

The data revolution in African economic history¹

JOHAN FOURIE²

Large datasets that analyze the behavior of millions of individuals or that capture the complexities of the galaxy have, during the last decade, yielded extraordinary scientific results. Improvements in computing power, increased interconnectedness and more advanced analytical techniques have heralded the era of Big Data in fields as diverse as astronomy³, economics⁴, biology⁵, and management⁶.

Yet in the halls of the history profession, this data revolution has gained limited traction. The obvious reason is the scale of data in the distant past. As the former CEO of Google, Eric Schmidt, succinctly put it in 2010, between the dawn of civilization through 2003, five exabytes of information was created. Only seven years later, that much information was created every two days.

Of course, to paraphrase a well-worn aphorism, not all data that is generated is useful, nor is everything that is useful captured in data. Indeed, it is often the historians' main duty to decipher the more useful data, qualitative and quantitative, from that which is meaningless. Too much information can be as problematic as too little.

It is the purpose of this paper, however, to show that the benefits from the data revolution in history can be consequential. The surge in computing power and access to data processing software and online resources have allowed historians over the last two decades to capture historical statistics on a much greater scale than what was possible before.⁷ The applicability of the data revolution to the past is especially valuable, I argue, in areas with limited written records, like sub-Saharan Africa. Here I document how a new generation of economists, geographers and historians have undertaken to rewrite African history based not only on thousands of individual archival records (colonial sources often written for purposes entirely

¹ I acknowledge the financial support from the Elite Project Fund of the Faculty of Economic and Management Sciences at Stellenbosch University, and the Oppenheimer Memorial Trust Sabbatical Study Fund. I am also indebted to my sabbatical study host Jan Luiten van Zanden at Utrecht University.

² LEAP, Department of Economics, Stellenbosch University. E-mail: johanf@sun.ac.za.

³ Feigelson, Eric D., and G. Jogesh Babu. "Big data in astronomy." *Significance* 9.4 (2012): 22-25.

⁴ Liran Einav and Jonathan Levin, *The Data Revolution and Economic Analysis*, NBER Working Paper, 2014.

⁵ Marx, Vivien. "Biology: The big challenges of big data." *Nature* 498.7453 (2013): 255-260; Howe, A. D., Costanzo, M., Fey, P., Gojobori, T., Hannick, L., Hide, W., ... Rhee, S. Y. (2008). Big data: The future of biocuration. *Nature*, 455(7209), 47-50.

⁶ McAfee, Andrew, and Erik Brynjolfsson. "Big data: the management revolution." *Harvard business review* 90 (2012): 60-6.

⁷ Manning, Patrick. *Big data in history*. Palgrave Macmillan, 2013.

orthogonal to current research questions), but also on geographical, climatic and demographic projections into the distant past, with surprising answers and new challenges.

A NEW AFRICAN ECONOMIC HISTORY

African history has always had to overcome the problem of limited written records. Henry Morton Stanley's quip of Africa as the 'Dark Continent' was, at least until the end of the colonial period, a reflection not only of its geography but also of its history. It was only in the 1960s when most African countries had gained independence that historians, within Africa and outside, borrowed from archeology, linguistics and oral history to investigate the continent's rich and varied past.

It was Africa's economic history, in particular, that was in high demand. The euphoria immediately following independence and the rapid economic growth rates during the late colonial and early post-colonial periods, raised unanswered questions about African societies' past fortunes, the impact of imperialism and colonialism, and the possibilities of future prosperity. Economic History departments were established in universities across the continent. Debates could be fierce, informed by the ideological battles between Marxism and liberal economics. Evidence in support of hypotheses came most often from colonial-era records, notably reports, letters and other such qualitative sources.

Sadly, by the 1980s, three forces had coalesced to push African economic history to the intellectual periphery.⁸ First, Africa's economies were in decline. The oil shocks of the 1970s and the consequent debt crises and IMF structural adjustment programmes had not only affected the resources available to African universities but had deflated interest in explaining past fortunes.⁹ Second, as postmodernism became fashionable, historians became more interested in cultural and social history rather than in general theories of economic development. Third, following the cliometric revolution of the 1970s, economics became more mathematical and statistical. A combination of these three forces meant that African economic history, both within and outside Africa, was in serious decline.

⁸ Green and Nyambara (2015) argue that African economic history only declined in Western institutions. See Green, Erik, and Pius Nyambara. "The Internationalization of Economic History: Perspectives from the African Frontier." *Economic History of Developing Regions* ahead-of-print (2015): 1-11. Austin (2015) suggests that it was also in decline in Africa. See Austin, Gareth. "African Economic History in Africa." *Economic History of Developing Regions* ahead-of-print (2015): 1-16.

⁹ Morten Jerven, 2010. "African growth recurring: An economic history perspective on African growth episodes, 1690-2010", *Economic History of Developing Regions*, 25, 2, 127-154.

It was not until the 2000s that African economic history would awake from its slumber. Led by prominent economists, the renaissance of African economic history would, much like the post-colonial period, coincide with the rise of Africa's economic prospects.¹⁰ The new African economic history, however, was also the result, I argue here, of another important change: the methodological and technological revolution within the economics profession. It was the dawn of the data revolution in African history.

