Detection of local interactions from the spatial pattern of names in France

Keith Head, Thierry Mayer

Detection of local interactions from the spatial pattern of names in France*

Keith Head† Thierry Mayer‡

May 24, 2007

Abstract

Using data on name distributions in 95 French *départements* observed from 1946 to 2002, we investigate spatial and social mechanisms behind the transmission of parental preferences. Drawing inspiration from recent work on social interactions, we develop a simple discrete choice model that predicts a linear relationship between choices by agents in one location and the choices made in neighboring areas. We explain the shares of parents that give their children Saint, Arabic, and American-type names. In a second exercise we examine the effect of distance between locations on differences in name-type shares. In our last exercise we consider dissimilarity in actual names rather than name-types. Using Manhattan Distances as our metric, we find a steady and substantial decline in the importance of geographic distance. Meanwhile, differences in class and national origins have increasing explanatory power.

JEL classification: D190, F150, R100
Keywords: Social economics, Cultural transmission, Diffusion, Conformity, Geography

*We appreciated the very helpful comments of participants at the "International Conference on the Empirical Methods for the Study of Economic Agglomerations" at the University of Kyoto, July 1, 2006. We are particularly grateful to Pierre-Philippe Combes, Henry Overman, and Tony Smith for suggesting feasible solutions to problems they identified in the first version of this paper. We thank Anne-Célia Disdier for assembling some of the data used in this paper.

†Corresponding Author: Sauder School of Business, University of British Columbia, 2053 Main Mall, Vancouver, BC, V6T1Z2, Canada. Tel: (604)822-8492, Fax: (604)822-8477, Email: keith.head@ubc.ca

‡Université de Paris I, Panthéon-Sorbonne, Paris School of Economics; CEPII; and CEPR.

1 Introduction

Proximity enhances a wide range of interactions. We see this in the high rents and congestion costs that individuals and businesses endure to locate in large cities.[1] We also see it in the attenuation of trade associated with larger distances between trade partners.[2] Introspectively, we see it in the willingness of academics to incur substantial travel costs to attend face-to-face conferences. Some of those interactions manifest themselves through price mechanisms, while others—so-called non-market interactions—are purely social.

Non-market interactions, while difficult to measure, merit increased attention because they offer potential explanations for a variety of important social phenomena. Glaeser and Scheinkman (2002) point to "stock market crashes, religious differences, the great depression, wildly different crime rates," as well as diffusion of new technologies and "mass cultural phenomena like the Hula Hoop and Harry Potter." Evidence is accumulating that suggests non-market interactions may influence the volume of market transactions. To return to the case of bilateral trade flows, empirical research finds distance and border effects that seem too large to be explained by observed impediments. Grossman (1998) conducts a back-of-the envelope calculation and concludes that freight costs are too small and have the wrong functional form to explain the large distance effects estimated in gravity equations. He argues "something is missing from our trade models... I suspect... imperfect information where familiarity declines rapidly with distance. Perhaps it is a model with very localized tastes..." These ideas implicitly invoke non-market interactions. Head and Mayer (2000) find that countries in Europe trade much less with each other than would be expected given the absence of tariffs. As measured non-tariff barriers cannot explain the levels or the changes in the trade-impeding effects of national borders, the authors infer that cultural differences might account for the apparent bias towards home-produced goods.

Studies of non-market interactions face two problems. First, the interactions themselves are almost always unrecorded in publicly available data. Moreover, even the imputed outcomes of the interactions are rarely measured in a consistent way across time and space. A second problem is that market and non-market interactions often combine to generate outcomes. For example, Glaeser and Scheinkman (2001, 2002) develop empirical techniques to investigate the influence of non-market interactions on urbanization and female work participation. However, large literatures in economic geography and labour economics focus on price-based determinants of these decisions. Hence, it would be very hard to disentangle empirically the role of social and pecuniary interactions.

This paper investigates the locality of non-market interactions using data on the spatial distribution of given names in France. While parental choice of names reflects idiosyncratic tastes, we hypothesize that these tastes also have systematic components. We have data on counts of babies' given names in 95 different *départements*

[1] See Lucas (2001) for a calibrated model of productive externalities within cities.

[2] Disdier and Head (forthcoming) compile 1500 estimates of the distance effect on bilateral trade and find that, on average, a 10% increase in distance lowers exports by 9%.

(hereafter translated as departments) for the whole Post-War period. We use it to quantify the effect of geography on the degree of similarity in name choice between locations at a point in time. By repeating the analysis for each year, we can observe trends in the degree to which proximity matters.

The selection of baby names is a practice that offers two key advantages as a laboratory for studying social interactions. First, data on names do not suffer from the sample selection and measurement problems that plague many other social decisions (e.g. drug use, criminal activity, sexual practices). All parents are required to file birth certificates and there are strong incentives *not* to mis-state the selected name.

A second advantage of studying social interactions using names is there is little danger of a confounding influence of market interactions. This is because no agent has a profit incentive to influence name use since intellectual property law does not apply to personal names. As pointed out by Lieberson (2000), "Unlike many other cultural fashions, no commercial efforts are made to influence our naming choices." In contrast, the waistlines on blue jeans and the colors of cars are potentially influenced by the price and advertising decisions of the designers and manufacturers. The current popularity of the name Jacob in the US differs from the popularity of Apple iPods; nobody owns Jacob so no firm stands to gain monetarily from influencing Jacob's popularity. When studying names, we do not have to worry that observed differences in naming practices derive directly from variation in business strategies across time and locations, i.e. we have a relatively pure case of non-market interactions.

Researchers in a number of different social sciences have explored the causes and consequences of the selection of first names. Sociologists have devoted particular attention to the question whether parents of different socio-economic status (SES) choose different names. Lieberson and Bell (1992, Table 2) report that mothers with higher education levels select significantly different names from lower education mothers. They also find (Table 10) that the high-education mothers tend to be early-adopters of new names and the low-education mothers tend to be followers. French sociologists have also considered class differences in naming preferences and what they call vertical diffusion of tastes (when lower classes copy higher class choices). The vituperative exchange between Besnard (1995) and Lieberson (1995) regarding the timing and quality of these studies makes entertaining reading.

To the extent that names are signals of unobserved individual attributes, the selection of a child's name may have the delayed result of influencing market interactions. Bertrand and Mullainathan (2004) find that "Black" names on resumes lead to less favorable appraisals by potential employers. Figlio (2005) finds teachers are less likely to classify a student as gifted if she has more identifiably Black name than her sibling. Fryer and Levitt (2004) argue that the most satisfactory story accounting for the divergence between White and Black names in recent decades in the US is a desire by some Black parents to express Black cultural "identity."

The paper proceeds as follows. First, we present a model of name selection with social interactions to motivate the subsequent empirics. Second, we introduce our data set, establishing some of the basic patterns of naming in France. Then we con-

sider how liberalization of naming laws may have affected naming practices. Next we quantify the importance of geography in a series of regression specifications. In the first two exercises, we aggregate individual name frequencies according to name-types. This decision was motivated in part by the fact that there are thousands of names that are in use in France in any given year, but most of which are used for small numbers of children and are therefore "rare" (less than 3 per year) in many, if not all, departments. This means that name-level share regressions would have massive missing data problems. We therefore examine three name-types that exhibit generic cultural issues: the maintenance of national tradition (Saint names), the introduction of foreign traditions via immigration (Arabic names), and the globalization of tastes, perhaps facilitated by the media (American names). We start by asking whether a given name type is more popular if it is common in neighboring departments. We then use contiguity and distance to explain the absolute differences in name types between pairs of departments. Finally, we use the full detail of the name distribution to calculate metrics of name dissimilarity between pairs of departments. In each exercise, we find that proximity leads to greater similarity. However, many of our results indicate that geographic separation has become much less important than it was 40 years ago. In contrast, a department's class structure, the determinant of preferences emphasized by sociologists, appears to be rising in importance.

2 Theoretical Framework

During the last decade, a new literature on social interactions has emerged. For reviews see Durlauf and Young (2001) and Scheinkman (forthcoming). The models share a number of common features. Agents have utility with private and social components. The social component is a gain (or loss) from matching behavior to that of a set of interacting agents. The agents form expectations on how others will act and then the model is solved for self-consistent equilibria. Although the models typically do not specify the underlying benefits of conformism, Young (2001) suggests three main sources: (a) direct desire to imitate, (b) coordination and conventions, and (c) learning successful practices from peers.

The models favored by Durlauf and co-authors draw inspiration from physics and treat social phenomena as an interacting particle system. These models are non-linear due to parametric assumptions made on the form of social interactions and individual heterogeneity. As commonly assumed in discrete choice theory, the functional form chosen by those economists for individual variations in preferences yields the logit model. The implicit solutions involving the hyperbolic tangent function are identical to those used in the Curie-Weiss model of magnetism.[3] Here we use a different set of parametric assumptions to achieve a tractable linear model of social interactions.

The Brock and Durlauf social physics approach can be thought of as a good model for interactions between peers, that is members of the same generation.

[3]See Brock and Durlauf (2001) for a complete treatment of the dichotomous choice model.

Cavalli-Sforza and Feldman (1981) adopted the epidemiological term of "horizontal transmission" to refer to peer-to-peer effects. A second strand in the theoretical literature emphasizes intergenerational transmission. Cavalli-Sforza and Feldman call the tendency of children to retain the traits of their parents vertical transmission.[4] Their model of social inheritance draws on ideas from population genetics.

In the social physics approach, agent heterogeneity is taken as exogenous. When it is small enough relative to positive interaction effects, there are multiple equilibria in which almost all members of the population take the same action. Bisin and Verdier (2001) explain how heterogeneity in choices within populations can be sustained endogenously. They do so by modeling the decision of parents to exert effort to socialize their children to replicate their traits (politics, religion). A key idea is that when the parents want to keep their children from adopting locally prevalent traits, they work harder at home to induce loyalty to parent traits. This offsets the tendency towards peer conformity. Thus models of vertical intergenerational interactions yield persistent heterogeneity *within* areas.

We incorporate the possibility of vertical transmission in a very crude way, so as to keep the model as simple as possible. Without modeling the dynamic process by which inter-group heterogeneity arose, we just assume that different groups have different means in their taste distributions. The groups will be operationalized in the empirical section based on national origins and socio-professional categories.

Households (denoted h) in each location (ℓ) choose between K names or name types (e.g. "traditional" or "modern"). The share of households in location ℓ selecting type k is given by $s_{\ell k}$. Households in each of the L possible locations are heterogeneous in two respects. First, they have their own idiosyncratic preferences over name types. Second, they are members of larger groups with different mean preferences. We assume G groups denoted with subscript g. The shares of each group are given by $x_{\ell g}$ with $\sum_{g=1}^{G} x_{\ell g} = 1$. In addition to exogenous group preferences, households care about the choices they expect from other members of the local community, ℓ, and of neighboring communities, n. Households put a weight $\nu_{\ell n}$ on neighbor regions, where $\nu_{\ell n}$ is assumed to vary between zero (localized interactions) and $1/(L-1)$ (global interactions). The upper bound decreases with the number of locations to prevent the aggregate amount of interactions from increasing too much as a result of finer spatial partitions.

