

## **METHODOLOGIES AND MODELS**



JUSTIN STOVER

THE CICERONIAN BOOK AND ITS INFLUENCE:  
A STATISTICAL APPROACH

The word “book”, or *liber*, is an equivocal term<sup>1</sup>. On the one hand it refers to a division of a larger work, as in the twelve *libri* that make up Virgil’s *Aeneid* or the 142 that made up Livy’s history. On the other, it is a physical object: a *volumen* or book-roll. These two senses were linked in antiquity. The *volumen* has certain soft, physical constraints – in theory a roll can be of any length, large or small, but if it is too big, its durability is compromised and as an object it becomes very difficult to handle. Hence, a long work was usually divided into individual physical volumes, and these tended to come at set places in the text. In theory of course, any individual copyist could move on to a new roll at will, but this would severely hamper the usability of the work, since no one would know where to find a particular passage in any given copy.

This essay offers a first foray into a statistical approach to the ancient literary book. Previous studies have used statistical methods for modelling our evidence for the ancient physical book by using the fragmentary papyrus remnants to reconstruct the physical features of ancient book-rolls<sup>2</sup>. This study approaches the problem from a different perspective: by statistically modelling the ancient works transmitted as individual books, we will be able to glean insights into the physical book-rolls that once contained them. The question we pursue is whether the vast literary output of Cicero represents an inflection point in the development of the Latin book.

---

\*I would like to thank Gabriel Nocchi Macedo for discussion and bibliographic suggestions on the ancient book; Alison John for working with me on collecting book lengths (and to the School of History, Classics, and Archaeology at the University of Edinburgh for funding her research assistantship); George Woudhuysen for reading a draft of this study; Tom Keeline for comments on an earlier draft; Dominic Berry for insight on the *status quaestionis* of Cicero and the *Rhetorica ad Herennium*; and the organizers of the *Cicero digitalis* conference for giving me the opportunity to present this work. I note that this is the one of the first sorties of a much larger project on book lengths in antiquity: another is on the Sallustian book (see Stover-Woudhuysen forthcoming). Translations are mine.

<sup>1</sup> The standard account remains Birt 1882. On the literary book, see Higbie 2010.

<sup>2</sup> Johnson 2004. See also Del Mastro 2014 for stichometric counts in rolls from Herculaneum.

## 1. *Methods*

First some preliminaries. The first problem is how to calculate book-length. For poetry, it is a simple matter: the length of the book is the number of individual verses it contains. Prose is much trickier. Modern methods have previously used totally unsatisfactory and imprecise methods, such as number of pages in a given edition. Thanks to the advent of electronic text corpora, we have much more precise methods available to us: not only can we count words automatically, we can even count characters without spaces with modern word processing. The ancients, however, used a different means of calculating the length of prose books: they counted units of sixteen syllables as a *versus* or *stichos* interchangeable with the (hexametric) poetic verse<sup>3</sup>. We will use both methods here: characters without syllables as the base measurement for the length of a book, and the conversion into stichometric verses where necessary for comparison with poetic books or to evaluate ancient stichometric counts. In order to statistically model a particular corpus of books, we record the following statistics:

**N** = the number of books in the corpus

**$\mu$**  = the arithmetic mean (sum of the characters in each corpus divided by **N**)

**$\sigma$**  = the standard deviation (calculated as the square root of the sum of the difference between each value and  **$\mu$**  squared divided by **N**)

**C<sub>V</sub>** = coefficient of variation, defined as  **$\sigma$**  divided by  **$\mu$** , which describes the dispersion of the data points. **C<sub>V</sub>** is particularly useful because it allows direct comparison of data from sets of different scales (e.g. characters of prose and lines of poetry). The lower a **C<sub>V</sub>** for a corpus is the better-defined the author's conception of the book.

**Low** and **High** = the smallest and largest books in the corpus respectively

**Range** = the difference between the shortest and longest books in a corpus

**Median** = the length of book at the midpoint of the distribution of book-lengths

**Q<sub>1</sub>** and **Q<sub>3</sub>** = the length of the books at the first and third quartiles

---

<sup>3</sup> The best general treatment of stichometry remains Ohly 1928, focused on the Greek material. Latin stichometry has been much less studied despite quite a bit of ancient evidence: see Rouse-McNelis 2000. Further, the long persistence of stichometry in Latin has not been adequately studied, but we have manuscript evidence for such counts in authors as diverse as Sedulius (see Colker 1962) and Firmicus Maternus (see Kroll-Skutsch-Ziegler 1968<sup>2</sup>, 2.XIII). What I present here is a very simplified account using only the sixteen-syllable versus for which we have ancient evidence; Diels 1882 argued for occasional employment of eighteen-syllable versus. My syllabic counts are based on the actual counting of the number of vowels and diphthongs using a Python script.

**IQR** = interquartile range, or the difference between  $Q_3$  and  $Q_1$ . This measurement of range is more useful than simple range which can be severely distorted by outliers.

**QCv** = quartile coefficient of variation, calculated here as the **IQR** divided by the sum of  $Q_1$  and  $Q_3$ . This is the most robust measure of variation since it is not as sensitive to outliers as range.

These measures together give us a very robust model for an author's ideal book and how well defined that idea of a book is. One final note: the texts used are freely available online via electronic text-corpora such as the *Latin Library*, the *PHI* database, and the *Perseus Project*. Obviously, the exact figures will depend on individual editing and coding decisions, but the differences between various texts are in no instance large enough to matter for the results<sup>4</sup>.

