Surnames, Genetic Selection, and Social Mobility in Pre-Industrial England

Gregory Clark, University of California, Davis
gclark@ucdavis.edu
January 19, 2009

A Farewell to Alms: A Brief Economic History of the World, argued controversially that in pre-industrial England the rich replaced the poor demographically, and that this helps explain why England became more "bourgeois" in these years: less violent, thriftier, more literate, more numerate. Here evidence from a different source, *surnames*, confirms the takeover of English society by the economically successful between 1600 and 1851, and the disappearance of the criminal and the poor. A man's economic success in pre-industrial England predicts a permanent increase of his surname frequency, and hence his gene frequency, by 1851. But the surnames also shows that pre-industrial England was a society of both downward and upward social mobility all the way from at least 1300, with no permanent upper class. In this respect it shows greater social mobility than modern societies such as the USA and Brazil.

Introduction: Surnames and Genetic Selection

A Farewell to Alms showed the selective pressures in pre-industrial English society in favor of the genes and culture of the economically successful, and against the genes and culture of the poor. This hypothesis has been controversial. Objections have included the idea that "regression to the mean" would mean that the children of the rich were little different from the general population, so that such selection could not change the average characteristics of the population.[1]

The current study shows evidence of selection from a completely different source, changes in rare surname frequencies over time. Rare surnames associated

[1] This argument is made by Bowles, 2007, and elaborated in McCloskey, 2008, and Pomeranz, 2008.

with rich men circa 1600 increased substantially in frequency relative to those associated with the poor and the criminal circa 1600.[2]

Surnames in pre-industrial England can be a measure of DNA frequencies because they propagated like the Y chromosome. They passed unchanged, except for mutations, from fathers to sons.[3] A recent study of 150 pairs of men in the modern Britain with a shared surname examined whether they had a common male ancestor in the patriline.[4] The study examined 17 markers which vary on the Y chromosome, a variation created long before the establishment of hereditary surnames in England around 1300. If two men share an ancestor in the male line in the recent past these markers would be identical on their Y chromosome, except for genetic drift. 16 of the 150 pairs showed identical markers. In another 20 pairs the markers were similar enough that the differences were probably due to genetic drift from a common ancestor in the patriline.

The probability of having a recent common male ancestor in the patriline was greater the rarer the name, even though the study deliberately avoided names held by less than 50 people in 1996, and excluded men known to be related. 15 of the 16 completely matched haplotype pairs were in the lower half of the name frequency distribution. Eight of the pairs of 15 least common names (50-186 occurrences in the population in 1996) showed evidence of a common male ancestor. This implies that for individuals with rare names in England there is a relatively high chance of an early common male ancestor in the male line. Surnames can serve as a proxy measure of selection of genetic types within pre-industrial England.

Here I identify two groups of rare surnames in England 1560-1640. The first was rare surnames held by economically successful men, as revealed by their leaving a will. The second group was rare surnames held by a man on the margins of society, someone indicted in the Essex courts in the years 1598-1620 for assault, burglary, theft, poaching, robbery and murder. The indicted were overwhelmingly from low socio-economic groups.

---

[2] I am grateful to Nicholas Wade of the New York Times for suggesting such a study as a test of the hypotheses of "survival of the richest."

[3] Large scale adultery, illegitimacy and adoption would break this connection between surnames and the Y chromosome. (Illegitimate children would typically bear their mother's surname). But in the seventeenth century England illegitimate births are estimated to be less than 2% of all births (Wrigley and Scofield, 1981). Adultery was thus likely also infrequent. Adoption was rare in pre-industrial England.

[4] King et al., 2006.

Table 1: Summary of the Results for England as a Whole

Group	Number of Rare Names 1560-1640	Median Occurrence 1841/51	Name disappeared by 1841/51 (%)
Indicted	337	27	21
Poorest Testators	159	70	15
Middling Testators	297	65	17
Richest Testators	206	115	8

For rare surnames a significant fraction of the holders will typically be related: brothers, cousins, second cousins. We know wealth and social status was strongly correlated between fathers, sons and brothers.[5] Thus the average man holding the same rare surname as a successful man in 1600 will be wealthy. The average man holding the same rare surname as someone indicted in 1600 will be poor. That is we can identify a subset of surnames where the typical holder was wealthy or poor in 1600.

As table 1 shows, the surnames of the rich of 1600 survived much better than those of the poor in the following 250 years. By 1851 there were at the median four times as many people bearing the surnames of the richest group in 1600 as those with the surnames of the indicted in 1600. But even among the rich, the richest, as would be expected from the results reported in *A Farewell to Alms*, had better reproductive success than the poorest.[6] The differential becomes even stronger when we concentrate on names held in by people in 1851 in the same geographic area as their ancestors, and most likely to actually be descendants of the man observed or his close relatives.

The implication is simple. Economic success by a man in 1600 substantially increased the share of their genes in the English gene pool by 1851. The genes of

[5] Clark, 2008.
[6] Clark and Hamilton, 2006.

the English in 1851 were composed disproportionately of those who succeeded economically in the pre-industrial era.

But it does not follow that pre-industrial society was divided into self-contained and persistent classes of the rich and the poor/criminal. Indeed the names evidence can also demonstrate that eventually the descendants of the rich and of the criminal, on average, converged to the same social status. "Survival of the richest" in pre-industrial England was compatible with strong social mobility.

Some of the hostility to the demonstration of "survival of the richest" in *A Farewell to Alms* seems to come from conflating two claims. The first, correct, claim is that the genes of the pre-industrial rich of any generation are overrepresented in the modern population. The second, incorrect, claim is that there was a persistent class of the rich in pre-industrial England, which eventually took over all the society through downward mobility.[7] While pre-industrial mobility was predominantly downward, there was also important upward mobility, as will be seen below.

I can also use surname evidence to show that social mobility even between 1300 and 1600 was complete also. Thus even in the middle ages pre-industrial England was a society of long run complete social mobility. In contrast surname evidence for the modern USA and Brazil suggests that there are persistent social classes.

