Intergenerational Mobility and the Informative Content of Surnames

Maia Güell (UPF)
José V. Rodríguez Mora (University of Southampton and UPF)
Chris Telmer (CMU)

January 2007
Summary (1/2)

- Introduction
- Motivation
- Intuition 1 (Inheritance)
- Intuition 2 (frequency, birth, death)
- Usefulness
- Why Surnames Inform
- A Model of Surnames and Income
  - Income
  - Surnames
  - Informative content of surnames
  - Objective
  - Procedure
  - Results
- Extensions of the model
  - Assortative Mating in Income or Education
  - Ethnicity
  - Inmigrants
Summary (2/2)

- Frequency of Surnames
- Spanish Surnames
- Data
  - The 2001 Catalan Census
  - Distribution of Surnames
  - Analyzed Population
- Empirical Results
  - Surnames are informative. Full Population
  - Large ICS for unfrequent surnames
  - Siblings
  - Evolution of Mean and standard deviation of Education
  - ICS increases with time.
  - Ethnicity and Background are BOTH more determinant
  - Increase of ICS
  - Assortative mating
  - Frequency
Introduction

- Goal: Study of inheritance using surnames.
- We construct a model that generates a joint distribution of surnames and income.
  - It explains why surnames carry information.
  - The more information surnames contain, the more prominent is inheritance in determining one’s position in society.
- We use the census of a large Spanish region. We show:
  - That Surnames do contain information
  - That it does so in the manner that it is perfectly consistent with the model.
  - That the amount of information contained in surnames has increased over time, suggesting a fall in the degree of intergenerational mobility.
Motivation

- It is difficult to measure mobility.
Motivation

• It is difficult to measure mobility.

• Traditionally. Panel data correlation of parents and children.
  • Difficulty gathering panel data (growth, LDC, history)
  • Difficult to get a direct measure of inheritance...
Motivation

- It is difficult to measure mobility.

- Traditionally, Panel data correlation of parents and children.
  - Difficulty gathering panel data (growth, LDC, history)
  - Difficult to get a direct measure of inheritance...

- But even if you have a panel (Solon 92,02,...):
  - Upward mobility bias because of noise.
  - Sample selection &Attrition.
Motivation

- It is difficult to measure mobility.

- Traditionally. Panel data correlation of parents and children.
  - Difficulty gathering panel data (growth, LDC, history)
  - Difficult to get a direct measure of inheritance...

- But even if you have a panel (Solon 92,02,...):
  - Upward mobility bias because of noise.
  - Sample selection & Attrition.

- Difficult to compare
Motivation

- It is difficult to measure mobility.

- Traditionally. Panel data correlation of parents and children.
  - Difficulty gathering panel data (growth, LDC, history)
  - Difficult to get a direct measure of inheritance...

- But even if you have a panel (Solon 92,02,...):
  - Upward mobility bias because of noise.
  - Sample selection & Attrition.

- Difficult to compare across countries
Motivation

- It is difficult to measure mobility.

- Traditionally, panel data correlation of parents and children:
  - Difficulty gathering panel data (growth, LDC, history)
  - Difficult to get a direct measure of inheritance...

- But even if you have a panel (Solon 92,02,...):
  - Upward mobility bias because of noise.
  - Sample selection & Attrition.

- Difficult to compare across countries and
Motivation

• It is difficult to measure mobility.

• Traditionally. Panel data correlation of parents and children.
  • Difficulty gathering panel data (growth, LDC, history)
  • Difficult to get a direct measure of inheritance...

• But even if you have a panel (Solon 92,02,...):
  • Upward mobility bias because of noise.
  • Sample selection & Attrition.

• Difficult to compare across countries and across time.
Motivation

- It is difficult to measure mobility.

- Traditionally. Panel data correlation of parents and children.
  - Difficulty gathering panel data (growth, LDC, history)
  - Difficult to get a direct measure of inheritance...

- But even if you have a panel (Solon 92,02,...):
  - Upward mobility bias because of noise.
  - Sample selection & Attrition.

- Difficult to compare across countries and across time.
  - and comparisons are the only meaningful thing.
Motivation

• It is difficult to measure mobility.

• Traditionally. Panel data correlation of parents and children.
  • Difficulty gathering panel data (growth, LDC, history)
  • Difficult to get a direct measure of inheritance...

