QUCEH WORKING PAPER SERIES
http://www.quceh.org.uk/working-papers

# AGE HEAPING AND NUMERACY: LOOKING BEHIND THE CURTAIN

Matthias Blum (Queen's University Belfast)
Karl-Peter Krauss (Institute for Danube-Swabian History and Regional Studies)

Working Paper 2017-05

QUEEN'S UNIVERSITY CENTRE FOR ECONOMIC HISTORY
Queen's University Belfast
185 Stranmillis Road
Belfast BT9 5EE
June 2017

# Age heaping and numeracy: looking behind the curtain

Matthias Blum[1] and Karl-Peter Krauss[2]

**Abstract**

Age heaping-based numeracy indicators have served as valuable tools to derive basic human capital estimates, especially for periods where other indicators are unavailable. However, the accuracy of *individual* age statements usually remains unknown, and due to the lack of precise information it can only assumed that excessive occurrence of multiples of five in age distributions reflects inferior numerical skills. We address this lacuna by identifying 162 individuals in two independent data sources, self-reported age statements and independently kept records which are based on family heritage books and church registers. This method allows us to identify individual misreporting and the *degree* of accuracy of each individual. We find that not everyone who reports a multiple of five reports an incorrect age, nor everyone who reports an age that is *not* a multiple of five reports an accurate age. In an empirical analysis we show that the commonly used binary numeracy indicator is correlated with the observed degree of accuracy in age statements, and that a more sophisticated occupational background reduces this inaccuracy. Our tentative results suggest that the commonly used binary indicator measuring age heaping is a valuable proxy of the numerical skills and occupational background in a population.

**Keywords:** Numeracy; ABCC; age heaping; human capital; inequality; economic history; skills; methodology; migration.

**JEL codes:** N01, N33, C43, O15.

Acknowledgements: We are indebted to Zoltán Csapo who provided a database containing these information. The authors thank the participants of the Budapest WEast workshop, Sara Horrell, Valeria Prayon and two anonymous reviewers for helpful comments and suggestions on earlier drafts of this article.

---

[1] Corresponding author: Address: Queen's Management School, Queen's University Belfast, 185 Stranmillis Road, Belfast BT9 5EE, United Kingdom. Email: matthias.blum@qub.ac.uk; Phone: +44 (2)289097 4428.

[2] Institut für donauschwäbische Geschichte und Landeskunde (Institute for Danube Swabian History and Regional Studies), Tübingen, Germany.

## Introduction

The excessive occurrence of rounded age statements, known as 'age heaping' (AH), has been observed in low skilled populations. Individuals in poorly-educated societies who do not know their ages or do not value this piece of information as relevant are likely to report an age ending in a multiple of five instead of reporting their accurate age.[3] Meanwhile, economic historians and development economists have started to use this phenomenon to derive estimates of basic numeracy of a population.

However, critics of this methodology argue that this metric should be treated with caution; for example, it has been argued that 'no one has as yet demonstrated for any society in the past that an *individual* with a lack of age awareness has a low degree of ability in quantitative reasoning'.[4] This lack of confidence and trust in this concept may originate in the estimating procedure since individual age statements remain unknown when conventional numeracy indicators are computed. Due to the lack of precise individual information it can only be assumed that excessive occurrence of ages ending in "5" or "0" reflects imprecision in age statements due to the absence of basic numeracy.[5] Without knowledge about an individual's true age, it is impossible to identify individual misreporting, regardless whether a multiple of five is reported or not. Moreover, even if inaccurate individual age statements could be unambiguously identified, a binary variable would not allow an assessment of the *degree* of imprecision since all misreported ages are treated as 'inaccurate'.

The findings of this study suggest that age heaping also correlates with precision in individual age statements. Our strategy is based on the use of two independently compiled datasets containing German-speaking migrants travelling via Vienna to the Kingdom of

[3] Bachi, *Tendency*; Myers, *Accuracy*; Myers, *reverse heaping*; Mokyr, *Ireland*.
[4] Davids, *Religion*, p. 62.
[5] Crayen and Baten, *Global trends*.

Hungary during the late eighteenth century; both datasets provide information about individual ages and allow us to cross-check their accuracy. We know the self-reported ages of these migrants from Habsburg migration lists and know their true age using family heritage books which relate to villages these migrants eventually settled in. Comparing these two independent data sources allows us to identify incorrect age statements and the degree of inaccuracy for each individual in our sample. In an empirical analysis we show that the commonly used binary numeracy indicator is correlated with the observed degree of accuracy in age statements, and that a more sophisticated occupational background is correlated with inaccuracy in age statements. There is, therefore, reason to believe that the commonly used binary AH indicator is a valuable proxy of basic numerical skills and occupational background in a population.

However, these findings must be considered tentative since the methodology used here makes the compilation of a large dataset difficult and this sample of 162 individuals is probably not fully representative. Yet, these findings, if applicable to other historical settings, support the use of AH and reinforce the findings of existing AH studies. For example, AH techniques have helped to assess periods and historical settings for which other metric are not available, such as sixteenth century Netherlands and England;[6] seventeenth century Cape Colony and Eastern Europe;[7] eighteenth to twentieth century Portugal;[8] nineteenth century Brazil, Ireland and Georgia;[9] Inca Indios' age heaping before the Spanish conquest;[10] the Roman province of Pannonia in the first century BCE.[11] Other studies assess human capital selectivity of migrants

[6] de Moor and van Zanden, *Every Woman Counts*; de Moor and Zuijderduijn, *Art of Counting*; Földvári et al., *How did women count?*.
[7] Baten and Fourie, *Numeracy of Africans*; Baten, Szoltysek and Campestrini, *Girl Power*.
[8] Stolz et al. 2013a, *Portuguese living standards*.
[9] Stolz et al. 2013b, *Growth effects*; Blum et al., *uncertain age*; Baten and Sirbiladze, *Elitegruppe*.
[10] Juif and Baten, *Inca Indios*.
[11] Baten and Priwitzer, *Numeracy in Pannonia*.

during the era of mass migration;[12] the effect of land inequality on numeracy in the long-run;[13] and numeracy in Europe in a comprehensive fashion between the 1790s and the 1880s.[14]

