Using Online Catalogs to Estimate Economic Development in Classical Antiquity

Charles de Dampierre*, Valentin Thouzeau and Nicolas Baumard

Institut Jean Nicod, Département d’études cognitives, ENS, EHESS, PSL University, CNRS

Abstract

Despite significant progress, economic development in Classical Antiquity remains difficult to study: proper economic data (e.g. agricultural production, wages) are scarce, estimates of urbanization, GDP or population remain highly uncertain, and the use of indirect markers of development such as shipwrecks or coin hoards is limited. Here, we propose a different approach based on the production of immaterial works (e.g. poems, philosophical treatises, musical pieces, scientific work). Immaterial works require time, energy, resources, and human capital to be produced, disseminated and appreciated, and thus indirectly reflect a wide range of economic processes. Moreover, their survival rate tends to be higher because they can be abstracted from their initial material incarnation (e.g. scrolls, manuscripts) and preserved throughout the centuries. We build a large database of cultural producers (painters, scientists, etc.) that exist in online catalogs (Library of Congress ID, GND ID, VIAF ID, Iranica ID etc) and create an estimate of immaterial production that is robust and consistent across cultures and sources. We show that immaterial production in Ancient Greece and Ancient Rome is closely related to economic development, and reveals important phases of economic development. Overall, immaterial production provides new insights into the roots and the evolution of economic development in the very long run in Classical Antiquity.

Keywords
humanities, online catalogs, computer science, Ancient Greece, Ancient Rome, immaterial productivity, economic development, Classical Antiquity

1. Introduction

Economic development is key to understanding the dynamic of ancient societies [1, 37, 12]. Yet, estimating economic development in pre-industrial societies is often difficult. One approach, initiated by Maddison [29] and subsequently improved by a large number of economists [3, 2, 13], has been to quantify the Gross National Product (GNP) using estimates of the agricultural and industrial production of a given society. This approach, however, is of limited use for the earliest periods, such as Classical Antiquity, where data is more fragile [41, 38]. Another solution consists of using a single marker of development such as urbanization [4, 5, 28], building construction [44, 46, 48], or shipwrecks [11]. Ian Morris notably built on these data to construct
a composite index that aggregates estimates of energy capture, urbanization, communication technologies, and warfare technologies [36]. These methods have considerably improved our assessment of the dynamics of pre-modern civilizations. However, they remain limited, both in their temporal resolution (typically, a century) and in their scope (often limited to the most developed periods but see recent progress in quantitative epigraphy [15, 22, 21]).

In this article, we propose a different approach based on immaterial production. immaterial production (e.g. poems, philosophical treatises, musical pieces) represents an enormous amount of information about social and economic development. Like all goods, cultural goods require resources to be produced (time, materials, technology, energy). To put it another way, behind every Plato, every Virgil, and behind the hundreds of thousands of other producers in our database, there are infrastructures, schools, teachers, consumer markets, institutions, that enable a significant level of immaterial production. In line with this idea, a higher level of economic and social development is associated with a higher level of immaterial production. For instance, scientific production, a sub-field of immaterial production, is well correlated with economic and social development, both in pre-modern [10] and modern societies [27, 30, 33]. Similarly, the production of manuscripts and books is strongly correlated with economic development [8, 9, 50]. Methodologically, using immaterial production as an indicator of the economic development of civilizations presents three advantages. First, immaterial works tend to survive very well throughout history. Cultural works have two modes of existence [14, 16]: a material mode (the physical support of the work, e.g., scrolls, parchment, codices) and an immaterial mode (the work itself, e.g., texts, sheet music, ideas) that can survive the destruction of the material support. For example, Kestemont et al. [26] recently demonstrated that while more than 80% of German medieval documents (e.g. a copy of Tristan and Iseult) have been lost over time, a much smaller proportion (about 20%) of their immaterial version (i.e. the text of Tristan and Iseult) are irretrievably lost. Second, immaterial production has been extensively studied by scholars in the humanities. immaterial producers have been the subject of numerous academic studies concerning their dating, authorship, and cultural significance (innovativeness, originality, sources of inspiration). To conclude, immaterial works can provide extensive, robust and high-quality data on economic development.