• But even if you have a panel (Solon 92,02,...):
  • Upward mobility bias because of noise.
  • Sample selection & Attrition.

• Difficult to compare across countries and across time.
  • and comparisons are the only meaningful thing.

• Our methodological point: surnames allow to study many longitudinal questions without panel data (nor family linkages between individuals)
• Surnames are a link between the present and the past.
Intuition 1 (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
Intuition 1  (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
- Surnames are inherited from parents.

Surnames do not affect your income... but are inherited along variables that affect your income. You do not observe these variables (how much economically meaningful inheritance did the individual get)... but you observe the surname. The more informative content of surnames, the less mobility.
Intuition 1 (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
- Surnames are inherited from parents.
- Surnames do not affect your income

...but a re inherited along va riables that aect y our income.

You do not observe these va riables (ho w much economically mean-

...but y ou observe the surname.

The mo re info rmation the surname has... it is b ecause the mo re
imp o rtance the OTHER va riables that a re inherited have in order
to determine so cio economic p osition.

Thus, mo re info rmative content of surnames, less mobilit y.
Intuition 1 (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
- Surnames are inherited from parents.
- Surnames do not affect your income

... but are inherited along variables that affect your income.
Intuition 1 (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
- Surnames are inherited from parents.
- Surnames do not affect your income
- ... but are inherited along variables that affect your income.
- You do not observe these variables (how much economically meaningful inheritance did the individual get)
Intuition 1 (Inheritance)

• Surnames are a link between the present and the past.
• To study of mobility is to study inheritance.
• Surnames are inherited from parents.
• Surnames do not affect your income
• ... but are inherited along variables that affect your income.
• You do not observe these variables (how much economically meaningful inheritance did the individual get)
• ... but you observe the surname.
Intuition 1 (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
- Surnames are inherited from parents.
- Surnames do not affect your income
- ... but are inherited along variables that affect your income.
- You do not observe these variables (how much economically meaningful inheritance did the individual get)
- ... but you observe the surname.
- The more information the surname has... it is because the more importance the OTHER variables that are inherited have in order to determine socioeconomic position.
Intuition 1 (Inheritance)

- Surnames are a link between the present and the past.
- To study of mobility is to study inheritance.
- Surnames are inherited from parents.
- Surnames do not affect your income
- ... but are inherited along variables that affect your income.
- You do not observe these variables (how much economically meaningful inheritance did the individual get)
- ... but you observe the surname.
- The more information the surname has... it is because the more importance the OTHER variables that are inherited have in order to determine socioeconomic position.
- Thus, more informative content of surnames, less mobility.
Intuition 2 (frequency, birth, death)

- It works because surnames have a very skewed distribution:
  - a few surnames are very common,
  - but the huge majority of surnames are very unfrequent,
  - and most people have quite unfrequent surnames.

For many people in society, those sharing their surname are substantially more likely to be family related than those not share the surname.

Surnames are a partition of society that is informative about family links, not only about ethnic origin or migration origin.

This is a consequence of the process of birth and death of surnames. It applies to most Western naming conventions.
• It works because surnames have a very skewed distribution:
  • a few surnames are very common,
  • but the huge majority of surnames are very unfrequent,
  • and most people have quite unfrequent surnames.

• For many people in society, those sharing their surname are substantially more likely to be family related than people who do not share the surname.
Intuition 2 (frequency, birth, death)

- It works because surnames have a very skewed distribution:
  - a few surnames are very common,
  - but the huge majority of surnames are very unfrequent,
  - and most people have quite unfrequent surnames.

- For many people in society, those sharing their surname are substantially more likely to be family related than people than does not share the surname.

- Surnames are a partition of society that is informative about family links.
Intuition 2 (frequency,birth,death)

- It works because surnames have a very skewed distribution:
  - a few surnames are very common,
  - but the huge majority of surnames are very unfrequent,
  - and most people have quite unfrequent surnames.

- For many people in society, those sharing their surname are substantially more likely to be family related than those that do not share the surname.

- Surnames are a partition of society that is informative about family links, not only about ethnicity or migration origin.
• It works because surnames have a very skewed distribution:
  • a few surnames are very common,
  • but the huge majority of surnames are very unfrequent,
  • and most people have quite unfrequent surnames.

