Online Appendix B: The dynamics of surnames

Miguel Angel Carpio and María Eugenia Guerrero

December 2021

The dynamics of surnames

In this appendix, we build a model for the dynamics of surnames and present simulations to illustrate the claims in the section of the paper in which we lay out the conceptual framework. This document is organized as follows: Section 1 introduces the basic setup; Section 2 presents formal properties that describe the evolution of surnames; Section 3 presents the simulations; finally, Section 4 presents formal proofs of these properties.

1 Model set-up

An important feature of surnames, found in both Anglo-Saxon and Hispanic naming conventions, is that they are inherited from the father (the surname of the father is passed down to his heirs). The model therefore does not include females. We consider two areas indexed by $j = [A, B]$, with a population size at time t denoted by P_t^j. Let $q_j \in [0, 1]$ be the time-independent probability of each agent born in area j to survive enough to reproduce. Conditional on surviving up to the age of reproduction, each agent has m sons.[1] We assume that agents can migrate from area A to area B, but not the other way around. Let $p \in [0, 1]$ be the probability of migrating from area A to B. Individuals live in one single period in which they migrate or do not migrate, reproduce or do not reproduce, and die. In this framework, the conditional expectations of population are

$$\begin{aligned} E\left[P_{t+1}^A | P_t^A\right] &= P_t^A(1-p)q_A m \\ E\left[P_{t+1}^B | P_t^A, P_t^B\right] &= P_t^A p q_A m + P_t^B q_B m \end{aligned} \tag{1}$$

We denote the fixed and discrete set of potential surnames by Ω and its number of elements by S. Each individual is associated with one surname $s \in \Omega$. All the elements of Ω are not necessarily active in period t, but only a subset Ω_t^j with its own number of elements $S_t^j \leq S$. The function $F_t^j(s) : \Omega_t^j \to [0, 1]$ represents the marginal distribution of surnames. Finally, let $N_t^j(s)$ be the number of individuals with the same surname s for area j in period t (notice that $N_t^j(s) = P_t^j F_t^j(s)$). Obviously, $S_t^j = \sum_{s \in S} I\left[N_t^j(s) > 0\right]$.

We now turn to describe the dynamics of surnames. In order to understand the process from $F_t^j(s)$ to $F_{t+1}^j(s)$, it is important recall that a surname s can be exclusive or common in the way it has been defined in the "Conceptual framework" section of the main document. In our model, the set of exclusive surnames in A at time t is given by $\Omega_t^A - \Omega_t^B$, the set of exclusive surnames in B is $\Omega_t^B - \Omega_t^A$, and the set of common surnames is $\Omega_t^A \cap \Omega_t^B$.

[1] The initial setup up to this point is similar to the one proposed by Guell et al. (2014), but with some differences. They propose a model for a single location where the number of surnames can increase if some individuals change their surname, that is, if some sons acquire different surnames than their fathers. They develop a methodology that relies on this "mutation" to analyze inter-generational mobility. In contrast, we propose two areas to emphasize the importance of migration to the dynamics of surnames. We do not allow for surname mutation.

2 The surname process

The appearance and disappearance of a surname within an specific area depends on a stochastic process. We begin by analyzing area A. A new surname cannot arise, because, by assumption, migration from area B to area A is not possible. However, a surname disappears in $t+1$ if all fathers with surname s bear zero offspring, or reproduce and migrate. This happens with probability $[(1-q_A)+q_Ap]^{N_t^A(s)} = [1-q_A(1-p)]^{N_t^A(s)}$. We now turn to consider the case of area B. A new surname appears in $t+1$ if an individual from A with an exclusive surname $s \in \Omega_t^A - \Omega_t^B$ migrates and reproduces. This happens with probability $1-(1-q_Ap)^{N_t^A(s)}$.[2] Finally, a surname can disappear from B in $t+1$ under two scenarios. First, a common surname $s \in \Omega_t^A \cap \Omega_t^B$ can disappear from B if no father in A with surname s migrates and reproduces, and concurrently all fathers s in B bear zero offspring. This happens with probability $(1-q_Ap)^{N_t^A(s)}(1-q_B)^{N_t^B(s)}$. Second, an exclusive surname $s \in \Omega_t^B - \Omega_t^A$ disappears simply if all fathers bear zero offspring. This happens with probability $(1-q_B)^{N_t^B(s)}$.

Our first observation is simple: if surviving and reproducing is a certain event and there is no migration, population may grow exponentially, while the number of surnames within a community remains the same over time.[3] The following property presents this idea formally.

Property 1. *Assume $q=1$, $p=0$ and $m>1$. While population exponentially grows in areas A and B, the number of surnames remains the same.*

Proof. The fact that population grows comes directly from applying the assumptions to equation 1. The stability of the number of surnames is a consequence that the probabilities for the appearance and disappearance of a given surname within A and B are equal to zero under the provided assumptions. ■

The evolution of surnames is different in a context of migration and mortality. In order to describe it, a previous step is computing the number of individuals with surname s migrating from A to B at t, which we denote by $N_t^m(s)$.

