

GDP and Convergence in Modern Times

Emanuele Felice*

Departament d'Economia i d'Història Econòmica, Universitat Autònoma de Barcelona, Bellaterra (Cerdanyola del Vallès), Barcelona, Spain

Abstract

In this chapter, I discuss historical estimates of GDP at both the national and the regional level and their application for assessing economic performance in modern times. Having been invented in (and conceived for) industrial capitalist societies, GDP has stronger informative power in those contexts where industry and services, and market exchange, retain the lion's share of production. In modern times, when comparing the series available for different countries, there are three major methodological problems to be acknowledged and possibly addressed: the dissimilarity of the quantity series and related proxies, deflation through purchasing power parities distant in time, and the differences in the base year used to construct GDP constant-price (Laspeyres) indices (the latter issue may be less widely recognized, but it may have a remarkable impact). The way the estimates are constructed also has a bearing upon the statistical tools and models we should use to interpret them; owing to the lack of reliable long-run series, cross-sectional techniques are often preferable to time series analysis; provided we have reliable estimates, growth accounting – decomposing GDP growth into productivity and industry mix effects – may provide important clues about the choice between theoretical approaches; not least for the quality of our data, cross-country convergence models based on conditioning variables should always be supplemented by historical information from qualitative sources and case studies. More generally, cliometricians should prove themselves capable of adapting their models to different historical contexts and relativizing findings to the limits of their estimates.

Keywords

GDP; Convergence; Purchasing power parity; Neoclassical school; Endogenous growth; New economic geography

Introduction

To the extent that economics should use facts to verify theories, history is precious, being the fieldwork where empirical information can be found. Of course, information must be reliable: potential mistakes but also methodological differences can affect the results to the point that data cannot serve the purpose, all the more so in international comparisons. When we deal with historical GDP estimates – the primary indicator of any macroeconomic reasoning – what may appear less obvious

Financial support from the Spanish Ministry of Economy and Competitiveness, project HAR2013-47182-C02-01, and the Generalitat de Catalunya, project 2014 SGR 591, is gratefully acknowledged.

*Email: emanuele.felice@gmail.com

is that in order to evaluate their soundness, we must rely not only upon historical knowledge but also on some basic expertise in quantitative techniques: economists may pick up a misleading series if they overlook the historical context, but nonquantitative historians can also accept the wrong figures if they are unable to assess the validity of the techniques used to produce them.

In this respect, quantitative economic historians – admittedly, a more comprehensive definition for cliometricians – are vital to both economics and more traditional history. From their historian backgrounds, they can provide a useful contribution to the former, insofar as they warn against a superficial approach to historical information (and estimates) based on the inattentive use of datasets and aprioristic assumptions about the past that do not meet the facts. They may even be able to contribute models that effectively account for historical change. Using their quantitative expertise, cliometricians may also help traditional historians understand why, and under which conditions, various models and estimates are useful descriptions of the past and tenable explanations for growth. In short, they can identify instances in which our historical interpretation should change according to the results proposed by quantitative history and economics. Such a double-sided task is not an easy one, because it implies that a good quantitative economic historian must have proficiency in both economics and history. However, the efforts have their rewards, as they may endow us with some of the most powerful instruments to understand the past.

GDP stands out among these instruments. It is virtually impossible for anyone studying economic growth to avoid using GDP estimates. Hence, it is important to understand how the series are constructed and what assumptions undergird the most popular growth models. However, it is also crucial to recognize that the choice of model and the interpretation of its results are informed and affected by the procedure employed to produce the figures. This chapter is dedicated to explaining and developing these issues. It reviews the procedures and uses of historical GDP estimates in modern times, roughly from the second half of the nineteenth century onward, at both the national and the regional level. In doing so, I highlight the main problems that can arise in terms of comparability between different estimates and make a case for improving explanatory models with an understanding of both the historical context and the GDP estimation procedures.

GDP: Concept, Limits, and Success

The production approach of calculating GDP considers it to be the sum of the final values of all the saleable goods and services produced within an economic system (a country or a region) over a certain period of time. Values are measured at market price, and they are final in the sense that they are net of the costs of intermediate goods and inputs to avoid duplication. According to the expenditure approach, GDP is the sum of consumption, investment, government spending, and net exports (exports minus imports). Finally, according to the income approach, GDP is the sum of all the incomes earned in that economic system (e.g., Lequiller and Blades 2006).¹ So many dimensions, into a single number: this is probably the ultimate reason behind its success. For instance, when divided by the number of inhabitants, total GDP corresponds to average income;² and when divided only by employment, it equals average per worker productivity. Production and expenditure, income, and productivity: the basics of any economic discourse cannot be addressed nowadays without GDP.

¹For a country, GDP includes the incomes earned by the individuals not officially living in that country. Gross national production (GNP) includes instead the incomes earned abroad by the citizens of that country.

²To be consistent with the definition of the previous footnote, GDP should be divided by the population de facto (present population) and GNP by the resident population.

Less widely known is the fact that the most important measure of economic performance is a recent invention, at least from an historical perspective. It was born in the United States during the Great Depression in order to monitor the impact of the 1929 crisis and the time and pace of recovery (Carson 1975). It was then elaborated in the National Bureau of Economic Research, a private institute of empirically oriented scholars directed by Wesley Clair Mitchell, one of the leading figures in institutional economics (Schumpeter 1950). Further, it should be credited mostly to the work of Simon Kuznets: under his authorship, the first official estimates were published in 1934, with reference to the US economy from 1929 to 1932 (Kuznets 1934). After World War II, in a western world governed by Keynesian policies (thus paying particular attention to cyclical fluctuations) and one strongly influenced by the economic and political power of the United States, GDP (and GNP)³ turned into official statistics in Europe⁴ and then throughout the world (although planned economies used a different system of national accounts). However, the origins of GDP should not be forgotten, at least from the point of view of cliometricians and economic historians, since they are essential in order to grasp the three basic features of the measure we are dealing with. First, GDP was conceived in an empirically oriented environment, as a sort of practical shortcut to solve the complex problem of how to monitor the economy, and thus it had strong theoretical limitations and even some related methodological contradictions. Second, it was born into an advanced industrial economy with the aim of measuring *that* economy, where industry (manufacturing) and services had by far the *magna pars* of national income to the detriment of agriculture (and mining) and where most of the production was sold and bought in the market. Third, it was created at a later stage in the history not only of the modern world but also of industrial capitalism as we have come to know it: it did not exist during the Industrial Revolution or in the first globalization era or at the time of World War I, not to mention medieval or ancient times.

There is now a vast literature on the theoretical limitations of GDP, which is of interest not only to economic historians and economists but also to social scientists and to an extent policymakers and the general public. Nevertheless, some confusion on this should be sorted out. Some of the limitations of GDP are neither theoretical nor the result of a methodological contradiction. For instance, GDP is neither a measure of well-being nor the standard of living: it excludes the nonmonetary dimensions of well-being (from clean air to free time to the quality of affective life), while including other items that do not contribute directly to well-being but at best prevent it from falling (such as the expenditures on defense or on the administration of justice), and it does not consider the impact of the distribution of income on personal utilities. But there is no contradiction on this: GDP simply was not born for this purpose. GDP cannot be a measure of “human development” – at least as intended in the capability approach by Sen (1985) that was developed half a century after the creation of GDP – since it does not allow for other fundamental dimensions of human development, namely, education and longevity.⁵ But again, GDP was never designed to be a comprehensive measure of all the desired goals a human being can nurture, and so there is no

³The United States used GNP instead of GDP as late as 1991. By that time, virtually all the other countries had already adopted GDP.

⁴The first official estimates for the United Kingdom were made in 1941 by Richard Stone and James Meade. The former also was the main contributor to developing a standardized system that since 1952 was implemented in OEEC (Organization for European Economic Cooperation) countries (Stone 1956, 1961).

⁵However, many others are equally excluded: take, for instance, political and civil freedoms. Nussbaum (2000) increases up to ten the number of basic capabilities: (1) life; (2) bodily health; (3) bodily integrity; (4) sense, imagination, and thought; (5) emotion; (6) practical reason; (7) affiliation; (8) other species; (9) play; and (10) control over one’s environment.

contradiction or theoretical limitation in this. Rather, limitations are in those who regard GDP as the ultimate icon of human fulfillment. But even then, it is only fair to acknowledge that there is still no agreement about alternative measures to GDP that would better monitor nonmonetary dimensions. Even the Human Development Index, which is gaining consensus among economic historians (Crafts 1997, 2002; Prados de la Escosura 2013a, b), is far from undisputed for what concerns its formula, weights, and components (Prados de la Escosura 2010; Ravallion 2012a, b), let alone its theoretical foundations. This may be the fundamental reason why GDP, although *it is not* a measure of well-being and human development, was and still is often *considered to be one* or at least a measure of economic progress, broadly defined.

Similar arguments can be raised to oppose another well-known accusation brought against GDP: it excludes unpaid work (Waring 1988). This can have paradoxical effects, such as the often quoted textbook insight that having grandparents take care of children, instead of hiring domestic help, may cause a fall in GDP. But we need to remember that GDP was conceived when policymakers needed to contrast official unemployment, not unofficial employment. Less known but particularly telling is instead what happens with the mining sector, which actually represents a theoretical limitation (and even a methodological contradiction). At the time GDP was invented, the US census didn't ask firms owning their mines to declare the value of their reserves (Fenoaltea 2008). As a consequence, GDP does not compute the net value of production or value added (total mining production minus an estimate of the depletion of natural resources), but only the value of outputs. In other words, the more you consume your reserves, the more GDP (artificially) increases. The mining sector is important by itself, of course, but also for being part of a major problem. GDP has serious theoretical limitations in dealing with the environment. Not only does it not account for air and water pollution or land contamination, but indeed all these phenomena can even indirectly increase GDP, as long as they lead to the creation of specific counter-pollution activities in the market economy. This is probably the most worrying issue, which in the future may negate the ability of GDP to measure economic progress, at least until it is modified to account for some costs of pollution and the consumption of the planet's resources.⁶ Of course, at the time GDP was invented, the concern for the environment was practically unknown in the United States or anywhere else.