• For many people in society, those sharing their surname are substantially more likely to be family related than people that do not share the surname.

• Surnames are a partition of society that is informative about family links, not only about ethnicity or migration origin.

• This is a consequence of the process of birth and death of surnames. It applies to most western naming conventions.
Usefulness

- This is a very general point, applicable to any problem of inheritance whenever:
  - you want to know the impact of the background (the parents) on a variable of the children.
  - but you do not observe the amount inherited, only the outcome on the children.
  - ... and you have available extensive information of the distribution of outcomes per surnames... which is quite common.

- **Drawback:** We can say only how much inheritance matter, not discern whether is of one type or another (nature vs. nurture).
Why Surnames Inform

- Surnames are a partition of the population that informs on the family relations of the individual.
Why Surnames Inform

- Surnames are a partition of the population that informs on the family relations of the individual.
  - Birth/Death process of surname creation generates skewed distributions.
  - For many individuals, sharing surname means close family relation.
  - The less mobility, the more family explains, the more info surname contains.
  - More unfrequent surnames should have more information.

- Surnames inform on the speed of assimilation of emigrants.
  - The slower the path of integration (less mobility), the more they matter.

- Differences in rates of birth/death of lineages according to income.
  - Frequency should also have info.
• Surnames are a partition of the population that informs on the family relations of the individual.
  - Birth/Death process of surname creation generates skewed distributions.
    - For many individuals, sharing surname means close family relation.
    - The less mobility, the more family explains, the more info surname contains.
    - More unfrequent surnames should have more information.
  - Assortative Mating is isomorphic to an increase of inheritance.
    - Increase the amount of information that the father has on the son... and surnames are transmitted through the male line.
Why Surnames Inform

- Surnames are a partition of the population that informs on the family relations of the individual.
  - Birth/Death process of surname creation generates skewed distributions.
    - For many individuals, sharing surname means close family relation.
    - The less mobility, the more family explains, the more info surname contains.
    - More unfrequent surnames should have more information.
  - Assortative Mating is isomorphic to an increase of inheritance.
    - Increase the amount of information that the father has on the son... and surnames are transmitted through the male line.

- Surnames are a partition that informs on ethnicity of the individual.

  If ethnicity matters, surnames do matter insofar ethnic characteristics are inherited through the male line.

- Surnames inform on the speed of assimilation of emigrants.
  - The slower the path of integration (less mobility), the more they matter.

- Differences in rates of birth/death of lineages according to income.
  - Frequency should also have info.
Why Surnames Inform

- Surnames are a partition of the population that informs on the family relations of the individual.
  - Birth/Death process of surname creation generates skewed distributions.
    - For many individuals, sharing surname means close family relation.
    - The less mobility, the more family explains, the more info surname contains.
    - More unfrequent surnames should have more information.
  - Assortative Mating is isomorphic to an increase of inheritance.
    - Increase the amount of information that the father has on the son... and surnames are transmitted through the male line.

- Surnames are a partition that informs on ethnicity of the individual.
  - If ethnicity matters, surnames do matter insofar ethnicity characteristics are inherited through the male line.
Why Surnames Inform

- Surnames are a partition of the population that informs on the family relations of the individual.
  - Birth/Death process of surname creation generates skewed distributions.
    - For many individuals, sharing surname means close family relation.
    - The less mobility, the more family explains, the more info surname contains.
    - More unfrequent surnames should have more information.
  - Assortative Mating is isomorphic to an increase of inheritance.
    - Increase the amount of information that the father has on the son... and surnames are transmitted through the male line.

- Surnames are a partition that informs on ethnicity of the individual.
  - If ethnicity matters, surnames do matter insofar ethnicity characteristics are inherited through the male line.
  - Surnames inform on the speed of assimilation of emigrants.
    - The slower the path of integration (less mobility), the more they matter.
Why Surnames Inform

- Surnames are a partition of the population that informs on the family relations of the individual.
  - Birth/Death process of surname creation generates skewed distributions.
    - For many individuals, sharing surname means close family relation.
    - The less mobility, the more family explains, the more info surname contains.
    - More unfrequent surnames should have more information.
  - Assortative Mating is isomorphic to an increase of inheritance.
    - Increase the amount of information that the father has on the son... and surnames are transmitted through the male line.