Property 2. *Assume $p>0$. The number of individuals with surname s migrating from A to B at t is a binomial random variable with support $[0, N_t^A(s)]$ and described by:*

$$\begin{aligned} N_t^m(s) &= N_t^A(s)p + \epsilon_{st} && if N_t^A(s) > 0 \\ &= 0 && if N_t^A(s) = 0 \end{aligned}$$

where the conditional mean of ϵ_{st} is zero and the conditional variance is $N_t^A(s)p(1-p)$.

[2] The probability that an individual from A does not migrate to B or does not reproduce is $1-q_Ap$ and hence the probability that all individuals from A do not migrate to B or do not reproduce is $(1-q_Ap)^{N_t^A(s)}$. The complement is the probability that at least one individual migrates and reproduces.

[3] The number of surnames can increase if some individuals change their surname, that is, if some sons acquire surnames other than those of their fathers (bringing their surname to the area). As Guell et al. (2014) states, this is an important generator of new surnames for a society. However, we believe this mutation could hardly explain large regional differences in the number of surnames and so we do not allow this possibility in the model.

The proof is in Section 4. Since the number of migrating individuals with a given surname is a binomial random variable, its mean and variance are easily computable. Therefore, the previous simple equation can describe its process.

We now analyze how surnames within area A evolve. The number of individuals with the same surname grows from those who do not migrate to B and reproduce. Note that, if a surname disappears in a given period, it does so forever. Finally, given that there are no migration flows from B to A, it is not possible that a new surname would appear. We state these ideas formally in the following property.

Property 3. *Assume $p > 0$, $q < 1$ and $m > 1$. The number of individuals with surname s for area A follows a random walk process with drift and an absorbing barrier at zero:*

$$\begin{aligned} N_{t+1}^{A}(s) &= \left[N_t^A(s) - N_t^m(s)\right] q_A m + \mu_{st} && if s \in \Omega_t^A \\ &= 0 && if s \in \Omega_t^B - \Omega_t^A \end{aligned}$$

where the conditional mean of μ_{st} is zero and the conditional variance is $\left[N_t^A(s) - N_t^m(s)\right] q_A(1-q_A)m^2$.

The proof is in Section 4. Basically, the number of individuals with a given surname that reproduce while staying in A is a binomial random variable with support $[0, N_t^A(s)m - N_t^m(s)m]$ whose parameters are easily computable.

The case of the surnames within B is more complex because this area receives migrants. The process from t to $t+1$ depends on the status of each specific surname in t, namely, whether s is exclusive in A, common, or exclusive in B. Consider the following property.

Property 4. *Assume $q < 1$, $p > 0$ and $m > 1$. The number of individuals with surname s for area B follows the next process with an absorbing barrier at zero:*

$$\begin{aligned} N_{t+1}^{B}(s) &= N_t^m(s) q_A m + w_{st}^A && if s \in \Omega_t^A - \Omega_t^B \\ &= N_t^m(s) q_A m + N_t^B(s) q_B m + w_{st}^A + w_{st}^B && if s \in \Omega_t^A \cap \Omega_t^B \\ &= N_t^B(s) q_B m + w_{st}^B && if s \in \Omega_t^B - \Omega_t^A \end{aligned}$$

where the conditional mean of w_{st}^A is zero and the conditional variance is $N_t^m(s) q_A(1-q_A)m^2$, whereas the conditional mean of w_{st}^B is zero and the conditional variance is $N_t^B(s) q_B(1-q_B)m^2$.

The proof is in Section 4. Basically, the number of individuals with surname s migrating from A that reproduce and the number of individuals in B that reproduce are both binomial random variables whose parameters are computable. The former is the only variable playing a role if s is exclusive in A, the latter is the only variable that matters if s is exclusive in B, and both play a role if s is common. Section 4 also shows the specific support of $N_{t+1}^B(s)$ in every case.

