Reconstructing The Past: The Measurement Of Aggregate Product

Fenoaltea, Stefano

January 2020

RECONSTRUCTING THE PAST:
THE MEASUREMENT OF AGGREGATE PRODUCT

Stefano Fenoaltea*
Fellow, Fondazione Luigi Einaudi, Turin

stefano.fenoaltea@uniroma2.it

January 2020

ABSTRACT

This paper summarizes the author's considerations on the measurement of a national economy's historical aggregate product. Given the sources from which we can start, and the objective we wish to reach, the proper path follows logically; the extant corpus of historical national accounts seems to follow a very different path, uninformed by due reflection.

Keywords: production, measurement, historical national accounts

JEL codes: C13, E01, N01

*The author thanks Brian A'Hearn, Alberto Baffigi, Alfredo Gigliobianco, Christopher Hanes, and Giovanni Vecchi for their often extended comments on an early draft; the usual disclaimer applies. This paper revises and extends earlier work, notably. Fenoaltea (2010) and (2019a), without further referencing the individual *reprises*.

RECONSTRUCTING THE PAST:
THE MEASUREMENT OF AGGREGATE PRODUCT

Peccata nominabo, peccatores rare.

1. ECONOMICS, ECONOMIC HISTORY, AND THE "DATA"

1.1 The scientization of economics

This paper presents a *discours de la méthode*; and it begins perforce with an intellectual history.

The discipline we call economics was born in Europe. By the early twentieth century economics was felt to comprise three essential strands, indissolubly linked: economic theory, economic history, and the history of economic thought. These were the core disciplines in the university training of economists, as it was then structured; a Luigi Einaudi, an Eli Heckscher, a Joseph Schumpeter moved seamlessly across all three.

Economics, so structured, so taught, was clearly not considered a "science": a medical doctor, for example, is considered fully trained even with no exposure to the history of medicine (Hippocratic humors) or to the history of medical problems (the Great Plague). Economics studied not the natural world but human behavior and institutions; it was clearly seen as more subtle, more complex, more political, *à tout prendre*, than a science. To use the language of our own day, economics was seen as a non-linear, multi-layered discourse that could not be understood by hearing only its most recent lines. To master economics, the logic ran, one has to understand how and why it has come to contain what it does, understand the problems that engaged past economists, the economic problems of their day, what to us is economic history; and to understand economic history, by the same token, one has to master economics.

The first half of the twentieth century was marked by a number of related changes. First and most broadly, the two World Wars Europe inflicted upon itself stripped it of its former hegemony, which passed to the United States, in cultural matters as in others. American graduate programs in economics used to mimic European models; now the reverse is true.

The Americanization of economics was not without consequence, for America has its quirks. Culture is not there highly regarded, it seems to be considered more a feminine adornment than stuff for real men.[1] America instead worships science, only America could spawn a Church of Scientology. In Europe the élite schools educated, the bright kids did classics, the scientific program was for second-raters; in America schools train, the scientific program carries the highest prestige.

We all value prestige; American economics, *quā* American, morphed into a "social science" (a poor oxymoron to anyone with any *nous*, but now, like the dollar, international

[1] When my then Department chairman at Duke University described his family's European visit he noted that they had abandoned the slow-moving guided tours, and boasted "We had done four cathedrals before breakfast!"

currency). The triad that composed economics was quickly dissolved: economic history and the history of economic thought were demoted from the core curriculum to fringe specializations, the histories of economic thought recast, impoverished, as the development of context-free abstract reasoning.[2] The discipline attributed to itself the cumulative nature of the (true) sciences: leading economists have been heard to brag that they have their students read nothing over three years old, in the early 1960s I myself heard Paul Samuelson declare that "any graduate student today is a better economist than Keynes." Economics became culture-less, and lost its sense of what it is.[3]

Between the wars, too, the developed world suffered the Great Depression: a challenge to the orthodox economic theory of the day, which could not begin to explain it. The well-known outcome was of course the Keynesian revolution, the commitment to maintain full employment (even in peacetime), the development of "national accounts" (and their continuous updating by the appropriate public Bureau) to diagnose the path of the economy and inform anticyclical intervention; on all this more below. A possible side effect was that the employment problem captured the attention of the profession, which with bigger fish to fry set economic history aside; and the void was filled by historians with no grasp of economics at all, even the best of whom penned interpretations that stood on end the hair of any economist who was not irretrievably ba(l)d. Whatever the reason, between the wars economic history became ever more history, ever less economics; the throwaway distinction between "economic historians" and "competent economists" by Lionel Robbins (1939, p. 9) was harsh but not unjustified.

That of course set the stage for the *reconquista*, the recovery of economic history by competent economists, the recovery specifically of the principle that economic historians must use the tools provided by up-to-date economic theory (Einaudi 1936a, p. 158). As we know, it happened first in the United States (and struggled to cross the Atlantic, overcoming only with difficulty the entrenched opposition of Departments of economic history, non-existent or virtually so in North America). It need not have – European economists could well have revived their own earlier tradition – but it did; and the "new" economic history that emerged owed and owes much to postwar American economics.

With a proto-Trumpian sense of subtlety, culture, and decorum the "new" economic history trumpeted its innovative "scientific" method, the construction of models and explicit counterfacts, the testing of hypotheses against the data. It provoked initial opposition, much

---

[2] To my mind one cannot understand the theory of international trade if one does not understand Ricardo. Ricardo is now portrayed as "the first abstract economist"; I see him as a thinker fully engaged in the political struggles and economic controversies of his day, and his *Principles* as a tract against the Corn Laws and the landed aristocracy that imposed them. His "model of trade" attacks tariffs in general with a carpet of dumb bombs; the smart bomb precision-guided to his target was his "model of growth" (a stealth characterization, to push the metaphor); see Fenoaltea 2011a, pp. 152–165, 252–257, 265–272.

[3] A telling episode is the discipline's reaction to the "prisoner's dilemma," greeted as a novel challenge to our faith in the "invisible hand." As far as I could tell from the early literature nobody pointed out that economists alone were ignorant of the issue, and of the solution (norms, ethical or legal, in the specific case at hand the "honor code" of *omertà*, and the certainty that if you confess to anything at all your family will be exterminated). The lesson of the "prisoner's dilemma" was common knowledge, and considered generally valid, from time immemorial; economics traces its tap-root to the physiocratic intuition that in certain ("economic") walks of life self-interested behavior need *not* be collectively maleficent (at least under certain conditions that have since been explored, e.g., free and informed consent, sufficient competition, ethical restraint on the abuse of asymmetric information, *und so weiter*), and that regulation is then unnecessary if not harmful.

of it directed at Robert Fogel's startling (and clearly wrong) claim that the railroads' contribution to America's economic growth was "not important"; but it was ineffective opposition, the obtuse complaint that counterfactuals were unhistorical (as if any statement that *a* affected *b* does not rely on one, at least implicitly), that economists' models were "too simple" (as if simplicity were not in itself a virtue that commands assent, and *excessive* simplicity were not to be determined case-by-case, depending on the model's use and purpose).[4] Brash enthusiasm met incompetence, a swift triumph was foregone.

I was in the United States, working on a doctorate in economics, when the "new" economic history burst upon the scene; I was present at the creation, I was born, as it were, a "new" economic historian. With my European background (and the cultural snobbery of the classically educated) I would as lief be called a hog farmer as a scientist; nor do I set particular store by quantification, let alone the econometric "testing" of hypotheses, called for in some cases but not in others.[5] The "new" economic history soon earned the sobriquet "cliometrics," but that was merely a joke that stuck, a name accepted perhaps because it played on the prestige of *Econometrica*, but not meant to capture the essence of the discipline. That essence, to my mind, is the return to economic history of competence in economics, and that alone. I do not believe much, but I do believe that without such competence one cannot hope to understand the economic past – just as one cannot hope to understand the past history of disease, say, if one is not versed in epidemiology, no more, no less. I see economic theory as the mistress of the cliometric household, quantification and econometrics as mere servants, to be summoned as needed.[6]

---

[4] The *human* mind, every last one of us, is convinced by the simplest possible explanation: think of the fellow who discards the complex set of hypotheses that supposedly explain his wife's repeated tardiness in returning home from work in favor of the single, powerful hypothesis that she is seeing somebody else. The complaint of unrealistic oversimplification was made *ad personas*, nobody extended it to the astronomers who model our entire world as a mere point with mass. Fogel played on the confusion of importance with necessity rather than sufficiency, arguing that Brutus (Cassius) was "not important" to Caesar's death because Cassius (Brutus) would have killed him anyway. He pretended to calculate the railways' marginal product, the extra output ("social saving") they allowed, as an economist would; but when we measure a factor's marginal product we do *not* contemplate substitution, if we did the marginal product of (homogeneous) labor would be identically zero (see Fenoaltea 2011a, pp. 168–169, and references therein). The early confusion between the differential and scarcity components of the rent of land comes to mind.

[5] Testing hypotheses is like skinning cats, the econometric way is not the only one. The hypothesis that gravity bends light was verified by taking advantage of a solar eclipse to observe a star (angularly near the sun) where we knew it wasn't: the theory correctly predicted that what we observed was not a fact at all. My own model of slave management correctly predicted that the observations in the secondary sources which contradicted it misrepresented the facts (Fenoaltea 1984, footnote 40).

[6] This was the shared view of the cliometric pioneers: at the 1985 meetings of the (American) Economic History Association, by then dominated by its cliometric Young Turks, the going prizes were swept by a paper that contained no algebra, no tables, and no econometrics, but for all that fully satisfied its cliometric judges because it developed an analysis only an economist could have authored. But we armed our children with econometric packages and let them loose upon the world, rather like Africa's boy soldiers; we may have done no more than ferry economic history from serious historians who were inadequate economists to serious econometricians who are inadequate historians and, in much of the work they do, sociologists/political scientists/whatever as embarrassingly amateurish as the "old" economic historians were economists. In a way, in fact, the discipline has come full circle: where ours was proudly problem-driven, that of our young is as source-driven as the historians' "old" economic history used to be.

1.2 Economics, measurement, and "data"

For all that, of course, the application of economics to real-world problems remains a naturally quantitative exercise, never more so than in the present context, the reconstruction of past economic growth. A hundred years ago, the aspiration of the profession to empirical relevance led to the creation, at Harvard, of *The Review of Economic Statistics*. That title contained a research program: we economists were asking ourselves what measures we wished to have, and how to construct them. Within thirty years that search was abandoned, that flag hauled down, that journal blandly renamed ("*The Review of Economics and Statistics*"): in the wake of the Keynesian revolution national and supranational bureaucracies took on the task of documenting the economy, and the profession sat back, happy enough to receive its data from an external source. At that very time (American) economics was redefining itself as a science; most sciences jealously reserve unto themselves the generation of their empirical evidence, but two do not. One is economics, the other, astrology.

The upshot was that "scientific" economics developed a schizoid attitude towards the empirical evidence. Much attention was paid to the quality of data-*processing*, to the exploration of the relationships among the numbers: econometric technique became ever more sophisticated, improving by leaps and bounds. Vanishing attention was instead paid to the quality of the data themselves. Data were simply taken as *data*, given (typically by public Bureaus), to be retrieved and uploaded (or, increasingly and even more simply, downloaded); "measurement" was viewed as no more than data *retrieval*, an activity perhaps time-consuming but quintessentially unskilled, suitably assigned to graduate students or farmed out to research assistants. We skilled economists would respect data *analysis* enough to do it ourselves, mere "data," "measurement" we would hold in contempt.

It is tempting here to return to an earlier mode of thought, to attribute personality to data and measurement, to say they resented that contempt; and that they exacted their revenge. At one point economists analyzed macroeconomic time series for the American economy going back into the nineteenth century, and found that the economy was far less stable then than it had since become. Numerous articles on "the stabilization of the American economy" quickly appeared in the leading journals; but that strand of the literature came to a sudden end. Christina Romer (1986), in the *American Economic Review*, informed the profession at large of what was, among the pioneering cliometricians, common knowledge: that the estimates for the more distant past were not comparable to the data for more recent times, that they could perhaps illuminate the longer-term movements of the American economy but certainly not its short-term variability. As Romer pointed out that early variability was much overstated, the economy's much-discussed "stabilization" never happened at all: it was "a figment of the data," data no one had bothered to validate before putting them to use.

I do not of course read astronomy journals, but somehow I cannot imagine a scholarly discussion of a newly discovered galaxy being terminated by a piece in the *Annual Review of Astronomy and Astrophysics* essentially saying, "uh, guys, hold up: we checked the original photograph, and what looked like an undiscovered galaxy was actually a reflection of the light on the newspaper boy's bicycle." In fact, nobody seems to read the journals of alien disciplines, and the economics profession washed its dirty linen in reasonable privacy; and that may sufficiently explain why economists were not laughed out of academia then and there.

