What Makes Maddison Right

Chinese Historic Economic Data

Jan Luiten van Zanden & Debin Ma

Key Points

- The 'Great Divergence debate' in economic history relates to the question of when China fell behind the levels of well-being in Western Europe.
- A recent paper published in this journal argues that existing historical data cannot answer this question and criticizes estimates of Angus Maddison of GDP per capita based on limited evidence.
- The authors believe, in contrast, that critiques, assessments and summaries on the state of the Great Divergence debate even if flawed are in the original spirit of the Maddison research.
- Maddison's work is less about right or wrong than about trying to achieve better or best estimates by overcoming the current constraint on data and methodologies over time.

Kent Deng and Patrick O'Brien recently published a review of the 'Great Divergence debate' in this journal, with the challenging title 'Why Maddison was wrong'. They presented their 'brief review' of this debate, which began in 2000 with the publication of Kenneth Pomeranz' (2000) seminal book *The Great Divergence: Europe, China and the Making of the Modern World*. Their take on this debate was that Pomeranz maintained 'that standards of living afforded by the economy for populations contained within the political boundaries of the Ming–Qing Empire of China (1368–1911) did not fall behind the levels of well-being afforded to the populations of the national economies of Western Europe until late in the eighteenth century' (Deng and O'Brien 2017, pp. 20–21). In other words, the divergence between Western Europe and China only began in the final decades of the eighteenth century—after 1780. It is a rather restricted interpretation of this debate—which, others would argue, is also about the causes of the divergence that did happen, and about the relevant unit of comparison, but those parts of the debate are not covered by the review (and therefore shall not be considered here either). They then discuss the debate about this hypothesis in highly subjective terms: the 'novel theses' by Pomeranz were subject to 'bombardment' by 'the heavy artillery of statistical-*cum*-econometric "tests"'. It is this discussion that we take issue with.

During the past two decades, the Pomeranz book and related publications by his colleagues from the California School have led to a flowering of new research into the economic history of China, with the aim of testing these ideas. In at least four different ways the hypothesis has been subjected to more detailed quantitative research, and in all cases the conclusion has been that the Pomeranz hypothesis cannot be proved right and is probably incorrect. These four tests are: (1) the development of GDP and GDP per capita in China (and Europe), (2) the development of real wages (of unskilled labourers), (3) the evolution of agricultural productivity and the consumption level of the population, and (4) the rate of urbanisation. All four literatures conclude that the Pomeranz hypothesis is probably too optimistic about the standard of living of the Chinese population compared with that of the most advanced parts of Western Europe.

One of the paradoxes of the recent contribution by Deng and O'Brien is that in another paper they presented a detailed overview and analysis of the third test, concluding that 'our clarified and recalibrated estimates suggest that, from the early seventeenth century onwards, the state, institutions and foundational culture of the Chinese empire were failing to cope nearly as well with 'the pressures of numbers and environmental degradation' as states and economies of Western Europe (an interpretation which comes close to that of Peer Vries (2013), who is quite critical of Pomeranz). Or, as the authors write themselves,

Commented [J1]: Where does this quote end? If at the parenthesis, presumably they are quoting 'the pressure of numbers and environmental degradation', so this should be in double inverted commas?

> [o]ur survey, critique, and recalibration of the data produced for this particular and altogether more promising line of historical enquiry rejects inference-derived statistical evidence that continues to be widely cited to support the revisionist claims of the California School. Recalibrated and tabulated here into kilocalories for

> purposes of reciprocal comparisons, that evidence suggests that although the population of early modern Jiangnan enjoyed "nutritional security," its standard of living was declining and falling to levels below standards afflicting England's laboring poor.

In sum, they are highly critical of the core hypotheses of the California School.

This is all valid economic historical research. But in the 'Why Maddison is Wrong' paper they develop an entirely different argument. The core argument is that 'the data for China accessible in secondary sources do not provide historical runs of estimates either for GDP or for total population, let alone for any purchasing-power-parity rates of exchange estimates' (p. 23). With this and similar arguments, they seek to criticise the conclusions from other research, focused on measuring GDP, real wages or the level of urbanisation, that arrive at similar conclusions to their research on agricultural productivity and consumption levels. In particular Angus Maddison is the bad guy in this saga, as he produced estimates of GDP per capita which were based on limited evidence. Moreover, his approach—linking time series to a benchmark estimate for 1990—has been criticised by 'a platoon of distinguished economists' who 'found his methods and estimates to be conceptually and statistically unacceptable as historical evidence' (Deng and O'Brien 2017, p. 21).