The Method

In the region this study focuses on, the south of England and East Anglia, already by 1350 the majority of people had surnames (McKinley, 1990, 32). The process of adoption of surnames was supposedly largely complete here by 1300-1400.[8]

While forenames in early England showed limited diversity, surnames exhibited from the earliest years astonishing variation. The 56 million people in England and Wales in 2002 were using nearly one million distinct surnames,

[7] I confess to have implicitly made that conflation myself in *A Farewell to Alms*.

[8] Surnames emerged in part because of the limited variety in forenames. The four or five most common male and female first names covered the majority of people from the middle ages on. So surnames became essential to identification, especially in a commercial and mobile society like pre-industrial England.

750,000 of which were held by fewer than 5 people.[9] This implies that in 2002 about 3% of the English population had surnames held by less than 5 people.

This may stem in part from emigration, and the creation of new surnames, but the evidence of the 1851 census suggests that even then there was an enormous variety of surnames. In 2002 the top 40 surnames covered only 13.1% of the population of England and Wales. In 1851 the top 40 surnames covered exactly the same 13.1% of the English population. There has always been a very long tale of rare surnames possessed by small numbers of individuals.

We have a good measure of what surnames were rare in England in 1601-2 through two books documenting the occurrences of surnames in 964 parish registers in England in 1601 and 1602, about 10% of all English parishes (Hitching and Hitching 1910, 1911). Someone's surname only appeared in the parish registers only if they had their baptism, wedding, or burial in these years. Thus the average person in the course of an average lifespan of 35 years, would appear three times in the registers. This implies that these registers contained a 1.8% sample of English surnames in 1601-2, about 73,000 names.

If this was a true random sample of names, a name held by as few as 400 people in England in 1601 would have a 99.9% chance of showing up on the list. Surnames held by as few as 41 people would still have an even chance of appearing. Only rare names, almost all with less than 200 holders, would escape this sieve.

In practice names are clustered by parish so that the sieve provided by these parish lists is less fine. Some quite common names will not be excluded. The name "Emery," for example, is not excluded even though there were more than 3,000 Emerys in England by 1841. To control for the inclusion of some not very rare names in our sampled from 1600 I look at the *median* occurrence of the surname 250 years later (rather than the mean). This avoids giving undue weight to common names that slipped through. But the typical name not excluded will be held by very few people. The name *Spyltimber*, for example, which showed up among the indicted, and which had disappeared by 1841, was excluded since it appeared in a register in 1601.

Since surnames passed from fathers to sons, the number of descendants from each of these groups in 1841/51, the first English censuses which recorded individual names, can be estimated just from the numbers of people in the 1841

[9] http://www.taliesin-arlein.net/names/search.php

and 1851 censuses bearing these surnames.[10] The records of these censuses have been transcribed and formed into a commercial database.[11]

The census returns were hand written, and that handwriting can be difficult to read. This produces errors in estimates of name frequencies in each census, which become apparent when we compare the frequencies of rare names in the 1841 and 1851 censuses. Some of these vary in implausible in the intervening 10 years. For example, 47 "Combers" listed in the 1841 census database, but only 6 for 1851. Inspection of images of the original returns shows that the 1841 "Combers" were transcribed in 1851 as "Comber." To reduce the transcription errors I used the average frequency of names in 1841 and 1851.

Another problem in categorizing surnames is that English spelling was highly irregular before the nineteenth century. The same surname would have many different variants. Johnson in 1601-2 was spelled Johnson, Johnnsone, Johnsone, Johnsonne, Jonson, Jonsson, Jhonson. "e" was added promiscuously to the end of names, without seemingly affecting the pronunciation. "y" and "i" were often interchangeable. To control for this I checked for variant spellings of surnames in 1601-2 and 1841/51 in determining their frequency in 1600 and 1841/51. Thus, for example, if a name ended in –y, I also checked for the same stem ending in –ie and –ey. If the name had a "ck" I also checked it with only a "k". Spelling variants introduce more errors, but not errors that should favor the names of the rich versus the poor. We can check this, however, in our data by looking at the relative frequency of spelling variants versus the originally spelled name in the case of the rich and the poor. This will test whether the names of the rich somehow were more fixed in their original form because of their greater literacy.

Another source of error is the mutation of surnames over time.[12] Partly this can occur because of shifts in the way names are pronounced, leading to a later shift in spelling. Thus the wills and court records for 1600 show a ratio of

---

[10] Since illegitimate children in England bore the surnames of their mothers, illegitimacy will not be a barrier to this test. Thus greater illegitimacy rates by the poor and the indicted would not affect the outcome here, since offsetting any loss from children of them or their sons not bearing the surname will be illegitimate children of their daughters who will bear the surname.

[11] http://www.ancestry.co.uk/

[12] As an extreme example, the surnames Birkenshaw, Bircumshaw, Burkimsher, Burtinshall, Brigenshaw, Buttonshaw, Brackenshaw, Buttinger, and Bruckshaw all apparently stem from the place name Birkenshaw (McKinley, 1990, 55).

"Clarks" of various stripes of 6:1 with "Clerks." By the 1841 census there were 73,049 "Clarks" and only 835 "Clerks" a ratio of nearly 100:1. Some of the "Clerks" must have evolved to become "Clarks." (Presumably because the pronunciation of clerk in modern English is clark).

If the rate of mutation is the same for rich and poor then this will just introduce more error into the data. However if the names of the rich adhered to them more exactly than to the poor, because they were more frequently literate, then this will bias the results in favor of the rich. I test below for such a bias, and show that it does not appear to occur.

Rare Surnames of the Rich and Poor, circa 1600

I get a sample of rare surnames held by at least one rich man with 1560-1639 from a database of 2,445 wills probated in these years, mainly in the counties of Essex and Suffolk.[13] 689 of these men, 28%, had names which did not appear on the parish registers lists for 1601-2. We can further divide these testators with rare names into rich (bequest of £250 or more), middling (£25-250), and poor (£0-25), where wealth is measured in 1630s prices.

Those leaving wills represent the upper end of the social scale and asset distribution in pre-industrial societies. Identifying rare surnames held by a man in the poorest strata of the society in socio-economic terms is more difficult. Most tax lists for pre-industrial England identify the propertied. The civil and manorial court records again tend to identify individuals with property to transact or dispute. One place where the poor do show up, however, is in criminal indictments. As in modern societies those accused of theft, forgery, assault, riot, robbery, murder, and desertion were disproportionately the poor.

