond the scope of the paper, and we will take them as given[6]

Table 4: The choice of an Arabic name : Full Sample

	(1)	(2)	(3)	(4)	(5)
Dep. Var:			Arabic name for baby		
at least one parent same name type	0.16^a	0.16^a	0.15^a	0.16^a	0.15^a
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
share of name type in block (aged 1-10)	0.07^a	0.07^a	0.07^a	0.08^a	0.07^a
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
occupation unemployment penalty	-0.30^a	-0.30^a	-0.30^a	-0.27^a	-0.27^a
	(0.09)	(0.09)	(0.09)	(0.09)	(0.09)
share of name type (aged 1 to 10) in par. occ.		-0.06			
		(0.24)			
share of name type in dept (aged 1-10)		-0.02			
		(0.10)			
share of name type in block (aged 11-25)			-0.01		
			(0.03)		
share of name type in block (aged 26-49)			0.04		
			(0.04)		
share of name type in block (aged 50+)			-0.03		
			(0.03)		
block unemployment penalty				-0.58^a	-0.25
				(0.22)	(0.22)
Observations	3387	3387	3304	3387	2524731
R^2					
Pseudo R^2	0.441	0.441	0.441	0.444	0.471
meanprob	0.09	0.09	0.10	0.09	0.09

Note: logit estimates (average marginal effects). Standard errors parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%.

Table 4 shows the benchmark estimates on the full sample of babies born over the 2003-2007 period. Our main estimation method throughout the paper is logit and we report the average

[6]Potential explanations include large political shocks that occurred in France in the years of our sample (2003-2007). The first is the 2002 election where the extreme right-wing party's candidate got selected for the second round of the presidential election. Second, and maybe related, the run-up to the climate of the presidential election of 2007, where a large share of the debate has been focused on security and immigration issues. We leave for future research the analysis of those factors.

marginal effects in the tables. We also control for spatial fixed effects at the department level. Column (1) shows that all three coefficients of interest are significant at the 1 percent threshold and have the expected sign. The t-stat of the coefficient on vertical transmission is 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 3.2.

The coefficient associated to the employment penalty is negative and statistically significant confirming that parents take into account the employment penalty of Arabic names in their own occupation when they transmit a first name: they are deterred from inflicting an economic cost to their offspring. We provide below several measures for the order of magnitude of this effect.

The horizontal transmission channel is also quite significant, and has an interesting spatial feature. In column (2) we add two components to the horizontal channel: i) the share of arabic names for kids under 11 in a larger area (called a *département*, 95 of those on continental France) and ii) the same share in the parents' occupation, independently of their location. Those two variables are intended to see how far does the horizontal transmission channel spans in space, the first one being a wider regional unit, the second considering the relevant network to be occupation-based, with no impact at all of space. 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 detailed geographical level, which is impossible with data usually at hand in the social interaction literature. Column (3) 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 cohort of kids under 11 years old has a statistically significant impact in their naming decision. This makes us confident when concluding that the coefficient of this variable captures horizontal influence and is unlikely to be massively contaminated by spatial sorting.

Column (4) reports the results when we add our measure of block-level penalty. The observable sources of variation that we exploit are variations in the composition by occupation of each block, given that this composition is likely to capture both the information set and the segments of the labor market parents (and future adults) have access to. This variable seems to add to the overall explanation of naming patterns, while altering very little the other coefficients. In column (5) we present the results for a weighted logit, where the individual representativeness weights reported in the LFS are applied. As discussed in Section 3, the labor force survey is stratified at the *département* level and representativeness is thus not guaranteed at the block level, our level of analysis. Our 3325 births in column (5) (the ones for which we are able to find LHS and RHS variables) therefore represent about 2.5 million births at the national level over the five years of our sample.[7] Most coefficients are only slightly affected, with the noticeable exception of the block unemployment penalty, which is divided by more than two and loses statistical significance.

[7] This represents an average number of births around half a million per year, for a true figure slightly under 800,000 in this period (`http://www.ined.fr/fr/pop_chiffres/france/naissances_fecondite/naissances_par_sexe/`).

5 Robustness results for the economic channel of cultural transmission

This section digs further into the economic determinants of cultural transmission. We test the robustness of the results when using the conditional unemployment rate differential between Arabic and non-Arabic name across occupation. This involves estimating Mincer-type equations from the Labor Force Survey to retrieve the conditional differential of unemployment rate controlling for a large list of individual and parental characteristics. We show that this conditional unemployment penalty is consistent with those found in audit studies on the French labor market. We then show that the magnitude of the economic channel remains unchanged with this new measure of the economic cost of Arabic names.