But what is less easy to understand is the discipline's own reaction to what should have been a sharp lesson: a total *lack* of reaction. The discipline was entirely unfazed, it did not plunge into introspection (as anthropology did, with a different but equally damning trigger, as succinctly recalled in Fenoaltea 2019a); the general attitude was "another day, another dollar, steady as she goes, nothing to see here, move along." Nothing happened;

some twenty years on Richard Easterlin (2004) would note, and eloquently lament, the low esteem in which economists (continue to) hold measurement, the reconstruction of the facts.

Economics – American, Americanized "scientific" economics – retains unchanged its cavalier attitude towards the evidence; and "scientific" economic history, *la cliométrie sa fille*, inherited that attitude in its genes. For all their emphasis on quantification, cliometricians' historical measurement is in the main unskilled work, poor work simply because it is not informed by due reflection on method; and that brings us back to methodology, to the very logic of our reconstructions of past aggregate product.

1.3 On method: the economic historian's craft

The historian's craft may overreach, *this* historian's craft may be the most I can attempt to describe. In courts of law, I gather, witnesses' credentials are examined before their testimony is taken; so let me introduce my professional persona.

Economists, American and Americanized, are in the mass still wedded to naïve nineteenth-century positivism: cut off from the cultural mainstream, a lost tribe in the intellectual jungle (Fenoaltea 2019a). I am in contrast a deep skeptic, unsure of the ultimate reality of my own person (Descartes be damned), let alone that of "the real world."[7] But even apart from that I am thoroughly convinced that what we call "facts," (supposedly) observable bits of reality, do not exist at all: even granting an external reality, what we can observe is not a fact. We both saw a chicken cross the road, did it not do so, is that not a fact? I would answer no, our eyes capture only patterns of light and color: we "see" "a chicken" "cross" "a road," every word in quotes is not an observed object or action but a construct, an *interpretation* of what our eyes actually perceive. What we call *facts* are no more than agreed-upon, shared interpretations: perhaps "real" enough for practical purposes, but only within a shared set of expectations, *culture-bound* expectations. "Facts," and a fortiori *evidence* of facts, are very slippery stuff.[8]

Past facts are a will o' the wisp, I cannot write history *wie es eigentlich gewesen*. All I can do is tell a story, whether in numbers or in words matters little; what matters is that I want to tell a *good* story. In Italian, as in French and German, the distinction between history and literature is merely that between *la storia* (*the* story, history) and *una storia* (*a* story, literature). Literature does not care to be "true" (and allows limited suspensions of disbelief, as with talking animals in fables or satire), history would like to be "true" but cannot be; neither is *vera*, but both can be, and to be any good must be, believable, *verisimilar*. My (hi)story will be good if it *rings* true (at least to those with a mind-set similar to mine), no more can be asked of it; the historian's craft is worlds away from the scientist's, "scientific history" is an oxymoron that proclaims, again, a lack of education, a lack of (contemporary) culture.

---

[7] I do know that the possibly imaginary world I perceive is characterized by consistency, by predictability (the bedside book that was there when I fell asleep was again there when I woke up, *les autres* that appear to me maintain their characteristics from appearance to appearance); and so I behave as if my world were real enough for practical purposes, as if there were such a thing as *the* world. If observed, I presumably appear, so to speak, perfectly normal: my skepticism would seem to be a private matter, a hidden hairshirt.

[8] That someone who doesn't believe in facts should turn to writing *history* may seem odd, but in my case the explanation is simple enough. I see myself first and foremost as a writer, but one denied the wit and skills to create literature, unable to invent a world out of whole cloth; forced by my own limitations to work on a story line that I am somehow given, I take my story line from the past.

For a (hi)story to "ring true" it must not clash with our strong beliefs; and I have two. I believe in the power of human logic, in our capacity for error-free deduction (Descartes be blessed); and I believe in the human aversion to pointless toil, to "the sweat of our brows." The one allows us to develop the ineluctable implications of the other: the discipline that does so, that investigates what we call "rational choice" (as if it could be anything else) is what we call "economics."[9] And that is why I see the recovery of competence in economic theory as the defining feature of the "cliometric" approach, why I see myself as a cliometrician, whether or not the problem I am addressing requires quantification; why I find no value in the "economic histories" that make no *economic* sense, the stories that are simply *not* possible if we believe, as I do, in the validity of human logic, and in the human aversion to pointless toil.

To ring true as *economic* history, in short, my story must be good economics. To ring true as economic *history*, quite analogously, it must be good history, it must sit well with "the facts." As noted, however, our cliometric *vulgata* glorifies "interpretation," the elucidation of the relationships among the facts, and reduces the elucidation of the facts themselves to "measurement," seen as the simple process of setting a yardstick next to an observed object, seen as only a simple-minded nineteenth-century positivist (or contemporary economist) could see it.

Whether or not we can "observe" the present, we certainly cannot observe the past, for it is *gone*; all we can observe are the traces it left behind.[10] The quantitative traces ("data") of particular interest to us are themselves not ("scientific") observations, but constructs – typically byproducts of other, and usually fiscal, public concerns – that must be deconstructed if their relation to the "facts" that interest us is to be understood at all. We cannot observe past objects, our sources reveal only their shadows, to reconstruct a shape from its shadows we must pinpoint the sources of light; to understand the data in the sources we must understand how, by whom, and to what purpose they were produced, we must learn to read our quantitative sources, or rather to read *through* them. This is the historian's stock-in-trade, but not even a glimmer in today's economists' mind's eye. That our quantitative sources present numbers clearly labeled in our native language, numbers that positively invite us to take them at face value, is a trap, a trap for beginners, a trap for *naïfs*: a trap a trained economic *historian* should readily recognize and avoid, a trap that should never have become a trap for our entire profession.[11]

---

[9] It follows that economics is relevant wherever choices have to be made, if only because the day is not infinitely long: those tiresome critics who consider our discipline relevant only to modern market economies utterly misapprehend it.

[10] I am reminded here of a book I had, that taught the reader to recognize the presence of unseen wildlife from their droppings. The title that sticks in my mind sounds like *Birds of North America*, but I know it wasn't that.

[11] Thus for example the long literature that takes Italy's industrial employment in 1911 directly from the "employment" data in that year's industrial census, never cottoning to what that census actually contains; see Fenoaltea (2015a, 2016) and references therein. The root problem is of course the profession's belief that "data" are merely to be collected, the lack of any sense that proper "measurement" requires that the sources be approached with suitable training, the sort of training doctoral programs in history provide as a matter of course. To the best of my knowledge the only graduate students in economics ever exposed to a formal course in "reconstructing the past" were those that took my course by that name at the Collegio Carlo Alberto in 2014.

What we call "measurement" is in fact a work of interpretation, no less than what we call "interpretation"; we quantitative historians cannot measure the past, we must interpret the sources to reconstruct it. And *this* interpretation is far more difficult than the subsequent "analysis," for three sufficient reasons. One is that it is not directly constrained, and thus facilitated, by codified (economic) theory and (econometric) technique; it is correspondingly the work not of fungible "scientists" but of non-fungible artisans, no two of whom will obtain the very same results from the very same materials.[12] Another is that the sources are opaque, that we learn to understand them only little by little (*poco a poco*, Einaudi 1936b, p. 7): that their hidden defects surface, rather like those of our former spouses, only with extended cohabitation. The third is so to speak that the evidence itself is often not evident. When it is indirect, as it so often is, it takes a good, experienced eye to identify it at all, to grasp its potential significance: it takes a Holmes to see clues where Watson sees nothing, an Indian scout to see tracks where the cowboy sees only dust.

We cannot observe the past, we cannot reconstruct it as it really was; we can only reconstruct it as to our eyes, in the light of everything we know, it most probably was, "it must have been." The obvious parallel is Viollet-le-Duc's "restoration" of Vézeley or Notre-Dame: a sobering parallel, sobering because such reconstruction clearly requires a taxing panoply of skills and breadth of knowledge, sobering too because it remains all the same highly personal, and can readily appear distorted by its author's fantasy.

The bottom line is that economists and cliometricians are right to believe that "interpretation" and "measurement" require different levels of skill and experience, but entirely wrong in their mapping. Our graduate students, our research assistants, well trained in economics and econometrics, are as well-equipped as we their elders for "analysis" and "interpretation"; it is in the creation of the "data," the reconstruction of the past, that they are challenged, challenged by a lack of training in *historical* scholarship, challenged even if so trained by a lack of familiarity with the sources, challenged above all by their lack of experience.[13] Only a contemporary economist, highly trained but badly undereducated, can be so *borné* as to entrust to a graduate student the restoration of a Medieval cathedral, the reconstruction of the past.

I have devoted much of my work on post-Unification Italy to "measurement," to my reconstruction of "the facts" (recounted in the framework of the national accounts, about which more below), as God gave me the light to see those facts.[14] Such reconstruction is not

---

[12] Because the derivation of production series is so personal, too, the constructed "data" can be verified, replicated, and improved only if their derivation from source to final estimate is exhaustively documented. I was taught by Gerschenkron (1962 [1955]) that such documentation was required to meet the minimal standards of serious scholarship; but his lesson was not widely heeded, and most historical reconstructions are accompanied by no more than general descriptions of the underlying sources and algorithms.

[13] Experience is necessary but of course not sufficient: some distinguished colleagues appear to have discovered the fountain of youth, and make beginners' mistakes decade after decade.

[14] It is, obviously, *my* story, *my* reconstruction, derived in Bayesian terms from the surviving data and *my* priors (Fenoaltea 2015b–h, 2019c); the measures I obtain put *my* posterior on the line. My "interpretation" story, my story of how the (apparent) facts (apparently) relate to each other, I have told elsewhere – repeatedly, and very differently, as the evidence accumulated and my understanding evolved (Fenoaltea 1969, 1988/2011a, 2017a).

a science but a craft, an *arte*: an *arte* with its *regole*, its rules of good practice.[15] The ur-rule is of course to tell as verisimilar a story as one can, a story verisimilar in its description of human behavior, and verisimilar too in its interpretation of the surviving sources, sources that are never complete and not infrequently inconsistent: in practice we maximize our story's overall plausibility only in an algebraic sense, we struggle in fact to minimize its overall implausibility. But the operative word is *overall*, for our implicit loss function is "quadratic," larger deviations from the norm carry more-than-proportional weight: a story the reader repeatedly glosses "maybe…" may receive serious consideration all the same, a single "never!" or "impossible!" will see it rejected out of hand.

The ur-rule can be developed into a set of less abstract recommendations. The next section offers a pentalogue that distils the fruit of my hard-won experience, an initial codification of the art of reconstructing economic growth; the examples are taken from my own work, my case study, but their import is general.

2. RECONSTRUCTING ECONOMIC GROWTH: *LE REGOLE DELL'ARTE*

2.1 Rule 1: Vet the data!

The first rule, already noted, is that the data in our sources cannot be taken at face value. We must see through them, gauge their relationship to the facts they ostensibly document, verify their credibility and potential usefulness; we must vet them, to use a trendier term deconstruct them.

In principle we should establish how each number in our sources was actually obtained, we should systematically investigate the underlying "data generating process." We who work on post-Unification Italy have the benefit of work done by Istat, then the Istituto centrale di statistica, for Italy's centenary: the lastingly useful multi-volume *Rilevazioni statistiche*, that documents the production of our "historical statistics," complete with reproductions of the forms circulated to gather the raw data. But even that is not enough, to understand what our data really are we would need to recover the instructions given to the enumerators, instructions possibly surviving in some archive, possibly never written down at all.[16]

Some of that we can certainly do: if production statistics were generated by a production tax, for example, we can and should go back to the enabling legislation, and learn what exactly was being taxed. Even the mere recognition that the production data are taxed-based can be instructive, and suggest for example that a sharp decline following a tax increase

[15] *A regola d'arte* is a common enough Italian phrase, without a direct equivalent in colloquial English: it harks back to medieval crafts (the *arti*), that guaranteed quality by imposing strict rules (the *regole*) on the materials and the manufacturing process. A literal translation is "[made] according to the rules of the trade."

[16] As is well known, Italy's 1881 census counted impossible numbers of female textile workers in the South. Tariffs on textiles had recently been hiked, and the textile factories were notoriously concentrated in the North; one wonders whether the census enumerators were "encouraged" to document that the industry was nation-wide, thus rescuing the tariff from the charge that it benefited the North alone. But even if this actually happened one hardly expects to find documentary confirmation.

registers a decline in *taxed* production only, less a decline in production than a shift to bootlegging. But a systematic approach to documenting the data-generation process is *ultra vires*, and, to the extent that it merely verifies the absence of evidence that it changed from one period to the next, not particularly efficient; most of the research we can and should do on the data-generating process is not *a priori* but *a posteriori*, triggered by signs that it changed provided by the data themselves. The identification of such signs is again a judgment call, a matter of interpretation; my guiding principle is Marshall's *natura non facit saltum*, it is the discontinuities in the series that attract my eye, that signal to me a change in the data-generating process.[17]

The eye should in fact look deep into the data. These may be found, already as well-arranged long time series, in abstracts of historical statistics (for Italy, the *Sommario*), abstracts which have typically lifted them from the statistical annuals of the day (the *Annuario*), annuals which in turn transcribe the more interesting figures from a cited battery of primary, narrowly focused sources. At times, a look at the final series is enough to warn us that it is not homogeneous over time; in the Italian context two examples come readily to mind. One concerns the official human-grain-consumption series, which displays an increase around the turn of the century that is simply not credible; it turns out to result from the interpolation of earlier (grossly underestimated) grain-production figures, disbelieved even when they were published, and later (far higher) figures based on much-improved production statistics (Fenoaltea 1969, pp. 97–98, 2010, pp. 83–85). The other concerns the official State-expenditure-on-public-works series, which looks perfectly normal save for an inexplicable downside outlier in 1870; research into the public budget's accounting rules revealed a change at that time, the practical result of which was that the figure for 1870 was only a partial one (Fenoaltea 1986, pp. 7–8).

