# Intergenerational linkages in consumption patterns and the geographical distribution of surnames[*]

M. Dolores Collado
(Universidad de Alicante)

Ignacio Ortuño-Ortín[†]
(Universidad Carlos III)

Andrés Romeu
(Universidad de Murcia)

## Abstract

This paper attempts to detect the existence of links in consumption patterns between generations. Preferences over consumption goods may be determined by the preferences of parents and/or by preferences arising from the environment. We propose an indirect methodology to overcome the lack of data on consumption choices of dynasties, i.e., parents and their adult offspring. This new approach is based on the analysis of the correlation between the geographical distributions of surnames and consumption choices. We show that there is no significant intergenerational link on consumption patterns for non-food goods. Our results also suggest that there is a link between parents' and children's preferences over food items.

*JEL classification*: **D12, R23**.

*Keywords*: Preference formation, migration, surnames, geographical consumption patterns, Mantel test.

# 1 Introduction

A central theme in the study of preference formation is the relative influence of parents versus the influence of the social environment in shaping a person's preferences. Thus, preference formation can be understood as emerging from a preference transmission from parents to offspring, and a preference transmission between any two individuals. The influence coming from parents is sometimes called "vertical transmission" and the influence from other people is called "horizontal transmission".[1] Vertical transmission of preferences can take place through children's imitation of their parents' tastes or through the parents' teaching of certain habits and values (and perhaps through some genetic inheritance). Horizontal transmission occurs when children adopt the habits of other agents outside the household, as for example when they learn values or preferences taught in school or when they imitate friends.

This paper provides a novel empirical approach to assess the importance of the parental and the environmental influence on the formation of children's preferences over *consumption goods*. Namely, we study the transmission of the rates of substitution for consumption goods as when a child adopts his parents' strong taste for eating meat rather than being vegetarian. However, we neither seek to describe the precise working of those channels of preference transmission nor attempt to provide any new theoretical explanation on the issue of preference formation. The paper just provides an empirical framework for detecting whether there is a significant intergenerational link on preferences over consumption goods.

Since preferences are not observable we will have to use data on consumption behavior instead. It will be argued that, after controlling for income and other variables that may affect consumption decisions, a correlation on the preferences of parents and their offspring implies a correlation on the consumption bundles chosen by parents and the consumption bundles chosen by their (adult) offspring. If there were data available on those consumption choices, assessing the significance of a intergenerational link would be an easy task. Unfortunately, unlike other types of intergenerational transmission, such as earnings or abilities, we are not aware of the existence of any survey that provides such information on a large number of consumption goods.

Thus, we propose an indirect strategy to overcome the data problem: we will compare the spatial distribution of consumption behavior and the spatial distribution of surnames. By studying whether regions with similar surname

---

[1] This terminology is based on Cavalli-Sforza&Feldman [5] (see also Bisin&Verdier [2]). These authors also distinguish between horizontal and oblique transmission.

distributions also have similar expenditure patterns, we will be able to draw some conclusions on the effect of parental versus environmental transmission of preferences.

The intuition behind this approach is simple: if parental transmission is very strong, preferences are transmitted from parents to children in a similar way to surnames. Consider an economy composed by geographical regions that originally were quite different in both the distribution of surnames and the distribution of consumption patterns. Suppose that there have been important recent migration flows among the regions. If parental transmission is perfect, so that surnames and preferences are transmitted in the same way, regions that have experienced large migration flows should be close both in the distribution of surnames and in the distribution of expenditure patterns. In contrast, if environmental transmission of preferences is the main force in the preference formation process, surnames and preferences evolve in different manner and regions that have experienced more migration flows will have more similar surnames distributions but not more similar consumption patterns. Thus, we will be able to detect intergenerational links by studying the spatial distributions of surnames and consumption patterns in different regions. More specifically, we will construct two matrices of distances between regions. The first matrix will be computed by calculating the distances in consumption patterns between regions and the second matrix by computing the distances in the distribution of surnames between regions. It will be argued that if parental transmission plays a significant role, after controlling for other relevant variables, those two matrices should be positively correlated. Unfortunately, there still remains an identification problem since a positive correlation need not imply parental transmission of preferences. If natives of the host region adopt the preferences of immigrants the two matrices would be positively correlated but this would be due to an environmental transmission of preferences. However, the case of no correlation between the two matrices will unambiguously indicate a strong environmental transmission effect. Moreover, the weaker the linkage between parent's preferences and children's preferences the faster immigrants adapt to their host societies. Thus, the analysis of the correlation of such matrices will enable us to draw some insights on both the general issue of preference formation and the more specific question on immigrant integration. Modern economies experience large migration flows. Research on the mechanism of assimilation of immigrants is important not only from a welfare point of view, but also to predict future trends in demand and consumption patterns.

It might happen that the correlation between the matrix of surnames distances and the matrix of consumption patterns is explained by other variables unrelated to the paternal/environment effect that we want to study

here. For example, it might happen that geographically closer provinces have similar surnames distributions and are also similar in the type of local agricultural produce available. The type of rural/urban environment might also condition consumption patterns because prices and the availability of certain goods depend on it. Thus, we will control for other factors that may have some explanatory power in the correlation between the two matrices. In particular, we will control for geographical distance, how urban/rural regions are, income differences, climate and household composition.

Using data from Spain, the paper shows that for non-food items there is no significant linkage between parents' preferences and those of their offspring. Regarding food consumption the paper suggests, but cannot unequivocally prove, that preferences are partially inherited from the parents. This result also suggests that in Spain, second generation immigrants adopt the consumption patterns of the host region for all goods except for food items.

An alternative explanation of our result for food items could be based on the household production model of Stigler and Becker[45]. Under such model we would be detecting transmission of skills in the household production of food related goods rather than transmission of preferences. Even though this distinction between preferences and technology could be seen as "a matter of semantics, not substance" (Pollak [37], page 375), we repeat our analysis differentiating between food items that are clearly cooking inputs and food items that are typically eaten without being further processed. In both cases we obtain again a positive correlation between the matrix of consumption distances and the matrix of surnames distances.

The statistical tool used to asses the possible correlation between distances in the distribution of surnames and distances in the distribution of consumption patterns is the multivariate Mantel Test, a method for testing (linear) correlation between distance matrices. The Mantel test has been applied to problems of Spatial Autocorrelation in Ecology and in Population Genetics (see Mantel [31], Sokal&Rohlf [43] and Legendre&Legendre [30]). However, to the best of our knowledge, this is the first time that this test has been applied in Economics (see Desmet, LeBreton, Ortuno-Ortin and Weber [17] for another recent application in Economics).

In this paper we use data from Spain, although the approach is general and the data needed is also available for many other countries. There are several reasons why Spain is an excellent case study. First, there were large internal migration flows in the last century.[2] Second, the number of for-

---

[2]In year 1960 18% of the Spanish citizens were living in a different region from the one where they were born. For years 1970, 1981 and 1991 the corresponding percentages were 22.6, 23.5 and 22.4 respectively (Carreras and Tafunell, 2005).

eign immigrants in Spain until the late 90's was very low compared to other countries. This is important because having a very high number of foreign immigrants would substantially complicate the analysis, as it would be necessary to have information on consumption patterns from the immigrants' countries of origin. Third, information on surnames is easily available as a computer database file using the telephone directory in electronic format. Fourth, the Spanish surveys on household consumption are of high quality as documented in Browning&Collado [4]. Finally, the regions considered are relatively small areas and therefore, we have a significant number of observation, and representative data are available at such aggregation level.