The second and third characteristics of GDP should be of particular concern to cliometricians and economic historians. GDP was born in order to monitor advanced industrial economies, where most of the production comes from industry and services. In these sectors, there are two factors of production, labor (L) and capital (K), meaning that the standard growth model starts from the following production function: $Y = f(L, K)$. A widely accepted specification of this function is the Cobb–Douglas form

$$Y = A \times L^\alpha \times K^\beta \quad (1)$$

and in particular the one with $\alpha + \beta = 1$ (i.e., with constant returns to scale)

$$Y = A \times L^\alpha \times K^{1-\alpha}. \quad (2)$$

In both Eqs. 1 and 2, α and β (or $1 - \alpha$) are the output elasticities of labor and capital, respectively, and in Eq. 2, assuming perfect competition, α and $\beta = 1 - \alpha$ also are their respective shares of output (Douglas 1976). A stands for total factor productivity (TFP), a factor measuring the

⁶In this direction, some progress has recently been made, but with little or no heed, thus far, in the systems of national accounts: see Boyd and Banzhaf (2007) and Ferreira et al. (2008).

efficiency with which capital and labor are employed in production: this captures both the technological change not incorporated in capital and the gains of efficiency in production processes due to the reallocation of activities from one sector to another (Solow 1957). Provided that we find values for α , or for α and β , and that we reconstruct the amount of labor (number of workers or, better, number of hours of work) and the value of capital (the physical capital stock, in turn composed of machinery, infrastructure, and equipment; means of transport; nonresidential construction; housing), the growth rate of GDP (Y) can be decomposed into the contributions of increases in labor (L) and in capital (K) and of improvements in their combination (A). And even if we don't have values for α and β , whose historical estimates are usually far from undisputed, the formula clearly indicates that capital deepening (K) and TFP growth (A) bring about an increase in GDP per worker (Y/L). According to the simple equation $Y/P = Y/L \times L/P$, GDP per worker is in turn one of the two determinants of GDP per capita (Y/P), the other being the percentage of workers in the total population (L/P). In short, this means that technological progress (in its broader sense) leads to a rise in GDP per worker and hence GDP per capita. Thus it follows that, other things being equal, countries with higher GDP are more technologically advanced.

These conclusions do not necessarily hold in a preindustrial world where agriculture maintains a significant share of the total output. The agricultural production function includes land as a third factor of production. Furthermore, similar to the problem with mining, GDP does not compute land as a cost (again, in part as a consequence of the specific context in which it was created): in agriculture, when passing from gross saleable production to value added, a figurative sum to account for the extension of the land used to produce agricultural goods is *not* detracted, as if land were an inexhaustible resource. All of this means that a rise in GDP, either per worker or per capita, can be due not only to technological progress but also to an extension of the land cultivated. In turn, this implies that in the preindustrial world, we can have countries with high GDP – or with high standards of living – that are not technologically advanced. They may be rich simply thanks to a favorable relation between land and population (because they have high land per capita), but that land can be inefficiently used: they would have low *per hectare* GDP (land productivity), but since they may rely upon a lot of land, relatively high *per worker* (and thus per capita) GDP. Obviously, in this situation, the standard coefficients of the Cobb–Douglas function do not hold. In addition, the assumption of perfect competition may be incorrect, at the very least because a significant proportion of preindustrial societies are not even market economies. These considerations make the use of GDP for eras and contexts radically different from ours, namely, for those preceding the Industrial Revolution, particularly problematic. At the very least, the interpretation we give to those GDP figures should be more cautious and not a mere replication of the interpretative framework we have assumed for the last stretch of human history. Because of such limitations, in turn I am limiting the present study to the use of GDP in modern times.

Even so, however, things are far from simple. And here we come to the third characteristic of GDP than any cliometrician or economic historian (but also any shrewd economist) should always have in mind. As discussed, the first official statistics of national income were produced in the United States in the 1930s. They progressively spread across the world only after World War II. For the previous periods, quantitative historians or applied statisticians – or “chipprephiles,” as Maddison (1994) once named himself – had to reconstruct their own historical series of GDP by making the best out of several different sources and hypotheses.⁷ When they were lucky, there could benefit from data on production, prices, labor force, and wages, but these data were not always comprehensive or

⁷They could, of course, take advantage of a long tradition of income and macroeconomic estimates, dating back to the seventeenth century (for an outline, see Maddison 2007, pp. 393–401).

exhaustive and often not even available. We may draw a line roughly at the mid-nineteenth century. For earlier epochs, available sources are scant, and GDP estimates often come from a handful of figures on urbanization and demography; related assumptions on the share of nonagricultural sectors (as most of Maddison's figures for the years before 1820), plus a few reliable series on prices and wages for a limited number of countries; and maybe some information about public revenues and tax collection. We cannot help warning once again against a too relaxed use of these shaky figures. Gregory Clark (2009, p. 1156) has efficaciously defined Maddison's pre-1820 estimates "as real as the relics peddled around Europe in the Middle Ages."⁸ However, a more in-depth discussion of these issues would go beyond the scope of this chapter.

For the years after the mid-nineteenth century, which also coincide with our period of concern, historical data are much more abundant and solid: they usually include production series that are complete, or nearly so, and at times also extended price series, plus reliable and highly detailed data on wages and employment in some benchmark years (those of official censuses). This is true for Europe, at least, where following the Enlightenment and Napoleonic wars in the course of the nineteenth century modern bureaucratic states replaced *ancien régime* governments. For other parts of the world, the colonial administrations notwithstanding, unless we are willing to use indirect procedures (such as import – export charts), more often than not we must wait until the second half of the twentieth century, when we are in the realm of the official GDP statistics.

In other words, historical GDP reconstructions are the result of ad hoc efforts by individual scholars⁹ who had to make the best possible use of the available incomplete sources. The available information typically changes from one country to another, and even within the same country, it changes across years and economic sectors. As a consequence, even for modern times, (country and regional) GDP series are often the product of different methodologies and hypotheses, and this has significant bearings on the results. Cliometricians need to be aware of the methodologies (and limitations) behind the GDP series they are using. The following section is intended to offer an outline of the main methodological problems we encounter when dealing with, and working on, historical GDP estimates in modern times.

Reconstructing GDP: Methods and Problems

In order to be able to assess the soundness of GDP figures, transparency is of course a preliminary condition: sources and methods must always be adequately described, ideally up to the point that results must be replicable. This may seem obvious, but actually it is not. For example, Italy's official historical series of GDP (beginning in 1861), one of the first in the world to be produced (Istat 1957), was a pioneering effort that also came to be famed for its lack of transparency in sources and methods, which did not help remedy the faults discovered by subsequent scholars (Federico 2003; Fenoaltea 2003; Felice and Carreras 2012). The original series has finally been replaced with a new one reconstructed almost entirely by economic historians (e.g., Baffigi 2013), more than half a century after it was originally published. Every country has its issues in this regard, and it would be impossible to review them all. The good news is that the standards have changed, and now an established rule of the scientific community is that GDP estimates must be transparent and replicable, which they are, for the most part. Maddison's magnum opus (1995, 2001, 2006), which

⁸However, some improvements on this are now on the way (Bolt and van Zanden 2014).

⁹For modern times, outstanding examples are Feinstein (1972) for the United Kingdom and Prados de la Escosura (2003) for Spain.

presents GDP figures spanning the past 2,000 years for most countries, also accomplishes this rule: although some of his assumptions for the nineteenth century are questionable – or may simply look too crude¹⁰ – an outline of the procedure is always provided, with further reference to the primary and secondary sources used; new contributions from the literature are also discussed and at times properly integrated.¹¹ The *Maddison project* was created¹² in 2010, the same year that Maddison died. Its aim is to revise and improve Maddison’s original dataset as new information becomes available. The first results have already been produced, and they incorporate a great deal of the new statistical evidence and historical estimates that had become available in the meantime (Bolt and van Zanden 2014).¹³ Other scholars are at work on comparative estimates for shorter periods of time or with a sectoral focus, producing data that can usefully complement and integrate those of Maddison. For its scope and accuracy, it is worth citing Williamson’s (2011) *Project on industrialization in the poor periphery*, which, after reviewing and harmonizing a number of primary and secondary sources, presents estimates of industrial output for the period 1870–1939 at constant prices for the European eastern and southern periphery (12 countries), Latin America (7 countries), Asia (7 countries), the Middle East (Egypt and the Ottoman empire), and Africa (South Africa), plus three leaders (Germany, the United Kingdom, and the United States). As these works progress, it is possible to imagine a future in which we may be able to take advantage of an international GDP dataset whose problems of reliability and comparability will have been progressively reduced and perhaps even become negligible.

However, reaching such a goal will not be an easy task, and it is only fair to acknowledge that we are still far from it: information is lacking and research is sparse not only for minor countries but also for the most important ones whose data surely look more robust. Moreover, even when we have *reliable* estimates, it is not assured that these are *comparable* between countries.

Indeed, comparability probably looms as the biggest challenge. At least three problems need to be recognized: one is about quantities, while the other two are about prices. However, at this point, before entering into further details, it may be useful to provide an outline of how GDP series are normally produced. As a general rule, since price data are not usually available throughout the period, but only for some reference years, GDP series are estimated at constant prices: a base year is

¹⁰Just a handful of examples: for Switzerland, per capita GDP growth from 1820 to 1951 is assumed equal to average for France and Germany (Maddison 2006, p. 409); for Italy, a “guesstimate” for 1820 is created, “assuming that GDP per capita grew at the same pace from 1820–1861 as from 1861–90” (Maddison 1991, p. 234; Maddison 2006, p. 408); but for this country, see Malanima (2006, 2011), his 2011 article having been incorporated in the updated version of Maddison’s database (Bolt and van Zanden 2014). For Albania, per capita GDP from 1870 to 1950 was assumed to move in the same proportion as the average for Bulgaria, Romania, Yugoslavia, Hungary (!), Czechoslovakia (!), and Poland (!); but what is more worrisome, this same average should work also for the entire Russian empire (Soviet Union territories) from 1820 to 1870 and for Greece from 1820 to 1913 (Maddison 2006, pp. 407, 469–471).

