



HAL
open science

New Cliometrics for Ancient History

Laurent Gauthier

► **To cite this version:**

Laurent Gauthier. New Cliometrics for Ancient History. Economics and Finance. Université Paris 8 - Vincennes-Saint-Denis, 2023. English. ⟨NNT : ⟩. ⟨tel-04230060⟩

HAL Id: tel-04230060

<https://theses.hal.science/tel-04230060v1>

Submitted on 5 Oct 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



HAL Authorization

Thèse de doctorat

**New Cliometrics
for Ancient History**

Laurent Gauthier

Soutenu le 22 septembre 2023

sous la direction d'Antoine PARENT

Discipline

Sciences économiques

École doctorale

ED 401, Sciences sociales

Laboratoire

Laboratoire d'Économie
Dionysien (LED)

Composition du jury

Steven DURLAUF Rapporteur
Professeur d'économie, Université de Chicago

François VELDE Rapporteur
Senior economist et research advisor, Federal
Reserve Bank de Chicago

Cécile BASTIDON Examinatrice
Maîtresse de conférences HDR en économie,
Université de Toulon

Josiah OBER Examineur
Professeur d'histoire ancienne et de sciences
politiques, Université de Stanford

Julien CHEVALLIER Président du jury
Professeur d'économie, Université Paris 8

Antoine PARENT Directeur de thèse
Professeur d'économie, Université Paris 8

noli turbare circulos meos

Eleonorae, Alexandrae, Iuliani Victoriaeque

L'Université de Paris 8 n'entend donner aucune approbation ni improbation aux opinions émises dans les thèses ; ces opinions doivent être considérées comme propres à leurs auteurs.

Résumé

Cette thèse propose de relier l'analyse économique formelle au détail des sources historiques anciennes, prolongeant ainsi la cliométrie traditionnelle, grâce aux développements récents qui ont associé cliométrie et complexité. La plupart des approches économiques de l'histoire ancienne relèvent des *analytic narratives*, plutôt que de la cliométrie, faute de données. Nous proposons un nouveau cadre épistémologique pour la cliométrie, reposant sur la complexité : nous soutenons qu'elle ne doit pas nécessairement se concentrer sur l'économie *per se*, mais peut opérer sur des sources historiques primaires, et peut aborder un ensemble beaucoup plus large de périodes et de sociétés, en s'appuyant sur des modèles microéconomiques. Nous nous concentrons sur le cas de la Grèce ancienne et montrons que nous pouvons placer de nombreux documents historiques, désormais disponible sous forme électronique, dans un cadre économique. Nous nous penchons ensuite de manière plus approfondie sur deux aspects particuliers de la société grecque antique : le choix des noms, et le choix des dieux. Nous développons un modèle économique qui rend compte de l'existence des noms, comme un moyen d'échange d'information, et nous étudions le choix optimal des noms dans ce contexte. L'impact du choix stratégique des noms a une influence sur la distribution de ces noms. En confrontant les résultats théoriques aux données empiriques des périodes archaïque et classique, nous montrons qu'un degré de conformisme dans les choix de noms explique la forme particulière de la distribution des noms dans les données. Enfin, nous nous intéressons au choix optimal de quel dieu adorer, en fonction de la stratégie présumée des dieux pour rendre des faveurs aux fidèles, et nous rapportons cela à une forme de mesure de l'efficacité divine. Les données d'un grand nombre de sources épigraphiques et littéraires sur les actes votifs des Grecs anciens montrent que les distributions de ces actes suivent effectivement des lois de puissance avec un haut degré de régularité, tel que le prédit le modèle.

Abstract

This thesis proposes to link formal economic analysis with the full detail of ancient historical sources, thereby extending traditional cliometrics, thanks to the recent developments that have associated cliometrics and complexity. Most economic approaches into ancient history have remained in the field of analytic narratives, rather than cliometrics, because of a lack of data. We propose a new epistemological framework for cliometrics, relying on complexity: we argue they do not have to be about the economy, but can operate on primary historical sources, and could address a much broader set of periods and societies, leaning on microeconomic models. We focus on the case of ancient Greek history, and show we can frame into economic terms the massive and complex cultural material now available in electronic form. We then drill further down and concentrate on two particular aspects of ancient Greek society: the choice of names, and which gods to pray to. We first develop an economic model for the existence of names, as a way of exchanging identification information. We study the optimal choice of names in these conditions and show that the impact of strategic naming on the distribution of names works as an alteration of existing mean-field approaches to name dynamics. Confronting theoretical results with empirical data from the archaic and classical periods, we show conformist strategic naming may account for the particular shape of the name distribution in Ancient Greece. Finally, we look into the optimal choice of which god to worship as a function of the presumed strategy of the gods for returning favors to the worshippers, and relate it to a form of divine efficiency measure. Data from a large volume of epigraphic and literary sources on actual acts of worships from the ancient Greeks shows that the distributions of these acts effectively follow power laws with a high degree of regularity, as predicted by the model.

Acknowledgments

Starting from the vague impression that there should be more quantitative perspectives into ancient history, I did not at first even know if there existed an epistemological framework in which to explore this idea. My intuition was that economics should be the right place, as opposed to quantitative history, or to statistical physics, or sociology, but this remained nothing more than an intuition for some time. Of the many reasons why I owe much to Antoine Parent, Professor of economics at Université Paris 8 - Vincennes Saint-Denis, and my thesis director, the first one should hence be the simple fact he initiated the Cliometrics and Complexity group, which turned out to be the perfect intellectual framework for me to pursue my interests. By combining many different perspectives, whether formal or not, this domain was a perfect fit for me, and I was able to explore all the ideas I wanted.

Another reason why I offer Antoine my warmest thanks is for his tremendous ability to resolve all bureaucratic problems and hassles. Not being physically present at Université Paris 8 most of the time was not an issue, and I was excused from following doctoral school training, thanks to him, which was essential to me, working full time in parallel. In the actual conduct of my research, Antoine's support, guidance and encouragements have been paramount, and I am indebted to him for that: he clearly told me when he thought what I was doing was right, and clearly told me when it was not the best approach. His advice has been essential for me to determine what to go after, what to drill into, and what was best left aside—still while leaving me all freedom to explore what I wanted. I also can't thank Antoine enough for having enrolled me in many conferences and seminars, and prompted me to take part in these as much as I could, so that I could present my research and interact with others in the field. When I approached this research, I had some background in history, in economics, in

mathematics, and some ideas about how to put it together, but I did not have a good understanding of how I could link it into existing research—a logical consequence of focusing on a subject that is, to say the least, somewhat off-the-run in economics. Antoine's help was instrumental in allowing me to see the connections that I needed to build, so that the epistemological side of my endeavor was not in a vacuum.

I wish to warmly thank my referees, Steven Durlauf, Professor of economics at the University of Chicago and François Velde, Senior economist and senior adviser at the Federal Reserve Bank of Chicago. I am honored that they accepted to examine, and report on, my work. François Velde's background, with a wide-ranging expertise from formal methods to historical economics, gives me an example of what to aspire to. Having had some occasions to discuss research ideas with Steven at several seminars, I have always been impressed with how he connected perspectives from the history of economic thought, from economic models, game theoretical models, and from economic history, making sense of ideas in a literary and quantitative sense at the same time.

Cécile Bastidon, Associate professor of economics with habilitation at Université de Toulon, is a founding member of the Cliometrics and Complexity group, and her works combining macro finance, network theory and economic history have provided me with an epistemological beacon, as they were essential for me in order to understand how to position my own research. I am very thankful that she has accepted to be a part of my jury.

Josh Ober, Professor of political science and classics at Stanford University, and a specialist of ancient Greek history, has agreed to be a part of my jury; it is an honor for me to have the input of such an eminent and innovative thinker. I am also particularly thankful to Josh, because he's coming from across the aisle: from history towards economics. Having tried to follow the same path (albeit from the other side), I have found it quite difficult, and I view Josh's work as a goal to aim for.

I also offer my thanks to Julien Chevallier, Professor of economics at Université Paris 8 - Vincennes Saint-Denis, and acting as president of the jury. This thesis makes use of a good amount of data, and I hope that the way they are processed and analyzed will be satisfactory to Julien, who is a specialist of econometrics.

Tackling a field that touched upon economics as well as history, I have benefited greatly from the help many historians and philologists offered me in my research, and I take this opportunity to thank them as well.

Christel Müller, Professor of ancient Greek history at Université Paris 10 - Nanterre, was my advisor for my research master's thesis in history, on the language of decision in Greek tragedy. Her guidance was essential to me in order to begin to understand how to think as an historian, and how to internalize the logic of a historian use of quantitative data, and also how to comprehend important distinctions between classical literature and history. These were essential inputs into my work in economics, and I offer her my thanks.

I also wish to extend my thanks to Julien du Bouchet, Associate Professor of Greek literature at Université de Montpellier 3 - Paul Valéry, and a specialist in ancient Greek lexicology, who was my advisor for my research master's thesis in classical literature, on the lexicology of debt in the Greek Antiquity. He has greatly helped me internalize the distinction between close and distant reading, and between the philological and quantitative readings of ancient texts.

My work on the religious aspects of Greek society, in particular votive acts, has significantly benefited from many discussions I have had with Corinne Bonnet, Professor of ancient Greek history, and Sylvain Lebreton, post-doc in ancient Greek history, from the Mapping Ancient Polytheisms group at Université de Toulouse 2 - Jean Jaurès. I thank them warmly for all these fruitful exchanges.

The analysis of names, onomastics, and the study of particular individuals, prosopography, are important fields of historical research, which often rely on network approaches. Karine Karila-Cohen, Associate Professor of ancient Greek history at Université de Rennes 2 has greatly helped me in my approach to this domain, and I thank her for our many discussions of the subject.

Aurélien Berra, Associate Professor of Greek literature at Université Paris 10 - Nanterre, is also a digital scholar, and he offered me much information and advice through our numerous exchanges, on various sources and processes relevant for ancient Greek texts. I am very grateful to him for his help in this regard.

Contents

- Résumé 5
- Abstract 7
- Aknowledgments 9
- Introduction 21

- I Complexity: An Epistemological Framework for the Cliometrics of Ancient Times 27**

- 1 The Political Economy of Ancient Greek *Polis* and *Demokratia* 29**

 - 1.1 The Birth of the *Polis* 30
 - 1.2 The Invention of *Demokratia* 34
 - 1.3 Analytic Narratives vs. Cliometrics 37
 - 1.4 Chapter Bibliography 39

- 2 Extending Cliometrics to Ancient History with Complexity 43**

 - 2.1 History, Economics and Cliometrics 45
 - 2.1.1 Defining History and Economics 45
 - 2.1.2 Cliometrics and Historical Economics 46
 - 2.2 Epistemological Distinctions Between History and Economics 49
 - 2.2.1 Stylized Facts and Primary Sources 49
 - 2.2.2 Nomothetic Perspectives 53
 - 2.3 New Cliometrics With Complexity 55

| | | |
|--|--|----------------|
| 2.3.1 | Cliometrics & Complexity Can Offer New Perspectives on Historical Data | 56 |
| 2.3.2 | Extending Cliometrics & Complexity Towards History | 58 |
| 2.3.3 | A New Notion of Corpus in Cliometrics & Complexity | 59 |
| 2.3.4 | Extending the Range of Cliometrics & Complexity to Ancient History | 62 |
| 2.4 | Conclusion | 63 |
| 2.5 | Chapter Bibliography | 65 |
| 3 | A Cliometrics and Complexity Perspective on Ancient Greek Culture | 71 |
| 3.1 | Identifying Distributional Patterns and Mechanisms | 73 |
| 3.1.1 | Some Useful Distribution Classes | 74 |
| 3.1.2 | Empirical Distribution Comparisons | 77 |
| 3.1.3 | Dealing with Sample Size | 79 |
| 3.2 | Ancient Greek Writings and Cliometrics | 81 |
| 3.2.1 | The Economy of Literary Genre | 82 |
| 3.2.2 | Epigraphy and Constraints | 86 |
| 3.2.3 | Gods and Market Shares | 90 |
| 3.3 | Networks, Complexity, and Naming Gods and Men | 94 |
| 3.3.1 | The Economy of Names: Onomastic Networks | 95 |
| 3.3.2 | How to Speak to the Gods? The Economy of Divine Formulae | 99 |
| 3.4 | Conclusion | 106 |
| 3.5 | Chapter Bibliography | 108 |
| II Cliometrics and Complexity Applications to Ancient Greek Society | | 115 |
| 4 | The Cliometrics of Name Choice in Ancient Greece | 117 |
| 4.1 | Ancient Greek Names in Context | 120 |
| 4.2 | Modeling the Use of Names | 124 |
| 4.2.1 | Equilibria with Perfect Information | 126 |
| 4.2.2 | Equilibria with No Information | 129 |
| 4.2.3 | Equilibria with Private Information | 130 |

| | | |
|----------|--|------------|
| 4.2.4 | Equilibria With Partially Common Information | 133 |
| 4.3 | The Dynamics of Naming Choices | 138 |
| 4.3.1 | Representing Strategic Name Choices | 139 |
| 4.3.2 | A Model for the Distribution of Names | 142 |
| 4.3.3 | Name Distribution Limiting Behavior | 146 |
| 4.4 | Data Analysis | 149 |
| 4.4.1 | Number of Name Observations | 149 |
| 4.4.2 | Empirical Name Distributions | 149 |
| 4.5 | Conclusion | 153 |
| 4.6 | Chapter Bibliography | 154 |
| 5 | Laws of Divine Power in the Greek Pantheon: The Cliometrics of God Choice | 157 |
| 5.1 | Some Aspects of Ancient Greek Religion | 161 |
| 5.2 | Worship Decision Model | 164 |
| 5.2.1 | People's Devotion Decisions | 164 |
| 5.2.2 | Optimal Divine Strategy | 169 |
| 5.2.3 | Human Drivers of Divine Performance | 174 |
| 5.2.4 | Random Growth and Divine Performance | 176 |
| 5.3 | Empirical Analysis of Veneration Acts | 180 |
| 5.3.1 | Epigraphic and Literary Sources Combined with <i>Polis</i> -level Data | 180 |
| 5.3.2 | Analyzing Cult Data | 184 |
| 5.3.3 | Model Test | 186 |
| 5.3.4 | Explaining Differences Across <i>Poleis</i> | 190 |
| 5.4 | Conclusion | 195 |
| 5.5 | Chapter Bibliography | 197 |
| | Conclusion: What Can History Bring to Cliometrics? | 201 |
| | Appendix | 205 |
| | Data Extraction and Analysis Algorithms | 205 |
| | Diorisis | 205 |
| | PHI | 206 |

| | |
|--|------------|
| LGPN | 214 |
| BDEG Data | 221 |
| Résumé long | 229 |
| 8.1 L'économie politique de la Grèce ancienne, cadre historique et revue de la littérature | 230 |
| 8.1.1 La naissance de la <i>polis</i> | 231 |
| 8.1.2 L'invention de la <i>dēmokratia</i> | 235 |
| 8.1.3 <i>Analytic narratives</i> et cliométrie | 237 |
| 8.2 L'extension de la cliométrie à l'histoire ancienne par la complexité | 238 |
| 8.2.1 L'importance de la distinction épistémologique entre histoire et économie | 239 |
| 8.2.2 Une nouvelle cliométrie associée à la complexité | 242 |
| 8.3 Une Perspective sur la Grèce ancienne fondée sur la cliométrie & complexité | 245 |
| 8.3.1 Approche empirique et données | 246 |
| 8.3.2 Résultats | 252 |
| 8.4 Analyse cliométrique du choix du nom en Grèce ancienne | 256 |
| 8.4.1 Modèle | 259 |
| 8.4.2 Résultats | 261 |
| 8.5 Analyse cliométrique du choix des dieux en Grèce ancienne | 263 |
| 8.5.1 Modèle | 265 |
| 8.5.2 Résultats | 267 |
| Full Bibliography | 271 |

List of Tables

| | | |
|------|---|-----|
| 3.2 | Summary Statistics on Distribution Fits | 79 |
| 3.4 | Summary Statistics Across Small Sample Distribution Fits | 81 |
| 3.7 | Summary Statistics on Distribution Fits Across Literary Genres | 86 |
| 3.10 | Summary Statistics on Distribution Fits | 90 |
| 3.12 | Summary Statistics on Distribution Fits for Votive Acts | 93 |
| 3.13 | Summary Statistics For Name Distributions Fits by Gender | 97 |
| 3.15 | Summary Statistics on Distribution Fits for the Number of In-Degrees in the Naming Network | 99 |
| 3.19 | Summary Statistics on Degree Distribution Fit | 104 |
| 4.7 | Summary Statistics on Distribution Fits Across Places and in Aggregate . | 152 |
| 5.1 | Summary Statistics on BDEG Data Across All <i>Poleis</i> | 183 |
| 5.2 | Summary Statistics on BDEG Data On <i>Poleis</i> with more than 25 Observations | 183 |
| 5.3 | Ranking of the First and Second Gods Most Commonly Mentioned in each <i>Polis</i> | 183 |
| 5.5 | Empirical Drivers of the Number of Gods For Which References Were Founds | 186 |
| 5.7 | Empirical Regressions for Predicted vs. Actual Probability of Worship for the Least Worshipped God | 188 |
| 5.8 | Summary Statistics on Distribution Fits Across All <i>Poleis</i> | 189 |
| 5.9 | Summary Statistics on Distribution Fits across <i>Poleis</i> | 190 |
| 5.10 | Summary Statistics on Distribution Fits across <i>Poleis</i> Excluding Observa- tions of Multiple BDEG Entries | 190 |
| 5.14 | Regressions of Distribution Parameters with All Variables | 194 |
| 5.15 | Regressions of Distribution Parameters with Selected Variables | 195 |

List of Figures

- 3.1 Comparison of Power Law, Exponential and Lognormal Distributions on a Large Sample 78
- 3.3 Comparison of Power Law, Exponential and Lognormal Distributions on Small Samples 80
- 3.5 Distribution of the Number of Words by Work Period 83
- 3.6 Log/Log Cumulative Distribution on Diorisis Corpus for Various Literary Genres 84
- 3.8 Log/Log Cumulative Distributions of the Number of Words in Inscriptions for Selected Areas 88
- 3.9 Log/Log Cumulative Distribution on Literary Texts and Inscriptions . . 89
- 3.11 Log/Log Cumulative Distribution of the Number of Votive Acts Among Gods, for Selected *Poleis* 92
- 3.14 Naming Network for the Oion Demes 98
- 3.16 Syntactic Network Representation of Testimony #100 101
- 3.17 Distance Network Representation of Testimony #100 102
- 3.18 Degree Distribution of Divinities in Distance-Weighted Formulae Networks 104
- 3.20 Log/Log Cumulative Distribution of Diorisis and MAP Corpora 105

- 4.1 Difference in Cheating Probability with Private Information and with No Information 133
- 4.2 Difference between the Cheating Probability with Partially Common Information and with No Information 136
- 4.3 Difference between the Cheating Probability with Partially Common Information and with Private Information 137

| | | |
|------|---|-----|
| 4.4 | Rank by Number of Carriers of the Optimal Name that Minimizes the Probability of Being Cheated | 140 |
| 4.5 | Log/log Plot of the Number of Distinct Men Names as a Function of Total Number of Mentions Across <i>Poleis</i> | 150 |
| 4.6 | Log/log Plot of Name Distributions Across <i>Poleis</i> for Men | 151 |
| 4.8 | Relationship Between Distribution and Number of Names | 153 |
| 5.4 | Log/log Plot of God Worship Distribution For Various <i>Poleis</i> | 185 |
| 5.6 | Predicted vs. Actual Probability of Worship for the Least Worshipped God | 187 |
| 5.11 | Log/log Plot of God Worship Distribution Grouped by Political Regime | 191 |
| 5.12 | Log/log Plot of God Worship Distribution Grouped by Hellenicity . . . | 192 |
| 5.13 | Log/log Plot of God Worship Distribution Grouped by <i>Polis</i> Size | 193 |
| 7.16 | Query for a Particular Entry in the PHI | 207 |
| 7.17 | Query for a Name in the LGPN | 215 |
| 7.18 | Query for a Particular Entry in the BDEG | 222 |

Introduction

Et je vous avais bien averti, pourtant, tout à l'heure encore : l'arithmétique mène à la philologie, et la philologie mène au crime¹. . .