There is an extensive literature dealing with intergenerational correlations in, for example, income (Solon [44], Mazumder [32]), wealth (Charles&Hurst [9]) abilities and aggregated consumption (Mulligan [34]), IQ (Daniels et al [14], Feldman et al [21]), political orientation (Jennings et al [26]), generosity (Wilhelm et al[47]) intertemporal preferences (Becker&Mulligan [1]), altruistic preferences (Mulligan [33]), contributions to public goods (Cipriani et al [10]), and risk and trust attitudes (Dohmen et al [18]). Waldkirch, Ng and Cox [46] deals with the type of intergenerational correlation considered in this paper. They study the intergenerational transmission of consumption preferences using data from the PSID that contain information on the *total* food expenditure of parents and of their adult offspring. They are the first to investigate potential intergenerational correlation in consumption beyond that induced through an intergenerational transmission of permanent income. After controlling for income and other relevant variables, they find a significant intergenerational transmission on tastes for food. Unfortunately their data set does not provide more disaggregated information on consumption choices that is the sort of information required to analyze intergenerational transmission of preferences over consumption bundles. Our paper contributes to this literature since we analyze intergenerational links between parents' and children's preferences over a disaggregated set of consumption goods. We find that parental transmission of preferences affects the shares of the different food products rather than the total amount spent on food. Simonson&Sela [41] using a twins study find a heritable effect on preferences for certain products. However, the sample size is very small and the number of consumer goods analyzed is quite limited.

The distribution of surnames in the population has been used to analyze several issues in areas such as Population Genetics and Health Sciences. This is because it contains relevant information about geographical mobility and the mating structure in a society. Since there are links between surnames and genotypes, scientists working in Population Genetics have incorporated the distribution of surnames into the analysis of population genetic diversity

(Lasker [29] and Jobbling [28]). In Health Science, surnames can be useful in studying the relationship between levels of inbreeding and prevalence of certain types of tumor and other diseases of genetic origin (see for example Holloway&Soafer [25]).

In Economics, the study of surnames has been mostly applied in the analysis of very specific discrimination and social integration problems and on questions of social mobility (see for example Einav&Yariv [19], Fryer&Levitt [22] and Goldin&Shim [23], Collado et al [13], Clark [11],[12]) but not in issues of intergenerational transmission of preferences. One possible reason why the information contained in surnames and its distribution has not been more intensively exploited in the literature could be the fact that, until very recently, large data sets were not available in electronic format. However, things have changed dramatically in the last few years, and in most developed countries telephone directories on CD-ROM support are now easily available and contain information about basically all households.[3]

## 2 Consumption Preferences and Surnames Distribution

Our goal is to assess the possible correlations between the preferences of parents and children. The paper, however, does not look at the specific channel of preference transmission from parents to children, and it is compatible with the possibility that children mimic particular consumption behavior of parents and/or that consumption behavior is attributed to genetic inheritance (see Rowe [39] and Harris [24] for the view that parental influence on child outcomes is limited, Cesarini et al [8] for the view that preferences for giving and risk taking are broadly heritable and Simonson&Sela [41] for a heritable effect on preferences for certain products). Since we do not observe preferences such possible correlation cannot be estimated directly. One possible approach to overcome this problem is to analyze the correlation between the consumption vectors of parents and their children. A correlation between preferences should, controlling for prices and income, be associated with a correlation on consumption vectors, i.e., parental transmission of preferences should imply an intergenerational link in consumption patterns.

Unfortunately, and contrary to the cases of transmission of earnings or abilities, except for some very limited surveys on consumption of specific

---

[3]In some countries, such as the UK, there are other comprehensive sources available, e.g. national census and electoral registers. However, in most countries the access for researchers to these data is restricted due to privacy concerns.

goods, there are no good data available on the consumption vector of parents and their (adult) offspring to check for such possible intergenerational link in consumption. Therefore, we follow an indirect approach in order to detect the existence of parental transmission. This approach is based on assessing how surnames and consumption patterns are distributed across different geographical regions of the country using a measure of geographical dissimilarity. The key idea is that surnames remain unaltered when transmitted vertically across generations, while consumption patterns may be determined by environmental and/or parental transmission.

An example might help to clarify the central idea of this paper. Consider two regions, $A$ and $B$, and let $w_I$ be the mean consumption vector and $x_I$ the vector of surname frequencies of agents in region $I$ ($I = A$, $B$). Let $d(\cdot)$ be a distance measure in the space of consumption patterns and surname frequencies. Assume that at a certain period in time regions $A$ and $B$ do not share any surname in common, so that if the $j^{th}$ element of $x_A$ is zero the same element of $x_B$ is not zero and vice versa.[4] Suppose that individual preferences are very different across regions, and therefore, consumption patterns are also very different.[5] These two assumptions imply that both $d(w_A, w_B)$ and $d(y_A, y_B)$ are large. Suppose that a representative agent moves from $A$ to $B$ and has a child born in $B$. In the next period the old agent dies and his child becomes an adult and stays in $B$. What are the implications of this migration in terms of $d(w_A, w_B)$ and $d(x_A, x_B)$? The distance in surnames, $d(x_A, x_B)$ will decrease since the child bears a surname that did not exist in region $B$ before. However, the distance in consumption $d(w_A, w_B)$ will decrease or remain constant depending on the importance of parental transmission. Under full parental transmission the child of the immigrant consumes the same vector as his father and this implies that $d(w_A, w_B)$ will also decline. This reasoning shows that, under full parental transmission, migration implies a decrease in surname and consumption distances[6]. Therefore, regions that are close in terms of surnames should also

---

[4]This is an extreme assumption, it is only required that the distribution of surnames differs substantially among regions.

[5]We are implicitly assuming that prices are the same in both regions and all agents have the same income.

[6]The reasoning above is just an example. In principle, migration between two regions with different distributions of surnames does not always reduce the distance in surnames. For example, suppose that in period 1 surname "x" only exists in region A and in period 2 all the people with such surname migrate to region B and nobody else migrates. In this case the distance in surnames between the two regions might remain the same in both periods. To rule out that possibility we need to assume that migration is not *surname biased* i.e., that there is no selection by surnames. Using the same reasoning we also have to assume that migration is not *consumption biased*

be close in terms of consumption patterns that implies a positive correlation between consumption distances and surname distances. In the event that parental transmission plays no role and the child of the immigrant acquires the preferences of region $B$, the distance $d(w_A, w_B)$ would remain constant. However, the children of the natives might also acquire the preferences of the immigrants and, in this case, the distance $d(w_A, w_B)$ would decrease. Therefore, horizontal transmission might imply both zero or positive correlation between surnames and consumption distances. Thus, a positive correlation between distances in surnames and distances in consumption is a *necessary but not sufficient* condition for parental transmission.

We will construct a matrix of surname distances and a matrix of consumption distances between the regions of mainland Spain. Based on the idea above, we claim that parental transmission should be reflected in a positive and significant correlation between these two matrices.

