# The Longevity of Famous People from Hammurabi to Einstein
# Online Appendix

David de la Croix
IRES and CORE, Université catholique de Louvain

Omar Licandro
IAE-CSIC and Barcelona GSE

May 28, 2015

**Abstract**

We build a new sample of 300,000 famous people born between Hammurabi's epoch and Einstein's cohort, including their vital dates, occupations, and locations from the Index Bio-bibliographicus Notorum Hominum. We discuss and control for selection and composition biases. We show using this long-running consistent database that there was no trend in mortality during most of human history, confirming the existence of a Malthusian epoch; we date the beginning of the steady improvements in longevity to the cohort born in 1640-9, clearly preceding the Industrial Revolution, lending credence to the hypothesis that human capital may have played a significant role in the take-off to modern growth; we find that this timing of improvements in longevity concerns most countries in Europe and most skilled occupations.

JEL Classification Numbers: J11, I12, N30, I20, J24.
Keywords: Longevity, Notoriety, Malthus, Elite, Compensation Effect of Mortality, Enlightenment, Europe.

# A Lifespan Precision

To measure the quality of the individual lifespan data, in this section we show two different statistics: the frequency of observations with imprecise vital dates and the heaping index.

The IBN adds the indications "c.", for *circa*, or "?" to the vital dates when the years of birth or death are not known with certainty. It may also be that more than one date is reported. We retained all the imprecise observations (taking the mean if there was more than one date), but created a discrete variable called *precision*, allocating a value of one when the lifespan was imprecise, zero otherwise. Figure A.1 shows the fraction of imprecise observations by decade. Individual lifespans measured by the IBN were highly imprecise until the end of the Middle Ages; the degree of imprecision then moves to zero as the sample reaches the 19th century.

When vital data are not known with certainty, biographers (or concerned persons themselves) often approximate them by rounding the year of death or birth to a number finishing in 0 or 5. Moreover, in the particular case of famous people, for obvious reasons, years of birth are likely to be more uncertain than years of death. The heaping index measures the frequency of observations with vital dates finishing in 0 or 5; it is commonly normalized by multiplying the frequency by 5. A heaping index close to unity shows that the vital data are very precise. Figure A.2 shows birth and death heaping indexes by decades up to 1879.[1] The death date heaping index is low, indicating that the dates of death of famous people were well known. Birth dates were much more uncertain, as the heaping index is about 2.5 before 1450, indicating that there are 2.5 times more dates finishing in 0 or 5 than there should be. Improvements in the birth year heaping index seem to start around 1450. This observation is consistent with the findings of De Moor and Zuijderduijn (2013) that numeracy levels among the well-to-do in the early modern period were very low (in the Netherlands). By 1700, the gap between birth and death heaping has decreased and both indexes fluctuate around one.

If, following A'Hearn, Baten, and Crayen (2009), we interpret the age heaping index as a measure of human capital (consistently with the robust correlation between age heaping and literacy at both the individual and aggregate level), our findings support the hypothesis that there was a major increase in human capital preceding the industrial revolution.

[1] Notice that heaping has no sense before 800, when the dating system starting at the birth of Jesus of Nazareth became widely used.