And I warned you, still, earlier again: arithmetics lead to philology, and philology leads to crime. . .

Mathematics leading to philology and philology to crime are supposed to make sense within the context of an absurdist theater play. This fact tells us, at the very least, that mathematics and philology are not supposed to go well together in a non-absurd world. Philology mostly conjures up the close reading of ancient texts, whereby one may painstakingly extract nuggets of information about the very distant world of Antiquity. What could be, indeed, further remote from the highly formal study of mathematical quantification? It would hence not be a big gap to then claim that linking ancient history and mathematical models is somewhat absurd. Still, there has been a good amount of research, relying on formal methods, that tackled ancient history, and in particular in economics, so that connecting these fields must not be entirely absurd. We will see, nevertheless, that most of the analyses that have been carried out in this way have not been very philological in nature, because they only leant on highly summarized information, and not on the fully detailed historical material that philologists typically work on. Further, cliometricians, who, among economists, would be the most likely to operate on detailed historical data, have not taken much of an interest in ancient history.

In this work, we will attempt to bridge the gap, and connect formal economic analysis with the full detail of ancient sources, thereby extending traditional cliometrics, thanks to the recent developments that have associated cliometrics and complexity analysis, developments better known as cliometrics & complexity. We structure our

¹*La Leçon*, Ionesco (1954).

approach into two broad parts. In Part I, we concern ourselves with the general epistemological framework in which one may apply cliometrics to ancient history. In this first part, there are three chapters. Each chapter includes its own bibliography.

In Chapter 1, we concern ourselves with the existing research in economics that has looked into ancient history, in particular into ancient Greece. We briefly review the history of archaic and classical Greece, focusing on the development of the *polis* and of ancient Greek democracy. We examine the political economy research that has provided a formal background for this historical development, and stress that it belongs to analytic narratives rather than to cliometrics. This first chapter is also a good occasion to introduce some aspects of ancient Greek history and culture, and situate some key dates and periods.

In Chapter 2, we propose to extend cliometrics in a certain way, so that they may encompass ancient history. Traditional cliometrics usually focus on economic data from the modern and contemporary periods, and do not have much to do with ancient history, mostly because of a lack of relevant data. Separately, the field of cliometrics and complexity, by looking at data in the light of complex systems analysis, gives access to a broader range of sources. Concentrating on the distinction between cliometrics and historical economics, we explore the epistemic gap between economics and history, which we reduce to two fundamental differences: the relationship to primary sources, and the presence of a nomothetic framework. Using this gap as a guide, we argue for a logical expansion of cliometrics and complexity: this way, they do not have to be about the economy, but can operate on primary historical sources, and could address a much broader set of periods, societies, and phenomena, leaning on microeconomic models. Redefining cliometrics in that way gives them access to the extensive corpora of historical material that digital humanities have produced. Working closer to primary sources contributes to bridging the epistemic gap between economics and history, and the systematic and explicit way in which cliometrics and complexity tackle data contributes to making historical research more scientific.

In Chapter 3, we take the framework from Chapter 2, and illustrate its application to ancient Greek culture, reviewing various important data sources in the process. By focusing on primary non-economic sources, thanks to complexity sciences, a cliometric

approach allows us to frame into economic terms the massive and complex cultural material we gathered from the Ancients. Concentrating on the generating mechanisms for various large-scale textual sources, in a complexity economics perspective, we consider how distinct constraints and objectives lead to measurable differences between judicial speeches, poetry, and epigraphic texts, for example. Based on thousands of inscriptions, we account for whom the Ancients worshiped with a shared-resources paradigm, and show how people's names and naming strategies exhibit patterns consistent with a high degree of conformity. Finally, we show that the religious formulae the Greeks used to address their gods, in spite of their short and condensed nature, may share the features of a language, and appear to be constructed as an optimal communication medium. Our economic analyses in this chapter remain on the surface, but they seek to illustrate how deeply one can drill into primary sources thanks to complex systems methods.

Part II relies on the new epistemology of ancient cliometrics we developed in Part I, and focuses on the economic and cliometric detailed analysis of particular facets of ancient Greek life. This second part includes the next two chapters and the conclusion.

In ancient Greece, people carried a single name, and some names were quite common, while others were very rare. If these names were used to distinguish people, why didn't everybody have a different name? In Chapter 4, looking into the manner in which the ancient Greeks picked names, we develop an economic model for the existence of names, as a way of exchanging identification information. Considering different information frameworks, we justify the exchange of names in this context as the best system in order to promote cooperation. We then study the optimal choice of names in these conditions and show that the impact of strategic naming on the distribution of names works as an alteration of existing mean-field approaches to name dynamics, which converge to power laws. Strategic naming adds a degree of freedom in the relationship between the observed number of names and the shape of the power law distribution of these names. Confronting these results to empirical data from the archaic and classical periods, we observe that a form of conformist strategic naming could account for the particular shape of the name distribution in Ancient Greece, which markedly differs from contemporary data.

The Greek pantheon is a well-known aspect of ancient culture; in ancient Greek polytheism, worshippers could indeed choose which gods they would address, and in doing so they expected some form of benefit in a *quid pro quo* relationship. In Chapter 5, we look into the optimal choice of which god to worship as a function of the presumed strategy of the gods for returning favors to the worshippers, and relate it to a form of divine efficiency measure. At the equilibrium, the model also shows that the least-worshipped god receives at least a certain volume of devotion. We propose two different approaches that may account for the assumed divine efficiency measure, one based on projecting characteristics of human performance onto the gods, and the other based on a random growth model for the benefits of addressing each god. Both approaches imply that observed acts of worship would follow a type of power law. We gathered data from a large volume of epigraphic and literary sources on actual acts of worships from the ancient Greeks, and it allows us to show that the distributions of these acts at the *polis* level effectively follow power laws with a high degree of regularity. The number of votive acts towards the least-worshipped gods also match the model's prediction. We test the extent to which the known characteristics of the *poleis*, the ancient Greek cities, affect the shape of these distributions, and find little explanatory power. The shape of the distribution of votive acts across gods hence appears to have followed a general law, at least for the Ancient Greeks.

The overall conclusion to this work is at the end of Part II. Finally, the Appendix details the computer programs that were used to access, extract, and structure the primary sources used in this work. The Appendix also includes a long summary of this work in French, and a full bibliography, while each chapter also has its own bibliography.

Note that particular attention has been paid to make the mathematical expressions in the text as explicit as possible, to avoid any ambiguities. This comes at the cost of a slightly more dense formalism. In addition, equations are not numbered: the intention has been that the text itself should be clear enough, to make the logical connections as explicit as possible. Also, in line with common practice in probability, some particular notations are sometimes used throughout the text, although they may not be standard in economics. These notations hopefully contribute to disambiguating mathematical expressions:

- For x and y quantities that can be ordered, $x \wedge y = \min\{x, y\}$ and $x \vee y = \max\{x, y\}$.
- If X is a random variable with a density f_X , then $\mathbb{P}[X \in dx]$ is defined such that for all measurable f , $\int f(x)\mathbb{P}[X \in dx] = \int f(x)f_X(x)dx$. For discrete variables, one can use the Dirac measure in order to simplify expressions: if $X \in \{a, b\}$, for example, and $\mathbb{P}[X = a] = p$, then $\mathbb{P}[X \in dx] = [p\delta_a(x) + (1 - p)\delta_b(x)] dx$.
- The indicator function \mathbb{I} is such that: $\mathbb{I}_A(x) = 1$ if $x \in A$ and 0 otherwise. In probabilistic expressions, \mathbb{I}_A typically means the function operating on $\omega \in \Omega$.
- In games, one often needs to write expressions that are conditioned by some information. For a conditional expectation, this is usually expressed as $\mathbb{E}[A|B]$, but it is sometimes more practical to use a notation of the form $\mathbb{E}_B[A]$.

Part I

Complexity: An Epistemological Framework for the Cliometrics of Ancient Times

Chapter 1

The Political Economy of Ancient Greek *Polis* and *Demokratia*²

The history of archaic (8th-6th century BCE) and classical (5th-4th century BCE) Greece can, in many respects, be centered around two critical innovations: the creation of the *polis*, the city-state, and the subsequent emergence of *demokratia* in Athens, the ancient Greeks' particular form of participative democracy³. Numerous research works have in fact sought to place these historical processes in a political economy framework. In this chapter, we propose a brief overview of these approaches, and also ask whether these endeavours should be qualified as cliometric, since they strive to analyze historical material with an economic perspective.

We will begin by defining the *polis*, and we will examine the conditions that appear to have favored its emergence in archaic Greece. We will also review the particular aspects that political economists have concentrated on in their analyses of the ancient *polis*. Then, we will turn to the Athenian democracy, and discuss some of the underlying trends that made it possible. We will, in turn, discuss the flourishing economic research that aims to explain many of the idiosyncratic features of this democracy. Finally, we will qualify these economic approaches in contrast with cliometrics.

²A version of this chapter is to appear in *Investigaciones de Historia Económica / Economic History Research*.

³While other democratic states existed in ancient Greece, Athens is by far the most documented historically.

1.1 The Birth of the *Polis*

The ancient Greek city-state, the *polis*, was an autonomous small-sized political community with an urban centre, and over a thousand of such *poleis* have been identified, each with their own specificity (Hansen & Nielsen 2004). In order to understand and qualify the emergence of the *polis*, we first look at the social conditions during the “dark ages” of ancient Greece, around the 10th century BCE, when there were no significant urban centres yet, and the communities could not be qualified as city-states. Between the 13th and the 11th centuries BCE, a large number of Mycenaean sites were abandoned, according to archaeological evidence (Murray 2017). Their complex socioeconomic structure gave way to a more egalitarian and cohesive one. As the centralized organization of the Mycenaean palaces, and their political power, disappeared, they stopped collecting agricultural production. There were therefore greater individual surpluses for the population, which transformed some of them into small landowners, with a self-interest in the long-term exploitation of their lands. The decline in population and the generalized loss of wealth early in the period also contributed to greater social and economic equality, which is observed through burial customs. Even if there were “big men”, chiefs, aristocrats or kings, they did not have the prerogative of great luxury, nor of great wealth, great influence, or military supremacy. Indeed, the *basileis*, the kings or aristocratic leaders described in the epics, did not rest on a very solid political structure, as the difficulties of Telemachus in the *Odyssey* illustrate⁴: since his father, the “big man”, is not there, his power is not automatically transmitted. In addition, Odysseus, who is described in the *Odyssey* as a *basileus* of infinite resources, owns a total of about thirty herds⁵, quite a different order of magnitude when compared to the riches of the Mycenaean palaces. We can speculate that this trend was reinforced by the disappearance of writing, and this contributed to greater cultural equality and to the reduction of possible attributes of differentiation, as between the rulers and magistrates of the Mycenaean palaces and the rest of the population. Linear B, the Mycenaean writing system, was in fact associated with the centralising power of the palaces, and served to implement the gathering of surplus from the population.

⁴Homeric texts were likely written in the 8th century BCE, but refer to earlier times, around the 10th century BCE, see Fowler (2004), p. 206-232.

⁵Homer, *Odyssey*, 14, 96-102.

The disappearance of most trade networks initially led to a scarcity of copper and tin, which had to be imported. This led to the development of iron metallurgy, which is more difficult but makes use of a much more widely distributed material. The proportion of bronze objects in burials between the Sub-Mycenaean and Late Geometric periods declined (Whitley 1991, 358). We know that in archaic Greece, many conflicts between emerging cities took place: Hansen & Nielsen (2004) identified dozens of total city destructions. Iron, more widely available than bronze, made it easier for people to arm themselves: according to the same source, the number of weapons per grave increased during the period. Access to weapons was therefore no longer the prerogative of a warrior class: a larger part of the population could acquire their own (iron made) military equipment for the defense of their city. The military organisation of the *phalanx* emerged, as a system of protection of the newly constituted city-states from any kind of foreign invasion or threat. It was composed of *hoplites*, those who could acquire or even produce themselves their equipment, generally free farmers. Working tools could also be improved: for example, according to Hesiod, it is an iron ax that cuts down the tree⁶, which illustrates widespread use of iron tools in the 8th century. With better tools, agricultural yields could increase.

In the 8th century, the acceleration of trade, the growth of the population and the expansion of numerous colonies are often considered as markers of the emergence of the *polis*. These colonies, and their own colonies in turn, constituted a dense network of commercial relations (Malkin 2011). In the archaic period, some products were manufactured specifically for distant markets (Osborne 1996). Colonies played a role in regulating growth, in addition to allowing access to distant resources: times of stress could trigger the founding of colonies. For example in the 6th century, according to Herodotus, during a drought lasting seven years, the inhabitants of Thera were forced to send settlers by choosing them at random, one out of every two brothers⁷. In times of civil unrest, or *stasis*, this same approach also made it possible to defuse fratricidal civil wars.

Writing was rediscovered by the Greeks through the Phoenician alphabet in the 8th century. In the context of the emergence of the *polis*, where differences in wealth were

⁶Hesiod, *Works and Days*, 420.

⁷Herodotus, 4, 151-153.

less pronounced than in the Mycenaean period, the ability to read and write may have been more widely distributed among the population, with significant variations across *poleis*—in the case of Athens, as Stoddart & Whitley (1988) puts it, “generalised social literacy”. This is reflected, among other things, in the wide range of texts produced in archaic Greece, which illustrate the widespread use of writing in contrast with the use of Mycenaean Linear B, purely for accounting purposes. In particular, written laws quickly emerged: according to Strabo⁸, the first written laws were due to the legislator Zaleucus in Locri Epizephyrii at the beginning of the 7th century. Writing laws presumably reduces the influence of the aristocrats, who were previously the only ones able to establish what was rightful or not.

Greater egalitarian principles could also be seen in war itself: The *phalanx* was indeed a particular method of fighting, relying on individual courage and group cohesion, as opposed to heroic duels, the form of combat displayed in Homeric texts. Archaic Greek society remained nevertheless strongly status-based, with some prominent aristocrats (Zurbach 2013). This effective equality in combat, which was not necessarily reflected in political participation, led to tensions between part of the population and the aristocrats. The majority of the Greek cities generally evolved in the second half of the archaic period from aristocratic regimes toward tyranny (Hansen & Nielsen 2004, 1338–1340), as these tyrants (usually themselves aristocrats) appealed to the hoplites to grab their arms, and shortcut the aristocracy. The tyrants, however, generally obeyed the laws. For example, according to Thucydides, the government of Hipparchus and Hippias in Athens at the end of the 6th century was reasonable and respected the multitude, and the city retained the full benefit of its existing laws⁹.

How can the emergence of these politically and economically integrated city-states, out of this particular substrate, be accounted for in a political economy perspective? There is a framework for the analysis of constitutional design and state formation in political economy, generally relying on game theory, as is described, for example, in Laffont (2001), and regarding the foundation for state capacity, by Besley & Persson (2009); the question of establishing an autocracy has also been considered by Myerson (2008). The emergence of the Greek *polis* has been examined in the light of institutional economics

⁸Strabo, 6, 1, 8.

⁹Thucydides, 6, 54.

by historians inspired by economic approaches (Ober 2015), and by economists with an interest in ancient history. Some have examined the general conditions in which the *polis* emerged, while others focused on specific aspects of the ancient Greek society. Lyttkens (2006) took a rational actors perspective, and analysed how the varying conditions of competition for power between elites led to the emergence of institutions after the Dark Ages. Lyttkens (2013) also considered gradual technological change after 1200 BCE and its impact on social organisation. Particular conditions for the emergence of the rule of law are required, according to Fleck & Hanssen (2019): there is a trade-off due to the commitment to follow the rule in the future, and a majority-driven decision process minimises the cost of this trade-off. In particular, Pitsoulis (2011) proposes a model capturing the impact of the greater role of *hoplites* on the birth of a more egalitarian system. The emergence of a new political landscape, turned towards equality, can also be related to the common ideals that were developed, linked to athletics, war and the gradual emergence of property rights protection, since the Archaic times onward, in the perspective of new institutional economics (Kyriazis & Economou 2015; Economou & Kyriazis 2017). More specifically, the emergence of the particular organisation of military combat as a *hoplite phalanx* is studied in Economou & Kyriazis (2019), p. 12-43. The role and cost of war in classical Athens were effectively both very significant (Lyttkens & Gerding 2022), and having to make decisions about war and peace may justify the emergence of a process to handle social choice (Economou & Kyriazis 2016). In archaic Athens, some laws addressed social conflict; Schwuchow & Tridimas (2022) examine the optimality of one of Solon's laws, which prescribed that in the case of *stasis*, one needed to pick a side.