# 3 Data on Surnames and Consumption

## 3.1 Surname Distribution in Spain

We consider the geographical surname distribution in Spain in 1999 extracted from the telephone directory. The directory can be purchased in a CD-ROM distributed for commercial purposes (INFOBEL, http://www.infobel.com). The directory contains 11.5 million domestic users and provides information on the full name and address of the subscriber, including the address and the zip code.[7] The total population of Spain in 1999 was around 40 million and the total number of main family residences was around 14 million. As mentioned in the introduction, the number of foreign immigrants in Spain until the late 90's was very low compared to other countries. Even in 1999, the number of foreign residents was 1.4% of the total population compared to 6.1% in France and 8.8% in Germany (see OECD [35]). Furthermore, because there have been no significant foreign migration inflows in the modern history of Spain the surnames analyzed here are largely of Spanish ancestry.

The naming convention in Spain is different from most Western countries, and similar to that prevailing in some Latin-American countries. The main distinctive feature is that the family name is formed by two surnames, the first being the father's first surname and the second the mother's first surname.[8] A second important feature is that married women keep their

---

[7]To the best of our knowledge this is the first paper using the information on the surnames of all the telephone users in Spain and not just a sample of them.

[8]There have been some changes in naming conventions during the past decades. The

maiden surname and do not adopt the husband's surname. This convention was legally established after the introduction of the Civil Registry in 1870 but it had been followed by a very large majority of the population since much earlier (see Salazar-Acha[40] and Fauré et al [20] for a History of surnames in Spain). Thus, for at least four or five generations the first surname of a person is the first surname of his/her father.[9] This is important because it allows us to conclude that population movements from one region to another permanently reduce the surname differences between the two regions.

Even though the telephone directory is possibly the best source available it has some problems: first, there may be duplicity of some surnames as an individual may have several telephone lines in different residences. Second, there is a potential bias towards people living in urban areas. Finally, some surnames have alternative spellings (e.g. Giménez and Jiménez) and it is difficult to decide which surnames are variant spellings and which are different surnames. Most variant spellings diverged many generations ago and thus we treat different spellings as different genealogical lines. Furthermore, the computer file on the CD-ROM records the paternal and maternal surname in a single field and in some cases it is difficult to disentangle the first from the second, particularly when compound names are involved. To overcome this problem, we use a computer program that implements an algorithm to separate the first (paternal) and the second (maternal) surname as accurately as possible[10] and we use only the first surname in our analysis. However, all the results in the paper remain unchanged if we either use the second surname instead of the first one, or if we pool together the first and second surnames.

Spain is divided into 52 districts called provinces. This geographical division was formally set in 1833 and in most cases followed the dominions of Medieval kingdoms, principalities and bishoprics, thus grouping people with a historically common institutional linkage. In this paper we only consider the 47 provinces in the Iberian Peninsula, leaving out the territories of the Canary archipelago, the Balearic Islands, and the two autonomous cities in the North of Africa, Ceuta and Melilla. The telephone directory provides the zip code of each user that is a much smaller area than the province. Even

---

law now allows for changes in the order of the surnames. This practice, however, is rather unusual and responds to personal motivations, for instance preserving the mother's surname in the next-to-present generation.

[9]There is one exception to this general rule. The law allows the combination of the two surnames of a person into a new one to be passed on to their offspring. However, this possible merger of surnames is not immediate and the total number of such double-barrelled surnames is very low (see Collado et al [13]).

[10]The program was coded in Q-BASIC and is available upon request. Our algorithm drops 3.1 percent of the total population. This includes foreign residents who have just one surname.

though a smaller geographical division leads to a bigger sample size, we will consider the province as the geographical unit of reference since the other data sources that we need are available only at provincial level.

Provinces are of similar geographical extension, though the population varies widely, ranging from Madrid with around 5.5 million inhabitants to Soria with just 95,000. Provinces are quite heterogeneous in population density and urban concentration, what made us concerned that the telephone directory might be over sampling predominantly urban provinces compared to those with dispersed rural populations. However, for all provinces, the number of individuals included in the telephone directory corresponds to approximately 30% of the population of each province[11] so we observe no particular bias towards under-sampling in rural provinces.

We found 132,882 different surnames in the whole country. The most common surname is García (3.5% of the population) and there are almost 60,000 surnames borne by just a single person. The 40 most common surnames cover 32.7 percent of the population, a figure higher than the 13.1 percent reported by Clark [11] for England and Wales. We found that the number of family names of the same size (the size of a family name is number of people bearing this surname) follows approximately a power law distribution with parameter -1.7, that is in accordance with the empirical findings for other countries and the theoretical predictions reported in Derrida, Manrubia and Zanette [15] and [16]. The most common surnames are German and Latin patronymics (Fernandez, Rodriguez, Martinez,...), geographic names (Navarro, Serrano,...), and names from personal characteristics (Blanco, Delgado,...). According to Rodriguez-Larralde et al [38] occupational names (Molina, Herrero,...) occur less frequent than in other European populations, and Arab surnames (Medina, Alcantara,...) are only relatively frequent in Andalusia.

## 3.2 Surname Distances

Our goal here is to construct a matrix of surnames distances among the Spanish provinces using the data from the telephone book described in the previous Section. We denote by $x_i$ the relative (national) frequency of surname $i$ and by $x$ the vector of such (national) frequencies. Similarly, we denote by $x_i^j$ the relative frequency of surname $i$ in province $j$ and by $x^j$ the vector of such frequencies in province $j$. We construct a *matrix of surname distances* between the provinces that is denoted by $D$.[12] The $(j, k)$ element

---

[11]The $R^2$ of the OLS regression is 0.98.

[12]This matrix and all the other matrices calculated in the paper, as well as the software programs used for the computation of correlation tests, are available from the authors

of the $D$ matrix represents the distance between province $j$ and province $k$ and is given by

$$d_{jk} = \left( \sum_{i=1}^{N} \left| x_i^j - x_i^k \right|^p \right)^{\frac{1}{p}} \tag{1}$$

where $N$ is the total number of different surnames in Spain (i.e., $N = 132,882$) and $p \geq 1$. We choose the Manhattan distance,[13] i.e. $p$ equal to 1, because is the only one among the distances defined in (1) satisfying the following desirable "anonymity" property: Consider two provinces A and B and assume migration from A to B. If this movement of population contributes to reduce (increase) the surname distance between the two provinces, the reduction (increase) in the distance is the same regardless of the surnames borne by the migrants.

Spain has experienced low immigration rates until very recently, therefore, matrix $D$ can be thought to contain aggregate information on the amount of interior migration flows (between provinces) over the last few centuries. Using this matrix we find that the average distance between provinces is 1.22. Notice the maximum value the Manhattan distance can take is 2, therefore, this result indicates that there are substantial differences in the surname distribution across provinces. The largest distance is 1.67 between Lugo and Huesca, two provinces far apart that have had very different historical ties, and the shortest distance is 0.52 between Seville and Cádiz that are next to each other and are considered very close from a historic and sociological point of view. As expected, closer provinces tend to have more similar surnames and the correlation between geographical and surname distances between provinces is 0.44. Despite this high average distance between provinces, the Spanish population is relatively uniform when compared to other European populations and, for example, the most common surnames in the country are also the most common surnames in most provinces (see Rodriguez-Larralde [38] ). We calculate the "center of gravity", i.e., the province that minimizes the distance to all the other provinces weighted by the population. We find that the primary center of gravity lies with Madrid, followed by Barcelona, i.e., the two major provinces in terms of population.

---

upon request.

