

The Data Revolution in African Economic History*

Johan Fourie[†]

October 22, 2015

Abstract

Large datasets that capture the complexities of the cosmos or the behavior of millions of individuals have, during the past decade, yielded extraordinary scientific results. Improvements in computing power, increased connectivity, and more advanced analytical techniques have heralded the era of Big Data in fields as diverse as astronomy, economics, biology, and management.¹ Yet in the history profession this data revolution has gained only limited traction. One obvious reason is that we do not have Big Data from the distant past. As the former CEO of Google, Eric Schmidt, put it in 2010, from the dawn of civilization through 2003, five exabytes of information were created. Only seven years later, that much information was being created every two days.²

Of course, to paraphrase the well-worn aphorism, not all data are useful, and not everything that is useful is captured in data. Indeed, it is often the historian's main duty to distinguish meaningful data from meaningless. Too much information can be as problematic as too little.

*I am indebted to Gareth Austin, Angus Dalrymple-Smith, Pim de Zwart, Felix Meier zu Selhausen, Ewout Frankema, Leigh Gardner, Erik Green, Kris Inwood, Martine Mariotti, Alex Moradi, Nonso Obikili, Auke Rijpma, Jan Luiten van Zanden, and the members of the LEAP team at Stellenbosch for valuable input and fruitful discussion during the writing process. I acknowledge 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 thank my sabbatical study host, Jan Luiten van Zanden, for generously hosting me in Utrecht.

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

¹Eric D. Feigelson, and G. Jogesh Babu, "Big Data in Astronomy," *Significance*, 9 (2012), 22–25; Liran Einav and Jonathan Levin, "The Data Revolution and Economic Analysis," NBER Working Paper No. 19035 (2013); Vivien Marx, "Biology: The Big Challenges of Big Data," *Nature*, 498 (2013), 255–260; A.D. Howe, M. Costanzo, P. Fey, T. Gojobori, L. Hannick, W. Hide, and S.Y. Rhee, "Big Data: The Future of Biocuration," *Nature*, 455 (2008), 47–50; Andrew McAfee and Erik Brynjolfsson, "Big Data: The Management Revolution," *Harvard Business Review*, 90 (2012), 60–6.

²The figures have been disputed (<http://readwrite.com/2011/02/07/are-we-really-creating-as-much>) but the point is made clearly enough – the quantity of data is increasing exponentially.

My purpose here, however, is to show that the data revolution can be consequential for the field of history. The surge in computing power and access to data processing software and online resources have enabled historians over the past two decades to capture historical statistics on a much larger scale than before.³ I argue that the data revolution is especially valuable when applied to regions where written records are fairly scarce, such as sub-Saharan Africa. In this paper I document how a new generation of economists, geographers, and historians are rewriting African history using archival records (colonial sources written for purposes unrelated to current research questions) along with geographical, climatic, and demographic projections into the distant past, and finding some surprising answers.

1 A NEW AFRICAN ECONOMIC HISTORY

African history has always had to overcome the problem of limited written records. Henry Morton Stanley's "Dark Continent" was, at least until the end of the colonial period, dark not only in its geography but also in its history. It was only in the 1960s, when most African countries had gained independence, that historians in Africa and elsewhere began to borrow methods 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 new questions about African societies' past fortunes, the effects 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 Marxist and liberal economics. Evidence in support of hypotheses came most often from colonial era records, notably reports, letters, and other such qualitative sources, although the abundant statistical records were not completely neglected. Many pioneers of African data collection, including Marion Johnson, Robert Szereszewski, Anthony Hopkins, Patrick Manning, David Eltis, and Ralph Austen, began to assemble datasets during the 1960s and 1970s.⁴

But sadly, by the 1980s three forces had combined to push African economic history to the intellectual periphery.⁵ First, Africa's economies were in decline.

³Patrick Manning, *Big Data in History* (Basingstoke, 2013).

⁴Marion Johnson, "The Cowrie Currencies of West Africa Part I," *Journal of African History*, 11 (1970), 17–49; Marion Johnson, "Cloth as Money: The Cloth Strip Currencies of Africa," *Textile History*, 11 (1980), 193–202; Robert Szereszewski, "The Process of Growth in Ghana, 1891–1911*," *Journal of Development Studies*, 1 (1965), 123–141; Anthony G. Hopkins, "Economic Imperialism in West Africa: Lagos, 1880–921," *Economic History Review*, 21 (1968), 580–606; Patrick Manning, "Slaves, Palm Oil, and Political Power on the West African Coast," *African Historical Studies*, (1969), 279–288; David Eltis, "The Export of Slaves from Africa, 1821–1843," *Journal of Economic History*, 37 (1977), 409–433; Ralph A. Austen, *Slavery among Coastal Middlemen: The Duala of Cameroon* (Wisconsin, 1977).

⁵Green and Nyambara (2015) argue that African economic history declined only in Western institutions. See Erik Green and Pius Nyambara, "The Internationalization of Economic

The oil shocks of the 1970s and the consequent debt crises and IMF structural adjustment programs had diminished the resources available to African universities and deflated interest in explaining past fortunes.⁶ Second, as postmodernism became fashionable, historians became more interested in cultural and social history than in economic development. Third, following the cliometric revolution of the 1970s, economics became more mathematical and statistical, which heightened the methodological barrier between economics and history. This combination meant that African economic history, both in and outside Africa, was in serious decline.