## 2. *The Latin book in the time of Cicero*

Famously, our ancient sources tell us that Aristarchus and his epigones in Alexandria divided the two long poems of Homer into forty-eight books, that is, the twenty-four books each of the *Iliad* and the *Odyssey* that we know today<sup>5</sup>. This testimony has not been universally accepted; nonetheless, we have very good evidence that the dividing of earlier books was considered an important job for the professional grammarian. This holds true for Latin as well. One of the earliest grammarians at Rome, C. Octavius Lampadio (s. II BC), Suetonius tells us, took the long poem of Naevius (s. III BC), the *Bellum Punicum*, and divided it into seven books (*gramm.* 2).

C. Octavius Lampadio Naevii "Punicum Bellum", quod uno volumine et continenti scriptura expositum divisit in septem libros.

*Gaius Octavius Lampadio divided into seven books Naevius' Punic War, which had been produced in a single volume, with uninterrupted writing.*

---

<sup>4</sup> As such, the numbers presented below offer only the illusion of precision. Depending on which electronic copy is used and how the text is processed, the method will produce slightly different results. Such variation, however, never exceeds the threshold of statistical relevance.

<sup>5</sup> See the symposium in *Symbolae Osloensis* 74 (1999), introduced by a provocative account by Jensen (5-91), with responses defending the ancient view; and the neglected evidence brought forward by Nünlist 2006, with further bibliography.

There is much to be learned here. First, it is not at all impossible that Naevius' epic was indeed contained in one *volumen* in its earliest circulation. Of course, we do not know how long it was, but we do know that seven Homeric books – the books made by the Alexandrian grammarians whom Lampadio believed he was imitating – can be as short as 2317 lines (seven times *Od.* 7, at 331 lines). This is not much longer than a single book of Naevius' Greek contemporary Apollonius of Rhodes, whose four books average about 1700 lines. Besides comedies – whose lengths are conditioned by factors beyond physical books – the one Latin work we have surviving by a contemporary is the *Res rusticae* of Cato the Elder. This is not transmitted with any book indications, nor quoted in such manner by later writers, which suggests that it too was written as a single *volumen*, consisting of approximately 84900 characters, 15650 words, and 37240 syllables. Now a *versus* or *stichos*, as we have seen, was the standard unit for measuring prose text in antiquity, meant to define the equivalent of a hexametric line of poetry as 16 syllables. Hence, Cato's work consists of approximately 2328 *versus*, or nearly exactly the minimum possible length of Naevius' epic. This offers strong *prima facie* support for the plausibility of Suetonius' account. Further confirmation is found in a fragment of Varro's contemporary, the grammarian Santra, preserved by Nonius Marcellus (*fr.* 5): *quod volumen unum nos lectitavimus et postea invenimus septemfarium divisum*, which must refer to Naevius, and suggests that in the first century BC his work circulated both a single massive *volumen* and as seven short books.

Around the same time, it seems that writers themselves were deliberately structuring their longer works into books. The anonymous *Rhetorica ad Herennium*, for example, which probably dates from the 80's, contains explicit discussion of the size of its books at the ends of books 1 and 2:

*Rhet. Her.* 1, 27, 3: Nunc quoniam satis huius voluminis magnitudo crevit, commodius est in altero libro de ceteris rebus deinceps exponere, ne qua propter multitudinem litterarum possit animum tuum defatigatio retardare.

*Now, since the size of this volumen has grown to a sufficient extent, it is more convenient to discuss the rest of these matters in a second liber, so that some weariness can't make your mind flag on account of the sheer amount of text.*

*Rhet. Her.* 2, 50, 2: Fere locos obscurissimos totius artificii tractavimus in hoc libro; quapropter huic volumini modus hic sit: reliquas praeceptiones, quoad videbitur, in tertium librum transferemus.

*We have treated almost all the most obscure topics of this craft in this book. Hence, let this be the limit to this volumen. We will carry over the remaining precepts, as will be seen, into a third book.*

One would think that we could use these statements to determine the *modus*, or proper measure, of a *liber* or *volumen*, in the author's conception. And indeed if we look at the first three books of the *Rhetorica*, we find a fairly well-defined concept of the book:

	Characters	Words
<i>Rhet. Her.</i> 1	22040	3713
<i>Rhet. Her.</i> 2	42491	7260
<i>Rhet. Her.</i> 3	33151	5499

The problem arises with the fourth book, which one will find in modern editions extending over 78546 characters, or 13617 words. At this length, this single book far exceeds the total of books 1 and 2 together (64531 characters), which seems to give the lie to the explicit statements of the author at the ends of books 1 and 2. The medieval manuscript tradition, however, transmits the work in six books, which gives a fairly consistent division.

	Characters	Words
<i>Rhet. Her.</i> 4	19561	3306
<i>Rhet. Her.</i> 5	31893	5272
<i>Rhet. Her.</i> 6	31395	5189

The reasons why editions print the text in four books are complex, and would take us to far afield to discuss here.<sup>6</sup> Let it suffice to note that

---

<sup>6</sup> Marx 1894, 4-5 provides the standard (and frankly inadequate) defence for printing the text in four books. In brief, the earliest editions print the text as only four books, since the Italian Renaissance manuscripts on which they were based do not provide incipits (although they do put an interstice) for books five and six, and end with an *explicit liber ultimus*. In favour of the four-book division is the fact that the author provides explicit signals in the text for books 1-4, and the fact that Cicero's *De inventione* consisted of four books. On the other side, we have the unanimous testimony of the medieval manuscript tradition, and the consistency of length between the six books.