[13]We have checked that all the results in this paper are robust to choice of $p$. Namely, we have repeated all our calculations for $p = 2, 5, 10, 100$ and the limiting case when $p \rightarrow \infty$, obtaining very similar results.

## 3.3   Consumption Expenditure

In this section, we construct an aggregated consumption vector for each province that will be used to compute a matrix of " preference distances" between provinces. Preferences are unobservable but consumption is not. Individuals choose their consumption profiles by maximizing their preferences subject to their corresponding budget restrictions. One option would be to use standard econometric regression methods to estimate the unknown parameters of the Engel curves that correspond to the unknown parameters of the individual preferences. However, we think that to pick up parental transmission of preferences it is very important to use a highly disaggregated classification of goods. Therefore, it would be unfeasible to estimate an Engel curve for each good considered in our analysis. For this reason, we use the raw vector of consumption shares to define a distance matrix. We are aware that variability in these shares responds not only to differences in preferences across provinces but also to heterogeneity in real prices, income and other factors. We control for these factors in the empirical analysis.

We calculate budget shares for each province using the 1990/91 consumer expenditure survey (Encuesta de Presupuestos Familiares, EPF). The EPFs are large surveys conducted every ten years by the Spanish statistics office (Instituto Nacional de Estadística, INE). These surveys use a representative sample of the Spanish population by province, providing very detailed information on household expenditure. They are used by officials to calculate consumption weights in the Consumer Price Index (IPC). In 1997 the EPF was replaced by the continuous consumer expenditure survey (Encuesta Continua de Presupuestos Familiares, ECPF) that uses a smaller sampling design and it is not representative at the provincial level. Thus, the EPF 1990/91 was the last large survey that is representative at provincial level, with a sample size of 21,155 households. We consider one hundred and ten composite goods that correspond to the subclasses defined by the Spanish Statistics Office. The description of the subclasses is listed in the Appendix.

We calculate the *budget share* of good $i$ in province $j$ as

$$w_i^j = \frac{\sum\limits_{h \,\in\, \text{province } j} c_{ih}}{\sum\limits_{h \,\in\, \text{province } j} c_h} \tag{2}$$

where $c_{ih}$ is the amount expended on good $i$ by household $h$, and $c_h$ is total expenditure by household $h$. Let $w^j$ denote the vector of budget shares for province $j$.

As we did in the case of surname frequencies, we now define the distance in consumption shares between province $j$ and province $k$ as

$$s_{jk} \;\; = \;\; \sum_{i=1}^{110} \left| w_i^j - w_i^k \right| \qquad (3)$$

The matrix of preference distances between provinces is denoted by $S$ and contains the distance $s_{jk}$ as the $(j,k)$ element.

Before proceeding, one might claim that the matrix $S$ is not capturing the differences in preferences between populations in the provinces for at least two reasons:

1. Our vector of budget shares, $w_i^j$, is an aggregated measure and some relevant information may be shaded by the way we aggregate.

2. As already claimed, consumption shares do not depend only on preferences but also on prices, income and other types of differences, such as weather, or the proportion of urban/rural population.

Regarding the aggregation problem, there is not much we can do since aggregating always implies a loss of information. Still, one could claim that the specific way used to compute the vector $w_i^j$ is not satisfactory and averaging over individual budget shares might be more desirable. This alternative, however, was not adopted due to the problem of infrequency of purchases of durable goods. In any case, we compute the average budget shares for each province using only food expenditures[14] to avoid the durable goods problem, and the correlation between the corresponding distance matrix and the $S$ matrix calculated with food goods is 0.966. Hence, both matrices contain essentially the same information and using one or the other would not affect our results.

In principle, the second problem raised could be solved by incorporating the necessary information on those variables. However, this is not an easy task since some of the required information, particularly about prices, is not available so that alternative and feasible approaches to reduce the severity of the problem have to be pursued. In our case, the following strategy is adopted:

- To reduce the problem of consumers facing different prices in different provinces we drop expenditures on housing from our analysis. Real estate prices vary widely across provinces and represent a very large component of the budget. Thus, we eliminate seven subclasses[15] from

---

[14]Food expenditures are divided into thirty-three subclasses ranging from subclass 110-A to 120-A (see Table 3).

[15]Subclasses 310-A to 320-B (see Table 3). Our main results do not depend on the exclusion of these subclasses.

the calculations, leaving 103 consumption items. Furthermore, it is expected that the law of one price applies for tradable goods, while significant price differences between regions could appear for non-tradable goods. For this reason we perform a robustness check distinguishing between tradable and non-tradable goods. We also control for the distance between provinces as a proxy for transportation costs.

- We control for differences in income across provinces as well as other factors such as the proportion of urban/rural population and differences in climate and household composition.

Thus, our matrix $S$ contains information on the distances between provinces on the aggregated vector of relative expenditures for 103 different consumption goods. Using this matrix we find that the average distance between provinces is 0.28, the largest distance is 0.46 and the shortest is 0.14. Since, as we mention above, the maximum value the Manhattan distance can take is 2, the differences in the distribution of consumption patterns across provinces are economically important. We interpret those distances as representing the consumption preference distances between provinces.

# 4 Main Results

## 4.1 A First Test

In this Section we test whether there is a positive correlation between the distance in terms of consumption shares and in terms of surname frequencies. We indeed find a positive correlation that suggests the existence of a significant parental transmission effect.

We ask the following question: Is there a statistically significant positive correlation between the matrix $D$ and the matrix $S$? This seems a priori an easy question to answer. There is, however, a " technical" problem: Since the elements of a distance matrix are not independent[16] we cannot use standard methods of least square estimation. To overcome this problem we use the Mantel test that is specially designed for testing linear correlation between distance matrices. The Mantel test is a non-parametric randomization procedure that can be used to test any linear relationship but is especially useful in the case of non independent data points. The Mantel's test statistic is the correlation coefficient, $r$, of the distance matrices $D$ and $S$, and its value range is $[-1, 1]$. The significance of the correlation is evaluated via random permutation of the rows and corresponding columns of $D$ and $S$. For

---

[16]This is due to the triangle inequality property.

each random permutation, the correlation $r$ is computed. After a sufficient number of iterations,[17] the distribution of values of $r$ is generated and the critical value of the test at the chosen level of significance is found from this distribution.

In our case, we find that the correlation coefficient between matrices $D$ and $S$ is 0.4198 and the hypothesis of non-positive correlation is strongly rejected on the basis of a Mantel test with $10,000$ replications (p-value zero). Tables 1 and 2 in the Appendix contain this and subsequent results.[18] Thus, we conclude that *closer provinces in terms of surnames also tend to be closer in terms of preferences.*

One might say that this result is not surprising at all and it only detects that provinces that have had much mixing one with another have similar preferences and similar surnames. Though this effect is clear for surnames, it is not obvious that population mixing should lead to more similar preferences as this depends on the attitudes of the newcomers (or locals) with respect to their new environment and their willingness to assimilate the values of the host region. Furthermore, it is even less obvious that provinces with similar preferences should have experienced intense population mixing. It might well be the case that two populations have similar preferences as a consequence of the spread of certain cultural and social views through channels that do not require population mixing.[19] Thus, we should control for other factors that may have some explanatory power in the correlation found between matrix $D$ and matrix $S$.