But the reconstructed series in the secondary sources typically mask far more than they reveal, and nothing catches the eye until one goes back to the primary sources themselves. These typically tell us far more about the data they report than the subsequent annuals and abstracts, and most of all they generally allow a literal deconstruction of their aggregates, the identification and reconstruction of their components; and even a beginner can then spot that the series' coverage may change over time, that some components appear or disappear. Such reconstructions can also signal an altogether subtler trap, when many components do not change at all from year to year: it means that the elementary data were *not* systematically updated, that the current issue of the primary source simply published the latest available figures as the best estimates of the current ones, altogether missing what may be significant growth (thus in the Italian case the "annual" quarry and non-metallic-mineral-product output data for 1901ff. in the *Rivista mineraria*: Fenoaltea 2015b, 2015c). This is in fact a trap that will catch all those who construct a cross-section for a given year by consulting only the sources for that particular year: the broader lesson is that cross-sections need to be validated by time-series evidence that places year-specific data in their intertemporal context.

The final point, *qui devrait aller sans dire*, is that the historical data are also to be vetted in the light of their broader context, what we know, or should know, of the relevant activities, institutions, and technology, of somehow related data in other historical sources:

[17] This is admittedly a one-sided test, I will not notice simultaneous discontinuities in the data-generating process, and in the underlying matter the data refer to, that nicely offset each other; the lame response is that such coincidences are presumably rare, and that in any case even a one-sided test is an improvement over the prevailing absence of any test at all.

the knowledge, the specific culture, that accumulates with experience.[18] The inexpert can fail badly, assuming for example that all State expenditure for public works was funded by the *Ministero dei lavori pubblici*, confusing for example natural silk and artificial silk, measures of volume with measures of weight (Fenoaltea 1986, pp. 6–7, 29; 2003, p. 1099; 2018, p. 302); my own early howler was the confusion of the engineering industry with the "machinery" industry, a confusion prompted by everything, the very little, I had read at the time (Fenoaltea 1967, 2017a).

2.2 Rule 2: Disaggregate!

Disaggregate, disaggregate, that is Moses and the prophets!

To a historian the impulse to disaggregate should be a natural one, a direct consequence of our curiosity about the past. An aggregate alone is like the low-resolution newspaper photographs of now long ago, which if magnified dissolved into a meaningless blur; the internet has made us used to photographs that when magnified reveal increasing detail, photographs that allow me to zoom in from my hemisphere to my continent, to my country, to my province, to my house, to add an Italian touch to my laundry drying in the sun. An aggregate ("chemicals") that allows us to zoom in on its immediate components ("electrochemicals") and then on theirs ("calcium cyanamide"), and even on the local sources of these, is thoroughly satisfying – a rare satisfaction, achieved only where the surviving documentation is unusually rich. But the impulse remains, we want disaggregated "data" because details are our *friandises*.

But the curiosity that drives us is by no means idle. In general, the path of an aggregate places only very loose constraints on the possible "interpretation" of its movements, as is well illustrated by the extant literature on post-Unification Italy; the likely validity of such interpretations can be gauged only by drawing out their specific implications, and verifying their likelihood in the light of more detailed "data." More specifically, when we limit ourselves to an aggregate we implicitly and naturally assume that its composition remained more or less unchanged; disaggregation can reveal the falsehood of that assumption, it can demolish the shared beliefs that underpin an entire literature (as that on the industrial investment cycle in post-Unification Italy, Fenoaltea 2017a).

Technically, of course, as far as aggregate-level "measurement" alone is concerned the purpose of disaggregation is reaggregation, the calculation of an improved aggregate that takes changes in composition into account; and to that purpose both vertical disaggregation (across "stages of production" within a production sequence) and horizontal disaggregation (across production sequences) are to be pursued. Disaggregation rarely reaches an objective limit: the practical limit comes from *vita brevis*, we disaggregate far less than we theoretically should but as much as we practically can. Even our elementary disaggregated series typically remain aggregates, aggregates we can only hope remain reasonably homogeneous over time; the point is simply that that hope is the less forlorn, the more extensively we actually disaggregate.

The point of vertical disaggregation is of course to distinguish the *different* time paths of successive stages of production; and these differ most readily in the presence of international trade in the partially processed goods that are the output of one stage of production and the input into the next. In practice, then, our vertical disaggregation will be

[18] Only acquired familiarity with the institutional framework will reveal what is perhaps the most subtle trap, the at times changing definition of the self-same unit of measurement: merchant ships in particular were measured "uniformly" in register tons, but the formula used to calculate tonnage from the ship's specific dimensions was repeatedly changed.

dictated by the major trade flows (for example of yarn, to distinguish the path of cloth production from that of yarn production), while we ignore or collapse the minor ones (making no attempt to distinguish, say, gray, bleached, and dyed yarn). The myriad steps that compose a full production sequence are thus, in practice, reduced to a few: *e pluribus unum*, or not much more than that.

But vertical disaggregation has a further aspect that bears notice. In general, we can calculate the quantity of product *A* from the "apparent consumption" (production plus net imports, ignoring undocumented inventory change) of its input *B*, and the corresponding input-output ratio (*B*/*A*); applied through the production sequence, this algorithm generates a set of series (one for each successive stage of production) that are locked together by the data on the intervening trade flows.[19] From this two things follow: most obviously, that any one "known" series can be used, with the trade data, to generate the rest of the set (as is not infrequently done, using e.g. net imports of raw cotton to estimate yarn production, and those figures plus net imports of yarn to estimate cloth production). In a data-poor environment, none of the series may be "known," all we know (from the trade data) is how they differ from each other. To fix any one is to fix all the others, to attribute a plausible path to one may imply an implausible path for another; in such cases the full set must be fixed with an eye to the *joint* (im)plausibility of the resulting estimates, for the implicit maximand is, as noted, the verisimilitude of the entire story we tell.[20]

Technically, horizontal disaggregation improves the aggregate exactly as vertical disaggregation does, in this case by distinguishing among goods, and processes, that differ in their value added per measured unit of product – typically because of qualitative differences, as for example between woollens and worsteds, or between battleships and submarines, at times because of economies of scale.[21] Heuristically, it involves very different considerations. A systematic approach to disaggregating an aggregate forces one to identify its components: a novice may consider Italian shipbuilding adequately documented by the "ships launched" series in the *Sommario*, but if we ask ourselves at all what "shipbuilding" covers we will soon discover that it includes the production of new ships, merchant and naval, and the

[19] The algorithm obviously extends to (apparent) final consumption, which is simply the production of the finished good plus the net imports thereof.

[20] Imagine, to clarify the point, the flax-linen production sequence, undocumented at any stage. A burst of flax imports can be interpreted as fueling a burst of linen consumption, with flat flax production – or as offsetting a flax harvest failure, with flat consumption. As noted, joint verisimilitude reflects a quadratic loss function: a number of mildly unlikely events, together, are less unlikely than a single highly unlikely event.

[21] As a rule, there is no substitute for horizontal disaggregation; but at times it can be finessed by suitably choosing the unit of measurement. The archetype here is the wartime measure of aircraft production in the United States: rather than counting aircraft and distinguishing, say, trainers, fighters, and medium and heavy bombers, a meaningful aggregate figure was obtained directly by counting airframe tons. In the Italian case, similarly, sufficiently detailed trade data allowed the conversion of the aggregate cotton yarn and cloth production figures from units of weight, which fail to reflect quality differences, to units of length (of yarn produced, of yarn woven into cloth), which directly capture them, obviating the need to disaggregate by quality (Fenoaltea 2019c). Again similarly, the water-supply industry includes the product of aqueducts, characterized by significant economies of scale; a synthetic measure could be obtained by measuring the aqueducts' yield not in tons (per unit period) but in equivalent tons, calculated as the actual tons to a power that captures the economies of scale (Fenoaltea 2015g).

maintenance of existing ships, again merchant and naval. The extant series' unit of measurement (register tons) is then enough to reveal that it refers to merchant ships alone (naval vessels were measured in displacement tons); and "ships launched" clearly refers to new ships alone, revealing that the series in the sources documents only one out of the industry's *four* basic activities (Fenoaltea 2015f). Horizontal disaggregation thus serves in the first instance the same purpose as explicit models and counterfacts, it brings out, and invites reflection on, assumptions that otherwise remain implicit and unexamined.

The present example was not selected without cause, for the maintenance of durable goods raises specific issues of its own. In the first place, it must clearly be distinguished from new production: the latter tracks the gross additions to a stock, maintenance varies in the first instance with the stock itself. Unlike new production, too, maintenance cannot be meaningfully measured by a physical output, because the attendant value added varies, for any given type of equipment, with the condition it happened to be in when brought in to the shop; in general, the best elementary "real" series one can construct is a measure in "constant" monetary terms (typically obtained as a benchmark value added estimate, extrapolated by an index of the maintained stock's activity, at worst of the stock itself). Third, and in the circumstances unsurprisingly, maintenance is even more sparsely documented than new production: ships, trains, and public infrastructure aside, typically, there is damn little to go on. The upshot is that the disaggregation to separate maintenance from new production is at once necessary, and difficult: in the (again typical) Italian case the censuses did not separate maintenance shops and workers from new-production shops and workers – which were often the very same ones, as blacksmiths, for instance, engaged indifferently in the one as in the other – and the two activities can only be disentangled with the aid of ancillary evidence.[22]

The separation of maintenance from new production entails in fact two further problems, of a different order. The minor problem is whether to attribute value added in maintenance to industry or to the services; the choice affects only the distribution of an unchanged aggregate, and the issue is no more than an irritant (attributable to the United Nations, as the *ISIC* is here thoroughly inconsistent). The more significant problem is whether to consider maintenance production of (reconditioned) durables, and thus capital formation that inflates GDP, or a cost of producing the goods and services that employ those durables, an intermediate product that finally cancels out of GDP (as do all non-durables consumed in further production); and on this there is no consensus, nor uniformity in the literature. These issues are here only noted, to be returned to below (Appendix A).

2.3 Rule 3: *Think* when you index!

"Indexation" is a catch-all term for filling gaps in "the data," for coping with time series that lack pieces or do not exist at all; it is so called because we use a "known" series as an "index" of (a proxy for the movements of) an unknown series. The above discussion of vertical disaggregation recalls a common form of indexation, the use of raw material

[22] Evidence that may be far removed, and not obvious a priori. To clarify the point with an example, estimates of Italy's blacksmiths' aggregate value added in 1911 can be derived from that year's census data. Per unit of value added, new production consumes far more metal than maintenance; given total metal consumption, and the value added/metal consumption ratio in new production revealed by market prices, the disaggregation of value added into new production and maintenance yields an implicit value added/metal consumption ratio in maintenance which must itself be reasonable next to the corresponding ratio in new production. The estimates for 1911 are much more tightly constrained by the relative ratios they imply for 1871, when total metal consumption per worker was much lower, than by the relative ratios they imply for 1911 itself (Fenoaltea 2015f).

consumption movements as an index of production movements: it is indexation based directly on a relatively tight technical relationship, there is no cause for complaint. But much indexation in the literature is utterly mindless, based on nothing other than bad precedent, unjustified and unjustifiable: proper measurement requires that we observe our third rule, that indexation be thought out.

If a series displays a gap, the latter can be filled in ("the series can be interpolated") in a variety of ways. The simplest index, the know-nothing index, is simply the passing of time: the interpolated values are obtained assuming constant growth.[23] Even here, however, some reflection is in order before proceeding with the calculation. Most of us quantitative historians live in semi-log space, we consider a constant growth *rate* (*x*% per year, log-linear or "geometric" interpolation) altogether more natural (verisimilar) than constant *absolute* growth (*x* tons or whatever per year, linear or "arithmetic" interpolation).[24] The obvious problem, however, is that while linear interpolations are additive, log-linear interpolations are not: the sum of the linear interpolations of *a* and *b* is the linear interpolation of (*a* + *b*), the sum of the log-linear interpolations of *a* and *b* is not the log-linear interpolation of (*a* + *b*). Imagine that from one end of our interpolation to the other *a* has grown sharply, and *b* declined sharply by just enough to offset that, so that (*a* + *b*) does not change at all. Linearly interpolating *a* and *b*, or (*a* + *b*), or log-linearly interpolating (*a* + *b*), we obtain, obviously, a flat interpolation, a constant value in each intervening year; if we log-linearly interpolate *a* and *b* and then aggregate, (*a* + *b*) will decline and then recover, displaying a cycle that is nowhere in the data and as a rule thoroughly unlikely on its face.[25] It may well make more sense, in such cases, to reverse the order of the estimates, first log-linearly interpolating (*a* + *b*), and then obtaining *a* and *b* by interpolating their shares of that total.[26]

---

[23] The procedure is so standard that this is the default meaning of "interpolation" if no specific interpolating variable is explicitly indicated.