5.1 Conditional Unemployment penalty

Table 5 shows the estimated employment penalty, controlling for the nationality at birth of the respondent and parent's respondent, other individual characteristics (age, age squared, education, occupation, marital status and number of children), departmental fixed effects and year dummies. The left-hand-side variable is equal to 1 if the respondent is employed, and 0 if unemployed.[8] Column (1) shows 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 9 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 discriminant 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. Since we want to estimate 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 Sub-Saharan Africa or Maghreb, and 0 otherwise. The estimated employment penalty associated with an Arabic name falls to 5 percentage points and remains highly statistically significant. Remarkably enough, the estimated employment penalty associated with an Arabic Name is of the same order of magnitude as the one associated with having an African or Maghrebian nationality, suggesting that the first name has a specific effect of its own on labor market outcomes.[9]

[8]We use LPM estimation for this intermediary step of our work. When constructing the conditional $c_{k,t}$ used in the second step, we run regressions year by year, with the most detailed occupation classification. Logit/Probit estimation then drop several of our Arabic name/occupation interaction terms which are perfect predictors, while LPM enables to estimate those. Furthermore, each of those year-specific regressions include a quite large proportion of fixed effects (department, occupation and occupation interacted with Arabic first name), which does not mix well with limited dependent variable techniques. Last, it is to be noted with Angrist and Pischke (2009) that marginal effects are generally very proximate when using LPM or Logit / Probit.

[9] Note that our results are different from Fryer and Levitt (2004) who find no effect of distinctively black names on labor market outcome. A potential explanation is that our database allows us to link *individual names with individual outcomes*. In contrast, Fryer and Levitt (2004) can only look at the effect of names on aggregate outcomes at the

Column (3) shows the employment penalty associated with an Arabic name for each broad occupational category. Since the reference category is executives, the dummy for Arabic name measures the unemployment penalty of bearing an Arabic name for executives. The sum of the Arabic dummy plus the interacted term between the Arabic dummy and each occupational dummy measures the employment penalty within the corresponding occupational category. Column (3) shows that while the employment penalty is not statistically significant among the executives, it becomes significant and economically sizable within the least skilled categories. Arabic name holders have a lower probability of being employed, relative to non-Arabic name holders, of 5 percentage points for intermediates, 7 percentage points for clerks, and 8 percentage points for unskilled blue collars.

We provide a simple "back of the enveloppe" calculation to assess the income loss associated with the employment penalty of Arabic names holders. We use Breuil-Grenier's (2001) detailed estimates of the income variation induced by a transition from employment to unemployment on the French labor market. The author takes into account the level of all the social benefits associated with an unemployment spell and the eligibility criteria. She finds an average income loss of 50 percent. The average monthly income of an employed individual in the 2003-2007 LFS is 1,488 euros. Unemployed people thus gain on average 744 euros per month. From Table 5 - Column (2), we know that every period, the conditional unemployment gap of Arabic name holders is 5 percentage points relative to non-Arabic name holder. They thus have an expected income loss of $0.05 \times 744 = 37.2$ euros per month. Since the average participation to the active population is 39 years in France, this means that the total income loss of typical Arabic name holder during their active life reaches $39 \times 12 \times 37.2 = 17,409$ euros, which is roughly equivalent to the loss of one year of average income.

Our estimates are comparable to estimates in the labor economics literature, using a more elaborate set of econometric methods. In particular, Duguet et al. (2008) use the pair auditing method 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 distinguishing the effect of being Muslim from the effect of being Maghrebi 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 Maghrebian first names (which received

level of the zip code, which is likely to mitigate the relationship.

Table 5: The penalty of an Arabic name

	(1)	(2)	(3)
Dep.Var:	emp. / unemp. status		
arabic name	-0.09^a	-0.05^a	0.01
	(0.01)	(0.01)	(0.01)
man	0.01^a	0.01^a	0.01^a
	(0.00)	(0.00)	(0.00)
age	0.01^a	0.01^a	0.01^a
	(0.00)	(0.00)	(0.00)
age squared	-0.00^a	-0.00^a	-0.00^a
	(0.00)	(0.00)	(0.00)
nationality at birth (self or parents) from Africa or Maghreb		-0.06^a	-0.05^a
		(0.01)	(0.01)
arabic name × intermediate			-0.05^a
			(0.01)
arabic name × clerks			-0.07^a
			(0.01)
arabic name × blue collar (skilled)			-0.08^a
			(0.02)
arabic name × blue collar (unskilled)			-0.08^a
			(0.02)
arabic name × craftman			-0.08^a
			(0.03)
arabic name × farmer			-0.06
			(0.09)
Observations	144524	144524	144524
R^2	0.038	0.040	0.045

Note: LPM estimates. Standard errors parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10%. Column (3) have executives as the baseline occupation group. All regressions include dummies for education level, occupation group, department 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%.

2.5 times less positive answers).

5.2 Results with conditional unemployment penalty

Table 6 replicates all the specifications of Table 4 using the conditional unemployment difference associated with Arabic names across occupations for the economic penalty. We still include parental occupation fixed effects in all regressions.