Table 2, for example, shows the distribution of the occupations of 494 men leaving estimated assets of at least £250 in England 1560-1640, compared to the distribution for 1,523 men indicted in Essex courts 1598-1620 for property crimes, assault and homicide. 54% of the indicted were classified as laborers or the equivalent, compared to 0.2% for the rich. Overall the bottom four social groups were 6% of the rich, 81% of the indicted.[14]

---

[13] Clark and Hamilton, 2006, describe how these data are constructed from the raw will transcripts.

[14] Those accused only of petty larceny were on average even lower in the social scale. 61% of them were laborers or the equivalent.

Table 2: Occupational Distribution of Rich Testators and the Indicted

Social Group	Fraction literate amongst all will makers	Bequest of £250 or more (%)	Indicted (%)
Gentry	0.94	17	2
Merchants/Professionals	0.88	8	1
Farmers/Yeomen	0.54	70	6
Traders	0.44	2	9
Craftsmen	0.43	2	13
Husbandmen	0.27	2	11
Laborers	0.17	0	54

For the reason that I am attempting to get a sample of the poorest and most violent, I excluded from this sample men indicted for what were crimes against regulations in restraint of trade or religious freedom: keeping an unlicensed alehouse, baking without license, erecting cottages on less than 4 acres of land, and recusancy.[15]

From this sample of 1,523 indicted men, we get 374 (25%) who have rare surnames, a similar percentage to that for the sample of will writers.

There is some overlap between rare names held by the indicted in this period and rare names held by will writers. This in part reflects some relatively common names escaping the parish register sieve. I thus use a second filter to form the final samples, which is to exclude from the wills sample any names found among the indicted, and from the indicted sample any names found among will makers.

[15] Recusants, those who refused to attend Church of England, tended to have upper class occupations. Since there were substantial numbers of recusants in these years an interesting parallel study would ask what their reproductive success was.

In the resulting smaller samples, whose numbers are reported in table 1, there are some names that occur more than once among both the indicted and the will writers. Sometimes these people are clearly related: brothers, or fathers and sons. But names which occur multiple times in 1600 also tend to show up in greater frequency in 1841/51, because they were more common all along. In table 1, and in the statistical tests below I include each occurrence of such names as an observation. This is done because otherwise the size of the initial sample matters in terms of the median frequency of the occurrence of names later. Smaller samples will contain proportionately more common names, and have higher median numbers later. Since we have unmatched sample sizes this is undesirable.

Table 3 shows a random sample of 10% of the names of the indicted and of 5% of the names of the rich, constructed by arranging them in alphabetical order and selecting each 10^{th}, or 5^{th}, name. There is nothing evident from this list that would suggest why the names on the second list would be far more common by 1841. The appendix gives a complete list of the names of each of these groups and their frequency by 1841/51 in order of frequency.

Table 4 shows the number of spelling variants of names from each group that were tested against the 1841/51 censuses, and the fraction of the matches that resulted from a spelling variant. This is a check on whether the names of the rich adhered more exactly to them because the rich were literate. The names of the poor were written down by sound by clerks. They might thus be liable to change more quickly over time. But if this was happening then we would expect that a smaller fraction of the matches for the poor would be to the original exact spelling as the name mutated away from that original. We see in table 4 no sign of such a pattern.

Table 3: A Random Sample of Names of the Indicted and the Rich

Names of the indicted	Names of the Rich
Abstan	Aldham
Banbricke	Ayliffe
Bittin	Base
Bradwyn	Birle
Cabwell	Breame
Cheveney	Bynder
Cockle	Cobbold
Creame	Coventry
Cutmore	Danbrook
Drinckall	Fatter
Elvis	Folkes
Fossett	Gatteward
Gillham	Godbold
Gullyes	Gooch
Heditche	Hazell
Hownell	Hunringdon
Kenwood	Ilger
Los	Kingsberie
Meese	Libbis
Mounson	Maynerd
Nouthe	Negus
Osteler	Overed
Pennocke	Playfere
Pollen	Raynberde
Reddyforde	Rosington
Sache	Scolding
Segrave	Spatchet
Shurly	Tokelove
Sticinger	Upston
Terlynge	
Thurland	
Uphavering	
Wendham	
Wrothman	

Table 4: Was Name Mutation Commoner with the Indicted?

Group	Rare names 1560-1640 (number)	Spelling variants of names tested	Fraction of tests variants	Fraction of 1841/51 matches to variants
Indicted	337	349	0.51	0.40
Poorest Testators	159	176	0.52	0.39
Middling Testators	297	317	0.52	0.42
Richest Testators	206	213	0.51	0.37

Surname Survival by Group

Table 1 shows the results for these various samples of rare names. For those indicted 21% of the names had disappeared by 1841, implying that at least a fifth of these men had no legitimate patrilineal male descendants. For the richest men the fraction of names disappearing was only 8%. For the indicted the median frequency of names by 1841/51 was only 27. Since population by 1841/51 was more than four times that of 1601, on average every name frequency should have quadrupled. Thus unless the median name in this sample was held by only 7 or fewer people in 1601, the median numbers of people bearing these names was declining as a share of the population.

To test the statistical significance of the median differences reported in table 1 I carry out two tests. The first looks just at the differences in the medians, and is a non-parametric test of the hypothesis that two samples were drawn from a distribution with the same median. Table 5 shows the results of this test for each of the four samples. The table reports the probability that the medians of the groups in the row and column are the same. These results indicate that the chances that each of the three wills samples have the same median as the indictments sample varies between 3 in 1000 and less than 0.5 in 1000. We cannot reject with any confidence, however, the hypothesis that the median was the same across all wealth levels of those leaving wills.