¹¹See, for example, the review of Good and Ma’s (1999) proxy measures of per capita GDP for six eastern European countries (Bulgaria, Czechoslovakia, Hungary, Poland, Romania, Yugoslavia) plus Austria, which are derived by regression by using three indicators (letters posted per capita, crude birth rate, and the share of nonagricultural employment in the labor force) and are accepted by Maddison only for some countries (Bulgaria, Poland, Rumania, Yugoslavia), owing to the lack of any other information (Maddison 2006, pp. 403–404, 471–472). For a comprehensive picture of Maddison’s amendments to his previous (2001) estimates, see Maddison (2006, p. 624).

¹²The project consists of a small working party of four established economic historians and a larger advisory board composed of 22 scholars from around the world. See the website of the project: <http://www.ggd.net/maddison/maddison-project/home.htm>.

¹³Despite the title of the article (“Re-estimating Growth Before 1820”), updated estimates referring to the last two centuries also are included.

taken (for which there are current-price GDP estimates) and that *current-price benchmark* becomes the year of the *constant-price series*. In order to do so, for each i sector and t year, it is assumed that

$$\text{GDP}^{ti}/Q^{ti} = \text{GDP}^{(t+1)i}/Q^{(t+1)i}, \quad (3)$$

where Q is the elementary physical series. In other words, it is assumed that for each elementary series, the relation between GDP and quantity, that is, unitary GDP, does not change throughout the years of the series, with respect to the unitary GDP of the baseline year. From Eq. 3, we obtain the formula used to produce constant (base year)-price estimates as

$$\text{GDP}^{ti} = (Q^{ti}/Q^{yi}) \times \text{GDP}^{yi}, \quad (4)$$

where y stays for the baseline year.

From this formula, the problem with quantities is almost self-evident. Ideally, the elementary physical series of each country must be taken at a similar level of decomposition. This in turn should be as high as possible, because within each country, the physical series should be homogeneous. For instance, we should not estimate textiles via the total amount of textiles produced; rather, we must include separately at least each major fiber (silk, cotton, wool, linen), and, indeed, even within a major fiber, at least the main production processes (spinning, weaving) should be broken down. On the basis of textiles then, one could argue that for each country it suffices to use the official series (of production and trade), which would then produce the finest comparable aggregate national series. But what about other sectors, such as mechanics, a sector with a non-negligible and growing impact on total GDP? In the long run, productions have changed enormously; even within a single subsector and a single production (e.g., automobiles), there are different types whose prices significantly vary from one model to another. And even the models could change: some disappear and we find them replaced by others, both backward and forward in time. As a consequence, in practice, for each country we must rely on a different methodology in order to produce the elementary physical series: not only the level of decomposition varies, but we also often resort to different proxies within the same sector or the same series (say, raw cotton instead of yarn cotton) with different hypotheses to cover the unknown productions (say, different elasticities between the other textiles, or the rest of cotton, and the chosen proxy). Even within each country, there may be problems of comparability between different periods. For example, a remarkable degree of decomposition has been reached for the Italian industry in the liberal age, for which about 200 elementary series have been produced (Fenoaltea 2003). But it was not possible to maintain the same level of decomposition in the interwar years, when “only” 90 industrial series could be produced (Felice and Carreras 2012). Moreover, how can Italy be compared with other countries for which only the major industrial sectors can be estimated?

Procedures also vary because there is no common rule to firmly guide us. One rule could be “disaggregate as much as you can,” but this inevitably results in many country-specific procedures, following differences in the systems of national statistics as well as the accidental availability of supplementary sources. Alternatively, it could be argued that if our goal is to compare the performances of countries, we should shift from the rule of disaggregating (which comes from a very national-centered estimating approach) to a “lowest common denominator” approach that would work for the highest number of countries. For example, we could decompose industry into a few major sectors, each one estimated through its aggregate total production (in quantities, say, tons, weighted with prices) or its most important product. However, not even this would solve the problem, simply because the most representative productions would also vary from one country

to another, with possible distortions. To sum up, we must resign ourselves to the fact that having perfect cross-country comparability in elementary series spanning long periods of time is all but a chimera. Once this human (social) limitation is accepted, we can look with more indulgence at the current state of the elementary series used to produce the available historical GDP estimates. That is, a disparate collection of what has been done in different countries during recent decades, by separate scholars concentrating on their own sources and problems, unworried by the need for a common aggregating methodology.

When dealing with constant-price series, however, comparability in prices may even be a more serious issue. In the choice of elementary price data, we encounter more or less the same problems briefly discussed above for physical quantities (although these are usually limited by the use of a few benchmarks instead of long series). However, there is also a further significant distortion due to the way in which relative prices vary over time. From Eq. 4, in fact, it is true that, for $i = 1 \dots n$ productions, total GDP (GDP^N) is

$$GDP^{tN} = \sum_{i=1}^n \left(\frac{Q^{ti}}{Q^{yi}} \right) \times GDP^{yi} \quad (5)$$

In Eq. 5, we can see that to each physical series, a GDP weight has been assigned, which is constant over time and corresponds to the GDP weight of that single production in the base year: this depends on the unitary GDP and the quantity produced, again in the base year. As Fenoaltea (2010, p. 91) efficaciously pointed out, such a bold assumption “is done . . . with a bad conscience but with good precedent: all sorts of scholars, similarly constrained, have done the same.” In short, Eq. 5 is a Laspeyres quantity index number, which uses the GDP weights of a base (fixed) year to convert the component quantities to comparable values and, at the same time, to weight them. Actually, most of the available GDP series are Laspeyres quantity indices.¹⁴ Of course, unitary GDP is the result of the price system in use that year, i.e., the relative price of that single production compared with the others, at a specific point in time. The problem is that relative prices (and thus unitary GDPs) do not remain constant over time. It is well known that prices and quantities are usually negatively correlated, on the demand as well as on the supply side, especially in the presence of technological progress, which reduces the unitary costs of production. Over the course of decades, in fact, some sectors and productions (e.g., chemicals and mechanics in the West between the late nineteenth and the twentieth century) grow faster than others thanks to technological progress. As a consequence, the early-weight price series, those based on a price system early in time (say, 1870 in an 1870–1913 GDP series), assign a higher weight to the sectors growing faster (whose quantities increase and relative prices decrease), and therefore, they grow more rapidly in the long run. For the same reason, the late-weight indices (say, a 1913-price series) grow less. This has become known as the “Gerschenkron effect,” since it was reasoned by Alexander Gerschenkron (1947), soon after World War II, when analyzing Soviet indices of industrial production. Today, it is also simply known as the “index number problem” (Feinstein and Thomas 2002, p. 513).

Of course, the Soviet Union in the interwar years was an extreme case of accelerated growth in heavy industrial sectors, and thus the distortion caused by the “Gerschenkron effect” was fundamental. However, it is worth stressing that the index number problem is also serious in countries that modernized at a slower pace. For example, Italy from 1911 to 1951 ranked more or less in the middle

¹⁴For a detailed discussion of Laspeyres indices and their properties as well as of the other main indices used in time series, see Feinstein and Thomas (2002, pp. 507–525).

among OECD countries.¹⁵ For Italy, three indices of industrial production at three different price bases are now available, all made up of the same elementary physical series (only the relative weights in the unitary GDP, which are 1911, 1938, or 1951, change). From 1911 to 1951, the 1911-price index of industrial value added more than triples from 100 to 362; the 1938-price index goes from 100 to 264; and the 1951-price index doubles from 100 to 210 (Felice and Carreras 2012, p. 447). It is clear that such a major distortion cannot be ignored when it comes to international comparisons. If large differences are observable in the *same* series (i.e., series constructed with the same methodology and proxies), which differ only in their benchmark years, then when it comes to comparing *different* series belonging to different countries, a minimum requirement is that their base years be the same or at least relatively close.

Nevertheless, this is barely the case. Actually, Maddison's GDP estimates put together a large collection of different price bases, following once again the national accounting systems of every country and the work of separate scholars. Even a brief examination of the price bases that Maddison reports to have used in order to produce constant-price GDP series offers a discomfiting picture: Austria, 1913 (for the 1820–1913 series) and 1937 (1913–1950); Belgium, 1913 (1913–1950); Denmark, 1929 (1820–1947); France, 1870 (1820–1870); Portugal, 1910 (1851–1910); Switzerland, 1913 (1913–1950); Australia, 1910/11 (1861–1938/39); the United States, 1929 (1890–1929) and 1987 (1929–1950); the Soviet Union, 1913 (1870–1928) and 1937 (1928–1950). And this is an incomplete list (Maddison 2006, pp. 403–409, 450–457, 471).¹⁶ This means, for instance, that for 1913–1951, Switzerland is barely comparable with Austria, and the same is true for Belgium in comparison with Denmark, for the Soviet Union in comparison with the United States, and so on.

It is worth noting that the “Gerschenkron effect” produces a distortion not only for what concerns international comparisons but also in terms of intra-sectoral comparisons within the same country: the GDP sectoral shares of a series at constant prices tend to remain very close to those of the base year, for obvious reasons (only quantities vary). Both these distortions (between- and within-country) would not be present if we were able to estimate GDP at current prices for each year of the series – as is done today. In order to have “real” GDP figures, current-price GDP series could then be deflated by using a single common deflator instead of sector-specific deflators as in Eq. 5: wages, for instance (Fenoaltea 1976). In this way, we would have constant-price series, unbiased toward the GDP composition of the baseline year, and comparable between countries. However, such a procedure is too data demanding, and in the end, it may also turn out to be a chimera, not least because the choice of the deflation system is far from undisputed (e.g., wages would ignore the share of GDP going to capital gains, while a consumer price index would ignore the price of investment goods). What can be reasonably done is to estimate as many current-price benchmark years as possible for every country. From these, short constant-price series can be created. Finally, a long-run constant-price series can be produced by connecting the shorter series through chain indices: ideally, a chain index rebased every year (a Divisia index) could be created. Alternatively, a Fisher Ideal index can be produced: the early-year and late-year indices can be combined through a geometric average, with weights inversely proportional to the distance between the year of the series and the price basis, according to the formula

¹⁵For updated international comparisons of Italy's GDP with the rest of the world, in 10-year intervals from the unification of the country (1861) until 2011, see Felice and Vecchi (2013, p. 28).