- Surnames are a partition that informs on ethnicity of the individual.
  - If ethnicity matters, surnames do matter insofar ethnicity characteristics are inherited through the male line.
  - Surnames inform on the speed of assimilation of emigrants.
    - The slower the path of integration (less mobility), the more they matter.

- Differences in rates of birth/death of lineages according to income.
  - Frequency should also have info.
A Model of Surnames and Income

- Non-Overlapping generation model of Income and Surnames transmission and generation.

- In each period generates a “census”:
  - pair \((surname, income)\) for each individual.

- **Income transmission process: Inheritance**

- **Surnames** are (almost always) inherited.
  - Death and Birth of lineages.

- **Informative content of surnames.** \(R^2\). Fake Surnames.

- **Objective**: See whether ICS increases with \(\rho\).

- **Procedure**

- **Results**
• Income generation process: $Y_t = \rho Y_{t-1} + u_t$

• $\rho$ is the measure of inheritance.

• Siblings have smaller conditional variance than population.
Surnames

- Model of only males (reproduction, but no females).
  - Surname inheritance via male line.

- Surnames (lineages) are born and die:
  
  - **DEATH**: The last **male** with the surname has no **male** children.
  
  - Probability of having (male) children: \( Q \)
    - Surname dies with prob \( 1 - Q \) if you are only one with it.

  - Conditional on having children, the number of male children is fixed and equal to \( M \)
    - The expected number of male children: \( E = QM \)

- **BIRTH**: “mutation”.

  - You have surname of your father, unless you have a mutation.
  
  - Exogenous, constant arrival rate, \( \mu \).
Informative content of surnames

- $R^2_{\text{surname}}$ of regression with \textit{income} in LHS, \textit{surname} in RHS.

- Surname distributions are very skewed. To control for strange things:
  - “Fake Surname”: Non family-related partition of the economy with the same distribution than surnames.
  - “Fake surname” taken from a distribution that is identical to the one of the simulated economy (skewed).

- $R^2_{\text{fake}}$ of regression with \textit{income} in LHS, \textit{fake.surname} in RHS.

- Define \textbf{Informational Content of Surnames} as $R^2_{\text{surname}} - R^2_{\text{fake}}$
  - If only one surname: $\text{ICS}=0$
  - If one individual per surname: $\text{ICS}=0$
  - If the information is because partition, not because family: $\text{ICS}=0$
Objective

• We want to see whether it is true that the informative content of surnames is larger the larger the degree of inheritance.

• $R^2_{surnames}$ is larger for larger $\rho$.

• $R^2_{fake}$ is not larger for larger $\rho$. 
Procedure

- Assume exogenous parameters \((\rho, V_u, Q, M, \mu)\)

- Start with uninformative surnames (random). Both, skewed and uniform initial surname distribution.

- Converge to a joint distribution of surnames and income. Skewed.

- Measure Informative Content of Surnames.

- Do it many times

- Do it for many values of \(\rho\)
• Surnames are informative, and their informational content increases with the degree of inheritance that there is in society.

• \( Q = \frac{1}{2}, \ M = 2, \ \mu = 0.02\%, \ V_e = 1 \)

• Irrespectively of

  • unconditional variance.
  • mutation rate.
  • or family size
Unconditional variance

- $V_e = 10$
- $V_e = 0.1$
Mutation rate

- $\mu = 0.2$
- $\mu = 0.002$
High Variance of Family Size

- **Adjusted R^2 of MUE surname dummies:**
  - $v_0 = 1.00, nD = 1000000, nMUE = (0.0250, 0.0200, 0.0200);$
  - $\varepsilon = (1.00, 1.00, 1.00); \theta = (0.25, 0.35, 0.25); \alpha = (4, 4, 4);$

- **Time series of Adjusted R^2 of MUE surname dummies:**

- **Average Adjusted R^2 of dummies (real and MUE) per rho:**

- **Standard deviation of Adjusted R^2 of dummies (real and MUE) per rho:**
Extensions of the model

- Assortative Mating in Income or Education
- Ethnicity
- Inmigrants
- Frequency of Surnames
Assortative Mating in Income or Education

- Extension of the model. Inheritance from both father and mother.

- Assortative Mating: rich people gets to marry rich people.