The evolution of political organisation in ancient Greece has also been related to economic growth or potential growth. Fleck & Hanssen (2006) used institutional economics to analyse the emergence of democracy in various Greek cities as a means of fostering personal investment in agriculture. They observed that in regions where the cultivation of the soil requires long-term investment, and therefore would imply an assurance of not being deprived from it in the future, the population may be demanding more guarantees from the aristocracy. Under their assumptions, a democratic regime is optimal when soil yields are low, because it induces better investment by farmers. This model takes into account the strong correlation that we observe between the richness of

the soil and the type of political regime: from the extremes of Thessaly with a very rich soil and a monarchical regime on one hand, to Athens with arid soil and a democratic regime on the other. Using a comparable framework, they also considered the transition through tyranny, and related it to growth potential (Fleck & Hanssen 2013, 2018). Effectively pointing out a form of virtuous circle, they showed in Fleck & Hanssen (2015) that the ancient Greek democratic system supported wealth creation. With another perspective on the creation of the democratic *polis*, McCannon (2012) considered that it was a way to alleviate the increase in wealth volatility across generations.

1.2 The Invention of *Demokratia*

We can define the Athenian *demokratia* as a participatory democracy by lot and without representation. In 507 BCE, Cleisthenes introduced initial reforms before the assembly which defined this *demokratia*, including *isonomia*, equal rights, and *isegoria*, equality of speech among all citizens. He relied on the existing written laws, in particular Solon's, who had set up the citizens' assembly and oligarchic census classes in 594 BCE (Stahl & Walter 2009). Cleisthenes was therefore also indebted to Peisistratus and Hippias, the tyrants who had continued to promote the functioning of the political system and respect for the law, as we pointed out earlier. The identity of the Athenian *polis* had been largely reinforced by the Peisistratid tyrants: temples were built, religious festivals and games instituted, literature centralised. Peisistratus instituted the Panathenaea, a festival and a set of games organised by Athens. These innovations helped establish a Panathenian identity above that of the traditional tribes and villages, which may have fostered collaboration among Athenians at large.

The democratic system promoted by Cleisthenes turned out to be totally different from what had been observed before. It was a substantial break in terms of political functioning: Vernant (1965) characterised democracy as the passage from the question of choice to that of how to choose, through the aggregation of the knowledge and skill of the population. Previously, in times of crisis, the *polis* appealed to a legislator or a tyrant. With *isonomia*, the system was designed to work and resolve crises on its own. Thanks to *isegoria*, the confrontation of ideas in the popular Assembly of citizens made it possible to exercise this rationality and to seek the best solutions. However, if everyone can give

their opinion, without necessarily having in-depth knowledge of the field in question, how can the city be properly managed? According to Vidal-Naquet ([1981] 2005), *hoplites* were not distinct from other combatants in inscriptions, and further points out that in his Funeral Oration, Pericles argues that it is not the preparation for war that distinguishes the Athenians, but bravery¹⁰. This stresses that it was not the technique or a particular expertise that took precedence, but effort and will. Ober (2008) looked into some aspects of how Athenian democracy worked, and in particular its ability to effectively aggregate information. For Ober, the government of Athens can best be viewed as a complex and efficient machine designed to identify and gather social knowledge and relevant techniques (Ober 2008, 118–167). By breaking traditional ties between tribal members and introducing randomness into the distribution of responsibilities, Cleisthenes' reforms forced Athenians to collaborate by sharing the private information each had. As a matter of fact, the development of a political economy thought in ancient Greece evolved hand in hand with the rise of *demokratia*, as a form of criticism of that political regime (Ober 1993; Carugati, Ober, & Weingast 2016; Canevaro 2018).

General economic research has addressed the logic of voting and the basic principles of democracy, such as in Davis, Hinich, & Ordeshook (1970) or Nurmi (1986). The design of the Athenian democratic system was innovative on many levels, however, and has raised a specific interest in institutional economics, with research works focusing its general functioning as well as on many of its particular features, such as the random selection of magistrates, ostracism, large judicial juries, or the voluntary contribution of the rich to the city's military or religious expenses, for example. The citizens' active engagement at every level of the *polis*, in spite of the fact they were not necessarily experts, has been shown to optimally reduce information asymmetry (Economou & Kyriazis 2022). While some positions were elected, many were selected by lot, and the optimality of this procedure has been shown to both ensure representativity and reduce the risk of corruption, but also reduced the effort necessary to choose (Tridimas 2011, 2012, 2022). Tangian (2008) analysed the specific set up of the Athenian assembly, the *boule* of 500 and its rotating committee of 50 in particular, and showed that the size of these representative bodies were well adapted to Athens. The frequency of votes in the Athenian participative democracy was high, and Tridimas (2017) showed

¹⁰Thucydides, 2, 39sq.

how, by increasing this frequency, citizens were obtaining a greater utility. The famous principle of ostracism, used up to 415 BCE, was a political device that allowed the citizens to banish whoever would receive the most vote against them in a particular session of the assembly. Tridimas (2016) looked into how these ostracism votes operated, and analysed them as a negative referendum on the most visible politicians. Another particular institution, set up after 415 BCE, was the *graphe paranomon*: the right that each citizen had to sue anyone who argued anything at the assembly by stating they made an anticonstitutional proposal. If the suer failed to gather a certain number of votes, they were subject to heavy penalties. Lyttkens, Tridimas, & Lindgren (2018) analysed the outcomes from a few of these votes and show that such proceedings could in fact be used by politicians to measure the support they had. Such stabilisation mechanisms, intended to reduce the risk of extreme democratic decisions such as the disastrous Sicily campaign, accounted for a significant evolution of the *polis* over time (Halkos, Economou, & Kyriazis 2022).

As part of the democratic process, judicial decisions were handled by large popular juries, of varying sizes, where the jurors were selected by lot. The specifics of this legal setup have been examined from an economic perspective, as there were some particular rules, warranting close economic examination. For example, in homicide trials, after evidence was presented, the accused could choose to be exiled and not risk anything more, or continue with the trial. McCannon (2010a) has examined, in a game theoretical setting, the conditions under which a person would chose one option or the other. The famous trial of Socrates has also been scrutinised from an economic perspective: McCannon (2010b) analysed it from the perspective of the median juror's decision, and Guha (2011) asks whether Socrates's strategy was rational. McCannon (2011) focuses on the optimality of jury size in Athens, since they were sometimes as large as 2000.

Finally, finances were handled by the assembly, and the Athenians generally followed a degree of "fiscal discipline" (Bitros, Economou, & Kyriazis 2021, 48–55). Liturgies, and in particular trierarchies, were semi-voluntary commitments by rich citizens to cover particular expenses related to the *polis*, such as theater plays or naval defense. Kaiser (2007) studied this particular mechanism, and established the conditions in which citizens could willingly commit to, or would rather skip, their obligations. Tridimas (2020) examines liturgies as a competitive system between the rich, leading to

effective income taxation. The process through which financial decisions were voted has been studied by Tridimas (2013), in the particular case of the attribution of newly found silver in Attic mines to the building of a large combat fleet in 483 BCE. As a consequence of the war fleet, the Athenian economy evolved from an agrarian type in the Archaic period to a maritime and commercially extroverted one in the classical periods. Kyriazis & Economou (2019) and Economou & Kyriazis (2019), p. 28-42, review this economic evolution more broadly. Still in Athens, Tridimas (2015) sees a relationship between wealth, war and democracy, as relying on the poor to man the war fleet, Athens effectively redistributed wealth towards them, which in turn induced a preference in the democratic regime for war.

As we have seen, the history of archaic and classical Greece is documented well enough that we possess detailed descriptions of many social and economic constructs the Ancients lived with. The manner in which they waged war, the way small communities coalesced into city-states, trading with each other, how they effected political choice, even sometimes choosing tyranny—all are interesting challenges for political economy to account for. Further, in the particular case of the ancient Athenian form of democracy, some aspects of the regime may appear close and understandable to us, such as voting, while others, such as ostracism rules, the size of juries, or the option to walk away from a murder trial, appear so remote that framing them in an economic perspective really helps coming to terms with their internal logic.

1.3 Analytic Narratives vs. Cliometrics

The large majority of economic research we have mentioned above, which we have loosely categorized as political economy, may be more precisely siloed into microeconomics, institutional economics, or even simply game theory. Nevertheless, since the works in question have to do with history, and with some specific historical transitions in particular, we may ask whether they may be qualified as cliometric. Etymologically, cliometrics refer to the measurement of history, and more precisely, it is a domain of economics dedicated to the application of econometric methods to historical data (Diebolt & Parent 2011), which stresses the importance of actually using some volume of data. While the economic lineage of the analyses we discussed above cannot be put

into question, the extent to which they use data is quite limited. In fact, a majority of the political economy modeling that has been applied to ancient Greece only relies on a few summarized elements drawn from historical material, and as a result it appears that it may not be appropriate to qualify this work as cliometrics.

Applying an economic framework to historical context outside of an econometric perspective, and without a large dataset with which to test may be, more appropriately, qualified as part of *analytic narratives*. Analytic narratives were originally defined in Bates et al. (1998), and more recently in Mongin (2016); they seek to explain historical patterns or events using economic thought or modeling (Alexandrova 2009). While cliometrics tend to operate from the underlying historical data, and constructs an economic understanding of that historical data, analytic narratives operate the other way around, and project an explanation into the historical events, based on economic models.

1.4 Chapter Bibliography

- Alexandrova, Anna.** 2009. "When Analytic Narratives Explain." *Journal of the Philosophy of History* 3 (1): 1–24.
- Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, & Barry R. Weingast.** 1998. *Analytic Narratives*. Princeton: Princeton University Press.
- Besley, Timothy, & Torsten Persson.** 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218–1244.
- Bitros, George, Emmanouil M. L. Economou, & Nicholas C. Kyriazis.** 2021. *Democracy and Money: Lessons for Today from Athens in Classical Times*. Routledge International Studies in Money and Banking. London: Routledge.
- Canevaro, Mirko.** 2018. "What Was the Law of Leptines' Really About? Reflections on Athenian Public Economy and Legislation in the Fourth Century BCE." *Constitutional Political Economy* 29 (4): 440–464.
- Carugati, Federica, Josiah Ober, & Barry R. Weingast.** 2016. "Development and Political Theory in Classical Athens." *Polis: The Journal for Ancient Greek and Roman Political Thought* 33 (1): 71–91.
- Davis, Otto A., Melvin J. Hinich, & Peter C. Ordeshook.** 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64 (2): 426–448.
- Diebolt, Claude, & Michael Hauptert.** 2020. "How Cliometrics Has Infiltrated Economics – and Helped to Improve the Discipline." *Annals of the Fondazione Luigi Einaudi* 54 (1): 219–230.
- Diebolt, Claude, & Antoine Parent.** 2011. *Essais cliométriques : capital humain, monnaie et finance, croissance et cycles*. Bern: Lang.
- Economou, Emmanouil M. L., & Nicholas C. Kyriazis.** 2016. "Choosing Peace Instead of War: A Lesson from Athenian Democracy." *Peace Economics, Peace Science and Public Policy* 22 (2): 191–212.
- . 2017. "The Emergence and the Evolution of Property Rights in Ancient Greece." *Journal of Institutional Economics* 13 (1): 53–77.
- . 2019. *Democracy and Economy: An Inseparable Relationship from Ancient Times to Today*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- . 2022. "Asymmetric Information, Social Choices, and Democracy." In *Democracy in Times of Crises: Challenges, Problems and Policy Proposals*, edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis, & Athanasios Platias, 117–136. Cham: Springer International Publishing.
- Fenoaltea, Stefano.** 2019. "Spleen: The Failures of the Cliometric School." *Annals of the Fondazione Luigi Einaudi* 53 (2): 5–24.
- Fleck, Robert K., & F. Andrew Hanssen.** 2006. "The Origins of Democracy: A Model with Application to Ancient Greece." *The Journal of Law and Economics* 49 (1): 115–146.
- . 2013. "How Tyranny Paved the Way to Democracy: The Democratic Transition in Ancient Greece." *The Journal of Law and Economics* 56 (2): 389–416.
- . 2015. "The Foundations of Wealth-Enhancing Democracy: Aristotle, Lindahl, and Institutional Design in Ancient Greece." Working Paper. Clemson: Clemson University. <https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=13727&tk=KZyieYK3>.

- . 2018. "What Can Data Drawn from the Hansen-Nielsen Inventory Tell Us about Political Transitions in Ancient Greece?" In *Ancient Greek History and Contemporary Social Science*, edited by Mirko Canevaro, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, 213–238. Edinburgh: Edinburgh University Press.
- . 2019. "Engineering the Rule of Law in Ancient Athens." *The Journal of Legal Studies* 48 (2): 441–473.
- Fowler, Robert L., ed.** 2004. *The Cambridge Companion to Homer*. Cambridge Companions to Literature. Cambridge: Cambridge University Press.
- Gonzales-Feliu, Jesus, & Antoine Parent.** 2016. "Clio-Combinatorics: A Novel Framework to Analyze Military Logistics Choices Using Operations Research Techniques." In *Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations*, edited by Alberto Ochoa-Zezzatti, Jöns Sánchez, Miguel Gastón Cedillo-Campos, & Margain de Lourdes, 79–101. Advances in Logistics, Operations, and Management Science. Hershey: IGI Global.
- Guha, Brishti.** 2011. "Preferences, Prisoners and Private Information: Was Socrates Rational at His Trial?" *European Journal of Law and Economics* 31 (3): 249–264.
- Halkos, George E., Emmanouil M. L. Economou, & Nicholas C. Kyriazis.** 2022. "Tracing the Optimal Level of Political and Social Change Under Risks and Uncertainties: Some Lessons from Ancient Sparta and Athens." *Journal of Risk and Financial Management* 15 (9): 416.
- Hansen, Mogens Herman, & Thomas Heine Nielsen.** 2004. *An Inventory of Archaic and Classical Poleis*. Oxford: Oxford University Press.
- Ismard, Paulin.** 2005. "Le désassemblage démocratique ou la cité antiplatonicienne." *Gestes*, 2005. <https://www.revue-geste.fr/articles/geste2/GESTE%2002%20-%20Assembler%20-%20Ismard.pdf>.
- Kaiser, Brooks A.** 2007. "The Athenian Trierarchy: Mechanism Design for the Private Provision of Public Goods." *The Journal of Economic History* 67 (02).
- Kyriazis, Nicholas C., & Emmanouil M. L. Economou.** 2015. "Macroculture, Sports and Democracy in Classical Greece." *European Journal of Law and Economics* 40 (3): 431–455.
- . 2019. "The Ancient Athenian Economy: A Review." In *The Influence of Hellenic Philosophy on the Contemporary World*, edited by John G. Dellis & Stephanos A. Païpetēs, 56–67. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Laffont, Jean-Jacques.** 2001. *Incentives and Political Economy*. Clarendon Lectures in Economics. Oxford: Oxford University Press.
- Lyttkens, Carl Hampus.** 2006. "Reflections on the Origins of the Polis: An Economic Perspective on Institutional Change in Ancient Greece." *Constitutional Political Economy* 17 (1): 31–48.
- . 2013. *Economic Analysis of Institutional Change in Ancient Greece: Politics, Taxation and Rational Behaviour*. Routledge Explorations in Economic History 58. London: Routledge.
- Lyttkens, Carl Hampus, & Henrik Gerding.** 2022. "Coping with Chronic Warfare: The Athenian Experience." In *Democracy and Salamis: 2500 Years After the Battle That Saved Greece and the Western World*, edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis, & Athanasios Platias, 181–200. Cham: Springer International Publishing.

-
- Lyttkens, Carl Hampus, George Tridimas, & Anna Lindgren.** 2018. "Making Direct Democracy Work: A Rational-Actor Perspective on the *Graphē Paranomon* in Ancient Athens." *Constitutional Political Economy* 29 (4): 389–412.
- Malkin, Irad.** 2011. *A Small Greek World: Networks in the Ancient Mediterranean*. Greeks Overseas. New York: Oxford University Press.
- Margo, Robert A.** 2018. "The Integration of Economic History into Economics." *Clometrica* 12 (3): 377–406.
- McCannon, Bryan C.** 2010a. "Homicide Trials in Classical Athens." *International Review of Law and Economics* 30 (1): 46–51.
- . 2010b. "The Median Juror and the Trial of Socrates." *European Journal of Political Economy* 26 (4): 533–540.
- . 2011. "Jury Size in Classical Athens: An Application of the Condorcet Jury Theorem." *Kyklos* 64 (1): 106–121.
- . 2012. "The Origin of Democracy in Athens." *Review of Law & Economics* 8 (2).
- Mongin, Philippe.** 2016. "What Are Analytic Narratives?" Working Paper ID 2796567. Rochester: Social Science Research Network. <https://papers.ssrn.com/abstract=2796567>.
- Murray, Sarah.** 2017. *The Collapse of the Mycenaean Economy: Imports, Trade, and Institutions 1300–700 BCE*. Cambridge: Cambridge University Press.
- Myerson, Roger B.** 2008. "The Autocrat's Credibility Problem and Foundations of the Constitutional State." *American Political Science Review* 102 (1): 125–139.
- Nurmi, Hannu.** 1986. "Mathematical Models of Elections and Their Relevance for Institutional Design." *Electoral Studies* 5 (2): 167–181.
- Ober, Josiah.** 1993. "Thucydides' Criticism of Democratic Knowledge." In *Nomodeiktēs: Greek Studies in Honor of Martin Ostwald*, 81–98. Ann Arbor: University of Michigan Press.
- . 2008. *Democracy and Knowledge: Innovation and Learning in Classical Athens*. Princeton: Princeton University Press.
- . 2015. *The Rise and Fall of Classical Greece*. Princeton: Princeton University Press.
- Osborne, Robin.** 1996. "Pots, Trade and the Archaic Greek Economy." *Antiquity* 70 (267): 31–44.
- Pitsoulis, Athanassios.** 2011. "The Egalitarian Battlefield: Reflections on the Origins of Majority Rule in Archaic Greece." *European Journal of Political Economy* 27 (1): 87–103.
- Schwuchow, Soeren C., & George Tridimas.** 2022. "The Political Economy of Solon's Law Against Neutrality in Civil Wars." *Public Choice* 192 (3-4): 249–272.
- Stahl, Michael, & Uwe Walter.** 2009. "Athens." In *A Companion to Archaic Greece*, edited by Kurt A. Raaflaub & Hans van Wees, 138–161. Blackwell Companions to the Ancient World. Malden: Blackwell Publishing.
- Stoddart, Simon, & James Whitley.** 1988. "The Social Context of Literacy in Archaic Greece and Etruria." *Antiquity* 62 (237): 761–772.
- Tangian, Andranik.** 2008. "A Mathematical Model of Athenian Democracy." *Social Choice and Welfare* 31 (4): 537–572.