It was not until the 2000s that African economic history awoke from its slumber. Led by prominent economists, the renaissance of African economic history coincided with the rise of Africa's economic prospects.⁷ 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.

Acemoglu, Johnson and Robinson are widely credited with rekindling interest in African economic history with their provocative claim that the institutional legacies of European settlers in Africa explain much of the disparity in levels of development today.⁸ Their main contribution was, however, not their use of a large, new dataset but rather the causal inference they made from it by using a novel instrumental econometric technique. Their claim about the persistence of institutions attracted the interest of a younger generation of economists, 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. Nunn used and expanded on an extensive dataset of African slaves exported to the slave markets in the Americas, a dataset that was obtained from the "original" Big Data project in African history begun by Philip Curtin in the 1960s⁹ and, following much debate and controversy,¹⁰ culminated in Eltis and David Richardson's Transatlantic Slave Trade Data-

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 Gareth Austin, "African Economic History in Africa," *Economic History of Developing Regions*, ahead-of-print (2015), 1–16.

⁶Morten Jerven, "African Growth Recurring: An Economic History Perspective on African Growth Episodes, 1690–2010", *Economic History of Developing Regions*, 25 (2010), 127–154.

⁷Gareth Austin and Stephen Broadberry, "Introduction: The Renaissance of African Economic History," *Economic History Review*, 67 (2014), 893–906; Johan Fourie and Leigh Gardner, "The Internationalization of Economic History: A Puzzle," *Economic History of Developing Regions*, 29 (2014), 1–14.

⁸Daron Acemoglu, Simon Johnson and James A. Robinson, "The Colonial Origins Of Comparative Development: An Empirical Investigation," *American Economic Review*, 91 (2001), 1369–1401.

⁹Philip D. Curtin, *The Atlantic Slave Trade: A Census* (Wisconsin, 1972).

¹⁰Joseph E. Inikori, "Measuring the Atlantic Slave Trade: An Assessment of Curtin and Anstey," *Journal of African History*, 17 (1976), 197–223; Paul E. Lovejoy, "The Impact of the Atlantic Slave Trade on Africa: A Review of the Literature," *Journal of African History*, 30 (1989), 365–394.

base.¹¹ Although the various versions of the slave database have over the past few decades contributed to a rich scholarship of the history of slavery in Africa,¹² it was Nunn’s demonstration of a causal link between the number of slaves exported and the level of African GDP in 2000 that suggested a new direction for African economic history.¹³ The paper later came under severe criticism—for example, that its premises and reasoning rested on fundamental factual errors and a highly selective use of secondary sources¹⁴—but most would acknowledge that it rekindled economists’ interest in African economic history.

After being recruited to Harvard, Nunn continued to ask questions about the African past: with Leonard Wantchekon, he identified levels of trust as one mechanism by which the slave trade still affects African societies; with Diego Puga, he showed how the slave trade pushed Africans into rugged areas, which still affects their ability to trade.¹⁵ These contributions laid the foundation for what would become a new trend in African economic history: identifying the causal mechanisms that explain the persistence into the present of the effects of past events.

Few things are, of course, more persistent than the effects of geography, which is why environmental information is an especially useful tool for linking past events to the present. Nunn and Puga use terrain ruggedness as an explanatory variable. Marcella Alsan uses measures of temperature and humidity to assess the effect of the tsetse fly on development in Africa today.¹⁶ Sambit Bhattacharyya shows that the prevalence of malaria is statistically the most important factor in explaining African poverty.¹⁷ James Fenske and Namrata Kala use an annual panel of African temperatures and slave exports to show how past environmental shocks affected the slave trade.¹⁸ Kostadis Papaioannou uses court cases and records of prisoners and homicides to show how deviations from the long-term rainfall pattern increased conflict.¹⁹ Papaioannou

¹¹David Eltis and David Richardson (eds.), *Extending the Frontiers: Essays on the New Transatlantic Slave Trade Database* (New Haven, 2008).

¹²Patrick Manning, *Slavery and African Life: Occidental, Oriental, and African Slave Trades, Vol. 67*. (Cambridge, 1990); Paul E. Lovejoy, *The African Diaspora: Revisionist Interpretations of Ethnicity, Culture and Religion under Slavery* (Boston, 1997).

¹³Nathan Nunn, 2008. “The Long-Term Effects of Africa’s Slave Trades,” *Quarterly Journal of Economics*, 123, 139–76.

¹⁴Ewout Frankema and Marlous Van Waijenburg, “Structural Impediments to African Growth? New Evidence from Real Wages in British Africa, 1880–1965,” *Journal of Economic History*, 72 (2012), 895–926. The criticism is more explicit in the Working Paper version, available at:

http://eh.net/eha/wp-content/uploads/2013/11/FrankemaWaijenburg_0.pdf

¹⁵Nathan Nunn and Diego Puga, “Ruggedness: The blessing of bad geography in Africa,” *Review of Economics and Statistics* 94 (2012), 20–36.

¹⁶Marcella Alsan, “The Effect of the TseTse Fly on African Development,” *American Economic Review*, 105 (2015), 382–410.

¹⁷Sambit Bhattacharyya, “Root Causes of African Underdevelopment,” *Journal of African Economies*, 18 (2009), 745–780.