**Methodological background**

Formula 1 shows how to calculate the Whipple Index (WI) of a population:

$$(1) \quad WI = \left( \frac{n_{25} + n_{30} + ... + n_{65} + n_{70}}{1/5 \times (n_{23} + n_{24} + n_{25} + ... + n_{72})} \right) \times 100 \; if \; WI \geq 100; \; else \; WI = 100.$$

A WI value of 500 suggests that all reported ages end in a multiple of five, while a WI value of 100 constitutes the lowest score possible, indicating no heaping and therefore high numeric abilities.[15] Since the WI scale contradicts intuition – high values are interpreted as low numeracy – a simple transformation resulting in a numeracy indicator that ranges between 0 and 100 has been suggested (formula 2).[16] This transformed index is known as the ABCC index, which can be interpreted as the estimated share of people in a population which are able to accurately report their age.

$$(2) \quad \text{ABCC} = \left( 1 - \frac{(\text{WI} - 100)}{400} \right) \times 100 \quad \text{if } \text{WI} \geq 100; \text{ else ABCC} = 100.$$

[12] Stolz and Baten, *Brain drain*.
[13] Baten and Juif, *landowners and math skills*.
[14] Hippe and Baten, *Regional inequality*.
[15] In general, all population samples are restricted to ages between 23 and 72 since among very high ages, small sample size and selective mortality are considered a potential bias. Also, among late teenagers and individuals in their early twenties age heaping on even numbers instead of 0 and 5 can has been observed (A'Hearn, Baten and Crayen 2009, Crayen and Baten 2010).
[16] A'Hearn et al., *Quantifying*.

## Historical background and data sources

During the eighteenth century, Prussia, the Habsburg Empire and the Russian Empire undertook immigration and settlement programmes, aimed at strengthening their economic and military basis. Key drivers for this wave of emigration were wars, harvest failures and periods of famine; main ‚pull‘ factors were recruitment campaigns by early modern states and landlords. In addition, craftsmen lamented falling prices for their services, and increasing food and land prices put significant economic pressure on indigent peasants.[17] The Kingdom of Hungary encouraged migration to repopulate regions that were severely affected by a series of wars between the Habsburg Empire and the Ottoman Empire during the seventeenth and early eighteenth centuries; also during the Kuruzzen wars (1703-11) large territories were devastated. Ethnic Germans originated mainly from southern and western territorial states of the Holy Roman Empire; predominantly German-speaking settlers also came from the French provinces of Alsace and Lorraine (a part of France since 1766).

This article considers the Josephinian colonisation, 1784 to 1786, which brought the influx of more than 7,000 mostly ethnic German families, totaling approximately 35,000 migrants in the historical regions of Banat and Bácska (Serbian: Bačka) in the modern states of Hungary, Romania and Serbia.[18] To encourage immigration, settlers received during this time considerable subsidies, such as: travel money; food supplies; agricultural land; livestock; equipment; and settlers were exempted from tax duties for ten years. This generous system of government subsidisation allowed less well-off families who had not been able to afford travelling to migrate; only a total lack of means constituted a financial barrier to migration.[19] Most of these migrants travelled from their places of origin via the Danube ports of Ulm,

[17] Fata, *Migration*, p. 211-212; Krauss, *Rahmenbedingungen*, p. 43-53.
[18] Fata, *Migration*, p. 258-260. Feldtänzer, *Ansiedlung*, p. 39 counted more than 9,000 families.
[19] Seewann *et al.*, *Ansiedlung.*

Günzburg and Regensburg in Germany, via Vienna and Budapest before they were directed to target regions in the Kingdom of Hungary. We are able to trace a sample of these migrants from their places of origin until their arrival and registration in Vienna.

One of the most important sources of information about the migrants is a dataset comprising the list of immigrant families and reports of the Hungarian Court Chancellery (*Hofkanzlei*), the highest representative of the Kingdom of Hungary in Vienna.[20] On the 14th of August 1783, Habsburg administrators were instructed to accept settlers from the Holy Roman Empire and from Lorraine. Settlers were supposed to carry documentation to qualify for travel subsidies; these information were checked by administrators of the Court Chancery. This procedure aimed at identifying family heads' occupations, places of origin, age, religion and assets.[21] In reality, however, administrators had to deal with a patchwork of documents rather than consistently designed passports. Moreover, the fact that administrators were encouraged to allow migrants to quickly continue their journey to limit travel costs as well as the sheer number of new arrivals – on the 13th of March 1784 the records list 70 newly arrived families – made a reliable verification of the household heads' statements virtually impossible. Instead, lists of the Court Chancellery, documenting individual information of this ongoing stream of settlers travelling through Vienna, are in all likelihood self-reported and unverified.[22]

The data used here cover approximately two and a half years of immigration, between the 18th of April 1784 and the 27th of October 1786, but a gap in the data exists for the nine-month period between November 1784 and July 1785. In total, approximately 4,800 individual records of migrants travelling through Vienna are available. We used family heritage books of

---

[20] Original Source: Magyar Nemzeti Levéltár-Országos Levéltár [Hungarian National Archive-Ungarisches Landesarchiv] (MNL-OL), Kancelláriai Levéltár Archive of the Hungarian Court Chancery], Magyar Királyi Kancellária [Archive of the Royal Hungarian Court Chancery], Acta generalia A 39.
[21] Feldtänzer, O., Joseph II., p. 51.
[22] Data source: Österreichisches Staatsarchiv, Finanz- und Hofkammerarchiv, Wien (FHKA), Neue Hofkammer, Galizische Domänen Akten, Ansiedlungsgeschäft, 151, 1783.08-1783.12.