[24] If we plot tons (say) against time, constant *absolute* growth yields a straight line, a constant growth *rate* a curved one; if we plot log(tons) against time, constant *absolute* growth yields a curved line, constant *relative* growth, a constant growth *rate*, a straight one (whence "log-linear" if the growth *rate* is held constant).

[25] The published series that display a U-shape between benchmarks signal that they were most probably constructed by log-linear interpolation of their components. The attribution of constant-growth paths to the individual components may seem reasonable if we look no further, but the implied U-shape of the aggregate remains implausible: if the above scenario were to hold across a series of benchmarks, a U-shape would link each successive pair, with the hardly credible implication that each and every observation happened to coincide with a local maximum.

[26] Shares are best interpolated linearly, as they then sum to one. Log-linearly interpolated shares do not (and the share obtained for *b* from the log-linear interpolation of *a* is not the share of *b* obtained by its own log-linear interpolation; log-linearly interpolating the shares of both *a* and *b* and then rescaling the results to sum to one seems pointlessly complex). Conversely, (input-output and other) ratios are best interpolated geometrically, as the interpolated values of (*a*/*b*) then equal the inverse of the interpolated values of (*b*/*a*); arithmetic interpolation yields different values, depending on which form of the same ratio is actually interpolated.

Again frequently, we interpolate a gap in our series for *a* by attributing to *a* the movements of a "known" related variable *b* ("using *b* as an index of *a*").[27] Extraordinary good fortune aside, however, over the relevant gap the relative changes in *a* and in *b* do not coincide, and must be reconciled. The standard ("automatic") solution is a trend correction, i.e., the elimination of the discrepancy by adding a constant to *b*'s annual percentage changes; but that algorithm need not be appropriate. When *b* grows much more than *a*, and its growth sharply accelerates, the trend correction may turn the years of relatively slow growth in *b* into years of decline in *a*; and that decline in *a* may again make no sense at all, for example if both *a* and *b* are responding, with different elasticities, to the same impulse (e.g., income growth). In such cases, the mere recognition of the problem points to its solution: not a trend correction but an elasticity correction, i.e., the elimination of the discrepancy by a *multiplicative* scaling of *b*'s annual percentage changes.[28] Alternative solutions may yield very different profiles; the point here is again that there *are* alternative solutions, and that the choice must be made with due consideration.

But the most damagingly mindless indexation occurs across production sequences, when entire industries are undocumented, the time path of their product "unknown"; and it occurs more often than not, for the surviving evidence is terribly partial. The standard procedure, in such cases, is to calculate the desired aggregate from its known components, up from aggregate to higher-level aggregate (using for example cotton and silk alone to represent the entire textile sector, and then the textile series, and those for the other thus reconstructed sectors, to represent all industry): so standard a procedure that it is simply followed, without discussion or justification (with a single well-known exception, returned to below).

Followed also without a thought, clearly, for any thought at all would have killed it before it became established. The procedure's first step is to attribute to unobserved production the path of observed production "of the same sector": a double absurdity, *le vice appuyé sur le bras du crime*. The "sectors" into which we classify the economy may have a logic, but they were not designed to support extrapolation across their components: from the present perspective they are simply arbitrary. Think of the rubber industry, indifferently considered a sector in its own right or part of the "chemicals (and related products)" sector, and imagine that its product is "unknown." With the latter classification rubber is attributed the path of the known elements of the "chemicals" group (in Gerschenkron 1962[1955], the production of sulphuric acid); with the former, the path of the known elements of industrial production in general (there milling, cotton, sulphuric acid, etc., all weighted by their value added and again by a coefficient that reflects the coverage of the individual industry groups). With the standard procedure the path attributed to rubber, and therefore to industry as a whole, is as arbitrary as the selected industrial classification.

The attribution of the path of observed production to unobserved production "of the same sector" is also nonsense because the components of a given sector may be independent,

---

[27] There is in such cases the obvious temptation to regress *a* on *b*, and to interpolate *a* using the resulting parameters; but these are typically so sensitive to the selected regressors and sample period that the resulting estimates are no less arbitrary than those obtained by the direct indexation with which we typically make do. The ultimate criterion, once again, is verisimilitude, the "reasonableness" of the result.

[28] In the case at hand, this is geometrically equivalent to forcing the interpolating curve through the desired end-point by flattening it, as opposed to rotating it.

or even rivals. Consider for example the textile industries, and imagine, for simplicity, that they process only cotton and linen; that neither is covered by output data; and that the apparent consumption of raw materials documents the growth of the first (because raw cotton is imported), but not of the second (because flax is home-grown). The growth of the textile sector is therefore represented by that of the cotton industry, in effect assuming that the linen industry matched its growth. But we know that the cotton industry was the first to be mechanized, that the linen industry was successfully mechanized over a century later: that technological change did not affect the cotton and linen industries together, but long favored the first *at the expense of* the second. The assumption that linen production grew as cotton production did could not be more palpably wrong-headed.

Consider too the case of the extractive industries, made up of the mining group and the quarrying group. In many (Continental) countries the subsoil belonged to the Crown: mining was regulated, and documented, as quarrying was not. In standard practice, the entire extractive group is indexed by mining alone, implicitly assuming that quarrying moved exactly like mining. Ask yourself how you would estimate quarrying production in its own right, and set yourself to the task: would you ever assume it moved like the mining sector, which operates in an essentially unrelated market? Would you not infer its movements from construction activity, whose materials quarrying provides, working back through the production sequence exactly as above (§2.2)? One wonders why that reasonable procedure is not normally applied across sectors as it is within them: it is as if quantitative economic historians were mesmerized by the Statistical Bureaus' partitioning of the economy into different sectors, like deer caught in the headlights of a fast-approaching car, and with equally gruesome results.

The procedure's second step is to take the path of the sum of the (partly) documented sectors to represent the aggregate, that is, to attribute to the (totally) undocumented sectors, together, the path of the (partly) documented sectors, together; and similar considerations apply, in spades. Some industries, typically those processing tropical products, were documented as noted by the general statistics on international trade; but direct evidence of production was gathered first and foremost where it was of particular interest, and relatively easy to obtain. On both counts, the sources tend generally to document the new factory industries far more than the traditional, much smaller-scale and far more dispersed, artisanal sectors: in general, the better-documented sectors were growing at the expense of the less-documented ones, the assumption that artisanal production grew as factory production did could not be, again, more palpably wrong-headed. There may be practical reasons that demand an immediate aggregate estimate, that warrant resort to guesswork rather than research; but nothing can justify *mindless* guesswork, the standard guess that defines the path of the undocumented sectors without so much as considering what those were, and how they differed from the documented ones.[29]

The roots of the standard procedure can only be inferred, as they are essentially unspoken; and three possibilities come immediately to mind. The first is the natural desire to produce an aggregate that includes only historical "data," to the exclusion of estimates,

[29] Contrast Fenoaltea (1972), p. 349: because documented manufacturing seemed essentially to cover new/factory industry and in particular the cyclical investment-goods sector, undocumented manufacturing was identified with the artisanal production of consumer goods, and attributed a simple slowly-growing (demographic) trend.

"guesses"; but that line of thought leads nowhere. One reason, familiar to those happy few who have actually approached the historical sources as historians, vetting them before copying "the data," is that the reported figures are not infrequently themselves guesses, the best guesses of the experts, or appointed officials, of the day. At times, as recalled above, the frequency of data publication exceeded the frequency of observation, and the best guess for "this year" was "same as the last time we looked, *x* years ago": in such cases the "data" are not only guesses, but not even guesses as good as those we can make, as we know what was discovered with the next actual observation. The other, overarching reason, which should be immediately obvious even to non-specialists, is that any aggregate series based on a partial set of ("observed") component series implicitly incorporates the very definite guess that the excluded ("not observed") series, together, moved exactly like included ones, again together. The surviving "data" are very partial, the only way to avoid including guesswork in the aggregate is not to produce the aggregate at all.[30]

This brings us to the second possibility: that the extrapolation from the documented subset to the aggregate is recognized as a guess, but considered a good guess, the legitimate attribution to the population of the time path of the sample. The problem here is that the sample is "random" only in a colloquial sense, when I started I had no idea what I would find, what I did find was what the sources happened ("randomly") to throw up.[31] A *statistically* random sample is made of sterner stuff, it must be *designed* to be representative of the underlying population; and the historical sources of production data were designed with other ends in view, they reflect the specific (and typically fiscal) concerns of the governments of the day. Statistical representativeness was neither here nor there, traditional artisanal industries in particular were of little interest and correspondingly "undersampled": to attribute to the aggregate the path of the documented subset is not only to guess but to guess, as a rule, demonstrably badly.

And that leads us to the third (and only documented) possibility, Sir Charles Feinstein's claim that there is nothing else one *can* do, that necessity is here the mother of the lack of invention (Feinstein 1972, p. 207). The central point here is that Feinstein was simply wrong, where direct evidence is lacking there is typically much indirect evidence that can be exploited, if only one recognizes it for what it is. The proper method is time-consuming, but ultimately simple, in fact simplex. Invent the series you seek to construct, your initial best guess; but don't stop there, the starting point matters little only if you move beyond it. Draw out the implications of your series as an applied economist would, recognizing technical relationships, the impact of trade, the substitution effects that can be inferred from the typically abundant evidence on relative prices, the income effects, where appropriate, that influence consumption; and set those implications next to the corpus of surviving "data," to

[30] As just noted, Fenoaltea (1972) treated the index constructed in Fenoaltea (1967) as an index of *documented* production alone. The earlier work did not address that issue: my concern there was rather to show that the growth rates obtained from aggregate indices were very sensitive to the way one weighted the component series, implicitly undercutting the argument in Gerschenkron (1955, pp. 365–366) from the growth rate generated by his own index.

[31] With the same logic one would attribute to the entire population the mean documented income, the income of the few rich enough to pay income tax and thus leave a record of their income. As far as I know nobody has ever done *that*, the procedure is obviously absurd; why its absurdity seems not to be obvious in the present context I cannot begin to explain.

the extent that you can master it, as an historian would. You will soon enough find that your initial estimates violate constraints that may be distant but are constraints all the same. Revise, rinse, and repeat; at the end of the process you will have a production series, for the "undocumented" industry at hand, that is reasonably tightly constrained by economic logic, and by the historical evidence. No more can be asked of it.

The bottom line is that *all* "undocumented" production too must be estimated, with suitable disaggregation, explicitly and in its own right – which should be enough to wean you from the standard "indexing" procedure, which no one can possibly follow *en connaissance de cause* – and above all with due research and reflection, identifying and exploiting the available indirect evidence. And once all not-directly-documented production too is properly estimated in its own right, the classification of economic activities becomes harmless: the estimate of rubber production is then what it is, whether we count rubber as a separate group or part of the chemical group affects only our groupings; the higher-level aggregates, like the elementary disaggregated estimates, are quite unaffected.

Indexation, the inference from the known to the unknown, must be thought out: we must think *before* we index, we must think *while* we index, we must think again, for good measure, *after* we index. It's a sad comment on the state of our ("intellectual") profession that we should have to be reminded to *think*.

2.4 Rule 4: Deflate all current-price values with the *same* deflator!

Our historical measures of value are born, inevitably, at current prices; to eliminate the distortions due to the changing purchasing power of the monetary unit, we "deflate" them into what we call "real" measures.[32]

The fourth rule is that deflation must be general and not activity-specific. The discussion can be technical (Fenoaltea 1976), but the essential point is simple enough. To construct aggregates, to compare their components, we need to reduce these to a common metric; in the measurement context at hand the obvious metric is "value added," that corresponds at once to the value of an activity's product, net of the materials it consumes, and to the value of the activity itself, the income accruing to the primary factors of production ("land, labor, and capital"). Three points bear notice. The first is that "value added" is the obvious metric only now that the concept is part of our standard intellectual baggage; we owe it to the United States Census Office, who developed it over the late nineteenth century, not without difficulty, to meet the perceived need for a *net* measure unaffected by vertical (dis)integration (ibid., p. 111).[33] Second, the value-equivalence of the results of activity and

[32] *Deflate*, because when the problem first presented itself the immediate need was to eliminate the distortion of different-year current-price measures caused by *inflation*. In general, the "real" measure R is obtained as, or equivalent to, the current-price ("nominal") measure V divided by a price index P, $R = V/P$. V is unambiguous, our concern here is with the deflator P.