- The more AM...
  - the more info the income of the husband has on the income of the wife.
  - the more info the income of the husband has on the inheritance than the wife passes
  - the more info the income of the husband has on the income of the kids

- More AM is identical to an increase in $\rho$ in our model.
• Imagine that belonging to a certain ethnic group may change the income process:

\[ y_t = \alpha_{eth} + \rho y_{t-1} \]

• If the surname contain information on the ethnicity, the surname contain info on the income.

• Surname contain info on ethnicity only in the measure
  • that ethnic characteristics are inherited through the male line
  • that there exists assortative mating according to ethnicity (judaism)
In the measure that immigrants have different surnames that population.

If by the surname we can characterize the origin of the migrant, we can determine the speed at which migrants integrate in the recipient society.
In the previous model the specific surnames had information, but their frequency did not.

Frequency has information if probability of death/birth of lineages differs across income groups.

"Hereu" effect: The rich want a male, $Q_{rich} = 1$, but $E = 1$ for all.

Average Fertility Differences: The rich (male) have more children in average.
Frequency in baseline model

- Time-series of frequency of real outcome
- Average t-statistic of the frequency of outcome (real and FAME) per rho
- Standard deviation t-statistic of the frequency of outcome (real and FAME) per rho
- Time-series of t-statistic of the frequency of FAME outcome
Hereu effect (1/3)
Hereu effect (3/3)
Average Fertility Differences (1/3)

ADJUSTED R^2 of REAL surname dummies:
Ve = 1.000; N0 = 10000000; mutations(M,m,p) = (0.0200, 0.0200, 0.0200);
M = (1.50, 1.00, 0.50); E = (0.50, 0.50, 0.50); M = (3, 2, 1);

Time series of Adjusted R^2 of FAKE surname dummies

Average Adjusted R^2 of dummies (real and FAKE) per rho

Standard deviation of Adjusted R^2 of dummies (real and FAKE) per rho
Average Fertility Differences (2/3)
Average Fertility Differences (3/3)
Spanish Surnames

- Spain is an ideal case for surnames studies.
- It is hard to change surnames (much more than, say, the US).
- Males and Females do not change surnames along their lives, regardless of marital status.
- Spaniards have two surnames, inherited from father and mother, and we can measure assortative mating, which is part of the story.
- We can locate siblings with quite a lot of accuracy, by the combination of two surnames.
- Almost all the immigration (before 2001) is between regions of Spain and we can control for this since we know the distribution of surnames in the country. Otherwise we would need data from other countries.
- Stable, simple and generalized orthographic rules. Few “mutations” due to spelling differentials.
Data

- The 2001 Catalan Census
- Distribution of Surnames
- Analyzed Population
The 2001 Catalan Census

- Census of the whole population, 6343110

- Variables:
  - 2 Surnames
  - Demographic characteristics (including birthplace)
  - Household characteristics

- Does not have income, but has relevant economic outcomes.
  - Education
  - House Inheritance
  - Availability of second residence.
• Very skewed. Gini .9

• Some interesting figures:
  • 61K different surnames.
  • On average, 34 people per surname.
  • Yearly 96 surnames die.
  • In Spain in 2001, 1.5K individuals had their surname approved for modification.

This amounts to a mutation rate of .26%. 107 new surnames.
Analyzed Population

- Aged 25 and above (full time education finished)

- Focus on information in surnames due to family links:
  - Spanish citizens living in Catalonia (no foreign immigrants)
  - “Ethnicity” component of surnames (regional origin within Spain),
    \[
    CatalanDegreeSurname(j) = \frac{\text{Number of Phones under surname } j \text{ in Catalonia}}{\text{Number of Phones under surname } j \text{ in the Spain}}
    \]
  - Captures the probability that an Spaniard holding name \( j \) is living in Catalonia.
  - **We use it as a proxy of the “catalonianess” of the holder.**
### CatDegree as proxy of ethnicity

#### Table 1: Comparison of Language Proficiency and Immigration Status

<table>
<thead>
<tr>
<th></th>
<th>All residents in Catalonia</th>
<th>Born in Catalonia before 1950</th>
<th>Born anywhere in Spain before 1950</th>
<th>Born anywhere in Spain after 1950</th>
</tr>
</thead>
<tbody>
<tr>
<td>Mean CatalanDegreeSurname2</td>
<td>0.344</td>
<td>0.5672</td>
<td>0.367</td>
<td>0.322</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>(0.302)</td>
<td>(0.3241)</td>
<td>(0.312)</td>
<td>(0.292)</td>
</tr>
<tr>
<td>Share with CatalanDegreeSurname2 &gt; 0.16</td>
<td>0.568</td>
<td>0.8365</td>
<td>0.596</td>
<td>0.542</td>
</tr>
</tbody>
</table>