3 Simulations

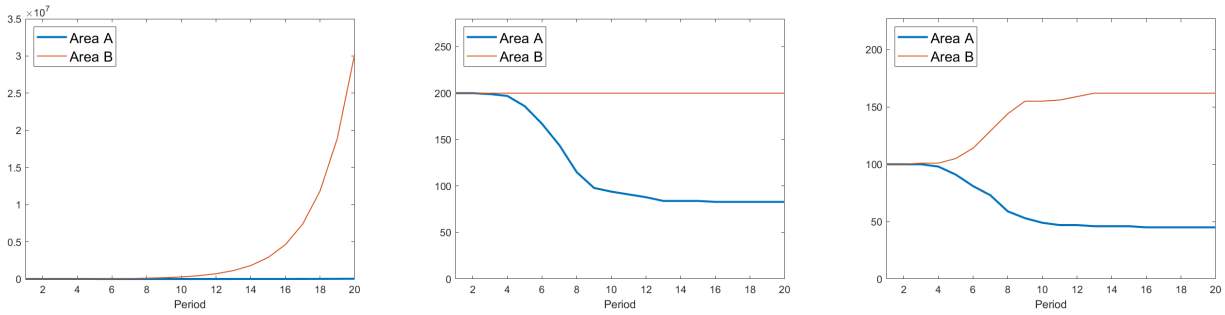
We present simulations to illustrate the four claims of our conceptual framework. We start by assuming an initial population structure: 100 A-exclusive surnames, 100 B-exclusive surnames and 100 common surnames, with 20 individuals per surname. Hence, the initial total population is 6,000. We then apply properties 3 and 4, which describe how the number of individuals with a particular surname evolves from one period to another in area A and B, respectively. Note that a surname can change status from period to period, and therefore the applicable equation of Property 4 for each surname may also change over time. Consider, for instance, the exclusive surname s in A that arrives at B at $t+1$. In that period, this surname can become common or exclusive in B.[4] Finally, we work with a number of generations equal to $T = 20$ and, for simplicity, we fix the number of sons to $m = 2$.

3.1 Claims

The *first claim* is that a mortality increase (or a fertility decrease), when it generates total deaths, leads to a decrease in the number of surnames in area A and it has no effect in area B. At the same time, this shock leads to a decrease in the number of exclusive surnames in area A and to an increase in area B. In order to illustrate this claim, we set the probability of surviving and reproducing to 0.8 over time in A and B, but we change this probability to 0.4 in A from period 2 to 8. Figure B.1 presents the time series of our indicators. The shock obviously generates a difference between population in area A and B. In the case of the number of surnames, it decreases in A because entire groups of individuals with the same surname disappear. This variable remains equal in B, because no shock was introduced there. In the case of the number of exclusive surnames, it decreases in A, because some surnames vanish due to total death. This number increases in B because total death in A for some families with common surnames makes these surnames exclusive in B.

[4] In general, a surname $s \in \Omega_t^A - \Omega_t^B$ can disappear, remain exclusive in A, become common or even become exclusive in B in period $t+1$. A surname $s \in \Omega_t^A \cap \Omega_t^B$ may also disappear, become exclusive in A, remain common or become exclusive in B. A surname $s \in \Omega_t^B - \Omega_t^A$ can only disappear or remain exclusive in A.

Figure B.1: Effect of mortality increase in A on time series of indicators



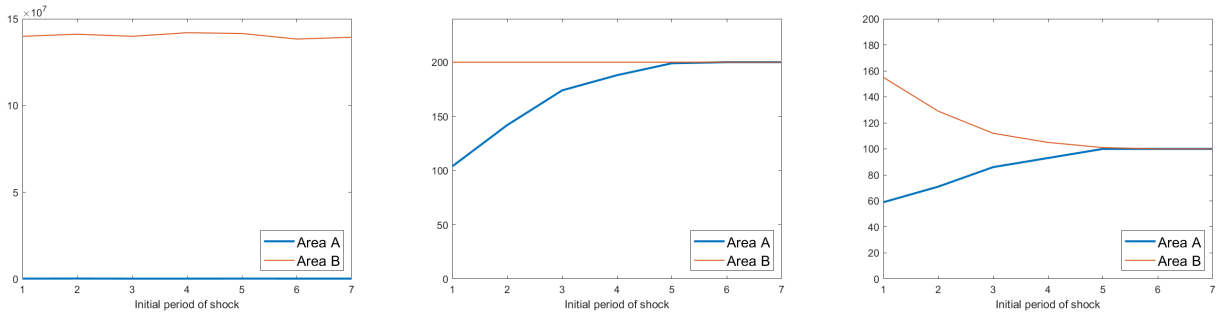
Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 0.8$, $p = 0$, $m = 2$ and $T = 20$. A 7-period shock is introduced from period 2 to 8, where $q_A = 0.4$. The graphs show time series for population size, number of surnames and number of exclusive surnames.

Source: Simulation using the model presented in the section titled "The surname process".

An important observation related to the first claim is that a mortality increase in A has to fulfill three conditions to produce total deaths. First, it must take place in a context of few people for each surname. Since the number of people per surname tends to grow exponentially, this last condition is fulfilled if the shock affects the first generations. Figure B.2 analyzes how the period of shock initiation affects the results. We set the probability of survival and reproducing to 0.9 in A and B from period 1 to period 20. We allow this probability to be equal to 0.4 in A for 7 periods, starting in a period that varies over the x-axis from 1 to 7. The y-axis of the three plots represents population size, number of surnames and number of exclusive surnames, respectively, in A and B at the last period. While shocks initiated during any period generate a difference between the last population size in A and B, only shocks initiated in the first periods open a gap between our surname indicators at the last period. In this particular simulation, the shock has to be initiated no later than period 4. The intuition is direct. During initial periods, total death may occur because families have a relatively smaller number of members, but it rarely takes place once a family reaches a certain size.