- Tridimas, George.** 2011. "A Political Economy Perspective of Direct Democracy in Ancient Athens." *Constitutional Political Economy* 22 (1): 58–82.
- . 2012. "Constitutional Choice in Ancient Athens: The Rationality of Selection to Office by Lot." *Constitutional Political Economy* 23 (1): 1–21.
- . 2013. "Homo Oeconomicus in Ancient Athens: Silver Bonanza and the Choice to Build a Navy." *Homo Oeconomicus* 30 (4): 435–458.
- . 2015. "War, Disenfranchisement and the Fall of the Ancient Athenian Democracy." *European Journal of Political Economy* 38: 102–117.
- . 2016. "Conflict, Democracy and Voter Choice: A Public Choice Analysis of the Athenian Ostracism." *Public Choice* 169 (1-2): 137–159.
- . 2017. "Constitutional Choice in Ancient Athens: The Evolution of the Frequency of Decision Making." *Constitutional Political Economy* 28 (3): 209–230.
- . 2020. "Modelling the Quest for Status in Ancient Greece: Paying for Liturgies." *Homo Oeconomicus* 37 (3-4): 213–236.
- . 2022. "Policy Making by Randomly Selected Citizens: The Perspective of Elected Politicians." In *Democracy in Times of Crises: Challenges, Problems and Policy Proposals*, edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis, & Athanasios Platias, 117–136. Cham: Springer International Publishing.
- Vernant, Jean-Pierre.** 1965. "Espace et organisation politique en Grèce ancienne." *Annales ESC* 20 (3): 576–595.
- Vidal-Naquet, Pierre.** (1981) 2005. *Le chasseur noir : Formes de pensée et formes de société dans le monde grec*. Poche 194. Paris: La Découverte.
- Whitley, James.** 1991. "Social Diversity in Dark Age Greece." *The Annual of the British School at Athens* 86: 341–365.
- Zurbach, Julien.** 2013. "The Formation of Greek City-States: Status, Class, and Land Tenure Systems." *Annales HSS* 68 (04): 615–657.

Chapter 2

Extending Cliometrics to Ancient History with Complexity¹¹

The etymological components of “cliometrics” refer to history on one hand, and more precisely to the Muse Κλειώ, who inspired poets when they sung the glory (κλέος) of past heroes, and to the taking of a measure on the other hand. Although there is no relation with economics in this etymology, cliometrics have nevertheless become a synonym for “quantitative historical economics”. As a consequence, where there is no economic data, there are no cliometrics; in particular, there is little research in cliometrics concerned with ancient history: the field’s manual, Diebolt & Hauptert (2019), does not even mention the Antiquity. One critic that has been raised towards cliometrics is also their lack of concern for history: this lack has been commented as a form of loss from the standpoint of economists (Fenoaltea 2019), and also as an opportunity for historians to try and use more economic tools in their analyses (Boldizzoni 2011). While cliometrics seemingly suffer from a lack of interest in history, economics interest themselves in an ever-increasing range of fields. Indeed, economics in general appear to seep into many aspects of social sciences, from toilet seat etiquette (Choi 2011) to the use of contraceptive sponges in *Seinfeld* (Dixit 2012), going much further than a strict focus on the economy. These various questions that economics may examine are often posed through simple stylized facts or observations, such as with both examples above. In cliometrics, however, one needs to have the relevant historical data, in the

¹¹This chapter is based in part on the article “Putting Clio Back in Cliometrics”, which has been published in *History & Theory* in 2022, and presented similar arguments from an historian perspective.

right form, in order to carry out any analysis. This necessity may account for the fact that cliometrics have not looked into historicizing questions of that sort.

New perspectives into data, in particular historical data, may constitute an opportunity to reassess what cliometrics are about, and what questions they may ask. The use of complex systems analysis in cliometrics, coalescing into the new field of cliometrics & complexity, has effectively set forth new hermeneutics to data, relative to traditional cliometrics. While complexity in economics may have many potential applications (Arthur 2021), it seems particularly well suited to the study of history through an economic perspective, offering the right tools to tackle the analysis of intricate systems over time, as Durlauf et al. (2023) suggest. For example, Abry et al. (2022) have looked into the history of market integration using dynamic networks. Relying on complexity analyses, one can allow the data's full granularity to percolate through to economic models, without having to reduce it to aggregate or stylized facts. Although research in the field has largely remained concentrated on the history of the economy, it has also expanded beyond, for example with an examination of the optimality of particular military allocation decisions (Gonzales-Feliu & Parent 2016).

In this chapter, we will try and understand how cliometrics effectively cover much less ground than economics in general, and, based on the differential epistemology between economics and history, propose an expanded version of cliometrics and complexity that could address these limitations. In the first section, we will focus on defining history in contrast with economics, and show how historical economics and cliometrics are defined along the same contrasting lines. The second section will more specifically concentrate on the epistemic differences between history and economics, focusing in particular on the use of primary sources as opposed to stylized facts, and on the nomothetic aspect of economics and cliometrics, which is not compatible with the historical method. Then, in the third section, we will examine the potential for new and expanded cliometrics & complexity, using the epistemological gap between economics and history as a guide. We will show how cliometrics could still be fully relevant to economics, but without having to be about the economy. Freed from this constraint, they can more directly operate on primary sources, and become epistemologically closer to history. In a world where large swaths of historical material have been digitized, and in which digital humanities have become a necessity, cliometrics & complexity, by

formalizing its operations on data, can further make the historiographical process more scientific. Finally, we discuss an example, in the field of ancient Greek religion modeling.

2.1 History, Economics and Cliometrics

In order to best understand and analyze the epistemic differences between history and economics, we begin by defining some of these terms. Equipped with these definitions, we will then turn to analyzing the differences between cliometrics and economic history.

2.1.1 Defining History and Economics

The positivists of the second half of the nineteenth century, starting with Auguste Comte, were the originators of modern definitions of history. Comte believed that history should be based on the techniques of the hard sciences, a viewpoint that was reflected in the first historiography course, which was offered in 1896 at the Sorbonne (Langlois & Seignobos [1898] 2014). In the 20th century, historians' perspectives on the definition of history changed and converged. Marrou stated simply: "history is the knowledge of the human past" (Marrou [1954] 2016, 24). This perspective on history differs from past positivist approaches because it places the human back into the science and stresses that historical knowledge is constructed from the ground up. History can be considered not as a science but as the scientifically developed narrative of humankind's actions and inventions, according to Lucien Febvre's definition of history, more general than Marrou's¹². Both definitions highlight history's aim toward science while also effectively isolating it from the hard sciences. This viewpoint is reinforced in Raymond Aron's seminal work on the philosophy of history¹³.

Economics can be understood to mean a fairly wide range of things. One manner to define economics is as "the social science that studies the production, distribution, and consumption of goods and services" (Krugman & Wells 2015). Implicitly defining

¹²See Febvre ([1952] 1992), "l'étude, scientifiquement conduite, des diverses activités et des diverses créations des hommes d'autrefois, saisis à leur date, dans le cadre des sociétés extrêmement variées et cependant comparables les unes aux autres (c'est le postulat de la sociologie)", and also "je qualifie l'histoire d'étude scientifiquement menée, et non pas de science".

¹³See Aron ([1938] 1991), p. 13: "notre livre conduit à une *philosophie historique* qui s'oppose au rationalisme scientifique en même temps qu'au positivisme." This is also in line with Marrou's understanding of what constitutes historical truth: "ni objectivisme pur, ni subjectivisme radical", in Marrou ([1954] 2016), p. 221.

the economy, Krugman & Wells (2015) also define economics as “the social science that studies the economy”. Microeconomics, which pertain to “the study of how people make decisions and how these decisions interact”, are a sub-field of economics but, thanks to their broad definition, appear to encompass a wider range of topics than economics itself. Microeconomics and economics in the classic sense are hence somewhat distinct fields of study, microeconomics being nevertheless sometimes employed as a tool to support economics in the traditional sense. The popular notion of “freakonomics” (Levitt & Dubner 2009), which apply microeconomic analysis to non-economic settings, such as sumo wrestling, gang member strategies and a whole gamut of other situations, have somewhat stretched the bounds of conventional economics, such as Choi (2011) and Dixit (2012) do. Although these topics would all be considered fairly typical areas of research in sociology or psychology, the “freak” element of freakonomics comes from the idea of linking these topics with microeconomics. Fine & Milonakis (2009) have examined this “freak” extension of economics into the social sciences in its historical context, and qualify it as a form of economic epistemological expansionism.

2.1.2 Cliometrics and Historical Economics

Cliometrics and historical economics are not the same thing. According to Diebolt & Parent (2011), as we pointed out in Chapter 1, cliometrics are a sub-field of economics that apply econometrics and economic theory to the data constructed by historical economics. Historical economics are concerned with the study of past economic occurrences. Cliometrics are seen by many as an extension of economic history rather than a separate science, the “new historical economics,” where “new” may be interpreted to imply “more formal” (Hauptert 2019). In particular, this perspective is apparent in Douglas North’s work (Brownlow 2010), and the shift in how economic history has been researched is well documented (Margo 2018), but for some, cliometricians have lost a certain historical perspective in this evolution (Cesarano 2006). The definition effectively states that cliometrics put historical economics within the context of economics. The importance of temporal seriality also distinguishes cliometrics from historical economics because, as Parent (2004) discussed, while economics in general may be nomothetic and postulate laws that are presumed to be universally valid, applying them in history has focused on changes through time more than on the specific circumstances at a

given point in the past, which is often more important to historians. Thus, to the risk of stating a tautology, we might say that one important distinction between historical economics and cliometrics is that the latter have been primarily studied by historians, whilst cliometrics have tended to remain the domain of economists.

One of the first significant historiographic schools to specifically focus on economic and social history was the French *Annales* endeavor. Four decades of history were dominated by the historical economics that the *Annales* promoted. Though on a smaller scale, historical economics had existed before to that; Hauptert (2019) situates their origins to the late nineteenth century in Germany and England. At the end of the 1960s, the significance of historical economics for historians was put into question, and it lost some of its prominence due to competition from other fields, such as anthropological history and the history of mentalities in particular. From the perspective of historians, using concepts and techniques from anthropology or psychology as opposed to orthodox economics started to provide more intriguing findings. From the standpoint of economists, historical economics has still remained a very active topic, which is why cliometrics and historical economics may now sometimes appear to be identical. Nevertheless, because they concern themselves with producing historical analysis, even if it is focused on the economy, and because they are assumed to employ the historical method, historical economics fundamentally make up a branch of history.

For Diebolt & Hauptert (2019), cliometrics “represent the quantitative projection of social sciences in the past.” Even if this may be etymologically correct, it is not actually the case, unless the social sciences were completely reduced to classical economics or the economy was the only meaningful quantification in the social sciences. Despite having a very broad meaning etymologically, cliometrics effectively only focus on the history of the economy, ignoring other elements of social interactions, even though many of these other aspects of social life may fall under an expanded purview of economics. Human capital, financial and monetary trends, or growth and cycles, in particular, have been important focuses in cliometrics and their study is presented as a means of resolving significant historical concerns. If they certainly are broad concerns, they are all nevertheless strictly related to the study of the economy. Even within ancient history, cliometrics are presented as fundamentally linked to the analysis of the economy. Diebolt (2011), for example, lays out the debate about applying cliometrics to ancient

history as revolving around whether there was a market economy or not. In cliometrics and complexity, the focus has for the most part also been on economic or financial phenomena, but the use of complex systems analysis has afforded a more systematic and more general perspective on the underlying data. Using complexity economics notions, whether they relate to networks, to time series, or to the emergence of patterns, has indeed renewed the cliometric perspective on economic or financial questions in the contemporary period (Bastidon et al. 2019, 2020; Bastidon & Parent 2022; Abry et al. 2022). Combining complexity economics, such as networks and non-linear econometrics, with institutional economics also gives us new angles into existing cliometric issues (Le Riche, Parent, & Zhang 2022). As we mentioned earlier, some work also extends beyond concerns about the economy, with for example Gonzales-Feliu & Parent (2016), who examined military strategy from the perspective of an optimal allocation on a road network, in order to analyze the logic that prevailed in some specific military choices.

It is useful to compare and contrast cliometrics with cliodynamics. Belonging neither in history nor in economics, cliodynamics do share certain methodological features with cliometrics, but they do not rely on economic models and concern themselves with a much wider range of subjects than the economy. Contrary to historical economics, cliodynamics approach history as a hard science, using mathematical models, often borrowed from the biological sciences and quantitative sociology, to general data and stylized facts (Turchin et al. 2013; Turchin [2003] 2018). Although cliodynamic technique does not involve theoretical economic models, it is not dissimilar to cliometrics from a methodological perspective, in that it frequently takes into account economic and demographic factors. It is possible to compare how cliodynamics approach history to how econophysics approach economics. Cliodynamics often use data that is very remote from the sources and spans a wide range of time periods and geographical regions. These data may comprise multiple categorizations, including population estimates, conflict statistics, or economic production, for example. Therefore, it is best categorized as a collection of stylized facts for each era or nation being studied. The sorts of problems investigated by cliodynamics recoup to a considerable degree with the study of meta-history started by Arnold Toynbee in the 1930s, as developed in Toynbee ([1954] 1987) and Toynbee ([1934] 1987), despite their methodologies being significantly different.

Historical economics may concern themselves with ancient societies, because, in the historical approach, one may study very particular aspects of economic relationships, such as the study of the setting of prices in ancient Greece (Chankowski 2020). In cliodynamics, one may study a very broad evolution, addressing multiple aspects of human society, ranging from the size of armies to territory distribution, for which there is some cross-sectional data, such as in the case of the Roman Empire (Roman & Palmer 2019). Cliometrics, however, has typically not covered these grounds because there is not enough data pertaining to the economy which may be studied in a systematic and serial manner. The new perspectives on data afforded by cliometrics and complexity, nevertheless, may constitute an opportunity to expand cliometrics towards ancient history.

2.2 Epistemological Distinctions Between History and Economics

At the core, the distinction that can be drawn between cliometrics and historical economics stems back from the one between history and economics, broader and more fundamental. The space between history and economics, in which cliometrics and historical economics reside, is indeed not a continuum. Fundamental epistemic differences prevent a smooth transition from one to the other. One may consider that the fact that economics are highly formal may deter historians, untrained in mathematics, from studying, and relying on, economics. Since the middle of the 20th century, economics have indeed developed to become more mathematical (Boulding 1948; Weintraub 2002). While this could be a contributing element, we also consider other, more essential causes, such as the use of stylized facts as opposed to actual raw data and the tendency to draw general conclusions as opposed to specific ones.

2.2.1 Stylized Facts and Primary Sources

Economics frequently use stylized facts: patterns, summaries of reality, that presumably capture the components of that reality deemed valuable for modeling, in order to test ideas or draw inspiration. For Hirschman (2016), stylized facts are indeed “empirical regularities in search of theoretical, causal explanations”, and they are useful heuristic tools in the social sciences in general, although in some cases their relevance has been opened for debate, such as in political sciences (Narang & LeVeck 2020). In their

methodological study of stylized facts, Arroyo Abad & Khalifa (2015) distinguished them from what they defined as base facts. Both base and stylized facts claim to describe phenomena, and both are expected to be explained by an economic theory. However, while base facts need to be “validly inferred from reliable data”, stylized facts do not, and are the results of a degree of interpretation or summarizing of the facts. According to these definitions, it appears that neither stylized nor base facts actually consist of the raw underlying information, they are, by definition, interpreted. Economics, when applied in a historical context, typically only rely on stylized facts. For example, in the case of ancient Greek history, some of the research we mentioned in Chapter 1 has looked into political regime transitions across various *poleis*, using aggregated data from secondary sources, which effectively are stylized facts (Fleck & Hanssen 2006, 2013; McCannon 2012).

Historical economics, in contrast, consist of the writing of elements of history that pertain to the economy, essentially using primary sources. In their own view, as we have seen, historians produce knowledge of the human past, following a method that aspires to being scientific, based on historical material. The deceptively simple notion of a fact, from a historical perspective, is in and by itself questionable. “It is not something buried that one just uncovers, ready for use¹⁴.” Marrou clarifies this further: a document does not constitute history (Marrou [1954] 2016, 45). Therefore, the foundation of any rigorous theory of history must be the manner in which documents are described, followed by a method for connecting that description to the justification that establishes it as historical fact. Carroll (2017) vividly illustrated the discrepancy between the data on violence in the medieval and modern periods and its actual meaning; if one simply considers stylized facts devoid of detailed historical context or one tries to serialize observations that are inherently different, the results are largely counter-factual.

The historical method, first and foremost, relies on historical documents, and in particular on a corpus of such documents, not on data or facts, because historical facts actually need to be constructed. Historical theory sees a continuum between the archive,

¹⁴See Febvre ([1952] 1992), p. 115, “Les faits, pensez-vous qu’ils sont donnés à l’histoire comme des réalités substantielles, que le temps a enfoui plus ou moins profondément, et qu’il s’agit simplement de déterrer, de nettoyer, de présenter en belle lumière à vos contemporains ?” Aron is also very clear: “il n’existe pas une *réalité historique*, toute faite avant la science qu’il conviendrait simplement de reproduire avec fidélité”, Aron ([1938] 1991), p. 147.

the document, the source and the corpus, and all essentially recoup with the notion of “documents produced by the actors of the history under study¹⁵”. A simple definition of a *corpus* is a set of documents assembled with a specific purpose. Constituting a corpus to carry out historical analysis is part and parcel of historical work and has even been considered to sufficiently define the very fact of writing history (Febvre 1934, 149). History writing has come to rely on constituting corpora as the centerpiece of its methodology: “the centrality of the archive” (Boutier 2014, 10–11). Since the very act of seeing events served as the basis for history writing, there could not be a notion of corpus or document in ancient historiography, as Bermejo Barrera (2001) noted. The importance of relying on one’s personal experience as a witness to events suggests that history writing was mostly focused on recent history and that texts were regarded as second-class witnesses (Boutier 2014, 12–14). Historians did not take into consideration the value of a text corpus until the history of the Church was recognized as a field of study. The direct testimony of Christ’s contemporaries bore more weight than the personal witnesses of the ancient Greek tradition, because this body of text had a material nature, much like sacred relics (Bermejo Barrera 2001, 194). At the core of the 18th and 19th century crystallization of a historical technique, close reading of the ancient texts evolved into hermeneutics. In the view of the 18th-century philosopher Pierre Bayle, as Gunther Pflug points out, the document’s predominance initially prevented any kind of deductive operation: “The scholar’s goal consisted of surveying the factual data, penetrating the historical givens, without attempting to impose any order unless it were for mere purposes of clarity” (Pflug 1971). Voltaire and later Turgot pulled history away from straight facts towards scientifically inspired analysis, relying on reason and common sense, thereby making the notion of pure document-based facts less central. Still critical and essential, the corpus then functioned with the application of reason, and inserted itself in the context of the question asked by the historian (Pflug 1971, 9–12). As the writing of history became professionalized in the 19th century, the methodology of source critique converged towards current practice (Offenstadt 2011, 70).