¹⁸James Fenske and Namrata Kala, “Climate and the Slave Trade,” *Journal of Development Economics*, 112 (2015), 19–32.

¹⁹K.J. Papaioannou, “Climate Shocks and Conflict: Evidence from Colonial Nigeria,” African Economic History Network Working Paper Series, No. 17 (2014).

and Michiel de Haas extend this analysis to include other parts of Africa too, finding similar results.²⁰ The use of climatic and other environmental information geocoded at the micro level enabled these scholars to test the persistence of the effects of historical events with a high level of statistical accuracy.

The data revolution has not only offered new variables to measure past events, but variables created from new data sources have been especially useful as contemporary outcome variables. In the absence of data on regional African economies, for example, Stelios Michalopoulos and Elias Papaioannou use light density at night obtained from satellite imaging to test the effect of precolonial ethnic institutions²¹ and the effect of the borders drawn during the Scramble for Africa²² on current economic performance. Although light density is now a popular outcome variable,²³ there has been, as with Nunn’s paper, strong criticism by economic historians of Michalopoulos and Papaioannou’s data sources and estimation techniques.²⁴

Contemporary survey data, now with geocoded observations, are also frequently used outcome variables. Nunn’s paper with Wantchekon is a case in point, using the Afrobarometer survey²⁵ to measure levels of trust. Martin Abel’s work on forced removals during apartheid also relies on the Afrobarometer.²⁶ Michalopoulos, Louis Putterman and David Weil match Demographic and Health Survey (DHS) data to information on the ancestral ethnicity of respondents to show that descendants of agriculturalists are wealthier and better educated than the descendants of pastoralists.²⁷

Not everyone agrees, though, that these broad-brush studies, linking an historical episode with contemporary data, are necessarily useful. The main criticism leveled at such an approach is that it “compresses history”, in short,

²⁰Kostadis J. Papaioannou and Michiel de Haas, “Climate Shocks, Cash Crops and Resilience: Evidence from Colonial Tropical Africa,” paper presented at the World Economic History Congress, Kyoto, 3–8 August 2015.

²¹Stelios Michalopoulos and Elias Papaioannou, “Pre-Colonial Ethnic Institutions and Contemporary African Development,” *Econometrica*, 81 (2013), 113–152; Stelios Michalopoulos and Elias Papaioannou, “Further evidence on the link between pre-colonial political centralization and comparative economic development in Africa,” *Economics Letters*, 126 (2015), 57–62.

²²Stelios Michalopoulos and Elias Papaioannou, “National Institutions and Subnational Development in Africa,” *Quarterly Journal of Economics*, 129 (2014) 151–213.

²³Timothy Besley and Marta Reynal-Querol, “The Legacy of Historical Conflict: Evidence from Africa,” *American Political Science Review*, 108 (2014), 319–336; Nonso Obikili, “An Examination of Subnational Growth in Nigeria: 1999–2012,” *South African Journal of Economics*, 83 (2015) 335–357.

²⁴Denis Cogneau and Yannick Dupraz, “Questionable Inference on the Power of Pre-Colonial Institutions in Africa,” PSE Working Papers, No. 2014–25 (2014).

²⁵The Afrobarometer survey is a series of surveys that measure public attitudes on economic, political, and social matters in several sub-Saharan African countries. More information available at: <http://www.afrobarometer.org/>.

²⁶Martin Abel, “Long-Run Effects of Forced Removal under Apartheid on Social Capital,” paper presented at African Economic History meetings, London School of Economics and Political Science, 25–26 October 2014.

²⁷Stelios Michalopoulos, Louis Putterman, and David N. Weil, “The Influence of Ancestral Lifeways on Individual Economic Outcomes in Sub-Saharan Africa,” paper discussed at NBER Reporter Number 4: Program and Working Group Meetings (2014).

that it oversimplifies the causal links from the past experience to the present.²⁸ Morten Jerven, in a recent book, criticizes the irresponsible manner in which economists have used econometrics to explain African development.²⁹ In reviewing this book, Alex de Waal, the director of the World Peace Foundation, calls for African economic history to be “liberated from the tyranny of econometricians”.³⁰

African economic historians, instead, have proposed an alternative data approach, one that focuses on unearthing and digitizing colonial era archival records. Over the past decade, scholars have ploughed through colonial Blue Books, tax censuses, voters’ rolls, marriage registers, and more, to gain insights into African societies’ population size,³¹ wages,³² incomes,³³ education,³⁴ fiscal

²⁸Gareth Austin, “The ‘Reversal of Fortune’ Thesis and the Compression of History: Perspectives from African and Comparative Economic History,” *Journal of International Development*, 20 (2008), 996–1027.

²⁹Morten Jerven, *Africa: Why Economists Get it Wrong*, (London, 2015); Morten Jerven and Deborah Johnston, “Statistical Tragedy in Africa? Evaluating the Data Base for African Economic Development,” *Journal of Development Studies*, 51 (2015), 111–115.