For tractability, we maintain the assumption that all households *within each location* interact homogeneously with each other, that is, we rule out finer spatial structure and group-based interactions. Denoting the portion of each group selecting type k as $s_{\ell gk}$, the aggregate share for a location is $s_{\ell k} = \sum_{g=1}^{G} s_{\ell gk} x_{\ell gk}$. The utility that household h of group g living in ℓ experiences from choosing a k name, $U_{\ell ghk}$, comprises a social and a private component:

$$U_{\ell ghk} = \overbrace{\lambda(s^e_{\ell k} + \sum_{n \neq \ell} \nu_{\ell n} s^e_{nk})}^{\text{social}} + \underbrace{\zeta_{ghk}}_{\text{private}} \tag{1}$$

[4]The biologists' use of vertical to refer to intergenerational transmission should not be confused with the sociologists' use of vertical to refer to transmission from higher to lower social classes.

The variables s^e_ℓ and s^e_n represent the expected shares of parents that choose type k names in the "local" and "neighbor" areas. The marginal utility from choosing type k as it becomes more popular is given by λ. The heterogeneity in private preferences regarding type k names is embodied in ζ. We include the g subscript in ζ because the model allows the mean preferences to vary by group. With a large population, the share choosing name k is given by the probability that name k yields utility higher than any other name:

$$s_{\ell gk} = \text{Prob}(U_{\ell ghk} > U_{\ell ghj}, \forall j \neq k). \tag{2}$$

A closed-form solution for equation (2) only exists under specific assumptions about the distribution of ζ. The best known case is the Type-I extreme value, which gives rise to the multinomial logit (MNL) form for this probability. Brock and Durlauf (2002) investigate the properties of the MNL model of social interactions. The problem with the MNL model is that its non-linearity makes it impossible to obtain analytical closed-form solutions for the self-consistent equilibrium, where $s_{\ell gk} = s^e_{\ell gk}$, except in special and simple cases.

We therefore focus on the case where $K = 2$, and on a uniform distribution of households' heterogeneity. The purpose of dichotomizing names (which we will do in three different ways) is both to allow for the development of a tractable estimating equation and to focus on aspects of names that we believe parents care about. The increase in utility due to choosing name-type 1 rather than type 2 is given by

$$V_{\ell gh} \equiv U_{\ell gh1} - U_{\ell gh2} = \lambda \left[(s^e_{\ell 1} - s^e_{\ell 2}) + \sum_{n \neq \ell} \nu_{\ell n} (s^e_{n1} - s^e_{n2}) \right] + \zeta_{gh1} - \zeta_{gh2}. \tag{3}$$

Defining $\epsilon_{gh} \equiv \zeta_{gh1} - \zeta_{gh2}$, and recognizing that with two name types we can drop subscripts and let $s_1 = s$ and $s_2 = 1 - s$, we re-express the above equation as

$$V_{\ell gh} = 2\lambda \left[(s^e_\ell - 1/2) + \sum_{n \neq \ell} \nu_{\ell n} (s^e_n - 1/2) \right] + \epsilon_{gh}. \tag{4}$$

The social contribution to household decisions is a combination of the expected frequency of the name-type and the scope of spatial interactions. In a symmetric equilibrium where $s^e_\ell = s^e_n = 1/2$, the social utility term is nullified and the choice depends only on private preferences.

The heterogeneity in private preferences regarding type-1 names is embodied in ϵ which we assume is a symmetric, uniformly distributed variable centered at β_g with upper bound $\beta_g + \alpha$ and lower bound $\beta_g - \alpha$. The probability that ℓ-residing household h from group g chooses a type-1 name is given by the probability that $V_{\ell gh} > 0$. With a large population, that probability equals the actual share, $s_{\ell g}$. Thus, we have

$$s_{\ell g} = \frac{\alpha + \beta_g + 2\lambda \left[(s^e_\ell - 1/2) + \sum_{n \neq \ell} \nu_{\ell n} (s^e_n - 1/2) \right]}{2\alpha}. \tag{5}$$

A self-consistent equilibria equalizes actual and expected shares: $s_\ell = \sum_{g=1}^{G} s_{\ell g} x_{\ell g} = s_\ell^e$ for all locations ℓ. The solution that arises is stable (assuming myopic dynamics) if social interactions are sufficiently small relative to individual heterogeneity. Specifically, for $\lambda(1 + \sum_{n \neq \ell} \nu_{\ell n}) < \alpha$ the unique interior solution is given by

$$s_\ell = \frac{\alpha + \sum_g \beta_g x_{\ell g} + 2\lambda \left[\sum_{n \neq \ell} \nu_{\ell n} (s_n - 1/2) - 1/2 \right]}{2(\alpha - \lambda)}. \tag{6}$$

For $\lambda(1 + \sum_{n \neq \ell} \nu_{\ell n}) > \alpha$ the equilibrium shown in (6) would be unstable. That is, for expectations given by lagged actual shares, a perturbation away from the equilibrium would send the system off in the direction of the perturbation. The stable equilibria in this case are $s_\ell = 0$ and $s_\ell = 1$. The corner solution values for the multiple equilibria are a consequence of assuming a uniform distribution for heterogeneity. One way to eliminate them would be to follow Brock and Durlauf (2001) in assuming logistic heterogeneity. The cost of that approach is the loss of the linear form for the solution—which renders analysis more difficult.[5]

Here we will focus on the case where social influences are small enough relative to heterogeneity that there is a unique interior equilibrium. Equation (6) is the basis for our first empirical exercise. To facilitate estimation, we assume in this specification that only contiguous locations interact and use $\mathcal{C}(\ell)$ to denote the set of locations that border on ℓ. Each location ℓ has $N(\ell)$ neighbors. The strength of interactions are specified as $\nu_{\ell n} = \nu / N(\ell)$ for $n \in \mathcal{C}(\ell)$ and 0 otherwise.[6] The neighbor share for each location ℓ will be defined as the average for all contiguous locations: $s_\ell^C = \left(\sum_{j \in \mathcal{C}(\ell)} s_j \right) / N(\ell)$. These assumptions imply that $\sum_{n \neq \ell} \nu_{\ell n} (s_n^e - 1/2) = \nu (s_\ell^C - 1/2)$.

Because the group composition shares add to one in each location, one group must be omitted from the regression, implying that the coefficients on the included groups should actually be interpreted as differences with respect to the excluded group ($g = 1$). We add an error term, e_ℓ, to incorporate unmeasured compositional differences as well as other deviations between model predictions and the data. Combining these assumptions, we can re-express equation (6) as

$$s_\ell = a + \sum_{g=2}^{G} b_g x_{\ell g} + c s_\ell^C + e_\ell. \tag{7}$$

The coefficients in this linear regression are related to the underlying parameters as follows:

$$a = \frac{\alpha - \lambda(1 + \nu) + \beta_1}{2(\alpha - \lambda)}, \qquad b_g = \frac{\beta_g - \beta_1}{2(\alpha - \lambda)}, \qquad \text{and} \qquad c = \frac{\lambda \nu}{\alpha - \lambda}.$$

Equation (7) relates closely to the linear-in-means model that is often used to estimate "neighborhood" and "peer" effects on continuous variables, such as test

[5] Another, less elegant fix, would be to add two masses of parents that have such strong individual preferences for and against type-1 names that they ignore social influences.

[6] Alternatively, one could specify $\nu_{\ell n}$ as a function of distance between locations ℓ and n. Estimating the parameters of that function would require non-linear methods, which we want to avoid here.

scores.[7] Manski (1993) points out that OLS estimates of this type of equation are biased for several reasons. First, the neighbor name-type shares in a given location $n \in \mathcal{C}(\ell)$ depend in part on s_ℓ. That is, there is a simultaneity issue which Manski labels the "reflection" problem. Fortunately, our theory also suggests a set of instrumental variables. The average class and origins composition variables in contiguous departments, $x^C_{\ell g} = \left(\sum_{n \in \mathcal{C}(\ell)} x_{gn}\right) / N(\ell)$ are assumed to be independent of name choice and only affect s_ℓ through the channel of affecting s^C_ℓ. The exogeneity and excludability of the $x^C_{\ell g}$ imply that we should be able to estimate equation (7) consistently via two-stage least squares (2SLS). Glaeser and Scheinkman (2001, p. 85) point out that 2SLS may also treat biases in OLS arising from omitted variables common to locations ℓ and $n \in \mathcal{C}(\ell)$ that determine both s_ℓ and s^C_ℓ.[8]

Our theory also generates predictions for *dissimilarity* in name-type shares for any pair of locations: $\mid s_\ell - s_n \mid$. By incorporating the s_n in the dependent variable, we avoid the econometric issues created by the reflection problem. Additionally, we can use this approach to measure the strength of interactions between non-contiguous locations.

Using equation (6) and the corresponding equation for s_n, and assuming symmetry in bilateral interactions ($\nu_{\ell n} = \nu_{n\ell}$), we can solve for $s_\ell - s_n$ in terms of $x_{\ell g}$ and x_{ng}. Adopting vector notation, $\sum_{g=1}^{G} \beta_g x_{\ell g} = \boldsymbol{\beta} \cdot \mathbf{x}_\ell$, we difference the reduced forms for s_ℓ and s_n and obtain the following expression for the absolute difference in name-type shares:

$$\mid s_\ell - s_n \mid = \frac{\mid \boldsymbol{\beta} \cdot (\mathbf{x}_\ell - \mathbf{x}_n) + 2\lambda(F_\ell - F_n) \mid}{2[\alpha - \lambda(1 - \nu_{\ell n})]}, \tag{8}$$

where $F_\ell = \sum_{i \neq \ell, n} \nu_{\ell i}(s_i - 1/2)$ and $F_n = \sum_{i \neq \ell, n} \nu_{ni}(s_i - 1/2)$. These terms represent the influence of third locations on the choices of ℓ and n. The denominator of (8) is positive since stability requires $\alpha > \lambda(1 + \sum_{i \neq \ell} \nu_{\ell i}) > \lambda(1 - \nu_{\ell n})$.

Direct estimation of equation (8) would be very difficult because of the non-linearities involved. However, three key implications of (8) can be implemented empirically:

1. Similar group composition promotes similar naming patterns. The two-group case may help build intuition. With $G = 2$, $\mid \boldsymbol{\beta} \cdot (\mathbf{x}_\ell - \mathbf{x}_n) \mid$ reduces to $\mid \beta_2 - \beta_1 \mid \cdot \mid x_{\ell 2} - x_{n2} \mid$. Differences in shares of each name type should be large if group shares are very different or groups differ substantially in their mean preferences.

2. Proximity promotes similar naming patterns. Since the denominator increases in $\nu_{\ell n}$, the spatial extent of social interactions decreases dissimilarity between locations. Specifying $\nu_{\ell n}$ as a function of geographic proximity, and holding

[7] See Durlauf (2004, p. 2205) for a comprehensive review of this literature.