While Acemoglu, Johnson and Robinson is widely credited with rekindling interest in African economic history with their provocative claim that the institutional legacies of European settlers in Africa explains much of the disparity in levels of development today, their main contribution was not the use of a large, new dataset but rather the causal inference they made through the use of a novel instrumental econometric technique. Their claim of the persistence of institutions attracted the interest of a younger generation of economists, though, eager to tackle the issues of African poverty and underdevelopment.

One of these was Nathan Nunn, then a PhD student at the University of British Columbia, who undertook an extensive study of the economic consequences of the slave trade. Using and expanding an already extensive list of African slaves exported to the slave markets in the Americas, Nunn proved the causal link between the number of slaves exported and the level of GDP in 2000.¹¹ After his appointment at Harvard, Nunn continued to ask questions about the African past: together with Leonard Wantchekon, Nunn identified trust as one mechanism how the slave trade still affects African societies; with Diego Puga, he showed how the slave trade pushed Africans into rugged areas, which still affect their ability to trade.

The use of environmental information is an especially useful tool to link past events to the present. As mentioned, Nunn and Puga use terrain ruggedness as explanatory variable. Marcella Alsan use measures of temperature and humidity to assess the impact of the Tsetse fly on current development outcomes in Africa.¹² And James Fenske and Namrata Kala use an annual panel of African temperatures and port-level slave exports to show how past environmental shocks affected the slave trade.¹³ The use of climatic and other environmental information

¹⁰ Austin, Gareth, and Stephen Broadberry. "Introduction: The renaissance of African economic history." *The Economic History Review* 67.4 (2014): 893-906; Fourie, Johan, and Leigh Gardner. "The internationalization of economic history: a puzzle." *Economic History of Developing Regions* 29.1 (2014): 1-14.

¹¹ Nunn, Nathan. 2008. "The Long-Term Effects of Africa's Slave Trades." *Quarterly Journal of Economics*, 123(1): 139-76.

¹² Alsan, Marcella. 2015. "The Effect of the TseTse Fly on African Development." *American Economic Review*, 105(1): 382-410.

¹³ Fenske, James, and Namrata Kala. "Climate and the slave trade." *Journal of Development Economics* 112 (2015): 19-32.

geocoded at the micro level allow these scholars to test the persistence of historical events with a high level of statistical accuracy.

The data revolution has not only offered new variables to measure past events, but it has been especially useful as a contemporary outcome variable. In the absence of data on regional African economies, for example, Michalopoulos and Papaioannou use light density at night obtained from satellite imaging to test the impact of precolonial ethnic institutions¹⁴ and the impact of the borders drawn during the Scramble for Africa¹⁵ on current economic performance. Others have followed this approach.¹⁶

Contemporary survey data, now with geocoded observations, are also popular outcome variables. Nunn's paper with Wantchekon on the origins of mistrust is a case in point, using the Afrobarometer survey to measure levels of trust. Martin Abel's work on forced removals during apartheid also relies on the Afrobarometer¹⁷, while Michalopoulos, Putterman and Weil match Demographic and Health Survey (DHS) data to information on the ancestral ethnicities of respondents to show that descendants of agriculturalists are wealthier and better educated than the descendants of groups that practiced pastoral farming.¹⁸

Not everyone agree, though, that these broad-brush studies, linking an historical episode with contemporary data, are necessarily useful. The main critique labeled against such an approach is that it 'compresses history', in short, that it oversimplifies the causal links from the past experience to the present.¹⁹ Morten Jerven, in a recent book, goes as far as to say that African economic history should be liberated from the tyranny of econometricians.²⁰

African economic historians, instead, have proposed an alternative data approach, one that focuses on unearthing and digitizing colonial-era archival records. Over the last decade, scholars

¹⁴ Michalopoulos, Stelios, and Elias Papaioannou. "Pre-Colonial Ethnic Institutions and Contemporary African Development." *Econometrica* 81.1 (2013): 113-152; Michalopoulos, Stelios & Papaioannou, Elias, 2015.

"Further evidence on the link between pre-colonial political centralization and comparative economic development in Africa," *Economics Letters*, Elsevier, vol. 126(C), pages 57-62.

¹⁵ Michalopoulos, Stelios, and Elias Papaioannou. 2014. "National Institutions and Subnational Development in Africa," *The Quarterly Journal of Economics*, Oxford University Press, vol. 129(1), pages 151-213

¹⁶ Besley, Timothy, and Marta Reynal-Querol. "The legacy of historical conflict: Evidence from Africa." *American Political Science Review* 108.02 (2014): 319-336; Obikili, Nonso. "An Examination of Subnational Growth in Nigeria: 1999-2012." *South African Journal of Economics* (2015).

¹⁷ Abel, Martin. "Long-run effects of forced removal under apartheid on social capital." Presented at: African Economic History meetings, London School of Economics and Political Science, 25–26 October 2014. Harvard University, Cambridge, MA, 2014.

¹⁸ Michalopoulos, Stelios, Louis Putterman, and David N. Weil. "The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa." (2014).

¹⁹ Austin, Gareth. "The "reversal of fortune" thesis and the compression of history: Perspectives from African and comparative economic history." *Journal of international development* 20.8 (2008): 996-1027.