[33] The measure that came more naturally to hand (to the census-takers that measured production, to the legislators that taxed it) was simply *value*, the firms' sales. But it was recognized that equal sales could correspond to very different levels of activity, as for example if two textile firms sold identical quantities of cloth at the same price, but one worked from the raw fiber, the other from purchased yarn; and that the aggregate sales of the firms in an industry (and the accompanying "turnover tax") could be radically reduced by vertical integration even though nothing changed on the shop floor, as the transfer of yarn from the yarn-producer to the weaver would pass from a sale on the market to a transaction internal to the now integrated firm.

of the activity itself is *in ipsis rebus*, given that activities are valued by their results; it is complicated by speculation, market power, and taxation, but these have been dealt with elsewhere (ibid.), and need not detain us here. Third, the objection that this equivalence holds only in zero-profit long-run equilibrium is based on the standard textbook model of short-run equilibrium, which contemplates non-zero profits; but that objection is as worthless as that model (ibid.; also Fenoaltea 2001).[34]

In any year, at that year's prices, we happily accept that if the value added of industry *a* is twice that of industry *b*, industry *a* produces twice as much, and is twice as big, as industry *b*. Problems arise in comparing "value added" across years, because the monetary unit in which we denominate our measures (e.g., a dollar) has the nasty habit of stretching, or more often shrinking, over time. To obtain a measure free of the attendant distortion we must "deflate" our current-price figures into what we call measures "in constant dollars" or more directly "in real terms." In the present context, as noted, the problem can be expressed as the need to deflate current-price value added to calculate "real value added." So far so good.

What, then, do we want of our "real" measures? At a minimum, surely, that they not generate wrong answers where we know the right one. Imagine, to illustrate the issue, a school's class photographs. Imagine that we have (as we do in the present context) an interest in relative magnitudes; and reduce it for simplicity to a merely ordinal interest in heights, we want to know only who is taller than whom, never mind by how much. So on the day the students come to school in flat shoes, each class arranges itself from tallest to least tall, and the photographs are taken. A further photograph is taken of the entire student body, similarly arranged. Clearly, if in the photograph of their common class Judy is taller than John, in the photograph of the entire student body Judy is again taller than John: how could she not be?

The class photographs correspond to our current-price value added measures, that establish relative rank in a limited context (the year); the student-body photograph, to our "deflated" measures, that we want to illustrate relative rank even across years.[35] Return to the metaphor: our problem is that we cannot actually take the student-body photograph, we must construct it by photoshopping, and merging, the pictures of the individual classes – which, as it happened, arrayed themselves at varying distances from the camera. So one class photograph can be taken as is, as the base; but to reconstruct the student-body photograph all the others must be scaled, the individuals extracted and slotted into the appropriate place in

[34] Any economic historian/historian of economic thought can readily see how that model emerged out of its British context, where industrial firms owned their machinery. Any economist should recognize that in a world of complete markets firms (can) rent their machinery as they rent their labor, that in the short run the stock of (industry-specific) machinery is given not for the firm but for the *industry*: with competing entrepreneurs the rental rate of machinery (its annual shadow price, if it is owned) varies to drive profit to zero even in the short run. Nor is that all: it should also be obvious that if an industry is the sole consumer of a raw material the possible variations in output may well be broad enough to affect the price of the raw material even with a fixed aggregate stock of machinery. The *only* difference between the (correctly understood) long-run equilibrium and the (misunderstood) short-run equilibrium is that in the latter *industry* supply is constrained by the given stock of equipment. The textbook dictum "the short-run industry supply curve is the horizontal sum of the firms' short-run supply curves" is simply *wrong*. We economists are as careless about the basic theory we teach as about the "evidence" we use: "scientists" indeed!

[35] The objective is to render every observation directly comparable to any other, say industry *a* in a given year to itself in a different year, to industry *b* in the same year, and to industry *b* in a different year.

the larger group. The merging is not easy – it is hard to tell how the tallest student in one class compares to the tallest in another – but one thing is clear: all the individual figures in each class picture are to be scaled in the *same* proportion. If that simple rule is not respected, our reconstructed picture of the student body may show our friend John as taller than his class-mate Judy; we already know that is wrong, and if that is how they appear in the student-body reconstruction the responsible photoshopper is clearly incompetent.

The reader unfamiliar with the literature may well be wondering why s/he had to suffer through the preceding paragraphs to reach a conclusion obvious to the meanest intelligence (a characterization on which I take the Fifth); the reader familiar with the literature will have grasped their import. The rule that deflation must be general and not activity-specific is the claim, in the terms of our metaphor, that Judy and John must be scaled in the same proportion, and not in different proportions specific to Judy on the one hand and John on the other. It is apparently not obvious to the profession, for *the standard "real" measure in the literature violates this elementary rule*, and deflates the value added of different activities with activity-specific deflators.

That measure is the Fabricant-Geary "double-deflation" measure (*SNA*, p. 295), calculated from the standard value-added formula using constant ("base year") prices. Let $v_{it} = p_{it}Q_{it} - z_{it}R_{it}$ represent the current-price value added of activity i in year t, where p_{it} and Q_{it} are the price and quantity of its output and z_{it} and R_{it} the price and quantity of its raw material(s); the standard measure of "real value added" at the prices of the base year o is $v_{rito} = p_{io}Q_{it} - z_{io}R_{it}$. Three things immediately hit the eye, and the fan. First, this measure is equivalent to the deflation of current-price value added by an activity-specific deflator: $v_{rito} = p_{io}Q_{it} - z_{io}R_{it} = (p_{it}Q_{it} - z_{it}R_{it})/[(p_{it}Q_{it} - z_{it}R_{it})/(p_{io}Q_{it} - z_{io}R_{it})]$, where the denominator in square brackets is a (current-year-quantity-weighted) index of the (output and input) prices specific to industry i; to return to our earlier metaphor, our photoshopper is clearly incompetent, the algorithm generates nonsense results. Second, as every economist should know, current-price value added can be indifferently measured as sales net of material costs, or payments to land, labor, and capital (above, footnote 34): $v_{it} = p_{it}Q_{it} - z_{it}R_{it} = r_{it}K_{it} + w_{it}L_{it}$, where K represents (land and) equipment in physical units, r is the rental value per unit, L is the labor consumed also in physical units, and w is the unit wage, all of course per unit time. If current-price value added is indifferently measured in two different ways, deflated ("real") value added too should be indifferently obtained from either one; but if we use quantities and base-year prices that will not be the case, in general $p_{io}Q_{it} - z_{io}R_{it}$ will *not* equal $r_{io}K_{it} + w_{io}L_{it}$.[36] Again, the measure yields nonsense results: its inventors and their imitators are attempting *economic* measurement with an inadequate grasp of economics, that the result should be rubbish is hardly surprising. Third, there is nothing in $p_{io}Q_{it} - z_{io}R_{it}$ that guarantees a positive outcome, measured "real value added" may well be negative – and it will be, if as we go back in time the input-output ratio becomes higher and higher, as it does in any industry marked by significant materials-saving technical progress (including fuel-saving progress, as for example in metallurgy). The measure's results are then obviously nonsense (strongly suggesting that they are always nonsense, even when not obviously so). The immediate problem is again bad economics: there is a logic to the price system, relative input and output prices reflect productivity, the input-output ratio; combining prices that reflect one technology and quantities that reflect another is absurd on the face of it.[37]

---

[36] The one deflates current-price value added by an index of output and raw material prices, the other by an index of labor and machinery prices (rental rates).

[37] Of these three problems, only the third was widely noted by the profession, because negative estimates soon turned up. Characteristically, the reaction was not to *think*, starting from first

The root problem seems twofold. On the one hand, the sheer intellectual sloppiness of our standard measures reflects our professional indifference to measurement, our refusal to think about it, seriously, as economists; it is yet another manifestation of our casual approach to the evidence. On the other, the particular form of the standard measures points to a lack of general education, of adequate literacy. "Real" measures have been taken to mean measures literally in (price-weighted) things, *res*, with no recognition that the technical meaning of the word is metaphorical. Even a casual acquaintance with the history of economic thought is enough to elucidate the matter: we called our deflated measures "real," thing-like, in the context of inflation, when things are "real" not because they are things but because they keep their value in exchange, and the currency does not. Imagine a world with a stable price level, with substantially unchanged relative prices, save that one good loses its value (because of exceptional technical progress, or mass conversion to a religion with dietary restrictions); in that world money is "real," and all goods are "real" except that one. The antonym of literal "real" is "unreal," the antonym of metaphorical "real" is "nominal": a clear enough signal, one would think, save for the verbally challenged, the "scientific" economists and economic historians, American and Americanized, who never learned how language works because they never struggled, in their formative years, with Latin and Greek. In this literature the only exceptions known to me, economists who saw *through* the *res* metaphor and advocated general rather than activity-specific deflation, are two Italians born early enough, and well enough, to have reaped as a matter of course the benefits of a classical education (Fenoaltea 1976, Fuà 1993); methinks it is not a coincidence.[38]

The bottom line is that to measure all production by the same unchanging standard what we actually want to calculate is not "real" value added but "real value" added: we want to deflate *all* current-price values by the *same* deflator, the price of "the" good that maintains its "real value." Here, sadly, the argument peters out without reaching closure, for no such good stands out. Setting aside extravagant suggestions (Fenoaltea 2010), the leading candidates are the early favorite, an hour of common labor (as "[the] value [of a nominal sum] is precisely equal to the quantity of labor which it can … purchase or command"), or the current standard, a broad basket of goods; but the first neglects the rising value of labor itself

---

principles, but to look for band-aids. Paul David (1962), in particular, proposed deflating value added by the output price alone: guaranteeing non-negative results, but maintaining activity-specific deflation, and violating the first condition that we want a value added measure to meet, that the aggregate be insensitive to vertical (dis)integration. Current-price value added in turning cotton fiber into cloth is the same whether we consider it one activity from end to end, or two activities, one producing yarn from fiber, the other cloth from yarn; David's index produces different results if all value added from fiber to cloth is deflated by the cloth price, or if the cloth price is used to deflate only the value added from yarn to cloth, and the price of yarn is used to deflate the value added in working fiber into yarn.

[38] Two comments may be added here. The first, to engage in counterfactual intellectual history, is that the profession's "real" measures might have followed a very different path had that poisoned metaphor been kept at bay, and the problem verbalized only as that of "deflating" current values into a time-invariant unit. The other is that the profession's lack of adequate verbal skills is confirmed by its failure to see through Fogel's attention-seeking word games, first on the "importance" of the railways (above, §1.1, footnote 4), and then again on the "efficiency" of slavery (Fenoaltea 1981), not to mention in the present context the Sims-Arrow claim that "real value added" does not even exist unless the production function is so separable that the primary factors of production alone combine to produce such a "thing," a thing that then interacts with the raw materials to produce the final product (Sims 1969, Arrow 1974).

as productivity increases, the second neglects the declining value of goods as they become more abundant.[39] Both seem to be limits to, rather than examples of, an intuitively appealing measure: our "real" measures appear to be defined, at best, up to the growth of the "real wage."

This ambiguity of our "real" measures is so to speak *in ipsis rebus*, there is no getting around it; and that is why our fourth rule goes no further than it does. A compelling standard of "real value" cannot be identified or prescribed; but whatever we select as our make-do standard it is clear that it must be used across the board, that deflation must be general and not activity-specific.[40]

2.5 Rule 5: Measure what you want to measure!

The above four rules all concern *how* product should be measured. A fifth rule concerns *what* we should measure, and it is *on ne peut plus Lapalissien*: we should measure what we want to measure. It follows, in the case at hand, that as we are interested in the economy's aggregate product, *that* is what we should measure.

Our standard measure of the economy's aggregate product is what we call "gross domestic product," familiarly, GDP. Long ago, when teaching Economics 1 in the United States, I would end my presentation of the national income accounts with the question, "why does the U.S. have the world's highest per-capita product?" (as it then was). The students answered with obvious references to advanced technology, abundant resources, "capitalist" efficiency (no comsymps there). Those reasons, I would answer, were true but superficial: "the real reason," I would say, "is that the measure was invented here." The point, of course, was that measured product was not a fact, something we observed, but a construct, one of many possible constructs.

That particular construct was defined by its particular genesis: who built it, to what purpose, and of what materials. The U.S. national accounts appeared *in utero* in the 1930s at Wesley Clair Mitchell's National Bureau of Economic Research, an institution marked at once by its atheoretical approach, and by its specific interest in cyclical fluctuations (e.g., Lerner 1947); they emerged as official statistics in the U.S. shortly thereafter, and world-wide, essentially on the American model, in the aftermath of the Second World War.[41] They came of age in a world marked by the Great Depression, when it was widely believed that mature capitalism tended inevitably to crisis and mass unemployment, that rearmament and war had been only momentary, dreadful remedies, that the next great slump was just around the corner. Governments therefore took on the task of stabilizing the business cycle, and maintaining employment, with the tools suggested by the *General Theory*; but to employ them to good effect they needed timely evidence on the path of the economy. The national accounts were to

---

[39] That decline may be limited, and tied as it were to pigovian diminishing marginal utility, or catastrophic, as and if goods increasingly become mere counters in a veblenesque zero-sum status game. The "goods" standard in particular is further burdened by the arbitrariness of any selected basket.