[8] Glaeser and Scheinkman also discuss a third source of bias in estimating equation (7): sorting. If individuals who like the same name-types endogenously choose to live near each other then the composition variables x can no longer be treated as exogenous. We think that feedback from naming preferences to the class and ethnic structure of a department is not likely to be very strong.

the difference in the composition of each location constant, the "distance" between name-type frequencies, $\mid s_\ell - s_n \mid$ should be increasing in geographic distance.

3. Third locations' naming patterns influence bilateral similarity. Consequently, estimation should attempt to neutralize these effects via the use of fixed effects.

Motivated by equation (8), we propose a linear estimation approach that embeds those three points:

$$\text{MDN}_{\ell n} = \theta \text{MDC}_{\ell n} + \rho \text{MDO}_{\ell n} + \gamma I_{n \in \mathcal{C}(\ell)} + \tau \ln D_{\ell n} + \delta_\ell + \delta_n + u_{\ell n}. \qquad (9)$$

The regression is estimated over the $L(L-1)/2$ set of distinct location pairings. The $\text{MD}_{\ell n}$ variables are metrics of the dissimilarity between two locations. In each case we use the so-called Manhattan Distance which sums over the absolute differences in shares. Thus, $\text{MDN}_{\ell n} = \mid s_{\ell 1} - s_{n1} \mid + \mid s_{\ell 2} - s_{n2} \mid$, where s_1 and s_2 are shares of type-1 and type-2 names. Social class dissimilarity, $\text{MDC}_{\ell n}$, and national origin dissimilarity, $\text{MDO}_{\ell n}$, are defined analogously using breakdowns from census data on occupational categories and citizenship.

The next two covariates examine the role of geography in determining similarity. We posit that $\nu_{\ell n}$, the factor determining the extent of interactions between locations, is decreasing in geographic distance $D_{\ell n}$ between locations. We expect $\hat{\tau} > 0$ because equation (8) shows that dissimilarity increases as interaction intensity declines. The specification includes a contiguity indicator, $I_{n \in \mathcal{C}(\ell)}$, to allow us to test whether the previous specification was justified in imposing a discontinuous elimination of interactions for non-contiguous locations. This extreme assumption would be supported by a positive estimate for γ and something near zero for τ.

Finally, regression specification (9) includes an error term $u_{\ell n}$ and a set of intercepts, δ_ℓ and δ_n, for each location. The fixed effects are designed to capture third-location effects on the pairwise differences.[9]

3 Names, name types, and naming regulations

The data we use in this paper come from the French statistical agency, INSEE. The data set, called the *Fichier des Prénoms*, is based on filings of birth certificates at the Civil Registry. Counts of births by name, sex, and department are available for all babies born in France from 1946 to 2002. Specific name counts are provided for all names given to at least three babies for a given department, sex, and year. The count of names given to just one or two babies are summed and coded as "rare."[10]

The number of distinct names reported in each department therefore depends on the total number of births. For example, the largest department, Nord (pop. 2.6m), reports 1439 names in 2002, and codes 11% of births as Rare. The smallest

[9] We thank Pierre-Philippe Combes for alerting us to this benefit of location fixed effects.

[10] There is also a national database containing name counts back to 1900. This data withholds particular names only if they were used less than two or fewer times in all France.

department, Lozère (pop. 0.08m), reports just 48 names and codes 62% of the births that year as rare. Using the non-rare names, we allocate the individual names into three name types: Saint names, Arabic names, and American names. Implicitly, we have to assume that rare names are allocated across types according to the same proportions as non-rare names.

The Saint name type draws on the history of French regulation of names. Legislation enacted in 1803 instructed civil registrars to permit only a narrowly defined set of names. The acceptable set included names in French calendars, names from ancient Greece and Rome, and names from the Bible. In practice this meant that most children were given names of Saints using French spellings. We therefore consider Saint names to be the traditional type of names. The list of Saint names from the French calendar was constructed using the website `http://nominis.cef.fr/`. We define a name as being Saint if it belongs to this list and non-Saint otherwise.

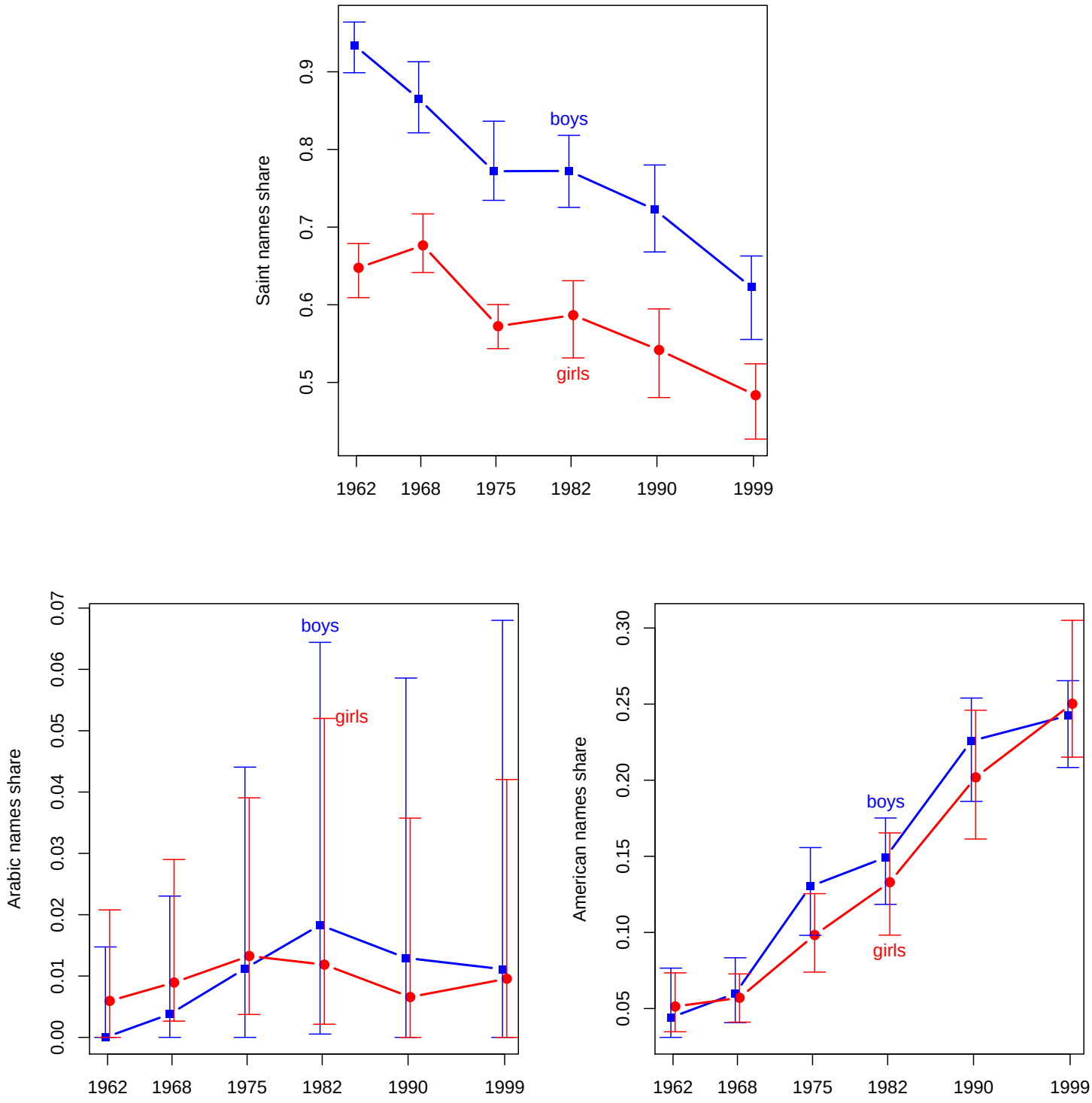
The birth registrars had some discretion to allow regional and foreign names as well as some spelling variations. In 1966 a ministerial directive called for increased permissiveness. The officials retained the right to make the initial decision, which the parents could then challenge in court. Legislation in 1993 dramatically shifted the rules. Now parents can choose any name and register it at birth. The civil officials can challenge names deemed to be contrary to the interest of the child in the courts.

We consider two sets of *non-traditional* names. The first is closely related to immigration. Large numbers of immigrants from the once colonized Maghreb (Algeria, Morocco and Tunisia) arrived in France in the 1960s, bringing with them a traditional set of Arabic names. Jouniaux (2001) provides a listing of names deemed to have Arabic origins. Although not driven by large-scale immigration, we also study the increase in French usage of names that are seen as typically American. Disdier, Head, and Mayer (2006) find evidence that one channel for the introduction of American names has been exposure to these names via the mass media of popular songs, TV shows, and movies. Here we are interested in whether there is evidence that American names are also being dispersed via spatial interactions between French parents.

The definition of "American" names is highly problematic. Most of the names associated with Americans (John, Robert, George) were brought by English colonizers whose ancestors were strongly influenced by French names and by common sources (e.g. the Bible). Hence, we define American names in this paper based on patterns of contemporary *usage* rather than etymology. The Social Security Administration tracks given names in the US and makes them available on its website, `www.ssa.gov/OACT/babynames/`. This site gives the top 1000 names by sex/decade back to 1900. To obtain the frequency of US names we need to divide by number of births by sex by year (or decade). Total births are available from the *Statistical Abstract of the United States* and this source also shows that the share of boys is 0.512. We define a name as being "American" when its share of total births over the whole 1900–2002 period was higher in the United States than in France.

Figure 1 graphs the spatial and temporal variation of the popularity of the three name types for boys and girls. We show medians and interdecile ranges (10%–90%)

Figure 1: Medians and interdecile ranges of name-types across French Departments in census years



across departments in census years (the sample used in the subsequent regressions). Several distinct patterns emerge. The upper panel shows the declining frequency with which French parents give names from the calendars of Saints. For boys this is a case where intertemporal variation swamps geographic variation since the interdecile ranges (IDRs) in the 1960s do not overlap with IDRs in the 1990s. The Saint case also illustrates differential popularity of name types for boy and girl babies.

The lower panels show the popularity of what we call Arabic and American names. The Arabic names are unique in that time-series variation and sex-differences are quite small compared to variation across departments. The regression estimates will confirm many readers' predictions that this arises mainly because of large variation in the share of immigrants from Arabic-speaking countries. With American names, the sex-differences remain minor but there is a strong trend up that leaves a large gap between the highest levels of popularity in the early years and the lowest levels of popularity in recent years.

Across all three name-types and both sexes, interdecile ranges have grown between 1962 and 1999. Checking other spread measures (5%–95%, 15%–85%, 25%–75%) we find roughly the same pattern. The interquantile ranges trend up for all female name-types and most male types. We see no instances of declining spread. The growing dispersion in name-type shares is surprising in light of our model of positive social interactions. We had expected that as barriers to long distance interactions have fallen, there would be great similarity in outcomes. This suggests there is something else at work and motivates the need for regression analysis.