Table 5: Difference of Medians Test

	Indictments	Wills-Poor	Wills-Middle	Wills-rich
Indictments	-	0.003	0.001	0.000
Wills-Poor		-	0.92	0.12
Wills-Middle			-	0.37
Wills-rich				-

Table 6: Difference in Distributions - Rank Test

	Indictments	Wills-Poor	Wills-Middle	Wills-rich
Indictments	-	0.0021	0.0000	0.0000
Wills-Poor		-	0.59	0.12
Wills-Middle			-	0.25
Wills-rich				-

The second test, that of Mann and Whitney, looks not just at the medians, but the whole rank of the observations. This tests not just the median, but whether the samples are from populations with the same distribution of values. Table 6 shows again that this test rejects even more strongly the possibility that the distribution of frequencies for the names of the indicted in 1841/51 is the same as that for any of the will samples. For the rich versus the indicted, for example, there are less than 0.5 chances in 10,000 that these samples were drawn from the same distribution. But again there is only weak evidence that the distribution of the wills of the rich is any different than that of the middling testators or the poor testators.

Regional Analysis

Though there was mobility in the English population in the pre-industrial era, people holding rare surnames in 1841/51 who were genetically related to those we observe circa 1600 would tend to live close to their ancestors. The data for the indicted is taken from Essex, and most of the wills come from Essex or the adjacent county Suffolk. Figure 1 shows these counties (Essex is 12, Suffolk 32). In another test of survivorship I thus define an area called the "South-East" that includes these and the adjacent counties – Norfolk (23), Cambridge (4), Hertford (16), Middlesex (22), Surrey (33) and Kent (18) – as indicated on the map.

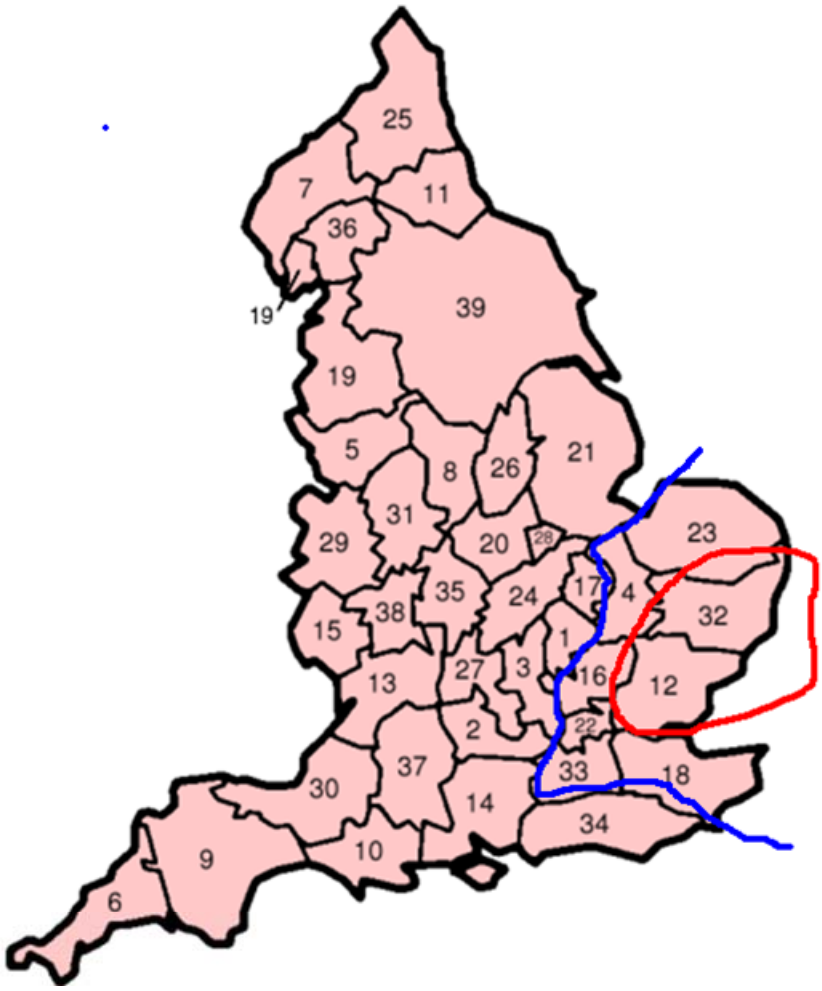
Surrey was included even though it is not contiguous to Essex, because the big destination of out migration of people from Essex and Suffolk before 1841 was the London area, part of which lay south of the river Thames in Surrey. In 1841 these counties had 28% of the population of England.

Under the hypothesis is that the differential survival and spread of rare surnames by the rich of 1600 is caused by the differential reproductive success of groups of people genetically related then this effect should be stronger if we concentrate on the South-East. By doing that we will be concentrating on the people in 1841/51 most likely to be actually related as opposed to be related by orthographic accident.

Table 7 shows the results for the medians and number of zeros for each group in the South-East in 1851. The differences between the indicted and will makers is now more marked than in table 1. The median number of occurrences of the names of the rich by 1851 is more than 7 times as great as for the indicted in the South-East (compared to a ratio of 4:1 for the country as a whole). This is because the fraction of the rare names for the indicted showing up in the South-East is much smaller than for any of the groups of will makers.

In contrast in the country outside the South-East the difference in name occurrence by 1851 between the will makers and the indicted, while still present, is greatly muted. Rare names of the rich show only twice the median number of occurrences as the rare names of the indicted. Table 8 shows these results.

Figure 1: English Counties in 1841



Note: This map is reproduced from
http://en.wikipedia.org/wiki/Historic_counties_of_England.

Table 7: Summary of the Results for the South East

Group	N South-East	Fraction of names 1851 in South East	South-East Median Occurrence 1851	Name disappeared by 1851 (%)
Indicted	337	0.46	9	35
Poorest Testators	147	0.62	36	21
Middling Testators	289	0.62	48	19
Richest Testators	204	0.67	67	17

Table 8: Summary of the Results for the rest of the Country

Group	N South-East	Fraction of names 1851 outside South East	Median Occurrence 1851	Name disappeared by 1851 (%)
Indicted	337	0.54	9	33
Poorest Testators	147	0.38	19	24
Middling Testators	289	0.38	22	24
Richest Testators	204	0.33	20	20

Social Mobility – Rare Names

Free market economists such as Gary Becker have argued that within a few generations social mobility is complete within modern societies. Thus:

> *Almost all earnings advantages and disadvantages of ancestors are wiped out in three generations. Poverty would not seem to be a "culture" that persists for several generations (Becker and Tomes, 1985, S32).*

This argument is based on the logic that since

$$ln(y_1) = bln(y_0) + u_1 \qquad b \approx 0.4$$

where y_1 is the income of the first generation, and y_0 the income of the original generation (mean 1), and u_1 a random error, then the income of the n^{th} generation will be

$$ln(y_n) = b^n ln(y_0) + u^*_n \approx u^*_n$$

where u_n^* is a cumulative error term with mean 0. b^n becomes very small very quickly if $b \approx 0.4$ as in modern societies. With this simple logic there are no persistent classes in modern societies, and equality of opportunity for all in the long run at least.