books 4-6 are very consistent with the first three books, and provide a consistent and well-defined statistical model of a book, unlike the four-book division:

Auth.	N	$\mu$	$\sigma$	$C_V$	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
<i>Rhet.</i> (6 bk.)	6	30089	8292	.28	19561	n/a	31644	n/a	42491	n/a	n/a	22930
<i>Rhet.</i> (4 bk.)	4	44057	24465	.55	22040	n/a	37821	n/a	78456	n/a	n/a	56506

We can get a more rounded picture of the Latin book before Cicero's influence by including three other authors contemporary with him. First Lucretius, who wrote six books in hexameters *De rerum natura*:

Auth.	N	$\mu$	$\Sigma$	$C_V$	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
Lucretius	6	1236	135.7	.11	1094	n/a	1230	n/a	1457	n/a	n/a	361

These are long books – the *Rhetorica ad Herennium*, by contrast, has on average about 825 *versus* per book. And yet they are nonetheless very well defined, with a coefficient of variation of just 11 per cent. We also have two prose authors with extant works transmitted in books: Caesar, for whom we have nine original books transmitted integrally (seven of the *Bell. Gall.* and two of the *Bell. Civ.*), and Varro, for whom we have transmitted integrally three books of the *De lingua latina* (5, 6, 7) and the three books of the *Res rusticae*:

	Characters	Words		Characters	Words
Varro, <i>ling. 5</i>	60515	11102	Caes. <i>Gall. 1</i>	50044	8518
Varro, <i>ling. 6</i>	34116	6276	Caes. <i>Gall. 2</i>	26164	4356
Varro, <i>ling. 7</i>	30170	5423	Caes. <i>Gall. 3</i>	22918	3755
Varro, <i>rust. 1</i>	80268	14437	Caes. <i>Gall. 4</i>	28544	4811
Varro, <i>rust. 2</i>	61214	10853	Caes. <i>Gall. 5</i>	47230	7757

<b>Varro,</b> <b>rust. 3</b>	57159	10157	<b>Caes.</b> <b>Gall. 6</b>	34854	5750
			<b>Caes.</b> <b>Gall. 7</b>	75457	12069
			<b>Caes.</b> <b>civ. 1</b>	70178	10991
			<b>Caes.</b> <b>civ. 2</b>	40467	6433

We can use this data (characters) to construct a statistical model of the Caesarian and Varronian books:

Auth.	N	$\mu$	$\sigma$	C <sub>v</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>v</sub>	RANGE
Caesar	9	49272	24350	.49	22918	39122	43849	65145	96867	35023	.37	45741
Varro	6	53907	18749	.35	30170	n/a	58837	n/a	80268	n/a	n/a	50098

Now, the evidence for Varro very much needs to be handled with extreme caution: not only are the books of the *Res rusticae* quite a bit longer than those of the *De lingua Latina* (although no less inconsistent), but also one of the books transmitted imperfectly (book 9) is in its surviving state quite a bit longer than two surviving integral books. Nonetheless, the evidence is clear enough that neither Caesar nor Varro had a well-defined notion of the book, and it is striking that the range in their respective corpora is itself longer than each of the first three books of the *Rhetorica ad Herennium*.

### 3. *The Ciceronian Book I: The Rhetorical Book*

Cicero's earliest work – written while still an *adulescens* – is the *De inventione*, a rhetorical handbook closely related to the *Ad Herennium*. It was written in four books, although only the first two have survived. The young Cicero seems to have been thinking very carefully about the length of his books, with explicit statements at the end of each one closely paralleling the statements at the end of the first two books of the *Rhetorica*:

Cic. *inv.* 1, 109: Sed quoniam satis, ut videmur, de omnibus orationis partibus diximus et huius voluminis magnitudo longius processit, quae sequuntur deinceps, in secundo libro dicemus.

*But since we seem to have discussed enough about all parts of a speech and the size of this volumen has gotten rather long, we will discuss what follows next in the second liber.*

Cic. *inv.* 2, 178: Quare, quoniam et una pars ad exitum hoc ac superiore libro perducta est et hic liber non parum continet litterarum, quae restant, in reliquis dicemus.

*On this account, since one part of this has reached the end in this and the preceding book, and this liber contains no small amount of text, we will discuss what remains in the following.*

The problem is that these are gigantic *libri*: book 1 contains 90673 characters, or 16020 words, and book 2 contains 101922 characters, or 17845 words. Indeed, book 2 by itself is longer than the first three books of *Rhetorica* put together (97682 characters). These *volumina*, then, are much more like one of Cato or Naevius. But genre may well have played a role as well: the only other work of rhetorical theory by Cicero in multiple books is the *De oratore* in three books. It too has gigantic books, of 107917 characters (18664 words), 155758 characters (27571 words), and 98524 characters (17114), respectively. *De oratore* 2 is in fact the longest Latin book of classical antiquity extant, rivalled only by Columella 12, at 157824 characters, or 26393 words, and definitively surpassed only by Tertullian *Adv. Marc.* 4, at 196810 characters or 33704 words. Taking these five books of rhetorical theory together, we can construct a statistical profile:

Auth.	N	M	$\sigma$	C <sub>v</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>v</sub>	RANGE
Cicero, <i>Rhet.</i>	5	110959	25807	.23	90673	n/a	101922	n/a	155758	n/a	n/a	65085

This suggests, despite the slight evidentiary basis, a fairly well-defined concept of a book, averaging just over 110000 characters, with a coefficient of variation of less than 25 per cent. These two are the only

rhetorical works by Cicero transmitted with book division, but it is consistent with two other works transmitted presumably as single books:

	Characters	Words
<i>Orator</i>	107590	18829
<i>Brutus</i>	145338	25445

Besides these four rhetorical works, we also have one collection of Ciceronian speeches transmitted as individual books. Speeches in general were not transmitted in books, for obvious reasons: their length was determined by factors external to the physical format in which they later circulated. The exception, the five speeches of the *secunda actio* against Verres, were (perhaps not coincidentally) never actually delivered. As a corpus, the books of the *II Verr.* offer a model identical with that found in the rhetorical books.