¹⁶For further details and more countries, reference must be made to the previous version of Maddison's work (1995, pp. 126–139) and to the country-specific sources cited by the author.

$$y \frac{i_{\max}-i}{i_{\max}-i_{\min}} \times y \frac{i-i_{\min}}{i_{\max}-i_{\min}} \quad (6)$$

i in *i*min prices *i* in *i*max prices

where *i* is the year of the series *y*, *i*min is the early benchmark, and *i*max is the late one.¹⁷

The third problem when comparing international GDP series comes with purchasing power parities (PPPs), which is not at all a minor issue (indeed, it is probably more easily recognizable than the Gerschenkron effect). With the ambitious goal of comparing not only income and production but also the standard of living, Maddison converted all his country estimates into Geary–Khamis PPP 1990 international dollars. It goes without saying that any purchasing power converter is different from the official exchange rate, since it allows for differences in the cost of living. The procedure is simple: (a) each national GDP series, expressed in constant prices at its own national currency, is converted into an index; (b) at the same time, for the baseline year 1990, each national GDP, expressed in its own national currency and at current prices, is converted into 1990 international dollars by using Geary–Khamis PPP deflators;¹⁸ (c) with the index in (a), a new national series in Geary–Khamis PPP 1990 international dollars is then created. By using this method, all series can be converted into a comparable unit of measurement without changing the growth rate of each national series. In order to estimate PPP converters, different multilateral measures (and methods) can be used, but it must be acknowledged that Geary–Khamis is a suitable one because it assigns each country a weight corresponding to the size of its GDP and considers the United States, the most important economy, as the *numeraire* country (i.e., the 1990 Geary–Khamis dollar has the same PPP as the US dollar has in the United States in 1990).¹⁹ However, of course, both the country weights and the purchasing power differences are those measured in 1990. In fact, Maddison’s entire magnificent edifice is based upon the situation recorded in 1990, as if the relative purchasing power of currencies (both domestic and international) was fixed, rather than changing over time, especially in the long run, as both the underlying forces (namely, the domestic and international flows of goods and services) that govern the movement of prices and the basket of goods and services used to construct the PPP converter change. This problem becomes more serious if we go further backward in our extrapolation, thus distancing ourselves from the baseline year. As Prados de la Escosura (2007, p. 18) put it:

As growth occurs over time, the composition of output, consumption, and relative prices all vary, and the economic meaning of comparing real product per head based upon remote PPPs becomes entirely questionable. Hence, using a single PPP benchmark for long-run comparisons implies the hardly realistic assumption that no changes in relative prices (and hence, no technological change) takes place over time.

¹⁷For an application of Divisia and Fisher Ideal indices, see Crafts (1985) for England, Prados de la Escosura (2003) for Spain (Fisher Ideal index), and Felice and Carreras (2012) for Italy (Fisher Ideal index). In Prados de la Escosura (2003, pp. 46–47), an application of the Paasche index can also be found: the Paasche index (which uses a changing set of prices to value the quantities) is used to produce price series, which are then combined with the Laspeyres quantity index to estimate GDP at current prices.

¹⁸The Geary–Khamis purchasing power converters for most countries can be found in Maddison (2006, pp. 189 (OECD countries), 190 (five East European countries and USSR), 199 (Latin America), 219–220 (Asia), 228 (Africa)). The reference year was always 1990 only for OECD, East European countries, USSR, Japan, and China; for the others, it varies from 1975 to 1993.

¹⁹Other multilateral measures either give all countries the same weight (such the EKS system used by Eurostat for political reasons), are a shortcut approach based on reduced information (such as ESCWA used for eight West Asian countries) or employ as a numeraire a currency different from the US dollar (such the ESCAP measure used for 14 East Asian countries, which takes as a reference the Hong Kong dollar) (see Maddison 2006, p. 172).

Even over a period of four decades, the distortions from the use of a baseline benchmark distant in time are large, “above 5 % and often much higher, while showing a high dispersion” (Prados de la Escosura 2000, p. 4). For these reasons, the use of a number of PPP converters at different points in time, following at least the main historical ages, would be preferable; but constructing PPP converters is a highly demanding task in terms of time and resources (Ahmad 1988) and one undermined in terms of feasibility (and reliability) by data scarcity for the period before World War II. Given the lack of reliable PPP converters for distant periods, Maddison’s approach, which was actually pioneered by Bairoch (1976), still represents a viable, if suboptimal alternative. It has been argued, for instance, that the distortion caused by comparing real products on the basis of long-run PPP projections can be larger than that generated by using current nominal exchange rates (Eichengreen 1986); thus, even simple exchange rates could turn out to be a more practical shortcut.

And yet there is indeed a superior shortcut, which is based on the reasonable assumption that price levels between a country and the rest of the world move according to some basic economic characteristics (e.g., the share of international trade, income, or population). By further developing the method originally envisaged by Kravis et al. (1978), Prados de la Escosura (2000) tested a number of variables against the 8 available PPP benchmarks (spanning 1950–1990) for 23 countries, through panel regressions. As a result, he proposed a structural relationship for each country between its price level (defined as the ratio between PPP and exchange rates), on the one hand (y , dependent variable), and its nominal GDP per capita plus an additional set of explanatory variables (ratio of commodity exports and imports to GDP, population, area, a periphery dummy indicating if the country’s nominal income represents half or less the US income), on the other ($x_1, 2, 3, 4$ and a dummy independent variables).²⁰ By applying the estimated parameters to the independent variables recorded in past times for the same countries (when available) as a second step, Prados de la Escosura could calculate additional PPP benchmarks, spanning 1820–1938 (and for some countries, previously uncovered, up to 1990), and then propose comparisons of real per capita GDP at current historical PPPs. The author is aware of the limitations of his method that “even for the same group of countries” is based on “the application of a structural relationship derived from advanced western economies over the past 50 years to earlier and different historical contexts.”²¹ Nevertheless, the potential error is minor compared with that residing in Maddison’s approach.²² The latter retropolates a PPP without any adjustment; in Prados de la Escosura, we still have retropolation, but with adjustments for changes in the underlying economic structure based on an empirically tested relationship. Thus far, the results from Prados de la Escosura’s method are available only for a limited number of countries. This may be the main reason why Maddison’s data continue to be so widely used, even in papers published in top economics journals: they are the only available long-run GDP series for many countries (or, in any case, those more easy to pick up), their patent unreliability notwithstanding. Bad conscience, but good precedent. To clean our conscience or to make it feel even more guilty, it is fair to warn against this habit.

²⁰In that article, an excellent discussion of the literature about these issues and the different shortcut methods is also provided (pp. 2–8).

²¹Prados de la Escosura (2000), p. 19.

²²As confirmed by the results. Just a couple of examples: in 1860, according to Maddison, Greece would have a per capita GDP higher than France (0.855 vs. 0.850), while according to Prados de la Escosura, France had a much higher GDP per capita (0.821 vs. 0.405) in 1860, 1870, and 1880. According to Maddison, Austria (at pre-World War I borders) would be above France, Germany, and Canada, while according to Prados de la Escosura, and much more plausibly, it would be below them (2000, pp. 24–25).

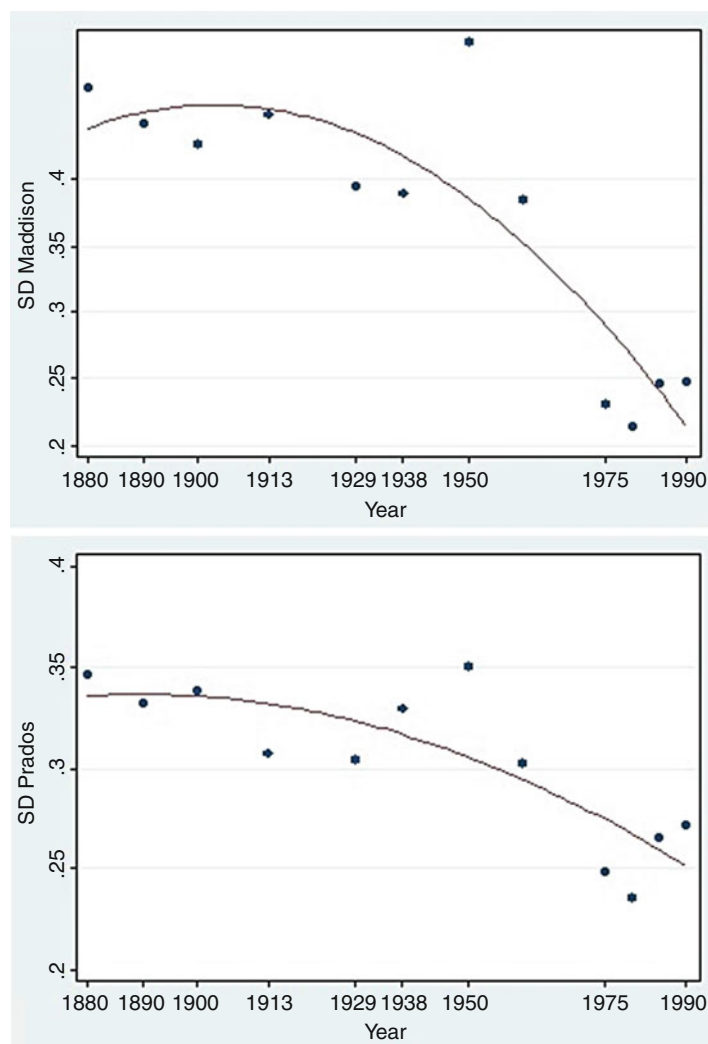


Fig. 1 σ -Convergence in GDP per capita from 1880 to 1990, according to different GDP estimates (Sources and notes: see text)

Convergence or Divergence? Measures and Models

Provided we have relatively sound estimates, we may then investigate the patterns of GDP growth in modern times. Did the country converge over time? A number of techniques are available to measure convergence, and different underlying theories are available to interpret the results. Techniques based on time series allow us to detect differences in cycles in trends and to identify country-specific break points. Unfortunately, they are more data demanding: any user should always check for the fact that the series at hand is not the result of some extrapolation or interpolation, as is often the case with historical estimates. Cross-sectional analyses allow us to test convergence when only a few benchmarks are estimated (possibly, each benchmark at its current prices) and therefore may result in more appealing long-run comparisons. Of course, they only consider the trend and for this can miss relevant information in between the two benchmarks.