[40] It bears notice that if current-price value added is uniformly deflated by a common deflator, the much-observed "Gerschenkron effect" simply disappears: it too reflects not an "index-number problem," but simply bad measurement (Fenoaltea 2019b).

[41] The success of the American model again owed more to hegemony than to technical merit. Istat's *Reddito nazionale* had followed the Italian conventions, and excluded *intermediate* government services from aggregate final product; the Fuà team was funded by the Ford Foundation, and Vitali's estimates included them (Fuà 1969), as do our more recent ones.

provide that evidence, with minimal delay: they had to be calculated quickly, even if approximately, using statistics that were already available or easily obtained; they were to document the current path of the economy, its likely impact on paid employment.

The official accounts were shaped by Simon Kuznets, a protégé of Mitchell's. In his measure Kuznets included all agricultural production, for the market and not, because the available data were based on observed acreages and yields. He included industrial production only for the market, and counted its value added, or its value, depending on what data were already provided by the Department of Commerce.[42] Of the services Kuznets again counted those sold on the (legal) market, but also the imputed rental value of owner-occupied housing, again because the underlying ready-made statistics refer, as in the case of agriculture, to the aggregate stock. *Nada mas*: Kuznets gave us an empirical aggregate to solve a practical problem, a creature of the Bureau with no theoretical basis at all. It is not a *measure* of anything, it is at best a rough index of paid-employment-generating production, an even rougher index of total product: and that in the short run, when the *ceteris paribus* clause may be a reasonable approximation.[43] It is not a fact, not an observation, but a construct, in fact a muddy one, good enough for government work. And government work it became: as noted above (§1.2) the profession hastily abandoned its pursuit of economic measures and happily took them, from then on, as issued by the relevant public Bureau.

Kuznets rendered the profession a great service, and a great *dis*service: he called his construct not "an index of predominantly market-oriented, paid-employment-generating economic activity," as he could and perhaps should have, not even "an index of gross domestic product," which seems the least demanded by intellectual honesty, but, notoriously, "gross domestic product" *tout court* (actually "gross national product," at the time, but that is here irrelevant). We all know that GDP falls if a man marries his housekeeper, even if there is no change in her activities (*honi soit qui mal y pense*), in her product, and therefore in *total* product, *ceteris paribus*; we all know, or should know, that "GDP" is not the measure its label suggests. But that has not stopped the profession from taking the label literally: because we do not take measurement seriously, perhaps once again because we are verbally challenged, perhaps also because we "social scientists" approach economics as a religion, proscribing heresy, accepting the dictates of the clergy, apparently believing that a statistic consecrated as

---

[42] For most industry, as noted (§2.4), the Department had evolved measures of value added; but the Department lacked information on the value of the sub-soil resources the extractive industries consumed, and Kuznets simply counted the mining firms' sales rather than their value added. The drawing-down of (underground) stocks is simply ignored; in strict logic, the mining sector is treated as if the goods it sells were created out of thin air rather than extracted (Fenoaltea 2005, pp. 306–307), whence of course the sky-high per-capita "product" of oil-producing deserts. To be precise, in the national accounts the mining firms' "value added" is computed by deducting from sales only the cost of purchased fuel and similar ancillary materials. An analogous "value added" for the transportation industries would deduct from the (c.i.f.) delivered value of the goods only the cost of purchased fuel and the like, and *include* the (f.o.b.) value of the goods at the point of origin. This mixing of value added and value demonstrates that the national accounts do not consistently measure production on a value added basis to avoid duplication (and sensitivity to vertical integration), as we tell our students: the underlying motivation was not theoretical but practical.

[43] The services of owner-occupied housing generate product but not paid employment; make-work projects, digging holes and refilling them, generate paid employment but no product; and so on, about which more below. A specialized index of paid employment and a specialized index of production are different tools; Kuznets' all-purpose Swiss army knife does everything, badly.

a measure of gross domestic product is transubstantiated into exactly that (Fenoaltea 2019a).[44] Whatever the reason, the result is clear enough: we accept "GDP" not for what it is but for what it says it is, an economist is one who uses a government-issue screwdriver to hammer nails because it says HAMMER right on the handle.[45]

We economic historians, in particular, have no interest in a rough-and-ready policy-guiding index of the economy's current path, in "GDP." We want to gauge the evolution of economies over decades and more, we want to compare them to each other as well as to themselves earlier or later; and to do that we need a proper measure of the economy's product, a measure of the opportunity set, in goods-space, it made available to those then alive (over their expected lives, at that, and not in any one year, think of the later fourteenth century). A number of considerations come immediately to mind. Market exchange and paid employment are, as such, simply irrelevant (Pollak 1985): our measure must count unpaid "family production" (typically the work of women, there is more than one battle to be fought here), the unpaid services of durables, including both consumer durables (not just owner-occupied housing but also, e.g., the appliances that allowed housewives to work also outside the home, Gordon 2016) and common-use infrastructure (the piazzas their Italian "owner-occupiers" enjoy daily, and Americans cross an ocean to see, which is of course where I came in), and obviously leisure (corrected for morbidity); and it must count the all-important gifts of nature, that vary from time to time and place to place. By the same token, our measure must exclude not just product-less make-work projects but "social intermediates" (armaments, by extension the police and the judiciary, perhaps the legal professions), and allow for negative externalities: production externalities (environmental costs, including if we want to count it here the reduction of our subsoil assets), and consumption externalities too, those caused both by congestion (the crush of tourists that has rendered our favorite piazzas quite unlivable) and by social rivalry (which turns increasing consumption into a zero-sum game, Veblen 1899, and may well destroy much of what we call "modern economic growth").[46]

---

[44] So entrenched has "GDP" become as our measure of "the economy" that even the few economists who pursued a better measure of total product felt they had to give it a different name (e.g., Nordhaus and Tobin 1972).

[45] It bears notice that the bureaucrats and the profession have parted company: where we take "GDP" at face value, at least in our empirical work, the United Nations emphasizes that what it (nonetheless) calls Gross Domestic Product is not that at all, that its "production boundary" is not all-inclusive (*SNA*, pp. 6–7). But they do not concede that the measure is what it is because Kuznets constructed it out of whatever statistics were readily available; their disingenuous claim is that it is a purpose-built measure of the part of the economy that is of interest to policy-makers, designed as it is to avoid "being swamped" by other values. They claim that it rightly "includes all production of goods for own use …, as the decision whether goods are to be sold or retained for own use can be made even after they have been produced, [and] excludes all production of services for own final consumption within households … because the decision to consume them within the household is made even before the service is provided" – as if a subsistence cultivator did not decide to consume the harvest before s/he planted it, or I could not charge my neighbor for cooking his steak on my barbecue. The further argument that traditional women's work must be excluded to maintain consistency with their own definition of "the labor force," which fails to recognize it, is magnificently solipsistic.

[46] This paragraph could easily be expanded into a book, but a few points bear immediate notice. One is that the flow account must be complemented by a stock account, with the former incorporating the per-period changes in the latter; the current product includes investment, by firms and households (as the present value of future services), and excludes disinvestment (the drawing down of stocks due to obsolescence, catastrophe, depletion, and depreciation: our fixation with gross rather than net product

Our backcasting of what we call "gross domestic product" is in fact intrinsically laughable. Imagine us in our Valhalla, imagine our conversation with economic historians yet unborn, imagine that they ask us what our generation did. Shall we be allowed to answer "We reconstructed the historical national accounts" ("Oh, wow!")? Or will Valhalla admit only the unvarnished truth? "We reconstructed the short-term indices of paid-employment-generating-production that would have helped past governments implement their stabilization policies, had they had our statistics and had they had such policies" ("You did *what???*").

The bottom line is that if we are interested in economic growth we should *not* construct that index of predominantly market-oriented, paid-employment-generating economic activity we have misnamed "gross domestic product," but measure instead the relevant aggregate, correcting "GDP" to include for example leisure, unpaid "family production," and the services of public and private durables other than houses, to exclude for example such "social intermediate goods" as the military, and to allow for the externalities of both production and consumption. A challenging agenda, to be sure, but no more so than the very different one that gave Kuznets his place in our intellectual history.

3. THE EXTANT HISTORICAL NATIONAL ACCOUNTS

3.1 *Genesis*

The above rules are not all on the same level. The first three – vet the data, disaggregate, *think* when you index – inform the construction of our elementary production series, typically in their natural physical units ("tons"): in the language of simple realism, they (aim to) recover the facts that could have been observed, the production that could have been measured, at the time. The series so obtained are constructs we carefully derive from the evidence in the sources; they are "elementary" only metonymically, as (our substitute for) the production "data" ideal sources would have handed down to us ready-made, and again computationally, as the activity-specific reconstructions that then enter the higher-level estimates. These three rules are addressed to the economic historian as historian, their aim is to ensure the quality, the historical verisimilitude, of the materials with which we erect our structures.

The fourth rule concerns the construction of "real value added" estimates meaningfully comparable across activities and time periods, and enjoins the deflation of current-price value added by a *common* deflator; the fifth concerns the construction of our final measure of aggregate product, and enjoins the adoption of a measure without the manifest deficiencies of what we continue to call "GDP." Both rules are addressed to the economic historian as economist, they are as relevant to the measurement of the present as to

---

may reflect the original concern with paid employment, or a deeper concern that the available depreciation data reflect tax-accounting rules rather than any underlying reality). Another is that the value of free goods cannot be gauged by their market price, sending us back to Dupuit. In the presence of free goods, it may be noted, our "GDP" figures vary in the wrong direction altogether: the opportunity set of people who must arm themselves against a threat, or heat their houses, is smaller than that of those who have no need to, *ceteris paribus*, but their "GDP" is greater. Our measure should grow, and not decline, as we approach Eden, or Marx's communism. A third is that veblenesque consumption externalities (footnote 39) may well validate the essential message of Easterlin (1974), despite the ambiguity of the evidence the author adduces (self-rated "happiness," again interpreted with no sense of what words actually do).

that of the past: their aim is to ensure that we use our materials to erect a satisfactory structure, a structure suited to the purpose at hand.

Together, these rules suggest a natural taxonomy, a sequence of generations. The historical national accounts produced by the pioneers, "first-generation" estimates in the obvious chronological sense, ignored all five rules: they are usefully considered "first-generation" estimates in the here relevant methodological sense as well.[47] The "second-generation" estimates take the first step on the road to a more verisimilar reconstruction of the past: they focus on the elementary "data," they observe the first three rules but, for the moment, only these. These estimates are a corpus of disaggregated production series covering the entire economy (at least as defined by the standard GDP "production boundary"): if we distinguish n elementary activities over t years, the second-generation effort yields n elementary physical-product series and $N = n \times t$ elementary year-specific estimates of physical product. "Third-generation" estimates also observe the fourth rule, that of applying a common deflator to the full set of current-price value added estimates; and the latter set is generated from the N second-generation estimates of physical product with the aid of a second set of N estimates, the (t) year- and (n) activity-specific estimates of current-price value added per unit of product. With the "fourth-generation" estimates, finally, we would move beyond conventional "GDP" and actually measure the economy's aggregate product: a consummation devoutly to be wished.

These generations succeed each other slowly, for some are as long-lived as the patriarchs. The supersession of the Kuznets-style index of paid-employment-generating production by a proper measure of the economy's aggregate product, the fourth-generation step for us historians, takes the prize: urgently required by both economists and policy-makers, after seventy-odd years it is still in the murky, indefinite future. The third generation should in contrast be brief enough, say a matter of a few years, as the necessary year- and product-specific estimates of current-price value added per physical unit do not seem to present particular difficulties; but the second generation is also dreadfully long-lasting.

The evidence here comes from the estimates for Italy from Unification (1861) to the Great War, in particular from my own reconstruction of industrial production over that half-century (Fenoaltea 2015b–h, 2019c). It takes time to identify the production data, time to acquire the technical and institutional knowledge needed to vet it, and time finally to transform it into the desired time series: time, but not an inordinate amount of time. But where production data are not available, where the surviving evidence is indirect and not immediately apparent, where the estimates themselves converge to their final form only through repeated iterations, months stretch into years, and years into decades.[48]

---

[47] In point of fact, my *regole* emerged from the first-generation corpus I had before me, from the evidence of what was, to my eyes, inappropriate procedure (e.g., Fenoaltea 1969, 1982). I concentrated of course on the Italian estimates, estimates constructed using the then-unchallenged international standard methodology, estimates criticized because they used it and not because they failed to: the claim that "the 'Istat-Vitali series' … were considered not up to international standards" utterly misrepresents the literature.