Figure B.2: Effect of a mortality increase in A on indicators at the last period, by initial period



Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 1$, $p = 0$, $m = 2$ and $T = 20$. A 7-period shock is introduced by setting $q_A = 0.4$ from an initial period that varies as shown by x-axis. The graphs show population size, number of surnames and number of exclusive surnames at the last period.

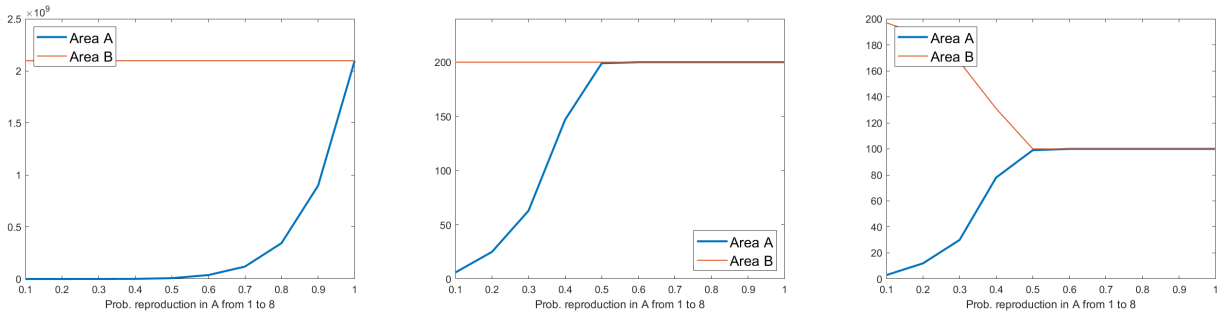
Source: Simulation using the model presented in the section titled "The surname process".

Second, for a mortality increase to produce total death, it must be large enough to include all the members of some families. Figure B.3 shows how the results depend on the magnitude of the shock. We set the probability of surviving and reproducing to 1 in A and B over time, and allow this probability to be different in A from period 2 to 8. The x-axis of the three plots of Figure B.3 show different levels for this probability, over a range that extends from 0.1 to 1.0 (i.e. the shock is equal to 1 minus the value in the x-axis). The y-axis of the three plots present the level of population size, number of surnames and number of exclusive surnames, respectively, in A and B at the last period ($t = T = 20$). Figure B.3 shows that all mortality decreases in A cause the last population size smaller in A than in B. However, small shocks do not open a gap between A and B in the final number of surnames or the final number of exclusive surnames. In this particular simulation, the mortality decrease must be larger than 0.5, that is, the probability of surviving and reproducing must be smaller than 0.5.

Third, the mortality increase must be spread over a number of generations until it reaches all family members. Figure B.4 shows that the results depend on the extension of the shock over time. We set the probability of surviving and reproducing to 1 in A and B from period 1 to 20, but change this probability to 0.4 in A from period 2 to a varying period number. Of course, this final period number determines different extensions for the shock. The x-axis of the three plots in Figure B.4 show these different extensions over a range that extends from 1 to 7 years. The y-axis of the three plots represents the level of population size, number of surnames and number of exclusive surnames, respectively, in A and B at the last period. Figure B.4 again shows that the final population size is always smaller in A than B. However, a shock in A does not open a gap between A and B in the surname indicators unless is lasts more than 3 periods.

In sum, an eventual difference in current population size may be caused by numerous shocks. In contrast, a difference in our indicators can only be generated by some specific mortality increases (or

Figure B.3: Effect of a mortality increase in A on indicators at the last period, by size

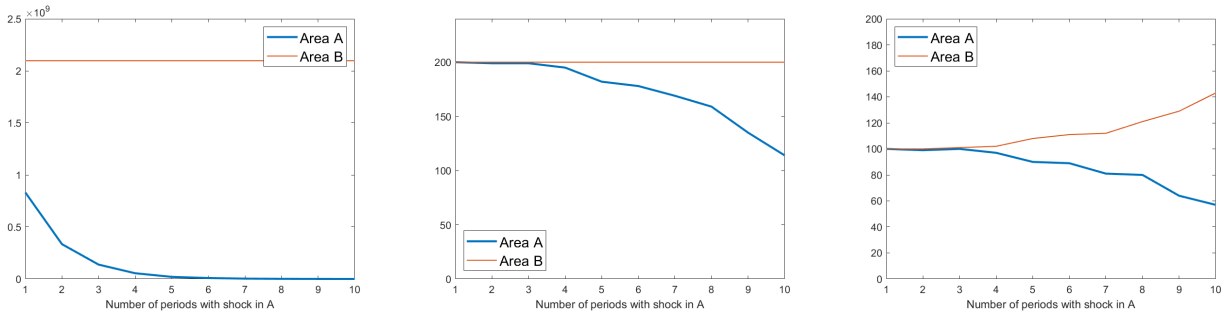


Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 1$, $p = 0$, $m = 2$ and $T = 20$. A 7-period shock is introduced from period 2 to 8 by varying q_A as shown by x-axis. The graphs show population size, number of surnames and number of exclusive surnames at the last period.