However, there is evidence for at least the USA and Brazil in the modern era that there are indeed, despite this logic persistent social classes. Thus if we look at one indicator of high socio-economic status in 1995, inclusion in *Who's Who*, there were 4.7 times as many people included with a surname that indicates Jewish patrilineal ancestry than there were such people in the general population. The great majority of the modern Jewish population in the USA arrived between 1860 and 1914, nearly three generations before 1995. Thus the Jewish population cannot be experiencing the general regression to the mean posited in this study. In contrast, those who identified as black were 0.15 times as likely to be included in *Who's Who* compared to their share in the general population. Those with surnames indicating a Scandinavian patrilineal ancestor were also overrepresented in *Who's Who*, though by much smaller margins (1.35 times as likely). The overwhelming majority of Scandinavian emigrants arrived between 1840 and 1914, so this again implies significant intergenerational persistence of economic and social status. Thus despite the overall evidence of substantial regression to the mean in incomes in the USA there are clearly subgroups both above and

below the mean who are not regressing as would be expected to the mean income level.[16]

In Brazil, similarly, in 1973 the list of prominent people in the society, indexed by appearing in the Brazilian version of *Who's Who*, was 8.8 percent of people with a German patrilineal ancestor. This despite the fact that the estimated numbers of those with German patrinlineal ancestry constituted only 2 percent of the population. German immigration to Brazil began in 1818, so that the majority of the population of German's in 1973 had arrived in the country 2-5 generations earlier. Yet somehow they had remained or become an elite group within Brazilian society.[17]

A second source showing surprising maintenance of ethnic differences in economic success are the questions on the 1990 and 2000 US censuses which ask people to state their ancestry. For example, those who stated Japanese ancestry in 1990 had average income 45 percent greater than average for the US.[18] Yet significant Japanese immigration to the US largely ended in 1907 with the "Gentlemen's Agreement" between the US and Japanese governments to restrict immigration to the female spouses of migrants already in the USA.

Using the information on rare surnames in 1600 and 1841/51 I can check the degree of social mobility in England in these years. As well as name frequencies, the 1851 census also supplies occupations by surname for men. Thus I can compare the average occupational status of the rare surname groups in both 1600 and 1851. Table 2 above shows that the rich were concentrated in high status occupations. 85% were listed as gentlemen, merchants, professionals, or land owning farmers (yeomen). In 1600 the indicted in contrast were overwhelmingly from lower-status occupations. Only 9% were in these higher status occupations.

How do the descendants of these two groups look in terms of socioeconomic status by 1851? Surprisingly there seems to be almost complete regression to the mean. Table 9 shows some measures of the socioeconomic status for a sample of both groups. While those descended from the rich show a slightly greater percentage in the top socio-economic groups, that result may well be sampling error. And at the bottom of the socio-economic scale, there are

---

[16] McDermott, 2002, 147.

[17] Nicholas and Snyder, 1981, 326.

[18] Darity et al., 1996. A problem with this measure, however, is that those with knowledge of their ancestry, particularly among the white population, are the more educated. Less educated white Americans are more likely to respond "American" to the ancestry question.

Table 9: Socioeconomic Status by Surname History, 1851

Status, 1851	Rich in 1600	Indicted in 1600
"Gentry/Professionals" (%)	6.1	4.1
"Farmers" (%)	4.7	3.7
"Laborers" (%)	31.5	28.6
Number in Sample	278	294

more of the descendants of the rich among "laborers" than there are descendants of the indicted.

If we compare these results to occupational distributions of England as a whole we find both groups have regressed to the mean. They are indistinguishable from each other and from the population as a whole. This implies both great downward mobility among the descendants of the rich, and modest upward mobility among the descendants of the indicted. The fraction of the descendants of the indicted who were among the lowest social group, the laborers actually declined from 54% circa 1600 to 29% in 1851.

The regression to the mean of both groups also shows up in the change in frequency of our rare surnames between 1841 and 1851. At this time English population as a whole grew by 12.7%. The rare surnames characteristic of the indicted of 1600 increased in median frequency from 26 to 29, a gain of 12%. The rare surnames characteristic of the will writers increased in median frequency from 79.5 to 89, a gain once more of just 12%. So by 1841 the reproductive success of these descendants of the lower and upper classes of 1600 was indistinguishable, and also indistinguishable from the general population.

Social Mobility – Common Names

The above results with rare names suggest complete social mobility between 1600 and 1850. I can also examine the degree of social mobility between 1200-1300, when surnames were first formed, and 1600, this time using common

names. A class of common surnames in England are those denoting occupations: Smith, Wright, Butcher, Baker, Clark and so on. Many of these occupation names denoted what were originally moderately high status occupations. These names became attached to some original male ancestor sometime between 1100 and 1300 when surnames were first formed in England in the Middle Ages.

We can group English occupation names into a rough socio-economic hierarchy by using the connection between actual occupations and wealth revealed by the wills circa 1560-1800. Table 10 shows this mapping. People describing their occupations in 1560-1800 as drapers, merchants or mercers, for example, had about ten times as much in assets as those describing their occupation as chapman. All of the men with any of these occupations were significantly more wealthy than men whose occupation was laborer. But they generally fell in the middle of the wealth range.

Assuming that this occupational hierarchy held also in the Middle Ages we can then ask what happened to the men in 1600 who had an ancestor with one of these higher status occupations, as witnessed by their surname. How were they distributed between those two extremes of the social hierarchy – wealthy testators and the indicted?