While Figure 1 can show the central tendency and dispersion of name types, we need maps to visualize spatial patterns in the name-type frequencies. Figures 2 and 3 illustrate the frequency of Saint names for girls and boys in 1972 and 2002. The figures show that name-types appear to be spatially correlated, with some major changes over the last three decades. In the area centered around Paris, Saint names are waning. The same appears be happening in the Southeast near the border with Italy. While the maps give some *prima facie* evidence for spatial dependence, we will need regression evidence to establish magnitudes and statistical significance.

4 Name-type regressions

The first set of regressions is an empirical implementation of equation (7), applied to the three name types detailed in the preceding section: Saint names, Arabic names and American names.

Each of those dependent variables will be explained by the socio-economic class structure (broken into six categories, with "intermediaries" as the omitted group), combined with the citizenship structure of inhabitants of the department (broken into five categories, with French nationals as the omitted category.) The source of this information is the French census. The French statistical data agency INSEE conducted population censuses in the years 1962 (5% sample), 1968 (25%), 1975 (25%), 1982 (25%), 1990 (25%), and 1999 (5%). We used individual level data to construct departmental measures of (a) the share of the active population of ages

Figure 2: Prevalence of Saint names for girls in 1972 and 2002

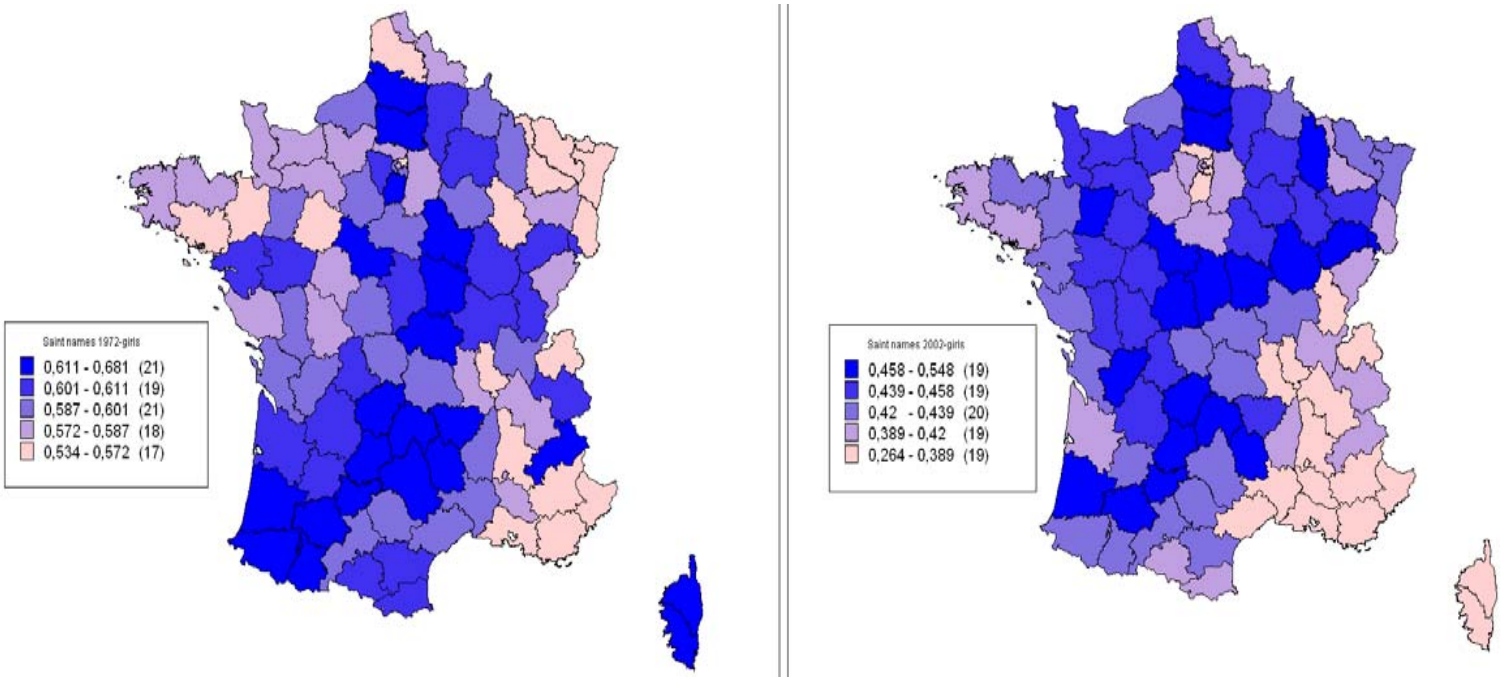
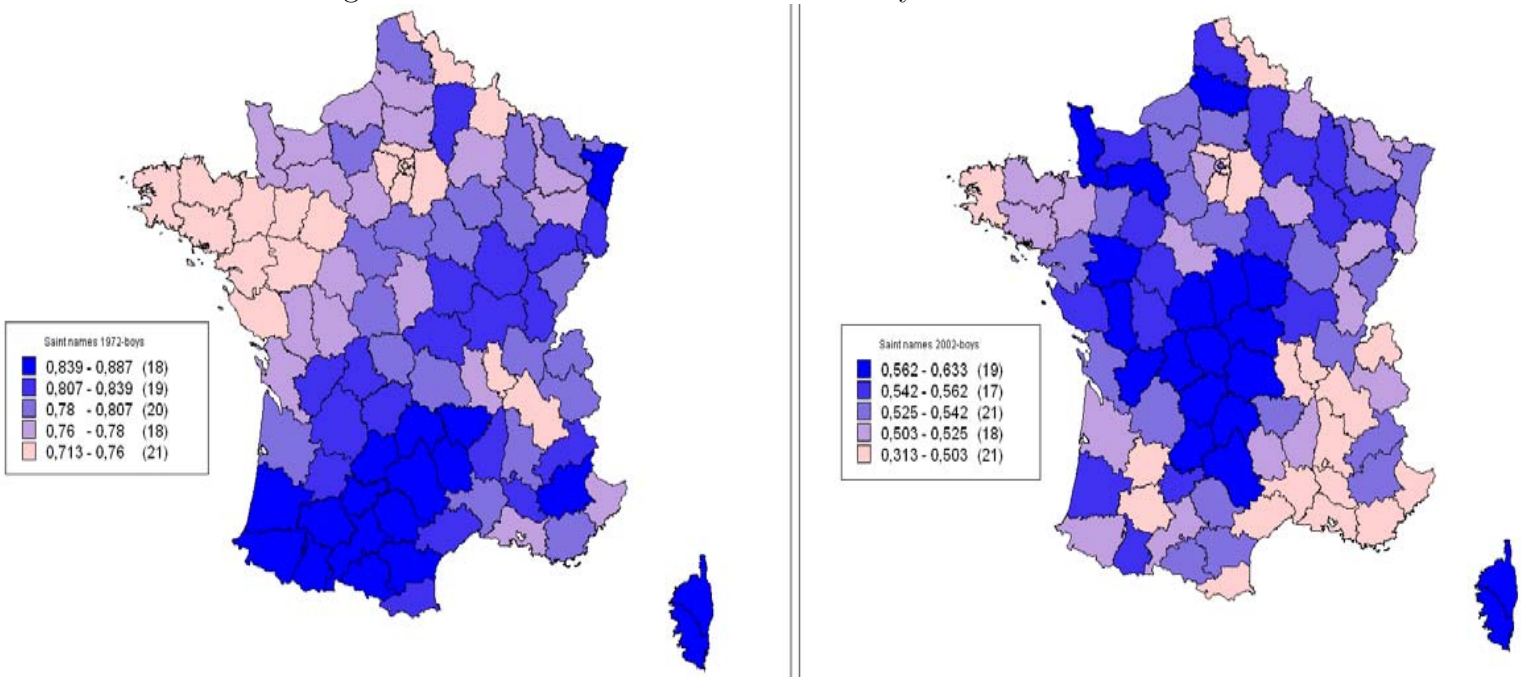


Figure 3: Prevalence of Saint names for boys in 1972 and 2002



20-44 in six socio-professional categories (farmers, business owners, professionals and managers, intermediate occupations, clerical workers, and manual workers); (b) the share of total population with birth nationalities France (includes Algerians prior to independence), Maghreb (Algeria after independence, Tunisia, Morocco), Sub-saharan Africa, the United States, and other birth nationalities.

Spatial interactions between members of the same generation are captured with the percentage of babies born with same name type in contiguous departments. As mentioned in the theoretical section, the simultaneity issue raised by this variable calls for two-stage least squares (2SLS), with the class and origin composition variables in contiguous departments being the natural instrumental variables. In all regressions we pool data over both sexes. To capture differences in the mean popularity of a name-type for boys and girls, we include a dummy for the observations corresponding to male shares. Finally we include an indicator for Corsica, which is the only island in our sample (Martinique, Guadeloupe and Réunion have missing census data for the class and origin variables). Insularity is expected to have an influence in the naming patterns, and particularly in spatial interactions, since it renders social interactions more difficult.[11]

Table 1 provides results for regressions explaining the "saintliness" of babies' names in different French departments, measured as the percentage of babies born with Saint names in the department. The first column shows pooled results over the six census years; annual results are shown in the following columns. The upper part of the table shows the 2SLS coefficients. The lower frame shows regression diagnostics, including the F-test for weak instruments and the Sargan test for instrument validity. For comparison with the IV coefficients, we also show the OLS coefficients for contiguous shares of Saint names.

Starting with class composition results we find very little in the way of stable relationships between the shares of the parent generation in each category and the share of babies given Saint names. Different classes appear to have different relative preferences over traditional names over time. For example, while farmers and manual workers seem to have a particular tendency to give Saint names to their children in 1999, the opposite was true in 1968. A much more stable composition effect is found for the share of the population with a Maghreb citizenship. The influence on traditional names is strongly negative and remains so over time.

The results that are the focus of our paper are the spatial interactions. The pooled coefficient of 0.34 is considerably smaller than the magnitudes prevailing before 1982. These results suggest a sharp decline of spatial interaction over time. During the sixties, the coefficient averaged about one, indicating that an increase in saintliness of names in the neighbors would be matched proportionately in the local naming patterns. At the end of the nineties, the coefficient is less than 0.2 and is not even statistically different from zero. Note that this trend in decreasing spatial interactions is also apparent in the OLS coefficients, which remain significant and are usually large in magnitude. We find evidence in the first and columns (1), (5), and (7) of the upward bias expected due to the reflection problem. This pattern is repeated in all but one of the 14 specifications for Arabic and American names.

[11] Omission of this dummy leads to big changes in the coefficient on contiguous name shares.