Source: Simulation using the model presented in the section titled "The surname process".

Figure B.4: Effect of a mortality increase in A on indicators at the last period, by duration



Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 1$, $p = 0$, $m = 2$ and $T = 20$. A shock is introduced by setting $q_A = 0.4$ from period 2 on, for a varying number of periods as shown by x-axis. The graphs show population size, number of surnames and number of exclusive surnames at the last period.

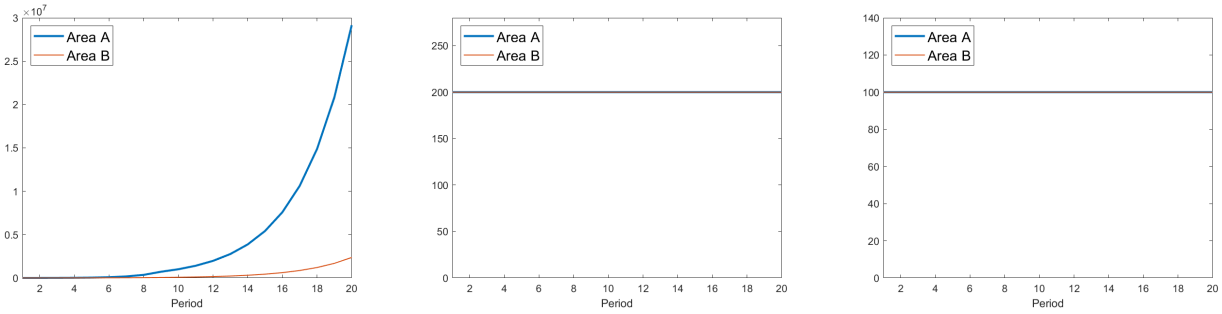
Source: Simulation using the model presented in the section titled "The surname process".

fertility decreases). In particular, shocks must be old, large and long. These exercises show that indicators based on surnames correctly identify a mortality increase with these characteristics (like the one caused by *mita*).

The *second claim* is that a mortality decrease in A (or a fertility increase) does not raise the number of surnames or the number of exclusive surnames in either location. We illustrate this claim by setting the probability of surviving and reproducing to 0.7 in A and B over time and by introducing a probability equal to 1 in A from period 2 to 8. Figure B.5 presents the time series of our indicators. The shock

obviously results in population size being greater in A than in B. However, since the stock of surnames is given, this shock does not increase the number of surnames or the number of exclusive surnames; it only results in the number of individuals per surname being larger in A than in B.

Figure B.5: Effect of a mortality decrease in A on time series of indicators



Population Surnames Exclusive surnames

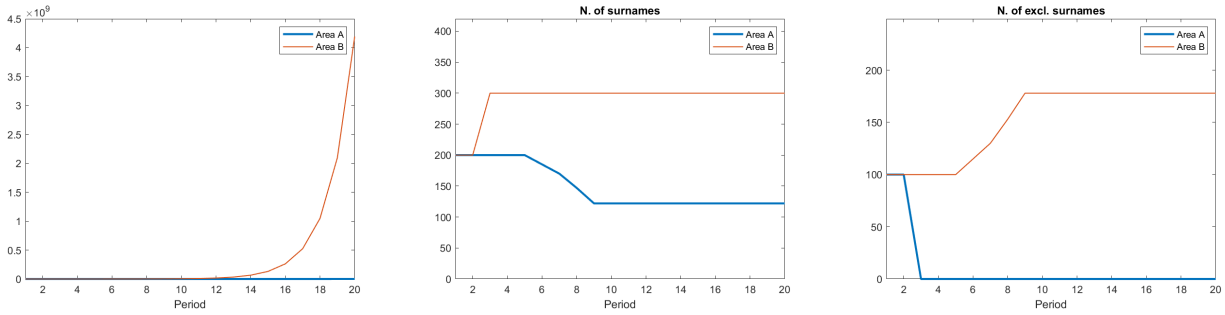
Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 0.7$, $p = 0$, $m = 2$ and $T = 20$. A 7-period shock is introduced from period 2 to 8, where $q_A = 1.0$. The graphs show time series for population size, number of surnames and number of exclusive surnames.

Source: Simulation using the model presented in the section titled "The surname process".

The *third claim* is that a migration flow from A to location different from B, when it generates total migrations, leads to a decrease in the number of surnames in area A and has no effect in area B. At the same time, this shock leads to a decrease in the number of exclusive surnames in area A and to an increase in area B. The figures used in the case of a mortality increase are useful for illustrating this migration flow: The reader only has to consider total migration instead of total death, and partial migration instead of partial death.