Column (1) shows that all three coefficients associated with the vertical, horizontal and economic channels remain significant at the 1 percent level. The magnitude of the coefficient associated with the economic cost is reduced by 16 percentage points compared to the previous estimate with the unconditional unemployment. But the coefficient remains statistically highly significant. The results are unchanged when we add two components to the horizontal channel: the share of arabic names for kids under 11 in the whole region (*département*) and the same share in the parents' occupation, independently of their location. Columns (4) and (5) show that the results are mostly preserved when we include the block penalty and when we use weighted estimates.

Column (5) of Table 6 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. For the employment penalty for instance, a ten percentage point increase in the estimated occupation unemployment penalty of Arabic names translates into a 2.5 points drop in the propensity to give an Arabic name. The last line of the table gives the mean predicted probability in the sample, around 9%, 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 14 points, when a 10% increase in the share of arabic names in births of the immediate neighborhood increases probability by "only" 0.7 points.

A way to quantify those effects relative to each other is to look at the model's predicted numbers of babies born with arabic names when shutting down each of the three channels in turn. This is done in Table 7, where different lines present different scenarios. The first line reports the true number of arabic names births (287, representing more than 200,000 babies nationally). The second line is the benchmark. We then remove the vertical channel, and are only able to predict 95 arabic naming decisions in that case, that is roughly a third of true births. The horizontal channel removes about 50 decisions, but removing the penalty has a much stronger effect of increasing the birth by almost 60%. The last line does a slightly different experiment. We make as if all blocks in the country had the same neighborhood composition and levels of unemployment penalty. That amounts to consider the predicted number of babies when averaging the horizontal and penalty variables, which induces a slight reduction in naming choices.

Table 6: Robustness result with conditional unemployment across occupations

	(1)	(2)	(3)	(4)	(5)
Dep. Var:			arabic name for baby		
at least one parent same name type	0.15[a]	0.15[a]	0.15[a]	0.15[a]	0.14[a]
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
share of name type in block (aged 1-10)	0.08[a]	0.08[a]	0.07[a]	0.08[a]	0.07[a]
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
occupation unemp. penalty (Mincer-based)	-0.25[b]	-0.26[b]	-0.26[b]	-0.22[b]	-0.25[b]
	(0.11)	(0.11)	(0.11)	(0.11)	(0.10)
share of name type (aged 1 to 10) in par. occ.		-0.06			
		(0.26)			
share of name type in dept (aged 1-10)		-0.04			
		(0.10)			
share of name type in block (aged 11-25)			-0.01		
			(0.03)		
share of name type in block (aged 26-49)			0.04		
			(0.04)		
share of name type in block (aged 50+)			-0.02		
			(0.03)		
block unemp. penalty (Mincer-based)				-0.56[b]	-0.14
				(0.23)	(0.24)
Observations	3325	3325	3243	3325	2477770
R^2					
Pseudo R^2	0.440	0.440	0.441	0.443	0.469
meanprob	0.09	0.09	0.09	0.09	0.09

Note: logit estimates (average marginal effects). Standard errors parentheses with [a], [b] and [c] respectively denoting significance at the 1%, 5% and 10%.

Table 7: The importance of the 3 channels

Scenario:	# babies with arabic name	weighted # babies with arabic name
true count	287	201,895
benchmark	287	206,276
no vertical	95	69,605
no horizontal	239	174,892
no penalty	452	322,159
no ghetto	258	187,072

Note: This table uses logit estimates (col 5 of Table 6). 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 3325 births (representing 2,477,770 nationally).

6 Robustness results for the horizontal channel of cultural transmission

As discussed in the model/identification section, the estimation of the horizontal channel of cultural transmission raises the issue of endogeneous 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. In our context, this phenomenon would imply that if people can choose the block where they live, they would rather move into neighborhoods where people have the same cultural attributes (some of which unobservable) and the same employment occupational status. Those two dimensions of horizontal transmission that we try to identify, namely the share of babies with Arabic names and the occupational composition of each block, are thus likely to be correlated with the error term. One way out of this traditional endogeneity issue is to study individuals who are assigned to their place of residence without consideration of their cultural and economic status. This section tests the robustness of our previous results by exploiting the social housing policy in France, which provides such source of exogenous location of households.

6.1 Exogeneous Allocation of parents in social housing

Due to a strong ideology (very typical of the philosophy founding the French republic) built into law, the French government allocates state-planned cheap rented apartments without concern of the cultural background of the individual, mixing people indiscriminately. The explicit goal of this urban policy is to avoid the formation of homogeneous ghettoes by allocating housing in a random manner, not permitting family networks to grow within housing units.

The social housing are mostly administrated at the department level, the French territory being geographically subdivided into 95 departments. This geographic unit gathers multiple towns and counties. Households who are eligible to social housing have to send their application to the departmental administration of social housings. This makes it unlikely that any personal contact between the applicants and the administrators affects the allocation procedure.

The consequence of this policy is that households who are eligible to social housing are allocated randomly, relatively to their cultural and occupational backgrounds, across the various social housings within a given department. Naturally, there is an endogeneous sorting among the population who are eligible to social housings. Since the rents are considerably lower in social housing than in private housing, there is an over-representation of individuals with lower occupational status and of immigrants. But among the eligible population within a given department, households should in principle be exogeneously allocated across the different blocks.