³⁰Alex de Waal, “Liberating African Economic History from the Tyranny of Econometric”, review of Morten Jerven, *Africa: Why Economists Get it Wrong*, (London, 2015), available at:

<http://africanarguments.org/2015/06/24/liberating-african-economic-history-from-the-tyranny-of-econometrics-by-alex-de-waal/>

³¹Patrick Manning, “Historical Datasets on Africa and the African Atlantic,” *Journal of Comparative Economics*, 40 (2012), 604–607; Ewout Frankema and Morten Jerven, “Writing History Backwards or Sideways: Towards a Consensus on African population, 1850–2010,” *Economic History Review*, 67 (2014), 907–931; Johan Fourie and Erik Green, “The Missing People: Accounting for the Productivity of Indigenous Populations in Cape Colonial History,” *Journal of African History*, 56 (2015), 195–215.

³²Ewout Frankema and Marlous Van Waijenburg, “Structural Impediments to African growth? New Evidence from Real Wages in British Africa, 1880–1965,” *Journal of Economic History*, 72 (2012), 895–926; Klas Rönnbäck, “Living Standards on the Pre-Colonial Gold Coast: A Quantitative Estimate of African Laborers’ Welfare Ratios,” *European Review of Economic History*, 18 (2014), 185–202; Sophia Du Plessis 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 (2012), 125–149.

³³Morten Jerven, “A West African Experiment: Constructing a GDP Series for Colonial Ghana, 1891–1950,” *Economic History Review*, 67 (2014), 964–992; Leandro Prados de la Escosura, “Output Per Head in Pre-Independence Africa: Quantitative Conjectures,” *Economic History of Developing Regions*, 27 (2012), 1–36; Johan Fourie and Jan Luiten Zanden, “GDP in the Dutch Cape Colony: The National Accounts of a Slave-Based Society,” *South African Journal of Economics*, 81 (2013), 467–49.

³⁴Leonard Wantchekon, Marko Klačnja, and Natalija Novta, “Education and Human Capital Externalities: Evidence from Colonial Benin,” *Quarterly Journal of Economics*, 130 (2015), 703–757; Johan Fourie and Dieter von Fintel, “Settler Skills and Colonial Development: The Huguenot Wine-Makers in Eighteenth-Century Dutch South Africa,” *Economic History Review*, 67 (2014), 932–963; Nonso, Obikili, “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; Jörg Baten, and Johan Fourie, “Numeracy of Africans, Asians, and Europeans during the Early Modern Period: New Evidence from Cape Colony Court Registers,” *Economic History Review*, 68 (2015), 632–656.

systems,³⁵ and transport networks.³⁶ Instead of discussing all these contributions, I illustrate the application of the data revolution to African history with reference to one type of archival document: military attestation forms. These forms are useful because they include individual-level observations of heights, an indicator that is widely used as a proxy for measuring living standards in the absence of other evidence.

2 HEIGHTS DATA UNLOCK THE AFRICAN PAST

The use of human heights or stature as a proxy for standards of living is now more than three decades old.³⁷ 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 changes in the living conditions of a society are borne out by a change in the average height of people in that society.³⁸

But while we would expect heights to correlate positively with incomes, economic historians discovered, when analyzing heights in Western Europe and North America, a surprising negative correlation between height and incomes during the early phase of industrialization in England and elsewhere.³⁹ This

³⁵Leigh A. Gardner, *Taxing Colonial Africa: The Political Economy of British Imperialism*, (Oxford, 2012); Johan Fourie, Ada Jansen, and Krige Siebrits, "Public Finances and Private Company Rule: The Dutch Cape Colony (1652–1795)," *New Contree*, 68, (December 2013), 1–22; Ewout Frankema and Marlous van Waijenburg, "Metropolitan Blueprints of Colonial Taxation? Lessons from Fiscal Capacity Building in British and French Africa, c. 1880–1940," *Journal of African History*, 55 (2014), 371–400; Ewout Frankema, "Colonial Taxation and Government Spending in British Africa, 1880–1940: Maximizing Revenue or Minimizing Effort?" *Explorations in Economic History*, 48 (2011), 136–149; Kostadis Jason Papaioannou and Angus Edwin Dalrymple-Smith, "Political Instability and Discontinuity in Nigeria: The Pre-Colonial Past and Public Goods Provision under Colonial and Post-Colonial Political Orders," *Economics of Peace and Security Journal*, 10 (2015), 40–53.

³⁶Remi Jedwab and Alexander Moradi, "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," *Review of Economics and Statistics*, (2015). Forthcoming.

³⁷Robert W. Fogel, Stanley L. Engerman, Roderick Floud, Gerald Friedman, Robert A. Margo, Kenneth Sokoloff, Richard H. Steckel, T. James Trussell, Georgia Villafior and Kenneth W. Wachter, "Secular Changes in American and British Stature and Nutrition," *Journal of Interdisciplinary History*, 14 (1983), 445–481; Robert W. Fogel, 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; Richard H. Steckel, "Height and Per Capita Income," *Historical Methods: A Journal of Quantitative and Interdisciplinary History*, 16 (1983), 1–7.

³⁸Richard H. Steckel, "Stature and the Standard of Living," *Journal of Economic Literature*, (1995), 1903–1940; George Alter, "Height, Frailty, and the Standard of Living: Modelling the Effects of Diet and Disease on Declining Mortality and Increasing Height," *Population Studies*, 58 (2004), 265–279; Angus Deaton, "Height, Health, and Inequality: The Distribution of Adult Heights in India," *American Economic Review*, 98 (2008), 468.