²⁰ Jerven, Morten, and Deborah Johnston. "Statistical Tragedy in Africa? Evaluating the Data Base for African Economic Development." *The Journal of Development Studies* 51.2 (2015): 111-115.

have ploughed through colonial Blue Books, tax censuses, voters' rolls, marriage registers and more to gain insights into African societies' population size²¹, wages²², incomes²³, education²⁴, inequality²⁵, fiscal systems²⁶ and transport networks²⁷. Instead of discussing all these contributions, I illustrate the application of the data revolution in African history with reference to one type of archival document: military attestation forms. These forms are useful because it includes individual-level observations of heights, an indicator that is widely used as a proxy for measuring living standards in the absence of other evidence.

A CASE STUDY OF HEIGHTS

The use of human heights or stature as a proxy for standards of living is now more than three decades old.²⁸ Social scientists agree that height generally reflects an individual's early-life living conditions, including access to nutrition and the prevalent disease environment, and that

²¹ Frankema, Ewout, and Morten Jerven. "Writing history backwards or sideways: towards a consensus on African population, 1850–2010." *The Economic History Review* 67.4 (2014): 907-931; Fourie, J. and Green, E. 2015. *The Missing People: Accounting for the productivity of indigenous populations in Cape Colonial History*. *Journal of African History*. Vol 56(2): 195-215.

²² Frankema, Ewout, and Marlous Van Waijenburg. "Structural impediments to African growth? New evidence from real wages in British Africa, 1880–1965." *The Journal of Economic History* 72.04 (2012): 895-926; Rönnbäck, Klas. "Living standards on the pre-colonial Gold Coast: a quantitative estimate of African laborers' welfare ratios." *European Review of Economic History* 18.2 (2014): 185-202; Du Plessis, Sophia, and Stan Du Plessis. "Happy in the service of the Company: the purchasing power of VOC salaries at the Cape in the 18th century." *Economic History of Developing Regions* 27.1 (2012): 125-149.

²³ Jerven, Morten. "A West African experiment: constructing a GDP series for colonial Ghana, 1891–1950." *The Economic History Review* 67.4 (2014): 964-992; Prados de la Escosura, Leandro. "Output per head in pre-independence Africa: quantitative conjectures." *Economic History of Developing Regions* 27.2 (2012): 1-36; Fourie, Johan, and Jan Luiten Zanden. "GDP in the Dutch Cape Colony: The National Accounts of a Slave-Based Society." *South African Journal of Economics* 81.4 (2013): 467-49.

²⁴ Wantchekon, Leonard, Marko Klačnja, and Natalija Novta. "Education and Human Capital Externalities: Evidence from Colonial Benin." *The Quarterly Journal of Economics* 130.2 (2015): 703-757; Fourie, Johan, and Dieter Fintel. "Settler skills and colonial development: the Huguenot wine-makers in eighteenth-century Dutch South Africa." *The Economic History Review* 67.4 (2014): 932-963; Obikili, Nonso. "Social Capital and Human Capital in the Colonies: A Study of Cocoa Farmers in Western Nigeria." *Economic History of Developing Regions* ahead-of-print (2015): 1-22; Baten, Jörg, and Johan Fourie. "Numeracy of Africans, Asians, and Europeans during the early modern period: new evidence from Cape Colony court registers." *The Economic History Review* 68.2 (2015): 632-656.

²⁵ Hillbom, Ellen, and Jutta Bolt. "Changing income inequality and structural transformation: The case of Botswana 1921-2010." *WIDER Working Paper* 2015 (2015); Fourie, J. and Von Fintel, J. 2011. *A History with Evidence: Income inequality in Dutch South Africa*. *Economic History of Developing Regions* 26(1): 16-48

²⁶ Gardner, Leigh A. *Taxing colonial Africa: the political economy of British imperialism*. OUP Oxford, 2012; Fourie, J., Jansen, A. and Siebrits, K. 2013. *Public finances and private company rule: The Dutch Cape Colony (1652-1795)*. New Contree. December 2013; Frankema, Ewout, and Marlous van Waijenburg. "Metropolitan blueprints of colonial taxation? Lessons from fiscal capacity building in British and French Africa, c. 1880–1940." *The Journal of African History* 55.03 (2014): 371-400; Frankema, Ewout. "Colonial taxation and government spending in British Africa, 1880–1940: Maximizing revenue or minimizing effort?." *Explorations in Economic History* 48.1 (2011): 136-149.

²⁷ Jedwab, Remi, and Alexander Moradi. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." *Review of Economics and Statistics* (2015).

²⁸ Fogel, Robert W., et al. "Secular changes in American and British stature and nutrition." *Journal of interdisciplinary History* (1983): 445-481; Fogel, Robert W., Stanley L. Engerman, and James Trussell. "Exploring the uses of data on height: The analysis of long-term trends in nutrition, labor welfare, and labor productivity." *Social Science History* (1982): 401-421; Steckel, Richard H. "Height and per capita income." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 16.1 (1983): 1-7.

changes in the average height of society are borne out by changes to the living conditions of those people.²⁹

But while we would expect that heights positively correlate with incomes, economic historians soon discovered, when analyzing heights in Western Europe and North America, a surprising negative correlation between height and output during the early phase of industrialization in England and elsewhere.³⁰ This became known as the ‘Early Industrial Growth Puzzle’ in Europe or the ‘Antebellum Puzzle’ in the US, and for most of the last two decades of the twentieth century scholars on both sides of the Atlantic would attempt to explain this puzzle.