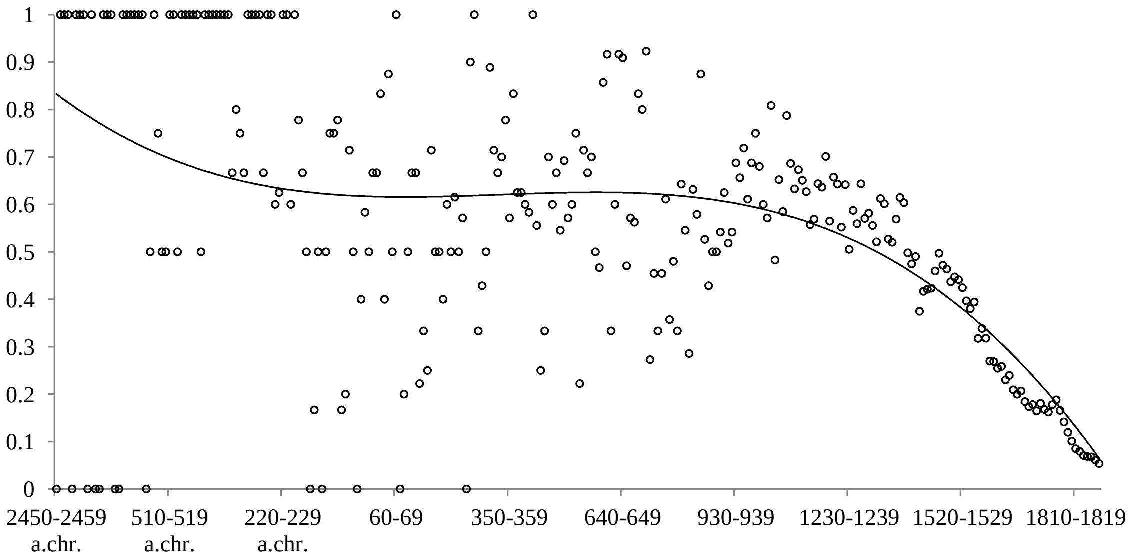


Figure A.1: Frequency of Imprecise Observations

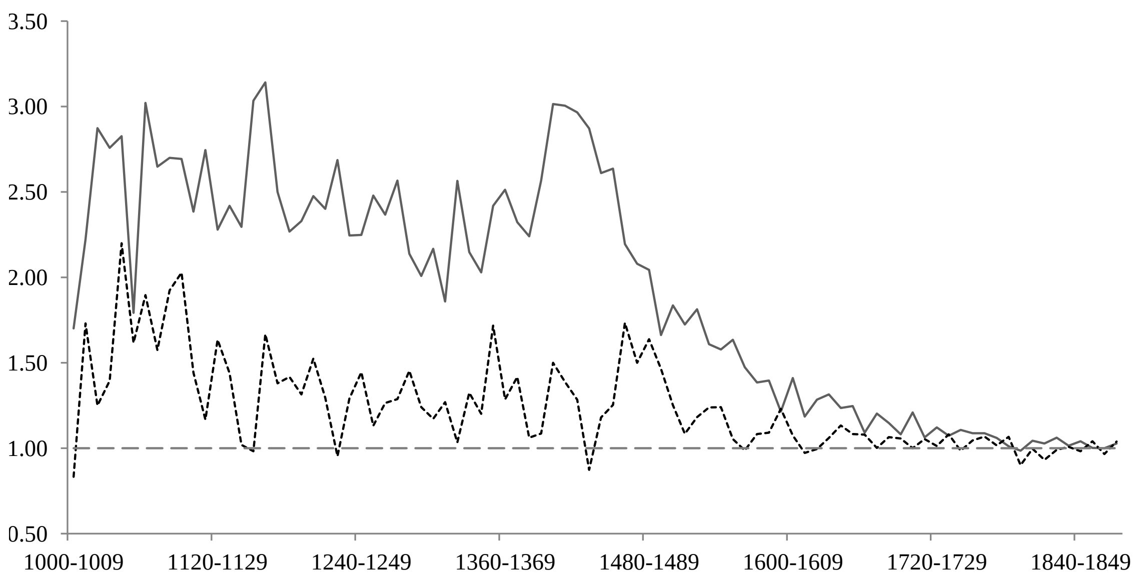


Figure A.2: Heaping Index. Birth year (solid line), death year (dashed line)

# B Longevity vs Life Expectancy

The population of currently famous people alive at the beginning to time $s$, $N_s$, evolves according to:

$$N_{s+1} = N_s - d_s + I_{s+1} \tag{1}$$

where $d_s$ is the number of deaths between age $s$ and age $s+1$, and $I_{s+1}$ is the number of new people gaining celebrity over the same age interval. When computing life expectancy, $N_s$ is the population at risk. Unfortunately, we do not observe $N_s$ since we do not know when people become famous (except in special cases for which a nomination is required) and $I_{s+1}$ is unobserved.

The life expectancy at age $a$ of this population is:

$$E_a = \sum_{s=a}^{T} (s-a) m_{s,a}$$

where $T$ is the maximum number of periods one can live, and $m_{s,a}$ is the true probability of dying at age $s$ conditionally on being alive at age $a$:

$$m_{s,a} = \frac{d_s}{N_s} \times S_{s,a}.$$

$S_{s,a}$ is the probability of reaching age $s$ if one has reached age $a$. It follows:

$$S_{s+1,a} = S_{s,a} \times \left(1 - \frac{d_s}{N_s}\right) = S_{a,a} \times \prod_{j=a}^{s} \left(1 - \frac{d_j}{N_j}\right) = \prod_{j=a}^{s} \left(1 - \frac{d_j}{N_j}\right)$$

Notice that, contrary to the case where all individual belong to the population at age $a$ and can be followed until death, $m_{s,a} \neq d_s / N_a$.

One can rewrite the population at risk as:

$$N_{s+1} = S_{s+1,a} \left( N_a + \sum_{j=a+1}^{s+1} \frac{I_j}{S_{j,a}} \right). \tag{2}$$

We denote the population of all famous people aged $s$ by $\hat{N}_s$. This population includes everyone that is or will become famous. Contrary to $N_s$, we observe $\hat{N}_s$. Its dynamics are given by

$$\hat{N}_{s+1} = \hat{N}_s - d_s \tag{3}$$

Using equations (1) and (3), we can compute the gap between total population $\hat{N}$ and population at risk $N$ as

$$\hat{N}_s - N_s = \hat{N}_{s+1} - N_{s+1} + I_{s+1}$$

Iterating forward, we get

$$\hat{N}_s - N_s = \hat{N}_T - N_T + \sum_{j=s+1}^{T} I_j = \sum_{j=s+1}^{T} I_j \tag{4}$$

where $T$ is the date at which all famous people have been discovered, i.e., $I_j = 0 \ \forall t > T$. It implies $\hat{N}_T = N_T$. The above equation reflects the idea that $\hat{N}_s$ incorporates all the people that are not yet famous but will be. We can define $\hat{m}_{s,a}$ as the observable probability of dying at age $s$ conditionally on being alive at age $a$:

$$\hat{m}_{s,a} = \frac{d_s}{\hat{N}_a}.$$

We have the following property:

$$\sum_{s=a}^{T} d_s = \hat{N}_a, \quad \sum_{s=a}^{T} \hat{m}_{s,a} = 1.$$

We can measure the mean lifetime conditionally on being alive at $a$:

$$L_a = \sum_{s=a}^{T} s \ \hat{m}_{s,a}$$

which we call “longevity”.