Table 1: Explaining shares of Saint names in each Department

	Dependent Variable: Saint share						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	0.50^a	-0.12	0.61^a	0.51^a	0.64^a	0.38^b	0.21
	(0.13)	(0.26)	(0.20)	(0.17)	(0.19)	(0.16)	(0.14)
male	0.13^a	-0.06	0.04	0.05^c	0.14^a	0.15^a	0.11^a
	(0.02)	(0.09)	(0.03)	(0.03)	(0.03)	(0.02)	(0.02)
Corsica	0.30^a	1.02^a	0.69^a	0.61^a	0.22^b	0.17^b	0.16^b
	(0.08)	(0.26)	(0.12)	(0.09)	(0.10)	(0.08)	(0.07)
% Farmers	-0.04	-0.06	-0.48^a	-0.35^b	-0.05	0.57^a	0.81^a
	(0.13)	(0.16)	(0.16)	(0.16)	(0.20)	(0.19)	(0.17)
% Craft	0.13	0.07	-0.43^b	-0.20	0.03	-0.04	-0.20
	(0.18)	(0.22)	(0.18)	(0.22)	(0.24)	(0.28)	(0.32)
% Superior	-0.23	-0.17	-1.34^a	-0.76^c	-0.47	0.26	0.40
	(0.27)	(0.50)	(0.47)	(0.41)	(0.49)	(0.40)	(0.32)
% Clerks	-0.10	0.07	-0.36^b	-0.48^a	-0.21	0.00	0.20
	(0.17)	(0.18)	(0.18)	(0.18)	(0.21)	(0.21)	(0.18)
% Manual	-0.06	-0.05	-0.57^a	-0.43^a	-0.28	0.16	0.31^b
	(0.14)	(0.17)	(0.17)	(0.16)	(0.19)	(0.19)	(0.16)
% African	0.12	-4.12	-9.98	-0.25	0.58	0.86	0.48
	(0.86)	(4.57)	(6.69)	(3.15)	(2.24)	(1.35)	(0.70)
% Maghreb	-1.44^a	-2.85^c	-1.17^a	-0.75^a	-1.02^a	-1.63^a	-1.15^a
	(0.18)	(1.72)	(0.36)	(0.24)	(0.23)	(0.28)	(0.21)
% US-born	-0.16	-0.43	9.40	3.08	-17.34	-16.85	-6.86
	(0.41)	(0.54)	(10.67)	(8.09)	(13.59)	(10.56)	(10.28)
% Other nat.	0.15^b	-0.10	0.09	0.19^c	0.00	0.18	0.10
	(0.06)	(0.07)	(0.08)	(0.10)	(0.10)	(0.11)	(0.10)
Contig % Saint (IV)	0.34^a	1.23^a	0.79^a	0.75^a	0.26^c	0.17	0.19
	(0.12)	(0.33)	(0.15)	(0.14)	(0.15)	(0.13)	(0.13)
observations	1080	180	180	180	180	180	180
R^2	0.953	0.97	0.962	0.958	0.943	0.947	0.918
RMSE	.03	.026	.02	.023	.025	.025	.024
Contig % Saint (OLS)	0.75^a	0.58^a	0.72^a	0.48^a	0.38^a	0.17^c	0.21^b
	(0.12)	(0.08)	(0.07)	(0.08)	(0.09)	(0.09)	(0.09)
F stat 1st stage	20.19*	2.54	9.07	31.87*	31.81*	26.43*	42.87*
Sargan stat	66.81^a	8.61	13.06	15.2^c	20.76^a	12.1	32.44^a

Note: Robust standard errors in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. * denotes significance at the 5% level for the weak IV bias test of Stock and Yogo (2002, table 1). Errors allow for clustering by department in the pooled regression, which also includes year fixed effects.

We regard the conflicting results as likely to be caused by the high standard errors inherent to IV estimation. With the exception of the first two years, we can see that the F statistics for our instruments are mainly much larger than the 11.5 critical value shown in Table 1 of Stock and Yogo (2002). This indicates that the contiguous composition variables are strong enough instruments to hold the relative bias of 2SLS below 10%, 95% of the time. The Sargan statistics raise concerns on the validity of our instruments for certain census years. The problem reoccurs for Arabic and American names, so we defer discussion to the end of this section.

Table 2 estimates the same regression but changes the dependent variable to be the share of Arabic names. As before, the effects of class composition are mixed across periods. Origin composition are the key determinants. The coefficient on the share of Maghreb-born in the department is 1.15 in the pooled regression, with a very clear decreasing trend over time. An interpretation of this trend is that immigrants from Arabic countries are the main group giving Arabic names in France, but that their attachment to origin-country traditional names is fading over time.

There is no persuasive evidence of positive spatial interactions for this cultural trait. An increase in the neighbor department share in Arabic names mainly seems to reduce the prevalence of such names locally. When significant (as it is in the pooled regression and two thirds of the census years), the 2SLS coefficient is around -0.25. Again, the difference between the OLS and 2SLS regressions is striking and goes exactly in the expected direction: the OLS coefficient on spatial interactions is overestimated by quite a wide margin, basically reversing the sign of the (statistically significant) relationship. Note the extremely high values of the first-stage F statistics, mostly arising from the very significant impact of contiguous Maghreb population on contiguous arabic name babies (a coefficient of 0.95 with a standard error of 0.06 in the pooled regression). As we saw for Saint names, Sargan statistics vary over time, in this case being statistically significant in four out of the six census years. Contrary to traditional Saint names, choices of Arabic names do not appear to exhibit social interaction between departments. It should be noted however that those interactions might take place at a narrower geographic level (quarters inside cities for instance).[12] The greater than one coefficient on Maghreb population share is suggestive of such interactions taking place, at least in the first census years, increases in the share of the Maghreb population have a more than one-for-one impact on the share of Arabic names.

Table 3 looks at the second set of "imported" practices in French cultural patterns, namely American sounding names. The effects of class composition are even less significant for this name type. Only 4 of a possible 35 (7 samples, 5 included class variables) coefficients are statistically significant at the 10% or better level. This compares to 12 out of 35 for both Saint and Arabic names. Origin composition is much more erratic than it was for the other two name-types. Despite the very small share of US-born inhabitants (the maximum was 2.2% in Indre in 1962, the median was one hundredth of the maximum), we see a positive impact in all but the last year (note however that the 1999 census was a 5% sample) and the pooled

[12]Alternatively, those results might have something to do with the relatively recent arrival of the Arabic community in France.

Table 2: Explaining shares of Arabic names in each Department

	Dependent Variable: Arabic share						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	-0.01	0.03	-0.09^b	-0.04	0.05	-0.19^a	0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.06)	(0.05)	(0.06)
male	0.00^a	0.00^a	-0.01^a	0.00	0.01^a	0.01^a	0.01^a
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Corsica	-0.03^a	0.00	-0.05^a	-0.07^a	-0.03^a	-0.01	-0.02^b
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
% Farmers	-0.01	-0.04	0.09^b	0.06	-0.07	0.16^a	-0.07
	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)	(0.06)	(0.07)
% Craft	-0.01	-0.04	0.16^a	-0.11^c	-0.08	0.15	-0.04
	(0.05)	(0.06)	(0.05)	(0.06)	(0.09)	(0.10)	(0.14)
% Superior	0.11^c	0.29^b	0.13	0.09	0.00	0.46^a	-0.02
	(0.06)	(0.13)	(0.12)	(0.10)	(0.19)	(0.13)	(0.14)
% Clerks	0.01	-0.07	0.08^c	0.03	-0.15^c	0.25^a	-0.01
	(0.05)	(0.05)	(0.05)	(0.05)	(0.08)	(0.07)	(0.08)
% Manual	0.03	-0.01	0.12^a	0.06	-0.01	0.17^a	-0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.07)	(0.06)	(0.07)
% African	0.53^b	0.54	6.53^a	2.21^a	1.06	-0.14	0.64^c
	(0.26)	(1.20)	(1.70)	(0.82)	(0.84)	(0.48)	(0.34)
% Maghreb	1.15^a	1.88^a	1.61^a	1.22^a	1.27^a	1.13^a	1.17^a
	(0.10)	(0.48)	(0.09)	(0.07)	(0.09)	(0.09)	(0.09)
% Other nat.	0.00	0.00	0.01	0.03	0.02	-0.09^b	-0.09^b
	(0.02)	(0.01)	(0.01)	(0.02)	(0.04)	(0.04)	(0.04)
% US-born	-0.29^b	-0.15	-14.75^a	-5.98^a	9.87^c	-3.90	2.32
	(0.11)	(0.14)	(2.72)	(2.06)	(5.06)	(3.66)	(4.47)
Contig % Arabic (IV)	-0.22^b	0.15	-0.39^a	-0.25^a	-0.20^b	0.02	-0.27^b
	(0.10)	(0.20)	(0.11)	(0.08)	(0.08)	(0.11)	(0.12)
observations	1080	180	180	180	180	180	180
R^2	0.751	0.533	0.826	0.874	0.838	0.835	0.810
RMSE	.009	.007	.005	.006	.009	.008	.01
Contig % Arabic (OLS)	0.18^b	0.29^b	-0.14	-0.28^a	-0.10	0.25^a	0.08
	(0.08)	(0.14)	(0.09)	(0.07)	(0.07)	(0.09)	(0.09)
F stat 1st stage	78.27*	12.78*	46.02*	65.52*	104.75*	45.32*	43.42*
Sargan stat	32.74^a	18.48^b	31.34^a	10.8	15.36	21.56^a	22.37^a

Note: Robust standard errors in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. * denotes significance at the 5% level for the weak IV bias test of Stock and Yogo (2002, table 1) Errors allow for clustering by department in the pooled regression, which also includes year fixed effects.