³⁹John Komlos, "Shrinking in a growing economy? The mystery of Physical Stature during the Industrial Revolution," *Journal of Economic History*, 58 (1998), 779–802; John Komlos, "Anomalies in Economic History: Toward a Resolution of the 'Antebellum Puzzle'," *Journal*

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 attempted to explain this puzzle.

Although the heights of nineteenth-century Africans were already the subject of Eltis’s papers in 1982 and 1990,⁴⁰ it was only in the mid-2000s that African attestations—and therefore heights—would receive the attention they deserve. 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 data;⁴¹ they used DHS data to construct historical birth-cohorts to investigate differences in heights.⁴² But the DHS data could be usefully employed in analysis only after the 1950s, at the end of the colonial era. Another source was needed to investigate the colonial and, it was hoped, the pre-colonial, era.

In 2008 Moradi published a paper that did exactly that.⁴³ He used a sample of 1,046 Ghanaian WWI and WWII recruits and 730 Kenyan WWII recruits, in addition to surveys, to measure the effect of colonial policies on African living standards. In follow-up work on the same question, his sample more than doubled.⁴⁴ This groundbreaking work stimulated interest in military records elsewhere in Africa, often with startling results. Gareth Austin, Baten and Bas van Leeuwen found that in nineteenth-century West Africa, Ghanaian and Burkinabe recruits were notably shorter than north-western Europeans but not shorter than southern Europeans during this period.⁴⁵ In twentieth-century West Africa, Denis Cogneau and Léa 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 to 1960, was almost as fast as the pace observed in France and Great Britain during the period 1875 to 1975.⁴⁶ In South Africa, Kris Inwood and Oliver Masakure found that Colored South Africans were on average six

of *Economic History*, 56 (1996), 202–214; John Komlos and Bjorn Alecke, “The Economics of Antebellum Slave Heights Reconsidered,” *Journal of Interdisciplinary History*, 26 (1996), 437–457.

⁴⁰David Eltis, “Nutritional Trends in Africa and the Americas: Heights of Africans, 1819–1839,” *Journal of Interdisciplinary History*, 12 (1982), 453–475; David Eltis, “Welfare Trends among the Yoruba in the Early Nineteenth Century: The Anthropometric Evidence,” *Journal of Economic History*, 50 (1990), 521–540.

⁴¹Others included David E. Sahn and David C. Stifel, “Urban–Rural Inequality in Living Standards in Africa,” *Journal of African Economies*, 12 (2003), 564–597.

⁴²Alexander Moradi and Joerg Baten, “Inequality in Sub-Saharan Africa: New Data and New Insights from Anthropometric Estimates,” *World Development*, 33 (2005), 1233–1265.

⁴³Alexander Moradi, “Confronting Colonial Legacies: Lessons from Human Development in Ghana and Kenya, 1880–2000,” *Journal of International Development*, 20 (2008), 1107–1121.

⁴⁴Alexander Moradi, “Towards an Objective Account of Nutrition and Health in Colonial Kenya: A Study of Stature in African Army Recruits and Civilians, 1880–1980,” *Journal of Economic History*, 69 (2009), 719–754.

⁴⁵Gareth Austin, 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 Faso,” *Economic History Review*, 65 (2012), 1280–1302.

⁴⁶Denis Cogneau 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 (2011), 55–82.

centimeters shorter than white South Africans at the start of the twentieth century, a significantly smaller difference than today's eight centimeters.⁴⁷ And Inwood, Martine Mariotti and I found that white heights increased significantly in the Boer Republics following the discovery of minerals, but found no concomitant increase for whites born in South Africa's Western Cape.⁴⁸ The findings from these papers are now being integrated into a larger literature on the living standards of indigenous populations across the world.⁴⁹

Information on heights is useful not only as a tool for measuring living standards over time but also to test the effects of different colonial policies. Cogneau and Moradi used a much larger sample of recruits from Ghana and Togo—11,940 observations—to test the effect of the partition of German Togoland after World War I on education outcomes.⁵⁰ The part of Togo that fell under French mandate had lower levels of literacy after the partition than the parts that were ruled by the British. Cogneau and Moradi attribute this to the French authorities' hostile attitude to missionary schools. Remi Jedwab and Moradi analyzed the effects of colonial railways on various economic outcomes including heights and found that railways boosted the production of cash crops like cocoa, improving incomes and increasing the heights of those born closest to the railways.⁵¹

The positive effect of income improvement on heights was also found by Mariotti in an investigation on the heights of black mineworkers in South Africa. A 1974 plane crash that killed 73 mineworkers provoked the Malawian government to ban the migration of mineworkers to South Africa, forcing mining authorities in South Africa to recruit workers from the Transkei, a homeland in South Africa. Mariotti showed that the the resulting improvement in household incomes for the newly recruited Transkei mineworkers increased the heights of

⁴⁷Kris Inwood and Oliver Masakure, "Poverty and Physical Well-Being among the Coloured Population in South Africa," *Economic History of Developing Regions*, 28 (2013), 56–82.

⁴⁸Johan Fourie, Kris Inwood and Martine Mariotti, "White Living Standards during South Africa's Mineral Revolution: Evidence from Attestation Forms," mimeo, Stellenbosch University (2015).