Although the heights of nineteenth-century Africans were already the subject of a David Eltis paper in 1982³¹, it was only in the mid-2000s that African attestations – and therefore heights – would receive the attention it deserves. Alexander Moradi and Jörg Baten were pioneers in using heights as a way to document the evolution of living standards of African peoples in an era of unreliable data³²; they used DHS data to construct historical birth-cohorts to investigate inequality in heights.³³ But the DHS data could only be usefully employed in analysis since the 1950s, which was at the end of the colonial era. Another source was necessary to investigate the colonial, and hopefully pre-colonial, era.

In 2008, Moradi published a paper in the *Journal of International Development* that did exactly that.³⁴ He used a sample of 1046 Ghanaian WWI and WWII recruits and 730 Kenyan WWII recruits, in addition to surveys, to measure the impact of colonial policies on African living standards. In follow-up work on the same question, his sample more than doubles.³⁵ Moradi’s groundbreaking work stimulated interest in military records elsewhere in Africa, often with startling results. In nineteenth-century West Africa, Gareth Austin, Jörg Baten and Bas van

²⁹ Steckel, Richard H. "Stature and the Standard of Living." *Journal of Economic Literature* (1995): 1903-1940; Alter, George. "Height, frailty, and the standard of living: Modelling the effects of diet and disease on declining mortality and increasing height." *Population Studies* 58.3 (2004): 265-279; Deaton, Angus. "Height, health, and inequality: the distribution of adult heights in India." *The American economic review* 98.2 (2008): 468.

³⁰ Komlos, John. "Shrinking in a growing economy? The mystery of physical stature during the industrial revolution." *The Journal of Economic History* 58.03 (1998): 779-802; Komlos, John. "Anomalies in economic history: toward a resolution of the ‘Antebellum Puzzle’." *The Journal of Economic History* 56.01 (1996): 202-214; Komlos, John, and Bjorn Alecke. "The economics of antebellum slave heights reconsidered." *Journal of Interdisciplinary History* (1996): 437-457.

³¹ Eltis, David. "Nutritional trends in Africa and the Americas: Heights of Africans, 1819-1839." *Journal of Interdisciplinary History* (1982): 453-475.

³² Others included Sahn, David E., and David C. Stifel. "Urban–rural inequality in living standards in Africa." *Journal of African Economies* 12.4 (2003): 564-597.

³³ Moradi, Alexander, and Joerg Baten. "Inequality in Sub-Saharan Africa: New data and new insights from anthropometric estimates." *World Development* 33.8 (2005): 1233-1265.

³⁴ Moradi, Alexander. "Confronting colonial legacies—lessons from human development in Ghana and Kenya, 1880–2000." *Journal of International Development* 20.8 (2008): 1107-1121.

³⁵ Moradi, Alexander. "Towards an objective account of nutrition and health in colonial Kenya: a study of stature in African army recruits and civilians, 1880–1980." *The Journal of Economic History* 69.03 (2009): 719-754.

Leeuwen find that Ghanaian and Burkinabe recruits were notably shorter than north-western Europeans but not shorter than southern Europeans during this period.³⁶ In twentieth-century West Africa, Denis Cogneau and Lea Rouanet found that the pace of increase in the heights of those born in Côte d'Ivoire and Ghana during the late colonial period (1925–1960) was almost as high as the pace observed in France and Great Britain during the period 1875 to 1975.³⁷ In South Africa, Kris Inwood and Oliver Masakure found that Coloured South Africans were 6cm shorter than white South Africans at the start of the twentieth century, a significantly smaller difference than today.³⁸ And Johan Fourie, Kris Inwood and Martine Mariotti find that white heights increased significantly in the Boer Republics following the discovery of minerals, while whites born in South Africa's Western Cape saw no concomitant increase.³⁹ These results can now be integrated into a larger literature on the living standards of indigenous populations across the world.⁴⁰

Information on heights is thus useful not only as tools to measure the level of living standards over time, but also to test the effects of different colonial policies. Cogneau and Moradi use a much larger sample of recruits from Ghana and Togo – 11940 observations – to test the impact of the partition of German Togoland after World War I on education outcomes.⁴¹ The part of Togo that fell under French mandate had lower levels of literacy following the partition than the parts that were ruled by the British. They attribute this to the French authorities' hostile stance towards missionary schools. Remi Jedwab and Moradi analyze the effects of colonial railways on

³⁶ Austin, Gareth, Joerg Baten, and Bas Van Leeuwen. "The biological standard of living in early nineteenth-century West Africa: new anthropometric evidence for northern Ghana and Burkina Faso1." *The Economic History Review* 65.4 (2012): 1280-1302.

³⁷ Cogneau, Denis, and Léa Rouanet. "Living conditions in Côte D'ivoire and Ghana, 1925–1985: what do survey data on height stature tell us?" *Economic History of Developing Regions* 26.2 (2011): 55-82.

³⁸ Inwood, Kris, and Oliver Masakure. "Poverty and physical well-being among the Coloured population in South Africa." *Economic History of Developing Regions* 28.2 (2013): 56-82

³⁹ Fourie, Johan, Kris Inwood and Martine Mariotti, "White living standards during South Africa's mineral revolution: evidence from attestation forms", Mimeo: Stellenbosch University, 2015.