Auth.	N	$\mu$	$\sigma$	$C_V$	LOW	$Q_1$	MEDIAN	$Q_3$	HIGH	IQR	$QC_V$	RANGE
Cicero, <i>II Verr.</i>	5	11008	20751	.19	88345	n/a	109274	n/a	143491	n/a	n/a	55146

These are quite well-defined books, with a coefficient of variation less than twenty per cent, and at every measure from mean to standard deviation to median to range almost indistinguishable from the rhetorical works. Hence, we can put together a dataset of twelve rhetorical and oratorical books, which presents a robust model of one sort of Ciceronian book:

Auth.	N	M	$\sigma$	$C_V$	LOW	$Q_1$	MEDIAN	$Q_3$	HIGH	IQR	$QC_V$	RANGE
Cicero, <i>Rhet./Orat.</i>	12	113180	22410	.20	88345	98551	107754	118950	155758	20399	.09	67413

Our expanded statistical profile here is especially illuminating. Obviously, the physical dimensions of the finished product were not the only factor Cicero had in mind when he composed his books and, for any number of other, internal reasons, he may have made individual books longer or shorter. Still, excluding the outliers by calculating the quartile

coefficient of variance reveals a very well-defined conception of the book, at less than ten per cent variance, and an interquartile range of just over twenty thousand characters. At 37.4 average characters per *versus*, this means that the book at the third quartile is just under 550 *versus* longer than that at the first quartile.

#### 4. *The Ciceronian Book II: The Philosophical Book*

So far we have only dealt with a few components of the Ciceronian corpus, the rhetorical works and the oratorical works transmitted in individual books. We have an even larger number of multibook works by Cicero dealing with philosophy. These are: the four books *De finibus*, the second book of the *De divinatione* (the first is not transmitted complete), the five books of *Tusculanae disputationes*, the three books *De natura deorum*, the first two books *De legibus* (the third is imperfect), and the three books *De officiis*. This gives us a total of eighteen books, a number sufficient to sustain a robust analysis:

Auth.	N	$\mu$	$\sigma$	C <sub>V</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
Cic., <i>Phil.</i>	18	58822	17334	.29	34985	48822	51641	69722	91427	20900	.18	56442

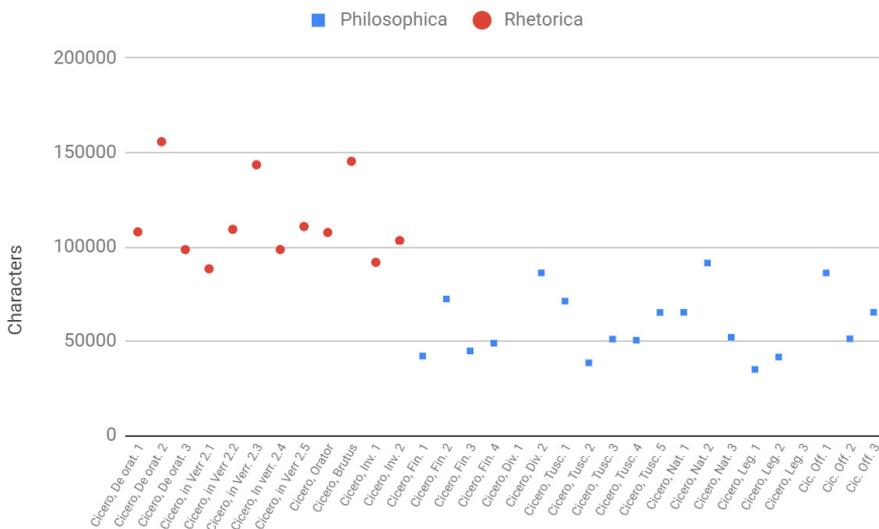
This corpus presents a well-defined concept of the book: while the coefficient of variation is middling at twenty-nine percent, the quartile coefficient of variation is relatively low at eighteen per cent. Hence what we have here is a second well-defined model of a book, which is different from the model of the rhetorical book. This model is consistent with our other evidence: Cicero also wrote six books *De re publica*, and their miserable state of preservation does not allow us to use their lengths in our modelling, but what we do have of books 1 and 2 is very much consistent with this model:

	Characters	Words
<b><i>Rep. 1 (imperfect)</i></b>	47335	8293
<b><i>Rep. 2 (imperfect)</i></b>	35927	6143

This sort of book is very similar to the books written by Cicero’s contemporaries, Caesar and Varro, just much better defined, that is to say with a closely comparable mean, median and range, but a much smaller (quartile) coefficient of variation:

Auth.	N	$\mu$	$\sigma$	C <sub>v</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>v</sub>	RANGE
Cic., <i>Phil.</i>	18	58822	17334	.29	34985	48822	51641	69722	91427	20900	.18	56442
Caesar	9	49272	24350	.49	22918	39122	43849	65145	96867	35023	.37	45741
Varro	6	53907	18749	.35	30170	n/a	58837	n/a	80268	n/a	n/a	50098

What this tells us about Cicero’s literary practice is that he consciously chose different types of books for different topics, according to identifiable pattern – a long book of about 3000 *versus* for rhetorical works, a more manageable *volumen* of 1500 *versus* for philosophical exposition. The striking consistency of these literary choices can be taken in at a glance if we put all thirty of the individual books into a scatterplot.