We firmly believe that critiques, assessments and summaries of the state of the Great Divergence debate and on the larger research agenda on assessing long-term living standards are very welcome and are also in the original spirit of the Maddison research. However, we feel that their characterisation of the Maddison works and others research related to the issue are incomplete and inaccurate, and in some cases misleading. Firstly, the Maddison estimates are less about right or wrong than about trying to achieve better or best estimates by overcoming the current constraint on data and methodologies over time. They are dynamic processes that require constant updating based on new data sources and methodologies. Secondly, the issues raised by Deng and O'Brien on the Maddison estimates by critiques have long been known to the scholars in the field of international comparison, including Maddison himself, and many of them are not peculiar to Chinese economic history. Indeed, some of their own summaries or assessments are not presented as meaningful or superior alternatives in terms of either methodology or empirics, or in many ways not markedly different from the existing works by Maddison and others.[1] Hence, we are puzzled by the highly critical and sometimes dismissive tones describing these works.

Their paper starts out with their assessment of the use of nominal and real wages, which again draws from their critique of others' work, mainly that of Allen et al. (Deng and O'Brien, 2016).[2] Allen et al. (2011) conclude that, already in the 18th century, real wages in Beijing, Canton and the Yangtze delta were significantly lower than in North-western Europe based on a large variety of sources from both the government (related for example to construction projects) and from companies (both foreign, like the VOC in Canton, and domestic firms). Allen et al. (2011) represents the first systematic attempt at constructing long-term wages for China and other Asian countries. So, we are happy to see this paper being given proper scrutiny by Deng and O'Brien (2016). As two co-authors of this paper, we find their meticulous critique very helpful to the debate despite numerous (perhaps unavoidable) misunderstandings. The team that worked on the Allen et al. 2011 paper included historians with a Chinese background and specialists on Chinese historical sources. The Allen et al. 2011 team has taken great care to lay out their data sources and estimation procedures, which—we are happy to see—have allowed other scholars to trace, reproduce and critique the result. In fact, the major issues that Deng and O'Brien discuss had already been raised in the Allen et al. (2011) paper; we alerted the readers to potential margins of error and, in many cases, tested these with robustness checks and alternative scenarios.

Commented [J2]: In this article there are phrases that mean 'the 2011 paper by ...' (where I do not put the date in brackets) and straightforward citations (where I do).

For example, the main argument that is now used against the results of this research is that labour markets in China were more marginal than in Western Europe. This may be correct (although there is perhaps a tension

[1] For example, their use of calories in their recalibration is a useful cross-check (Deng and O'Brian, 2015). But the method of using calories is already part of constructing the consumption basket in the Allen et al. 2011 paper, which they critiqued. The Allen et al. (2011) consumption basket used a combination of calories and proteins with an actual consumption basket and commodities precisely because calories alone are an extremely limited indicator for measuring living standards.

[2] Strangely, the Allen et al. 2011 paper, upon which much of their critique is based, is either missing from their citation in this journal, or misquoted as Allen et al. 2005: see p. 22.

between this line of defence and the original argument of *The Great Divergence*, that institutional structures in China were very similar to those of Europe), but the Allen et al. (2011) paper (p. 29) compared the daily wage rates (of short-term labourers) with the bottom of the income pyramid and argues that the purchasing power of the wages of unskilled labourers may be a good guide to the standard of living of a much larger group of those on lower incomes.

On similar grounds the new research into the development of the urbanisation ratio shows a strong decline of this measure of economic transformation from the early Qing (c. 12%) to the late 18th century (about half that level). So we have the results of four tests of the Pomeranz hypothesis, based on different sets of sources, all with their own problems and limitations, but also all pointing in the same direction, that is, that this hypothesis (as summarised by Deng and O'Brien) is not correct. In fact, in their own original, data-based contribution, they arrived at the same conclusion.

Back to Maddison, and his contribution, where we have problems with the Deng and O'Brien characterisation. Firstly, in the 1990s, way before the publication of the Pomeranz book, Maddison began doing research on Chinese economic development, not because he wanted to address the 'Great Divergence' thesis, but because he was working on his grand synthesis of the growth of the world economy during the past two millennia, and China was obviously a very large part of that story. At the time (the 1990s), however, very few people were doing quantitative economic historical research on China, which would result in the historical GDP series that he needed for his synthesis. Not shying away from the challenge, he decided to do the research himself.[3] He travelled to China, spoke to the most eminent economic historians and tried to collect the data for his first set of estimates, making use of state of the art literature. Maddison's OECD book on China contains a thorough re-evaluation of the entire 50-year Communist era GDP and the components of their production accounts, including separate purchasing-

[3] Deng and O'Brien suggest that he published them in 2007, as a response to the Pomeranz book of 2000; he actually published his results in 1998, before the Great Divergence debate started (Maddison, 1998).

power-parity estimates for agriculture and industry. For the pre-Communist era, Maddison examined a variety of works going way beyond Liu and Yeh (1965). Hence it is highly unfair to characterise all this work as 'Madison utilized just two estimates for China's GDP in current prices'.