#### Table 2: Logistic Regression Results

**LHS: Knowledge of Catalan language**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatalanDegreeSurname2</td>
<td></td>
<td>0.639 (0.003)</td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-2248320.8</td>
<td>-2219774.9</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.2387</td>
<td>0.2483</td>
</tr>
</tbody>
</table>

**LHS: Immigrant**

<table>
<thead>
<tr>
<th></th>
<th>(1)</th>
<th>(2)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatalanDegreeSurname2</td>
<td>-4.121 (0.006)</td>
<td></td>
</tr>
<tr>
<td>Log likelihood</td>
<td>-1418949.8</td>
<td>-903746.31</td>
</tr>
<tr>
<td>Pseudo $R^2$</td>
<td>0.0052</td>
<td>0.3664</td>
</tr>
</tbody>
</table>
Empirical Results

- That the informational content of surnames is **large**, beyond ethnicity.
  - Also among Very Catalan surnames; those born in Catalonia; and Catalan born before 1950.
- Calibrating the model, the $\rho$ that matches the ICS for those born in Catalonia (3.19%) is 0.4.
- That, as the model predicts,
  - the ICS is **larger for surnames that have less individuals**, as the surname partition is then more related to family.
  - The **frequency** of surnames is also informative.
- That when we use *both* surnames (in practice grouping **Siblings** together) the informational content of surnames is even larger.
- Increase in provision of public education results in **more education**
- In spite of it:
  - Also **ICS increases with time**. Once controlling for ethnicity.
  - **Family Background and Ethnicity are BOTH more determinant**
  - This is very robust. **among least frequent, very Catalan surnames, siblings**.
- **Explanation**: The amount of **Assortative Mating** did increase one generation earlier.
## Surnames are informative. Full Population

<table>
<thead>
<tr>
<th>LHS: years of education</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatalanDegreeSurname2</td>
<td>1.692(0.007)</td>
<td>1.017(0.008)</td>
<td>1.692(0.007)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surname Dummies</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fake Surnames Dummies</td>
<td>Yes</td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.3363</td>
<td>0.3440</td>
<td>0.3653</td>
<td>0.3440</td>
<td>0.3629</td>
<td>0.3363</td>
</tr>
<tr>
<td>Surnames jointly significant* (p-value)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td>0.000</td>
<td>0.384</td>
</tr>
</tbody>
</table>

**Notes:** Individual controls include gender, age dummies & county of birth.

Number of surnames is 38,024.

ICS= 2.13%.
Large ICS for unfrequent surnames

Figure 1: Food for thought - the graph illustrates the relationship between the percentile of individuals per surname and the informational content of surnames. The x-axis represents the percentile, while the y-axis shows the number of individuals per surname. The graph indicates a decreasing trend in the number of individuals per surname as the percentile increases, suggesting a higher informational content for more frequent surnames.
## Siblings

<table>
<thead>
<tr>
<th>Dependent variable: years of education</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Adjusted $R^2$, Surname Dummies</td>
<td>0.5486</td>
<td>0.5416</td>
<td>0.5375</td>
<td>0.5326</td>
<td>0.5283</td>
<td>0.4696</td>
</tr>
<tr>
<td>Adjusted $R^2$,Fake Surnames Dummies</td>
<td>0.3244</td>
<td>0.3224</td>
<td>0.3218</td>
<td>0.3228</td>
<td>0.3232</td>
<td>0.3354</td>
</tr>
<tr>
<td>Informational Content of Surnames</td>
<td>0.2242</td>
<td>0.2192</td>
<td>0.2157</td>
<td>0.2098</td>
<td>0.2051</td>
<td>0.1342</td>
</tr>
<tr>
<td>Observations</td>
<td>774,788</td>
<td>1,315,853</td>
<td>1,664,717</td>
<td>1,900,652</td>
<td>2,067,590</td>
<td>3,695,479</td>
</tr>
<tr>
<td>Number of surnames</td>
<td>387,394</td>
<td>567,749</td>
<td>654,965</td>
<td>702,152</td>
<td>729,975</td>
<td>811,502</td>
</tr>
<tr>
<td>Max number of people per surname</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>6</td>
<td>All sample</td>
</tr>
</tbody>
</table>
Evolution of Mean and standard deviation of Education

Mean

Standard deviation
ICS increases with time.