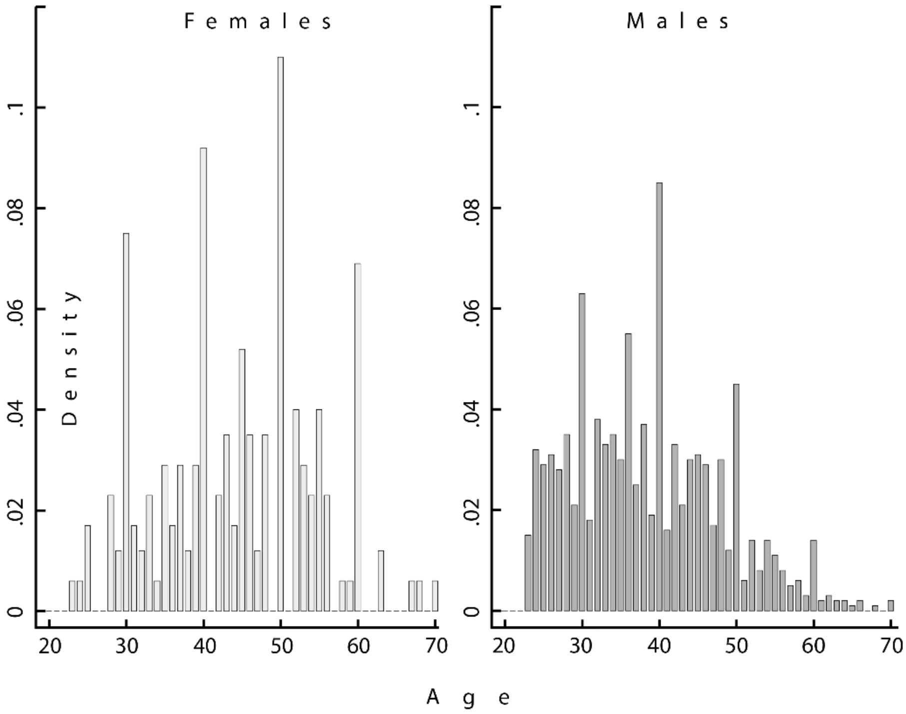
individual villages and towns in the Banat and Bácska regions as a reliable, independent source of information to cross-check and verify 162 of these self-reported age statements. These family heritage books have been compiled by genealogists who used local archival materials in modern-day Hungary and Serbia as well as church registers in settlers' home towns in modern-day Germany. The final dataset we use to verify age statements is limited due to the small intersecting set of observations that can be found both sources including all relevant information. These data pertain to the Catholic villages of Srpski Miletić / Milititsch,[23] and Kruščić / Weprowatz[24] as well as the Protestant villages of Liebling,[25] Vrbas / Neu-Werbaß[26] and Crvenka / Tscherwenka.[27] We used the information about an individual's full name and place of origin to link both datasets. In addition, we use information on an individual's denomination as this piece of information was reported in Vienna, and we can independently infer an individual's denomination by identifying his final destination in the Kingdom of Hungary. Family heritage books often provide an individual's date of birth, which was researched by genealogists using church registers in migrants' places of origin in Germany. This piece of information in combination with the date of arrival in Vienna allows us to cross-check self-reported ages. Finally, we used emigrant registers for German source regions to identify information on the migrants' occupations.[28]

[23] Schuy, *Ortssippenbuch Miletitsch*.
[24] Scherer, *Familienbuch Weprowatz*.
[25] Möhler, *Ortssippenbuch Liebling*.
[26] Hein, *Neuwerbaß in der Batschka*.
[27] Hefner, *Tscherwenka*. The most detailed information are available for the predominantly Protestant village of Crvenka / Tscherwenka. The Protestant domination among the 162 individuals in this database are a result of the wealth of information on places and dates of birth for this particular source.
[28] Hacker, *Hohenzollern*; Hacker, *Upper Neckar*. The use of this additional source was necessary since family heritage books usually do not provide occupational information; self-reported occupations in Vienna were in all likelihood biased since most settlers reported a combination of occupations, often combining 'farmer' with what was allegedly their true occupation, or an occupation they expected to bring about favourable treatment by Habsburg administrators, such as a generously allocated piece of land. Administrators in target regions in the Kingdom of Hungary complained that some settlers obviously were obviously unqualified farmers, despite having reported to be farmers. There is a consensus that many settlers made false statements regarding their occupations to obtain a piece of land. Hacker, in contrast, reports these migrants' occupations as they were noted in settlers' home regions emigration lists, a source we consider reliable.

The sample used here is potentially selected due to the self-selection of migrants and the data sources available in the matching process (a distribution of self-reported ages is shown in figure 1). The self-selection of migrants seems to be modest, or even absent (see figure 2). The matching process, however, may have introduced bias; we used family heritage books of villages that were founded during the Josephinian settlement period to verify self-reported statements. The population in these newly founded villages were probably not representative of the entire German-speaking migrant population in the Kingdom of Hungary as these newly founded locations were populated by disproportionately large agriculturalist population, i.e. we are more likely to observe farmers and their labourers in this matched sample. In contrast, the general migrant population also contained white collar and professional occupations which are underrepresented in predominantly rural locations. Similarly, we had to rely on a given set of family heritage books, with the sources for formerly Protestant villages being the richest sources. Conversely, we believe that the provision of travel subsidies such as cash payments enabled migration of less well-off classes and therefore reduced sample selection.[29]

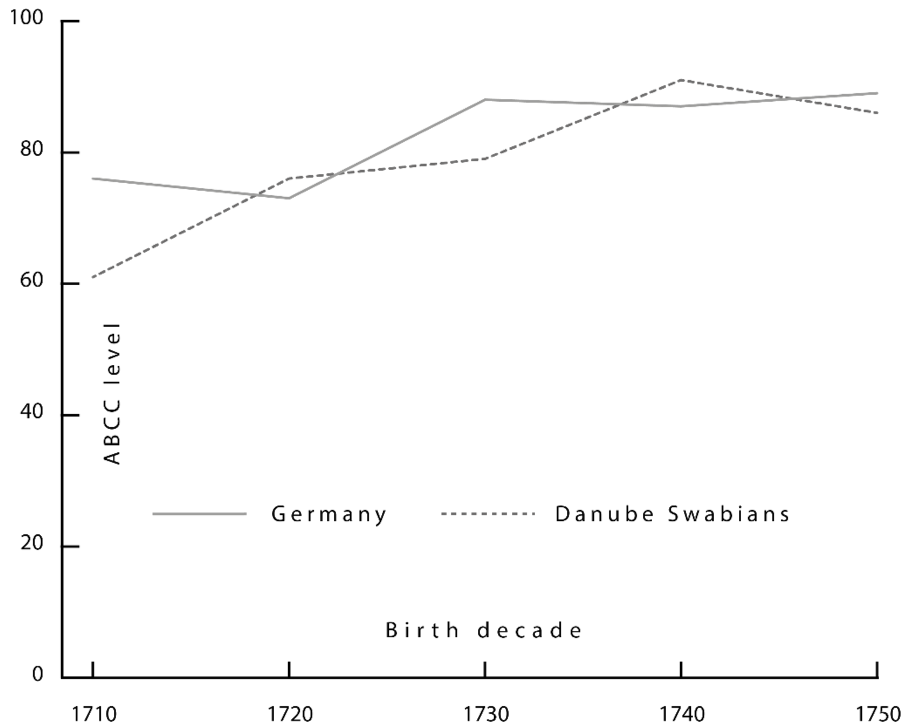
[29] A reviewer correctly noted that this list of potential selection factors is not exhaustive since we naturally can only identify observables. Statistically speaking, potential unobservable factors leading to selection bias could relate to genealogists‘ decision or ability to trace some villages (or individuals) and ignore others; status and prominence of villages and individuals; and continuous existence of a settlement.