[48] The estimates for the construction industry, for example, took three years of full-time research. One year yielded the initial set of estimates, for railway construction (from mileage data), for other infrastructure (from the public-works budgets), and for private buildings (from buildings-tax data); but (much like Istat's) my estimated aggregate value added in 1911 fell far short of that implied by that year's census. Searching for what I (and Istat before me) may have missed, I discovered that many public works appeared in other budgets (e.g., schools, in the education budget); a second year through

Because the second-generation effort is so time-consuming, and after half a century still incomplete, we cannot move quickly from the first generation to the third, to our proper reconstruction at least of past GDP; in the interim, to satisfy our curiosity, we periodically take stock, simply combining what we have into a preliminary set of second-generation national accounts, accounts that are in fact preliminary for a double set of reasons. On the one hand, obviously, the aggregate estimates are preliminary because they combine the second-generation series compiled for only a subset of activities, and crude indices for the rest; as the work progresses and such indices are replaced by proper sets of elementary series, these second-generation national accounts are periodically revised (e.g., Fenoaltea 2005, 2017b). On the other hand, the second-generation aggregates are preliminary because the elementary physical product series are combined very simply, with no more than base-year estimates of value added per physical unit: merely *n* such estimates, rather that the *N* required for the third generation.[49] In this second sense the second-generation aggregates remain preliminary even if calculated from a full set of proper elementary physical-product series, preliminary because they incorporate *de facto* activity-specific rather than general deflation, preliminary in short because fail to respect our fourth rule.

The second-generation national accounts, sneak previews justified only by their reduced information needs, fail to recognize changes in relative prices; and that means that they let us see the economy's past performance only in a distorting mirror, as illustrated by the John-and-Judy show described above. But we know how relative prices change, mainly in response to differences in productivity growth, and approximate measures of such differences are readily obtained: we know in general *how* our mirror distorts, our second-generation aggregates are naturally accompanied by corrected series, not yet third-generation estimates but at least controlled conjectures as to their likely shape.[50] After half a century of work, the Italian literature can offer no more than that: for the second-generation, only interim estimates, for the third, no more than conjectures.

---

the sources produced estimates that gave for 1911 a total that was higher than before, but still not high enough to match the census. The still-missing component turned out to be privately financed non-railway infrastructure (e.g., the hydroelectric dams built by the power utilities); and this was recovered with a third year of work. The time absorbed by other major sectors is not so readily established, as they were studied, set aside, and then returned to; but such complex and largely ill-documented sectors as textiles or engineering each easily absorbed, over time, half a dozen full-time-equivalent years. Gerschenkron (1962 [1955]), his first-generation index of Italian industrial production, incorporated a few dozen series mostly found ready-made in the *Annuario*; it should not have taken more than a few months.

[49] The second-generation elementary "real value added" series are base-year-value added-weighted physical product series, in the above notation $v_{rito} = (v_{io}/Q_{io})\ Q_{it}$. Expanding this last, $v_{rito} = (v_{io}/Q_{io})\ Q_{it} = (p_{it}Q_{it} - z_{it}R_{it})/[(p_{it}Q_{it} - z_{it}R_{it})/((p_{io} - z_{io}(R_{io}/Q_{io}))\ Q_{it})]$: the implicit value-added deflator, in square brackets, is obviously activity-specific. The calculated "real" value-added relatives are therefore distorted – like the first-generation "double-deflated" relatives, albeit typically less so (Fenoaltea 1976).

[50] E.g., Fenoaltea (2011b). The second-generation aggregate is there accepted, but its *composition* is tentatively recalculated to allow for plausible trends in relative prices, yielding conjectural time series for the economy's major sectors not *at 1911 prices*, but *at the 1911 price level*.

3.2 *Corpus delicti*

The pioneering historical national accounts, for the major European economies, were a mere handful, easily reviewed (Fenoaltea 1982). Since then, the corpus of historical national accounts has exploded – the Maddison project website provides no fewer than twenty-three such national sets, along with Angus Maddison's own famous world-spanning reconstruction – and I cannot claim to have mastered it.[51] But as far as I have been able to tell, all are absolutely standard replicas of the pioneers' work, the methodology appears not only unchanged but unquestioned.[52] Those accounts appear to have been done quickly, ignoring the above time-consuming rules, and when done, *done*, as would have satisfied Macbeth.[53] Above all, they appear to have been produced *serenely*, with none of Macbeth's own anguished soul-searching: produced with the confidence and untroubled conscience bestowed by an unquestioning faith in orthodoxy, the orthodoxy maintained by institutional design, by the system of heresy-rejecting peer review that Galileo also encountered.[54]

This mass failure to follow *le regole dell'arte* scandalizes *me*, because my professional ethics are those of an old Italian craftsman; but quality standards can arguably be excessive, Italy's seventeenth-century industrial decline has in fact been attributed precisely to attitudes like mine (Cipolla 1959, Rapp 1975, but see also Fenoaltea 1999). My readers, I suspect, are cut from a different cloth: pragmatic, Missouri-born, with no thought of rending their clothes, they are waiting to be shown that the first-generation corpus is not good enough *in practice*, that the later generations are worth the effort and the time.

The case has in part already been made. For one, as recalled above, inadequate disaggregation means that the first-generation estimates cannot help us distinguish between alternative interpretations ("hypotheses") consistent with the same aggregate. Again, the failure to adopt a common deflator means that the first-generation "real" estimates cannot capture the economy's evolving structure: such estimates reveal nothing under the aggregate's surface, they are practically exoskeletal.

The remaining question is whether the first-generation estimates adequately represent at least the shape of that aggregate, the minimal aggregate, conventional GDP. In principle nothing can be said, it is a straightforward matter of luck; and even in practice very little can be said, as the only evidence to hand is that provided by the Italian case. That evidence is here summarized by Figure 1, which plots together the annual per-capita product attributed to

---

[51] The national accounts are collected at https://www.rug.nl/ggdc/historicaldevelopment/na/, Maddison's figures at http://www.ggdc.net/maddison/oriindex.htm (accessed December 2019). I will not repeat, yet again, my reasons for questioning both Maddison's competence, and his intellectual honesty; see most recently Fenoaltea (2019a, footnote 49).

[52] In the international corpus the only exception I have found is the Egyptian reconstruction by Ulaş Karakoç, who describes it as "first-generation" work in the above taxonomy (Karakoç 2017, p. 61); the others appear to produce first-generation work *sans le savoir*.

[53] Some efforts mobilized a large team and absorbed man-decades rather than mere man-years; but experience is not additive, a number of cooperating apprentices can no more substitute for a single master than nine cooperating women can deliver a baby in one month.

[54] Only a religious commitment to orthodoxy can account for the collection not just of rejections, but of spluttering, *indignant* rejections, I have collected over the years.

post-Unification Italy by the first-generation Istat-Vitali estimates (Fuà 1969), by their amended version from Maddison (1991), by the preliminary second-generation estimates (Fenoaltea 2005), by the sesquicentennial estimates (Baffigi 2011), and by my own recent, revised second-generation estimates (Fenoaltea 2017b).[55]

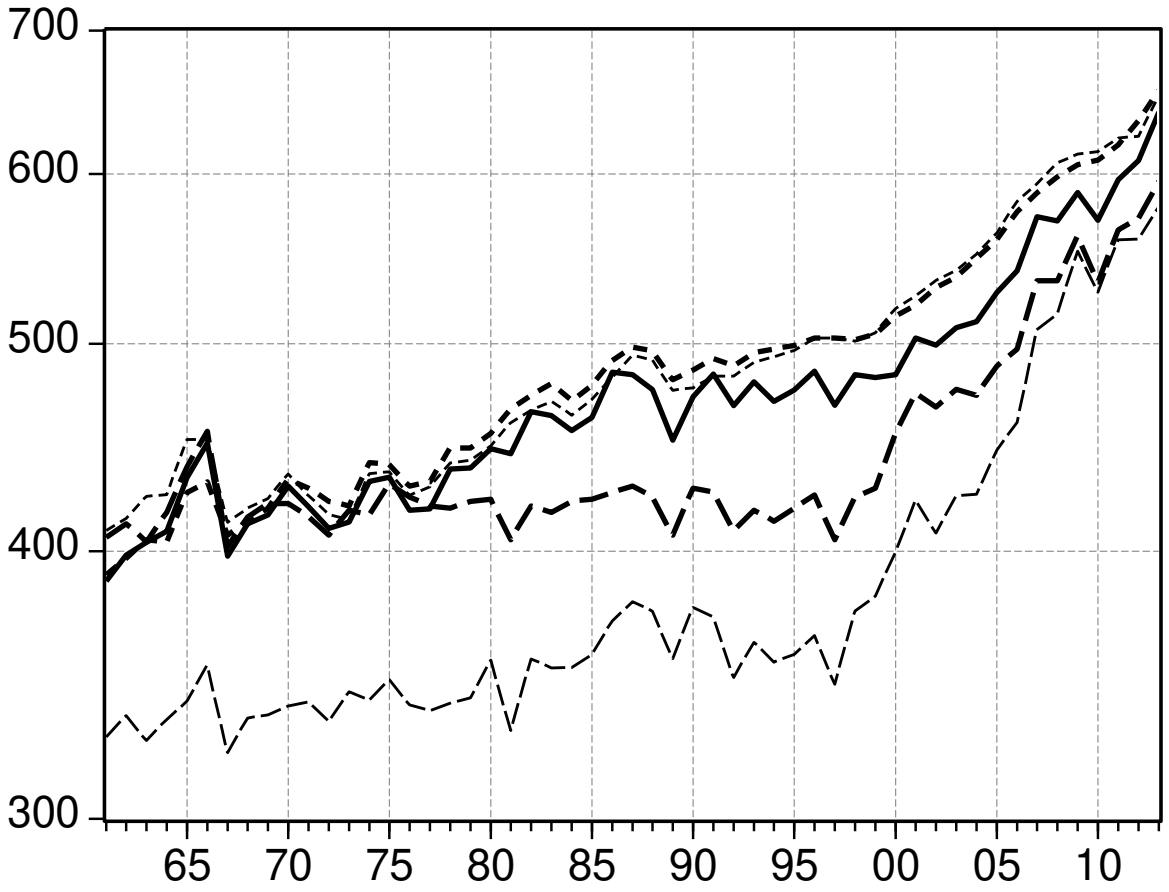
Ugliness too is in the eye of the beholder, and I have a dog in this fight; but it is a fair statement that even the earliest second-generation estimates prompted the profession to look for stories very different from those suggested by the estimates of the first generation (Fenoaltea 2011, ch. 1, and 2017a), the estimates produced with the international standard methodology. The first Italian estimates are as good, or as bad, as (the bulk, at least, of) those now available for other countries: the to my mind inescapable conclusion is that the extant corpus of historical national accounts is to say the least unreliable.

The broad corpus of historical national accounts is very much first-generation stuff, of unknown, but certainly suspect, fidelity to (what we can know of) the facts. Economists, and the cliometricians they train, hold measurement in contempt, and happily exploit these *comptes fantastiques* – the more *fantastiques*, in fact, the further back they go – in studies that span the ages and the continents. These reconstructions, and the analyses that famously exploit them, are neither science nor history; we can appreciate them, perhaps, as literature.

[55] Baffigi's estimates (2011) are close kin to my earlier ones (2005); my latest (2017) differ primarily because they allow for harvest fluctuations (increasing short-term variability), and, more significantly, because they remove hitherto unnoticed gross errors in the estimates for the services he and I inherited, to a varying degree, from colleagues who worked specifically on that sector.

Figure 1

Per-capita income at 1911 prices,
Italy, 1861-1913: alternative estimates (lire)



first-generation estimates (Istat-Vitali 1969)
Maddison estimates (1991)
preliminary second-generation estimates (2005)
sesquicentennial estimates (Baffigi 2011)
revised second-generation estimates (2017)

APPENDIX A. RECKONING WITH MAINTENANCE

A.1 Industry v. services

The quantitative reconstruction of the past is organized by an accounting framework. The *ISIC* (p. 29) treats maintenance with the lack of uniformity, not to say common sense, we have come to expect of the United Nations: the repair of motor vehicles and the repair of personal and household goods are considered services (respectively trade, division 45, and other services, division 95), the repair of other machinery and the repair of buildings and other structures are considered industry (respectively manufacturing, division 33, and construction, division 43).

Logically, surely, maintenance is either the one thing or the other. As to which it is, a repairman surely provides a maintenance service; and equally surely that is entirely irrelevant, for factors of production always provide services, and the relevant criterion is whether or not those services yield a commodity, a good that exists in its own right and can be stored and resold.[56] New production of goods takes commodities with certain physical characteristics and transforms them into commodities with more desirable ("valuable") physical characteristics, maintenance takes commodities with certain physical characteristics and transforms them into commodities with more desirable ("valuable") physical characteristics; there is no meaningful difference between the two, the only sensible solution is to consider maintenance activity uniformly as "industry" rather than "services."