The *fourth claim* is that a migration flow from A to B leads to a decrease in the number of surnames and number of surnames in area A, and to an increase in both indicators in area B. We set the probability of migration from A to B equal to 0 from periods 1 to 20. However, we increase it to 0.6 from period 1 to period 8. Figure B.6 presents the time series by location. The number of surnames in A decreases, because some families experience total migration. The number of surnames in B, in contrast, increases due not only to total migration of families with exclusive surnames, but also to their partial migration. Since the appearance of a surname in B requires only one individual from a family with an exclusive surname in A to migrate, the number of surnames in B rapidly reaches its maximum possible value. The number of exclusive surnames drops in A due to total and partial migration: if only one individual from a family with an exclusive surname in A migrates, then this surname becomes common. The number of exclusive surnames in B rises due to total migration: only if all members of a family out-migrate from A does their surname becomes exclusive in B. In sum, migration from A to B makes both the number of surnames and the number exclusive surnames smaller in A than in B.

Figure B.6: Effect of a migration flow from A to B on time series of indicators



Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 1.0$, $p = 0$, $m = 2$ and $T = 20$. A 7-period shock to migration is introduced from periods 2 to 8, where $p = 0.7$. The graphs show time series for population size, number of surnames and number of exclusive surnames.

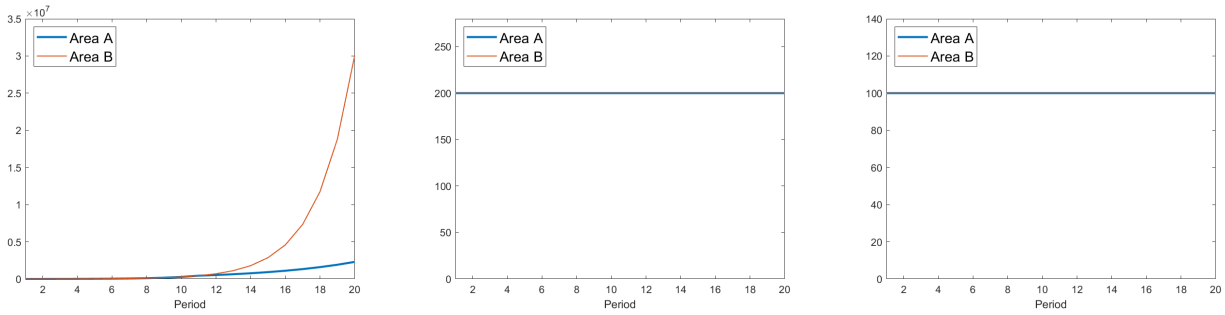
Source: Simulation using the model presented in the section titled "The surname process".

3.2 Scenarios

The article states that we are not permitted to draw conclusions about past population size from observing this variable today, and it identifies two scenarios in which current population may lead to error and surnames would not. One scenario assumes a shock that did not generate a mortality increase in the past but in recent times as a consequence of differentiated living conditions (*Scenario 2*). We illustrate this case by setting the probability equal to 0.8 in A from periods 1 to 10 and then to 0.60 from period 11 to 20. Figure B.7 shows the time series of our indicators. Notice that, in the last period, we observe a gap in population size but no gap in the surname indicators. In fact, the surname indicators do not change at all.

The other case assumes a shock that generated a mortality increase in the past, and a posterior mortality decrease or a posterior fertility increase (*Scenario 4*). We illustrate this by introducing two successive shocks. In a context of a probability of surviving and reproducing equal to 0.6 in A and B, we set this probability equal to 0.4 in A from periods 2 to 8 and then to 0.85 from period 9 to 16. Figure B.8 shows the time series of our indicators. The two shocks compensate for each other in terms of population, but the first opens a gap between the number of surnames and the number of exclusive surnames that lasts forever.

Figure B.7: Effect of a recent shock to mortality in A on time series of indicators

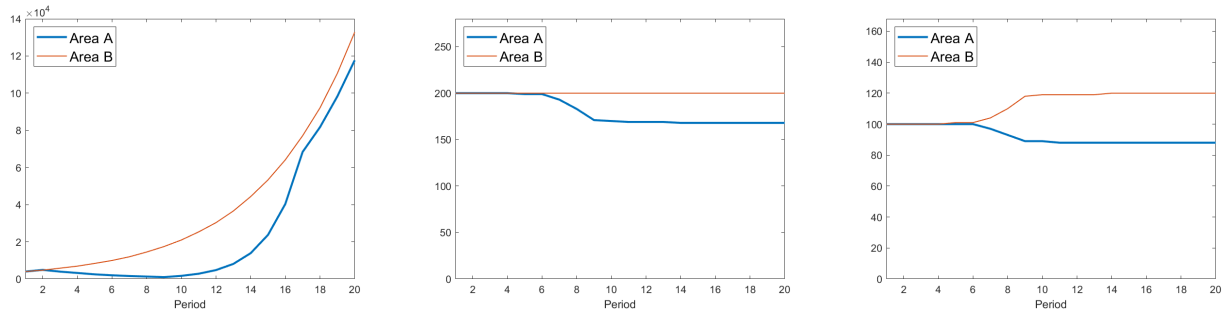


Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 0.8$, $p = 0$, $m = 2$ and $T = 20$. A 10-periods mortality increase is introduced from period 11 to 20, where $q_A = 0.6$. The graphs show time series for population size, number of surnames and number of exclusive surnames.