Table 1 shows the summary statistics of the social housing units in the Labor Force Survey. The survey comprises 1,535 social housings, with an average of 45.12 social housing blocks by department. The average number of households within a block is 17.99, which is similar to the whole sample. We observe 618 new born babies in the social housing sector, which will be used as our sample for the estimates centered on social housing.

Table 8 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 social 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 a Maghrebian nationality at birth, or who have at least one parent with a Maghrebian nationality at birth, is 16.8 percent in social 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 social 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 social housing, and 55 percent in the whole sample) and an under-representation of executives (3 percent in social housing, 12 percent in the whole sample). This table shows that there is a clear selection of individuals into the social housings, since the eligibility is based on the socio-economic characteristics. However, our key identification strategy relies on the exogeneous allocation of individuals *across the different social housings within departments*. An assumption we now formally investigate.

Table 9 provides a formal statistical test for this exogeneity in the allocation of households in the social housing sector. Since social housings are administrated at the department level, we check whether households are allocated randomly across the different social housings of a given department. For that purpose, we regress households' characteristics on fixed effects associated with the different social housings within each department. Our random allocation test consists in performing standard F-test on the null hypothesis that the fixed effects are jointly not statistically different from zero. In the case of endogeneous residential sorting in some social housings, the fixed effects associated with those blocks should be statistically significantly correlated with the household characteristics, and the F-test will be rejected. We run those regressions on the sample of household "heads" who are living in social housings, for each department taken separately. We test the absence of residential sorting along two main households' characteristics that drive the horizontal

Table 8: Descriptive statistics of individuals

	Total sample Mean (std)	Social 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)
Maghrebian Nationality at Birth	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)

transmission of naming patterns in our model. The first characteristic is the Arabic nature of the respondent's name, or 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). The second characteristic is the occupation of the respondent. We regress the probability of being a blue collar (skilled or unskilled) on the social housing dummies. This second test is important since we will identify the influence of the socio-economic backgrounds of the neighbors on the naming decision by using the distribution of employment occupations within the social housings.

Column (1) of Table 9 shows the share of departments for which we accept the null-hypothesis that the social housing dummies are not statistically correlated with the household characteristics. In 86 percent of the departments, the characteristic "Arabic name holder" is not correlated with the social housing fixed effects. In 93 percent of the departments the occupational status is not correlated with the social housing fixed effects. In 96 percent of the departments, the educational level (measured by "No elementary school", this educational level being over-represented in social housing) is not correlated with the social housing fixed effects. In Column (2), we run the same exogeneity tests but on the whole sample of household heads, including both those who live in the public and private housing sectors. In this case, endogeneous residential sorting becomes statistically more significant. In around one third of the departments we fail to reject the null-hypothesis of no statistical correlation between individual characteristic and the block fixed effects.

Table 9: Test of random residential allocation

% departments with random residential allocation relative to households' characteristics

	Social Housing	Total Sample
Household's characteristics		
Arabic names	86.53	67.85
French Nationality at Birth	83.01	66.66
Occupation: blue collar	96.72	67.02
Elementary education	98.36	64.89

Note: The tables report the share of departments for which F-tests reject the null-hypothesis of a statistically significant correlation at the 5 percent level between individual characteristics and social housings fixed effects. The F-test are based on Logistic regression of each individual characteristic on social housing fixed effect within each department. The sample includes household heads aged over 15 years old.

6.2 Results within social housing

In Table 10 we replicate all the specifications of Table 6 on the subsample of households living in social housing. Random allocation in social housing should eliminate the estimation bias resulting from spatial sorting such that the variations at the block level are now exogenous to the parental characteristics.

We include the same controls as in the previous section, namely spatial fixed effects at the *département* level, dummies coding for parental occupation and educational level, as well as year fixed effects. Controlling for spatial fixed effects is all the more important in this section that the design of the social housing policy is implemented at the *département* level. Households living in social housing are free to choose the *département* where they want to live but are then randomly allocated within the chosen *département*. From an econometric point of view, the exogenous source of variations in name patterns that we exploit are within *département*, across-block variations. And spatial fixed effects capture all the unobserved heterogeneity in spatial sorting (for example clustering of new migrants around big cities).

Remarkably, the coefficients remain precisely estimated despite the fivefold reduction in sample size. Qualitatively and quantitatively the results are robust to this more demanding identification strategy. One important difference is the mean predicted probability which is now around 25%, signaling that social housing has an over-representation of immigrants of Arabic origin in general in France. Note however, that we are not comparing social housing with private housing here, but different blocks of social housing inside a region, in which households are randomly allocated by the administrative decision structure. The impact of the vertical channel is now even more impressive, adding 30 points on average to the propensity to give an arabic name. The horizontal channel seems also stronger, although less statistically significant. Concerning the unemployment penalty, the estimate of column (5) reveals that a 10% increase in the unemployment differential now translates into a 19 points fall in the naming probability.