What this means is that were we to come across a new Ciceronian book we would be able to identify it as either philosophical or rhetorical in content with more than ninety percent confidence *solely on the basis of its length*.

### 5. The Influence of the Ciceronian Book

We can find the Ciceronian book reflected in later authors, in a way that is consistent with these generic divisions. We have surviving two multibook rhetorical works dating from after Cicero's lifetime, by Seneca the Elder and Quintilian. For Seneca's *Controversiae*, we have five books out of the original ten transmitted in complete form (the rest are transmitted only as *Excerpta*), 1, 2, 7, 9, and 10, and for Quintilian, we have the twelve books of the *Institutio* transmitted basically entire:

	Characters	Words		Characters	Words
<b>Cont. 1</b>	105328	17987	<b>Inst. 1</b>	103357	17583
<b>Cont. 2</b>	95740	16256	<b>Inst. 2</b>	84051	14560
<b>Cont. 7</b>	100984	16989	<b>Inst. 3</b>	84415	14406
<b>Cont. 9</b>	82137	13963	<b>Inst. 4</b>	72225	12335
<b>Cont. 10</b>	76150	12994	<b>Inst. 5</b>	68737	11905
			<b>Inst. 6</b>	70529	12195
			<b>Inst. 7</b>	74378	13150
			<b>Inst. 8</b>	78361	13495
			<b>Inst. 9</b>	112462	19462
			<b>Inst. 10</b>	73098	12396
			<b>Inst. 11</b>	92319	15817
			<b>Inst. 12</b>	78650	13469

These books are a bit shorter than the Ciceronian rhetorical books, but present very similar models, albeit better defined:

Auth.	N	$\mu$	$\sigma$	Cv	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QCv	RANGE
Cicero, <i>Rhet./Orat.</i>	12	113180	22410	.20	88345	98551	107754	118950	155758	20399	.09	67413
Sen. <i>Cont.</i>	5	92067	12458	14	76150	n/a	95740	n/a	105328	n/a	n/a	29178
Quint. <i>Inst.</i>	12	82715	13682	.17	68737	72880	78506	86391	112462	13511	.08	43275

Tracing the influence of Cicero's philosophical books is much more difficult. First, we have very few philosophical corpora from Latin antiquity, and the two we have, by Seneca the Younger and Apuleius, follow a very different conception of the book, as we shall see later on. Second, as we have already observed, the philosophical books of Cicero are not especially distinct from books by his contemporaries, save in that they are better defined. Therefore, while we can tentatively identify Cicero's innovation in defining his books more precisely than his predecessors and contemporaries, we can be much less sure when we turn to later works with similar sorts of books that they are actually imitating Ciceronian practice. With those caveats, we can indeed find some later expository works of diverse genre which present a book model similar to that which we find in the Ciceronian book. Valerius Maximus, for example, presents an extremely well-defined conception of the book, which is very close to that of Cicero, as do the nine books of his contemporary Columella on agriculture<sup>7</sup>:

Auth.	N	$\mu$	$\sigma$	C <sub>V</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
Cicero, <i>Phil.</i>	18	58822	17334	.29	34985	48822	51641	69722	91427	20900	.18	56442
Caesar	9	49272	24350	.49	22918	39122	43849	65145	96867	35023	.37	45741
Varro	6	53907	18749	.35	30170	n/a	58837	n/a	80268	n/a	n/a	50098
Valerius Maximus	9	57179	4072	.07	49447	56391	57474	59475	63794	3084	.03	14352
Columella	9	54177	6906	.13	43751	50075	56165	57634	62904	7559	.07	19153

Whether or not we can detect the influence of Cicero here, at the very least we can see the tendency for books to become better defined over time, varying within a much tighter range.

---

<sup>7</sup> We exclude book 10 because it is partially in prose and partially in verse (that is, in verse with a prose preface); books 11 and 12 (which are vastly longer than 1-9) are actually separate works (Goodyear 1982, 669).

## 6. *The Books of Cicero's Epistulae*

The final component of the Ciceronian corpus transmitted in individual books are the *Epistulae*. There is good reason for this: unlike the other works, we can be certain that the books of letters as we have them were not put together and published by Cicero himself, but only sometime after his death. Hence, in order to analyze them, it is necessary to have another brief excursus into the post-Ciceronian book.

Alongside the prose authors of the late Republic and Augustan period we have looked at so far, we have a parallel and even more striking development among Latin poets of a single, well-defined model of the book. This is not a novel observation: the Latin poetry book has received vastly more scholarly attention than the prose book. Even so, it has rarely been explored with a statistical approach. Fortunately, between Virgil, Horace, and Ovid, we have three poets with large extant corpora that permit us to model their books with a high degree of accuracy.