2. Material and Methods

2.1. Material: Online catalogs

Online catalogs are managed by different types of institutions. First, by national specialized institutions dedicated to collecting and curating information: the Library of Congress is the most known example and comprises an extensive assortment of individuals’ information, books, periodicals, manuscripts, maps, and audiovisual materials for a total of around 170 million items. Some other notable examples of national catalogs’ identifiers include the GND ID (Germany), the BNF ID (France), the ctext (Chinese Text) data entity ID, the National Library of Israel ID, Iranica ID (Iran), the National Library of Korea ID (Korea), TDV İslam Ansiklopedisi ID (Turkey) for the Islamic Encyclopedia created by the Turkish religion authority Diyanet (Erünsal, 2007) and many others. Second, they are managed by international institutions such as the Virtual Authority File (VIAF), the International Standard Name Identifier (ISNI) or the Consortium of
European Research Libraries (CERL). Third, they are managed by collective institutions such as Wikipedia or Wikidata \[47\].

Table 1
Number of individuals in the top catalogs

<table>
<thead>
<tr>
<th></th>
<th>Greek World</th>
<th></th>
<th>Latin World</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>n</td>
<td>%</td>
<td>n</td>
<td>%</td>
</tr>
<tr>
<td>Google Knowledge Graph ID</td>
<td>1034</td>
<td>99</td>
<td>413</td>
<td>98</td>
</tr>
<tr>
<td>VIAF ID</td>
<td>838</td>
<td>81</td>
<td>344</td>
<td>82</td>
</tr>
<tr>
<td>GND ID (Germany)</td>
<td>764</td>
<td>73</td>
<td>326</td>
<td>77</td>
</tr>
<tr>
<td>CERL Thesaurus ID</td>
<td>738</td>
<td>71</td>
<td>320</td>
<td>76</td>
</tr>
<tr>
<td>CIRIS author ID (France)</td>
<td>573</td>
<td>55</td>
<td>293</td>
<td>69</td>
</tr>
<tr>
<td>ISNI</td>
<td>506</td>
<td>49</td>
<td>275</td>
<td>65</td>
</tr>
<tr>
<td>Treccani ID (Italy)</td>
<td>476</td>
<td>46</td>
<td>242</td>
<td>57</td>
</tr>
<tr>
<td>Library of Congress authority ID (US)</td>
<td>466</td>
<td>45</td>
<td>284</td>
<td>67</td>
</tr>
<tr>
<td>Perseus author ID</td>
<td>457</td>
<td>44</td>
<td>249</td>
<td>59</td>
</tr>
<tr>
<td>IdRef ID (France)</td>
<td>431</td>
<td>41</td>
<td>233</td>
<td>55</td>
</tr>
<tr>
<td>Nationale Thesaurus voor Auteurs ID (Netherlands)</td>
<td>421</td>
<td>40</td>
<td>246</td>
<td>58</td>
</tr>
<tr>
<td>FAST ID</td>
<td>402</td>
<td>39</td>
<td>239</td>
<td>57</td>
</tr>
<tr>
<td>Treccani’s Enciclopedia Italiana ID (Italy)</td>
<td>393</td>
<td>38</td>
<td>198</td>
<td>47</td>
</tr>
<tr>
<td>National Library of Israel J9U ID (Israel)</td>
<td>392</td>
<td>38</td>
<td>210</td>
<td>50</td>
</tr>
<tr>
<td>Bibliothèque nationale de France ID (France)</td>
<td>364</td>
<td>35</td>
<td>215</td>
<td>51</td>
</tr>
<tr>
<td>Oxford Classical Dictionary ID</td>
<td>354</td>
<td>34</td>
<td>210</td>
<td>50</td>
</tr>
<tr>
<td>Vatican Library VcBA ID (Vatican City)</td>
<td>349</td>
<td>34</td>
<td>209</td>
<td>50</td>
</tr>
<tr>
<td>Total unique individuals (among all catalogs)</td>
<td>1040</td>
<td>100</td>
<td>422</td>
<td>100</td>
</tr>
</tbody>
</table>

The richness and depth of catalogs allow researchers to access a wealth of primary and secondary sources, enabling them to explore a variety of social phenomena and conduct interdisciplinary analyses. Another main advantage of catalogs are the rigorous standards and principles they are built on, ensuring high-quality metadata for the resources they contain. There has been ongoing works over the years to unify sources, such as author names, publication dates, subject headings, and descriptive annotations. This early quality control promotes data consistency and accuracy, thereby enhancing the reliability and reproducibility of computational social science research \[45\]. To retrieve authoritative data on individuals and their works, we employed Wikidata as a gateway within the domain of online catalogs. Wikidata encompasses both collaborative and automated components, the latter of which extracts information from diverse catalogs and harmonizes knowledge. For example, a notable figure such as Plato is directly associated with over 200 catalogs worldwide through unique identifiers on Wikidata. There are a total of 548 catalogs in our database distributed in 60 countries on all continents.