For the earlier period, Maddison understandably relied on much scattered and dispersed evidence, which includes some population and agricultural output series, all of which have their problems. In many cases, Maddison relied on his intellectual intuition, based on a huge amount of reading on Chinese and global history sources, to come up with estimates for benchmark GDP series in 1990 prices. On numerous occasions, he explicitly termed them 'guess-estimates'. These estimates were clearly not intended to be the final word in this regard. Maddison's research strategy was aimed at challenging people, and inviting them, if they knew better sources and methods, to make improved estimates of the development of GDP. He was very open in welcoming the result of all scholarly work in the field, with the aim of constantly improving the quality of the estimates of GDP, population and GDP per capita of his dataset. That this was a highly dynamic process, and that it was very successful in the long run, is clear from the dataset, which started, in the 1960s, with observations for a dozen or so industrial countries going back (sometimes) to 1870, and ended, in his synthesis of 14,000 data points, with estimates for more than a hundred countries covering, in a few cases, the entire two millennia. We agree that by publishing his 'guess-estimates' in tabulated form, Maddison made his data vulnerable to the risk of abuse by undiscriminating users who have not followed historical works carefully. But for people who had done so, the critiques by Deng and O'Brien of these estimates may be useful but rather superfluous.

The last decade has seen a new wave of works by various authors (Xu et al. 2016; Broadberry et al. 2014), who have produced papers with new estimates for the development of GDP during the Qing (and sometimes going back to Ming and Sung times). What all this new work has in common is that it is broadly consistent with the estimates Maddison made in his grand synthesis, and do not confirm the optimistic view of the California School. In fact, most recent estimates are more pessimistic and show a decline of GDP per capita during the period 1600–1800, where Maddison assumed stability.

It seems in the end there are a lot of good things to be said about Maddison's initial intuition. Like Maddison's own works, various issues arise in relation to all these works. The essence of the Maddison estimates was therefore that it was not supposed to be a static end-product of research, but that it had to be updated frequently in order to incorporate new research and improved estimates.

The second issue which Deng and O'Brien critiqued is the use of a 1990 price benchmark. We should start out by saying that the issue of projecting backwards from 1990 is one of which Maddison and his team have been keenly aware for a long time, and is part of a larger indexing problem for historical national accounts for all countries. The entire ICOP (International Comparison of Output and Productivity at www.rug.nl/ggdc/projects/icop?lang=en), which was originally founded under Maddison's leadership at the University of Groningen, engages in the systematic and large-scale construction of current-price purchasing power parities (PPP), both as cross-checks and as a basis for new updates. Maddison himself has been highly encouraging of others conducting new current-price PPP estimates. Various research along this line actually reveals mixed results from comparing current-price with backward-extrapolated estimates. For example, Fukao, Ma and Yuan (2007) confirmed that the 1930s current-price PPP estimates for China and Japan were not that far from the 1990 backward extrapolation used by Maddison. Li and Van Zanden (2012), for example, compared the two most advanced parts of Western Europe with China in the 1820s, and came to the conclusion that even the most urbanised parts of China clearly had a much lower level of GDP per capita than the Netherlands, one of the two comparators. So, in the end, Maddison retained the use of a single 1990 benchmark, extrapolated backwards, after weighing the costs and benefits of using a single benchmark *versus* multiple benchmarks or current price estimates. Clearly, as we move further and further away from 1990, it becomes increasingly necessary to make updates. But it is a matter of choice rather than right or wrong, as Deng and O'Brian have led us to believe.

When Maddison passed away in 2010 (it is a bit odd that Deng and O'Brien criticise him for not incorporating the new ICP results of 2011, which were published the year after his death), it was decided that a group of scholars—close colleagues and friends—would set up 'the Maddison project' to continue his work (he had been very much in favour of this idea) (Bolt and van Zanden, 2014). This has resulted in a number of updates, the first one published in 2014, the second now about to be launched, which have 'in his spirit' dealt with some of the problems mentioned by Deng and O'Brien (and probably by other members of the platoon as well). In the latest update, the 1990 benchmark which has been heavily criticised, in combination with the method used to extrapolate time series linked to this one benchmark back in time, has been replaced by a more sophisticated approach combining benchmarks with time series, as developed by Penn World Tables (Feenstra et al. 2015). Both the 2014 and the 2017 updates include new estimates of GDP per capita for China, made by a team of authors (Xu, Shi, van Leeuwen, Ni, Zhang, and Ma, 2016; also Ma and De Jong, 2017), which tend to arrive at results which are similar to the old set of 'guess-estimates' (it is again rather odd that Deng and O'Brien did not mention these works).