Auth.	N	$\mu$	$\sigma$	$C_V$	LOW	$Q_1$	MEDIAN	$Q_3$	HIGH	IQR	$QC_V$	RANGE
Horace	10	764	242	.32	476	575	725	984	1083	409	.26	607
Virgil	17	760	141	.19	514	705	804	871	952	166	.11	438
Ovid	38	784	91	.11	578	734	787	847	968	113	.07	390

It is plain to see that these are identical books: their separate means are within 24 lines of one another, differing only in range and definition<sup>8</sup>. Where did this book model come from? It does not seem to have come from the Hellenistic poets: Apollonius' books are more than double these in length, and although Callimachus makes much of his short books of poetry in the prologue to the *Aitia*, they seem to have contained at least a thousand lines each in their four books, and perhaps even more<sup>9</sup>. More likely they took inspiration from the Homeric books, which are fairly well-defined, and quite a bit shorter than the books of the Hellenistic poets:

<sup>8</sup> Van Sickle 1980, with references to earlier literature.

<sup>9</sup> Janko 2000 offers a perceptive comment on the lengths of Apollonius' books. The four books of the *Aitia* contained perhaps 6000 lines according to Harder 2011, 64.

Auth.	N	$\mu$	$\sigma$	C <sub>V</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
<i>Iliad</i>	24	654	161	.25	423	521	611	811	909	291	.22	486
<i>Odyssey</i>	24	505	112	.22	331	434	495	569	847	135	.13	516
Total:	48	579	157	.27	331	469	547	623	909	154	.14	578

Much less well-known, however, is that prose authors of the same period began to write books according to the same model<sup>10</sup>. Vitruvius, the author of ten books *De Architectura*, written probably in the 20s BC, wrote compact and elegant books, which I give here in both characters and in hexametric *versus* measured by syllables:

Auth.	N	$\mu$	$\sigma$	C <sub>V</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
Vitruvius (char.)	10	35472	7745	.22	24842	31967	35737	37465	51830	5498	.08	27767
Vitruvius ( <i>versus</i> )	10	948	207	.22	661	858	954	1002	1383	144	.08	722

These are a bit longer than the standard book of Augustan poetry, but very much within the same range, unlike the previous prose books we have examined. Indeed, the only comparable work in terms of book-lengths is the *Rhetorica ad Herennium*. They are also extremely well defined, with a quartile coefficient of variation of just eight per cent.

Vitruvius also provides us with an explicit discussion of book lengths in the preface to book 5. He first discusses how historians and poets could write rather long books, counting on the interest of their subject matter and the formal qualities of their writing to hold readers' attention (*De arch.* 5. *praef.* 1). Books of history are a topic we have not discussed here since we have no historiography by Cicero; but it is certainly true that Livy wrote quite long books (N=32,  $\mu$ =90936,  $\sigma$ =13274), and we have

---

<sup>10</sup> Vasaly 2002 offers a partial exception to this general neglect; she provides a stimulating look at how the structure of Livy's first five books replicates the structure of an Augustan poetry book, but does not take into account the physical features of the volumes.

every reason to believe that Sallust wrote even longer ones before him<sup>11</sup>. The mention of poetry is more surprising, but Lucretius did write books on average thirty percent longer than those of Vitruvius, and long books, such as Naevius' *Bellum Punicum* (before it was divided by Lampadio) were certainly still in circulation in Vitruvius' lifetime, as Santra informs us. Vitruvius then goes on to explicitly discuss in numerical terms, the ideal "Pythagorean" book (5. *praef.* 3):

paucis iudicavi scribendum, uti angusto spatio vacuitatis ea legentes breviter percipere possent. Etiamque Pythagoras quique eius haeresim fuerunt secuti, placuit cybicus rationibus praecepta in voluminibus scribere, constitueruntque cybum *CC et L* versus eosque non plus tres in una conscriptione oportere esse putaverunt.

*I decided it should be written in just a few words, so that those who read it can quickly take it in during their brief intervals of free time. Pythagoras, too, and those attached to his sect, chose to write their teachings in volumina according to the nature of cubes; they established a cube of 250 verses and judged it appropriate to place no more than three of them in a single composition.*

The problem is, as Fra' Giocondo perceived half a millennium ago, that 250 is not a cube. He proposed to emend the text to CCXVI (216), which is indeed the cube of the perfect number 6. Almost every edition since has printed his conjecture, until the recent Budé edition by Saliou, which restores the manuscript reading on the basis that 250 is double 125, which is the cube of 5<sup>12</sup>. Scholarship has extrapolated from this passage that Vitruvius is saying that a perfect book ought to be 648 lines, or three times 216<sup>13</sup>. This interpretation, however, takes the rare word *conscriptio* as equivalent to *volumen*, which seems to go against the sense of the passage, and is not consistent with Vitruvius' usage

---

<sup>11</sup> On the *Jugurtha*, see Stover-Woudhuysen 2015; on the likely length of the lost books of the *Histories*, see Stover-Woudhuysen 2021.

<sup>12</sup> Saliou 2009, 2.

<sup>13</sup> Mondin 2019, 694-695. Mondin argues that the *Liber de virginitate* in Avitus of Vienne's *Historia spiritalis* consists of 648 lines if one excludes the preface which thereby confirms Giocondo's conjecture. This is totally unnecessary: given that we have nearly a thousand extant books of Latin poetry from Antiquity, and 648 is with a couple of hundred lines of the mean length, it is more than likely that at least one book will randomly end up at 648 lines (as indeed book 3 of Martial's *Epigrams* does, and no one considers that *jeu d'esprit* somehow an instantiation of a perfect Pythagorean book). There is simply no evidence for the "Pythagorean" book outside of Vitruvius, and no evidence that Avitus had Vitruvius in mind.

elsewhere (7. *praef.* 1). Figuring out the precise meaning is not necessary for our purposes here, however: what is important is simply that Vitruvius provides an explicit defence of, and a model illustrating, short prose books.