Figure 1: Age heaping among female and male migrants to the Kingdom of Hungary




Note: Both histograms shown above refer to the full migrant dataset of approximately 4,800 German-speaking settlers arriving in Vienna. Figure 1 is based on N=173 female and N=4,625 male observations. Only family heads were included. We identified 162 of these migrants in an independent data source and use these observations in the analysis outlined below.

Figure 2: ABCC values for Germany and German migrants, 1710s-1750s




Data sources: Germany: A'Hearn et al. (2009), Baten et al. (2008); 'Danube Swabians': see text.
ABCC trends are organised by birth cohort to allow a direct comparison.

Descriptive statistics presented in table 1 suggest that the average year of birth of the 162 migrants is 1745, with the majority (44 per cent) being 33 to 42 years of age at the time of migration; 27 per cent and 17 per cent are in the age cohorts of 23 to 32 and 43 to 52 years. In total, only 12 per cent of these 162 migrants are found to be older than 52 years of age. The average reported age (38.7 years) is slightly lower than the true age (39.5 years). Migrants in our sample are predominantly Protestant (85 per cent); this figure is upwardly biased due to the availability of predominantly Protestant family heritage books.[30] All migrants were male because only household heads reported to Habsburg authorities, applying for travel subsidies on behalf of their families. Also, for a small sub-sample of migrants (N=52) we know the occupational background. We categorise all individuals according to the Armstrong-taxonomy, using occupational titles to differentiate between professional occupations (e.g. doctor), semi-professionals (e.g. watchmaker), skilled (e.g. weaver), semi-skilled (e.g. miner), and unskilled workers (e.g. labourers).[31] Each category is rank-ordered; unskilled workers are given a rank of 1 while professionals are given a rank of 5. Due to the lack of observations, we combine unskilled and semi-skilled occupations into one category and test the difference between this 'low occupation' group and their 'high occupation' counterpart group, which combines skilled, semi-professional and professional occupations. Using this categorisation, we find that 46.2 per cent of this subset of migrants is in the 'low' category, while 53.8 per cent of individuals is found to have a 'high' occupational background.

[30] On 21st of December 1781, Joseph II. enacted a 'tolerance decree', allowing Protestants to officially settle in the Kingdom of Hungary. Despite this opportunity, the majority of settlers remained to be Roman Catholic; the large share of Protestants in this sample is the result of the detailed information provided in the Family Heritage Books of the Protestant villages of Vrbas and Crvenka, allowing us to identify a disproportionately large number of Protestant settlers.

[31] Armstrong, *Information about Occupation*.

Table 1: Data characteristics of 162 migrants in the sample (in % unless otherwise stated)

| | **Average** |
|---|---|
| Birth year (true) | 1745 |
| Age in years (true) | 39.5 |
| Age in years (reported) | 38.7 |
| | |
| | **%** |
| **Religion** | |
| Protestant | 84.6 |
| Catholic | 15.4 |
| | |
| **Age group** | |
| 23-32 | 26.5 |
| 33-42 | 44.4 |
| 43-52 | 16.7 |
| 53-62 | 8.6 |
| 63-72 | 3.7 |
| | |
| **Occupation (N=52)** | |
| Low (Armstrong 1 to 2) | 46.2 |
| High (Armstrong 3 to 5) | 53.8 |
| | |
| *N* | *162* |

**Results**

We analyse the age heaping pattern of all 162 individuals in our sample to obtain basic numeracy estimates based on two independent datasets (table 2); 58 migrants (36 per cent) reported a rounded age upon arrival in Vienna, while 104 (64 per cent) of these self-reported ages do not end in a multiple of five. These age statements correspond with a Whipple index of 179 and an ABCC index of 80, suggesting that the individuals we identified in church records have a higher tendency to report a rounded age compared to the migrant population as a whole. We use ages that are accurately computed on the basis of family heritage books to obtain the true distribution of ages. This alternative assessment suggests that only 18.5 per cent of these migrants actually have an age that is a multiple of five, while for 81.5 per cent the opposite is true. If these age statements are used to compute numeracy indicators, both the Whipple (100.0) and ABCC indices (100.0) suggest the absence of age heaping. This is not a surprising result since these ages are not self-reported by migrants but are computed on the basis of accurate family heritage books and church records.

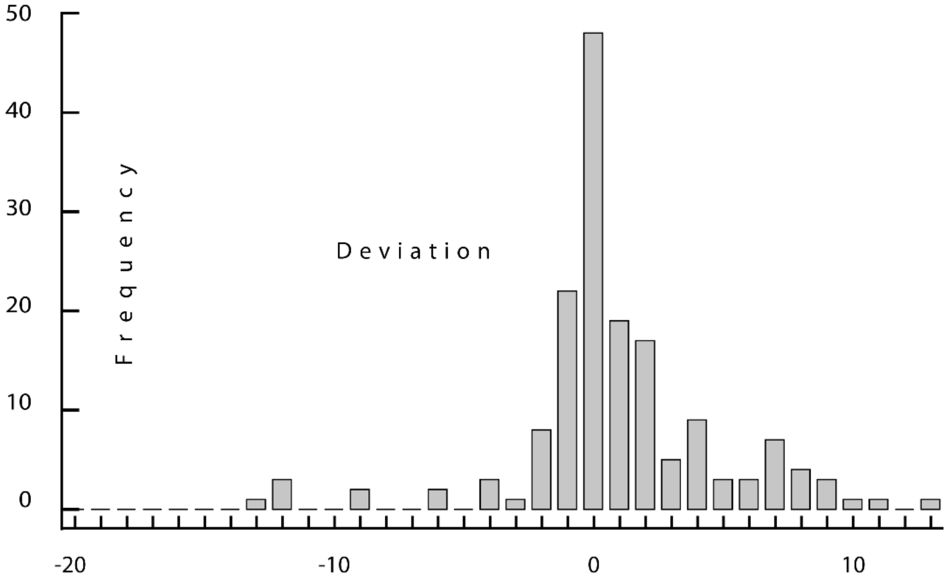
Table 2: Rounding and numeracy: self-reported ages vs. verified age statements