## 4.2   Controlling for Other Variables

So far, we have considered a simple correlation between surname distances and preference distances. In this Section we show that such correlation also holds after controlling for some relevant variables. This result is considered as a check of the robustness of the significance of the parental transmission effect on consumption behavior.

As we mention above, it is clear that the correlation between matrix $D$ and matrix $S$ might depend on a number of different facts that include:

- Geographical distance between provinces

---

[17]The number of iterations for all the runs in this article was 10,000.

[18]All our results are robust to the exclusion of the most common surnames or the very rare ones.

[19]This is somewhat similar to an old question in Population Genetics about the demic versus the cultural transmission of technological changes. See Cavalli-Sforza et al. [6] and Jobling et al [27]

- How urban/rural provinces are

- Income differences

- Climate

- Household composition

**Geographical distance:** The geographical distance between provinces accounts for transportation costs and could also be very closely correlated with other variables, for example the type of local agricultural produce available, that might explain some of the differences in consumption. Thus, we will define the matrix of geographical distances between provinces, $G$, where the element $g_{jk}$ indicates the distance, in kilometers, between the capital of province $j$ and the capital of province $k$. Since provinces are relatively small in area and in most cases the majority of the population is concentrated around the provincial capital, this distance is a good index of the geographical distance between the whole populations in the provinces.

**Urban/Rural:** It is natural to assume that consumption shares in urban environments might differ from consumption shares in rural areas, even if household preferences were identical, as urban and rural households face different prices and the set of available goods may be different. In order to control for this, we classify municipalities into eight groups[20] and assign each household in the EPF sample to the corresponding group. We denote by $u_i^j$ the percentage of households in group-size $i$ in province $j$. Then the "urban" distance between province $j$ and province $k$ is given by

$$u_{jk} = \sum_{i=1}^{8} \left| u_i^j - u_i^k \right|$$

and the corresponding matrix of distances is denoted by $U$.

**Income:** The EPF contains information on the total income reported by household members. In principle, this information could be used to control for the income effect. However, since this variable is provided by the survey respondent herself, it is natural to suspect of under-reporting. For this reason we decided to use total expenditure instead. Total expenditure is

---

[20] Groups are defined in terms of population intervals: 2,000 or less, from 2,000 to 5,000, from 5,000 to 10,000, from 10,000 to 20,000, from 20,000 to 50,000, from 50,000 to 100,000, from 100,000 to 500,000 and more than 500,000.

highly correlated with household income and it does not suffer from significant measurement error. Thus, denoting by $m^j$ the mean household total expenditures in province $j$, the "income" distance between province $j$ and province $k$ is given by $m_{jk} = |m^j - m^k|$ and the corresponding matrix is denoted by $M$.

**Climate.** The weather in Spain varies greatly across regions. In general, the North is cold and rainy and the South warm and dry. Because of these differences, people with the same preferences might need to consume different goods depending on their province of residence. To control for this a matrix of "climate distances", $T$, is defined in the following way: for each province $j$ we compute the vector $t^j = (t_1^j, ..., t_{12}^j)$, where the $t_i^j$ element indicates the average temperature [21] during month $i$. The climate distance between province $j$ and $k$ is given by

$$t_{jk} = \sum_{i=1}^{12} \left| t_i^j - t_i^k \right|$$

and the matrix $T$ is formed with these elements.

**Household composition.** Finally, household composition may be a major explanatory factor in determining differences in consumption patterns. Rural provinces contain a higher proportion of elderly, retired people, whose consumption pattern differs substantially from that of a middle-aged family with children, for instance. The EPF also contains information on household composition in the form of a categorical variable for the fourteen different types of households. In Table 4 in the Appendix we describe the different type of household defined in the EPF. We compute the vector $h^j$ for province $j$ containing the proportion of households of each type. Thus, matrix $H$ denotes the distances in terms of household composition.

Summing up, we have the matrices $S$, $D$, $G$, $U$, $M$, $T$ and $H$, the distance matrices of, respectively, consumption shares, surnames, geographical distances, proportion of urban/rural population, income per capita, climate and household composition. The question now is how to extend the bivariate Mantel test to our context of multiple control variables. Smouse et al. [42] propose the following three-step technique:

- OLS estimation of $D$ on $G$, $U$, $M$, $T$ and $H$

---

[21]These temperatures are averages for 1997-2002. These data are available from the website of the Spanish Statistics Office, http://www.ine.es/.

- OLS estimation of $S$ on $G$, $U$, $M$, $T$ and $H$

- Bivariate Mantel test using the residuals of the previous two regressions.

Therefore, we perform a multivariate Mantel test to determine the significance of the correlation coefficient of the $D$ and $S$ matrices, controlling for $G$, $U$, $M$, $T$ and $H$. The correlation is now 0.2277 and is significantly greater than zero as the p-value is 0.0042. Thus, after controlling for how close provinces are in income, urban/rural environment, geographical distance, climate and household composition, we still find that provinces that are similar in the frequencies of surnames tend to be similar in their consumption preferences.

## 4.3    Different Groups of Consumption Goods.

We now look at differences in the correlation depending on the type of consumption good considered. Parental transmission might play a more important role in, for example, the formation of preferences over food than in the formation of other types of preferences. Or, equivalently, for some types of consumption good such as food, the offspring of immigrants can be less prompted to adopt the habits of the host province than for other consumption goods. This new exercise will also help to analyze whether the positive correlation between surname and consumption distances is due to what we call a ghetto grouping effect. If immigrants tend to concentrate into some particular areas, newcomers will be living in the same environment as in their province of origin. In this case, the preferences of the parent would coincide with those of the environment, and therefore, we should find a large correlation coefficient for all consumption goods both under parental and environmental transmission.

We create two new distance matrices in consumption shares. The first matrix, $S_f$, includes exclusively all food goods that are the first thirty–three items on the list used to compute $S$. [22] The second matrix, $S_{nf}$, contains the remaining seventy items. Then we repeat our previous multivariate Mantel test twice, firstly replacing matrix $S$ by matrix $S_f$ and secondly replacing $S$ by $S_{nf}$. The results of these two tests indicate a striking difference between food and non-food cases. The test for food items (matrix $S_f$) shows a significant correlation coefficient of 0.3735 (p-value 0). The correlation coefficient when the matrix with non-food items, $S_{nf}$, is used is 0.0465 and non-significant (p-value 0.3190).

---

[22]Subclasses 110-A to 120-A in Table 3.

The non-significance of this coefficient indicates that newcomers assimilate the preferences of locals in non-food consumption. This fact is difficult to interpret under the assumption of "ghetto grouping" and no parental transmission. Thus, our findings on the correlation coefficients are more consistent with the following statement: *There is full environmental transmission on preferences over the non-food consumption goods, and therefore perfect integration of immigrants to their host environment. Regarding preferences over food items the result show lack of integration and suggests, but cannot unambiguously conclude, the existence of a strong parental transmission effect.*[23]

This possible parental transmission for food preferences is related to the results provided in Waldkirch et al [46]. It is important to notice that these authors provide a result on possible parental transmission for the total amount spent on food. Our analysis, however, is much more disaggregated and focuses on the expenditure shares of thirty–three different food items. Moreover, after controlling for the State of residence of parents and their offspring they obtain no correlation in the total amount spent on food. Interestingly, we have checked and found that a similar result holds in our case, i.e. after controlling for income and geographical distance the total amount spent on food is not related to surname distances. Thus, our findings also suggest that parental transmission of preferences affects the shares of the different food products rather than the total amount spent on food.