⁴⁰ Prince, Joseph M., and Richard H. Steckel. "Nutritional success on the great plains: nineteenth-century equestrian nomads." *Journal of Interdisciplinary History* 33.3 (2003): 353-384; Bodenhorn, Howard. "The mulatto advantage: the biological consequences of complexion in rural antebellum Virginia." *Journal of Interdisciplinary History* 33.1 (2002): 21-46; Guntupalli, Aravinda Meera, and Joerg Baten. "The development and inequality of heights in North, West, and East India 1915–1944." *Explorations in Economic History* 43.4 (2006): 578-608; Inwood, Kris, Les Oxley, and Evan Roberts. "Physical growth and ethnic inequality in New Zealand prisons, 1840–1975." *The History of the Family* ahead-of-print (2015): 1-21; Baten, Joerg, Ines Pelger, and Linda Twrdek. "The anthropometric history of Argentina, Brazil and Peru during the 19th and early 20th century." *Economics & Human Biology* 7.3 (2009): 319-333; Baten, Jörg, Mojgan Stegl, and Pierre van der Eng. "The biological standard of living and body height in colonial and post-colonial Indonesia, 1770–2000." *Journal of bioeconomics* 15.2 (2013): 103-122; Baten, Joerg, and Matthias Blum. "Growing tall but unequal: new findings and new background evidence on anthropometric welfare in 156 countries, 1810–1989." *Economic History of Developing Regions* 27.sup1 (2012): S66-S85

⁴¹ Cogneau, Denis, and Alexander Moradi. "Borders that divide: Education and religion in Ghana and Togo since colonial times." *The Journal of Economic History* 74.03 (2014): 694-729.

various economic outcomes including heights, finding that railways boosted the production of cash-crops like cocoa, improving incomes and the heights of those born closest to the railways.⁴²

A positive income shock on heights is also found by Martine Mariotti in an investigation on the heights of black mine workers in South Africa. A 1974 plane crash provoked the Malawian government to ban the migration of mine workers to South Africa, forcing mining authorities in South Africa to recruit from the Transkei, a homeland in South Africa. Mariotti show that the exogenous shock in household incomes to the newly recruited mineworkers of the Transkei increased the heights of children born during or immediately following 1974. This increase in heights happened to only those districts where mine workers were recruited from.⁴³ Mariotti together with Taryn Dinkelman have now turned their attention to the impact of the plane crash and the sudden stop to labour migration on Malawian households.⁴⁴

The value of heights data is that it provides a snapshot of living standards in the absence of other individual-level records. Colonial authorities were often less inclined to collect accurate data on indigenous populations. Where information is available, it comes with colonial-era biases that are difficult to disentangle. There is however no reason to suspect that military heights information would be subject to systematic biases; its use for researchers are orthogonal to the purposes for which it was collected.

Yet attestation forms have their own selection issues. Not everyone qualified for military service. Often minimum height requirements were imposed, which means that attestation heights are often truncated to the left. Such truncation can be checked statistically (through the use of truncated regression models, for example) but that does not entirely allay the fears that some other unobservable selection is present. These fears are indeed the subject of several new research papers by Howard Bodenhorn, Timothy Guinnane and Thomas Mroz.⁴⁵ They claim that the declining heights observed during the Industrial Revolution – the ‘Early Industrial Growth Puzzle’ – does not reflect a decline in living standards but instead a change in selection into the military. In short, they argue that the best men (i.e. tallest) would have chosen the military as a career when returns in the private sector was small (at the start of the Industrial Revolution). As

⁴² Jedwab, Remi, and Alexander Moradi. "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa." *Review of Economics and Statistics* (2015).

⁴³ Mariotti, Martine. "Fathers' Employment and Sons' Stature: The Long-Run Effects of a Positive Regional Employment Shock in South Africa's Mining Industry." *Economic Development and Cultural Change* 63.3 (2015): 485-514.

⁴⁴ Dinkelman, Taryn, and Martine Mariotti. *Does labor migration affect human capital in the long run? Evidence from Malawi*. Mimeo, Dartmouth College, Hanover, NH, 2014.

⁴⁵ Bodenhorn, Howard, Timothy W. Guinnane, and Thomas A. Mroz. *Sample-selection biases and the "industrialization puzzle"*. No. w21249. National Bureau of Economic Research, 2015; Bodenhorn, Howard, Timothy Guinnane, and Thomas Mroz. *Caveat Lector: Sample Selection in Historical Heights and the Interpretation of Early Industrializing Economies*. No. w19955. National Bureau of Economic Research, 2014.

the economy grew, however, returns to private sector employment would have increased, with no concomitant increase in the military. This would have meant that weaker men (i.e. shorter) would have selected a military career as the best men found employment in the private sector. This type of unobservable selection may also arise, Johan Fourie, Kris Inwood and Martine Mariotti show using South African recruits to World War I, because of changes in military technology.⁴⁶

Despite these concerns, heights provide a useful proxy for individual-level living standards in the absence of other measures. Information on heights show how the data revolution in African history have provided new insights into the welfare impact of colonial-era policies on subjects that remained outside the remit of colonial-era recordkeeping.