In order to identify immaterial producers, we used Wikidata’s advanced ontology system. Each new ontology (for example the creation of a category such as "painter" or "geographer") must be validated by the Wikidata community to ensure the validity of the information. This ontology system consists of a hierarchy of categories and sub-categories. For example, a
"Painter" is a subcategory of "Visual Artist" which itself is a subcategory of "Artist". This system of assigning occupations ensures that we build the dataset without any particular hypothesis in mind, leaving some of the methodological decisions to the contributors and administrators in the Wikidata community. We selected individuals who have at least one occupation that is a subcategory of ‘writer’, ‘scientist’ or ‘artist’ as described in the Wikidata Ontology. Table 1 contains the cumulative count of individuals within the top catalogs.

2.2. Geographical and temporal granularity

To measure the immaterial productivity of each region, we take as a first rule that an individual contributes to a region if he or she is born in a place belonging to this region. This is technically convenient since the place of birth can be linked to geographic coordinates in Wikidata. Thus, even in cases where the name of the place of birth no longer exists, it is easily identifiable through its geographic data and can be extracted and then assimilated to a country existing today. When information about the birth place is not present, we consider the nationality of the individuals as indicated in Wikidata and linked it to today’s countries’ coordinates. We then manually classify some individuals accounting for the fact that individuals move from a region to another during their lifetime and when they move, they often contribute more to the region they move to than to the one where they were born. We divide the Mediterranean territory into two parts: the Latin World and the Greek World, using the Jireček line [19]. In the Latin World, we include individuals born between -300 and 500 in Italy and between -100 and 500 in Tunisia, Algeria, Morocco, Romania, Croatia, Serbia, Bosnia and Herzegovina, Slovenia, France, United Kingdom, Germany, Switzerland, Austria, Spain and Portugal. For the Greek World, we include individuals born between -800 and 500 in Ukraine, Albania, Montenegro, Kosovo, Turkey, Greece, Bulgaria, Romania, France (only from -800 to -300), Italy (only from -800 to -300), Libya, Egypt, Israel, State of Palestine, Lebanon, Syrian Arab Republic, Jordan, Cyprus and Iraq. We use the date of birth of individuals to locate them in time. We assume that a region has high immaterial productivity at a specific time if individuals born during that time are more likely to engage in creative or scientific activities during their lifetime than at other times.

2.3. Creating the immaterial production index

Using the catalogs, we create an index of immaterial production by counting the number of individuals for every region every 50 years. We first plot the immaterial index for the different catalogs that exist in our database in order to assess the consistency of the different sources across different regions of the world to avoid biases. We chose the top 20 catalogs assigned to a country and removed international catalogs. We found that the overall trends across catalogs are strongly correlated (mean person correlation $p = 0.80$ for the Greek World and $p = 0.88$ for the Latin World) indicating that different catalogs in different countries converge towards a fairly same pattern of the past (see Fig. 1).

We then plot a unique trend for the Greek World and the Latin World combining information from the diverse catalogs. In order to leverage every piece of information, we included all individuals found in all catalogs: if an individual exists in different catalogs, it is counted just
once to ensure that every individual has the same weight in the index, whatever its source: for instance, belonging to the Library of Congress or to the National Library of Korea weights the same. In order to smooth the trend, we use a loess regression. We display individuals on the graphs based on a proxy of notability ie the number of catalogs they are referred in (See Fig. 2, 4.5).
3. **Immaterial production captures economic development in pre-industrial civilizations**

Our results are in line with the dynamics highlighted by previous work. Firstly, we highlight an initial period of strong growth between 800 BCE and 500 BCE. From 800 to 300 BCE, economic indicators suggest that affluence increased in the Greek world (Greece, the Aegean, Sicily and Southern Italy) (See Fig. 2). During this period, population increased by 1500% (from 0.5 to 8.5 millions) [32, 38, 17], urbanization went from 0% to 25% [38, 17, 35] and GDP per capita probably tripled from $400 to $1200 [38, 36, 34]. These quantitative indexes are confirmed by other indicators such as hoards size or house size [36, 38]. For instance, while the median Greek house in the 9th century BC was small, it became both much bigger and much better built over the next 500 years. Looking at square footage alone, the change in the size of the median house is over 350%—from ca. 80 square meters to ca. 360 square meters [36, 38].