Three additional comments are called for here. First, it may be that our test does not detect parental transmission for non-food goods because the classification of these goods is less disaggregated than the classification of food goods. For example, it might be the case that children have the same preferences as parents for going to the theatre instead of to the cinema. However, both theatre and cinema expenditure belong to the same category and therefore the transmission of this sort of preference cannot be detected with our data. We have repeated all the previous tests using the maximum level of disaggregation available in the EPF (246 food items and 523 non-food items) and the results remain unchanged.[24] **S**econd, our result could be due to the fact that preferences for non-food goods are basically the same across provinces. Our data, however, do not support this view since the non-food consumption patterns are different across the Spanish provinces even after controlling for income, geographical distance, household composition and climate (the unexplained variability of the regression is 79.2%). Third, following

---

[23]This ambiguity derives from the discussion in Section 1.2 where we have shown that a positive correlation between surnames and consumption distances is a necessary but not sufficient condition for parental transmission.

[24]The results of these tests and the detailed specifications are available upon request. The disaggregated list of EPF goods and services can be found in http://www.ine.es/

the approach in Stigler and Becker[45] an alternative explanation of our result for food goods is that we are not detecting transmission of preferences but *transmission of skills* in the household production of food related goods. Cooking is a production process that is learnt within the family environment and children and young adults might easily acquire their parents' cooking skills or "human capital". Since parents and their offspring share similar cooking skills it is natural to expect them to buy similar "cooking inputs" as well, i.e. similar food, and this fact is reflected in a similar composition of their food consumption shares. Although we cannot perfectly discriminate between this "skill transmission" hypothesis and our "preference transmission" hypothesis, we tried to shed some light on this issue by dividing food items in two groups and running separate regressions. The first group includes food items that are clearly cooking inputs, like salt, sugar, vinegar, rice, flour, eggs, butter, different types of oil, of milk, of pasta, of meat, of fish, of fresh vegetables, etc. The second group includes food items that are typically eaten without being further processed, like coffee, tea, chocolate, ice creams, different types of fresh fruits, of bread, of pastries, of ham and sausages, of cheese, etc.[25] We find a positive and significant correlation coefficient for both groups (see columns 1 and 2 in Table 2). The correlation coefficients are 0.3168 (p-value 0) for the first group and 0.3698 (p-value 0) for the second. This result does not rule out the "skill transmission" hypothesis but shows that it is not the unique channel. Finally, the main caveat of our approach is that we cannot perfectly control for differences in prices across provinces. We control for how urban/rural regions are which partially explains differences in prices. We also control for geographical distance that accounts for transportation costs for tradable goods. However, for non-tradable goods there might be differences in prices between provinces that we have not accounted for, and these differences in prices could be driving our results for non-food goods because, although we have exclude housing from the analysis, non-food goods include some non-tradable services that could have different prices in different provinces. To check whether our results for non-food goods are driven by differences in prices, we have calculated our correlations for non-food goods using only tradable goods (see column 3 in Table 2). Tradable non-food goods include books, magazines, towels, blankets, carpets, cutlery, jewelry, computers, different types of clothes, of furniture, of small appliances, of cleaning products, of toiletries, of vehicles, of toys, etc.[26] The correlation coefficient for tradable goods is 0.1389 (p-value 0.1), a bit larger than for all non-food items but still non-significant.

---

[25]The grouping is based on the detailed classification that considers 246 food items. The detailed list of goods on each group is available upon request.

[26]The classification into tradable and non-tradable is based on the detailed classification

This result confirms that the absence of correlation for non-food items is not driven by differences in prices across provinces.

## 4.4   Controlling for Recent Migration

In this Section we show that controlling for recent migration flows does not alter the finding of the previous Section. Thus, the possible importance of parental transmission only for food consumption remains true even when we leave out recent migrants.

One might claim that immigrants' adaptation to their host environment takes place during the second generation (see Borja [3]). In this case, since there were some significant migration flows in Spain in the 60's and 70's, our previous conclusion relating a significant correlation coefficient as a necessary condition for a parental transmission effect might be misleading. To see this, suppose that the first generation of adult immigrants keep their original preferences. Their offspring, however, adopt the preferences of the host province, i.e. environmental transmission is the only factor in the preference formation process. However, the matrix of surname distances and the matrix of consumption shares could be correlated because of the recent migration, leading us wrongly to deduce the possibility of a significant parental transmission.

To control for this possibility we should disentangle the contribution of recent migration to the consumption and surname vectors and this requires information on the net migration flows between Spanish provinces during the last generation. Unfortunately, this information is not available. However, the Spanish labor force survey (Encuesta de la Poblacion Activa, EPA) also conducted by the INE, contains information on the current province of residence and the birthplace of individuals. The EPA is a quarterly survey and the sample size is around 190,000 individuals. Households are interviewed for six consecutive quarters. In this study we merge two waves of the EPA: First quarter 1999 and third quarter 2000.[27] Thus, using the data from the EPA, we construct a matrix, $E$, of "migration" distances in the following way: we associate the vector $b^j = (b^j_1, ..., b^j_{47})$ to province $j$ such that the $b^j_i$ element contains the percentage of people living in province $j$ who were born in province $i$. We construct the matrix $E$ in a similar way to the previous matrices. The element $e_{jk}$ is the distance between province $j$ and $k$, i.e. $e_{jk} = \sum_{i=1}^{47} \mid b^j_i - b^k_i \mid$ . Since geographical mobility over the last 15 years in Spain has been very low we are confident that a large majority of the people born in province $i$ and currently living in province $j$ did not live

---

that considers 523 non-food items. The detailed list of tradable goods is available upon request.

[27]We use these two waves to avoid duplicity of households.

in another province before migrating to $j$. Thus, our matrix $E$ may be seen as a good approximation to the matrix of recent migration flows between Spanish provinces.

We repeat our previous multivariate Mantel test now including the additional matrix $E$. More precisely, we test the significance of the correlation coefficient between matrix $S$ and matrix $D$, controlling for $G$, $U$, $M$, $T$, $H$ and $E$. We obtain that the correlation coefficient is 0.1669 (see Table 1), lower than before but still large and significant (p-value 0.0359). We also carry out the test when consumption goods are divided into food (matrix $S_f$) and non-food groups (matrix $S_{nf}$) as in the previous subsection. It is reassuring to observe that the correlation coefficient when only food is considered is still high (0.3154) and significant (p-value 0.0002), and the coefficient for the non-food case remains low (0.0180) and clearly non-significant (p-value 0.4191), indicating that for non-food goods parental transmission plays no role in the preference formation process.

We were concerned that the significance of our results were due to a small number of "outlier" provinces. Hence, we have checked the robustness of our test by sequentially dropping groups of three provinces. Then, we run the test for all possible subsets of provinces (16.215) leaving out one triplet in each run. The results shows that when we consider all goods we find a non-significant correlation in about one third of the runs. In the case of food consumption the correlation coefficient in all the runs is positive and significant at the 5% level. For non-food items the correlation coefficient is not significantly greater than zero at the 5% level in any of the runs. Hence, we are confident that our results are not driven by a small number of extreme observations.

We also checked that the results presented in the previous section regarding cooking inputs and tradable non food goods remain valid after controlling for recent immigration (see Table 2).