There is a gap $G_a$ between the expected length of life $a + E_a$:

$$G_a = L_a - E_a - a = \sum_{s=a}^{T} (s-a)(\hat{m}_{s,a} - m_{s,a}).$$

This gap comes from the fact that we cannot compute the correct mortality rates $m_{s,a}$ as we do not know the population at risk. Replacing $\hat{m}_{s,a}$by its value, we obtain

$$G_a = \sum_{s=a}^{T} (s-a) \left( \frac{d_s}{N_s} \frac{N_s}{\hat{N}_a} - m_{s,a} \right) = \sum_{s=a}^{T} (s-a) m_{s,a} \left( \frac{N_s}{\hat{N}_a S_{s,a}} - 1 \right)$$

Using (2) we get:

$$G_a = \sum_{s=a}^{T}(s-a)m_{s,a}\left(\frac{S_{s,a}\left(N_a + \sum_{j=a+1}^{s}\frac{I_j}{S_{j,a}}\right)}{\hat{N}_a S_{s,a}} - 1\right)$$
$$= \sum_{s=a}^{T}(s-a)m_{s,a}\left(\frac{N_a + \sum_{j=a+1}^{s}\frac{I_j}{S_{j,a}}}{\hat{N}_a} - 1\right)$$

Using (4) we get:

$$G_a = \sum_{s=a}^{T}(s-a)m_{s,a}\left(\frac{\hat{N}_a - \sum_{j=a+1}^{T} I_j + \sum_{j=a+1}^{s}\frac{I_j}{S_{j,a}}}{\hat{N}_a} - 1\right)$$
$$= \sum_{s=a}^{T}(s-a)m_{s,a}\left(\sum_{j=a+1}^{s}\frac{I_j}{\hat{N}_a S_{j,a}} - \sum_{j=a+1}^{T}\frac{I_j}{\hat{N}_a}\right).$$

Here is how the gap depends on the process leading to notoriety $\{I_j\}_{j=a..T}$:

$$G_a = \sum_{s=a}^{T}(s-a)m_{s,a}\left(\sum_{j=a+1}^{s}\frac{I_j(1-S_{j,a})}{\hat{N}_a S_{j,a}} - \sum_{j=s+1}^{T}\frac{I_j}{\hat{N}_a}\right).$$

To illustrate the effect of a change in the age at which people become famous, suppose that a proportion $\mu$ of all famous people are already famous at age $a$ and that the proportion $1-\mu$ gets famous at age $f > a$. Then the bias is:

$$B_a = \sum_{s=a}^{T}(s-a)m_{s,a}\left(\frac{(1-\mu)\hat{N}_a}{\hat{N}_a S_{f,a}}\right) = \frac{1-\mu}{S_{f,a}}E_a.$$

The bias is therefore proportional to life expectancy, with the proportionality factor increasing in $f$ (as $S_{f,a}$ is decreasing in $f$) and decreasing in $\mu$. If age at notoriety $f$ changes, we have:

$$\frac{\partial B_a}{\partial f} = -\frac{1-\mu}{S_{f,a}^2}E_a\frac{\partial S_{f,a}}{\partial f}.$$

The derivatives depends on the slope of the survival function at age $f$. If it is not too decreasing at $f$ ($\frac{\partial S_{f,a}}{\partial f}$ is small), for example when $S$ is concave and $f$ is low enough, the effect on the bias will be small.

# C Occupation categories

*Arts and métiers*: actor, artist, cantor, collector, composer, designer, dramatist, engraver, goldsmith, illustrator, kapellmeister, lithograph, musician, organist, painter, pewterer, pianist, poet, regisseur, sculptor, singer, violinmaker and violinist.

*Business*: antiquary, bookseller, banker, printer, publicist, businessman, director, editor, farmer, librarian, industrialist, merchant, trader, manufacturer and wholesaler.

*Education*: author, academician, dean, lecturer, professor, rector, scholar, student, teacher and writer.

*Humanities*: archaeologist, classicist, economist, historian, journalist, orientalist, pedagogue, philologe, philosopher and translator.

*Law and government*: administrator, adviser, ambassador, bailiff, beamter, congressman, consul, councillor, deputy, diplomat, governor, inspector, judge, jurist, lawyer, magistrato, mayor, minister, notary, politician, prefect, president, procureur, secretary, senator and sheriff.

*Military*: admiral, brigadier-general, captain, colonel, commander, fighter, general, lieutenant, lieutenant-colonel, major, major-general, marshal, military, officer and soldier.