7 Quantification and Counterfactual Experiments

7.1 The dynamic version of the model

In this section we are interested in assessing the long run implications of various counterfactual experiments based on our structural estimates. To this purpose we consider a simple dynamic extension of our structural model described in Section 2. Let us consider a block populated by a large number of agents $\mathcal{N}$. As a first pass, 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 2. 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

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

Table 10: The choice of an Arabic name : Social Housing

	(1)	(2)	(3)	(4)	(5)
Dep. Var:			arabic name for baby		
at least one parent same name type	0.32^{a}	0.32^{a}	0.31^{a}	0.31^{a}	0.31^{a}
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
share of name type in block (aged 1-10)	0.13^{b}	0.14^{b}	0.18^{b}	0.12^{b}	0.13^{b}
	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)
occupation unemp. penalty (Mincer-based)	-1.03^{b}	-0.97^{b}	-0.92^{b}	-0.91^{b}	-1.15^{a}
	(0.43)	(0.43)	(0.42)	(0.42)	(0.43)
share of name type (aged 1 to 10) in par. occ.		-1.11			
		(1.13)			
share of name type in dept (aged 1-10)		-0.54^{c}			
		(0.30)			
share of name type in block (aged 11-25)			0.03		
			(0.08)		
share of name type in block (aged 26-49)			0.04		
			(0.11)		
share of name type in block (aged 50+)			-0.13		
			(0.09)		
block unemp. penalty (Mincer-based)				-1.88^{c}	-1.19
				(1.01)	(1.01)
Observations	607	607	589	607	402097
R^2					
Pseudo R^2	0.475	0.480	0.485	0.479	0.528
meanprob	0.25	0.25	0.25	0.25	0.24

Standard errors in parentheses
^{c} p¡0.1, ^{b} p¡0.05, ^{a} p¡0.01

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

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

where $\tanh(x) \equiv (e^x - e^{-x})/(e^x + e^{-x})$ and ΔV_A with $A \in \{0,1\}$ is the observable utility differential retrieved from equation (5). In ΔV_A, whatever the type of expectations (i.e. rational or backward-looking), 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 where they are identified under the standard normalization assumption $\sigma = 1$. We thus have

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

where $\mathcal{C}$ corresponds to the employment penalty of Arabic name holders.

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

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

First it is worth to notice 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 our analysis on the steady-state only, abstracting from any consideration on the transition dynamics. Secondly, 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.[10] We consequently rely on numerical computations of (11) to characterize the set of solutions $\mu(\mathcal{C})$.

This analysis can be refined by increasing the number of individual and block characteristics we condition the probabilities of transmission $(\mathbb{P}_{0,t}, \mathbb{P}_{1,t})$ upon. This is important given that our benchmark specification (Column 5, Table 4) includes several covariates and fixed-effects for the purpose of filtering out various estimation biases. Under the hypothesis that there is no inter-block migration (or, in a less drastic way, that migration is a random process orthogonal to block characteristics; a pattern which is observed within the subpopulation of individuals living in social housing) we can generalize in a simple way our previous approach. The fixed point equation for a specific block k becomes

$$\mu_k = \frac{1}{2} + \int \tanh\left[\frac{1}{2}(\hat{\alpha}_0 + \hat{\alpha}_2\mu_k + \hat{\alpha}_3\mathcal{C} + \mathbf{X}'\hat{\beta} + \mathbf{Z}_\mathbf{k}'\hat{\gamma})\right] d\mathcal{F}_k(X), \tag{12}$$

[10]In absence of vertical transmission, i.e. $\hat{\alpha}_1 = 0$, our model would be included in the class of model analyzed in 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).

where $(\mathbf{X}, \mathbf{Z_k})$ denote individual and block characteristics, $(\hat{\beta}, \hat{\gamma})$ denote their respective point estimates, and $\mathcal{F}_k(.)$ is block k-specific distribution of individual characteristics. The binary variable coding for vertical transmission is implicitly included in $\mathbf{X}$ (and so is $\hat{\alpha}_1$ included in $\hat{\beta}$). For each block k we calibrate $\mathcal{F}_k(.)$ on the empirical distribution of observable characteristics retrieved from our sample over the 2003-2007 period.