# 5    Final Comments and Further Work

We have developed a novel indirect approach to analyze the existence of intergenerational linkages in consumption preferences. This new approach is based on the analysis of the correlation between the geographical distributions of surnames and consumption choices. It is important to stress here that the use of surnames, though novel in this field, is not indispensable to our method: the spatial distribution of any other individual characteristic that is perfectly transmitted from parents to offspring could play the same

22

role. However, surnames fit perfectly for this purpose and by using the spatial distribution of surnames obtained from the telephone directory we have shown that only for food items there might be an intergenerational linkage.

Information on geographical distribution of surnames is becoming easily available for many other countries where there are also good surveys of consumption behavior. Applying our methodology to a second country would be an important exercise in investigating the robustness of our results and making international comparisons.

# References

[1] Becker, G.S. and C.B. Mulligan (1997), "The Endogenous Determination of Time Preference," *Quarterly Journal of Economics*, **112**, 729-758.

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

[3] Borjas, G. (1994), "Long-run Convergence of Ethnic Skill Differentials: the Children and Grandchildren of the Great Migrations", *Industrial & Labor Relations Review*, **47**, 553-573.

[4] Browning, Martin and M. Dolores Collado (2001), "The Response of Expenditures to Anticipated Income Changes: Panel Data Estimates" , *American Economic Review*, **91**, 681-92.

[5] Cavalli-Sforza, L.L. and M.W. Feldman (1981), *Cultural Transmission and Evolution: A Quantitative Approach*, Princeton: Princeton University Press.

[6] Cavalli-Sforza, L.L., Menozzi P.and A. Piazza (1994), *The History and Geography of Humans Genes*. Princeton: Princeton University Press.

[7] Carreras, A. and X. Tafunell, (2005), *Estadísticas históricas de España*. Bilbao: Fundación BBVA, eds.

[8] Cesarini, D., Dawes, C.T., Johannesson, M., Lichtenstein, P. and B. Wallace, (2009), "Genetic Variation in Preferences for Giving and Risk Taking," *Quarterly Journal of Economics*, May, 809–42.

[9] Charles, K.K. and E. Hurst (2003), "The Correlation of Wealth across Generations." *Journal of Political Economy*, **111**(6):1155-82.

[10] Cipriani, M., Giuliani, P. and O. Jeanne (2007), "Like Mother Like Son? Experimental Evidence on the Transmission of Values from Parents to Children," *IZA Discussion Paper* n. 2768.

[11] Clark, Gregory (2009), "The indicted and the wealthy: Surnames, reproductive success, genetic selection and social class in pre-industrial England", mimeo.

[12] Clark, Gregory (2010), "Regression to mediocrity? Surnames and social mobility in England, 1200-2009," mimeo.

[13] Collado, M.D., Ortuño-Ortin, I. and A. Romeu (2008): "Surnames and Social Status in Spain," *Investigaciones Economicas* **32**(3), 259-287.

[14] Daniels. M., Devlin, B. and K. Roeder (1977), "Of Genes and IQ" in B. Devlin, Fienberg, S., Resnick, D. and K. Roeder (eds.) *Intelligence, Genes, and Success.* New York: Springer-Verlag, 45-70.

[15] Derrida, B., Manrubia, S.C. and D.H. Zanette (2000), "On the Genealogy of a Population of Biparental Individuals"*Journal of Theoretical Biology*, **203**, 303-315

[16] Derrida, B., Manrubia, S.C. and D.H. Zanette (1999), "Statistical Properties of Genealogical Trees"*Physical Review Letters*, **82**, 1987-1990.

[17] Desmet, K., Le Breton, M., Ortuno-Ortin, I. and S. Weber (2009), "The Stability and Breakup of Nations: A Quantitative Analysis," *Journal of Economic Growth,* **16**, 183-213.

[18] Dohmen, T., Falk, A., Huffman, D. and U. Sunde (2011), "The Intergenerational Transmission of Risk and Trust Attitudes", *Review of Economic Studies*, forthcoming.

[19] Einav, L. and L. Yariv (2006), "What's in a Surname? The Effects of Surname Initials on Academic Success" , *Journal of Economic Perspectives*, **20**, 175-188.

[20] Fauré, R., Ribes, M. and A. García (2001), *Diccionario de Apellidos Españoles*, Espasa, Madrid.

[21] Feldman, M.W., Otto, S. P. and F.B. Christiansen (2000), "Genes, Culture, and Inequality."in K. Arrow, S. Bowles and S. Durlauf (eds.) *Meritocracy and Economic Inequality.* New Jersey: Princeton University Press. 61-85.

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

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

[24] Harris, J.R. (1998), *The Nurture Assumption: Why Children Turn Out the Way They Do.* The Free Press.

[25] Holloway, S.M. and J.A. Sofaer (1992), " Coefficients of Relationship by Isonomy among Registrations for Five Common Cancers in Scottish Males." *Journal of Epidemiology and Community Health*, **46**, 368-372.

[26] Jennings, M.K., L. Stoker and J. Bowers (2001), *Politics Across Generations: Family Transmission Reexamined.* Institute of Governmental Studies, University of California, Berkeley.

[27] Jobling, M.A., Hurles, M.E. and C. Tyler-Smith (2004), *Human Evolutionary Genetics. Origins, People & Disease.* New York: Garland Science.

[28] Jobling, M.A. (2001), "In the Name of the Father: Surnames and Genetics." *Trends in Genetics*, **17**(6): 353-57.

[29] Lasker, G.W. (1985), *Surnames and Genetic Structure*, Cambridge University Press.

[30] Legendre P. and L. Legendre (1998), *Numerical Ecology*, Elsevier, New York.

[31] Mantel N. (1967), "The Detection of Disease Clustering and a Generalized Approach," *Cancer Research*, **27**, 209-220.

[32] Mazumder, B. (2005), "Fortunate Sons: New Estimates of Intergenerational Mobility in the United States Using Social Security Earnings Data," *The Review of Economics and Statistics*, **87**,2, 235-255.

[33] Mulligan, C. (1997), *Parental Priorities and Economic Inequality.* Chicago: University of Chicago Press.

[34] Mulligan, C. (1999), "Galton vs. Human Capital Approaches to inheritance" *Journal of Political Economy*, **107**(6), 184-224.

[35] OECD (2002), *Trends in International Migration*, Publications Service, Paris.

[36] Pollak, R.A. (1970), "Habit Formation and Dynamic Demand Function" *Journal of Political Economy*, **78**, 745-63.

[37] Pollak, R.A. (1978), "Endogenous Tastes in Demand and Welfare Analysis" *American Economic Review*, 68(2), 374-379.

[38] Rodriguez-Larralde, A., A. Gonzales-Marin, C. Scapoli, and I. Barrai (2003), "The Names of Spain: A Study of the Isonymy Structure of Spain," *American Journal of Physical Anthropology*, **121**, 280-292.

[39] Rowe, D. (1994), *The Limits of Family Influence: Genes, Experience and Behavior*. New York: The Guilford Press.

[40] Salazar-Acha, J. (1991), *Génesis y Evolución Histórica del Apellido en España*, Real Academia Matritense de Heráldica y Genealogía, Madrid.

[41] Simonson, I. and A. Sela (2011), "On the Heritability of Consumer Decision Making: An Exploratory Approach for Studying Genetic Effects on Judgment and Choice," *Journal of Consumer Research*, **37** (6), 951-966.