![Figure 2: Immaterial production in the Greek World. Every bar represents the number of individuals referenced in at least one online catalog every 50 years. The line (left axis) represents the lowess curve (span = 0.4) of the left axis. The top 30 notable individuals are located on the timeline at their birthdate (scatter point). The score of an individual (right axis) is computed by counting the number of catalogs an individual is referenced in.](image-url)
From 350 BCE onward, we observe a period of marked decline. In line with this result, historians have established that Core Greece experienced a period of decline [36, 38]. However, the decline was not uniform, and there is some economic growth in the newly Hellenized kingdoms in Egypt, Asia Minor or Syria with cities such as Alexandria (300,000 inhabitants in 300 BC, and 400,000 in 200 BC), Antioch (250,000 inhabitants) and Pergamon (200,000 inhabitants) (Morris, 2006b). This period of decline seems to end around 200 BCE. The level of economic development in the eastern Mediterranean remains relatively low until the end of the ancient period, with an optimum during the 2nd century CE, which is also found in the western part [18, 6]. Overall, the pattern we get from the study of immaterial production is strikingly similar to the pattern inferred by Ober using more qualitative indicators (see Fig. 3).

![Figure 3: Population and consumption estimates, core Greece, 1300 BCE–1900 CE from Ober, 2015][38]

Notes: Population estimated in millions. Median per capita consumption estimated in multiples of subsistence minimum. LBA = Late Bronze Age. EIA = Early Iron Age. EH = Early Hellenistic. LH = Late Hellenistic. ER = Early Roman. LR = Late Roman. EB = Early Byzantine. MB = Middle Byzantine. EO = Early Ottoman. LO = Late Ottoman. Ind = Independent Greek state. Core Greece = Territory controlled by the Greek state 1881–1912.

For the Latin world, we observe a period of strong growth from 300 BCE to 100 BCE (see Fig. 4). Quantitative archaeological series confirm this trend showing an increase in shipwrecks (a proxy of international trade), wood finds, coins in circulation, lead pollution, building inscriptions, water mills and papyruses [5, 20, 36, 31]. Thereafter economic development remained...
relatively stable, with some well-known periods of decline such as the 3rd century (Crisis of the Third Century), and some periods of relative affluence such as the second century (Nerva-Antonine dynasty) and the 4th century (Constantinian dynasty). This period of decline is well documented. International trade declined, size of building, public inscription, size of cows, etc. [5, 48]. After 200 BCE, all economic indicators (shipwrecks, building inscriptions, lead pollution, wood finds, coins in circulation, etc.) demonstrate that the level of development of the Mediterranean area started to decline importantly around the 3rd c. CE, although this was probably less important and delayed until the 4th c. in Egypt or Northern Africa [5, 48].

Figure 4: Immaterial production in the Latin World. Every bar represents the number of individuals referenced in at least one online catalog every 50 years. The line (left axis) represents the lowess curve (span = 0.4) of the left axis. The top 30 notable individuals are located on the timeline at their birthdate (scatter point). The score of an individual (right axis) is computed by counting the number of catalogs an individual is referenced in.

Overall, the pattern we get from the study of immaterial production is strikingly close to the patterns that can be inferred from material proxies such as shipwrecks (see Fig. 5) [49] or lead pollution [31, 11] (for a discussion, see [36, 40, 49]).

These estimations also allow us to compare the level of development of the Greek and Roman golden age. While the population of the Latin World during the 1st century was probably three of four times more important than the population of the Greek world during the Classical period (20 to 25 millions against 5 to 7 millions [42, 38, 7]), the immaterial production of the Latin world is lower than the immaterial production of the Greek world (see Fig. 6). This is consistent with recent works emphasizing the unique level of Greek development in terms of urbanization, per capita production, living standards during the Athenian period [37, 38, 43].
4. Discussion

In this paper we have used immaterial production to quantify economic development throughout classical antiquity. Our results are broadly consistent with the academic literature. In the Greek world, we observe a period of strong growth from 800 BCE to 450 BCE, followed by a steep decline and then stagnation for the rest of antiquity. In the Roman world, we also observe a period of strong growth from 300 BCE to 50 BCE, followed by a steep decline and then stagnation until the fall of the Roman Empire. These results suggest that immaterial production could indeed serve as a proxy for economic development. Such a correspondence between immaterial production and economic development replicates and extends to earlier periods the results of previous studies showing a relationship between scientific immaterial production and GDP per capita in modern Europe (1300-1850) [10].