At that juncture, historical knowledge acquired “a new configuration thanks to the introduction of two notions: that of document [. . .]; and that of the scientific method” (Bermejo Barrera 2001, 198). This perspective effectively established the document and

¹⁵“Les documents produits par les acteurs de l’histoire étudiée”, see Offenstadt (2011), p. 68.

the aspiration to a scientific approach as two facets of the same coin. Indeed, in a Foucauldian approach, analysis that is specifically historical as well as a more general form of analysis common in the social sciences at large both stem from the same source: “the questioning of the *document*¹⁶”. Bermejo Barrera stresses the primordial place that the corpus holds in current historiography: “History builds its object starting from the constitution of its documentary corpora; it then develops different methods of reading and interpreting the texts, methods that are sometimes contradictory and that are not reducible to a common factor” (Bermejo Barrera 2001, 204). This view largely recoups with Foucault’s perception of historiography’s position with respect to the document: it seeks not to interpret it, but to work it from the inside and elaborate it¹⁷; hence the document should not to be seen as inert material. Foucault defines the writing of history as the manner in which a mass of documents is organized¹⁸. In this perspective, the actual historical work carried out in the context of economic history stops once the historical documents have been contextualized and organized, which is the step at which economic “facts”, and cliometrics, begin.

Hence, by not operating directly on primary sources, cliometrics are somewhat disconnected from history. Resorting to stylized information, or facts in the economic sense, eliminates the massive amount of meta-information that comes with establishing any historical “fact”. As a result, from the viewpoint of the historian, the findings are not historically valid since they may be seen as a restatement of some historians’ prior claims. Some cliometricians have come to realize that cliometrics had drifted away from the history-grounded field of historical economics, and in one such critic, as we pointed out earlier, Fenoaltea (2019) regretted the increasing distance between cliometricians and the sources. In their *riposte* to Fenoaltea, Diebolt & Hauptert (2020) argued, among other things, that modern cliometrics had helped further both economics and history by providing carefully grafted datasets, thanks to the importance of data for the cliometricians’ uses, and thanks to their advanced econometric techniques. From a historian perspective, we can see that the issue is not so much in the range of data

¹⁶“La mise en question du *document*”, see “Foucault (1969), p. 13.

¹⁷“[L’histoire] a changé sa position à l’égard du document : elle se donne pour tâche première, non point de l’interpréter, non point de déterminer s’il dit vrai et quelle est sa valeur expressive, mais de le travailler de l’intérieur et de l’élaborer [...]”, see Foucault (1969), p. 14.

¹⁸“L’histoire, c’est une certaine manière pour une société de donner statut et élaboration à une masse documentaire dont elle ne se sépare pas.” Foucault (1969), p. 14.

that may have been gathered by economists, but in whether this was done according to historical method. Although the association of cliometrics with complexity may appear as a simple praxeological distinction, it has therefore deep epistemological consequences. Using complex systems approaches makes it possible to directly operate on very raw data, and hence typically on primary sources, that do not have to be pre-processed by historians or historical economists. In that sense, cliometrics and complexity approaches can exploit the data's full granularity, which makes them by nature more historical. Epistemologically, cliometrics & complexity could hence be placed between history and traditional cliometrics.

2.2.2 Nomothetic Perspectives

While the relationship to primary sources contributes to distinguishing economics from history, and in turn cliometrics from historical economics, it is not the only such distinction. In fact, cliometrics search for universal rules, either deductively, beginning with general economic theory, or more inductively, searching for the formal model that best fits the data. They look for more or less universal set of principles, or try to verify them. This search for a set of rules is carried out through the use of formal modeling, which is based on data organized in a way that illustrates the connections under research. Hence, a fundamental distinction between history and cliometrics, which can make their approaches unintelligible to each other, is one between nomothetic and idiographic visions of history. This distinction does rely on the nature of the relationships with data: the search for universality in economics requires the use of manageable data, and cliometrics must operate on entities that have been pre-defined in some way, and in relation to a theory, and operates on base or stylized facts. Both base or stylized facts in economics are only derived from underlying data, and they are both conceived of within the perimeter of a theory. What economists may call base or stylized facts may in fact be called "general laws" by historians, and invoke the concept of a Weberian *Idealtypus*. These stylized facts are not mechanically derived from raw data or information, they necessarily are the result of some analysis and interpretation, when they are immersed in a theoretical apparatus. Economic "facts", in the sense of Arroyo Abad & Khalifa (2015), when they pertain to history therefore are best viewed as the output from some historical work. Hence, traditional cliometrics typically use these

base or stylized facts, which have been theoretically oriented.

However, one should build structures, rather than hypothesize them, in history (Marrou [1954] 2016, 166), a perspective which mostly rejects any “grand explanation” as well as the use of generic notions. Indeed, it cannot be assumed that a broad model of the world would provide a satisfactory explanation for all aspects of reality. Since history is defined as the study of distinctive and singular phenomena, Marrou believes that it is useless to develop historical “laws”. Parallels and analogies do not make up these rules; they are simply examples that draw on a few common characteristics. The purpose of social sciences, and especially of economics, is to reduce a complex situation down to a simple model. According to Ober, an historian of ancient Greece: “the social scientist’s goal of theory testing, aimed at some more general understanding of human behavior, may be strictly irrelevant for the historian who remains focused intensively on the Greek past” (Canevaro et al. 2018, 6). Modern economic thought is used by economists as a framework for analysis, which goes against the historiographic culture of the majority of historians. It is indeed erroneous to try to generalize behavior models based on the specific experience that we are familiar with, according to Bourdieu’s anthropological view of economics, in particular his viewpoint on the theories of rational action¹⁹. Therefore, the pursuit of universal rules is irrelevant to history. Febvre’s critique of Toynbee’s book is also quite revealing. Febvre asks²⁰, from the perspective of an historian, “why bother?” Indeed, whether one analyzes meta-history with formal models or with the perspective of classical sociology, it is, for most historians, non historical.

Cliometrics tend to restrict themselves to the modern and contemporary periods, because economic data for earlier period is quite sparse; there are few series of base facts. For example, Hobson (2014) discussed the debate between neo-primitivists, who saw the Roman empire as a “developing country,” and modernists, who see the Roman empire as having benefited from institutions that provide incentives for economic success.

¹⁹In Bourdieu (2017), p. 16: “En partant du cas particulier de l’action économique dans des sociétés d’un type particulier comme les nôtres, et plus précisément de régions particulières de ces sociétés particulières, ils commettent, me semble-t-il, l’une des erreurs les plus funestes en sciences sociales : celle qui consiste à universaliser sans le savoir le cas particulier, c’est-à-dire à donner pour universelles des particularités d’un cas particulier qui s’ignore comme tel.”

²⁰See Febvre ([1952] 1992), p. 134: “Si on résiste à la séduction du magicien ; si on se refuse à l’attitude sentimentale du croyant assistant au culte ; si on examine les idées froidement, et les conclusions : quoi de neuf, en tout ceci ; quoi de vraiment neuf et qui puisse, historiens, nous inciter à un retour sur nous-mêmes, à une condamnation de nos méthodes, à l’adoption de méthodes neuves ?”

He emphasizes that this conflict has polarized the discussion on the ancient Roman economy and draws attention away from the reality that it is unrealistic to assume that we could accurately know and evaluate every aspect of the Roman economy. The scarcity of data on the subject is such that there is an impossibility of knowledge. This focus on the modern and contemporary periods is also a consequence from the fact that cliometrics, as we stressed earlier, deal with traditional economic questions: this is cliometrics' natural focus. In economics, most of the theoretical apparatus deals with issues pertaining to the economy. Hence, coming from this body of knowledge, when one considers the application of economics to history, this application must concern itself with the study of the economy, which mechanically restricts the range of periods and cultures under study.

The epistemic differences between economics and history are substantial and fundamental enough that it is difficult to imagine research that could truly further both fields at the same time. The discontinuity, in particular in relation to the nomothetic aspect of economic modeling, cannot be simply bridged. Nevertheless, it can provide cliometrics with a new angle, and room for expansion, if we consider this epistemic distinction as a direction towards which to progress, in particular starting from cliometrics & complexity.

2.3 New Cliometrics With Complexity

We propose to extend the definition of cliometrics, by making them broader in terms of their relationship to economics. We will first discuss how cliometrics can expand the range of questions they address by not restricting themselves to purely economic questions. Doing so widens the gamut of historical data that may be studied, and, in particular, allows cliometrics to directly drill into primary sources, connecting the field with complexity sciences. In that sense, the extension of cliometrics to primary sources and to a broader notion of economic questions, by nature, corresponds to cliometrics & complexity. Then, we look into how the epistemological gap between history and economics may be bridged, in some cases, by being able to focus on the same historical material as historians. Next, since primary sources are available in large quantities in the form of electronic data, we examine how corpora of historical material are constructed by

historians. We point out how, through a more systematic and programmatic approach, and through the perception of corpora as processes rather than data, cliometrics & complexity can make historiography more scientific. Finally, we discuss an example of these new cliometrics, in an application to the study of ancient Greek religious practice.

2.3.1 Cliometrics & Complexity Can Offer New Perspectives on Historical Data

The association of cliometrics with the study of the economy is strong, to the point that even those who seek to criticize the field take it as a given and do not question it. When he raised some strong criticism against cliometrics, Boldizzoni (2011) still remained within a paradigm of cliometrics strictly applied to the economy. For Boldizzoni, the use of microeconomics could broaden the applications of cliometrics, but only as (Boldizzoni 2011, 87): “an investigation of the past from a micro-economic point of view, with an analysis of decisions taken with regard to production, consumption, and exchange at the level of producers and households, and naturally of the consequences of these decisions.” A collection of essays in Rawski et al. (1996), in order to “broaden and deepen the exchange of ideas between economists and historians”, proposed to “show how to apply the core ideas and methods of economics to a wide range of historical issues.” However, the subjects addressed from an economic perspective in this book, ranging from economic trends to international economics, would be fairly irrelevant and in-exploitable for anything else than European societies after the Middle Ages.

This focus on the economy is not, however, a proper reflection of the range of domains in which economic models, and in particular microeconomic models, have brought new theories, paradigms and methods in the social sciences. After Maynard Smith (1972) applied game theory to biology through evolutionary stable strategies, he revolutionized the study of evolution, not only in biology but also in numerous evolutionary aspects of social sciences (Boyd & Richerson 2005). This approach, indeed, was further applied to behavioral ecology, evolutionary psychology, and to the notion of cooperation (Axelrod & Hamilton 1981; Axelrod 2006). Within social sciences, economics-grounded approaches have reached across the board. For instance, issues related to religion have been explored with a microeconomic angle, both by economists (Iannaccone 1998), and by specialists of religious studies, for example developing costly signal models for rituals (Henrich 2009; Bulbulia & Frean 2010). A model accounting for

the advantages of group identification through symbols (Carr & Landa 1983) and the definition of a generalized utility function accounting for identity-related preferences (Akerlof & Kranton 2000) are two examples of how the concept of identity has been investigated from an economic perspective. Game theory has also been applied to the study of literature: Brams (1994) offered a review on the subject, and de Ley (1988) developed some examples of applications to contemporary French literature. As evidenced by Posner ([1973] 2014), the economic analysis of law, first used by Coase (1960), has also shed a new light on the understanding of the legal environment and its evolution. Microeconomic tools have also been used to investigate argument structure and rhetorics (Beigman Klebanov & Beigman 2010; Lipman & Seppi 1995; Matthews 1989). All these examples of fields in which economics have been applied may benefit from a historical perspective, which would devolve to cliometrics.

Cliometrics, just like economics in general, do not have to be about the economy. They can use economics to address numerous historical questions that are not, *per se*, economic, but on which one can shed light through the application of microeconomic models. In consequence, when analyzing historical documents, one may use these approaches in order to enhance one's understanding of these documents. This idea's fundamental tenet is that, by design, microeconomic methods should be applied to historical materials that are as close as possible to original sources, not to the end product of the historians' interpretations. By not being tied to specific models about the economy, the range of data that may be analyzed in cliometrics becomes more varied, and may pertain to all aspects of human life. As a result, the same documents that historians work on may be used, albeit with a different angle. Such historical material may be both particular and general at the same time. Each individual atomic component of historical data can be used independently of the rest, but it can also be used in conjunction with other related components, which is, in essence, the logic underlying the analysis of complex systems. Through the features of the distribution of these components, the meta-analysis of the set of elements as a whole contains information beyond the mere sum of the information belonging to each element. Cliometrics & complexity can be used to identify the generative processes that may have produced the distributions one can observe in the data. Considering generating mechanisms fundamentally connects cliometrics with complex systems analysis, which fundamentally focus on explaining

the emergence of patterns (Boccaro 2010). Cliometrics, relying on complexity modeling in conjunction with microeconomic models, can therefore help determine, following a constructive view, according to what process the data has been generated. Taking this process into consideration allows for a much improved understanding both from an economic modeling perspective and from an historian perspective.

2.3.2 Extending Cliometrics & Complexity Towards History

While cliometrics do not necessarily have to be fungible with history, the extension of the field to non-economic questions and to the use of primary sources afforded by a complex systems perspective can, in certain conditions, constitute an epistemological bridge. In effect, the use of atemporal economics-based approaches on historical phenomena is compatible with establishing structures, rather than postulating them, provided that the economic approaches in question are constructed from historical data rather than being postulated as an application of a particular economic theory. This corresponds to seeing microeconomics from a historian's perspective as an auxiliary instrument rather than as an application of particular economic theories. Since the historian must in any case make a fundamental assumption of rationality in order to understand the actors, the economic approach does not need to make an additional assumption. The notion of an optimality of behavior, which is what rationality presumably should pursue, does not strictly require that the actors be self-aware. Indeed, Foley (1985) showed that optimality is a useful transverse concept in order to describe the behavior of early humans. This is also an important aspect of Bourdieu's argument against *homo oeconomicus*²¹: accounting for behavior perceived as rational does not require the assumption of conscious reasoning. Maynard Smith (1982) too provided numerous examples of optimal behavior not as the result of an explicit or conscious calculation, but attained by some evolutionary process.

Trying to perceive the rationality of the actors in a historical situation is, in fact, precisely how Aron (1981) defined "understanding", as opposed to "explaining". Understanding the optimality of some people's or groups' behavior in the past based on some data, a behavior that may be relative to any aspect of life, using a microeconomic model, would hence not only pertain to cliometrics, but also to Aron's notion of

²¹See Bourdieu (2017), p. 15: "je défendrai une anthropologie tout à fait différente, fondée sur l'idée que, pour rendre compte des conduites perçues comme rationnelles, il n'est pas besoin de faire l'hypothèse qu'elles ont la raison, ou l'intention consciente de rationalité, pour principe."

understanding in historiography. It is conceivable to use a model that explains the behavior of these actors and from which one might infer links between these observed “facts” whenever circumstances involving rational agents would result in the generation of historical “facts”. Cliometrics & complexity may provide historians with a more in-depth understanding of the actors as well as the historical sources on which they base their study with this approach. According to Parent (2004), the focus of economic theory, and microeconomics in particular, is often not on how things may have developed but rather on how they are, in some equilibrium. These models do not require that causes and effects be seen from a temporal dynamic perspective.

We can hence expand the territory of cliometrics to the application of economics, and in particular microeconomics, to all facets of human life that involve rationality and optimal behavior, directly through the use of historical documents, thanks to the use of complexity sciences. The models that would be applied to this larger territory, to the extent that they are close to the data and only rely on minimal rationality arguments are, at least in principle, epistemologically compatible with the field of history.

2.3.3 A New Notion of Corpus in Cliometrics & Complexity

Historians work on historical documents, arranged in corpora, often in electronic form, off of which they produce their historical analyses. Cliometricians who would want to analyze a particular set of historical documents as a complex system would, presumably, operate in the same way. If the economic models used by the cliometricians are to be applied to some data, that data, most likely, is in electronic form. Digital humanities have penetrated every level of historians’ work, and thanks to many large-scale efforts in the digitization of historical material, a wealth of historical documents is now available electronically. These digitization efforts do not systematically require the formal modeling of complex links between the elements that compose the data; sometimes, they mainly consists of ensuring the quality of the resulting electronic text or data, and in the storage of all relevant metadata pertaining to the original documents. In that sense, the electronic material is often raw. In ancient history, and for ancient Greece in particular, essentially all literary texts, a majority of inscriptions and artifacts, such as vases, and numismatic finds have been digitized (Barker & Terras 2016; Crane

2012; Pantelia 2020). Cliometrics should, however, consider the processing of these historical documents in a different fashion than historians do.

For Philippe Rygiel, historians are “hypertextual polygraphs, who dissimulate most of the inscriptions they produce²²”, stressing the fact that the majority of the historian’s work is not visible from the results or the analyses they publish. In Rygiel’s view, the historian’s annotations, essentially in textual form, constitute the core of their work, their production. In this perspective, the historical inquiry becomes the definition of a corpus, augmented with these annotations (Rygiel 2011, 34). The fact that the underlying data may exist in electronic form does not change the situation: in history, the creation of a digital corpus appears to be a complex exercise for which there is no clear and unique epistemological framework or methodology (Gibbs & Owens 2013; Hoekstra & Koolen 2019).

It is nevertheless fundamentally beneficial for the design of a corpus and the processing of its information to be automated, and leverage its electronic nature when possible. One can indeed make an argument linking the reproducibility of research, gained by digital processing, to its scientificity. For McGillivray, Wilson, & Blanke (2019), collecting and processing historical material with computational methods “would be a science if we could learn to automate it”. In their view, historians should clearly delineate between what they define as “evidence” and what they define as “claims”, so that one may separate evidence-based findings from other statements, thanks to the systematic analysis of evidence. There is a benefit being able to precisely understand the underpinnings of any statement, when they can be made explicit. Making processes automatic, hence, may not, *per se*, serve a fundamental purpose, but it would have the advantage of making the creation of a corpus and some of its processing reproducible, and open it to critique. Reproducibility makes any statement falsifiable, which is a fundamental feature of any scientific statement according to the Popperian logical framework (Popper [1934] 1992, 78 sq). The way in which the electronic nature of the underlying data is leveraged by historians through their use of textual research is however not made explicit. Cliometricians, when they rely on complex systems analyses

²²“L’historien contemporain apparaît alors d’abord comme un polygraphe hypertextuel dissimulant aux regards l’essentiel des inscriptions qu’il produit”, see Rygiel (2011), p. 32.

on primary sources, by being more focused on the processing of the data, can enhance the scientificity of historical inquiry.