Two concepts of convergence (Barro and Sala-i-Martin 1991) are generally accepted and used mostly – especially the second one – with benchmark data. σ -Convergence is a measure of dispersion in per capita GDP between different countries. The σ prefix comes from the standard deviation, which is

used to quantify it. A simple test of σ -convergence is provided in Fig. 1. This figure displays the standard deviation of the logarithm of real per capita GDP for 20 countries, in selected benchmarks, from 1880 to 1990; the benchmarks are those for which the estimates by both Maddison (upper quadrant) and Prados de la Escosura (lower quadrant) are available for an unchanged minimum number of 20 countries.²³

As can be seen, the results can differ significantly. For the same countries, σ -convergence is much stronger using Maddison's estimates than it is when using Prados de la Escosura's estimates. This should not come as a surprise, given that differences in nominal GDPs and differences in PPPs are usually positively correlated. Both authors record convergence in GDP per capita, and this means that differences in this variable are lower in later periods; for this very reason, Maddison's differences in PPPs (which are for 1990), when reprojected, may also be lower than the real (historical) ones (say, for the nineteenth century). The latter are those estimated and employed by Prados de la Escosura, who then makes use of higher differences in PPPs in the early periods. This means that in early periods, the cost of living was lower in poorer countries than that supposed by Maddison, and therefore poorer countries had at that time higher real GDP; as a consequence, they converge less.²⁴ However, it is also worth noting that both authors report similar trends: most of the convergence took place from 1950 to 1975, but then it came to a halt and even reversed.

The fact that poorer countries grow faster than richer ones is usually regarded as a precondition for a decrease in dispersion. Technically, this is known as β -convergence, which can be conditional or unconditional. The prefix in this case derives from the coefficient of the regression model used to measure it (Eq. 8). β -Convergence can be tested by regressing the growth rate of per capita income with its initial level; if there is a negative correlation, then countries with higher per capita GDP are growing less. It is worth noting, however, that when we have unconditional (or absolute) β -convergence, we may not necessarily also have σ -convergence. For example, the initial GDP of a country may pass from 0.6 to 1.6 (the average being 1), implying β -convergence but also an increase in dispersion (σ -divergence). The opposite, however, is not true, namely, if we have σ -convergence, we always have β -convergence. If a country goes from 1.6 to 0.6, we record both σ - and β -convergence. For the same countries as in Fig. 1, β -convergence is tested in Fig. 2, where the growth rate from 1880 to 1990 in real per capita GDP is regressed on the logarithm of initial income.

As expected, β -convergence is stronger in Maddison's than in Prados de la Escosura's estimates: in the former, R^2 is considerably higher, and, as a consequence, the standardized β -coefficient is also more elevated (-0.869 vs. -0.738). Figure 2 also provides information about the relative performances of individual countries, namely, which grew above average, *given their initial income*, and which grew below; the former position themselves above the fit line, whereas the latter are below. For example, in both cases, Argentina records a disappointing performance, while Japan is the big winner. The entire European northern periphery (Sweden, Finland, Norway, Denmark) has been growing above the average, and today it is no longer periphery. Instead, the southern periphery (Portugal, Spain, Greece), with the exception of Italy, is below the average.

Why do some countries converge more than others? Economic theory is replete with elaborate models to explain the observed patterns. In the space of a few pages, it is impossible to review all of

²³The countries are Argentina, Australia, Austria, Belgium, Canada, Denmark, Finland, France, Germany, Greece, Italy, Japan, the Netherlands, New Zealand, Norway, Portugal, Spain, Sweden, the United Kingdom, and the United States. The benchmarks are 1880, 1890, 1900, 1913, 1929, 1939, 1950, 1960, 1975, 1980, 1985, and 1990. GDP per capita is expressed in 1990 international dollars, but in the case of Prados de la Escosura, the figures are rescaled with his current-price PPPs (2000, pp. 24–31).

²⁴It should be reminded that all are expressed in logs. In absolute terms, the standard deviation of real GDP per capita increased in both Maddison and (more) in Prados de la Escosura.

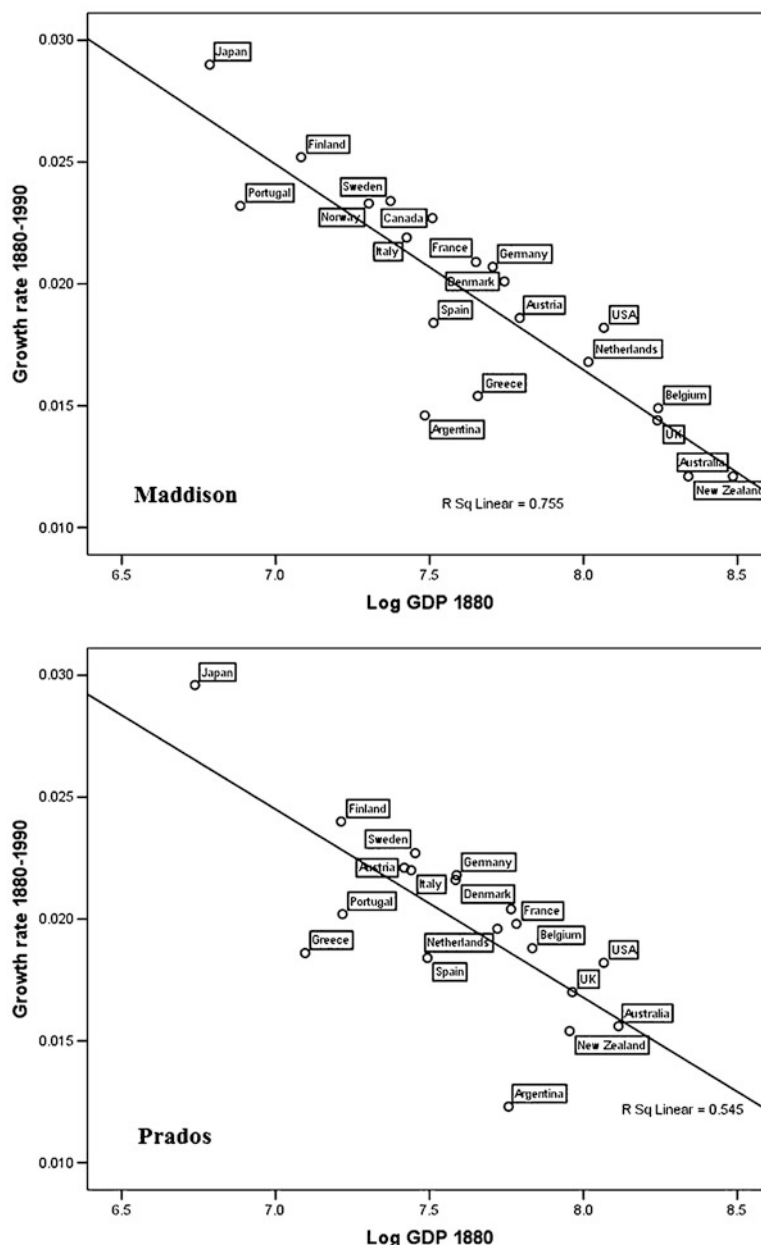


Fig. 2 β -Convergence in GDP per capita from 1880 to 1990, according to different GDP estimates (Sources and notes: see text)

them thoroughly, but we may provide a sketch of the most important (and popular) ones. β -Convergence, both conditional and unconditional, can easily be incorporated in the neoclassical approach. This is based on the assumption of diminishing returns to capital or, in other terms, the downward slope of the savings curve. According to Solow (1956) and Swan (1956), in a closed economy where savings are equal to gross investments, the growth rate of capital stock would be

$$\gamma_k = s * Af(k)/k - (\delta + n) \quad (7)$$

where s is the constant savings rate, ranging from 0 to 1, k is the capital stock per person, $Af(k)$ is the production function in per capita terms, δ is the depreciation rate of the capital stock, and n is

the exogenous rate of population growth. Thus, $\delta + n$ is the depreciation curve, a horizontal line, and $s \cdot Af(k)/k$ is the savings curve, a downward-sloping line. The argument for convergence holds that, given diminishing returns to capital, each addition to the capital stock generates higher returns when the capital stock is small. Of course, the capital stock determines per capita GDP, or income, via productivity. Thus, output and income should grow faster in countries or regions with smaller capital, i.e., with smaller income. It is worth stressing, however, that in order to satisfy this condition, the neoclassical model needs many collateral qualifications: the most important ones are that all economies must have a similar technology (considered in a broader sense to include taxation, property rights, and other institutional factors) as well as similar savings and population growth rates. These assumptions are anything but realistic in long-run cross-country comparisons. This is not a problem in itself, provided we always remind ourselves to use the models as they should be used: not as something true or false to be verified, in order to corroborate a theory, but as an analytical instrument useful to describe facts in a simplified way. In other words, we must always remind ourselves that theories are confirmed by facts, not by models, and that models rather serve us to draw the contours of the most relevant facts.

Using a Cobb–Douglas form of the production function, following Barro (1991), cross-country growth regressions may be expressed as

$$\gamma_i = \beta \log y_{i,0} + \psi X_i + \pi Z_i + \varepsilon_i \quad (8)$$

where γ_i is the growth rate of an i country, $y_{i,0}$ is its initial level of per capita GDP, X_i represents other growth determinants suggested by the Solow model apart from the initial level of income, and πZ_i represents those determinants not accounted for by the Solow model.

We have unconditional β -convergence (as seen in Fig. 2) when

$$\gamma_i = \beta \log y_{i,0} + \varepsilon_i \quad (9)$$

with the negative sign of the coefficient β .