Obviously, this study is not without limitations. First, it relies on catalogs and Wikidata, both of which are imperfect and biased sources. Second, the geographical granularity of the study is limited. For example, it would be interesting to follow more closely the trajectories of important economic regions such as Egypt or North Africa, which do not have the same economic trajectory as Greece or Rome. Thus, more work needs to be done to accurately link the location of individuals to the regions studied. In this study, we used the place of birth as a proxy for the region (Latin or Greek) in which people lived and worked. But individuals may have traveled (especially during the Imperial period), and this may bias our estimates. Third, the study is subject to survivor bias, focusing primarily on notable individuals that exist in modern catalogs, which doesn’t account for the entirety of immaterial production. This may
Figure 6: Immaterial production in the Greek (green) and Latin (red) worlds and overall immaterial production (dashed black line). Every bar represents the number of individuals referenced in at least one online catalog every 50 years. The lines (left axis) represent the lowess curve (span = 0.4). Individuals are located on the timeline at their birthdate (points). The top 60 notable individuals are displayed on the graph. The score of an individual (right axis) is computed by counting the number of catalogs an individual is referenced in.

Explain why our estimate of the peak economic development for the Latin world (100 BCE) is earlier than other methods (150 CE). Future research should address these limitations by systematically comparing estimates based on immaterial production and estimates based on material production [29, 3, 2]. Another possibility would be to explore methods for estimating individuals who are not remembered [26].

Regardless of the limitations of this methodology, it remains true that we observe a strong association between economic development and immaterial production. This relationship is all the more surprising because we did not control for population size. In principle, we should expect larger populations to produce more immaterial production. For example, we should predict that the immaterial production of the Greek world during the Hellenistic and Roman periods (when more than 20 million people lived in a Greek-speaking area) is necessarily more important than the immaterial production of the Greek world during the Classical period (when
about 5 million people lived in a Greek-speaking area). But this is not the case. This suggests that immaterial production is not driven by the size of the economy (GDP), but by the productivity of individuals (GDP per capita).

The one surprising result we observe is the fact that, in the Latin world, economic decline seems to begin very early, as early as the first century BC, and not later, during the first century CE [5] or at the end of the Nerva-Antonine dynasty [48, 40]. There are several ways to interpret these results. The first is that our data are biased towards the first century BC. Indeed, this period corresponds to the golden age of Latin literature (e.g. Cicero, Julius Cesar, Virgil, Ovid). Our index could be biased by the fact that these authors were canonized by subsequent generations, notably during the medieval period. Because works by non-canonized authors were less likely to be preserved, we might have an important bias in favor of the golden age of Latin literature.

This explanation, however, is not really an explanation at all. It does not explain why the authors of the golden age were canonized. In particular, it does not explain why Latin culture was unable to demonstrate an equivalent or higher level of cultural dynamism in subsequent periods producing authors that later generations would have wanted to include in the canon. Indeed, there’s nothing deterministic about literary canonization or the emergence of “classical” periods in literary history. For instance, Japanese literature experienced a first golden age during the Heian period (794 to 1185), which saw the emergence of some great classics (e.g. The tale of Gengi) [23, 24, 25]. However, Japanese literature subsequently experienced other periods of cultural dynamism, such as the Edo period (e.g. Ihara Saikaku, Matsuo Bashō, Chikamatsu Monzaemon). Similarly, French literature experienced several periods of cultural dynamism, such as the 12th century (e.g. Chrétien de Troye), the 16th century (e.g. Rabelais), during which French asserted itself as a literary language [51]). However, canonized authors (Chrétien de Troye, Rabelais) did not prevent the emergence and canonization of other authors (Victor Hugo, Balzac, etc.).

Thus, it may be the case that economic development peaked in Rome in the 1st century BC. In fact, the proxies often used to explain economic development (number of shipwrecks, quantity of money) probably reflect more economic production at the global level than at the individual level. For example, it is possible that the peak of Roman GDP occurred in the 1st or 2nd century AD, but that the peak of GDP per capita occurred earlier, during the conquest of the Roman Republic. Future studies should look more closely at the relationship between material and immaterial production.

5. Data and Related code

GitHub Project: Cultura Project

References


[22] V. Kaše, P. Heřmánková, and A. Sobotková. “Division of labor, specialization and diversity in the ancient Roman cities: A quantitative approach to Latin epigraphy”. In: Plos one 17.6 (2022), e0269869.