In the end, we want to argue that the research strategy which lay behind Maddison's figures is quite sound. As mentioned already, the years since 2000 have seen the emergence of the modern economic historical research that was carried out for other continents (Europe and North America in particular) after the rise of 'new economic history' in the 1960s and 1970s. Both Chinese and non-Chinese scholars have profited from the opening up of Chinese archives, the growth of universities in China, increased international academic exchange and the growing interest in the history of China, all resulting directly or indirectly from the spectacular economic performance of the country since about 1980. We are sure that this is just the beginning of a boom in the quantitative study of Chinese history.

The Chinese state with its large bureaucracy has produced vast quantities of historical data—some of a quantitative nature—which scholars are now beginning to exploit. Of course, as Deng and O'Brien stress, all sources have their problems, and nothing can be taken at face value; but it is the core business of economic historians to deal with these problems, which are not unique to China (although Deng and O'Brien want us to believe this is perhaps the case). Scholars working on European economic history have struggled with similar problems, and the many results of the quantification of European economic development are also still contested (as, for example, recent exchanges about economic growth in England demonstrate (Clark, 2017)). European scholars, however, have been doing this kind of research for a much longer time, and have therefore built up more experience and credibility. Such a track record requires time and hard work, and it is perhaps natural that established scholars from older generations are critical of the new kid on the block—the 'new economic history' also met such resistance in the 1960s and 1970s (Drukker, 2006).

Angus Maddison was never afraid of being considered wrong, especially when it comes to estimates for China, partly because he wanted to encourage or 'provoke' other scholars with more specialist knowledge to come up with better estimates down the road. So, in that sense, he was right, perhaps not so much because he always estimated Chinese economic performance correctly (although his intuition seemed quite good in this regard as well), but because he developed the right research strategy to make consistent progress in the extension and refinement of his historical dataset of GDP and GDP per capita. This is his true achievement.

References

Allen, R. C., J-P. Bassino, D. Ma, C. Moll-Murata, and J. L. van Zanden (2011).Wages, prices, and living standards in China, 1738–1925: in comparison with Europe, Japan, and India. *Economic History Review* **64**, pp. 8–38.

Bolt, J. and J. L. van Zanden (2014). The Maddison Project: collaborative research on historical national accounts. *Economic History Review* **67**, pp. 627–51.

Broadberry, S., H. H. Guan, and D. D. Li (2014). China, Europe and the great divergence: a study in historical national accounting. Available from: http://eh.net/eha/wp-content/uploads/2014/05/, accessed 6 January 2016.

Commented [J3]: Probably there, but subscriber-only site.

Clark, G. (2017). Growth or stagnation? Farming in England, 1200–1800. *Economic History Review*. doi:10.1111/ehr.12528

Deng, K. and P. O'Brien (2015). Clarifying data for reciprocal comparisons of nutritional standards of living in England and the Yangtze Delta (Jiangnan), c.1644–c.1840. *Journal of World History*, **26**, 2, pp. 233–67.

Deng, K. and P. O'Brien (2016). Establishing statistical foundations of a chronology for the Great Divergence: a survey and critique of the primary sources for the construction of relative wage levels for Ming–Qing China. *Economic History Review* **69**, 4. pp. 1057–82.

Deng, K. and P. O'Brien (2017). Why Maddison was wrong. The Great Divergence between Imperial China and the West. *World Economics*, **18**, 2, pp. 21–41.

Drukker, J. W. (2006). *A Revolution that Bit its Own Tail*. Amsterdam: Amsterdam University Press.

Feenstra, R. C., R. Inklaar and M. P. Timmer (2015). The next generation of the Penn World Table. *American Economic Review* **105**, 10.

Fukao, K., D. Ma and T. Yuan (2007). Real GDP in pre-war East Asia: a 1934–36 benchmark purchasing power parity comparison with the US. *Review of Income and Wealth* **53**, 3, pp. 503–37.

Li, B., and J. L. van Zanden (2012). Before the Great Divergence? Comparing the Yangzi Delta and the Netherlands at the beginning of the nineteenth century. *Journal of Economic History*, **72**, 4, pp. 956–89.

Ma, Y. and H. J. de Jong (2016). Unfolding the turbulent century; a reconstruction of China's historical national accounts, 1840–1912. *Review of Income and Wealth* **63**.

Maddison, A. (1998) *Chinese Economic Performance in the Long-Run*. Paris: OECD Development Centre.

Pomeranz, K. (2000). *The Great Divergence. China, Europe and the Making of the Modern World Economy*. Princeton, NJ: Princeton University Press.

Vries, P. (2013). *Escaping Poverty. The Origins of Modern Economic Growth.* Vienna: V&R unipress GmbH.

Xu, Y., B. van Leeuwen and J. L. van Zanden (2015). Urbanization in China, ca. 1100–1900. Centre for Global Economic History working paper 63.

Xu, Y., Z. Shi, B. van Leeuwen, Y. Ni, Z. Zhang and Y. Ma (2016). Chinese national income, ca. 1661–1933. *Australian Economic History Review* (forthcoming).