| | Self-reported | % | Verified | % |
|---|---|---|---|---|
| Non-rounded age | 104 | 64.2 | 132 | 81.5 |
| Rounded age | 58 | 35.8 | 30 | 18.5 |
| WI | 179.0 | | 100.0 | |
| ABCC | 80.2 | | 100.0 | |

Note: The 'verified' WI and ABCC values were actually 92.6 and 101.9, but were changed to 100 according to formulas 1 and 2.

In a further step, we calculate the deviation between accurate and reported ages. Negative values represent individuals who are found to be younger than their self-reported age and vice versa (figure 3). Approximately 30 per cent of all individuals reported an accurate age, and approximately 25 per cent reported an age that deviates from the accurate age by only one year. The distribution generally suggests a fairly high level of accuracy, but there is also a standard deviation of 3.9 around the mean deviation. It also appears that this distribution is not distributed normally; 44 per cent and 26 per cent of migrants underestimate and overestimate the accurate age, respectively. We do not know to what degree this inaccuracy is caused by deliberate deviations from an individual's true age; for example, it is plausible that some elderly migrants systematically under-reported their age in order to signal physical strength and to have an edge in the competition for large pieces of land in the land allocation process. We address this concern by including age-specific control variables in some of the regression models below.

Figure 3: Deviation between self-reported age and accurate age




Note: The deviation is defined as accurate age minus reported age. A negative value therefore indicates the individual being younger (reported an older age) than the reported age and vice versa. A value of zero indicates that reported age and verified age are identical.

Finally, we investigate to what degree rounding, i.e. reporting an age that is a multiple of five, corresponds with inaccurate age statements (table 3). A simple cross table provides empirical evidence that supports this rationale: 33 out of 162 individuals (20 per cent) report an accurate age that is not a multiple of five, and 43 out of 162 individuals (27 per cent) report an incorrect age rounded to a multiple of five. Conversely, we find that 15 out of all 162 individuals (nine per cent) correctly report a multiple of five, while 71 out of 162 individuals, equaling 44 per cent of the total sample, are found to misreport their age while reporting an age that is *not* a multiple of five.

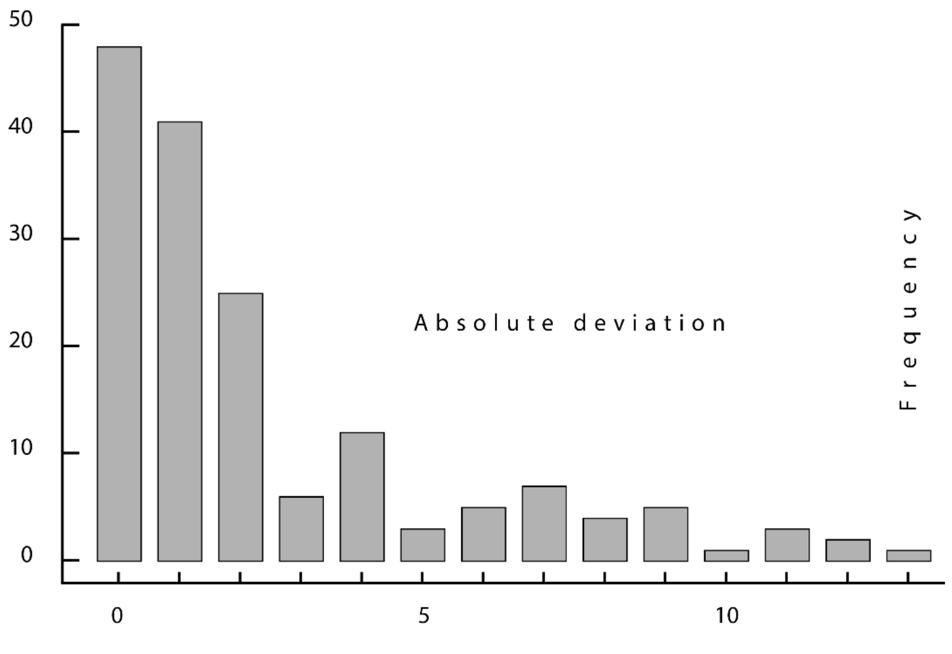
Table 3: Expected and actual relationship between accuracy and multiples of five

| | | Age statement is | | |
|---|---|---|---|---|
| | | Accurate | Inaccurate | Total |
| Reported a multiple of 5? | No | 33 | 71 | *104* |
| | Yes | 15 | 43 | *58* |
| | | *48* | *114* | *162* |

**OLS and Poisson regression results**

Evidence presented in previous sections suggest that not all multiples of five are the result of low numeric abilities, and that age statements that are *not* multiples of five do not necessarily indicate accurate age statements. This finding appears inconsistent with the implications of the conventional interpretation of age heaping and numeracy indicators. We address this finding by assessing the correlates of accuracy in a set of regression models; we are especially interested in the relationship between the commonly used binary numeracy indicator and the degree of accuracy we observe in our sample. If the conventional numeracy indicators, such as WI and ABCC, reflect basic numeracy in a population, we expect to find a correlation between migrants' heaping patterns and the degree of inaccuracy in their age statements.

Figure 4: Absolute deviation between self-reported age and verified age




We investigate this relationship by using absolute differences between accurate age and reported age to assess the degree of accuracy in age statements. A histogram presented in figure 4 illustrates that the dependent variable follows a right-skewed distribution. Approximately one third of all age statements and computed ages are identical; most self-reported ages deviate from the accurate age by one or two years with some few individuals showing extreme deviation of up to thirteen years.