MORE THAN JUST LARGE DATASETS

The case study on heights has illustrated how the collection of large numbers of numeric data, particularly in regions like sub-Saharan Africa with few written records, can provide new insights into the past. But the availability of large, unexamined archival datasets is only one aspect of the data revolution; new tools of data transcription, augmentation and collaboration allow for faster and less expensive digitization, more rigorous investigations and the replicability and comparability of results. The data revolution is not only a shift to large datasets but a framework by which computational power and statistical techniques are employed to expand the scale and scope of research questions.

Let us investigate each of these aspects in turn. Data transcription – the manual rewriting of archival sources often into a digital format like Excel – is a laborious and expensive exercise. As an example we again turn to the attestation forms. Figure 1 shows the attestation of Diederick Alfred Joseph Yates, an Englishmen who signed up for the South African Constabulary in 1902. Although the form is standardized and the text clearly eligible to a skilled historian trained in reading late nineteenth-century handwriting, there is no possibility to automate the transcription of these attestations. Computer science algorithms are as yet unable to capture the information contained in these attestations with a high level of accuracy.

The historian must therefore spend countless hours transcribing these attestations into information that can be analyzed, or do so as part of a research team. Both options are costly. It often means that historians have to select samples, much like Kris Inwood and Oliver Masakure

⁴⁶ Fourie, Johan, Kris Inwood, and Martine Mariotti. Can historical changes in military technology explain the industrial growth puzzle?. Mimeo., London School of Economics, 2014.

did when investigating South African soldiers fighting in World War I and World War II. Although several hundred thousand of these attestations are available in the Defense Force archives, time and funding constraints meant that they could only transcribe 10 000 attestations for each of the Wars. Such sampling introduces additional biases that must be considered when analyzing the results. It also contradicts the methodology now popular in the era of Big Data, which advocates the use of all available data.

Figure 1: An attestation form from the South African Constabulary

The image shows an open document with two pages. The left page is titled 'ATTESTATION PAPER' and features a royal coat of arms at the top. Below it, the text reads 'E. R. ATTESTATION PAPER.' followed by a formal declaration in the name of 'EDWARD THE SEVENTH, His Majesty King' regarding the enlistment of a soldier. The soldier's name is 'Edward Alfred Joseph Vols.' The text continues with details about the enlistment date (1902) and the conditions of service. At the bottom of the left page, there is a section for 'SO HELP ME GOD' with a signature and a date '1902'. The right page is a form titled 'SOUTH AFRICAN CONSTABULARY. RECORD OF CONDUCT and SERVICE of Edward Alfred Joseph Vols.' The form is divided into several sections: 'PERSONAL DESCRIPTION' (with fields for height, eyes, hair, complexion, marks, general appearance, country, and where recruited), 'PROMOTIONS', 'TRANSFERS', and 'REDUCTIONS'. The 'REDUCTIONS' section has a large 'Discharged' stamp written vertically. There is also a section for 'Re-engagement' and 'Transfer to the Reserve' at the bottom of the right page. The form is filled with handwritten entries, including the soldier's name, rank, and dates.

There is hope, however, that the costly transcription of archival sources such as attestation forms could soon be replaced by the less expensive algorithms of computer scientists. Optical character recognition (OCR) is a rapidly developing technology that already has wide applicability in history, most notably in the Google Books project to scan all distinct book titles. The digitization of historical editions of newspapers has gained momentum too.⁴⁷ Note that these books and newspapers are printed which simplifies the OCR technology required to digitize and transcribe the content. Yet advances are being made in OCR of natural images which

⁴⁷ Gupta, Maya R.; Jacobson, Nathaniel P.; Garcia, Eric K. (2007). "OCR binarization and image pre-processing for searching historical documents." (PDF). Pattern Recognition 40 (2): 389.

contain written text.⁴⁸ It is only a matter of time before an image such as Figure 1 can be read by an algorithm with a high level of accuracy. Such a technological breakthrough in computational linguistics, with the aid of historians, can have large consequences for the field of economic history.

The second aspect of the data revolution is data augmentation. Data augmentation is the addition of information from internal or external sources with the aim of adding value to the original data. Although it has always been possible to add new information to existing data, it has almost always had to be done manually. Computerization and the development of matching algorithms have made this process much easier.⁴⁹ To return to the attestation case, it is now possible to match recruits' names and birth dates to, for example, genealogical records. This allows researchers to not only gain new insights into the recruits themselves, but also answer questions that were previously unsolvable, like calculating intergenerational mobility using height as an indicator of living standards. There are countless such new avenues which data augmentation can explore.

The third aspect of the data revolution is the ability to share and augment across research networks. There are already several examples in Economic History where data sharing have resulted in collaborative projects that aim to compare historical measures. At a macroeconomic level, series such as the Penn World Tables⁵⁰ and the Maddison project⁵¹ have provided scholars with cross-country comparisons back in time. University of Pittsburgh professor Patrick Manning's Big Data in History project aims to "create a world-historical archive that will trace the last four centuries of historical dynamics and change".⁵² And Jan Luiten van Zanden at Utrecht University has undertaken the Clio-infra project, which aims to establish "a set of interconnected databases has been set up containing worldwide data on social, economic, and institutional indicators for the past five centuries, with special attention to the past 200 years".⁵³ This project has now evolved into the CLARIAH project, consisting of a consortium of more than 40 partners who store and share all types of data related to the arts and humanities in a standardized and user-friendly format.