⁴⁹Joseph M. Prince and Richard H. Steckel, "Nutritional Success on the Great Plains: Nineteenth-Century Equestrian Nomads," *Journal of Interdisciplinary History*, 33 (2003), 353–384; Howard Bodenhorn, "The Mulatto Advantage: The Biological Consequences of Complexion in Rural Antebellum Virginia," *Journal of Interdisciplinary History*, 33.1 (2002), 21–46; Aravinda Meera Guntupalli and Joerg Baten, "The Development and Inequality of Heights in North, West, and East India 1915–1944," *Explorations in Economic History*, 43 (2006), 578–608; Kris Inwood, 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; Joerg Baten, Ines Pelger, and Linda Twrdek, "The Anthropometric History of Argentina, Brazil and Peru during the 19th and early 20th Century," *Economics & Human Biology*, 7 (2009), 319–333; Jörg Baten, 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 (2013), 103–122; Joerg Baten 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 (2012), S66–S85.

⁵⁰Denis Cogneau and Alexander Moradi, "Borders that Divide: Education and Religion in Ghana and Togo since Colonial Times," *Journal of Economic History*, 74 (2014), 694–729.

⁵¹Remi Jedwab and Alexander Moradi, "The Permanent Effects of Transportation Revolutions in Poor Countries: Evidence from Africa," *Review of Economics and Statistics*, (2015). Forthcoming.

children born during or immediately after 1974, and only in those districts from which mineworkers were recruited.⁵² Mariotti and Taryn Dinkelman have now turned their attention to the effects of the plane crash and the sudden prevention of labor migration on Malawian households.⁵³

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 than on Europeans. Where information is available, it often comes with colonial-era biases whose effects are difficult to exclude. There is, however, no reason to suspect that military heights information would be subject to such systematic biases.

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 (by using truncated regression models, for example) but that does not entirely allay our fear that some other unobservable selection is present. This fear is indeed the subject of several new research papers by Howard Bodenhorn, Timothy Guinnane, and Thomas Mroz.⁵⁴ They claim that the declining heights observed during the Industrial Revolution—the “Early Industrial Growth Puzzle”—reflects not a decline in living standards but rather a change in selection into the military. In short, they argue that the stronger men (i.e. the tallest) would have chosen the military as a career when returns in the private sector were small (at the start of the Industrial Revolution). As the economy grew, however, more would have returned to private sector employment, with no concomitant increase in the military. This would have meant that the weaker men (i.e. the shorter) would have selected a military career as the stronger, and taller, men found employment in the private sector. And Inwood, Mariotti and I, using data on South African recruits in World War I, show that this type of unobservable selection may also arise because of changes in military technology.⁵⁵

Despite these concerns, however, heights provide a useful proxy for individual-level living standards in the absence of other measures. Comparing results derived from military attestation forms and those from survey data, for example, suggest that selection into the military may not have been strong enough to effect earlier results. The studies that have been done on heights are an example of how the data revolution in African history has provided new insights into

⁵²Martine Mariotti, “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 (2015), 485–514.

⁵³Taryn Dinkelman, and Martine Mariotti, “Does Labor Migration Affect Human Capital in the Long Run? Evidence from Malawi,” mimeo, Dartmouth College, Hanover, NH (2014).

⁵⁴Howard Bodenhorn, Timothy W. Guinnane, and Thomas A. Mroz, “Sample-selection biases and the ‘industrialization puzzle,’” NBER Working Paper No. 21249 (2015); Howard Bodenhorn, Timothy Guinnane, and Thomas Mroz, “Caveat Lector: Sample Selection in Historical Heights and the Interpretation of Early Industrializing Economies” NBER Working Paper No. 19955 (2014).

⁵⁵Johan Fourie, Kris Inwood, and Martine Mariotti, “Can Historical Changes in Military Technology Explain the Industrial Growth Puzzle?” Mimeo, London School of Economics (2014).

how colonial-era policies affected the welfare of subjects who remained outside the remit of colonial-era recordkeeping.

3 MORE THAN JUST LARGE DATASETS

The case studies described above show how analyzing large quantities of numeric data, particularly in regions like sub-Saharan Africa with few written records, can give new insights into the past. But the use of large archival datasets, until now never examined as a whole, is only one aspect of the data revolution. New tools for data transcription, augmentation, and collaboration allow for faster and less expensive digitization, more rigorous investigations, and better replicability and comparability of results. The data revolution is not just new material, in the form of large datasets, but a new method too: the use of computational power and statistical techniques to expand the scale and scope of research questions.

Let us look at the steps of this new method: transcription, augmentation, and collaboration. Data transcription—the manual copying 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 Englishman who signed up for the South African Constabulary in 1902. Although the form is standardized and the text clearly legible to a skilled historian trained in reading late nineteenth-century handwriting, there is no possibility of automating the transcription of these attestations. Optical character recognition (OCR) software is 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 train a research team to do it. Both options are costly. It often means that historians have to select samples, much as Inwood and Masakure did when investigating South African World War I and World War II soldiers. Although several hundred thousand of these attestations are available in the Defense Force archives, time and funding constraints meant that Inwood and Masakure could transcribe only 10,000 attestations for each of the Wars. Such sampling introduces additional biases that must be considered when analyzing the results. It also does not fully conform to the methodology now popular in the era of Big Data, which advocates the use of all available data.