When historians work on historical sources that are available electronically, they effectively carry out combinations of data and operations; they create their own corpus off of which they write their analysis. In practice, historians often follow such a pipeline: for example, obtain some data (in a spreadsheet), transform it, save the clean version, compute some aggregates. The resulting corpus, in spite of it being the result of a process, is considered as a *thing*, as data, focusing on the result of a series of operations rather than on the operations themselves. Hence, only the end result survives, and the details of all the steps are lost to everyone else. If this entire process is coded as a pipeline or as an algorithm more generally, every single assumption, explicit or not, becomes visible. One should therefore realize that a corpus, used for historical analysis, or for cliometric modeling, is not data, but rather a set of operations. One can consider the notion of a *pipeline*, in the context of modern data analytics, or in computational linguistics, comparable to the data analytics suites used in the hard sciences (Jockers & Thalken 2020; Wickham & Grolemund 2017, 261–268). In such a pipeline, the raw input at each stage is transformed in place, and serves as the input for the next stage. This code can be analyzed, and it can be run in whole or in part by anyone. Purely seen as a dataset, a corpus cannot be properly analyzed from an external perspective, but once its construction is made entirely explicit, then this process can be fully subject to critique.

Hence, once defined as computer code, or as a pipeline, the corpus contains and makes explicit all the decisions, small and large, made in cleaning, filtering, completing, or arranging the raw information, in the most concise manner possible: the corpus's definition is the process, and reciprocally. Modern data analysis and text edition tools have largely converged, so that there is not such a strong distinction between the two anymore: the text a researcher produces and the computer code that gathers and processes data exist in the same document. Considering the data organization, processing, analysis and write-up of a historical corpus as a continuum in a seamless process naturally leads to reproducible research²³. Given the nature of cliometric work,

²³The RMarkdown language, for example, combines the data and statistical modeling infrastructure of the R language with the editorial capabilities of the Markdown syntax and LaTeX system; it has been suggested as a good framework for reproducible research, see Calero Valdez (2020).

most likely requiring data manipulations, the joining of various sources, and fitting some models, treating the entire process as a pipeline presumably comes more naturally to cliometricians than to historians. The fact that complexity approaches allow for the use of raw primary sources hence permits cliometrics to fully and systematically describe, through algorithms, the details of their analyses from the ground up. This is not possible when operating on base or stylized facts which are already remote from the sources.

2.3.4 Extending the Range of Cliometrics & Complexity to Ancient History

As we have pointed out earlier, economic data for ancient history is quite sparse, but there is a certain amount of digitized data, a lot of it textual, that is available. In order to illustrate how the expanded view of cliometrics we have describe above can operate, we discuss an example pertaining to the study of ancient Greece, in particular votive acts, which will be further developed in Chapter 3, and fully analyzed in Chapter 6.

In ancient Greece historiography, the study of divine onomastic sequences, inscriptions mentioning the gods to whom people would appeal, started at the end of the 19th century, and is an active field of research (Brulé & Lebreton 2007; Lebreton et al. 2014; Bonnet & Lebreton 2019). The raw information, listing all such sequences that have been found, is usually accessed by historians at an atomic level; the philological analysis of a handful of inscriptions can provide historical insight on the perceived relations between divine entities (Bonnet 2021). The BDEG database (Lebreton et al. 2014), containing over 11,000 records of votive acts, is in fact not designed for any access other than on a one-by-one basis. Each entry represents a group of attested observations of invocations of the gods in Greek language. Creating a dataset that can lend itself to cliometrics & complexity analysis, in this case, can be done by automatically extracting all the relevant data from the BDEG's website, joining it with some other sources of information on ancient Greek *poleis*, and tabulating the occurrences of god names as needed. Going from the historical sources to the analytic results, in this case, can be clearly seen as a process, expressed as a replicable algorithm, such that if any of the inputs changed, all the resulting analyses would be updated accordingly.

A cliometrics & complexity-based perspective on epigraphic observations of votive acts can rely on a microeconomic model, using game theory, addressing individual

choices to worship one god rather than another at a point in time, if the benefit from these acts are considered as a shared resource. Indeed, in this case, it is a form of “Kolkata Restaurant Problem” or “El Farrol Bar Problem” (B. K. Chakrabarti 2007). Simple randomized strategies are optimal, based on the gods’ perceived strength. Using the formal model, one can derive an expected distribution shape for the votive acts across gods, which expresses how, under minimal assumptions on atomic choices, the numbers of observations for each god stack relative to each other. The empirical form of the distribution, at the *polis* level, can be obtained from the raw data and compared with the model’s predictions. It is also possible to try and relate certain characteristics of each *polis* to its local votive distribution. In this instance, the distributional information emerging from the raw data, which historians do not generally take into account, has the potential to become fact similarly to other historical material, when seen through the lens of microeconomics. The new data might then be questioned: for instance, are there any times or places where the distributions significantly deviate from the baseline expectation? Are the systematic differences caused by factors at the *polis*-level? The data, in fact, shows surprisingly regular patterns for a large range of *poleis*, although the most popular god names vary significantly. Using distribution fitting techniques from complexity sciences, we can show that local votive acts follow power laws, whose generating mechanisms can be modeled in different ways.

2.4 Conclusion

Cliometric methodology and ancient history, do not, at first sight, appear to have much in common. We have argued that, in fact, they should, through complex systems analysis. The study and close reading of ancient sources is a prerogative of historians, but the digital tools that have been developed for these purposes are often underutilized from a data analytics perspective, and the new approaches developed in the field of cliometrics & complexity are particularly well adapted to these data. Cliodynamics and quantitative history, to some extent, exploit this data but without a theoretical framework that seeks to explain the underlying phenomena. If one considers that cliometrics do not have to be about the economy *stricto sensu*, then large swaths of primary sources, in particular in ancient history, can be analyzed in the light of microeconomic models.

Progress in the digital humanities has brought a large quantity of historical documents to an electronic form. This extends the notion of cliometric analysis in two ways: first, cliometrics could concern themselves with any kind of human behavior, as far as there are grounds to believe there is some optimality in it, with a kind of “cliofreakonomics”; and second, by not having to be specific to the economy, the data can be directly drawn from primary sources, thanks to complexity sciences. The lesser the distance to the sources, the lesser the epistemological distance to historiography, which makes cliometrics more intelligible to historians. A cliometrics & complexity approach looks at historical material in a way that is fundamentally different from historians, potentially bringing new light on a range of historical issues. This focus on data processing may also significantly improve the manner in which historical document corpora are processed, by making systematic and explicit many aspects of data analysis that are usually hidden by historians, hence reinforcing the scientific validity of statements. New cliometrics & complexity may in fact be the answer to Lucien Febvre’s definition of history as an endeavor aspiring to scientificity.

2.5 Chapter Bibliography

- Abry, Patrice, Cécile Bastidon, Pierre Borgnat, Pablo Jensen, & Antoine Parent.** 2022. "The Contours of a Cliophysics: How Can Econophysics Enrich Cliometrics? Case Studies in Debt Issues and Global Capital Markets." *Frontiers in Physics* 10: 894400.
- Akerlof, George A., & Rachel E. Kranton.** 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–753.
- Alexandrova, Anna.** 2009. "When Analytic Narratives Explain." *Journal of the Philosophy of History* 3 (1): 1–24.
- Aron, Raymond.** 1981. "Quelques remarques sur la compréhension et l'explication." *Revue européenne des sciences sociales* 19 (54-55): 71–82.
- . (1938) 1991. *Introduction à la philosophie de l'histoire : essai sur les limites de l'objectivité historique*. Tel 58. Paris: Gallimard.
- Arroyo Abad, Leticia, & Kareem Khalifa.** 2015. "What Are Stylized Facts?" *Journal of Economic Methodology* 22 (2): 143–156.
- Arthur, W. Brian.** 2021. "Foundations of Complexity Economics." *Nature Reviews Physics* 3 (2): 136–145.
- Axelrod, Robert M.** 2006. *The Evolution of Cooperation*. 2nd ed. New York: Basic Books.
- Axelrod, Robert M., & William D. Hamilton.** 1981. "The Evolution of Cooperation." *Science, New Series*, 211 (4489): 1390–1396.
- Bastidon, Cécile, Michael Bordo, Antoine Parent, & Marc Weidenmier.** 2019. "Towards an Unstable Hook: The Evolution of Stock Market Integration Since 1913." Working Paper w26166. Cambridge: National Bureau of Economic Research. <https://doi.org/10.3386/w26166>.
- Bastidon, Cécile, & Antoine Parent.** 2022. "Cliometrics of World Stock Markets Evolving Networks." *Annals of Operations Research* To appear.
- Bastidon, Cécile, Antoine Parent, Pablo Jensen, Patrice Abry, & Pierre Borgnat.** 2020. "Graph-Based Era Segmentation of International Financial Integration." *Physica A: Statistical Mechanics and Its Applications* 539: 122877.
- Beigman Klebanov, Beata, & Eyal Beigman.** 2010. "A Game-Theoretic Model of Metaphorical Bargaining." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 698–709. Uppsala: Association for Computational Linguistics.
- Bernheim, B. Douglas.** 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–877.
- Boccaro, Nino.** 2010. *Modeling Complex Systems*. Graduate Texts in Physics. New York: Springer.
- Boldizzoni, Francesco.** 2011. *The Poverty of Clío: Resurrecting Economic History*. Princeton: Princeton University Press.
- Bonnet, Corinne, ed.** 2021. *Noms de Dieux : Portraits de Divinités Antiques*. Essais - Histoire. Toulouse: Anacharsis.
- Bonnet, Corinne, & Sylvain Lebreton.** 2019. "Mettre les polythéismes en formules ? À propos de la base de données Mapping Ancient Polytheisms." *Kernos* 32: 267–296.
- Boulding, Kenneth E.** 1948. "Samuelson's Foundations: The Role of Mathematics in Economics." *Journal of Political Economy* 56 (3): 187–199.

- Bourdieu, Pierre.** 2017. *Anthropologie économique : cours au Collège de France, 1992-1993*. Raisons d’agir. Paris: Seuil.
- Boyd, Robert, & Peter J. Richerson.** 2005. *The Origin and Evolution of Cultures*. Evolution and Cognition. Oxford: Oxford University Press.
- Brams, Steven J.** 1994. “Game Theory and Literature.” *Games and Economic Behavior* 6 (1): 32–54.
- Brownlow, Graham A.** 2010. “Structure and Change: Douglass North’s Economics.” *Journal of Economic Methodology* 17 (3): 301–316.
- Brulé, Pierre, & Sylvain Lebreton.** 2007. “La Banque de données sur les épicleses divines (BDDE) du Crescam : sa philosophie.” *Kernos* 20: 217–228.
- Bulbulia, Joseph, & Marcus Frean.** 2010. “The Evolution of Charismatic Cultures.” *Method & Theory in the Study of Religion* 22 (4): 254–271.
- Canevaro, Mirko, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, eds.** 2018. *Ancient Greek History and Contemporary Social Science*. Edinburgh Leventis Studies 9. Edinburgh: Edinburgh University Press.
- Carr, Jack L., & Janet T. Landa.** 1983. “The Economics of Symbols, Clan Names, and Religion.” *The Journal of Legal Studies* 12 (1): 135–156.
- Carroll, Stuart.** 2017. “Thinking with Violence.” *History and Theory* 56 (4): 23–43.
- Cesarano, Filippo.** 2006. “Economic History and Economic Theory.” *Journal of Economic Methodology* 13 (4): 447–467.
- Chakrabarti, Bikas K.** 2007. “Kolkata Restaurant Problem as a Generalised El Farol Bar Problem.” In *Econophysics of Markets and Business Networks: Proceedings of the Econophys-Kolkata III*, edited by Arnab Chatterjee & Bikas K. Chakrabarti, 239–246. New Economic Windows. Milano: Springer.
- Chankowski, Véronique.** 2020. “Juste prix, prix libres et prix fixés dans les cités du monde grec antique : histoire d’un malentendu.” In *Les infortunes du juste prix : marchés, justice sociale et bien commun de l’Antiquité à nos jours*, edited by Véronique Chankowski, Clément Lenoble, & Jérôme Maucourant, 25–68. Documents. Lormont: Le Bord de l’eau.
- Choi, Jay P.** 2011. “Up or Down? A Male Economist’s Manifesto on the Toilet Seat Etiquette.” *Economic Inquiry* 49 (1): 303–309.
- Cline, Diane Harris.** 2020. “Athens as a Small World.” *Journal of Historical Network Research* 4: 36–56.
- Coase, Ronald H.** 1960. “The Problem of Social Cost.” *The Journal of Law and Economics* 3: 1–44.
- Diebolt, Claude.** 2011. “The Stakes of Cliometrics in Ancient History.” *Historical Social Research / Historische Sozialforschung* 36 (3): 350–361.
- Diebolt, Claude, & Michael Hauptert, eds.** 2019. *Handbook of Cliometrics*. 2nd ed. Springer Reference. Cham: Springer International Publishing.
- . 2020. “How Cliometrics Has Infiltrated Economics – and Helped to Improve the Discipline.” *Annals of the Fondazione Luigi Einaudi* 54 (1): 219–230.
- Diebolt, Claude, & Antoine Parent.** 2011. *Essais cliométriques : capital humain, monnaie et finance, croissance et cycles*. Bern: Lang.
- Dixit, Avinash.** 2012. “An Option Value Problem from *Seinfeld*.” *Economic Inquiry* 50 (2): 563–565.

- Durlauf, Steven, Antoine Parent, Cécile Bastidon, Patrice Abry, Pierre Borgnat, & Pablo Jensen.** 2023. *Complexity and Cliometrics*. Edited by Steven Durlauf & Antoine Parent. Vol. forthcoming. Elements in Complexity and Agent-Based Economics 1. Cambridge: Cambridge University Press.
- Febvre, Lucien.** (1952) 1992. *Combats pour l'histoire*. L'ancien et le nouveau 12. Paris: Armand Colin.
- Fenoaltea, Stefano.** 2019. "Spleen: The Failures of the Cliometric School." *Annals of the Fondazione Luigi Einaudi* 53 (2): 5–24.
- Fine, Ben, & Dimitris Milonakis.** 2009. *From Economics Imperialism to Freakonomics: The Shifting Boundaries Between Economics and Other Social Sciences*. London: Routledge.
- Fleck, Robert K., & F. Andrew Hanssen.** 2006. "The Origins of Democracy: A Model with Application to Ancient Greece." *The Journal of Law and Economics* 49 (1): 115–146.
- . 2013. "How Tyranny Paved the Way to Democracy: The Democratic Transition in Ancient Greece." *The Journal of Law and Economics* 56 (2): 389–416.
- Foley, Robert.** 1985. "Optimality Theory in Anthropology." *Man* 20 (2): 222–242.
- Gonzales-Feliu, Jesus, & Antoine Parent.** 2016. "Clio-Combinatorics: A Novel Framework to Analyze Military Logistics Choices Using Operations Research Techniques." In *Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations*, edited by Alberto Ochoa-Zezzatti, Jöns Sánchez, Miguel Gastón Cedillo-Campos, & Margain de Lourdes, 79–101. Advances in Logistics, Operations, and Management Science. Hershey: IGI Global.
- Hansen, Mogens Herman, & Thomas Heine Nielsen.** 2004. *An Inventory of Archaic and Classical Poleis*. Oxford: Oxford University Press.
- Hauptert, Michael.** 2019. "History of Cliometrics." In *Handbook of Cliometrics*, edited by Claude Diebolt & Michael Hauptert, 3–32. Cham: Springer International Publishing.
- Henrich, Joseph.** 2009. "The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution." *Evolution and Human Behavior* 30 (4): 244–260.
- Hirschman, Daniel.** 2016. "Stylized Facts in the Social Sciences." *Sociological Science* 3: 604–626.
- Hobson, Matthew S.** 2014. "A Historiography of the Study of the Roman Economy: Economic Growth, Development, and Neoliberalism." *Theoretical Roman Archaeology Journal* 2013: 11–26.
- Iannaccone, Laurence R.** 1998. "Introduction to the Economics of Religion." *Journal of Economic Literature* 36 (3): 1465–1495.
- Ionesco, Eugène.** 1954. *La cantatrice chauve : anti-pièce ; suivi de La leçon ; drame comique*. Collection folio 236. Paris: Gallimard.
- Jackson, Matthew O., & Brian W. Rogers.** 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review* 97 (3): 890–915.
- Karila-Cohen, Karine.** 2016. "Prosopographia Attica 2.0 : base de données et raisonnement prosopographique." *Revue historique* 680 (4): 869–904.
- . 2018. "Le graphe, la trace et les fragments : l'apport des méthodes quantitatives et des outils numériques à l'étude des élites civiques athéniennes." *Annales HSS* 73 (4): 785–815.