Table 3: Explaining shares of American names in each Department

	Dependent Variable: American share						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	-0.04	0.03	-0.02	-0.16	-0.05	-0.13	-0.01
	(0.07)	(0.07)	(0.09)	(0.10)	(0.11)	(0.14)	(0.13)
male	0.00^a	0.00^c	0.00^c	0.01^b	0.00	0.01	-0.01^b
	(0.00)	(0.00)	(0.00)	(0.01)	(0.00)	(0.01)	(0.01)
Corsica	0.04^a	0.04^a	0.03^b	0.05^c	0.09^a	0.10^b	0.12
	(0.02)	(0.01)	(0.02)	(0.02)	(0.03)	(0.05)	(0.07)
% Farmers	0.03	-0.02	0.03	0.12	-0.09	-0.14	-0.17
	(0.07)	(0.08)	(0.09)	(0.11)	(0.13)	(0.14)	(0.15)
% Craft	-0.11	-0.07	-0.05	0.04	0.10	0.09	0.18
	(0.11)	(0.11)	(0.11)	(0.16)	(0.15)	(0.24)	(0.37)
% Superior	0.01	-0.12	0.42	0.32	0.06	0.27	0.08
	(0.13)	(0.24)	(0.27)	(0.29)	(0.32)	(0.32)	(0.31)
% Clerks	0.17^b	0.03	-0.04	0.25^b	0.15	0.43^b	0.30
	(0.09)	(0.09)	(0.10)	(0.12)	(0.14)	(0.18)	(0.18)
% Manual	0.09	0.04	0.10	0.28^b	0.11	0.26	0.09
	(0.07)	(0.08)	(0.09)	(0.11)	(0.13)	(0.16)	(0.16)
% African	0.91^b	6.61^a	0.35	1.39	0.39	-0.56	0.29
	(0.43)	(2.16)	(3.66)	(2.28)	(1.46)	(1.21)	(0.73)
% Maghreb	0.22^a	0.11	-0.10	-0.21	-0.17	0.09	0.10
	(0.08)	(0.82)	(0.19)	(0.16)	(0.14)	(0.22)	(0.18)
% Other nat.	0.08^b	0.15^a	0.12^a	0.02	0.05	0.04	0.18^c
	(0.04)	(0.03)	(0.03)	(0.06)	(0.07)	(0.10)	(0.11)
% US-born	1.05^a	1.15^a	6.93	10.48^c	10.23	10.87	-6.69
	(0.40)	(0.26)	(5.78)	(5.50)	(8.76)	(9.02)	(10.81)
Contig % Amer. (IV)	0.35^a	0.04	0.30	0.58^a	0.73^a	0.49^b	0.50^c
	(0.11)	(0.15)	(0.18)	(0.16)	(0.15)	(0.21)	(0.26)
observations	1080	180	180	180	180	180	180
R^2	0.935	0.568	0.53	0.686	0.626	0.593	0.519
RMSE	.02	.013	.011	.016	.016	.021	.023
Contig % Amer. (OLS)	0.75^a	0.44^a	0.64^a	0.72^a	0.97^a	0.84^a	0.84^a
	(0.12)	(0.09)	(0.10)	(0.10)	(0.09)	(0.10)	(0.09)
F stat 1st stage	12.99*	12.06*	15.61*	25.59*	16.21*	8.5	7.78
Sargan stat	48.26^a	21.48^a	16.9^b	15.88^b	18.52^b	7.13	13.22

Note: Robust standard errors in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. * denotes significance at the 5% level for the weak IV bias test of Stock and Yogo (2002, table 1) Errors allow for clustering by department in the pooled regression, which also includes year fixed effects.

coefficient is significant and has a reasonable magnitude. The pooled results suggest that all foreign-born groups are more inclined towards American names than the French-born. The relative antipathy of the French-born towards American names and the Maghreb-born towards Saint names is consistent with the Bisin and Verdier (2001) model of parental socialization.

Spatial interactions seem much more prevalent for this name type than for Arabic names. The pooled estimate is 0.35, which is remarkably similar to the pooled coefficient for Saint names (0.34). In contrast to the Saint names, we observe no decline in spatial interactions for American names. As with the other two name types, 2SLS appears successful in correcting the substantial (more than doubling) bias of the OLS coefficients. The caveat raised earlier about the erratic behavior of the overidentification tests still applies.

The implications of the mixed results of Sargan tests over time are unclear. It is hard to understand why composition variables would be endogenous in a given year and exogenous in another. One possibility is that composition variables in contiguous locations have a direct effect on local naming patterns, which does not enter through the impact on contiguous names. The concerns raised by the Sargan statistics reinforce the attractiveness of an approach that does not require instrumental variables, such as the one described in equation (9), and implemented in section 5.

5 Name-type dissimilarity

In the regressions of name-type shares on contiguous averages, we maintained an easy-to-estimate linear specification while imposing an unrealistic structure on the geographic form of spatial interactions. Indeed we had to include a dummy for Corsica so that its lack of contiguous neighbors would not bias the results. It seems more plausible (and consistent with the literature on gravity equations) to specify interactions as a continuous function of distance. While it would have been possible to implement such an approach in a name share specification, we opt instead to introduce a new specification inspired by equation (8). The idea is to explain dissimilarity in name-types with measures of social dissimilarities and geographic distance (see the theoretical section for more detail). The dissimilarity approach has the additional benefit of moving the endogenous neighbor shares, s_n in the model, over to the left-hand side of the regression, eliminating the need for instrumental variable methods.

Our measure of dissimilarity between locations ℓ and n is the Manhattan distance, i.e. the sum of the absolute differences in shares. For each dependent variable, we have a very simple definition of name type, which can be for instance either Saint or not. This yields

$$\text{MDN}^{\text{st}}_{\ell n} = \mid s^{\text{st}}_{\ell} - s^{\text{st}}_{n} \mid + \mid s^{\text{nst}}_{\ell} - s^{\text{nst}}_{n} \mid = 2 \mid s^{\text{st}}_{\ell} - s^{\text{st}}_{n} \mid,$$

where s^{st}_{ℓ} is the share of Saint (st) names in location ℓ, while s^{nst}_{ℓ} is the share of non-Saint (nst) names in this same location.

We attempt to explain name dissimilarity between departments using information on other aspects of dissimilarity among the parents of the two departments

Table 4: Explaining differences in name-type shares: Saint names

	Dependent Variable: Dissimilarity in Saint shares						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	-0.071^a	-0.019^a	-0.124^a	-0.147^a	-0.050^a	0.035^a	-0.012
	(0.005)	(0.006)	(0.008)	(0.008)	(0.009)	(0.009)	(0.009)
male	0.010^a	-0.005^a	0.015^a	0.035^a	0.002	0.000	0.016^a
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
contiguous	0.000	0.000	0.007^a	0.013^a	-0.003	-0.012^a	-0.012^a
	(0.002)	(0.002)	(0.003)	(0.003)	(0.004)	(0.004)	(0.004)
ln distance	0.012^a	0.010^a	0.022^a	0.025^a	0.008^a	-0.003^b	-0.004^b
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)
MD: class	0.040^a	-0.002	0.014^b	0.051^a	0.105^a	0.143^a	0.153^a
	(0.005)	(0.004)	(0.006)	(0.008)	(0.010)	(0.011)	(0.011)
MD: origins	0.123^a	-0.020^b	-0.038^a	-0.068^a	0.230^a	0.447^a	0.349^a
	(0.007)	(0.008)	(0.010)	(0.010)	(0.013)	(0.014)	(0.012)
observations	48060	8010	8010	8010	8010	8010	8010
R^2	0.238	0.221	0.312	0.407	0.355	0.5	0.569
RMSE	.055	.04	.043	.046	.053	.053	.047

Note: Department fixed effects included. Robust standard errors in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. Errors allow for clustering by department-pair in the pooled regression, which also includes year fixed effects.

under investigation. Other dissimilarity metrics can be calculated as

$$\text{MDC}_{\ell n} = \sum_{g} \mid x_{g\ell} - x_{gn} \mid,$$

where g are levels of social class (agriculture, craftsmen and entrepreneurs, professionals, intermediates, clerical workers, and manual workers), and

$$\text{MDO}_{\ell n} = \sum_{g} \mid x_{g\ell} - x_{gn} \mid,$$

where g are countries of origin and x_g are category shares of child-bearing age population. Summary statistics for the name-type and group composition variables are show in an appendix as Table 10.

The results in Table 4 broadly corroborate the findings for Saint name shares in Table 1. First, geographic proximity tends to promote name similarity. In the pooled results and the 1962–1982 census years, the further apart two departments are, the more different is their share of Saint names. The reported coefficients are semi-elasticities and not easy to interpret. Alternatively, one can scale by relative standard deviations and say that a one standard deviation increase in distance raises Saint dissimilarity by $0.012 \times (0.588/0.063) = 0.112$ standard deviations.[13]

In the pooled results, contiguity does not matter after controlling for distance (non-contiguous departments are on average about 5 times further apart than contiguous ones). The estimates suggest that the effect of distance on Saint name dissimilarity is declining. One hard-to-interpret result is that in the last two census years the sign of distance flips (to be perversely negative) while contiguity comes in with the expected negative sign.[14] If we re-estimate without the contiguity dummy, the distance effect is about the same for all years except the final two, where it comes in as 0.000 and insignificant. The takeaway is that geographic proximity was once a fairly important influence on similarity in the propensity to name children after Saints but this effect has disappeared.

The pooled results confirm that differences in class and origin composition widen the differences in Saint name usage. The raw coefficient on class is one-third of that on origins. However, class differences between pairs of departments exhibit more variation than origin differences. The standardized coefficients reveal that one-standard deviation increases in class and origin dissimilarity raise Saint name dissimilarity by 0.081 and 0.146 standard deviations (respectively), magnitudes that are comparable to the pooled standardized distance effects. As distance effects have waned, both composition variables have become much more important over time. In 1999 the standardized coefficients are 0.201 for class and 0.393 for origins.

[13]Standardized coefficients are obtained by multiplying the coefficients reported in the regression tables by the ratio of the standard deviations of the explanatory and dependent variables. Standard deviations required for these calculations are provided in Table 10.

[14]The switch in the sign of the effects of contiguity and distance in the late years for Saint regressions is suggestive of a change in the scope of spatial interactions. However, this inversion only applies to Saint names regressions.

Table 5: Explaining differences in name-type shares: Arabic names

	Dependent Variable: Dissimilarity in Arabic shares						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	0.000	0.003	0.007^a	0.008^b	0.060^a	0.006	0.016^a
	(0.003)	(0.002)	(0.003)	(0.004)	(0.006)	(0.005)	(0.005)
male	0.009^a	-0.005^a	-0.001^a	0.007^a	0.013^a	0.019^a	0.020^a
	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)	(0.000)
contiguous	0.000	-0.001	-0.002	0.001	0.002	-0.003	-0.004^b
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
ln distance	-0.002^a	0.000	-0.002^a	-0.001^b	0.000	-0.004^a	-0.006^a
	(0.001)	(0.000)	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)
MD: class	0.036^a	0.022^a	0.031^a	0.054^a	0.038^a	0.060^a	0.090^a
	(0.002)	(0.002)	(0.002)	(0.004)	(0.006)	(0.005)	(0.007)
MD: origins	0.179^a	0.016^a	0.082^a	0.174^a	0.312^a	0.267^a	0.268^a
	(0.004)	(0.003)	(0.003)	(0.005)	(0.009)	(0.007)	(0.008)
observations	48060	8010	8010	8010	8010	8010	8010
R^2	0.598	0.802	0.818	0.691	0.612	0.718	0.753
RMSE	.023	.009	.011	.018	.026	.021	.023

Table 6: Explaining differences in name-type shares: American names

	Dependent Variable: Dissimilarity in American shares						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	-0.028^a	-0.003	-0.004	-0.047^a	-0.015^a	-0.003	0.009^c
	(0.003)	(0.004)	(0.003)	(0.005)	(0.005)	(0.006)	(0.005)
male	-0.008^a	0.005^a	0.008^a	0.001^c	-0.011^a	-0.019^a	-0.034^a
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
contiguous	0.001	0.000	-0.002	0.006^a	0.001	0.000	0.001
	(0.001)	(0.001)	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)
ln distance	0.008^a	0.002^b	0.001	0.013^a	0.013^a	0.006^a	0.010^a
	(0.000)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
MD: class	0.043^a	0.045^a	0.042^a	0.076^a	0.037^a	0.072^a	0.087^a
	(0.002)	(0.002)	(0.002)	(0.005)	(0.006)	(0.007)	(0.007)
MD: origins	0.055^a	0.087^a	0.070^a	-0.045^a	-0.004	0.141^a	0.134^a
	(0.004)	(0.006)	(0.004)	(0.006)	(0.008)	(0.008)	(0.007)
observations	48060	8010	8010	8010	8010	8010	8010
R^2	0.246	0.514	0.369	0.369	0.382	0.421	0.496
RMSE	.04	.024	.022	.033	.034	.041	.042

Note: Department fixed effects included. Robust standard errors in parentheses with ^{a}, ^{b} and ^{c} respectively denoting significance at the 1%, 5% and 10% levels. Errors allow for clustering by department-pair in the pooled regression, which also includes year fixed effects.