There is hope, however, that the transcription of archival sources such as attestation forms may soon become less costly with the use of OCR software. This is a rapidly developing technology, most notably in the Google Books project to scan all distinct book titles (estimated in 2010 to be about 130 million). OCR already has wide applicability in history, for example in the digitization of historical editions of newspapers.⁵⁶ But books and newspapers are printed

⁵⁶Maya R. Gupta, Nathaniel P. Jacobson, and Eric K. Garcia, “OCR Binarization and Image Pre-Processing for Searching Historical Documents,” *Pattern Recognition*, 40 (2007),

texts, which simplifies the OCR technology required to digitize and transcribe the content. Yet advances are being made in OCR of natural images which contain written text.⁵⁷ It is only a matter of time before an image such as Figure 1 will be read by an OCR algorithm with a high degree of accuracy. Such a technological breakthrough in computational linguistics, with the aid of historians, could have far-reaching consequences for the field of economic history.

The second step of the new method is data augmentation. This is the addition of information from internal or external sources to add 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.⁵⁸ To return to the attestation forms, it is now possible to match recruits' names and birth dates to, for example, genealogical records. This allows researchers not only to learn more about the recruits themselves but also solve problems that were previously unsolvable, such as calculating intergenerational mobility using height as an indicator of living standards. There are countless such new avenues which data augmentation can explore.

The third step is collaborating across research networks. There are already several examples in Economic History where data sharing has resulted in collaborative projects that aim to compare historical measures. At a macroeconomic level, series such as the Penn World Tables and the Maddison project⁵⁹ 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".⁶⁰ The Global Collaboratory on the History of Labour Relations, based at the International Institute of Social History in Amsterdam, has assembled a team of scholars to collect quantitative benchmarks for a global labor history, including population size and composition, workforce size and composition, occupations, and types of labor, using 1500, 1650, 1800, 1900, and 2000 as benchmark dates (1950 is included for Africa).⁶¹ And Jan Luiten van Zanden at Utrecht University has undertaken the Clio-infra project, which aims to establish "a set of interconnected databases . . . containing worldwide data on social, economic, and institutional indicators for the past five centuries,

389–397.

⁵⁷Sergey Milyaev, Olga Barinova, Tatiana Novikova, Pushmeet Kohli, and Victor Lempitsky, "Image Binarization for End-to-End Text Understanding in Natural Images," paper presented at 12th International Conference on Document Analysis and Recognition (ICDAR), Washington, DC, 25–28 August 2013.

⁵⁸See, for example, James J. Feigenbaum, 2015. "Automated Census Record Linking," mimeo, Harvard University, available at: <http://scholar.harvard.edu/files/jfeigenbaum/files/feigenbaum-censuslink.pdf>

⁵⁹Robert Feenstra, Robert Inklaar, and Marcel Timmer, "The Next Generation of the Penn World Table," NBER Working Paper No. 19255 (2013); Jutta Bolt and Jan Luiten Zanden, "The Maddison Project: Collaborative Research on Historical National Accounts," *Economic History Review*, 67 (2014), 627–651.

⁶⁰Manning, *Big Data*.

⁶¹See their website: <https://collab.iisg.nl/web/labourrelations>

with special attention to the past 200 years”.⁶² 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.

In African history, Ewout Frankema of Wageningen University and his students have taken the lead to digitize and transcribe the colonial Blue Books, to make it possible to do research on colonial education, public finance and population.⁶³ In an award-winning study, Frankema and Marlous van Waijenburg showed that African real wages were significantly above subsistence levels, and rising for most of the colonial period.⁶⁴ In some places, wages were much higher than comparable figures for Asia. They suggest that the idea of sub-Saharan Africa as always having been poor is not supported by the evidence.⁶⁵

In a similar vein, Morten Jerven has reconstructed African GDP measures and found discrepancies between World Bank, IMF and Penn World Table estimates.⁶⁶ In two highly acclaimed books, Jerven emphasizes that Africa’s image as the “hopeless continent”, famously encapsulated by the cover of *The Economist* of 13 May 2000, is to some extent a construction of the 1980s and 1990s.⁶⁷

Jutta Bolt and Leigh Gardner are compiling fiscal revenue and expenditure data at the local level for the late colonial period.⁶⁸ They pair these new records with evidence on pre-colonial states to assess the extent to which the fiscal capacity of local government units reflects pre-colonial state centralization. And in an attempt to understand the evolution of inequality during the colonial era (in the absence of information on income at the individual level), Bolt and Ellen Hillbom are compiling colonial-era records of occupations and labor market structures.⁶⁹

⁶²Jan Luiten van Zanden, Joerg Baten, Marco Mira d’Ercole, Auke Rijpma, Conal Smith, and Marcel Timmer (eds.), *How Was Life? Global Well-Being since 1820* (Paris, 2014).

⁶³Frankema and Jerven, “Writing History Backwards or Sideways; Ewout Frankema, “The Origins of Formal Education in Sub-Saharan Africa: Was British Rule More Benign?” *European Review of Economic History*, 16, (2012), 335–355; Ewout Frankema, “Colonial Taxation and Government Spending in British Africa, 1880–1940: Maximizing Revenue or Minimizing Effort?” *Explorations in Economic History*, 48 (2011) 136–149, January.