- Karila-Cohen, Karine, Claire Lemerrier, Isabelle Rosé, & Claire Zalc.** 2018. "Nouvelles cuisines de l'histoire quantitative." *Annales HSS* 73 (4): 773–783.
- Kenna, Ralph, Máirín MacCarron, & Pádraig MacCarron, eds.** 2017. *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. Understanding Complex Systems. Cham: Springer International Publishing.
- Krugman, Paul R., & Robin Wells.** 2015. *Economics*. 4th ed. New York: Worth Publishers.
- Kydros, Dimitrios, Panagiotis Notopoulos, & Georgios Exarchos.** 2015. "Homer's Iliad – A Social Network Analytic Approach." *International Journal of Humanities and Arts Computing* 9 (1): 115–132.
- Langlois, Charles-Victor, & Charles Seignobos.** (1898) 2014. *Introduction aux études historiques*. Bibliothèque idéale des sciences sociales. Lyon: ENS Éditions.
- Le Riche, Antoine, Antoine Parent, & Lei Zhang.** 2022. "Introduction." In *Institutional Change and China Capitalism: Frontier of Cliometrics and Its Application to China*, edited by Antoine Le Riche, Antoine Parent, & Lei Zhang, xxvii–xliv. Transformations in Banking, Finance and Regulation 2. Hackensack: World Scientific.
- Lebreton, Sylvain, Jean-Baptiste Barreau, Karine Karila-Cohen, & Pierre Brulé.** 2014. "Banque de Données des Épicleses Grecques (BDEG)." 2014. <https://epiclesesgrecques.univ-rennes1.fr/?lang=en>.
- Lemerrier, Claire.** 2012. "Formale Methoden der Netzwerkanalyse in den Geschichtswissenschaften: Warum und Wie?" *Österreichische Zeitschrift für Geschichtswissenschaften* 23 (1): 16–41.
- Levitt, Steven D., & Stephen J. Dubner.** 2009. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: Harper Perennial.
- Ley, Herbert de.** 1988. "The Name of the Game: Applying Game Theory in Literature." *SubStance* 17 (1): 33–46.
- Lipman, Barton L., & Duane J. Seppi.** 1995. "Robust Inference in Communication Games with Partial Provability." *Journal of Economic Theory* 66 (2): 370–405.
- Malkin, Irad.** 2011. *A Small Greek World: Networks in the Ancient Mediterranean*. Greeks Overseas. New York: Oxford University Press.
- Margo, Robert A.** 2018. "The Integration of Economic History into Economics." *Cliometrica* 12 (3): 377–406.
- Marrou, Henri-Irénée.** (1954) 2016. *De la connaissance historique*. Histoire H21. Paris: Seuil.
- Matthews, Steven A.** 1989. "Veto Threats: Rhetoric in a Bargaining Game." *The Quarterly Journal of Economics* 104 (2): 347.
- Maynard Smith, John.** 1972. *On Evolution*. Edinburgh: Edinburgh University Press.
- . 1982. *Evolution and the Theory of Games*. 1st ed. Cambridge University Press.
- McCannon, Bryan C.** 2012. "The Origin of Democracy in Athens." *Review of Law & Economics* 8 (2).
- Mongin, Philippe.** 2016. "What Are Analytic Narratives?" Working Paper ID 2796567. Rochester: Social Science Research Network. <https://papers.ssrn.com/abstract=2796567>.
- . 2018. "A Game-Theoretic Analysis of the Waterloo Campaign and Some Comments on the Analytic Narrative Project." *Cliometrica* 12 (3): 451–480.

-
- Narang, Neil, & Brad L. LeVeck.** 2020. "Stylized Facts and Comparative Statics in (Social) Science Inquiry." *International Studies Quarterly* 64 (2): 482–485.
- Ober, Josiah.** 2008. *Democracy and Knowledge: Innovation and Learning in Classical Athens*. Princeton: Princeton University Press.
- . 2015. *The Rise and Fall of Classical Greece*. Princeton: Princeton University Press.
- Parent, Antoine.** 2004. "L'analyse économique est-elle intemporelle ?" *Temporalités* 1: 98–106.
- Parker, Robert, Jean-Baptiste Yon, & Mark Depauw.** 1996. "Lexikon of Greek Personal Names." 1996. <http://www.lgpn.ox.ac.uk>.
- Posner, Richard A.** (1973) 2014. *Economic Analysis of Law*. 9th ed. Aspen Casebook Series. New York: Wolters Kluwer Law & Business.
- Rawski, Thomas G., Susan B. Carter, Jon S. Cohen, Stephen Cullenberg, Peter H. Lindert, Donald N. McCloskey, Hugh Rockoff, & Richard Sutch.** 1996. *Economics and the Historian*. Berkeley: University of California Press.
- Roman, Sabin, & Erika Palmer.** 2019. "The Growth and Decline of the Western Roman Empire: Quantifying the Dynamics of Army Size, Territory, and Coinage." *Cliodynamics* 10 (2): 1–23.
- Toynbee, Arnold J.** (1954) 1987. *A Study of History*. Vol. 2. 2 vols. New York: Oxford University Press.
- . (1934) 1987. *A Study of History*. Vol. 1. 2 vols. Oxford: Oxford University Press.
- Tridimas, George.** 2011. "A Political Economy Perspective of Direct Democracy in Ancient Athens." *Constitutional Political Economy* 22 (1): 58–82.
- Turchin, Peter.** (2003) 2018. *Historical Dynamics: Why States Rise and Fall*. Princeton Studies in Complexity. Princeton: Princeton University Press.
- Turchin, Peter, Thomas E. Currie, Edward A. L. Turner, & Sergey Gavrilets.** 2013. "War, Space, and the Evolution of Old World Complex Societies." *Proceedings of the National Academy of Sciences* 110 (41): 16384–16389.
- Vernant, Jean-Pierre, & Pierre Vidal-Naquet.** (1972) 2001. *Mythe et tragédie en Grèce ancienne*. Vol. 1. 2 vols. Paris: La Découverte.
- Weintraub, E. Roy.** 2002. *How Economics Became a Mathematical Science*. Science and Cultural Theory. Durham: Duke University Press.

Chapter 3

A Cliometrics and Complexity Perspective on Ancient Greek Culture

Advances in cliometrics have not offered much help to ancient history, because of the scarcity of economic data for the period, as we have discussed in Chapter 2. On the other hand, we have seen that the conjunction of cliometric perspectives with complexity sciences into a new intellectual endeavor, cliometrics & complexity, can significantly widen the reach of economic history in general, as Durlauf et al. (2023) suggest. Using complexity economics notions, whether they relate to networks, to time series, or to the emergence of patterns, has renewed the cliometric perspective on economic or financial questions in the contemporary period (Bastidon et al. 2019, 2020; Bastidon & Parent 2022; Abry et al. 2022). While the use of networks is not something new in contemporary history (Lemercier 2012), the cliometrics & complexity approach, by considering complex generating mechanisms based on the agents' behavior, offers deeper insights. Combining complexity economics, such as networks and non-linear econometrics, with institutional economics also offers new insights in existing cliometric issues (Le Riche, Parent, & Zhang 2022).

Staying in the path of cliometrics & complexity, we propose in this chapter an economics-inspired analysis of various facets of ancient Greek culture, serving as an illustration of using new methodological angles in the study of the past through economics. We consider that advanced econometric methods do not have to be restricted to the study of the economy, but instead that large amounts of data can always benefit from quantitative scrutiny and be put into perspective through economic models. We

apply econometric and economic analyses, inspired by complexity models, to several important datasets pertaining to ancient Greece, directly derived from the Ancients' literary or epigraphic production. We have examined the epistemic distinctions between economics and historiography in Chapter 2, and one of the salient conclusions was that if economics are to contribute results that can be intelligible to historians, and further their understanding of the past, it is important to focus on primary rather than secondary sources. Focusing on primary sources, which are large datasets, recoups well with cliometrics and complexity approaches, designed to analyze the emergence of patterns. Historians of ancient Greece usually consider primary sources at the atomic level. In ancient history, the philological analysis of a single inscription, or the occurrence of a particular name in a text, can often give rise to a detailed commentary, based on the historian's knowledge of the context in which the historical documents were produced. Economic analysis of a particular situation, typically using game theory, may too shed light on such historical analysis: this is the focus of analytic narratives, as we saw in Chapter 1, such as Mongin (2018), looking at strategy at the battle of Waterloo, or Crettez & Deloche (2018), considering Caesar's decision to step into the Senate where he was killed. However, cliometrics can offer a new perspective by focusing on primary sources through the lens of complexity economics, based on fundamentally different hermeneutics. Given the categorical nature of the data, one can indeed examine it in a distributional perspective, and ask what generating mechanism may have produced it, relying on complexity theory. The analysis of generating mechanisms, and the identification of the conditions for their optimality in terms of human behavior, nevertheless belong to economic analysis.

Given the importance of properly identifying the nature of the distributions observed in the data, we begin by discussing some common distributions applicable to categorical data, their generating mechanisms and practical aspects in their fitting. Then, we will turn to the analysis of the main ancient Greek textual sources, literary texts and inscriptions, as well as a particular subset of inscriptions concerning votive acts. We will examine them under the light of economics, either applied to linguistics or to religion sciences. Finally, we will examine some primary sources that can be structured into networks: the records of people's names, along with parent-child relationships, and full religious invocations, known as divine onomastic sequences, recorded from inscriptions,

which create a network of gods and qualifiers. This overview of some ancient Greek data will serve as an introduction to the types of phenomena we will examine in much greater detail, the following chapters, in Part II. The list of primary sources pertaining to ancient Greece we discuss here is not intended to be fully comprehensive, but covers a fairly wide range of domains. For the most part, the primary sources of interest to us are not designed to be electronically available in bulk: this is a reflection of the manner in which classicists use the data. As a result, it is necessary to program specialized software to exploit the websites where the information is available as very large numbers of narrow subsets. Simply extracting and arranging these datasets requires programming expertise and powerful systems; this may explain why, so far, they have not often been exploited in bulk by historians.

3.1 Identifying Distributional Patterns and Mechanisms

Most of the data sources we discuss here contain large amounts of data of a categorical nature. Before we begin to drill into them, and try to understand various aspects of ancient Greek culture under this light, it makes sense to establish what kind of patterns one may naturally expect from such data. Indeed, one could observe apparent regularities; what is the most direct manner in which to visualize them? It is worth noting that the analysis of distributions is not particularly common in cliometrics, nor in relation with complexity economics, as the data is rarely categorical in nature. The empirical analysis of distributions of unstructured data representing complex social or physical phenomena is nevertheless part and parcel of complexity sciences (Boccaro 2010). Most primary sources for ancient history being categorical or textual, the most direct handle to approach them is through a distributional perspective. In some cases, when dating is available, this may be combined with a diachronic analysis, but in this chapter we will restrict ourselves to synchronic analyses for the sake of simplicity. Since most traditional economic data is precisely situated through time, the question of a particular regime change, such as in Bastidon et al. (2020) or Abry et al. (2022), can lead to interesting insights, but we do not follow this angle here.

3.1.1 Some Useful Distribution Classes

Given the nature of the data at hand, we will typically seek to analyze the relationship between the size of some category (whether it be the number of times a word appears in a text corpus, or the number of votive acts a given god received, for example) and its rank. This can also equivalently be understood as considering the counter cumulative empirical probability: for an item $i \in [1..I]$ (a word, a god) with s_i observations, its rank r_i is the number of items with a size greater than s_i , so that $r_i = |\{j \in [1..I] : s_j \geq s_i\}|$. Hence the empirical cumulative probability for the random variable representing the size S is $\mathbb{P}[S \geq s_i] = \frac{r_i}{I}$.

The kind of metrics we observe tend to be positive, and for the most part, their frequencies are strictly decreasing as a function of the observed values. While there is an infinite number of random distributions that may account for these patterns, such positive observations are often compared with power laws, exponential distributions or truncated power laws, because these distributions are found in a large number of empirical phenomena and have strictly decreasing densities. These distributions are defined as follows:

- If X follows a power law of parameter²⁴ α and minimal value x_{\min} , we have

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x > x_{\min}} \frac{\alpha}{x_{\min}} \left(\frac{x}{x_{\min}} \right)^{-1-\alpha} dx.$$

A particular case of power law is Zipf's law, where $\alpha = 1$ and in which case the cumulative probability $\mathbb{P}[X > x]$ has the form $\frac{1}{x}$;

- If X follows an exponential distribution of parameter λ , shifted to take values above x_{\min} , we have

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x > x_{\min}} \lambda e^{-\lambda(x-x_{\min})} dx;$$

²⁴We follow the convention, more common in economics than in physics, of specifying the parameter as the exponent of the cumulative distribution, rather than that of the density. We hence have the same convention as Gabaix (1999), but that is different from that of Clauset, Shalizi, & Newman (2009), for example.

- If X follows a truncated power law of parameters α and λ , with minimal value x_{\min} , we have

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x > x_{\min}} \frac{\lambda^{-\alpha}}{\Gamma(-\alpha, \lambda x_{\min})} e^{-\lambda x} x^{-1-\alpha} dx.$$

When considering distributions for which the frequency may not be decreasing, the lognormal distribution, as the exponential of a Gaussian, is a good candidate:

$$\mathbb{P}[X \in dx] = \mathbb{I}_{x > 0} \frac{1}{x\sigma\sqrt{2\pi}} e^{-\frac{(\ln(x)-\mu)^2}{2\sigma^2}} dx.$$

The lognormal distribution can be shown to be asymptotically equivalent to a power law, for intermediate values (Saichev, Malevergne, & Sornette 2010, 12).

Power laws are fat-tailed. When one talks of a fat-tailed distribution for some positive variable X , it means that $\mathbb{P}[X > x]$ does not go quickly to zero as x grows; there is some non-negligible probability of extreme values for X . A fat-tailed distribution for a continuous variable X translates into a particularly wide distribution. When we consider categorical data, such as names or words, then it is not the data *per se* that follows a certain distribution, but the number of observations in each category. If S represents the number of observations in each such bucket, then saying that S is fat-tailed means that there should be particularly few buckets with a large number of observations, and many with very few observations. Hence, it is important to note that, for categorical data, a fat-tailed distribution of occurrences rather translates into heavily concentrated observations.

Power laws are commonly used in complexity science, because they frequently appear in nature and in many human phenomena, and may be produced by numerous types of underlying processes (Boccaro 2010; Mitzenmacher 2004). Power laws have indeed been applied in many aspects of economics (Gabaix 2016). The generating mechanisms for these distributions are typically associated with preferential attachment in network formation (Barabási & Albert 1999), or with random growth processes (Gabaix 1999). Power laws can also be shown to correspond to the limiting behavior of extreme fluctuations (Alfarano & Lux 2010). One particular case of power law is Zipf's law, with parameter α set to 1. Zipf initially observed that the use of words in human

language, measured as the relationship between the frequency of occurrence of words in natural language and their frequency rank, followed this particular distribution (Zipf 1949, 19–55). This pattern has been extensively studied in computational linguistics, and many models accounting for the emergence of this pattern in language have been proposed, although some have argued they cannot truly explain why language exhibits it (Piantadosi 2014). Power laws, including Zipf’s law, fundamentally recoup with economics in the sense that they can be shown to emerge as the result of optimization. This first came to light when Mandelbrot (1953) showed that Zipf’s law was optimal for communication, by considering there is a cost of using a word depending on the number of characters in it, and that the informational content in the word depends on its probability of use. More complex modeling of the communication process, factoring in the cost of encoding and decoding information, confirms the optimality of Zipf’s law for the use of words, as it reduces coding errors (Ferrer i Cancho & Solé 2003).

Making the generating mechanisms more general, for example in network formation, results in a richer class of distributions, for which the straight power laws form approximations (Jackson & Rogers 2007). The more general truncated power law, with an exponential tail, can be associated with random group formation. Indeed, Baek, Bernhardsson, & Minnhagen (2011) show that forming groups off of any kind of categorical data in order to minimize the cost of identifying a given element results in a truncated power law distribution for the group size. Exponential distributions are generally associated with random times, or surprise, due to their “lack of memory” through conditioning. Ely, Frankel, & Kamenica (2015) modeled the notion of surprise maximization from an economic perspective, but do not characterize the optimal distribution of surprising events in general. Borwein, Borwein, & Maréchal (2000), who solve for the so-called “hangman’s paradox” (how to best surprise a prisoner on the day of their execution), show that over long periods of time, the maximum surprised is reached with a partially exponential distribution. Connections have been found between exponential and power law distributions: the emergence of power laws can also be related to exponential processes by combining random event times with random observation times (Reed & Hughes 2003a). Exponential distributions can also emerge in networks: in a non-equilibrium network, continuously growing, if the new nodes appearing link to the existing nodes following a uniform distribution, then the network’s

degree distribution is exponential, with a parameter driven by the average degrees in the network (Deng et al. 2011).

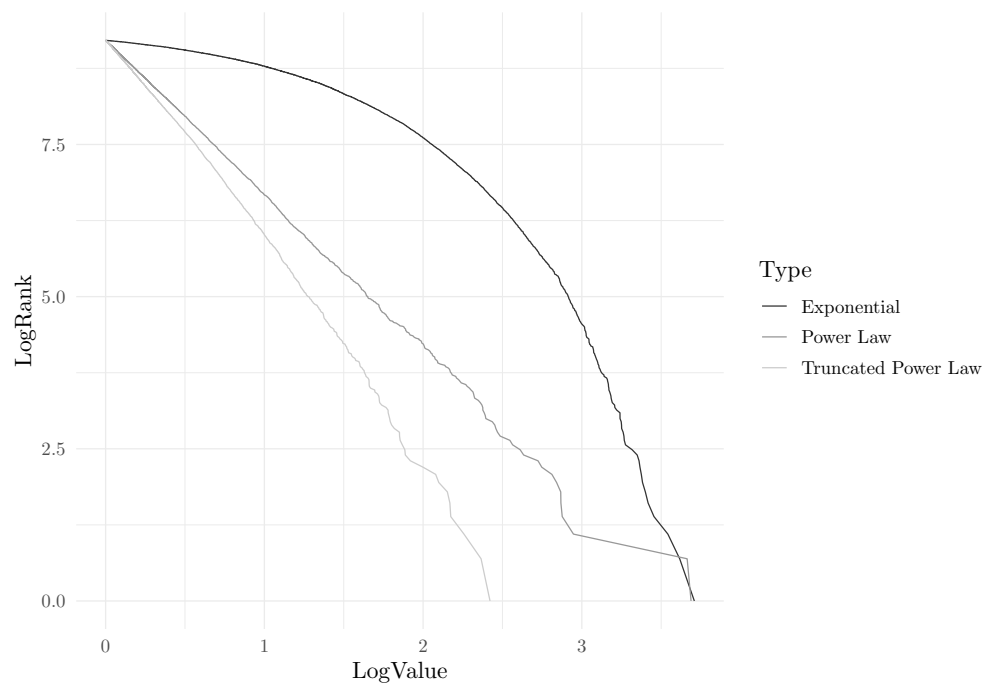
Considering the lognormal distribution, which we mentioned earlier, simply as the exponential of a normal distribution does not help much in understanding it, unless one has a specific reason for the exponentiation, such as exponential growth. Limpert, Stahel, & Abbt (2001) have noted the widespread occurrences of the lognormal distribution in biological phenomena and growth mechanisms, and point out that it is a good candidate for situations where the minimal size cannot be the most commonly occurring. Some economic data that were initially thought to follow Zipf's law have been shown to follow lognormal distributions, such as firm and city sizes (Saichev, Malevergne, & Sornette 2010, 2–5).