Otherwise, we do not have *unconditional* convergence. We can still have *conditional* convergence, however, if after adding other variables to Eq. 8, the β coefficient becomes negative (Barro and Sala-i-Martin 1992). The basic idea behind conditional convergence is that differences in per capita incomes are not permanent only because of cross-country structural heterogeneity, that is, because the model does not satisfy collateral qualifications. This can be due to different resource endowments, institutions, and migration rates, as well as to human and social capital disparities, among others things. In the growth regressions, each of these factors can be a conditioning variable, coming either from within the Solow model variable group X_i (i.e., human capital, institutions or social capital, if we consider technology in its broadest sense) or from outside the Solow model from the Z_i variable group (think of climate, but usually variables of this kind are much less common in the literature, while spanning an impressive range of categories). Once we have checked for the effects of structural heterogeneity, there can still be convergence; however, this is not convergence to a single common steady state, but rather the convergence of every country to its own steady state, given its own conditioning variables (i.e., *conditional* convergence). It has been called *convergence*, but truly this model does not measure convergence across regions or countries, since different regions or countries may have different steady states.

A major problem with this framework is the multiplicity of possible regressors, given that the conditioning variables that can be run are practically countless. Durlauf et al. (2005) classified about 150 independent variables used in growth regressions (in almost 300 articles) plus about

100 instrumental variables. In short, the number of possible regressors exceeds the number of cases, thus “rendering the all-inclusive regression computationally impossible” (Sala-i-Martin et al. 2004, p. 814). One reason for the multiplicity problem may lie in the analytical and theoretical weakness of the Cobb–Douglas function, which is valid only in the presence of a vast number of assumptions and has been verified only in a limited number of cases (namely, for the United States in the interwar years). There are two approaches to cope with the multiplicity problem: one is to take advantage of information from qualitative and case study research, while the alternative is to resort to econometrics in order to automatically sort out the irrelevant regressors. Bayesian models, which attach probabilities to each regressor, are an answer to the multiplicity problem safely within the second approach. Among these, the Bayesian Averaging of Classical Estimates (BACE) model, which makes use of the classical ordinary least squares (OLS) estimation, may be the most appealing technique. However, results from BACE models are far from convincing. To date, probably the most comprehensive exercise has been carried out by Sala-i-Martin et al. (2004), who proposed a BACE approach in order to sort 67 explanatory variables in cross-country regressions. Some of their findings look reasonable: for instance, they found primary school enrolment to be the second most important explanatory variable for 1960–1990 GDP growth rates. However, others don’t. According to their model, the most significant explanatory variable was the dummy for East Asian countries. This outcome can be accepted only if we recognize that these regressions indicate a simple correlation; but if we are in search of an explanation (i.e., causation), what the model tells us is that South Korea grew because. . . it was South Korea. And there are more problems with the results from that BACE model. For example, the socialist dummy is not correlated with (negative) growth. While the authors apparently do not note the tautology about the East Asian dummy, in the case of the socialist dummy they discuss the unpersuasive result and specify that it “could be due to the fact that other variables, capturing political or economic instability such as the relative price of investment goods, real exchange rate distortions, the number of years an economy has been open, and life expectancy or regional dummies, capture most of the effect” (Sala-i-Martin et al. 2004, p. 829). Nevertheless, the ultimate determinant of most of these variables, as well as the particular political and economic features of those countries, was the socialist regime and the correlated planned economy: an econometric model concealing this evidence may lead to distorted interpretations of both history and the determinants of economic growth. These examples have been made to illustrate that the first approach must not be overlooked and, indeed, is often preferable. Historical knowledge, sensitivity to case studies, and country-specific characteristics should serve as a compass in order to choose among conditioning variables as well as be seen as an indispensable complement to any econometric analysis.

There remains the possibility that countries do not converge because initial conditions determine different outcomes in the long run. That is, the hypothesis that there are no decreasing returns to capital, for example, because the production function is not of a Cobb–Douglas form. A simple linear technology AK , instead of the neoclassical technology $Af(k)$, would transform equation Eq. 7 into

$$\gamma_k = s^*A - (\delta + n) \quad (10)$$

where the savings curve is no longer downward sloping, but a horizontal line, just like the depreciation curve. Thus, two economies with different initial capital stocks would not converge even with all other conditions being equal. If technology or other parameters differ as well, these economies could still converge, but indeed they could also further diverge. They would converge if A or s are systematically higher in the poorer economy, if the depreciation line is systematically

lower, or if other determinants of growth not included in the model are systematically higher as well; however, to quote Sala-i-Martin (1996, p. 1344), “there is no a priori reason why this should be the case.” On the contrary, there is evidence that the savings curve is not even horizontal, but upward sloping. For example, because of economies of scale, increasing returns to capital have frequently been called into question to account for the rise in the United States during the second half of the nineteenth century or the rise of China in recent decades.

With the hypothesis of increasing returns to capital, we have entered the field of cumulative approaches. Following Myrdal (1957), this approach claims that growth is a spatially cumulative process that requires a minimum threshold of resources in order to start and thus may indeed increase cross-country disparities. Different schools refer to cumulative approaches. Among those worth mentioning are endogenous growth models (Romer 1986) that can still be regarded as a derivation from the neoclassical approach and link economic growth to levels of human capital. Also important is new economic geography (NEG) (Krugman 1991), where the key determinants are either the economies of agglomeration (divergence) or the costs of congestion (convergence), and thus the size of the market plays a central role.

In practice, it is not easy to distinguish the increasing returns of endogenous growth models from the lack of collateral conditions of traditional (exogenous) neoclassical models. When there is no convergence, it may be difficult to conclude whether the traditional neoclassical model can still be valid with some qualifications to be satisfied or, on the contrary, that cumulative endogenous growth should be regarded as more suitable. Moreover, in historical analyses, crucial data, such as estimates of capital, are often lacking or unreliable. Furthermore, the models of increasing returns can easily be extended to predict convergence, such as in Eq. 10 by endogenizing the savings rate on the assumption that it would decrease with higher levels of capital (Sala-i-Martin 1996). Such a hypothesis is not at all implausible: think again of the opposite cases of China and the United States (the latter with higher capital but a lower savings rate). Thus, a unified long-term production function based on increasing returns could still be plausible in the case of convergence. On the other hand, some conditioning variables, such as the stocks of human (or even social) capital, can be seen alternatively as initial conditions in exogenous growth models, such as by decomposing K into physical and human capital (Mankiw et al. 1992).

Attempting to distinguish between the NEG approach, on the one hand, and the two neoclassical approaches, on the other, might be a more fruitful approach. Indeed, in terms of implications, it may be even more appealing. Broadly speaking, NEG models focus on the demand side. The resulting divergence in per capita GDP should be due to differences in “within-sector” productivity, brought about by economies of scale. The other two models, both the exogenous and the endogenous growth versions, are instead based on the supply side, namely, on imbalances in factor endowment. Divergence should refer to the “industry mix” effect, i.e., differences in the allocation of the working force between economic sectors. A simple algebraic calculation that decomposes GDP per capita into the product between GDP per worker (productivity) and workers per capita (employment rate), and then in turn decomposes the growth of productivity into “within-sector” productivity and the “industry mix” effects, may provide us with an (approximate) answer. This is also appealing in terms of interpretation given that, arguably, NEG growth can be explained by forces beyond human control (position, population density, and infrastructures that impact transportation costs, though they are at least in part the result of human decisions) in a larger portion than exogenous or endogenous neoclassical growth, which would typically include human capital, social capital or culture, and institutions as conditioning variables. Caution is warranted once again, since a significant proportion of the results may depend not only on the reliability of the estimates but

also on the level of sectoral decomposition: within-sector productivity differences may be present between single industrial productions, but concealed at the aggregate level.

Besides these empirical difficulties, all three approaches seem to have theoretical limitations. Although the shift from divergence to convergence is usually allowed for and widely accepted, the economics literature has mostly neglected the possibility of a further reversal of fortune, namely, convergence to be followed by divergence again. This is common to all three models: there may be divergence at the beginning, because of either conditional exogenous variables, endogenous differences in factor endowments, or economies of scale, but then at a certain point convergence begins. Because differences in conditioning variables have been removed, factor endowments have converged or congestion costs have exceeded economies of scale. Once progress is at work, it should go on until convergence is achieved. A renowned paper by Robert Lucas (2000) may be taken as paradigmatic of this frame of mind. Lucas argued that sooner or later a country will start industrial development and then converge. The problem is only to establish when, not if. However, once a region has embarked on economic growth, the process of convergence (in the long run) cannot be reversed. Nevertheless, how tenable is this argument? Many examples suggest that convergence may be stopped or even overturned. Take the cases of western Europe and Japan (toward the United States) in the past two decades or of southern Europe (toward northern Europe) in recent years. The reason for the inadequacy of theoretical models can be explained by the fact that they are all static, in the broadest sense of the word: they are all based upon a single production function, which is supposed to be valid throughout the period of analysis. However, in reality – and especially over the long term – the shape of the production function may modify, following, for instance, technological progress. Think of human capital. Primary education was surely a fundamental ingredient of growth in the initial phases of the Industrial Revolution, while higher education may have made the difference in later phases. Other conditioning variables may change, too: natural resources may still have been important in the first Industrial Revolution, as testified by the geographical distribution of industries in nineteenth-century Europe. However, social capital has probably become more important in the current post-Fordist age, as far as it helps reduce transaction costs among a multiplicity of small firms. Some institutions (namely, authoritarian ones) may be effective in promoting growth at their early stages, but not at more advanced ones. Generally speaking, dynamic economics seem to be reconcilable with history better than static ones, since history too is essentially dynamic. But there is little or no use of dynamic models in the long-run analyses of GDP convergence.