We use OLS and Poisson regression techniques to assess the correlates of this 'inaccuracy' in age statements (table 4). In a set of five OLS regression models (models one to five) we ignore the fact that the dependent variable is right-skewed and follows a discrete distribution; a second set of Poisson regressions (models six to 10) serves as an alternative test to address these very features of the dependent variable. Generally, we design models one and six as simple tests on whether the degree of imprecision in age statements is correlated with the likelihood of being 'numerate', i.e. reporting an age not ending in zero of five. These models are designed as a test of one of the central assumptions of numeracy indicators in that they test for a relationship between multiples of five and inaccurate age statements. If we were to find a negative correlation between 'numeracy' and inaccuracy, this evidence would suggest that excessive occurrence of multiples of five in an age distribution may be actually the result of an individual's inaccuracy.

These results support the claim that age heaping-based numeracy indicators reflect accuracy, since the coefficients in both regression specifications indicate a negative and statistically significant relationship between numeracy and the degree of imprecision. In the remaining models we test whether the degree of rounding is correlated with a set of socioeconomic variables. All models testing for religion-specific differences suggest that Roman Catholics report their age systematically less accurately compared to Protestants in the sample. A set of binary variables identifying different age groups is supposed to capture any changes in numeracy over time. Also, a control for age-related effects to control for the phenomenon that 23 to 32 year-olds tend to be better at reporting a correct age due to their proximity to important events in life, such as marriage and military service. The results suggest that there are indeed differences in terms of inaccuracy between different age groups. The group of 23 to 33 year-olds shows the highest accuracy, while we find that 63 to 72 year-olds are characterised by the lowest accuracy. In addition, we add a series of binary variables identifying the unit place of each individual's true age. In our case, we observe predominantly heaping on 0 rather than on 5, we therefore expect precision to improve with high and low values such as 0, 1 and 9.[32] The rationale behind this is that if rounding due to lacking numeracy skills occurs, someone with an age near a multiple of five might be naturally more precise compared to someone with an age that is further away from a multiple of five; but an age ending in zero or nine might come with a higher proneness to round compared with other digits. In other words, a true age ending in 1 or 0 would increase the probability of reporting a heaped age by quite a bit, which reduces the accuracy of reported ages. But the degree of inaccuracy in these cases is small; the two factors tend to cancel each other out. Accordingly, neither regression coefficients nor levels of statistical significance in the main specifications testing this hypothesis (models three and eight) suggest such a pattern.[33] We test the difference between individuals in the 'low occupation' group and the 'high occupation' group in terms of accuracy. Results presented in models 4 and 5 (OLS) as well as 9 and 10 (Poisson) indeed suggest that migrants with a 'low' occupational background tend to report a less accurate age compared with their 'high occupation' peers.

[32] Age-group and terminal digit variables are based on true, verified ages.

[33] We would like to thank an anonymous reviewer for pointing this out.

Table 4: Correlates of inaccuracy in age statements

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) | (10) |
|---|---|---|---|---|---|---|---|---|---|---|
| Method | OLS | | | | | Poisson | | | | |
| Dependent variable | Imprecision (absolute deviation) | | | | | Imprecision (absolute deviation) | | | | |
| **Numeracy** | | | | | | | | | | |
| Numerate | -1.01* | | | | | -0.39** | | | | |
| (i.e. age not ending in 0 or 5) | (-1.92) | | | | | (-2.02) | | | | |
| Non-numerate | *reference* | | | | | *reference* | | | | |
| (i.e. age ending in 0 or 5) | | | | | | | | | | |
| **Religion** | | | | | | | | | | |
| Roman-Catholic | | 1.75** | 1.56** | 3.67*** | 3.30* | | 0.59*** | 0.48** | 0.70** | 0.60 |
| | | (2.17) | (2.05) | (4.07) | (2.02) | | (2.76) | (2.37) | (2.00) | (1.45) |
| Protestant | | *reference* | *reference* | *reference* | *reference* | | *reference* | *reference* | *reference* | *reference* |
| **Age group** | | | | | | | | | | |
| 23 to 32 | | *reference* | *reference* | *reference* | *reference* | | *reference* | *reference* | *reference* | *reference* |
| 33 to 42 | | 1.09** | 1.23** | 0.39 | 0.44 | | 0.51** | 0.55** | 0.31 | 0.37 |
| | | (2.14) | (2.32) | (0.67) | (0.60) | | (2.01) | (2.21) | (0.86) | (1.16) |
| 43 to 52 | | -0.15 | 0.15 | 0.26 | 0.44 | | -0.07 | 0.08 | 0.21 | 0.32 |
| | | (-0.24) | (0.24) | (0.38) | (0.53) | | (-0.19) | (0.22) | (0.45) | (0.72) |
| 53 to 62 | | 1.47* | 1.76** | 1.02 | 1.04 | | 0.63** | 0.75** | 0.61* | 0.65* |
| | | (1.80) | (2.07) | (1.48) | (1.45) | | (2.04) | (2.35) | (1.67) | (1.90) |
| 63 to 72 | | 4.56** | 5.07*** | 4.16** | 3.66* | | 1.39*** | 1.60*** | 1.32*** | 1.16*** |
| | | (2.60) | (2.82) | (2.49) | (1.79) | | (3.83) | (4.21) | (3.72) | (3.00) |
| **Age ending in** | | | | | | | | | | |
| "0" | | | *reference* | | *reference* | | | *reference* | | *reference* |
| "1" | | | 0.35 | | 1.55 | | | 0.29 | | 0.77 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | (0.37) | | (1.31) | | | (0.49) | | (0.97) |
| "2" | | | 1.53* | | 2.58* | | | 0.75 | | 1.44* |
| | | | (1.66) | | (1.96) | | | (1.40) | | (1.77) |
| "3" | | | 0.23 | | 2.74 | | | 0.22 | | 1.41* |
| | | | (0.25) | | (1.62) | | | (0.38) | | (1.72) |
| "4" | | | -0.08 | | 2.87** | | | 0.01 | | 1.56* |
| | | | (-0.10) | | (2.11) | | | (0.03) | | (1.93) |
| "5" | | | 1.14 | | 1.22 | | | 0.59 | | 0.60 |
| | | | (1.01) | | (1.00) | | | (0.99) | | (0.71) |
| "6" | | | 2.14 | | 1.84 | | | 0.85 | | 0.98 |
| | | | (1.53) | | (1.37) | | | (1.50) | | (1.20) |
| "7" | | | 0.64 | | 2.10 | | | 0.37 | | 1.06 |
| | | | (0.72) | | (1.12) | | | (0.70) | | (1.26) |
| "8" | | | 1.12 | | | | | 0.60 | | |
| | | | (0.95) | | | | | (0.98) | | |
| "9" | | | 1.52 | | 1.08 | | | 0.72 | | 0.17 |
| | | | (1.41) | | (0.67) | | | (1.28) | | (0.16) |
| **Occupational background** | | | | | | | | | | |
| Unskilled & Semi-skilled | | | | 1.40* | 1.91* | | | | 0.61** | 0.87*** |
| | | | | (1.83) | (2.01) | | | | (2.07) | (2.65) |
| Skilled & Professional | | | | *reference* | *reference* | | | | *reference* | *reference* |
| **Constant** | 3.12*** | 1.44*** | 0.50 | 0.91*** | -1.35 | 1.14*** | 0.40* | -0.11 | 0.03 | -1.23 |
| | (6.94) | (3.81) | (0.66) | (3.12) | (-0.98) | (7.92) | (1.75) | (-0.21) | (0.12) | (-1.50) |
| **Observations** | 162 | 162 | 162 | 52 | 52 | 162 | 162 | 162 | 52 | 52 |
| **R-squared** | 0.03 | 0.14 | 0.18 | 0.39 | 0.45 | | | | | |