Source: Simulation using the model presented in the section titled "The surname process".

Figure B.8: Effect of mixed shocks to mortality in A on time series of indicators



Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = S_B(0) = 200$, 100 common surnames, 20 individuals per surname, $q_A = q_B = 0.6$, $p = 0$, $m = 2$ and $T = 20$. A 7-period mortality increase is introduced from period 2 to 8, where $q_A = 0.4$, and another 7-period mortality decrease from period 2 to 8, where $q_A = 0.85$. The graphs show time series for population size, number of surnames and number of exclusive surnames.

Source: Simulation using the model presented in the section titled "The surname process".

3.3 Galton-Watson process

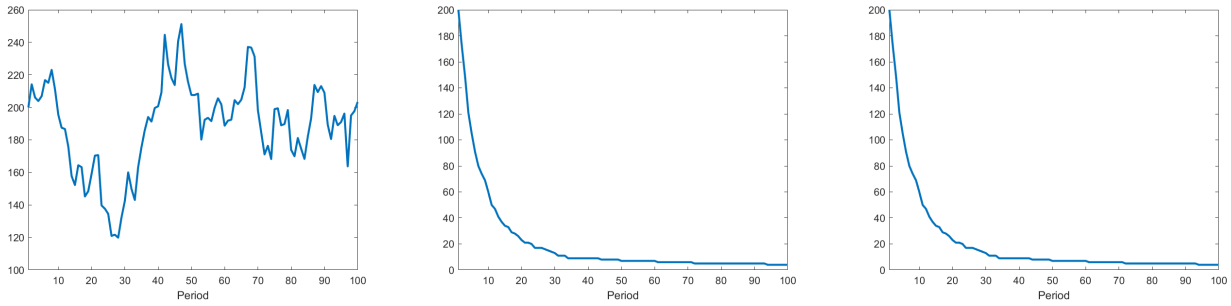
The question posed by Francis Galton refers to an initial situation with a large number N of males, each with a different surname. It also assumes that six fixed fractions of the males have different numbers of male children (from 0 to 5) in each generation. A main conclusion of the GW process is that, if the population growth rate is positive, the number of surnames decreases overtime up to a fixed number. If the population growth rate is zero, this fixed number is equal to one and the population eventually disappears.

The model we present is compatible the Galton-Watson (GW) process. The population growth rate

of our simulations is usually greater than zero, except for the temporary shocks in q we introduce. This can be easily verified by considering that the growth population rate is given by $1 - mq$ and that we assume individuals to have two children ($m = 2$). Our graphs show convergence to a fixed number as predicted by the GW process. Consider, for instance, the time series for the number of surnames in Figure B.1. This variable converges to a fixed number in area A from period 3 to 14, and then it is steady; and it is constant for area B.

We go further by testing the case of growth population rate equal to zero. We introduce a few simplifications in our simulations: one individual per surname, a null probability of migration ($p = 0$), and a probability of surviving and reproducing equal to 0.5 ($q = 0.5$). As $m = 2$, the growth population rate is zero. We assume a number of periods equal to 100. Figure B.9 shows the time series of our indicators. We obtain that the number of surnames converges to 1, as predicted by GW process.

Figure B.9: Galton-Watson process



Population Surnames Exclusive surnames

Notes: Simulation with baseline parameter values: $S_A(0) = 200$, 0 common surnames, 1 individual per surname, $q_A = 0.5$, $p = 0$, $m = 2$ and $T = 60$. The graphs show time series for population size, number of surnames and number of exclusive surnames.

Source: Simulation using the model presented in the section titled "The surname process".

4 Proofs of model's properties

Proof of property 2

Consider a surname $s \in \Omega_t^A$. We define $N_t^m(s)$ as the number of individuals with surname s migrating from A to B at t. Conditional on $N_t^A(s)$, $N_t^m(s)$ is a binomial random variable with support $[0, N_t^A(s)]$:

$$P\left(N_t^m(s) = k\right) = \binom{N_t^A(s)}{k} p^k (1-p)^{N_t^A(s)-k}$$

$$\text{where } \mathbb{E}_t\left[N_t^m(s)\right] = N_t^A(s)p \text{ and } \mathbb{V}_t\left[N_t^m(s)\right] = N_t^A(s)p(1-p).$$

Then:

$$N_t^m(s) = N_t^A(s)p + \epsilon_{st}$$

$$\text{where } \mathbb{E}_t\left[\epsilon_{st}\right] = 0 \text{ and } \mathbb{V}_t\left[\epsilon_{st}\right] = N_t^A(s)p(1-p).$$