A Further Step: From National to Regional Estimates (and Models)

In recent times, the reconstruction of GDP has been extended from the nation state to its regions and provinces. For these cases, the same caveats illustrated for national accounts apply, while methodological problems (and differences) are often even more serious due to the lack of data at the subnational level. A common methodological framework has been proposed and applied to produce comparable regional GDP figures for Europe (Rosés and Wolf 2014). The method elaborates on an idea originally put forward by Geary and Stark (2002): at a sectoral, and hopefully sub-sectoral, level, national GDP is allocated by regional employment; the preliminary results are then corrected through regional nominal wages, which should approximate differences in per worker productivity; then, to have real GDP estimates, nominal figures should finally be rescaled by differences in the cost of living. Such a procedure is based on the assumption that capital gains are distributed along the lines of incomes from labor, namely, that the elasticity of substitution between capital and labor is

equal to one. Moreover, the method is all the more effective the higher the degree of sector decomposition. For the reasons exposed in previous sections (namely, the Gerschenkron effect), the national GDP to be allocated should be at current, rather than constant, prices. Another issue is that detailed figures on regional employment before World War II are available only from official censuses, which are usually taken at 10-year intervals. As a consequence, the production of regional GDP series is almost impossible. And even so, for some important sectors, census data may be misleading. For example, agricultural production may significantly vary from one year to the next, especially at the local level, without significant changes in the official labor force. At least in the primary sector, direct estimates (which are not impossible to find, even at the subnational level)²⁵ should be preferred.

The analytical tools are also similar to those briefly examined in the previous section, with only a few differences. First, at the subnational level, techniques based on benchmark estimates are de facto the only ones utilizable, at least for international comparisons. Since the subnational series of GDP for periods before World War II are often a product of interpolation,²⁶ with some possible exceptions at the sectoral level,²⁷ time series econometrics should be avoided. Second, when we measure σ -convergence, it may be useful to weight the regions with their population.²⁸ As long as we are interested in discussing the performance of national economic policies, we may treat different countries as statistical units with the same weights (thus giving the same importance to each national policy) and, at the same time, treat regions within a country with different real weights (thus measuring the overall dispersion of income within a national polity).

In addition, we can compare regions using an extension of the intercountry comparison models, with only a few qualifications. From a neoclassical perspective, the search for convergence within nation states should be simplified by the fact that here structural heterogeneity plays a minor role, given the usually common macroeconomic and institutional context. In fact, neoclassical scholars tend to be more optimistic about regional convergence than they are about convergence at the national level. For instance, Sala-i-Martin (1996) investigated unconditional β -convergence by applying the Solow–Swan growth model in five large European countries (Germany, the United Kingdom, France, Italy, Spain), plus Canada, Japan, and the United States, mostly for the years running from 1950 to 1990,²⁹ and found a similar rate of convergence, around 2 % per year. Of course, this may not always be the case: at times structural heterogeneity can be hard to overcome, even within nation states, as is arguably the case for Italy, whose regional rate of β -convergence in the long run (1871–2001) has been found to be lower, barely 1 % (Felice 2014);³⁰ on the other hand,

²⁵See Federico (2003) for Italy.

²⁶See, for example, the regional series for Italy estimated by Daniele and Malanima (2007), which have been produced by interpolating through the available regional benchmarks the national cycles of agriculture, industry, and services.

²⁷For instance, the industrial production of Italy in the liberal age (1861–1913) (e.g., Ciccarelli and Fenoaltea 2009, forthcoming). In fact, time series techniques have been applied to the Italian regional construction movements during the liberal age (Ciccarelli et al. 2010). Even in this case, however, it must be pointed out that although the regional series by Ciccarelli and Fenoaltea running from 1861 to 1913 are indeed very accurate, they are estimated at constant 1911 prices, with possible distortions in interregional comparisons for the early years.

²⁸Different population-weighted standard deviation measures are available and can be used, from the Williamson (1965) to the Theil (1967) index.

²⁹Data for the United States run from 1880 to 1990, those for Canada from 1961 to 1991, and those for Spain from 1955 to 1987; data for Japan start in 1955.

³⁰The results from panel models, for the years 1891–2001, are even lower: 0.5 % (random effects GLS regression) (Felice 2011). The growth rate of convergence increases to 2 % only once fixed effects are considered, that is, when we pass from unconditional to conditional convergence (Felice 2012).

the forces of NEG should work better within a nation state provided there are no institutional barriers. The neoclassical approach of equalization in factor endowments and the increasing returns of the NEG have been both tested and compared for the Spanish regions by decomposing historical estimates (1860–1930) of regional per capita GDP in productivity and industry mix effects. The results suggest that they somehow reinforced each other, following the model by Epifani (2005), which combines both: from 1860 to 1930, the between-sector component was predominant, but NEG forces were gaining momentum in the last stretch, once industrialization had arrived in a considerable number of regions (Rosés et al. 2010). Furthermore, it should be considered that within nation states there may be regional development policies at work. Thanks to the common institutional framework, these can be more effective than development policies carried out at the international level (for least developed countries), and they may significantly change the pace of convergence, at least in specific periods.³¹

The descriptive model proposed by Williamson (1965) can be used to illustrate the observed patterns at the regional level. This is an extension of the Kuznets model (1955) of the evolution of personal income within a nation state. As for the personal income distribution, the relationship between national income and inequality would take a functional inverted U shape and a subsequent double movement: rising in the first phase, when industrialization begins and tends to concentrate in the strongest areas, then decreasing as industrialization spreads to the rest of the country. Williamson was mainly concerned about industrialization and structural change, and therefore his model focused on the supply side and it is more easily reconcilable with the neoclassical approach (differences in conditioning variables would prevent industrialization spreading until they are removed). However, from a NEG perspective, the pattern would be similar (with rising inequalities due to economies of scale and then decreasing inequality due to congestion costs). There is some confirmation of the Williamson inverted U shape for the United States. The estimates suggest divergence between the late nineteenth and early twentieth century, as industrialization increased in the northeast and spread mainly to the northern and central regions. In the second half of the twentieth century, the southern and western states industrialized as well and thus converged (Kim 1998). When looking at Europe, we have confirmation for the Spain case, with divergence from 1860 to 1920 and then convergence from 1920 to 1980 (Martínez-Galarraga et al. 2014). For Italy, however, the inverted U shape is observed in the Center-North, but not when the southern regions are included (Felice 2014). Moreover, in many cases, regional convergence seems to have come to a halt in recent decades. This finding suggests that the long-run evolution of regional inequality may follow an N movement (divergence, then convergence, followed again by divergence) (Amos 1988), but on this issue, both empirical investigation and theoretical models have barely begun.

Concluding Remarks

The paper reviewed the most common methods employed to produce historical GDP estimates at the national and the regional levels, and the use of GDP to compare economic performances in the long run. The first point to be highlighted is that GDP estimates, even when relatively sound and well informed, are more suitable for measuring economic performance from the Industrial Revolution onward. GDP was born in the United States in the aftermath of the 1929 crisis, within an empirically oriented environment. It was designed for industrial advanced economies and may not correctly

³¹For Italy, the western country where the most impressive regional policy (in terms of expenditures as a share of GDP) was carried out, see again Felice (2010).

approximate material standards of living in preindustrial societies, where most production is from agriculture (for which the amount of land is a fundamental ingredient that GDP does not consider) and a non-negligible proportion is not even exchanged in the market (and thus is not included in GDP accounting). Moreover, for preindustrial societies, we often lack the minimum information required to produce reliable national accounts.

In modern times, when making cross-country comparisons, we should always make sure that the adoption of different estimating methodologies does not significantly affect the results. By themselves, national estimates can be reliable or made so given the available information, but this is not the point. For cross-country (and cross-regional) GDP comparisons, it is crucial that three basic conditions are satisfied. First, the decomposition level of the quantity series must be relatively homogeneous from one country to another. Even more important, and less generally acknowledged, the base year of constant-price series must be relatively close. Third, when considering real GDP figures, the PPPs used to compare countries must be as close as possible to the period of concern: as a consequence, the renowned Maddison estimates at 1990 PPP international dollars may be not reliable for years before World War II, as illustrated by a contrast with the alternative PPPs proposed by Prados de la Escosura (2000).

Convergence tests may be significantly affected by the cumulative effect of these distortions. The way estimates are constructed also impacts upon the models used to interpret and describe the results. For instance, in international and (even more so) interregional comparisons, cross-sectional techniques are preferable to time series analysis, because the former are less data demanding even though they may be less informative. Provided we have reliable estimates, decomposing GDP growth into productivity and industry mix effects may yield important clues for distinguishing between the role of factor endowments and structural heterogeneity, on the one side, and market access, on the other. However, such clues should always be handled with care, for example, by searching for confirmation in the patterns of individual countries or regions. It also needs to be stressed that given the quality of the data, convergence models based on conditioning variables as well as more statistically refined ones such as the BACE techniques can be trustworthy only up to a certain point. They should always be supplemented by sound historical information, including qualitative sources and case studies, which should also help sort among the best conditioning variables to be tested, given the multiplicity of possible predictors.

In short, cliometricians should make an effort not to rely exclusively on statistical tools when searching for the determinants of growth, but to complement them with historical expertise. They should also have a broad view of the available models, from exogenous to endogenous growth to NEG (and others that may or may not combine ideas from the three we have outlined), and be flexible enough to adapt both the models and the statistical techniques to the different historical settings and to the quality of their data.