Table 5 shows that geographic separation does not increase dissimilarity for Arabic name shares. The weakly perverse effect of distance is consistent with the negative coefficients found for contiguous Arabic shares in Table 2. Also corroborating earlier results, we find much stronger results for origin composition differences than class composition differences: the standardized coefficients for origins are three times higher than for class. In contrast to the share regressions, the dissimilarity regressions show evidence that both composition effects are stronger in the last three census years than in the pooled results.

The results shown in Table 6 provide the most consistent support for distance promoting dissimilarity in naming patterns. The standardized coefficient in the pooled regression is 0.102, slightly less than the corresponding effect for Saint differences. Unlike the case for Saints, but corroborating what we saw in Table 3, distance effects for American names are not declining over time. Origin effects are smaller for this name-type with the effect of a standard deviation change actually being smaller for origins (0.090) than class (0.120) using the pooled regression coefficients.

6 Name dissimilarity results

The Manhattan distance metric allows us to use the full richness of the name distribution to measure dissimilarity in choices. It also obviates the need for dichotomous classifications. Let $s_{\ell k}$ now equal the frequency of name k in department ℓ in a given year. The dissimilarity between names in two departments ℓ and n is given by the Manhattan distance:

$$\text{MDN}_{\ell n} = \sum_k \mid s_{\ell k} - s_{nk} \mid . \tag{10}$$

$\text{MDN}_{\ell n}$ is considered a metric because it meets certain conditions seen as desirable for distance measures. Most importantly, it equals zero for pairs of departments with identical choices. We also considered the correlation between name frequencies and a measure of "overlap" calculated as $\sum_k s_{\ell k} \times s_{nk}$. Correlation meets the zero criteria since one minus the correlation equals zero for sets of identical frequencies. However, since it measures the strength of a linear relationship, correlation is somewhat difficult to interpret as a similarity measure. Overlap has the interpretation of being the probability that two children born in different locations receive the same names. It suffers from the defect of overlap with self not being one. This is because overlap combines information on similarity in choice with heterogeneity (or variety) of choice. Overlap could fall over time if preferences became more heterogenous even if locations were becoming more similar.[15]

One problem in calculating name-level dissimilarity is that the frequency of a name can be known in one location but not in another. This arises in our data because INSEE does not tabulate names given to two or one child in a given year. Although this data problem makes it impossible to calculate the exact manhattan distance, we can calculate upper and lower bounds for dissimilarity.

[15]We thank Tony Smith for pointing out the flaw in overlap and suggesting Manhattan Distance as an alternative.

The upper bound is straightforward. It assumes that the two locations use entirely different sets of rare names. This implies replacing missing frequencies with zeros. Thus, if a name's frequency is known in only one location, we add the frequency there to the sum of absolute differences. For the remaining counts of names that are rare in both locations, we add $s_\ell^R + s_n^R$ to the sum.

For the lower bound we allocate names that are rare in one department to non-rare names in the other department. The most one can allocate, of course, are two. Thus the element contributed to the sum would be $\mid s_{\ell k} - \underline{s}_{nk} \mid$, where $\underline{s}_{nk}$ equals 2 divided by the number of babies born in department n.

We regress $\text{MDN}_{\ell n}$ (calculated as the average of the upper and lower bound values MDN can take) on geographic distance between ℓ and n. As before, we control for differences in class and origin composition. We also include department level fixed effects.

Table 7: Manhattan Distance in names

	Dependent Variable: Manhattan Distance						
Sample:	All	1962	1968	1975	1982	1990	1999
intercept	0.166[a]	-0.023	0.079[a]	0.210[a]	0.304[a]	0.347[a]	0.421[a]
	(0.010)	(0.016)	(0.011)	(0.011)	(0.010)	(0.011)	(0.013)
male	-0.036[a]	-0.044[a]	-0.021[a]	-0.023[a]	-0.040[a]	-0.047[a]	-0.040[a]
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
contiguous	0.001	0.002	-0.001	-0.001	-0.002	-0.004	-0.002
	(0.004)	(0.006)	(0.004)	(0.004)	(0.004)	(0.004)	(0.005)
ln distance	0.055[a]	0.086[a]	0.069[a]	0.052[a]	0.044[a]	0.030[a]	0.023[a]
	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
MD: class	0.120[a]	0.119[a]	0.115[a]	0.235[a]	0.209[a]	0.309[a]	0.377[a]
	(0.008)	(0.009)	(0.008)	(0.012)	(0.011)	(0.012)	(0.015)
MD: origins	0.223[a]	0.119[a]	0.171[a]	0.186[a]	0.282[a]	0.244[a]	0.240[a]
	(0.012)	(0.018)	(0.013)	(0.014)	(0.014)	(0.015)	(0.015)
observations	48060	8010	8010	8010	8010	8010	8010
R^2	0.741	0.748	0.766	0.768	0.819	0.864	0.841
RMSE	.069	.069	.055	.053	.048	.049	.057

Note: Department fixed effects included. Robust standard errors in parentheses with [a], [b] and [c] respectively denoting significance at the 1%, 5% and 10% levels. Errors allow for clustering by department-pair in the pooled regression, which also includes year fixed effects.

Table 7 shows that by moving away from name-dichotomies, we seem to obtain much clearer patterns in the results. Most notably, distance effects are twice as large: the pooled standardized coefficient is 0.238 (compared to 0.112 and 0.102 for Saint and American names). Table 8 compares standardized coefficients on distance, class, and origins for all the name dissimilarity measures.

The distance effects for name dissimilarity exhibit a steady decline over time, while nevertheless remaining significant in the final sample (the standardized coeffi-

Table 8: Standardized coefficients from the pooled dissimilarity regressions

Dep. var.:	Indep.var.: (Std. dev.)	Log distance (0.588)	MD: Class (0.128)	MD: Origins (0.075)
MD: Saint	(0.063)	0.112	0.081	0.146
MD: Arabic	(0.037)	-0.032	0.125	0.363
MD: American	(0.046)	0.102	0.120	0.090
MD: Names	(0.136)	0.238	0.113	0.123

cient falls to 0.096 in 1999). Contiguity is always small and insignificant, supporting the hypothesis of continuous effects of spatial separation. Class and origin composition both have strong effects and both are rising in importance over time, with the standardized coefficient on class differences actually more than doubling from 0.113 to 0.249. Thus, we see a remarkable transposition in the relative strengths of geography and class composition.

A decrease in the influence of distance on inter-departmental interactions could occur because improved transportation and communication infrastructure lowers the cost of long-distance direct interactions or because a wider share of the population is exposed to a common set of media interactions. The regression specification shown in Table 7 can only be estimated in census years. However, we would like to assess the evolution of the distance effect over the full range of data for which we can calculate $\text{MDN}_{\ell n}$. Hence we dispense with the composition controls and estimate year-by-year regressions with just distance, department fixed effects and a dummy for males. Fortunately, the resulting coefficients on distance, shown as the dark line in Figure 4, are quite similar to the coefficients obtained in Table 7.

Figure 4 illustrates the decreasing estimated effect of geographic distance on name dissimilarity. Since we would like to interpret this as a consequence of interactions becoming less localized over time, it may be informative to see whether the declining impact of geography matches up in time with other changes that might have been influencing spatial interactions. Using a second scale on the right-axis we show the penetration of television and car-ownership in French society. Those two variables are natural candidates for measuring the increase in mass-media influence and the decrease in transport costs, both likely to yield an increase in the average distance of social interactions in recent years compared to the fifties. With improved long-distance communication means, the locality of cultural patterns is likely to fall over time. Note that the simultaneity of the time trends in Figure 4 should be seen as illustrative and not implying causation, which would (at least) imply the huge task of constructing bilateral measures of those two variables over the time range under consideration. Overall, the spatial proximity factor in naming practices is falling dramatically over time, a pattern that is not inconsistent with the rise of mass-media and easier long-distance travel.

Faced with the fading impact of spatial proximity in naming similarities within France, one is naturally inclined to ask whether the globalization of interactions

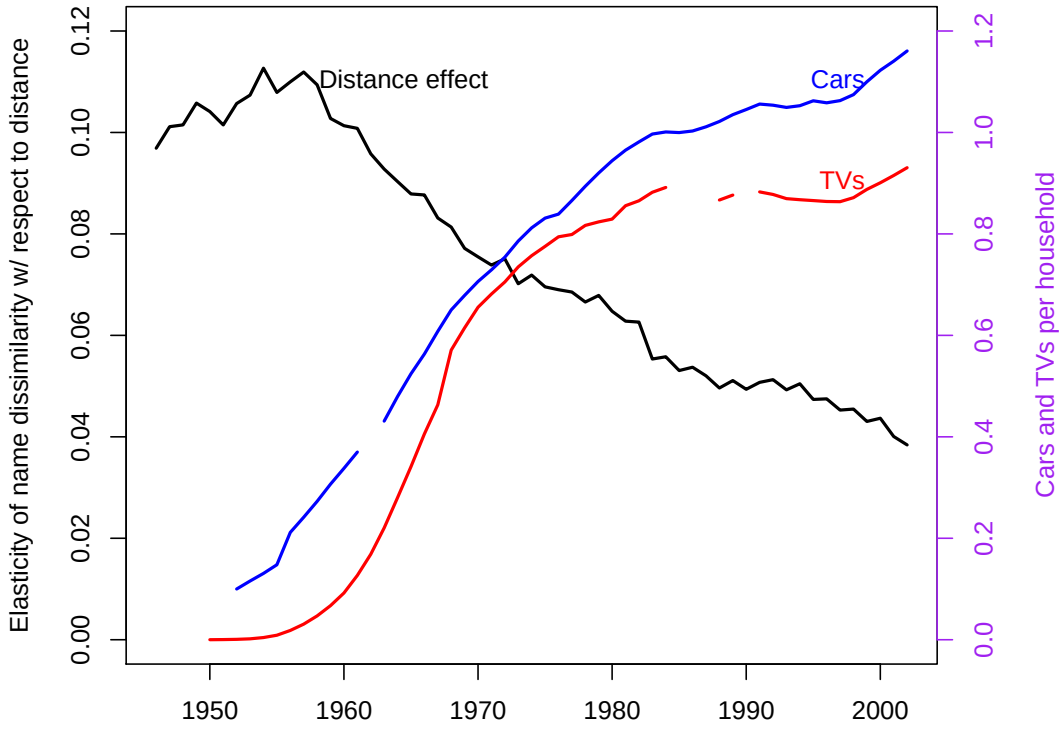


Figure 4: The coefficient on distance in year-by-year fixed effects regressions of bilateral Manhattan Distance on geographic distance, compared with the penetration of TVs and cars.

in cultural patterns extends beyond France's borders. While collecting the same data for a large set of countries and comparable time period seems out of reach, it is possible to obtain the list of most popular names in recent years for certain countries that are interesting to compare with France. Table 9 provides pairwise Manhattan Distances between top-100 names in France, certain remote regions of France (Corsica and overseas departments, DOM), the two different linguistic parts of Belgium (Wallonia where French is spoken, and Flanders where Flemish is spoken), different Canadian provinces (French-speaking Quebec and English-speaking British Columbia), and the United States. The bottom left triangle gives figures for girls' names and the upper right one for boys'. The number of names common to both places is shown in parentheses.