⁴⁸ Milyaev, Sergey; Barinova, Olga; Novikova, Tatiana; Kohli, Pushmeet; Lempitsky, Victor (2013). "Image binarization for end-to-end text understanding in natural images." Document Analysis and Recognition (ICDAR) 2013. 12th International Conference on. Retrieved 2 May 2015.

⁴⁹ See, for example, Feigenbaum (2015). James J. Feigenbaum, 2015. "Automated Census Record Linking", Mimeo: Harvard University. Available online: <http://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf>.

⁵⁰ Feenstra, Robert, Robert Inklaar, & Marcel Timmer, 2013. The next generation of the Penn World Table. NBER Working Paper 19255. National Bureau of Economic Research.

⁵¹ Bolt, Jutta, and Jan Luiten Zanden. "The Maddison Project: collaborative research on historical national accounts." *The Economic History Review* 67.3 (2014): 627-651.

⁵² Manning, Patrick. 2013. *Big Data in History*. Palgrave Macmillan.

⁵³ Van Zanden, Jan Luiten, et al. *How was life?: Global well-being since 1820*. OECD Publishing, 2014.

In African history, Ewout Frankema of Wageningen University and his students have taken the lead to digitize and transcribe the colonial Blue Books, allowing research on colonial education, public finance and population estimates.⁵⁴ In an award-winning study, Frankema and Marlous van Waijenburg show that African real wages were significantly higher than subsistence levels, and rising for most of the colonial period. In some places, wages were much higher than comparative figures for Asia. They suggest that the idea of sub-Saharan Africa is persistently poor is not supported by the evidence.⁵⁵

In a similar vein, Morten Jerven has reconstructed African Gross Domestic Product measures, finding large irregularities between existing World Bank, IMF and Penn World Table estimates.⁵⁶ In two highly-acclaimed books, Jerven emphasized that Africa's image as a 'hopeless continent' is really only a construction of the 1980s and 1990s.⁵⁷

Jutta Bolt and Leigh Gardner are compiling large amounts of fiscal revenue and expenditure data at the local level for the late colonial period.⁵⁸ They pair these new records with anthropological records on pre-colonial states to assess the extent to which the fiscal capacity of local government units reflected pre-colonial state centralization.

These reinterpretations of the African past are the consequence of digitizing, transcribing and also sharing the vast amounts of data available in colonial archives. And such projects are ongoing. One of these is a project led by Felix Meier zu Selhausen at Sussex University to digitize and transcribe vast quantities of demographic records kept in missionary station archives. These records are not without problems – selection into formal Christian mission stations, for example

⁵⁴ Ewout Frankema & Morten Jerven, 2014. "Writing history backwards or sideways: towards a consensus on African population, 1850–2010," *Economic History Review*, Economic History Society, vol. 67(4), pages 907-931, November; Ewout H.P. Frankema, 2012. "The origins of formal education in sub-Saharan Africa: was British rule more benign?," *European Review of Economic History*, Oxford University Press, vol. 16(4), pages 335-355, November; Frankema, Ewout, 2011. "Colonial taxation and government spending in British Africa, 1880-1940: Maximizing revenue or minimizing effort?," *Explorations in Economic History*, Elsevier, vol. 48(1), pages 136-149, January.

⁵⁵ Frankema, Ewout & Waijenburg, Marlous Van, 2012. "Structural Impediments to African Growth? New Evidence from Real Wages in British Africa, 1880–1965," *The Journal of Economic History*, Cambridge University Press, vol. 72(04), pages 895-926, December

⁵⁶ Jerven, Morten. "African growth recurring: an economic history perspective on African growth episodes, 1690–2010." *Economic History of Developing Regions* 25.2 (2010): 127-154; Jerven, Morten. "For richer, for poorer: GDP revisions and Africa's statistical tragedy." *African Affairs* (2012): ads063.

⁵⁷ Jerven, Morten. *Poor numbers: how we are misled by African development statistics and what to do about it*. Cornell University Press, 2013; Jerven, Morten. *Africa: Why Economists Get It Wrong*. 2015.

⁵⁸ Jutta Bolt & Leigh Gardner, 2015. "De-compressing History? Pre-colonial institutions and local government finance in British colonial Africa," Paper presented at the Economic History Association meetings, Nashville, 11 September 2015.

– but they do provide a glimpse of African demographic change and living standards unrecorded in the colonial Blue Books.

Of course, more detailed records are available of settlers and their activities. At Stellenbosch University in South Africa, Erik Green, Dieter von Fintel and I are constructing an annual panel data set of several thousand settler farmers over more than 140 years of settlement. Once complete, the dataset, augmented with information from probate inventories and genealogical records, should provide a thorough account of eighteenth- and nineteenth century life in colonial South Africa, and provide economists with a dataset to understand the economic consequences of exogenous shocks such as the abolishment of slavery or smallpox epidemics. Yet these records also provide detailed demographic information on the Khoesan, a population group almost completely neglected in the quantitative sources.⁵⁹

The scale of the project – several thousand individuals are recorded annually – does not allow for matching individuals manually across years. Instead, a self-learning algorithm identifies unique individuals, and links them to an individual ID. The same individual can thus be observed over his entire lifetime. By adding genealogical records, the algorithm can be expanded to also link families across generations. Such an intergenerational panel would not have been infeasible without the advent of large computing power and powerful statistical techniques.