References

- Ahmad S (1988) International real income comparisons with reduced information. In: Salazar-Carrillo J, Prasada Rao DS (eds) *World comparisons of incomes, prices, and product*. North-Holland, Amsterdam, pp 75–92
- Amos OM Jr (1988) Unbalanced regional growth and regional income inequality in the latter stages of development. *Reg Sci Urban Econ* 18(4):549–566
- Baffigi A (2013) National accounts, 1861–2011. In: Toniolo G (ed) *The Oxford handbook of the Italian economy since unification*. Oxford University Press, Oxford, pp 157–186

- Bairoch P (1976) Europe's gross national product: 1800–1975. *J Eur Econ Hist* 5(2):273–340
- Barro RJ (1991) Economic growth in a cross section of countries. *Q J Econ* 106(2):407–443
- Barro RJ, Sala-i-Martin X (1991) Convergence across states and regions. *Brook Pap Econ Act* 1:107–182
- Barro RJ, Sala-i-Martin X (1992) Convergence. *J Polit Econ* 100(2):223–251
- Bolt J, van Zanden JL (2014) The Maddison Project: collaborative research on historical national accounts. *Econ Hist Rev* 67. doi: 10.1111/1468-0289.12032.
- Boyd J, Banzhaf S (2007) What are ecosystem services? The need for standardized environmental accounting units. *Ecol Econ* 63(2–3):616–626
- Carson CS (1975) The history of the United States national income and product accounts: the development of an analytical tool. *J Income Wealth* 21(2):153–181
- Ciccarelli C, Fenoaltea S (2009) La produzione industriale delle regioni d'Italia, 1861–1913: una ricostruzione quantitativa. 1. Le industrie non manifatturiere. Banca d'Italia, Roma
- Ciccarelli C, Fenoaltea S (2014) La produzione industriale delle regioni d'Italia, 1861–1913: una ricostruzione quantitativa. 2. Le industrie estrattivo-manifatturiere. Banca d'Italia, Roma
- Ciccarelli C, Fenoaltea S, Proietti T (2010) The effects of unification: markets, policy, and cyclical convergence in Italy, 1861–1913. *Cliometrica* 4(3):269–292
- Clark G (2009) Review essay: Angus Maddison, contours of the world economy, 1–2030 AD: essays in macro-economic history. *J Econ Hist* 69(4):1156–1161
- Crafts NFR (1985) British economic growth during the industrial revolution. Cambridge University Press, Cambridge
- Crafts NFR (1997) The human development index and changes in standards of living: some historical comparisons. *Eur Rev Econ Hist* 1(3):299–322
- Crafts NFR (2002) The human development index, 1870–1999: some revised estimates. *Eur Rev Econ Hist* 6(3):395–405
- Daniele V, Malanima P (2007) Il prodotto delle regioni e il divario Nord-Sud in Italia (1861–2004). *Rivista di Politica Economica* 67(3–4):267–315
- Douglas PH (1976) The Cobb-Douglas production function once again: its history, its testing, and some new empirical values. *J Polit Econ* 84(5):903–916
- Durlauf SN, Johnson PA, Temple JRW (2005) Growth econometrics. In: Aghion P, Durlauf SN (eds) *Handbook of economic growth*, vol 1A. Elsevier, Amsterdam, pp 555–677
- Eichengreen B (1986) What have we learned from historical comparisons of income and productivity? In: O'Brien P (ed) *International productivity comparisons and problems of measurement, 1750–1939*. 9th international economic history congress, Session B6, Bern, pp 26–35
- Epifani P (2005) Heckscher–Ohlin and agglomeration. *Reg Sci Urban Econ* 35(6):645–657
- Federico G (2003) Le nuove stime della produzione agricola italiana, 1860–1910: primi risultati e implicazioni. *Rivista di Storia Economica* 19(3):359–381
- Feinstein CH (1972) National income, expenditure and output for the United Kingdom, 1855–1965. Cambridge University Press, Cambridge
- Feinstein CH, Thomas M (2002) Making history count. A primer in quantitative methods for historians. Cambridge University Press, Cambridge
- Felice E (2010) Regional development: reviewing the Italian mosaic. *J Mod Ital Stud* 15(1):64–80
- Felice E (2011) Regional value added in Italy, 1891–2001, and the foundation of a long-term picture. *Econ Hist Rev* 64(3):929–950
- Felice E (2012) Regional convergence in Italy (1891–2001): testing human and social capital. *Cliometrica* 6(3):267–306

- Felice E (2014) Regional income inequality in Italy over the long-run (1871–2001). Patterns and determinants. In: Rosés JR, Wolf N (eds) *Europe's regions, 1900–2010. A new quantitative history of the economic development of Europe*. Routledge, New York
- Felice E, Carreras A (2012) When did modernization begin? Italy's industrial growth reconsidered in light of new value-added series, 1911–1951. *Explor Econ Hist* 49(4):443–460
- Felice E, Vecchi G (2013) Italy's growth and decline, 1861–2011. *CEIS Tor Vergata. Res Pap Ser* 11(13):293
- Fenoaltea S (1976) Real value added and the measurement of industrial production. *Ann Econ Soc Meas* 5(1):113–139
- Fenoaltea S (2003) Notes on the rate of industrial growth in Italy, 1861–1913. *J Econ Hist* 63(3):695–735
- Fenoaltea S (2008) A proposito del PIL. *Italianieuropei* 8(1):165–169
- Fenoaltea S (2010) The reconstruction of historical national accounts: the case of Italy. *PSL Q Rev* 63(252):77–96
- Ferreira S, Hamilton K, Vincent J (2008) Comprehensive wealth and future consumption: accounting for population growth. *World Bank Econ Rev* 22(2):233–248
- Geary F, Stark T (2002) Examining Ireland's post-famine economic growth performance. *Econ J* 112:919–935
- Gerschenkron A (1947) The Soviet indices of industrial production. *Rev Econ Stat* 29(4):217–226
- Good D, Ma T (1999) The economic growth of central and eastern Europe in comparative perspective, 1870–1989. *Eur Rev Econ Hist* 3(2):103–137
- Istat (1957) Indagine statistica sullo sviluppo del reddito nazionale dell'Italia dal 1861 al 1956. *Annali di statistica* 8(9):1–271
- Kim S (1998) Economic integration and convergence: U.S. regions, 1840–1987. *J Econ Hist* 58(3):659–683
- Kravis IB, Heston A, Summers R (1978) Real per capita income for more than one hundred countries. *Econ J* 88:215–242
- Krugman P (1991) Increasing returns and economic geography. *J Polit Econ* 99(3):483–499
- Kuznets S (1934) *National income 1929–1932*. National Bureau of Economic Research, New York
- Kuznets S (1955) Economic growth and income inequality. *Am Econ Rev* 45(1):1–28
- Lequiller F, Blades D (2006) *Understanding national accounts*. OECD, Paris
- Lucas R (2000) Some macroeconomics for the 21st century. *J Econ Perspect* 14(1):159–168
- Maddison A (1991) A revised estimate of Italian economic growth, 1861–1989. *Banca Nazionale del Lavoro Q Rev* 44(177):225–241
- Maddison A (1994) Confessions of a chiffrephile. *Banca Nazionale del Lavoro Q Rev* 47(189):123–165
- Maddison A (1995) *Monitoring the world economy 1820–1992*. Development Centre Studies, OECD, Paris
- Maddison A (2001) *The world economy: a millennial perspective*. Development Centre Studies, OECD, Paris
- Maddison A (2006) *The world economy. A millennial perspective and volume II: historical statistics, vol I*. Development Centre Studies, OECD, Paris
- Maddison A (2007) *Contours of the world economy I-2030 AD*. Oxford University Press, Oxford
- Malanima P (2006) Alle origini della crescita in Italia 1820–1913. *Rivista di Storia Economica* 22(3):306–330
- Malanima P (2011) The long decline of a leading economy: GDP in central and northern Italy, 1300–1913. *Eur Rev Econ Hist* 15(2):169–219

- Mankiw NG, Romer D, Weil DN (1992) A contribution to the empirics of economic growth. *Q J Econ* 107(2):407–437
- Martínez-Galarraga J, Rosés JR, Tirado D (2014) The evolution of regional income inequality in Spain, 1860–2000. In: Rosés JR, Wolf N (eds) *Europe's regions, 1900–2010. A new quantitative history of the economic development of Europe*. Routledge, New York
- Myrdal G (1957) *Economic theory and underdeveloped regions*. Hutchinson, London
- Nussbaum M (2000) *Women and human development: the capabilities approach*. Cambridge University Press, Cambridge
- Prados de la Escosura L (2000) International comparisons of real product, 1820–1990: an alternative data set. *Explor Econ Hist* 37(1):1–41
- Prados de la Escosura L (2003) *El progreso económico de España (1850–2000)*. Fundación BBVA, Bilbao
- Prados de la Escosura L (2007) When did Latin America fall behind? In: Edwards S, Esquivel G, Márquez G (eds) *The decline of Latin American economies: growth, institutions, and crises*. University of Chicago Press, Chicago
- Prados de la Escosura L (2010) Improving human development: a long-run view. *J Econ Surv* 24(5):841–894
- Prados de la Escosura L (2013a) Human development in Africa: a long-run perspective. *Explor Econ Hist* 50(2):179–204
- Prados de la Escosura L (2013b) World human development, 1870–2007. Working papers in economic history 13-01, Universidad Carlos III
- Ravallion M (2012a) Mashup indices of development. *World Bank Res Obs* 27(1):1–32
- Ravallion M (2012b) Troubling tradeoffs in the human development index. *J Dev Econ* 99(2):201–209
- Romer P (1986) Increasing returns and long run growth. *J Polit Econ* 94(5):1002–1037
- Rosés JR, Wolf N (eds) (2014) *Europe's regions, 1900–2010. A new quantitative history of the economic development of Europe*. Routledge, New York
- Rosés JR, Martínez-Galarraga J, Tirado DA (2010) The upswing of regional income inequality in Spain (1860–1930). *Explor Econ Hist* 47(2):244–257
- Sala-i-Martin X (1996) Regional cohesion: evidence and theories of regional growth and convergence. *Eur Econ Rev* 40(6):1325–1352
- Sala-i-Martin X, Doppelhofer G, Miller RI (2004) Determinants of long term growth: a Bayesian averaging of classical estimates (BACE) approach. *Am Econ Rev* 94(4):813–835
- Schumpeter JA (1950) Wesley Clair Mitchell (1874–1948). *Q J Econ* 64(1):139–155
- Sen AK (1985) *Commodities and capabilities*. Oxford University Press, Oxford
- Solow RM (1956) A contribution to the theory of economic growth. *Q J Econ* 70(1):65–94
- Solow RM (1957) Technical change and the aggregate production function. *Rev Econ Stat* 39(3):312–320
- Stone R (1956) *Quantity and price indexes in national accounts*. OEEC, Paris
- Stone R (1961) *Input–output and national accounts*. OEEC, Paris
- Swan T (1956) Economic growth and capital accumulation. *Econ Rec* 32(2):334–361
- Theil H (1967) *Economics and information theory*. North Holland, Amsterdam
- Waring M (1988) *If women counted. A new feminist economics*. Macmillan, London
- Williamson JG (1965) Regional inequality and the process of national development: a description of the pattern. *Econ Dev Cult Change* 13(4):3–84
- Williamson JG (2011) Industrial catching up in the poor periphery 1870–1975. CEPR discussion paper no 8335