The most striking feature of this table lies with the linguistic border. While countries/regions speaking the same language often have small MDs and large numbers of common names in the top-100 list, it is generally not the case for the combinations of countries/regions using different languages. France's Manhattan distance to Wallonia is much smaller than its distance to Flanders even though their is little difference in their geographic differences. Physical distance can be seen to play as important a role as political borders. Names in Wallonia are more proximate to the set of names used in France than names used in Corsica or in the overseas departments. Names used in Quebec are far more different (although they show almost no relationship whatsoever with names used in Anglophone countries/regions, even BC). Names used in British Columbia are very close to names used in the United states. Overall, while the spatial overlap in naming practices within France has become very large over time, it is certainly not the case that this level of similarity extends simply to international comparisons. There, distance, borders, and language would appear to remain as strong barriers to convergence in tastes.

7 Conclusion

Most parents we have spoken to regard the choice of their child's name as an idiosyncratic decision determined by personal tastes and, in some cases, family histories. We have presented evidence that—on the contrary—there are systematic forces at work governing naming practices. Social class and national origins matter, as do decisions by other parents. We also showed how those complex determinants can be articulated in a coherent and simple model of social interaction, that encompasses idiosyncratic tastes, group preferences, and the influence of spatially proximate agents. Despite the relatively large scale of the geographic areas in our study, we find strong evidence of spatial interactions. In our pooled two-stage least squares estimates, a 10 percentage point increase in the popularity of a Saint or American-type names in neighboring departments increases local shares of those name-types by 3.4 and 3.5 percentage points, respectively. In contrast, the Arabic names brought by Maghreb immigrants seem to be transmitted only through the vertical (intergenerational) channel, and do not diffuse to nearby places with low levels of immigration. Indeed, some of our estimates indicate that use of Arabic names in nearby departments

Table 9: Manhattan Distances and in-common names for top 100 names in 2000

	France	DOM	Corsica	Wallonia	Quebec	Flanders	BC	USA
France		0.47	0.53	0.36	0.88	0.88	1.08	1.04
		(71)	(56)	(72)	(44)	(27)	(20)	(24)
DOM	0.43		0.62	0.53	0.77	0.79	0.91	0.87
	(62)		(48)	(58)	(48)	(22)	(21)	(22)
Corsica	0.48	0.55		0.6	0.93	0.99	1.17	1.12
	(59)	(48)		(54)	(36)	(15)	(12)	(14)
Wallonia	0.29	0.47	0.55		0.85	0.85	1.03	1
	(75)	(57)	(53)		(48)	(29)	(25)	(27)
Quebec	0.87	0.76	0.99	0.84		1.07	1.01	0.93
	(36)	(36)	(27)	(36)		(23)	(41)	(43)
Flanders	0.69	0.67	0.79	0.66	0.96		1.02	0.99
	(35)	(24)	(28)	(38)	(19)		(14)	(15)
BC	0.96	0.82	1	0.96	0.86	0.86		0.36
	(19)	(21)	(17)	(18)	(37)	(20)		(73)
USA	0.98	0.81	1.01	0.97	0.9	0.87	0.29	
	(15)	(18)	(14)	(14)	(30)	(15)	(75)	

Note: The bottom left panel gives figures for baby girls and the upper right for baby boys.

causes a decline in local use.

There are intriguing trends in the strength of spatial interactions. Dissimilarity in usage of Saint names and specific name dissimilarity were more significantly affected by distance between location pairs in past decades. The effect of a one standard deviation increase in distance on name dissimilarity has declined markedly (from 0.4 standard deviations in 1962 to 0.1 standard deviations in 1999) and disappeared altogether for Saint-name dissimilarity. Nevertheless, American names show signs of strong spatial interactions that are not disappearing over time. These names might be introduced by the small numbers of American-born living in France but it seems more likely that they enter via travel by the French-born and their exposure to foreign media. Regardless of the original source of these names, it would seem that interactions promote their diffusion.

The methods and results presented here should provide some value for social scientists, even if they have no particular interest in naming patterns. First, our linear model combining composition effects and social interactions could prove useful in a variety of contexts. Second, we have shown a case where theoretically motivated instrumental variables do what they are supposed to do and drive down the estimated coefficients on neighborhood effects. Third, we introduce a technique for estimating effects of distance and composition differences on dissimilarity. This technique suggests that continuous distance effects may be preferable over discontinuous contiguity effects even at the geographic scale of a nation. Finally, we think the evidence points to a declining role for spatial separation and an increasing role for social, ethnic, and linguistic separation. These findings would be of broader significance if they were corroborated in studies of other expressions of household preferences beside names.

Appendix: Summary Statistics

Table 10: Descriptive statistics for variables in dissimilarity regressions

Variable	Mean	St. Dev.	Min	Max
Pooled (48060 obs)				
MD: Saint	0.08	0.063	0	0.526
MD: Arabic	0.034	0.037	0	0.231
MD: American	0.053	0.046	0	0.459
MD: Names	0.625	0.136	0.209	1.342
Log distance	5.856	0.588	2.823	7.138
MD: class	0.224	0.128	0.011	0.888
MD: origins	0.104	0.075	0	0.493
1962 (8010 obs)				
MD: Saint	0.06	0.044	0	0.301
MD: Arabic	0.016	0.021	0	0.099
MD: American	0.039	0.034	0	0.241
MD: Names	0.581	0.137	0.209	1.234
Log distance	5.856	0.588	2.823	7.138
MD: class	0.296	0.166	0.014	0.888
MD: origins	0.102	0.081	0	0.429
1999 (8010 obs)				
MD: Saint	0.091	0.071	0	0.469
MD: Arabic	0.046	0.046	0	0.231
MD: American	0.067	0.058	0	0.459
MD: Names	0.721	0.141	0.403	1.342
Log distance	5.856	0.588	2.823	7.138
MD: class	0.181	0.093	0.011	0.605
MD: origins	0.112	0.08	0.001	0.493

References

Bertrand, M. and S. Mullainathan (2004), "Are Emily and Greg more Employable than Lakisha and Jamal? A Field Experiment on Labor Market Discrimination," *The American Economic Review*, 94(4), 991–1013.

Besnard, P. (1995), "The Study of Social Taste Through First Names: Comment on Lieberson and Bell," *American Journal of Sociology* 100(5), 1313–1317.

Bissin, A. and T. Verdier (2001),"The economics of cultural transmission and the dynamics of preferences," *Journal of Economic Theory* 97, 298–319.

Brock, W.A. and S.N. Durlauf (2001), "Discrete Choice with Social Interactions" *Review of Economic Studies* 68(2), 235–260.

Brock, W.A. and S.N. Durlauf (2002), "A Multinomial-Choice Model of Neighborhood Effects" *American Economic Review* 92(2), 298–303.

Cavalli-Sforza, L.L. and M.W. Feldman, (1981), *Cultural Transmission and Evolution: A quantitative approach*, Princeton University Press.

Disdier, A-C and K. Head, forthcoming, "The Puzzling Persistence of the Distance Effect on Bilateral Trade," *Review of Economics and Statistics*.

Disdier, A-C, K. Head, and T. Mayer (2006), "Exposure to foreign media and changes in cultural traits: Evidence from naming patterns in France" CEPR DP #5674.

Durlauf, S.N. (2004), "Neighborhood Effects" in Henderson V. and J.F. Thisse (eds.) *Handbook of Regional and Urban Economics* Volume 4, Amsterdam: Elsevier, 2174–2234.

Durlauf, S.N. and H.P. Young (2001), "The New Social Economics," in Durlauf, S.N. and H.P. Young, (eds.), *Social Dynamics*, Brookings Institution Press, Washington, D.C.

Figlio, D.N., 2005, "Names, Expectations and the Black-White Test Score Gap," NBER Working Paper # 11195.

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

Glaeser, E. and J. Scheinkman (2002), "Non-market interactions" in *Advances in Economics and Econometrics: Theory and Applications, Eight World Congress*, M. Dewatripont, L.P. Hansen, and S. Turnovsky (eds.), Cambridge University Press, 2002 (Also NBER WP #8053, December, 2000).

Glaeser, E. and J. Scheinkman, 2001, "Measuring social interactions" in Durlauf, S.N. and H.P. Young, (eds.), *Social Dynamics*, Brookings Institution Press, Washington, D.C.

Grossman, G. M., (1998), "Comment," in Frankel J.A. (ed), *The Regionalization of the World Economy*, NBER Project Report, The University of Chicago Press.

Head, K. and T. Mayer (2000), "Non-Europe: The Magnitude and Causes of Market Fragmentation in the EU," *Weltwirtschaftliches Archiv* 136(2):285–314.

Jouniaux, Léo, 2001, *Les vingt mille plus beaux prénoms du monde* Hachette Pratique, Paris.

Lieberson S. and E.O. Bell (1992), "Children's First Names: An Empirical Study of Social Taste," *American Journal of Sociology* 98, 511–554.

Lieberson, S. (1995), "Reply to Philippe Besnard" *American Journal of Sociology* 100(5), 1317–1325.

Lieberson, S. (2000), *A Matter of Taste: How names, fashions, and culture change*, Yale University Press: New Haven.

Lucas, Robert E. Jr. (2001), "Externalities and Cities," *Review of Economic Dynamics* 4, 245–274.

Manski, Charles F. (1993), "Identification of Endogenous Social Effects: The Reflection Problem" *The Review of Economic Studies*, 60(3), 531–542.

Scheinkman, J. (forthcoming), "Social Interactions," in the *New Palgrave Dictionary of Economics and Law.* `http://www.princeton.edu/~joses/wp/socialinteractions.pdf`

Stock, J. and M. Yogo (2002), "Testing for Weak Instruments in Linear IV Regression," NBER Technical Working Paper 284.

Young, H.P. (2001), "Dynamics of Conformism," in Durlauf, S.N. and H.P. Young, (eds.), *Social Dynamics*, Brookings Institution Press, Washington, D.C.