THE DIFFUSION OF THE DATA REVOLUTION

One may legitimately ask whether the data revolution has contributed to a better understanding of the African past, and how historical factors still shape the destinies of its people. While it is perhaps too early to judge the contribution of the data revolution, the new approach has certainly generated interest far beyond the narrowly-defined disciplinary borders of economic history. Econometricians, geographers, evolutionary biologists, linguists, demographers, sociologists and computer scientists are beginning to exploit the rich quantitative history of Africa in new ways using their own methodological approaches. This should be welcomed by economic historians, not only because it generates an audience for their previous work but because it provides new ways to test earlier conjectures and hypotheses.

The risk, however, is that economic historians are left behind by the data revolution, anchored to their own approaches which emerged in an era without readily available quantitative data. This

⁵⁹ Fourie, J. and Green, E. 2015. The Missing People: Accounting for the productivity of indigenous populations in Cape Colonial History. *Journal of African History*. Vol 56(2): 195-215; Baten, J. and Fourie, J. 2015. Numeracy of Africans, Asians, and Europeans during the Early Modern Period: New Evidence from Cape Colony Court Registers. *Economic History Review*. 68(2): 632-656.

is especially pertinent for scholars with limited access to the statistical tools the data revolution requires. Often these scholars are based at universities in Africa, creating inequality within the discipline between those able to adopt the new methods due to their frequent interdisciplinary interaction and those geographically isolated from such interactions.⁶⁰

A recent debate within *Economic History of Developing Regions* summarizes this tension best. In a summary of African economic history research, Erik Green and Pius Nyambara writes⁶¹:

[E]conomic history research at African universities is not only strong, but remained vibrant even when African economic history was on the decline at universities elsewhere. The lack of visible output in major economic history journals is thus not a sign of weakness. Instead it is an effect of the increased methodological specialization of economic history in the Western world. There is a danger that this specialization may led to regional isolation and we thus urge economic historians in the Western world to further engage in the work by African scholars.

In a response, Gareth Austin suggests that limitations within African universities are likely the reason for slow adoption of quantitative techniques by African scholars⁶²:

Resource constraints are a major hurdle, but research and teaching in the field has been inhibited by institutional constraints and intellectual priorities as well. The overwhelming priority that economics departments in Africa rightly give to the study of current problems does not seem to be combined with an awareness of the uses of history in fulfilling this mission. Meanwhile, the institutionalization of the humanities/social sciences divide in many universities has made it less likely that history graduates will be equipped to combine qualitative and quantitative techniques, let alone focus on the latter.

These are serious limitations to equipping the next generation of African economic historians with the quantitative tools to partake in the data revolution. It is not entirely clear which incentives, and by whom, will minimize the likelihood of continued isolation by African scholars. One approach suggested by both Green and Nyambara and Austin is to encourage collaborative work through joint funding applications and research programmes. Here, though, the danger of an unequal partnership, with the non-African partner in the dominant position driving the research agenda, remains real.

⁶⁰ A case in point is the contributors to a special issue on the Economics of Apartheid published in *Economic History of Developing Regions* in 2014. See Mariotti, Martine, and Johan Fourie. "The economics of apartheid: An introduction." *Economic History of Developing Regions* 29.2 (2014): 113-125.

⁶¹ Erik Green & Pius Nyambara, 2015. "The Internationalization of Economic History: Perspectives from the African Frontier", *Economic History of Developing Regions*, 30, 1, 68-78.

⁶² Gareth Austin, 2015. "African Economic History in Africa", *Economic History of Developing Regions*, 30, 1, 79-94.

Another approach, favoured by me, is for European and US universities to actively recruit and subsidize PhD students from Africa. After completing their studies, these students, equipped with the tools necessitated by the data revolution, can return to their countries-of-origin, and begin the slow but sustainable process of training and teaching colleagues and students. That is the most likely way, in my opinion, to bridge the data revolution divide that threatens to stall the renaissance in African Economic History on the African continent.

CONCLUSION

African economic history has already gained much from the data revolution of the last two decades. The tools of data transcription, augmentation and collaboration associated have revealed much we did not know about the African past, and its impact on the present. The continuing efforts to transcribe and digitize large numbers of colonial and post-colonial records suggest that we are likely to learn much more about Africa's economic past over the coming decade.

But to avoid the deep division between those equipped in the tools of the data revolution and those outside its purview, more African students must be assisted to study under the guidance of those at the technical and methodological frontier. This may require a commitment by non-African scholars at European or US universities to invest time and resources in establishing long-run partnerships with African universities.

Equipping African scholars with the tools to partake in the data revolution is not only necessary to redress the inequalities of the past but an essential building block towards a thriving academic discourse, both within Africa and between Africa and the rest of the world. 'The past is never dead, it is not even past', William Faulkner once said. This is nowhere more appropriate than on a continent where the effects of colonialism, as the new African economic history show, still permeate. An inclusive data revolution which builds on the lessons of the past promises to make the Dark Continent a lighter shade of grey.