[42] Smouse P.E., Long J.C. and R.R. Sokal (1986), "Multiple Regression and Correlation Extensions of the Mantel Test of Matrix Correspondence." *Syst. Zool*, **35**, 627-632.

[43] Sokal, R.R. and F.J. Rohlf (1995), *Biometry: The Principles and Practice of Statistics in Biological Research*. W.H. Freeman and Company, New York.

[44] Solon, Gary (1992), "Intergenerational Income Mobility in the United States." *American Economic Review*, **82**(3),393-408.

[45] Stigler, G.J. and G.S. Becker (1977), "De Gustibus non Est Disputandum" *American Economic Review*, **67**(2), 76-90.

[46] Waldkirch, A., Ng, S. and D. Cox (2004), " Intergenerational Linkages in Consumption Behavior," *The Journal of Human Resources*, **39**(2), 355-381.

[47] Wilhelm, M.O., Brown, E., Rooney, P.M. and R. Steinberg (2008), "The Intergenerational Transmission of Generosity," *Journal of Public Economics* 92, 2146-2156

# Appendix: Tables

**Table 1.** Mantel Test Statistics.

| Control variables. | All EPF | Food | Non Food |
|---|---|---|---|
| None | 0.4198 | 0.5391 | 0.1991 |
| | (0.0000) | (0.0000) | (0.0089) |
| Urban, Income, Climate, | 0.2277 | 0.3735 | 0.0465 |
| Kms, Household Composition | (0.0042) | (0.0000) | (0.3190) |
| Urban, Income, Climate, | | | |
| Kms, Household Composition, | 0.1669 | 0.3154 | 0.0180 |
| Recent immigration | (0.0359) | (0.0002) | (0.4191) |

Right tail p-values in parenthesis

**Table 2.** Mantel Test Statistics.

| Control variables. | Cooking inputs | Other food goods | Non Food (tradable goods) |
|---|---|---|---|
| None | 0.5393 | 0.5730 | 0.2005 |
| | (0.0000) | (0.0000) | (0.0136) |
| Urban, Income, Climate, | 0.3168 | 0.3698 | 0.1389 |
| Kms, Household Composition | (0.0000) | (0.0000) | (0.1000) |
| Urban, Income, Climate, | | | |
| Kms, Household Composition, | 0.2404 | 0.3141 | 0.0599 |
| Recent immigration | (0.0006) | (0.0002) | (0.2995) |

Right tail p-values in parenthesis

**Table 3.** Subclasses of consumption goods in the EPF survey.

| Class | Subclass |
|---|---|
| 110 | A. Rice  B. Flour and lightly processed cereals  C. Bread<br>D. Pastry-cooked products  E. Pasta products and other cereal based products |
| 111 | A. Beef  B. Veal  C. Pork  D. Sheep meat<br>E. Poultry  F. Cooked pork  G. Canned and processed meat<br>H. Other meats and meat offal |
| 112 | A. Fresh and frozen fish  B. Dried, smoked, canned and processed fish<br>C. Fresh and frozen crustaceans and molluscs |
| 113 | A. Liquid milk  B. Preserved milk  C. Cheese and other dairy products<br>D. Eggs |
| 114 | A. Butter and margarine  B. Edible oils |
| 115 | A. Fresh fruit  B. Nuts and raisins, olives, canned fruit and fruit juices<br>C. Fresh vegetables  D. Dried vegetables  E. Frozen, preserved and canned vegetables |
| 116 | A. Potatoes and their by-products |
| 117 | A. Sugar |
| 118 | A. Coffee, cocoa, infusions and substitutes |
| 119 | A. Chocolate and chocolate substitutes  B. Other food products |
| 120 | A. Non-alcoholic beverages |
| 130 | A. Spirits  B. Wine  C. Beer  D. Other alcoholic beverages |
| 140 | A. Tobacco |
| 210 | A. Men's clothes  B. Men's underwear  C. Women's clothes<br>D. Women's underwear  E. Children's clothes<br>F. Children's and babies' underwear  G. Clothing complements and repairs |
| 220 | A. Men's footwear  B. Women's footwear  C. Children's and babies' footwear |
| 221 | A. Footwear repair |
| 310 | A. Housing rentals  B. Expenses related to property<br>C. Repair and maintenance of rented housing  D. Repair and maintenance of property |
| 311 | A. Water supply |
| 320 | A. Electricity and gas  B. Heating fuels |

**Table 3.** *(Continued)*

| Class | Subclass |
|-------|----------|
| 410 | A. Furniture for kitchen and bathroom  B. Other furniture and decorative ornaments for the household  C. Floor coverings and repairs |
| 420 | A. Household textiles  B. Other furniture goods and repairs |
| 430 | A. Refrigerators, washing machines, dishwashers and irons  B. Cookers  C. Heating appliances  D. Other electrical appliances and repairs |
| 440 | A. Glassware, tableware, cutlery and their repairs  B. Other kitchen and household equipment and their repairs |
| 450 | A. Goods for cleaning and routine household maintenance  B. Other non-durable household goods |
| 451 | A. Household services, except domestic service |
| 460 | A. Domestic service |
| 510 | A. Medical products  B. Other pharmaceutical products |
| 520 | A. Therapeutic appliances and equipment and repairs |
| 530 | A. Medical, nursing and other out-patient services |
| 540 | A. Hospital services and similar |
| 550 | A. Health insurance |
| 610 | A. Motor cars  B. Other vehicles |
| 620 | A. Tyres, spare parts and repairs |
| 621 | A. Fuels and lubricants |
| 622 | A. Other goods related to personal transport |
| 630 | A. Local transport |
| 631 | A. Long-distance transport |
| 640 | A. Postal services and communications |
| 710 | A. Radio and television equipment and repairs  B. Other audio-visual equipment |
| 711 | A. Photographic equipment, computers and others |
| 712 | A. Equipment for sport  B. Games and toys  C. Other recreational goods |
| 720 | A. Cinema, theatre, football and others performances |
| 721 | A. Recreational services |
| 730 | A. Books newspapers and magazines |
| 740 | A. Pre-primary and primary education  B. Secondary education  C. Expenses related to education  D. Secondary education  E. Education not definable by level |

**Table 3.** *(Continued)*

| Class | Subclass |
|-------|----------|
| 810 | A. Personal care services |
| 811 | A. Durable personal care goods B. Non-durable personal care goods |
| 820 | A. Jewellery, imitation jewellery and their repairs |
| 821 | A. Other personal effects |
| 822 | A. Stationery materials |
| 830 | A. Restaurants, bars and cafes |
| 831 | A. Hotels and other accommodations |
| 840 | A. Tourist services |
| 850 | A. Financial services |
| 860 | A. Other services |

**Table 4.** Type of household considered

| Column | Type of household |
|--------|-------------------|
| 1 | One adult aged 65 or older without children |
| 2 | One adult younger than 65 without children |
| 3 | One adult with one or more children |
| 4 | Couple without children (head aged 65 or older) |
| 5 | Couple without children (head younger than 65) |
| 6 | Couple with one child |
| 7 | Couple with two children |
| 8 | Other households with two adults (wihtout children) |
| 9 | Other households with two adults |
| 10 | Three adults (without children) |
| 11 | Three adults (with children) |
| 12 | Four adults or more (without children) |
| 13 | Four adults or more (with children) |