Table 11 shows this distribution. There are signs that these men who had an ancestor somewhere in the middle of the social hierarchy had moved almost equally upward and downward in social ranks. 9 percent of the rich had surnames from these occupational groups, but so did 7.2 percent of the indicted. This implies almost complete social mobility over the course of only 300-400 years. All the more remarkably these 300 to 400 years included a substantial period which was the later Middle Ages, which many presume was a society that consisted only of a super rich lordly class and a servile and oppressed peasantry.

Pre-industrial England all the way from 1300 to 1850 was thus a very fluid society, with families moving up and down the social scale across each generation. There was no persistent upper class or lower class in this society. In contrast the evidence for the US suggests that even if we looked over periods of hundreds of years we would find incomplete convergence to the mean by both the descendants of those at the top and the bottom of the income distribution.

Table 10: Occupations and Assets circa 1600

Type of Occupation	Associated Surnames	Ave Asset Incomes, all wills 1560-1800 (£)
Merchants	Draper, Merchant, Mercer	91.4
Manorial Officials	Aylward, Butler, Chamberlain, Hayward, Reeve, Steward	-
Retailers	Baker, Butcher, Chandler, Cook	23.7
Artisans	Smith, Wright, Cooper, Mason, Tiler	17.8
Clark	Clark	16.1
Minor artisans	Barker, Fuller, Potter, Tanner, Taylor	12.7
Petty Retailers	Chapman	9.4
Laborers	-	6.4

Table 11: Occupational Name Shares among the Rich and the Indicted, c. 1600

Occupation Names	Wills (>£250) 1590-1639 (%)	Indicted 1598-1620 (%)
Merchants: Draper, Merchant, Mercer	0.25	0.11
Manorial Officials: Aylward, Butler, Chamberlain, Hayward, Reeve, Steward	0.88	0.33
Retailers: Baker, Butcher, Chandler, Cook	1.88	1.25
Artisans: Smith, Wright, Cooper, Mason, Tiler	2.75	3.41
Clark	0.63	0.70
Minor artisans: Barker, Fuller, Potter, Tanner, Taylor	2.25	1.03
Petty Retailers: Chapman	0.38	0.38
ALL	9.02	7.21
N	800	1,845

Interpretation

The surname evidence confirms a permanent selection in pre-industrial England for the genes of the economically successful, and against the genes of the poor and the criminal. If someone was economically successful in England in 1600 then their surname is more common than would be expected in 1841/51. There genes are thus more common than would be expected in 1841/51. For the indicted the reverse holds. There extra reproductive success had a *permanent* impact on the genetic composition of the later population.

Both Samuel Bowles and Deirdre McCloskey have objected to the possibility of such effects being significant on the grounds of regression to the mean in the

children of the rich and the criminal.[19] This, combined with mating that is only imperfectly assertive, means that within a modest number of generations the descendants of successful males will be indistinguishable from the general population. Between 1600 and 1841/51 there are roughly 7 generations. An original male founder would contribute 1/128 of the genetic material of his descendants by the 7th generation. The direct genetic effect the abilities or criminality of the man we observe in 1600 would have on his descendants of 1841 would thus be minimal, even allowing for some assortative mating. Tables 9 and 11 seem to bear out this prediction.

I am not denying the impact of regression to the mean. The names evidence indeed suggests that there is no persisting upper class even in societies like pre-industrial England.

But while this is correct, it does not follow that "survival of the richest" did not permanently change the nature of the population. Indeed the names evidence demonstrates that it did. For before we get to the long-run where the descendants of rich and poor are largely indistinguishable, the descendants of prosperous men gain a permanent advantage in numbers that is never erased by the later regression of the characteristics of their descendants to the mean of the population.

Each economically successful male in pre-industrial England contributed, on average, more of the genetic material of the population stock in 1851 or 2008 than each economically unsuccessful or criminal male. "Capitalist" genes were thus increasing as a share of the population over time. Regression to the mean does not imply that the mean characteristics of the population cannot change over time – in that case we would end up with a reduction ad absurdum that would prove all evolutionary change impossible.

As long as regression to the mean takes a few generations for rare names one economically successful generation will have statistically a permanent and significant effect on the later frequency of the name, and similarly one criminal or indolent generation. As long as this process was repeated in each generation for a new set of economic winners and losers, the characteristics of the population would permanently change in favor of the characteristics that make people rich, and against the characteristics that make them poor or criminal.

The study of wills reported in *A Farewell to Alms* implied that economic competition could change the genetic composition of the English population over

[19] Bowles, 2007, McCloskey, 2008.

time. This study of rare surnames shows that indeed economic success in 1600 by a man could permanently increase the relative frequency of his surname, and by implication of his genes. This does not demonstrate that these genetic changes had significant impacts in changing the behavior of the average person in England by 1800. But Clark (2008) shows that economic success in modern societies has at its roots a significant genetic component.

References

Becker, Gary S. and Nigel Tomes. 1986. "Human Capital and the Rise and Fall of Families." *Journal of Labor Economics*, *4(3), Part 2*: S1-S39.

Bowles, Samuel. 2007. "Genetically Capitalist?" *Science, 318* (Oct 19): 394-5.

Clark, Gregory. 2007. *A Farewell to Alms: A Brief Economic History of the World.* Princeton: Princeton University Press.

Clark, Gregory. 2008. "In Defense of the Malthusian Interpretation of History." *European Review of Economic History*, *12(2)* (August).

Clark, Gregory and Gillian Hamilton. 2006. "Survival of the Richest. The Malthusian Mechanism in Pre-Industrial England." *Journal of Economic History***,** *66*(3) (September): 707-36.

Cockburn, J. S. 1978. *Calendar of Assize Records, Essex Indictments, Elizabeth I.* London: Her Majesty's Stationary Office.

Cockburn, J. S. 1982. *Calendar of Assize Records, Essex Indictments, James I.* London: Her Majesty's Stationary Office.

Darity, William Jr., David K. Guilkey, and William Winfrey. 1996. "Explaining Differences in Economic Performance among Racial and Ethnic Groups in the USA: The Data Examined**."** *American Journal of Economics and Sociology, 55(4):* 411-425.