Proof of property 3

The variable of interest $N_{t+1}^A(s)$, conditional on $N_t^m(s)$, is a binomial random variable with support $\left[0, \left[N_t^A(s) - N_t^m(s)\right] m\right]$:

$$P\left(N_{t+1}^A(s) = km\right) = \binom{N_t^A(s) - N_t^m(s)}{k} {q_A}^k (1-q_A)^{N_t^A(s) - N_t^m(s) - k}$$

$$\text{where } \mathbb{E}_t\left[N_{t+1}^A(s)\right] = \left[N_t^A(s) - N_t^m(s)\right] q_A m$$
$$\text{and } \mathbb{V}_t\left[N_{t+1}^A(s)\right] = \left[N_t^A(s) - N_t^m(s)\right] q_A(1-q_A)m^2.$$

Then:

$$N_{t+1}^A(s) = \left[N_t^A(s) - N_t^m(s)\right] q_A m + \mu_{st}$$

$$\text{where } \mathbb{E}_t\left[\mu_{st}\right] = 0 \text{ and } \mathbb{V}_t\left[\mu_{st}\right] = \left[N_t^A(s) - N_t^m(s)\right] q_A(1-q_A)m^2.$$

Proof of property 4

We present this proof on a case by case basis.

Exclusive surnames of area A

Consider a surname $s \in \Omega_t^A - \Omega_t^B$. Conditional on $N_t^m(s)$, $N_{t+1}^B(s)$ is a binomial random variable with support $[0, N_t^m(s)m]$:

$$P\left(N_{t+1}^B(s) = km\right) = \binom{N_t^m(s)}{k} {q_A}^k (1-q_A)^{N_t^m(s) - k}$$

$$\text{where } \mathbb{E}_t\left[N_{t+1}^B(s)\right] = N_t^m(s) q_A m \text{ and } \mathbb{V}_t\left[N_{t+1}^B(s)\right] = N_t^m(s) q_A(1-q_A)m^2.$$

Then:

$$N_{t+1}^B(s) = N_t^m(s) q_A m + w_{st}^A$$

$$\text{where } \mathbb{E}_t\left[w_{st}^A\right] = 0 \text{ and } \mathbb{V}_t\left[w_{st}^A\right] = N_t^m(s) q_A(1-q_A)m^2.$$

Exclusive surnames of area B

Consider a surname $s \in \Omega_t^B - \Omega_t^A$. Conditional on $N_t^B(s)$, $N_{t+1}^B(s)$ is a binomial random variable with support $[0, N_t^B(s)m]$:

$$P\left(N_{t+1}^B(s) = km\right) = \binom{N_t^B(s)}{k} {q_B}^k (1-q_B)^{N_t^B(s) - k}$$

$$\text{where } \mathbb{E}_t\left[N_{t+1}^B(s)\right] = N_t^B(s) q_B m \text{ and } \mathbb{V}_t\left[N_{t+1}^B(s)\right] = N_t^B(s) q_B(1-q_B)m^2.$$

Then:

$$N_{t+1}^B(s) = N_t^B(s) q_B m + w_{st}^B$$

$$\text{where } \mathbb{E}_t\left[w_{st}^B\right] = 0 \text{ and } \mathbb{V}_t\left[w_{st}^B\right] = N_t^B(s) q_B (1 - q_B) m^2.$$

Common surnames

Consider a surname $s \in \Omega_t^A \cap \Omega_t^B$. Conditional on $N_t^m(s)$ and $N_t^B(s)$, $N_{t+1}^B(s)$ is the sum of two binomial random variables g_1 and g_2, whose supports are $[0, N_t^m(s)m]$ and $[0, N_t^B(s)m]$, respectively:

$$P\left(g_1 = km\right) = \binom{N_t^m(s)}{k} {q_A}^k (1 - q_A)^{N_t^m(s) - k}$$

$$P\left(g_2 = km\right) = \binom{N_t^B(s)}{k} {q_B}^k (1 - q_B)^{N_t^B(s) - k}$$

$$\text{where } \mathbb{E}_t\left[g_1\right] = N_t^m(s) q_A p m \text{ and } \mathbb{V}_t\left[g_1\right] = N_t^A(s) q_A (1 - q_A) m^2,$$
$$\text{and } \mathbb{E}_t\left[g_2\right] = N_t^B(s) q_B m \text{ and } \mathbb{V}_t\left[g_2\right] = N_t^B(s) q_B (1 - q_B) m^2.$$

Then:

$$\begin{aligned} N_{t+1}^B(s) &= g_1 + g_2 \\ &= N_t^m(s) q_A m + N_t^B(s) q_B m + w_{st}^A + w_{st}^B \end{aligned}$$

$$\text{where } \mathbb{E}_t\left[w_{st}^A\right] = 0 \text{ and } \mathbb{V}_t\left[w_{st}^A\right] = N_t^m(s) q_A (1 - q_A) m^2,$$
$$\text{and } \mathbb{E}_t\left[w_{st}^B\right] = 0 \text{ and } \mathbb{V}_t\left[w_{st}^B\right] = N_t^B(s) q_B (1 - q_B) m^2.$$