A.2 Net production/capital formation v. cost of production

All this involves only what is counted where; but maintenance involves a more serious issue, the issue of whether it should ultimately be counted at all. The issue turns on whether maintenance is considered production (of reconditioned durables), or a cost of production (of the goods and services that use those durables). The United Nations is, on the subject, unusually guarded: the *SNA* asserts that ordinary maintenance is to be considered a cost of production, while more extensive maintenance is to be considered production – but grants the possible objection that all maintenance should be considered production.[57]

The objection appears valid: painting a hull yields a painted hull, to consider it production of a durable good (and capital formation) in some cases and a cost of production (of seaborne transportation) in others seems frankly quixotic. Maintenance produces a commodity (a newly reconditioned good) that wasn't there before, just as new production produces a commodity (a brand-new good) that wasn't there before; and the former

---

[56] Services proper cannot be, which is why the providers of transportation or medical services can price-discriminate as commodity-producers cannot.

[57] "Ordinary maintenance and repairs undertaken by enterprises to keep fixed assets in good working order are treated as intermediate consumption. However, major improvements, additions or extensions to fixed assets, both machinery and structures, which improve their performance, increase their capacity or prolong their expected working lives count as gross fixed capital formation. In practice it is not easy to draw the line between ordinary repairs and major improvements, although the SNA provides certain recommendations for this purpose. Some analysts, however, consider that the distinction between ordinary repairs and maintenance and major improvements and additions is neither operational nor defensible and would favour a more 'gross' method of recording in which all such activities are treated as gross fixed capital formation" (*SNA*, pp. 8–9).

commodity is clearly a durable good, just as the latter is. Logically, maintenance is industrial production, and capital formation; *secundum non datur*.[58]

A.3 The "logic" of the United Nations

The national accounts were less concerned with production in general than with paid-employment-generating production in particular; one consequence was the neglect of household production (and the attendant consumption), essentially traditional "women's work," the women's work that is never done. A further consequence was the neglect of households' consumption of the services of the durables they owned (save housing), and, derivatively, of the corresponding stock of durables. The accounting framework does not allow households to invest in durable goods (save housing, as will not be repeated): everything households purchase is treated as a non-durable, households are assumed to consume not the transportation services of their vehicles but the vehicles themselves, to eat not just the food in the refrigerator but the refrigerator too.

Choices are path-dependent. In the national accounts firms are recognized as owners of durable goods, their additions to their stocks are recognized as investment; the maintenance of their assets could be considered industry, as seemed sensible (there is not much difference between the shops in which locomotives are assembled, and those in which they are disassembled and reassembled), and it could as noted be taken to produce either a final good or, as the United Nations prefers, an intermediate good consumed in the production of other goods and services. But maintenance could not be taken as an intermediate good in the production of the services of household durables: because these durables were altogether ignored, there was no imputed value-of-product from which this input could be deducted. But the product of firms that maintain household durables had to be counted *somewhere*; and the "clever" solution was to pretend that those firms' activity did not yield (altered) commodities, that those firms were not "industry" but "services." And that is where, in the production accounts, they are supposedly to be counted: a blacksmith repairing a farmer's plow works in industry, but the moment he turns to repairing a household's andirons he migrates to a different sector, only to return when that job is done. I kid you not.

[58] Logically, too, the "durability" of a good depends not on a convention related to how long it lasts ("one year"), but on the facts of the case: whether use implies the (destructive) consumption of the good itself (food, a raw material, an hour of labor), which disappears, or the consumption (only) of the services of the good, which survives (a refrigerator, a tool, a laborer). From this perspective the exclusion of clothes-washing from durable-good-producing industry is arguably not an exception, as while the textile products themselves are clearly durables, their spotlessness and odorlessness are not: the national income and product accounts are a hair-splitter's playground.

REFERENCES

Annuario (year). Direzione generale della statistica. *Annuario statistico italiano* (1878 ff.).

Arrow, K. J. (1974). The measurement of real value added. In P. A. David and M. W. Reder, eds., *Nations and households in economic growth: essays in honor of Moses Abramowitz* (New York: Academic Press), pp. 3–20.

Cipolla, C. M. (1959). Il declino economico dell'Italia. In C. M. Cipolla, ed., *Storia dell'economia italiana* (Turin: Einaudi), pp. 605–624.

David, P. (1962). The deflation of value added. *Review of Economics and Statistics* 44, 148–155.

Easterlin, R. A. (1974). Does economic growth improve the human lot? Some empirical evidence. In P. A. David and M. W. Reder (eds.), *Nations and households in economic growth: essays in honor of Moses Abramowitz* (New York: Academic Press), pp. 89–125.

Easterlin, R. A. (2004). *The Reluctant Economist. Perspectives on Economics, Economic History, and Demography*. Cambridge: Cambridge University Press.

Einaudi, L. (1936a). Lo strumento economico nella interpretazione della storia. *Rivista di storia economica* 1, 149–158.

Einaudi, L. (1936b). Teoria della moneta immaginaria nel tempo da Carlomagno alla rivoluzione francese. *Rivista di storia economica* 1, 1–35.

Feinstein, C. H. (1972). *National income, expenditure and output of the United Kingdom, 1855-1965*, Studies in the national income and expenditure of the United Kingdom, vol. 6 (Cambridge: Cambridge University Press).

Fenoaltea, S. (1967). Public policy and Italian industrial development, 1861–1913. Unpublished Ph.D. dissertation, Harvard University.

Fenoaltea, S. (1969). Decollo, ciclo, e intervento dello Stato. In A. Caracciolo, ed., *La formazione dell'Italia industriale* (Bari: Laterza), pp. 95–114.

Fenoaltea, S. (1972). Railroads and Italian industrial growth, 1861-1913. *Explorations in Economic History* 9,. 325-352.

Fenoaltea, S. (1976). Real value added and the measurement of industrial production. *Annals of Economic and Social Measuremen*t 5, 111–137.

Fenoaltea, S. (1981). The slavery debate: a note from the sidelines. *Explorations in Economic History* 18, 304–308.

Fenoaltea, S. (1982). The growth of the utilities industries in Italy, 1861-1913. *Journal of Economic History* 42, 601-627.

Fenoaltea, S. (1984). Slavery and supervision in comparative perspective: a model. *Journal of Economic History* 44, 635–668.

Fenoaltea, S. (1986). Public works construction in Italy, 1861-1913. *Rivista di storia economica* 3, *International issue* (1986), 1–33.

Fenoaltea, S. (1988). International resource flows and construction movements in the Atlantic economy: the Kuznets cycle in Italy, 1861–1913. *Journal of Economic History* 48, 605–638.

Fenoaltea, S. (1999). Lo sviluppo economico dell'Italia nel lungo periodo: riflessioni su tre fallimenti. In P. Ciocca and G. Toniolo (eds.), *Storia economica d'Italia, vol. I. Interpretazioni* (Milan-Rome-Bari: Cariplo/Laterza), pp. 3–41.

Fenoaltea, S. (2001). *Lezioni di economia politica.* http://fenoaltea.altervista.org/ (accessed August 2019).

Fenoaltea, S. (2003). Notes on the rate of industrial growth in Italy, 1861–1913. *Journal of Economic History* 48, 695–735.

Fenoaltea, S. (2005). The growth of the Italian economy, 1861–1913: preliminary second-generation estimates. *European Review of Economic History* 9, 273–312.

Fenoaltea, S. (2010). The reconstruction of historical national accounts: the case of Italy. *PSL Quarterly Review* 63, no. 252, 77–96.

Fenoaltea, S. (2011a). *The reinterpretation of Italian economic history: from Unification to the Great War* (New York: Cambridge University Press).

Fenoaltea, S. (2011b). On the structure of the Italian economy, 1861–1913. *Rivista di storia economica* 27, 61–72.

Fenoaltea, S. (2015a). Industrial employment in Italy, 1911: the burden of the census data. *Rivista di storia economica* 31, 225–246.

Fenoaltea, S. (2015b). Italian industrial production, 1861–1913: a statistical reconstruction. B. The extractive industries. *Carlo Alberto Notebooks* n. 413.

Fenoaltea, S. (2015c). Italian industrial production, 1861–1913: a statistical reconstruction. C. The non-metallic mineral products industries. *Carlo Alberto Notebooks* n. 414.

Fenoaltea, S. (2015d). Italian industrial production, 1861–1913: a statistical reconstruction. D. The chemical, coal and petroleum products, and rubber industries," *Carlo Alberto Notebooks* n. 415.

Fenoaltea, S. (2015e). Italian industrial production, 1861–1913: a statistical reconstruction. E. The metalmaking industries," *Carlo Alberto Notebooks* n. 416.

Fenoaltea, S. (2015f). Italian industrial production, 1861–1913: a statistical reconstruction. F. The engineering industries," *Carlo Alberto Notebooks* n. 419.

Fenoaltea, S. (2015g). Italian industrial production, 1861–1913: a statistical reconstruction. J. The utilities industries," *Carlo Alberto Notebooks* n. 421.

Fenoaltea, S. (2015h). Italian industrial production, 1861–1913: a statistical reconstruction. K. The Construction Industries," *Carlo Alberto Notebooks* n. 422, Moncalieri, 2015.

Fenoaltea, S. (2016). Fenoaltea on industrial employment in 1911: a rejoinder. *Rivista di storia economica* 31, 113–117.

Fenoaltea, S. (2017a). The fruits of disaggregation: the engineering industry, tariff protection, and the industrial investment cycle in Italy, 1861–1913, *Banca d'Italia – Economic History Working Papers* n. 41.

Fenoaltea, S. (2017b). The growth of the Italian economy, 1861–1913: revised second-generation production-side estimates, *M.P.R.A.* n. 87962.

Fenoaltea, S. (2018). Italy in the market for seagoing vessels, 1861−1913: domestic production, imports, and exports. *Rivista di storia economica* 34, 293−355.

Fenoaltea, S. (2019a). Spleen: the failures of the cliometric school. *Annals of the Fondazione Luigi Einaudi* (in press).

Fenoaltea, S. (2019b). Of economics and statistics: the "Gerschenkron effect." *PSL Quarterly Review* 72, no. 290, 195−205.

Fenoaltea, S. (2019c). Italian industrial production, 1861–1913: a statistical reconstruction. H. The textile, apparel, and leather industries. *Carlo Alberto Notebooks* n. 583.

Fuà, G., ed. (1969). *Lo sviluppo economico in Italia*, vol. 3 (Milan: Franco Angeli).

Fuà, G. (1993). *Crescita economica. Le insidie delle cifre* (Bologna: Il Mulino).

Gerschenkron, A. (1955). Notes on the rate of industrial growth in Italy, 1881–1913. *Journal of Economic History* 15, 360–375.

Gerschenkron, A. (1962 [1955]). Description of an index of Italian industrial development, 1881−1913. In A. Gerschenkron, *Economic backwardness in historical perspective* (Cambridge MA: Harvard University Press), pp. 367–421.

Gordon, R. J. (2016). *The rise and fall of American growth: The U.S. standard of living since the Civil War* (Princeton: Princeton University Press).

ISIC: United Nations. *International standard industrial classification of all economic activities*, Statistical Papers, Series M, No. 4, Rev. 4. (New York: United Nations, 2008).

Karakoç, U. (2017). Industrial growth in interwar Egypt: first estimates, new insights. *European Review of Economic History* 22, 53–72.

Lerner, A. P. (1947). Review of *Measuring business cycles*, by Arthur F. Burns and Wesley C. Mitchell. *Journal of Economic History* 7, 222–226.

Nordhaus, W. D., and Tobin, J. (1972). Is growth obsolete? In W. D. Nordhaus and J. Tobin, eds., *Economic research: retrospect and prospect, Vol. 5, Economic growth* (New York: NBER), pp. 1–80.

Pollak, R. A. (1985). A transaction cost approach to families and households. *Journal of Economic Literature* 23, 581–608.

Rapp, R. (1975). The unmaking of the Mediterranean trade hegemony: international trade rivalry and the Commercial Revolution. *Journal of Economic History* 35, 499–525.

Reddito nazionale: Istat (Istituto centrale di statistica). *Indagine statistica sullo sviluppo del reddito nazionale dell'Italia dal 1861 al 1956*, *Annali di statistica*, serie VIII, vol. 9 (Rome: Istat, 1957).

Rilevazioni statistiche: Istat (Istituto centrale di statistica). *Le rilevazioni statistiche in Italia dal 1861 al 1956*, *Annali di statistica*, serie VIII, vols. 5–8 (Rome: Istat, 1957–59).

Rivista mineraria: Corpo delle miniere. *Relazione sul servizio minerario* (1877–82); *Rivista del servizio minerario* (1883 ff.).

Robbins, L. (1939). *The Economic Basis of Class Conflict* (London: Macmillan).

Romer, C. (1986). Is the stabilization of the postwar economy a figment of the data? *American Economic Review* 76, 314–334.

Sims, C. A. (1969). Theoretical basis for a double deflated index of real value added. *Review of Economics and Statistics* 51, 470–471.

SNA: United Nations. *System of national accounts 2008* (New York: United Nations, 2008).

Sommario. Istat (Istituto centrale di statistica). *Sommario di statistiche storiche italiane, 1861–1955* (Rome: Istituto poligrafico della Stato, 1958).

Veblen, T. (1899). *The theory of the leisure class: an economic study in the evolution of institutions* (New York: Macmillan).