Robust t-statistics in parentheses; *** p<0.01, ** p<0.05, * p<0.1.

## Conclusion

We investigate whether age heaping-based indicators reflect basic numeracy and human capital. Numeracy indicators are based on the relative frequency of excessive occurrence of ages that are multiples of five. In empirical analyses tendencies and correlations between the likelihood to report a multiple of five allows establishing statistical relationships between socioeconomic backgrounds and basic numeracy. However, this procedure does not allow testing for the accuracy of an *individual* age statement, nor is it possible to assess the *degree* of inaccuracy of an inaccurate age statement. Due to the lack of precise information it is generally assumed that excessive occurrence of ages ending in "5" or "0" reflects the lack of numeracy in a population.

We address this gap of knowledge and provide empirical evidence that traditional age heaping metrics indeed reflect a form of basic human capital. We know self-reported ages of a sample of migrants travelling from the Holy Roman Empire to the Kingdom of Hungary in the 1780s. We are also able to compute these migrants' true ages, using church records and family heritage books. Comparing these two independent data sources allows us to distinguish correct and incorrect age statements and to identify the degree of inaccuracy for each individual in our sample. The majority of individuals shows modest inaccuracies; approximately 30 per cent of all individuals reported an accurate age, while approximately 25 per cent reported an age that deviates from the accurate age by only one year. Some few self-reported ages are found deviate by up to thirteen years from the true age. We also find that not everyone who reports a multiple of five reports an incorrect age, nor everyone who reports an age that is *not* a multiple of five reports an accurate age. In fact, we find that nine per cent of all individuals correctly report a multiple of five, while 44 per cent of the total sample, are found to misreport their age while reporting an age that is *not* a multiple of five. This finding is inconsistent with the underlying assumptions of age heaping which imply that the excessive occurrence of multiples of five are due to innumerate individuals rounding off to more attractive numbers, such as multiples of five. We assess the correlates of this 'age accuracy' metric in a set of OLS and Poisson regressions and show that it is correlated with the traditional ‘age heaping’ metric. We also use information about these migrants’ occupations to show that a more sophisticated occupational background reduces inaccuracy in age statements, and that Protestantism is associated with higher accuracy.

A limitation of this study is the absence of evidence for an individual’s reason to report an inaccurate age. It is plausible that some elderly migrants systematically under-reported their age in order to signal physical strength and to have an edge in the competition for large pieces of land in the land allocation process. Moreover, this sample is potentially selected due to the matching process, leading to a disproportionate large number of Protestants and agriculturalists in the sample. We also cannot rule out that there is selection on unobservable factors that are related to the data compilation and matching process, appearance of family names in official documents, or the decisions to compile family heritage books of certain villages in the first place. Finally, we are able to explain some variation in inaccuracy in age statements, but there is still space for other factors influencing this accuracy. Future research may help disentangling the roles of numeric abilities and other forces, and the awareness of the importance of numeracy and accuracy more generally.

## References


A'Hearn, B., Baten, J. and Crayen, D., 2009. Quantifying quantitative literacy: Age heaping and the history of human capital. *The Journal of Economic History*, *69*(03), pp. 783-808.

Armstrong, W.A., 1972. The use of information about occupation. *Nineteenth-century society: essays in the use of quantitative methods for the study of social data*, pp. 191-310.

Bachi, Roberto, 1951. The Tendency to Round off Age Returns: Measurement and Correction. *Bulletin of the International Statistical Institute,* 33, pp. 195- 221.

Baten, J., Crayen, D. and Voth, H.J., 2014. Numeracy and the impact of high food prices in industrializing Britain, 1780–1850. *Review of Economics and Statistics*, *96*(3), pp. 418-430.

Baten, J. and Juif, D., 2014. A story of large landowners and math skills: Inequality and human capital formation in long-run development, 1820–2000. *Journal of comparative economics*, *42*(2), pp. 375-401.

Baten, J. and Fourie, J., 2015. Numeracy of Africans, Asians, and Europeans during the early modern period: new evidence from Cape Colony court registers. *The Economic History Review*, *68*(2), pp. 632-656.

Baten, J. and Priwitzer, S., 2015. Social and intertemporal differences of basic numeracy in Pannonia (first century BCE to third century CE). *Scandinavian Economic History Review*, *63*(2), pp. 110-134.

Baten, Joerg, Szoltysek, Mikołaj and Campestrini, Monica, 2016. "Girl Power" in Eastern Europe? The human capital development of Central-Eastern and Eastern Europe in the seventeenth to nineteenth centuries and its determinants. *European Review of Economic History*.

Baten, J. and Sirbiladze, T., 2014. War der Adel eine Elitegruppe? Die ‚Numeracy' von Sozialgruppen im Georgien des 19. Jahrhunderts. *Vierteljahrschrift für Sozial- und Wirtschaftsgeschichte*, *101*(2), pp. 125-138.