Hitching, F. K. and S. Hitching. 1910. *References to English Surnames in 1601.* Walton-on-Thames: Charles Bernau.

Hitching, F. K. and S. Hitching. 1911. *References to English Surnames in 1602*. Walton-on-Thames: Charles Bernau.

Jobling, Mark A. 2001. "In the name of the father: surnames and genetics." *TRENDS in Genetics, Vol. 17, No. 6* (June): 353-357.

King, Turi E., Stéphane J. Ballereau, Kevin E. Schürer, and Mark A. Jobling. 2006. "Genetic Signatures of Coancestry within Surnames." *Current Biology, 16* (21 February): 384-388.

McCloskey, Deirdre. 2008. "You know, Ernest, the rich are different from you and me: A comment on Clark's *Farewell to Alms*." *European Review of Economic History*, *12(2): ---*.

McDermott, Monica. 2002. "Trends in the Race and Ethnicity of Eminent Americans." *Sociological Forum, 17(1)*: 137-160.

McKinley, R. A. 1990. *A History of British Surnames.* London: Longmans.

Nicholas, Glenn A. and Philip S. Snyder. 1981. "Brazilian Elites and the Descendants of the German, Italian, and Japanese Immigrants." *Journal of Interamerican Studies and World Affairs, 23(3)*: 321-344.

Pomeranz, Kenneth. 2008. "Featured Review: *A Farewell to Alms*" *American Historical Review*, 113(3) (June): 775-779.

Wrigley, E. A. and R. S. Schofield. 1981. *The population history of England, 1541-1871 : a reconstruction.* Cambridge: Cambridge University Press.

Appendix

Below are listed in order of frequency in 1841/51 the rare surnames of the indicted and the rich (with the average frequency in brackets). Where a name appeared more than once in each sample that is indicated by a superscript giving the number of observations. The most common names on this list by 1841/51 were held by less than 0.01% of the population.

Rare Surnames of the Indicted

Abstan (0), Adrinon (0), Adyen (0), Allegant (0), Berdsell (0), Caboule (0), Cabwell (0), Callingswood (0), Carrudder (0), Cheveney (0), Chopan (0), Cleefes (0), Clovell (0), Clovile (0), Culpacke (0), Cunsden (0), Cuppledike (0), Curtopp (0), Derryfall (0), Drakwood (0), Eatney[2] (0), Eggesfield (0), Fawchett (0), Filbrick (0), Fitzgarratt (0), Fromfairefield (0), Furbench (0), Gannocke (0), Girord (0), Golesman (0), Gynnericke (0), Hewthett (0), Hinckhorne (0), Homsfield (0), Johnjohn (0), Kyttar (0), Lygeatt (0), Malbroke (0), Marborow (0), Michaelfield (0), Nynnam (0), Olster (0), Pafelyn (0), Pennoll (0), Penyall (0), Pettiepoole (0), Quanterell (0), Sansham (0), Sawdry (0), Selfscall[2] (0), Selscall (0), Sheepbotham (0), Slaterford (0), Spratborowe (0), Sticinger (0), Straunge (0), Strechie[2] (0), Surbote (0), Totnam (0), Uphavering (0), Vynold (0), Wakeringe (0), Whitekyrtle (0), Withar (0), Wrotheram (0), Wrothman (0), Wuthers (0), Wysbiche (0), Yecupp (0), Colwye (0.5), Littoll (0.5), Murcock (0.5), Offington (0.5), Pamphelyn (0.5), Pickroft (0.5), Toyse (0.5), Twyers (0.5), Wendam (0.5), Dudsbury (1), Frunt (1), Glyberie (1), Harridance (1), Pypall (1), Wystocke (1), Banbricke (1.5), Jeffarye (1.5), Mosier (1.5), Mounck[4] (1.5), Selon (1.5), Thimble (1.5), Walgrave (1.5), Yarrett (1.5), Blossom (2), Mounson (2), Ridland[2] (2), Sawkyn (2), Brockas (2.5), Claysbye (2.5), Cocksett (2.5), Lydcott (2.5), Romball (2.5), Terlynge (2.5), Inifer (3), Oath (3), Ole (3), Nouthe (3.5), Shatbolt (3.5), Gullyes (4.5), Pecham (4.5), Saffold (4.5), Warnor (4.5), Grynhill (4.5), Snellock (5), Dason (5.5), Dowdale (5.5), Goldingham (5.5), Bittin (6), Clanford (6), Dednam (6), Gunvyll (6), Hinnis (6), Hownell (6), Seckington (6), Bardney (7), Gervase (7.5), Thurger (8), Heditche (8.5), Worrett (8.5), Theedam (9), Strachie (9.5), Hovill (10), Elleott (10.5), Elrick (10.5), Fellford (10.5), Mullox (10.5), Jurdan (11), Paken (12), Hoyton (13.5), Rombold (13.5), Brussell (14), Chittam (14.5), Bickner (15.5), Earlinge (15.5), Reddyforde (16), Bradwyn (16.5), Pontifex (18), Chatwell (19.5), Paulter (20), Nowlinge (20.5), Byrchnall (21), Glydewell (21.5), Lawten (21.5), Halpeny (23.5), Tewse (23.5), Pordage (24), Combers (26.5), Stubben (26.5), Handler (27), Fromant[2] (29), Thurland (29.5), Boath (30), Los (30), Trowton (31), Adwicke (32), Offyn[2] (33), Tunge (34), Serritt (36), Blighton (36.5), Staughton (36.5), Backen (37), Newyn (37), Eminge (39), Stanwood (40), Duche (42.5), Catmore[2] (43), Hye (47.5), Benefield[2] (49), Dunse (50.5), Stidman (52.5), Gyllian (58), Marleborrowe