*Nobility*: baron, chamberlain, duke, earl, king, knight, lord, noble, prince and queen.

*Religious*: abbot, archbishop, archdeacon, capuchin, cardinal, clergyman, deacon, franciscan, jesuit, martyr, missionary, pastor, piarist, preacher, priest, rabbi, theologian and vicar.

*Sciences*: agronomist, architect, astronomer, botanist, builder, cartographer, chemist, doctor, engineer, geographer, geologist, inventor, mathematician,naturalist, pharmacist, physician, physicist, surgeon and zoologist.

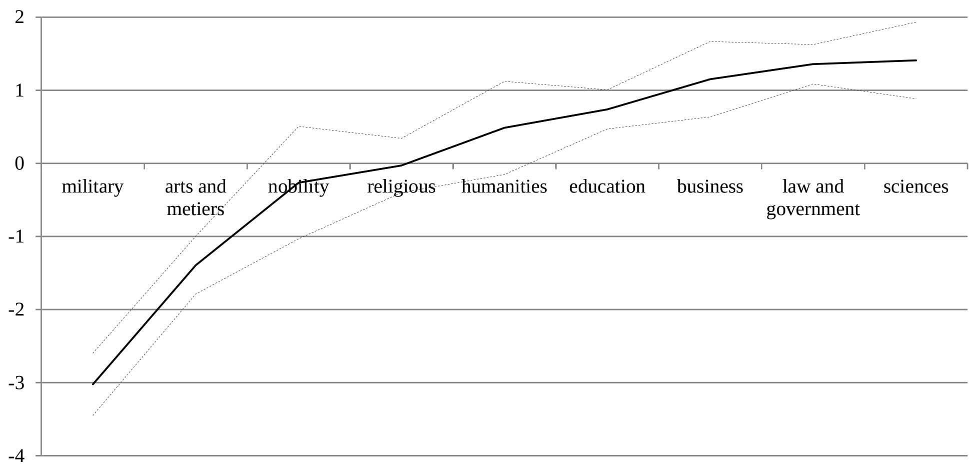


Figure A.3: Conditional Longevity. Main occupational groups

# D Analysis of the Residuals

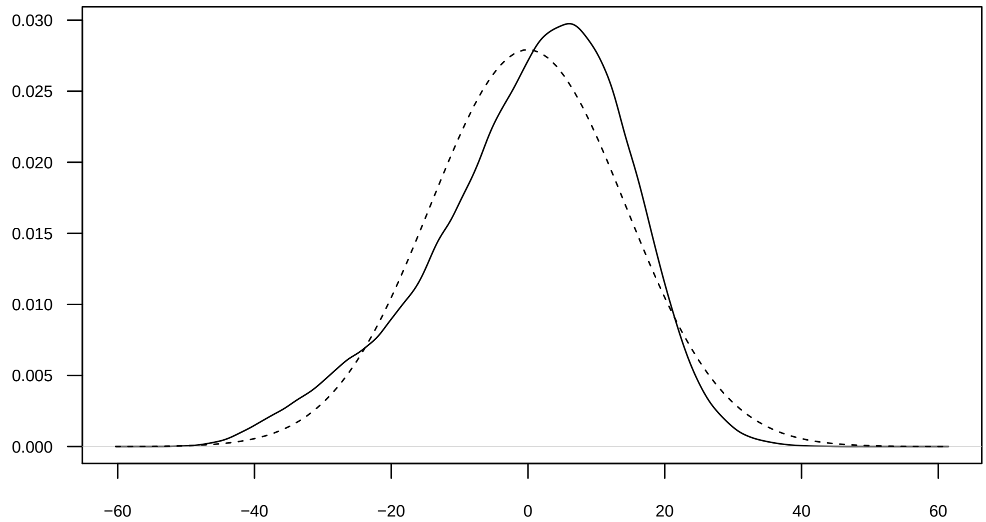


Figure A.4: Kernel Density of the Residuals (solid) and Normal density (dashes)


Figure A.5: Standard Deviation of Residuals by Decade, and 95% confidence interval