<table>
<thead>
<tr>
<th>LHS: years edu. all</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
<th>(5)</th>
<th>(6)</th>
</tr>
</thead>
<tbody>
<tr>
<td>CatalanDegreeSurname2</td>
<td>0.896</td>
<td>0.594</td>
<td>0.897</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Surname Dummies</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Fake Surnames Dummies</td>
<td>Yes</td>
<td></td>
<td>Yes</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Adjusted R-squared</td>
<td>0.2313</td>
<td>0.2335</td>
<td>0.2533</td>
<td>0.2335</td>
<td>0.2524</td>
<td>0.2313</td>
</tr>
<tr>
<td>Surnames jointly significant* (p-value)</td>
<td>Yes</td>
<td>No</td>
<td>Yes</td>
<td>No</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>0.000</td>
<td>0.679</td>
<td>0.000</td>
<td>0.688</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Born before 1950 (ICS= 1.98%)

Born after 1950 (ICS= 3.5%)

| CatalanDegreeSurname2 | 2.143 | 1.271 | 2.145 |
| Surname Dummies | Yes  |      | Yes  |
| Fake Surnames Dummies | Yes  |      | Yes  |
| Adjusted R-squared | 0.1002 | 0.1183 | 0.1534 | 0.1184 | 0.1481 | 0.1002 |
| Surnames jointly significant* (p-value) | Yes | No | Yes | No |    |    |
|  | 0.000 | 0.260 | 0.000 | 0.421 |      |      |
Ethnicity and Background are BOTH more determinant

Evolution of ICS over time

Evolution of parameter of Catdegree
45% Most Catalan Surnames
50% Least Frequent Surnames
Increase of ICS (3/4)

ICS of the complete surname (2 surnames)
Siblings. 45% Most Catalan Surnames

Siblings. 50% Least Frequent Surnames
### Assortative mating (1/2)

#### Education

<table>
<thead>
<tr>
<th>EduSurname1</th>
<th>EduSurname2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Old&quot;</td>
<td>&quot;Young&quot;</td>
</tr>
<tr>
<td>0.170</td>
<td>0.303</td>
</tr>
<tr>
<td>Observations</td>
<td>2,041,044</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.3410</td>
</tr>
</tbody>
</table>

#### Catalan degree

<table>
<thead>
<tr>
<th>CatDegreeSurname1</th>
<th>CatDegreeSurname2</th>
</tr>
</thead>
<tbody>
<tr>
<td>&quot;Old&quot;</td>
<td>&quot;Young&quot;</td>
</tr>
<tr>
<td>0.217</td>
<td>0.328</td>
</tr>
<tr>
<td>Observations</td>
<td>2,041,044</td>
</tr>
<tr>
<td>$R^2$</td>
<td>0.5110</td>
</tr>
</tbody>
</table>
Assortative mating (2/2)
<table>
<thead>
<tr>
<th>LHS: years of education</th>
<th>(1)</th>
<th>(2)</th>
<th>(3)</th>
<th>(4)</th>
</tr>
</thead>
<tbody>
<tr>
<td>FrequencySurname1</td>
<td>-30.157(0.309)</td>
<td>-23.696(0.309)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>FrequencyFakeSurname1</td>
<td></td>
<td>0.148(0.301)</td>
<td>0.107(0.299)</td>
<td></td>
</tr>
<tr>
<td>CatalanDegreeSurname2</td>
<td>1.636(0.007)</td>
<td></td>
<td>1.692(0.007)</td>
<td></td>
</tr>
<tr>
<td>R-squared</td>
<td>0.3378</td>
<td>0.3449</td>
<td>0.3363</td>
<td>0.3440</td>
</tr>
</tbody>
</table>
Intergenerational Mobility and the Informative Content of Surnames

Maia Güell (UPF)
José V. Rodríguez Mora (University of Southampton and UPF)
Chris Telmer (CMU)

January 2007