(58), Tynge (60.5), Alvyn (63), Elvis (65), Marryan (68), Marty (70.5), Meese (71), Creame (72.5), Forby (74), Boreman (82), Moxley (82.5), Vere (83), Croxon (83.5), Pollen (84), Armond (86.5), Thredder (89), Pecker (89.5), Kenwood (93), Raffe (94.5), Okeman (95), Bushie (99), Mullock[2] (99), Cremer (99.5), Laman (100), Pleasante (106), Clithero (109), Tytman (109.5), Cadge (113), Hunley (116), Stammer (118.5), Garnsey (120.5), Petchie[2] (121), Samford (122), Sunman (125), Lummys (125.5), Shurly (126.5), Tarver (136.5), Curryer (139), Sames (140.5), Sache[2] (141), Rond (141.5), Liget (144), Fannynge (144.5), Fossett (144.5), Deeringe (146.5), Curbye (148.5), Drinckall (148.5), Muche (154), Patient (161), Treherne (162.5), Carewe (167), Curtyn (172), Hackley (176.5), Ratley (182.5), Saward (191), Bundocke (195.5), Pawlin (198.5), Devenishe (205.5), Lindsell (206), Wooddy (213), Tier (222.5), Luce (223.5), Bindley (225), Woofe (229), Bycroft[2] (238), Fernes (238), Woodthorpe (241.5), Waterfall (251.5), Cranford[3] (252), Boker (254), Plaile (254), Cockerton (260), Cockle (261.5), Garlinge (261.5), Roose (269.5), Cakebred (287.5), Cowland[2] (292), Dearman (292.5), Berysford (302), Vynson (305.5), Borley (310), Shadbolt (310), Segrave (314.5), Sells (317), Woolsey (320), Cutmore (322), Motley (325.5), Hornsey (327.5), Hollowell[2] (332), Enys (341.5), Hatten (359), Merell[2] (360.5), Tubbs (362), Carder (378.5), Albert (385), Hewer (394), Kidman (398.5), Pennocke (409), Osteler[2] (409.5), Powe (424.5), Pynnocke (433), Rudland (445), Stebbinge[2] (474.5), Grout[2] (477), Boreham[2] (528.5), Munt (530), Rankin (530.5), Pidgeon (545), Botting (553), Greenhill (614), Rootes (615), Wakelyn (649), Burchall (730.5), Keeley (748), Whitney (757.5), Thurgood[4] (784), Kirkland (812), Harlowe (835.5), Gillham (952), Cracknell (1,047.5), Seeres (1,096.5), Knapp (1,106.5), Adkyns (1,336.5), Hynes (1,447), Denham (1,524.5).

Rare Surnames of the Rich

Antleby (0), Arwaker (0), Brighthall (0), Bundich (0), Dirifall (0), Downsdale (0), Glamfield[2] (0), Glozer (0), Harlakenden (0), Monnynges (0), Peperton (0), Salthorne (0), Selsden (0), Tovill (0), Typtott (0), Whitnam (0), Grenling (0.5), Hoxon (0.5), Innold (0.5), Leffingwell (0.5), Mawndry (0.5), Convers (1), Enyver (1), Ignes (1), Shawbery (1.5), Benold (2), Berriff (2.5), Hursteler (3), Mellsopp[2] (3), Ridnall (3), Damron (3.5), Gages[2] (4.5), Palsey (4.5), Pickys (4.5), Rowninge (4.5), Jower (5), Tokelove (5.5), Baas (6.5), Hompstede (7.5), Maynerd (7.5), Budley (8), Chacer (8), Coggeshall (8), Popley (8), Ilger (8.5), Fatter (10), Marcall (11.5), Bulbrooke (14.5), Gosnold (15), Spatchet (15), Drywood (15.5), Sandcroft (16), Barlyman (17.5), Westhropp (17.5), Keagle (18), Roath (21), Kingsberie (22.5), Casborne (24), Danforth (24), Libbis (24), Danbrook (25.5), Overed (26.5), Raynberd[2] (27), Playfere (27.5), Pitches (28.5), Derslye (29), Scolding (30), Birle[2] (30.5), Flowerdew (31), Banoke (38), Turnidge[2] (38.5), Berker (45.5), Scotchmere

(57.5), Gilbard (59.5), Clodd (60), Huntingdon (61.5), Soame (64.5), Traye (64.5), Spencely (68), Tillott (70.5), Huggon[2] (71.5), Faulke (73.5), Rutterford (80), Verdon (82), Rosington (84), Goldson (86), Manthorpe (91), Upston (91.5), Leaguy (95), Wyard (95.5), Bloyse (96.5), Cheesewright (100.5), Goymer (103.5), Aldham (111), Wace (111.5), Whiter (115), Soane (115.5), Stonham (116.5), Raneham (119.5), Riseing (124), Revett (124.5), Beart (129), Breame (129), Brother (130), Oxborowe (137), Pennyng (140.5), Base (147.5), Grimwade (152), Gatteward (159), Blosse[2] (159.5), Shale (161.5), Clench (163), Debnam (163.5), Bobbett (167.5), Letton (176.5), Hagon (190), Culham (193), Bridon (195.5), Hovell[2] (199.5), Buckenham (201), Daynes[2] (201.5), Bynder (207), Brille (213.5), Bardwell (218), Hammand (219), Wyeth (220.5), Punchyarde (222), Felgate (234), Denington (237), Boycott (240), Meene[2] (245.5), Lany (253), Cobbold (262), Jaggard (265.5), Noblett (266), Crowne (267.5), Rosier (275), Ayliffe (278.5), Greengrasse (282.5), Godbold[3] (293), Bunnyng (310.5), Marvyn (311), Firman (324), Folkard (333), Folkes[2] (344.5), Botwright[2] (356.5), Pawsey (372.5), Burlynge (373.5), Hurrey (381), Voyce (381), Jenney (401), Copsey (415), Syer (441), Kingsbury (447.5), Hynson (489), Clover (499), Rackham (514), Fincham (537), Coventry (544.5), Everard[2] (550.5), Negus (558.5), Sheldrake (564.5), Biles (633.5), Aldous[3] (644.5), Copping[3] (729.5), Welton (818.5), Creasey (887.5), Canham (953.5), Noone (980), Ryxe (995), Thoebald (1,000.5), Pett (1,086), Ryece (1,103.5), Keble (1,103.5), Starling (1,301.5), Mace (1437.5), Mayhew (1,481.5), Newson[3] (1564.5), Hazell (1,656.5), Gooch[3] (1,657), Buntinge (1,926.5).