The aggregate steady state share of Arabic name holders is then obtained by summing μ_k across all blocks weighted by block relative size

$$\mu = \sum \frac{\mathcal{N}_k}{\mathcal{N}} \times \mu_k, \tag{13}$$

7.2 Long run steady states and Counterfactual Experiments

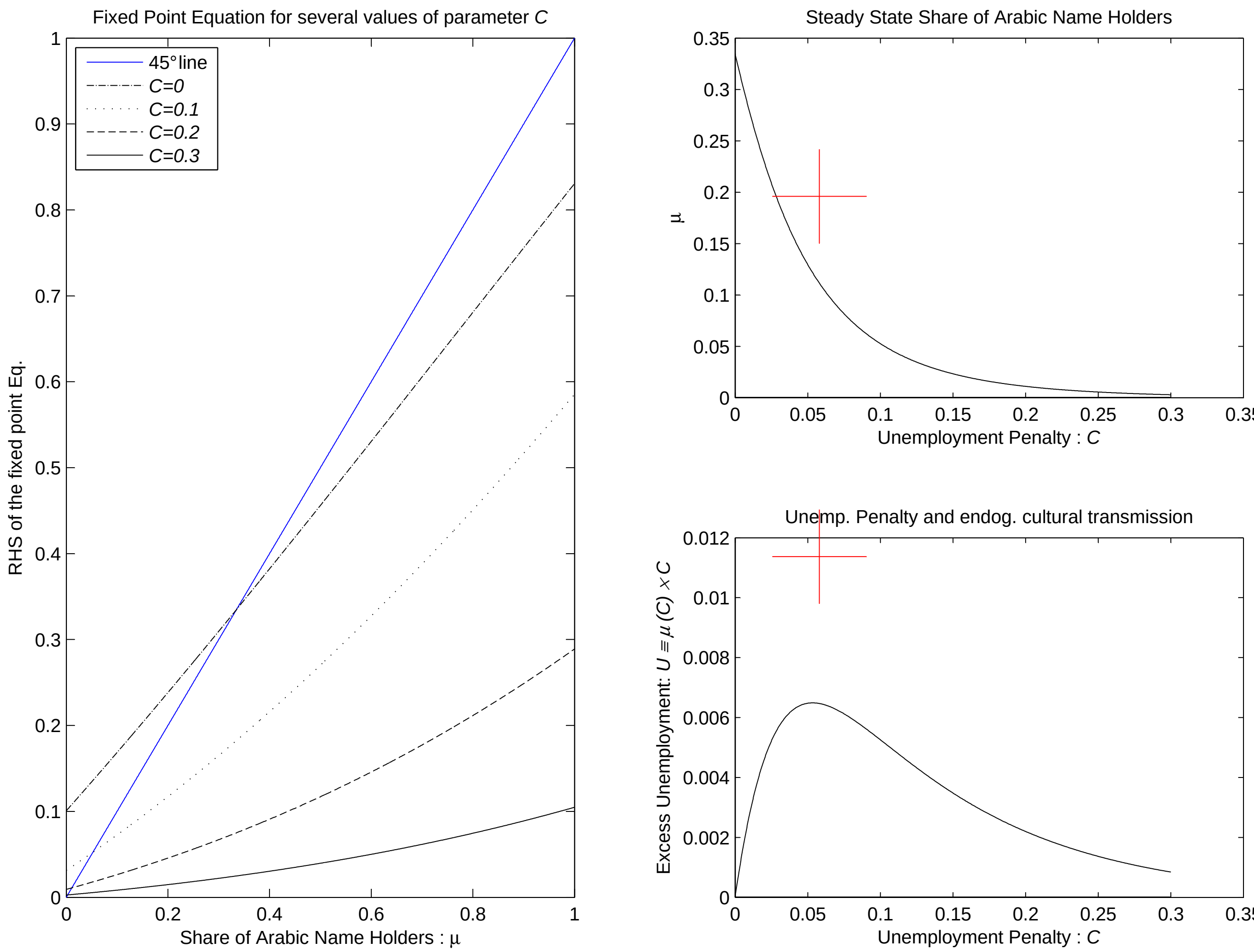
We start by solving numerically the previous model when name type is the only dimension of cross-individual observable heterogeneity. The parameters $(\hat{\alpha}_0, \hat{\alpha}_1, \hat{\alpha}_2, \hat{\alpha}_3)$ correspond to the point estimates of a specification similar to our benchmark where the sample is restricted to social housing (Column 5, Table **??**) except that we remove all the covariates and fixed-effects which are not directly related to our structural model (5). We solve numerically the fixed-point equation (11) for values of $\mathcal{C}$ spanning the range $[0, 0.3]$; 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 toward Arabic name holders

$$\mathcal{U}(\mathcal{C}) \equiv \mu(\mathcal{C}) \times \mathcal{C}, \tag{14}$$

The results are reported on figure 1 for $(\hat{\alpha}_0 = -2.19, \hat{\alpha}_1 = 3.06, \hat{\alpha}_2 = 0.72, \hat{\alpha}_3 = 12.45)$. The left panel depicts the fixed-point equation (11) for various values of $\mathcal{C}$ in the range $[0, 0.3]$. We can check visually that the equilibrium is unique. The top right panel reports the steady-state value of $\mu(\mathcal{C})$; the red cross represents the actual values of the unemployment penalty (equal to 0.058) and the actual share of Arabic name holders (equal to 0.196) observed in our sample of individuals living in social housing over the 2003-2007 period. We observe that for $\mathcal{C} = 0.058$ the steady-state share of Arabic name-holders predicted by our structural model is $\mu(0.058) = 0.124$ which is smaller than the actual one. This feature might be explained by the fact that the Maghrebian migration wave is a pretty 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. On the bottom right panel we report the predicted excess-unemployment $\mathcal{U}(\mathcal{C})$. We observe a non-linear relationship and this confirms 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).