⁶⁴Ewout Frankema and Marlous van Waijenburg shared the 2013 Arthur H. Cole prize, awarded for the best paper published in the *Journal of Economic History*.

⁶⁵Ewout Frankema and Marlous van Waijenburg, “Structural Impediments to African Growth? New Evidence from Real Wages in British Africa, 1880–1965,” *Journal of Economic History*, 72 (2012) 895–926.

⁶⁶Morten Jerven, “African Growth Recurring: An Economic History Perspective on African Growth Episodes, 1690–2010,” *Economic History of Developing Regions*, 25 (2010), 127–154; Morten Jerven, “For Richer, For Poorer: GDP Revisions and Africa’s Statistical Tragedy,” *African Affairs*, 112 (2012), 138–147.

⁶⁷Morten Jerven, *Poor Numbers: How we are Misled by African Development Statistics and What To Do About It*, (Ithaca, NY, and London, 2013); Morten Jerven, *Africa: Why Economists Get It Wrong* (London, 2015).

⁶⁸Jutta Bolt and Leigh Gardner, 2015. “De-compressing History? Pre-colonial Institutions and Local Government Finance in British Colonial Africa,” paper presented at the Economic History Association meeting, Nashville, 11 September 2015.

⁶⁹Jutta Bolt and Ellen Hillbom, “Potential for Diversification? The Role of the Formal Sector in Bechuanaland Protectorate’s Economy, 1900–65,” *Economic History of Developing Regions*, 30 (2015), 1–30.

These reinterpretations of the African past are the consequence of digitizing and transcribing the vast amounts of data available in colonial archives. And such projects are ongoing. A project led by Felix Meier zu Selhausen at the University of Southern Denmark is digitizing and transcribing vast quantities of demographic records kept in missionary station archives. These records are not without problems—selection into formal Christian mission stations, for example—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 the European settlers and their activities. At Stellenbosch University in South Africa, Erik Green, Dieter von Fintel and I are constructing an annual panel dataset of several thousand settler farmers over more than 140 years of settlement. Once complete, this dataset, augmented with information from probate inventories and genealogical records, will offer a wide-ranging and informative account of eighteenth- and nineteenth-century life in colonial South Africa, and provide economists with data for investigating the economic consequences of events such as the abolition of slavery or the smallpox epidemics. And despite the fact that these statistical records were collected for the European colonial administration, they contain detailed demographic information on the Khoesan, a population group almost completely neglected in other sources.⁷⁰

The scale of our project—several thousand individuals were 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 or her entire lifetime. By adding genealogical records, we can expand the algorithm to link families across generations as well. Such an intergenerational panel would not have been feasible without the advent of large computing power and powerful statistical techniques.

4 THE DIFFUSION OF THE DATA REVOLUTION

We may legitimately ask whether the data revolution has given us a better understanding of the African past and the way historical factors still shape African destinies. It is perhaps too early to pass judgement. What we can confidently state is that the new approach has 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 methods. This should be welcomed by economic historians, not only because it generates an audience for their work but because it provides new ways to test conjectures and hypotheses.

The risk, however, is that some economic historians may be left behind by

⁷⁰Fourie and Green, “The Missing People”; Baten and Fourie, “Numeracy of Africans”.

the data revolution, anchored to their own methods which were created in an era that lacked readily available quantitative data. This 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. This creates inequality in the discipline between those whose frequent interdisciplinary interaction enables them to adopt the new methods and those who are geographically isolated from such interactions.⁷¹ To give a very real example: of the more than 50 authors cited in the section above on the new economic history of Africa, fewer than 10 are from Africa. Only four are black.

A debate in a recent issue of *Economic History of Developing Regions* summarizes this problem best. Green and Pius Nyambara write that economic 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 lead to regional isolation and we thus urge economic historians in the Western world to further engage in the work by African scholars.⁷²

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

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.⁷³

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

Another approach, which I favor, is for European and US universities to

⁷¹A case in point is the list of contributors to a special issue on the economics of apartheid published in *Economic History of Developing Regions* in 2014. See Martine Mariotti and Johan Fourie, “The Economics of Apartheid: An Introduction,” *Economic History of Developing Regions*, 29 (2014), 113–125; Gardner and Fourie also discuss this problem for other developing regions: Fourie and Gardner, “The Internationalization of Economic History”.

⁷²Green and Nyambara, “The Internationalization of Economic History”, p. 68.

⁷³Austin, “African Economic History in Africa”, p. 91.

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 colleagues and students in the new methods. I believe this is the most promising way to bridge the data revolution divide that threatens the renaissance of African Economic History on the African continent.

5 CONCLUSION

African economic history has already gained much from the data revolution of the past two decades. The tools of data transcription, augmentation, and collaboration have revealed much we did not know about the African past and its echoes in the present. The continuing projects to transcribe and digitize large numbers of colonial and post-colonial records are likely to teach us much more about Africa's economic past over the coming decade.

But to bridge the divide between those equipped with 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-term partnerships with African universities and their faculty.

Equipping African scholars with the tools to take part in the data revolution is not only necessary to redress the inequalities of the past but also an essential step towards building a thriving academic discourse, both in 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 here in Africa, where the effects of colonialism, as the new African economic history shows, still permeate. An inclusive data revolution which builds on the lessons of the past promises a brighter view of the Dark Continent.