# E Longevity per Year of Death

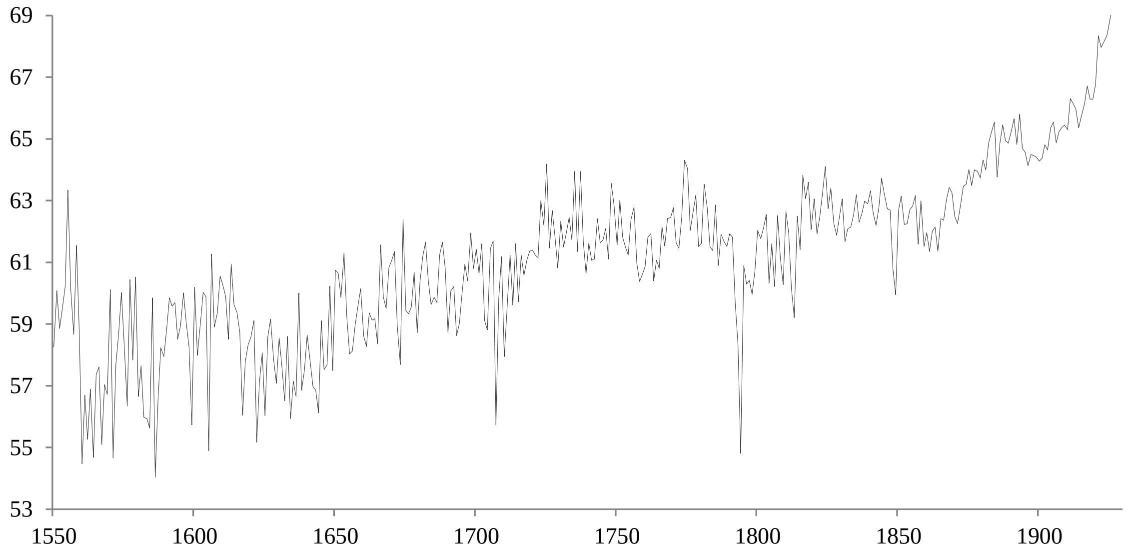


Figure A.6: Longevity per Year of Death

## F 1600 cohorts

| cohort # | years | # obs |
|---|---|---|
| 1 | 2450 BCE - 1040 CE | 1,611 |
| 2 | 1041 - 1254 | 1,602 |
| 3 | 1255 - 1360 | 1,634 |
| 4 | 1361 - 1415 | 1,617 |
| 5 | 1416 - 1450 | 1,737 |
| 6 | 1451 - 1481 | 1,619 |
| 7 | 1482 - 1502 | 1,600 |
| 8 | 1503 - 1520 | 1,676 |
| 9 | 1521 - 1534 | 1,636 |
| 10 | 1535 - 1546 | 1,675 |
| 11 | 1547 - 1557 | 1,641 |
| 12 | 1558 - 1566 | 1,627 |
| 13 | 1567 - 1575 | 1,765 |
| 14 | 1576 - 1583 | 1,708 |
| 15 | 1584 - 1590 | 1,660 |
| 16 | 1591 - 1597 | 1,723 |
| 17 | 1598 - 1603 | 1,773 |
| 18 | 1604 - 1610 | 1,926 |
| 19 | 1611 - 1616 | 1,616 |
| 20 | 1617 - 1622 | 1,740 |
| 21 | 1623 - 1628 | 1,743 |
| 22 | 1629 - 1633 | 1,621 |
| 23 | 1634 - 1639 | 1,870 |
| 24 | 1640 - 1644 | 1,660 |
| 25 | 1645 - 1649 | 1,621 |
| 26 | 1650 - 1654 | 1,731 |
| 27 | 1655 - 1659 | 1,663 |
| 28 | 1660 - 1664 | 1,862 |
| 29 | 1665 - 1669 | 1,710 |
| 30 | 1670 - 1674 | 1,846 |
| 31 | 1675 - 1679 | 1,730 |

*continued on next page*

| **cohort #** | **years** | **# obs** |
|---:|:---:|---:|
| 32 | 1680 - 1683 | 1,609 |
| 33 | 1684 - 1688 | 1,888 |
| 34 | 1689 - 1693 | 1,903 |
| 35 | 1694 - 1697 | 1,713 |
| 36 | 1698 - 1701 | 1,872 |
| 37 | 1702 - 1705 | 1,641 |
| 38 | 1706 - 1709 | 1,680 |
| 39 | 1710 - 1713 | 1,946 |
| 40 | 1714 - 1717 | 2,023 |
| 41 | 1718 - 1720 | 1,662 |
| 42 | 1721 - 1724 | 2,106 |
| 43 | 1725 - 1727 | 1,763 |
| 44 | 1728 - 1730 | 1,879 |
| 45 | 1731 - 1733 | 1,892 |
| 46 | 1734 - 1736 | 2,046 |
| 47 | 1737 - 1739 | 1,994 |
| 48 | 1740 - 1742 | 2,208 |
| 49 | 1743 - 1745 | 2,298 |
| 50 | 1746 - 1748 | 2,284 |
| 51 | 1749 - 1750 | 1,874 |
| 52 | 1751 - 1752 | 1,795 |
| 53 | 1753 - 1754 | 1,793 |
| 54 | 1755 - 1756 | 1,901 |
| 55 | 1757 - 1758 | 1,834 |
| 56 | 1759 - 1760 | 1,907 |
| 57 | 1761 - 1762 | 1,840 |
| 58 | 1763 - 1764 | 2,042 |
| 59 | 1765 - 1766 | 2,125 |
| 60 | 1767 - 1768 | 1,978 |
| 61 | 1769 - 1770 | 2,231 |
| 62 | 1771 - 1772 | 2,054 |
| 63 | 1773 - 1774 | 2,074 |