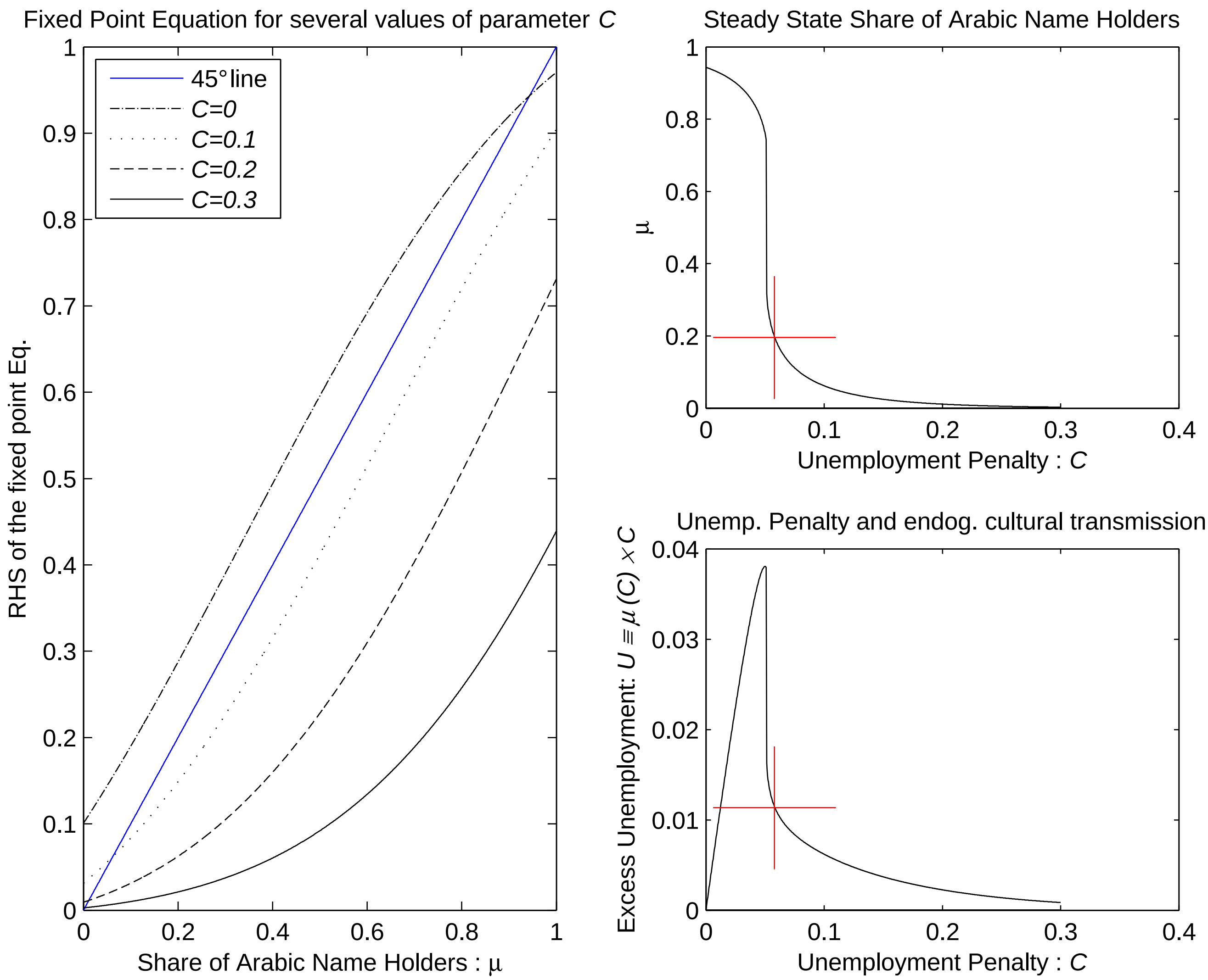
Our simulation shows the interesting result that the overall discrimination starts by rising when the estimated penalty (intensity of discrimination) decreases from a high initial level.

Figure 1: First simulation



In the previous analysis the coefficient of horizontal transmission, $\hat{\alpha}_2 = 0.72$ is not large enough to generate multiplicity of steady-state multiplicity. To illustrate what would happen if the horizontal channel was to be larger, we redo in figure 2 the same exercise after setting $\hat{\alpha}_2 = 2.62$ a value where the actual share of Arabic name holder is equal to its steady-state value. With such a counterfactual coefficient we see that multiplicity may show up for some intermediate range of $\mathcal{C}$. We also notice how extreme is the non-linearity of $\mathcal{U}(\mathcal{C})$. Note however that this horizontal transmission channel is probably much too high, predicting an implausibly large share of Arabic names in the population when the penalty $\mathcal{C}$ goes to 0.