The Vitruvian book would go on to have great success during the principate. We have, for example, 37 integral books by Seneca the Younger, six books of *Naturales quaestiones* (book 4 is imperfect), two books *De ira* (the first book is imperfect), one book *De clementia* (book 2 is imperfect), seven other books of *Dialogi* (1, 2, 6, 7, 9, 10, 12), 13 books of letters to Lucilius (books 1-10, 16, 19, 20). Statistically, these books are very similar to those of Vitruvius:

Auth.	N	$\mu$	$\sigma$	C <sub>v</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>v</sub>	RANGE
Vitruvius	10	35472	7745	.22	24842	31967	35737	37465	51830	5498	.08	27767
Seneca	37	38634	7893	.20	23102	35499	38369	45392	55621	9893	.12	32519

Like Vitruvius, Seneca also expresses definite opinions about the ideal book length. In a letter to Lucilius (*Ep.* 93, 11), Seneca contrasts the *paucorum versuum liber* with the overlong work of an old Republican historian Tanusius:

Et paucorum versuum liber est et quidem laudandus atque utilis: Annales Tanusii scis quam ponderosi sint et quid vocentur.

*And the liber of few versus is indeed both praiseworthy and useful. You know how bulky the Annales of Tanusius are, and what they are called.*

Seneca, of course, is not specifically discussing the lengths of his own books – although surely a reader holding the *volumen* in his hand could be expected to make the inference.

This same book model we find in Cicero's epistolary corpus. This corpus consists of 36 books transmitted integrally: three books *Ad Quintum*, one book *ad Brutum* (book 2 is imperfect), sixteen *Ad Atticum*, and sixteen *Ad familiares*:

Auth.	N	$\mu$	$\sigma$	C <sub>V</sub>	LOW	Q <sub>1</sub>	MEDIAN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
Cicero, <i>Epp.</i>	36	39867	9758	.24	18302	34987	38689	46245	71382	11258	.14	53080

This is the same as the Senecan book, with a larger range, but a statistically identical mean, median, interquartile range, and quartile coefficient of variation. Indeed, if we look to just the genre of epistolography, we can find something approaching a normative book model. For this, we have the thirteen books of letters of Seneca, and the nine written by Pliny the Younger (book 10 of the letters was put together after Pliny's death).

Auth.	N	$\mu$	$\sigma$	C <sub>V</sub>	LOW	Q <sub>1</sub>	MEDI- AN	Q <sub>3</sub>	HIGH	IQR	QC <sub>V</sub>	RANGE
Cicero, <i>Epp.</i>	36	39867	9758	.24	18302	34987	38689	46245	71382	11258	.14	53080
Seneca, <i>Epp.</i>	13	37621	6533	.17	26870	32575	38185	41002	49532	8427	.11	22662
Pliny, <i>Epp.</i>	9	38064	1579	.04	36315	37114	37491	38526	41042	1412	.02	4727

Now when Cicero's letters were arranged into books and published is a notoriously difficult question<sup>14</sup>. We know it must have happened before Seneca, however, and plausible datings have ranged from the end of Cicero's life to the reign of Nero. For the *Ad Atticum* in particular, much depends on the text and interpretation of a disputed passage in Cornelius Nepos' life of Atticus<sup>15</sup>. What this analysis can potentially add to the debate is that (1) there is no particular reason why the books of letters could not have been assembled as early as the time of Vitruvius, and

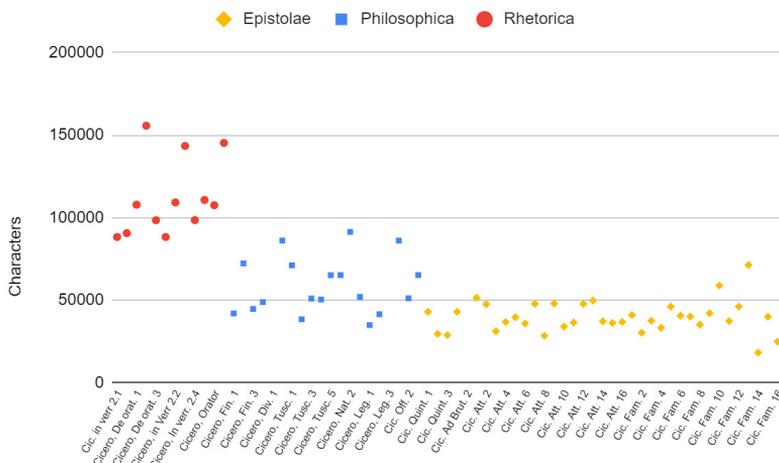
<sup>14</sup> Shackleton Bailey 1965, 59-75 (arguing for a later date); Setaioli 1976 (arguing for an earlier date); both with extensive reference to earlier literature. There has been increasing attention to the careful design of the letter-books: e.g. Beard 2002; Grillo 2015. See also Gibson 2012.

<sup>15</sup> Nep. *Att.* 16, 3: *Ei rei sunt indicio praeter eos libros, in quibus de eo facit mentionem, qui in vulgus sunt editi, undecim volumina epistularum ab consulatu eius usque ad extremum tempus ad Atticum missarum; quae qui legat, non multum desideret historiam contextam eorum temporum.*

that (2) Seneca's letters (and the subsequent tradition of Latin epistolography) is likely much more deeply indebted to Cicero's collection than the relatively sparse references would have us believe. Going further than this would take us too far beyond the matter at hand. Let it suffice here to note that the Senecan book, which is roughly comparable to the Augustan poetry book, if a little longer (ca. 1000 *versus* against 750), would go on to have great success, and become the standard prose book in the age of Apuleius, Fronto, and Gellius<sup>16</sup>.