*continued on next page*

| **cohort #** | **years** | **# obs** |
|---|---|---|
| 64 | 1775 - 1776 | 2,041 |
| 65 | 1777 - 1778 | 2,060 |
| 66 | 1779 - 1781 | 2,039 |
| 67 | 1781 - 1782 | 1,867 |
| 68 | 1783 - 1784 | 1,997 |
| 69 | 1785 - 1786 | 2,146 |
| 70 | 1787 - 1788 | 2,255 |
| 71 | 1789 - 1790 | 2,390 |
| 72 | 1791 - 1792 | 2,240 |
| 73 | 1793 - 1794 | 2,436 |
| 74 | 1795 - 1796 | 2,485 |
| 75 | 1797 - 1798 | 2,620 |
| 76 | 1799 - 1800 | 2,950 |
| 77 | 1801 - 1802 | 3,061 |
| 78 | 1803 - 1804 | 3,017 |
| 79 | 1805 - 1806 | 3,152 |
| 80 | 1807 - 1808 | 3,161 |
| 81 | 1809 - 1810 | 3,322 |
| 82 | 1811 | 1,688 |
| 83 | 1812 | 1,743 |
| 84 | 1813 | 1,611 |
| 85 | 1814 | 1,642 |
| 86 | 1815 | 1,795 |
| 87 | 1816 | 1,624 |
| 88 | 1817 | 1,849 |
| 89 | 1818 | 1,838 |
| 90 | 1819 | 1,805 |
| 91 | 1820 | 1,863 |
| 92 | 1821 | 1,705 |
| 93 | 1822 | 1,731 |
| 94 | 1823 | 1,770 |
| 95 | 1824 | 1,709 |

*continued on next page*

| **cohort #** | **years** | **# obs** |
|---:|---:|---:|
| 96 | 1825 | 1,845 |
| 97 | 1826 | 1,685 |
| 98 | 1827 | 1,760 |
| 99 | 1828 | 1,746 |
| 100 | 1829 | 1,795 |
| 101 | 1830 | 1,890 |
| 102 | 1831 | 1,692 |
| 103 | 1832 | 1,693 |
| 104 | 1833 | 1,786 |
| 105 | 1834 | 1,775 |
| 106 | 1835 | 1,785 |
| 107 | 1836 | 1,835 |
| 108 | 1837 | 1,873 |
| 109 | 1838 | 1,860 |
| 110 | 1839 | 1,931 |
| 111 | 1840 | 2,069 |
| 112 | 1841 | 1,958 |
| 113 | 1842 | 2,001 |
| 114 | 1843 | 1,972 |
| 115 | 1844 | 1,951 |
| 116 | 1845 | 2,048 |
| 117 | 1846 | 1,953 |
| 118 | 1847 | 1,966 |
| 119 | 1848 | 2,067 |
| 120 | 1849 | 1,882 |
| 121 | 1850 | 2,097 |
| 122 | 1851 | 2,022 |
| 123 | 1852 | 2,079 |
| 124 | 1853 | 1,901 |
| 125 | 1854 | 1,956 |
| 126 | 1855 | 2,056 |
| 127 | 1856 | 2,134 |

*continued on next page*

| **cohort #** | **years** | **# obs** |
|---:|---:|---:|
| 128 | 1857 | 2,055 |
| 129 | 1858 | 2,292 |
| 130 | 1859 | 2,239 |
| 131 | 1860 | 2,286 |
| 132 | 1861 | 2,212 |
| 133 | 1862 | 2,304 |
| 134 | 1863 | 2,300 |
| 135 | 1864 | 2,267 |
| 136 | 1865 | 2,296 |
| 137 | 1866 | 2,273 |
| 138 | 1867 | 2,245 |
| 139 | 1868 | 2,410 |
| 140 | 1869 | 2,367 |
| 141 | 1870 | 2,301 |
| 142 | 1871 | 2,128 |
| 143 | 1872 | 2,288 |
| 144 | 1873 | 2,296 |
| 145 | 1874 | 2,286 |
| 146 | 1875 | 2,381 |
| 147 | 1876 | 2,327 |
| 148 | 1877 | 2,267 |
| 149 | 1878 | 2,309 |
| 150 | 1879 | 2,349 |
| Total | | 297,651 |