Baten, J., Crayen, D. and Manzel, K., 2008. Zahlenfähigkeit und Zahlendisziplin in Nord-und Westdeutschland, 16.-18. Jahrhundert [Numerical Abilities and Numerical Discipline in Northern and Western Germany, 16th to 18th Centuries]. *Jahrbuch für Wirtschaftsgeschichte/Economic History Yearbook*, *49*(2), pp. 217-229.

Blum, M., Colvin, C. L., McAtackney, L. and McLaughlin, E., 2016. Women of an uncertain age: quantifying human capital accumulation in rural Ireland in the nineteenth century. *The Economic History Review, 70*(1), pp. 187-223.

Crayen, D. and Baten, J., 2010. Global trends in numeracy 1820–1949 and its implications for long-term growth. *Explorations in Economic History*, *47*(1), pp. 82-99.

Davids, K. 2012. *Religion, technology, and the great and little divergences: China and Europe compared, c. 700-1800* (Vol. 32). Brill.

De Moor, T. and Zuijderduijn, J., 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), pp. 41-56.

De Moor, T. and van Zanden, J.L., 2010. "Every Woman Counts": a gender-analysis of numeracy in the Low Countries during the Early Modern Period. *Journal of interdisciplinary history, 41*(2), pp. 179-208.

Fata, M., Migration im kameralistischen Staat Josephs II. Theorie und Praxis der Ansiedlungspolitik in Ungarn, Siebenbürgen, Galizien und der Bukowina von 1768 bis 1790 [Migration in the cameralistic state of Joseph I. Theory and Practice of the colonisation policy in Hungary, Transylvania, Galicia and Bukowina,1768-1790], Münster 2014.

Földvári, P., Van Leeuwen, B. and Van Leeuwen-Li, J., 2012. How did women count? A note on gender-specific age heaping differences in the sixteenth to nineteenth centuries1. *The Economic history review*, *65*(1), pp. 304-313.

Feldtänzer, O., Joseph II. und die donauschwäbische Ansiedlung. Dokumentation der Kolonisation im Batscherland 1784-1787 [Joseph II. and the Danube Swabian colonisation in the Bačka], München 1990.

Hacker, W., Auswanderung aus dem Raum der späteren Hohenzollerischen Lande nach Südosteuropa im 17. und 18. Jahrhundert [Emigration from Hohenzollern to South-East Europe in the 17th and 18th centuries], Sigmaringen 1969

Hacker, W., Auswanderer vom Oberen Neckar nach Südosteuropa im 18. Jahrhundert [Emigrants from the Upper Neckar region to South-East Europe in the 18th century], München 1970.

Hein, G., Die Einwohner von Neuwerbaß in der Batschka 1785-1944 und ihre Herkunft. Die Familien nach dem reformierten Kirchenbuch Neuwerbaß [The inhabitants of Neuwerbaß in the Bačka, 1785-1944 and their origins]. Sindelfingen 2004.

Hefner, A., Tscherwenka. Cservenka-Crvenka, Batschka 1785-1944. Tscherwenkaer Familien [The families of Crvenka]. CD-ROM. 2nd edition, Karlsruhe 2002.

Hippe, R., 2012. How to measure human capital? The relationship between numeracy and literacy. *Economies et sociétés*, *45*(8), pp. 1527-1554.

Hippe, R. and Baten, J., 2012. Regional inequality in human capital formation in Europe, 1790–1880. *Scandinavian Economic History Review*, *60*(3), pp. 254-289.

Juif, D.T. and Baten, J., 2013. On the human capital of Inca Indios before and after the Spanish Conquest. Was there a "Pre-Colonial Legacy"? *Explorations in Economic History*, *50*(2), pp. 227-241.

Krauss, K.-P.: Wirtschaftliche Rahmenbedingungen der Auswanderung aus Württemberg nach Russland 1817. In: Absmeier, Christine; Röder, Annemarie (Eds.): Flucht vor der Reformation. Täufer, Schwenckfelder und Pietisten zwischen dem deutschen Südwesten und dem östlichen Europa.Stuttgart 2016, pp. 43-53.

Möhler, J., Ortssippenbuch Liebling im Banat [Local Heritage Book of Liebling in the Banat], Altshausen 1979.

Myers, R.J., 1954. Accuracy of age reporting in the 1950 United States census. *Journal of the American Statistical Association*, *49*(268), pp. 826-831.

Myers, R.J., 1976. An instance of reverse heaping of ages. *Demography*, pp. 577-580.

Mokyr, J., 1983. *Why Ireland starved: a quantitative and analytical history of the Irish economy, 1800-1850*. Routledge.

Scherer, P., Familienbuch Weprowatz in der Batschka [Local Heritage Book of Weprowatz in the Bačka], Volume 1 and 2. Karlsruhe 1998.

Schuy, J., Ortssippenbuch Miletitsch [Local Heritage Book of Miletitsch], Schönaich 1987.

Seewann, G., Krauss, K.-P., Spannenberger, N., 2010. Die Ansiedlung der Deutschen in Ungarn. Beiträge zum Neuaufbau des Königreiches nach der Türkenzeit. R. Oldenbourg, München.

Stolz, Y., Baten, J. and Reis, J., 2013a. Portuguese living standards, 1720–1980, in European comparison: heights, income, and human capital1. *The Economic History Review*, *66*(2), pp. 545-578.

Stolz, Y. and Baten, J., 2012. Brain drain in the age of mass migration: Does relative inequality explain migrant selectivity? *Explorations in Economic History*, *49*(2), pp. 205-220.

Stolz, Y., Baten, J. and Botelho, T., 2013b. Growth effects of nineteenth-century mass migrations: "Fome Zero" for Brazil? *European Review of Economic History*, *17*(1), pp. 95-121.

**Archival sources**

Magyar Nemzeti Levéltár-Magyar Országos Levéltár (MNL-MOL) [Ungarisches Nationalarchiv-Ungarisches Landesarchiv], Kancelláriai Levéltár [Archiv der Ungarischen Hofkanzlei], Acta Generalia, A 39.

Österreichisches Staatsarchiv, Finanz- und Hofkammerarchiv, Wien (FHKA), Neue Hofkammer, Galizische Domänen Akten, Ansiedlungsgeschäft, 151, 1783.08-1783.12.