Figure 2: Second simulation



8 Conclusion

This paper documents the effect of economic incentives on cultural transmission in the context of naming decisions. We estimate a structural model disentangling three channels for the transmission of first names: a vertical channel from parental culture, an horizontal channel from the neighborhood culture, and an economic channel associated to the labor market penalty from holding a first name at odds with the dominant culture. We estimate those channels on the French Labor Force Survey, taking advantage of the randomized allocation of parents in social housing to analyze the spatial interactions in naming decisions within blocks. We show that the economic incentives have a first order effect in the parental decision to transmit a first name that belongs to their own culture. Parents from a culturally discriminated minority would accept an expected one year income loss for their offsprings by transmitting their cultural trait. More generally, this paper takes a first step in providing a framework to analyze the effect of public policy on the transmission of cultural values. Based on our structural model, we estimate the interplay between cultural transmission and anti-discriminatory policies on the labor and housing markets. We show that those policies might have non-trivial effects when we take into consideration the endogeneous reaction of individuals in their transmission of cultural values.

9 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), 384-390.

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

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, 4-30.

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

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

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

Blume, L. Brock, W., Durlauf, S. and Ioannides, Y. (2010): Identification of Social Interactions. in Benhabib, Bisin and Jackson, *Handbook of Social Economics.*

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

Duguet, E., N. Landri, Y. L'Horty and P. Petit (2010): Are young French jobseekers of ethnic immigrant origin discriminated against? A controlled experiment in the Paris area *Annales dÉconomie et de Statistique*, pp. 99–100.

Fernandez, R. (2008): Culture and Economics *New Palgrave Dictionary of Economics*

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

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

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

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

Jouniaux, L. (2001): *Les 20,000 plus beaux prnoms du monde.* Hachette edts.

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

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

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

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

10 Appendix - Additional Robustness checks

This Appendix provides additional robustness analysis. All robustness checks are provided for both the full sample and for the social housing one. Results are reported in Table 11. In columns (1) and (2), we take into account measurement error in the first-stage estimation by weighting the point estimates of our discrimination measures retrieved from the auxiliary equation by the inverse of their standard deviation. This is a standard procedure which is to some extent a substitute for boostraping. The coefficients keep their statistical significance and this makes us confident that previous results are not driven by measurement errors on discrimination. However the magnitude of the coefficients is hardly interpretable given the weighting.

In columns (3) to (6), we test whether our effects depend on the gender of the babies by separating the samples between baby girls and boys. Gender does not affect the results except for the horizontal transmission channel which is slightly more significant for baby boys in the social housing case. In columns (7) and (8) we investigate whether the results are different for babies whose parents are citizens from a country belonging to Maghreb. Despite the drastic fall in the number of observations, results are striking. The vertical effect is very much the same, the horizontal is stronger, and the block penalty impact jumps to a very high level: Social housing parents faced with a 1 point larger difference in expected unemployment rate, lower their likelihood of giving an arabic name by 8%.

Table 11: The choice of an Arabic name - robustness

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Dep. Var:	arabic name for baby							
at least one parent same name type	0.15^a	0.32^a	0.18^a	0.20^a	0.34^a	0.43^a	0.28^a	0.35^a
	(0.01)	(0.02)	(0.02)	(0.01)	(0.03)	(0.04)	(0.05)	(0.10)
share of name type in block (aged 1-10)	0.08^a	0.13^b	0.08^a	0.10^a	0.17	0.15	0.18^b	0.37^b
	(0.02)	(0.06)	(0.03)	(0.03)	(0.13)	(0.11)	(0.09)	(0.17)
block unemp. penalty (Mincer-based, s.e. corrected)	-0.01	-0.06						
	(0.01)	(0.05)						
occupation unemp. penalty (Mincer-based, s.e. corrected)	-0.01^c	-0.04^c						
	(0.01)	(0.02)						
occupation unemp. penalty (Mincer-based)			-0.47^b	-0.19	-1.72^c	-1.49^b	-0.47	-1.92
			(0.21)	(0.17)	(0.94)	(0.71)	(0.61)	(1.19)
block unemp. penalty (Mincer-based)			-0.53	-0.63^c	4.02	-3.24^c	-2.82^c	-6.74^c
			(0.45)	(0.38)	(2.97)	(1.67)	(1.54)	(3.70)
Parents origin	All	All	All	All	All	All	Maghreb	Maghreb
Baby gender	All	All	girl	boy	girl	boy	All	All
Social housing	No	Yes	No	No	Yes	Yes	No	Yes
Observations	3322	607	1322	1327	233	225	459	178
Pseudo R^2	0.441	0.478	0.378	0.557	0.518	0.513	0.247	0.366
Mean pred. prob.	0.09	0.15	0.10	0.13	0.18	0.19	0.22	0.27

Standard errors in parentheses
[c] p¡0.1, [b] p¡0.05, [a] p¡0.01