# G Estimation of the Gomperz-Makeham Law

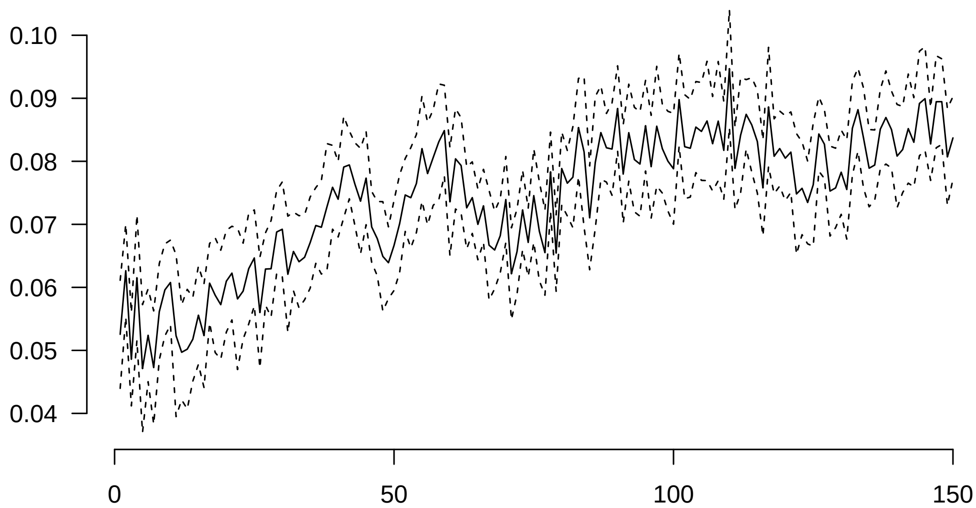


Figure A.7: Estimated $\hat{\alpha}$

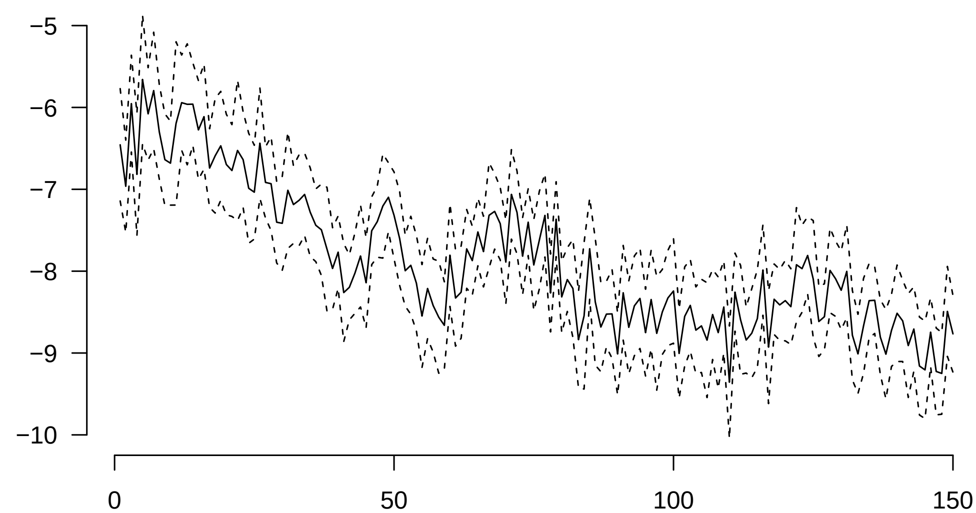


Figure A.8: Estimated $\hat{\rho}$

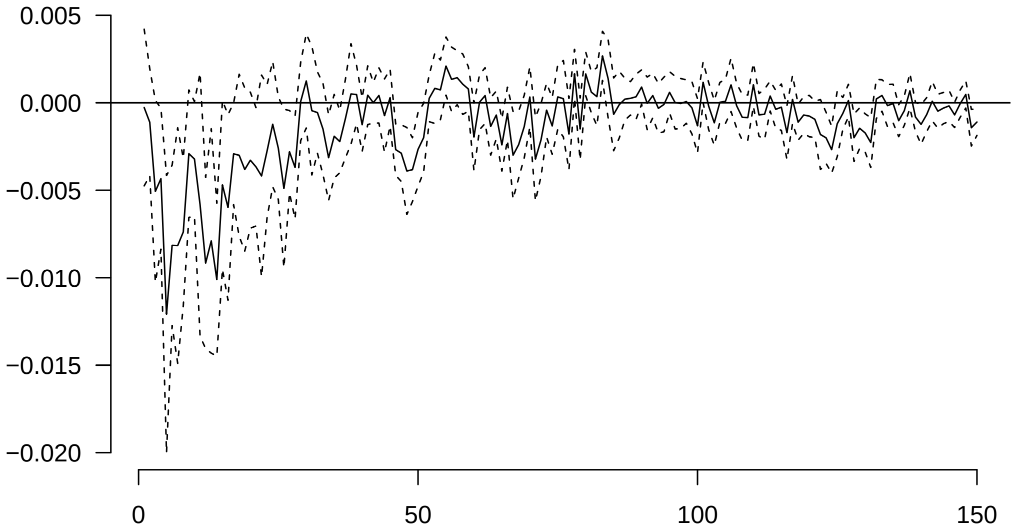


Figure A.9: Estimated $\hat{A}$

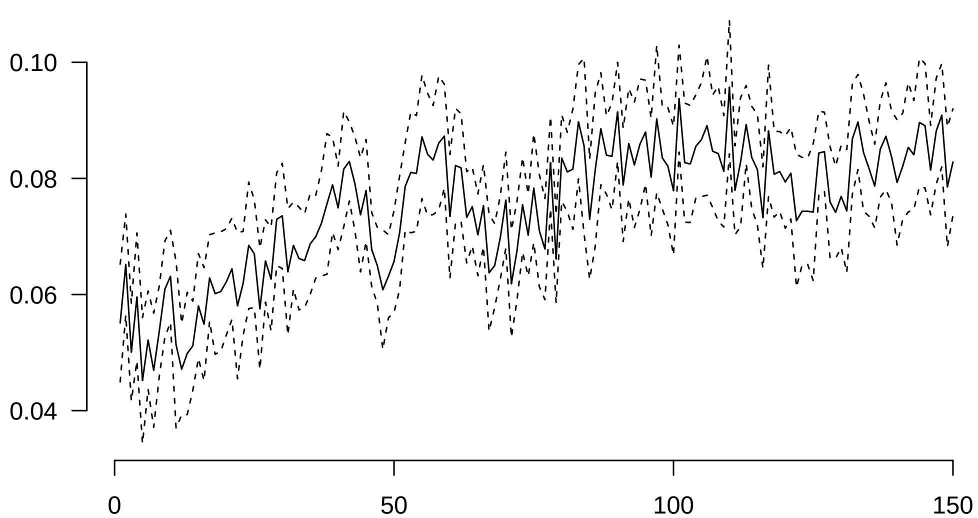


Figure A.10: Estimated $\hat{\alpha}$ - Notoriety Bias Corrected

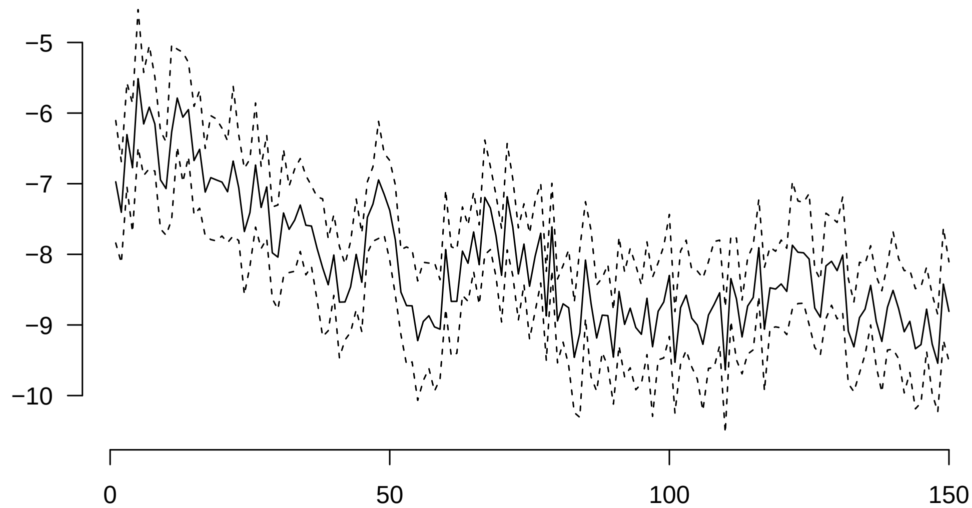

Figure A.11: Estimated $\hat{\rho}$ - Notoriety Bias Corrected

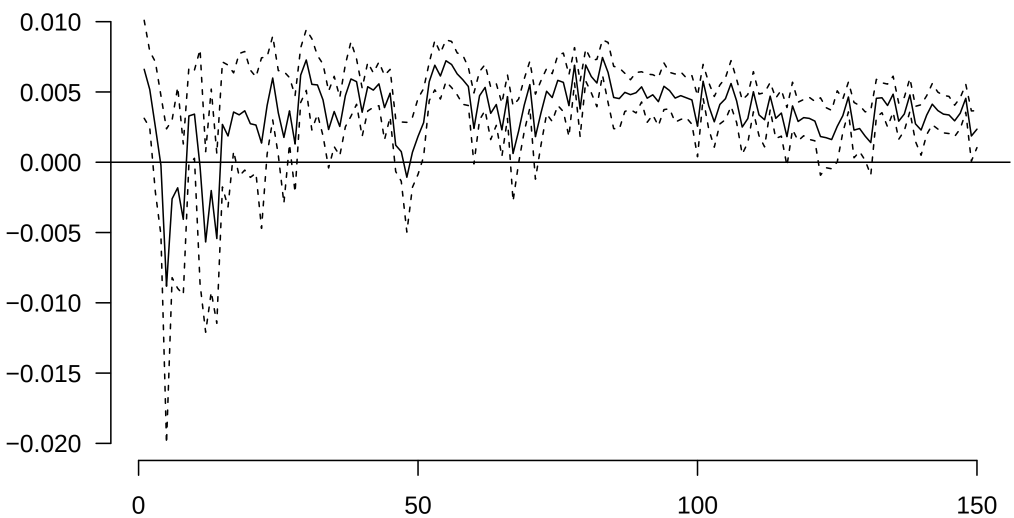

Figure A.12: Estimated $\hat{A}$ - Notoriety Bias Corrected

# References

A'Hearn, Brian, Joerg Baten, and Dorothee Crayen. 2009. "Quantifying quantitative literacy: age heaping and the history of human capital." *Journal of Economic History* 69:783–808.

De Moor, Tine, and Jaco Zuijderduijn. 2013. "The Art of Counting: Reconstructing Numeracy of the Middle and Upper Classes on the Basis of Portraits in the Early Modern Low Countries." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 46 (1): 41–56.