### 7. Three Ciceronian books

This has been no more than a first foray into new territory. As mentioned at the outset, until two decades ago, it was not possible to accurately measure the length of prose works, at least not in any significant number. Future scholarship will find this a rich territory for exploration; and may indeed come upon new findings which force us to question settled consensus (for example, the original structure of the *Rhetorica ad Herennium*). For Cicero, we have achieved a remarkably consistent set of results, which show that his corpus very neatly divides into three groups based solely on book length, and that these three groups correspond almost perfectly to the three genres of rhetoric, philosophy, and epistolography.



As Vitruvius said, decades after Cicero's death, different lengths of book are appropriate for different genres of writing, and what works for one is not necessarily appropriate for another. This is what we find strikingly illustrated in the *corpus Ciceronianum*.

### Works cited

- Birt 1882: T. Birt, *Das antike Buchwesen in seinem Verhältnis zur Literatur*, Berlin 1882.
- Colker 1962: M. Colker, *Stichometry that does not tally*, «*Scriptorium*» 16, 1962, pp. 85-89.
- Del Mastro 2014: G. Del Mastro, *Titoli e annotazioni bibliologiche nei papiri greci di Ercolano*, «*Cronache Ercolanesi*», Supplementi 5, Napoli 2014.
- Diels 1882: H. Diels, *Stichometrisches*, «*Hermes*» 17, 1882, pp. 377-384.
- Gibson 2012: R. Gibson, *On the Nature of Ancient Letter Collections*, «*Journal of Roman Studies*» 102, 2012, pp. 56-78.
- Goodyear 1982: F. Goodyear, *Technical writing*, in E. Kenney, W. Clausen (eds.), *The Cambridge History of Classical Literature*, vol. 2, Cambridge 1982, pp. 667-673.
- Harder 2011: A. Harder, *Callimachus as Fragment* in B. Acosta-Hughes, L. Lehnus, S. Stephens (eds.), *Brill's Companion to Callimachus*, Leiden 2011, pp. 63-80.
- Higbie 2010: C. Higbie, *Divide and Edit. A Brief History of Book Divisions*, «*Harvard Studies in Classical Philology*» 105, 2010, pp. 1-31.
- Janko 2000: Review of Jensen 1999, «*Bryn Mawr Classical Review*», <https://bmcr.brynmawr.edu/2000/2000.01.25/>.
- Jensen 1999: M. Jensen, *Dividing Homer: When and How were the Iliad and the Odyssey divided into songs?*, «*Symbolae Osloenses*» 74, 1999, pp. 5-91.
- Johnson 2004: W. Johnson, *Bookrolls and Scribes in Oxyrhynchus*, Toronto 2004.
- Kroll-Skutsch-Ziegler 1968<sup>2</sup>: W. Kroll, F. Skutsch, K. Ziegler (eds.), *Julii Firmici Materni Matheseos libri VIII*, 2 vol., revised edition, Leipzig 1968.
- Marx 1894: F. Marx, *Incerti auctoris De ratione dicendi ad C. Herennium libri IV: (M. Tulli Ciceronis Ad Herennium libri VI)*, Leipzig 1894.
- Mondin 2017: L. Mondin, *Quattro note al De virginitate di Avito di Vienne*, «*Bollettino di studi latini*» 47, 2017, pp. 674-695.
- Nünlist 2006: R. Nünlist, *A Neglected Testimonium on the Homeric Book-Division*, «*Zeitschrift für Papyrologie und Epigraphik*», 157, 2006, pp. 47-49.
- Ohly 1928: K. Ohly, *Stichometrische Untersuchungen*, Leipzig 1928.

- Rouse-McNelis 2000: R. Rouse, C. McNelis, *North African literary activity: a Cyprian fragment, the stichometric lists and a Donatist compendium*, «Revue d'histoire des textes» 30, 2000, pp. 189-238.
- Setaioli 1976: A. Setaioli, *On the date of publication of Cicero's letters to Atticus*, «Symbolae Osloenses» 51, 1976, pp. 105-120.
- Shackleton Bailey 1965: D. Shackleton Bailey (ed.), *Cicero: Letters to Atticus*, vol. 1, Cambridge 1965.
- Stover forthcoming: J. Stover, *The Nature of Apuleius' De Platone: An Isagoge?*, in A. Motta, F. Petrucci (eds.), *Isagogical crossroads in the post-Hellenistic age and late antiquity*, Leiden forthcoming.
- Stover-Woudhuysen 2015: J. Stover, G. Woudhuysen, *Aurelius Victor and the Ending of Sallust's Jugurtha*, «Hermathena» 199, 2015 [published 2020], pp. 93-134.
- Stover-Woudhuysen forthcoming: J. Stover, G. Woudhuysen, *Libri quinque historiarum: Hegesippus and the Histories of Sallust*, «Histos», forthcoming.
- Van Sickle 1980: J. Van Sickle, *The Book-Roll and Some Conventions of the Poetic Book*, «Arethusa» 13, 1980, pp. 5-42.
- Vasaly 2002: A. Vasaly, *The Structure of Livy's First Pentad and the Augustan Poetry Book*, in D. Levene, D. Nelis (eds.), *Clio and the Poets*, Leiden 2002, pp. 275-290.