3.1.2 Empirical Distribution Comparisons

Figure 3.1 plots simulations of the first three types of distributions mentioned above, for a large number of random draws. The parameters are chosen so that the curves are relatively close to each other on the logarithmic scale. Their shapes are clearly distinguishable: the straight line of the power law and the curvature of the exponential stand out. The truncated power law's distribution plot shows a straighter part on the left side (for high frequency, low value cases), and a curved part on the right side, where x is large and the exponential behavior of the exponential dominates the density. The exponential's effect hence affects the tail, the largest observations, while the lower values retain a power law-like behavior. In this example, one may not need advanced statistical methods to establish the nature of these distributions: we can clearly see that the log/log plot, while heavily compressing the scale of the ranks and outcomes, does not betray the underlying density. It is not just any randomly generated numbers that, once ranked and scaled, will exhibit the patterns visible in Figure 3.1.

However, Clauset, Shalizi, & Newman (2009) offered a detailed treatment of the statistical methodology that can help empirically determine the type of probability law followed by some data, specifically in the case of power laws. They have shown that it is paramount to carry out proper distribution fit comparisons using a maximum likelihood approach, rather than simply carrying out a linear regression on the data on a

Figure 3.1: Comparison of Power Law, Exponential and Lognormal Distributions on a Large Sample



Note: The parameters for the distributions are $\alpha = 2.5$ and $\lambda = 0.25$. The samples contain 10000 draws.

logarithmic scale. According to their analysis, fitting a regression line on a log/log plot, such as was done in Gabaix (2016), for example, results in a gross over-estimate of the probability that the data follows a power law. Using the methodology, as implemented by Alstott, Bullmore, & Plenz (2014), we can compare the fits of the three distributions, as shown in Table 3.2.

We can see that the parameters estimates are very close to their true values, if we know what distribution to look up in each case. The goodness-of-fit ratios are clearly able to disambiguate the choice between power laws or truncated laws and the exponential. When we simulate a truncated power law, the goodness-of-fit also indicated a preference for that form over a simple power law. However, when we simulate a power law, the tests cannot cleanly distinguish it from a truncated power law. In this case, however, the estimated parameter in either case is close to that of the underlying simulation, and the decay rate is very small. Given that the truncated distribution with a small λ can come arbitrarily close to the straight power law, this is not surprising.

Table 3.2: *Summary Statistics on Distribution Fits*

| Statistic | Exponential | Power Law | Truncated Power Law |
|----------------|-------------|-----------|---------------------|
| Lambda Exp | 0.247 | 1.543 | 2.090 |
| Alpha Pow | 0.742 | 2.512 | 3.013 |
| Alpha Trunc | 0.000 | 2.429 | 2.478 |
| Lambda Trunc | 0.097 | 0.031 | 0.260 |
| Trunc vs Pow R | 78.353 | 1.138 | 3.147 |
| Trunc vs Pow p | 0.000 | 0.057 | 0.000 |
| Trunc vs Exp R | -17.306 | 11.510 | 8.632 |
| Trunc vs Exp p | 0.000 | 0.000 | 0.000 |
| Pow vs Exp R | -44.643 | 11.264 | 7.184 |
| Pow vs Exp p | 0.000 | 0.000 | 0.000 |

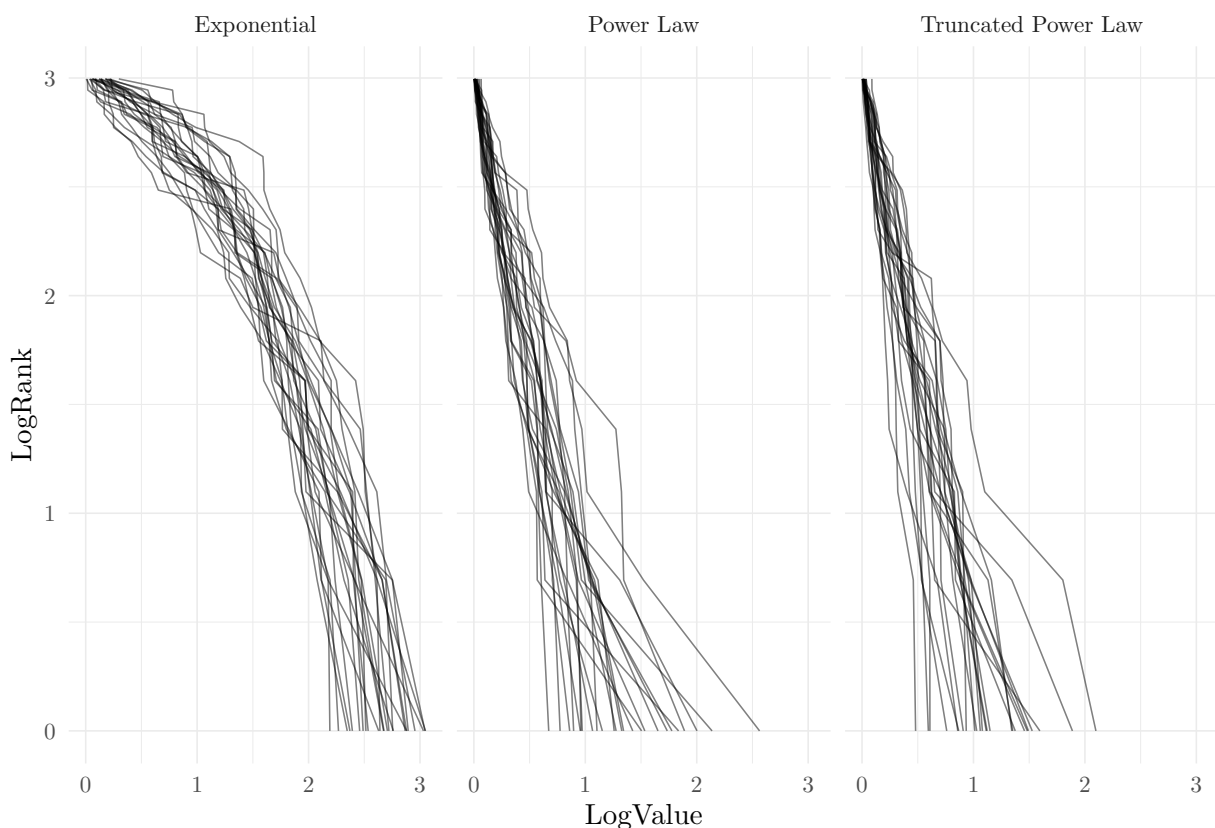
Note: The distribution names in the tests are abbreviated as follows: *Exp* = exponential, *Pow* = (pure) power, *Trunc* = power law with exponential decay. *R*: ratio of goodness-of-fit; a positive number means that the first law of the two is preferred. *p*: significance level; the probability that the preference would be due to randomness. The same abbreviations are used in other comparable tables.

3.1.3 Dealing with Sample Size

If the categories we are interested in are words in a large text corpus, or individual persons in some historical record, or individual physical inscriptions across a large geographic

space, then the number of these categories is very large, and sample sizes as well. However, in some cases one may not find so many different categories, when considering individual gods worshiped in a given city, or distinct characters across ancient Greek theater, for example. Figure 3.3 shows the empirical cumulative distributions for a series of small samples from the same distributions as those represented in figure 3.1. Even with few observations in each case, and the variability across samples, we can see that the exponential distribution and power laws are markedly different. The behavior of the exponential for low ranks and low values gives it away. Since the effect of truncation on the power law is only apparent on the much scarcer high value and high rank outcomes, the distinction between the truncated and pure power law is naturally more difficult.

Figure 3.3: *Comparison of Power Law, Exponential and Lognormal Distributions on Small Samples*



Note: The parameters for the distributions are $\alpha = 2.5$ and $\lambda = 0.25$. The 25 samples for each law contain 20 draws.

Distribution fits can be applied to each small sample, and the resulting statistics once aggregated are displayed in Table 3.4. The median significance levels for the

goodness-of-fit are not very high across the board: there is a high probability of the difference in distributions to be fortuitous. For power law and exponential distribution draws, the median parameter estimate comes out close to the real value, but that is not the case for the truncated power law; here again a consequence of the additional degree of freedom in fitting to a small number of observations. These results illustrate that, with a reduced sample, it may still be possible to distinguish between exponential and power laws (just as one could see it on the chart, in Figure 3.3), but the more flexible nature of truncated power laws makes them more difficult to determine.

Table 3.4: *Summary Statistics Across Small Sample Distribution Fits*

| Statistic | Exponential (Med.) | Exponential (St. Dv.) | Power Law (Med.) | Power Law (St. Dv.) | Truncated Power Law (Med.) | Truncated Power Law (St. Dv.) |
|----------------|-----------------------|--------------------------|---------------------|------------------------|----------------------------------|-------------------------------------|
| Lambda Exp | 0.250 | 0.056 | 1.949 | 0.926 | 2.215 | 0.680 |
| Alpha Pow | 0.735 | 0.081 | 2.815 | 0.911 | 3.102 | 0.738 |
| Alpha Trunc | 0.000 | 0.000 | 1.179 | 1.347 | 1.614 | 1.505 |
| Lambda Trunc | 0.105 | 0.037 | 0.902 | 0.994 | 0.821 | 0.752 |
| Trunc vs Pow R | 4.004 | 1.267 | 0.698 | 0.761 | 0.862 | 1.203 |
| Trunc vs Pow p | 0.006 | 0.006 | 0.542 | 0.327 | 0.577 | 0.344 |
| Trunc vs Exp R | -1.228 | 1.051 | 0.554 | 1.144 | 0.317 | 1.223 |
| Trunc vs Exp p | 0.164 | 0.155 | 0.404 | 0.325 | 0.379 | 0.281 |
| Pow vs Exp R | -2.749 | 1.880 | 0.166 | 1.419 | -0.076 | 1.565 |
| Pow vs Exp p | 0.060 | 0.084 | 0.380 | 0.262 | 0.388 | 0.313 |

Since the mechanics that can generate the observed distributions can be quite different, and correspond to distinct optimal behavior by agents, it is important to be able to differentiate between these distributions. The analysis of the controlled experiments above has shown us to what extent we may, in practice, rely on the statistical tests proposed by Clauset, Shalizi, & Newman (2009).

3.2 Ancient Greek Writings and Cliometrics

Texts, more than any other artifact from the past, are the prime raw material of historiography. The most natural primary source for the study of ancient Greece, logically, is the corpus of all recorded literature, covering the range from the Archaic period to late Antiquity. These texts have reached us, for the most part, through manuscripts which were copied through time; there is essentially no instance where we possess original literary writings. There are also numerous inscriptions, many with lacuna or missing characters, but they nevertheless constitute quite a large sample.

Hence, the corpus of ancient Greek texts available today is the result of a combination of chance findings as well as specific choices that were made over more than 2,500 years about which works deserved conserving, and which ones did not. In spite of these layers of selection and filtering, the texts we have today belong to a fairly diverse set of genres. In this section, we will concentrate in turn on two large ancient Greek text corpora: the Diorisis, which gathers literary works, and the PHI, which gathers inscriptions. Then, we will discuss votive acts data extracted from inscriptions.

It is sometimes helpful to leverage historical work in order to categorize data from primary sources. One important resource for this purpose is the voluminous compilation by Hansen & Nielsen (2004). These authors gathered information on a large variety of categorizations and metrics applicable to each *polis*, for instance such as its size, political regime, or affiliation. The POLIS database (T. Johnson & Ober 2014) is the computerized and augmented version of this inventory, covering over 1,000 *poleis* of the Greek world. The majority of *poleis* listed in the POLIS database have geographic coordinates and in many cases the primary sources can be mapped to the polity-level information from POLIS²⁵.

3.2.1 The Economy of Literary Genre

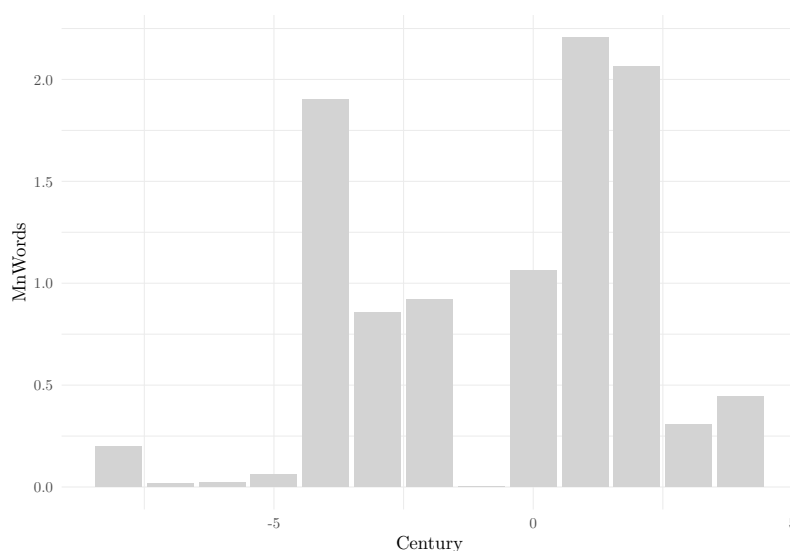
There are multiple electronically available resources compiling ancient Greek texts, such as the Thesaurus Linguae Graeca, or TLG, (Pantelia 2020), or the large Perseus website hosted by Tufts University (Crane 2012). These resources are nevertheless generally not absolutely comprehensive, do not allow users to download the data as one unique set, and only contain the raw text. For inflected languages such as ancient Greek, morphological inflection makes the identification of variations of a same word difficult. In order to associate any instance of a word to its lemma, the noninflected root, and hence to identify each word's role in a sentence, the words need to be categorized, with a so-called part-of-speech tagger (Ide 2004). The Diorisis is a centralized and comprehensive corpus of ancient Greek text that is already tagged (Vatri & McGillivray 2018b). In total, this corpus gathers 820 different works; it was originally composed in order to analyze semantic change in ancient Greek over time, with computational

²⁵The POLIS data itself, although many layers removed from primary sources, has been independently used in economics research: for example Fleck & Hanssen (2018) focused on political transition using this data exclusively.

linguistics methods, and is aimed at historians and classicists. The Diorisis data contains one row per word or punctuation sign, for a total of over 14m rows. A lemma is mapped to each word, representing its root: the nominative form for a noun, or the singular first person of the present indicative for a verb in most cases. The particular declension or conjugation of the word is also specified. All the corpus's data is available online, in the form of one file per book (Vatri & McGillivray 2018a).

Since the Diorisis contains over 800 different works, it is useful to be able to categorize these works. A practical catalog has been compiled for that purpose by Tauber (2021) For each one of the works, it gives a year estimate, and a two-level literary genre category. Combining the catalog data with the Diorisis corpus, we can carry out a wide range of analyses. Figure 3.5 shows the distribution of the number of words in Diorisis as a function of the year when the work is estimated to have been composed. We can see the large concentrations in the traditional classical period (5th and 4th centuries), in Hellenistic Greece afterwards, and during the Roman Empire in the early Christian period.

Figure 3.5: *Distribution of the Number of Words by Work Period*

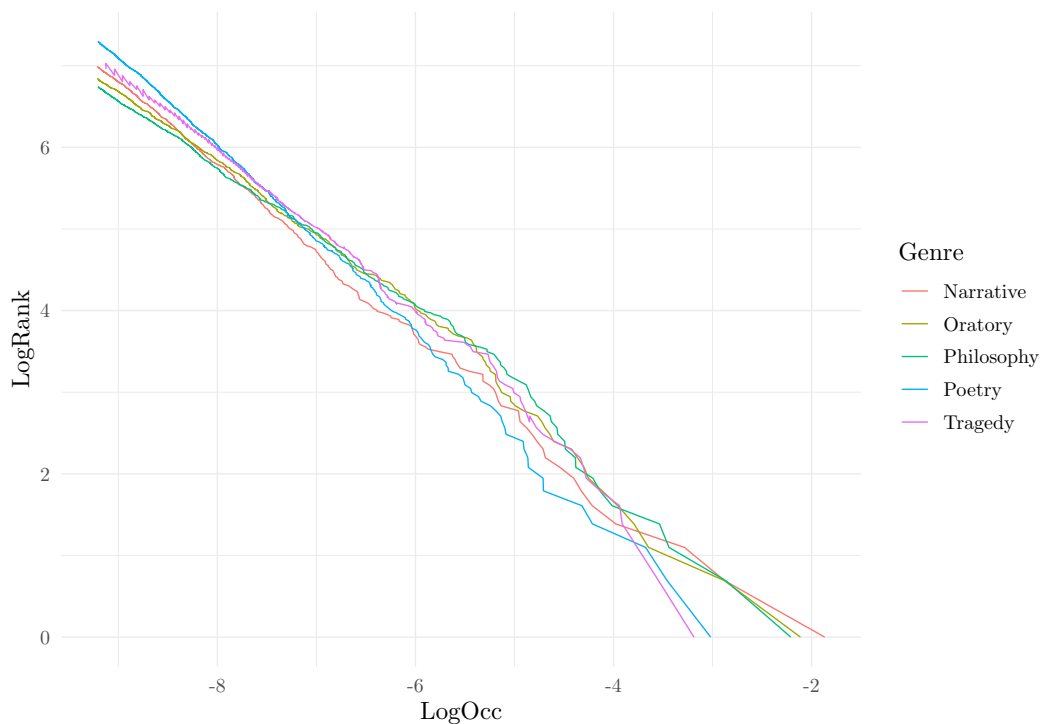


Note: The number of words is given in millions.

Typical linguistic patterns can also be observed on the data, such as the relationship between word rank in frequency of use, and probability of occurrence. Figure 3.6 illustrates this relationship in the Diorisis corpus, grouped according to literary genre

according to Tauber's catalog. The left hand side of the curves, representing the tail of the distribution, are indeed quite straight and with a slope close to 1, and it appears reasonable to call these distributions Zipf laws. We know that Mandelbrot (1953) showed the optimality of the general pattern of Zipf's law in language, and Ferrer i Cancho & Solé (2003) have shown that power laws may emerge from a least-effort optimization in communications, and that its emergence may be associated with the use of symbolic concepts. We can note, however, that there are some variations in the power law coefficients in our data. Poetry, comedy and tragedy all seem to exhibit thinner tails than the other genres, with their most frequent words being less common than for other genres.

Figure 3.6: *Log/Log Cumulative Distribution on Diorisis Corpus for Various Literary Genres*



Note: The data includes the lemmas for each genre that have more than 0.01% of occurrences. The horizontal axis is the logarithm of normalized frequency of each lemma, and the vertical axis is the logarithm of the lemma's rank.

Table 3.7 shows the application of the distribution statistics discussed in Section 3.1.2 to the data displayed in Figure 3.6. We can see that in all five cases, the distribution is identified as a power law or truncated power law, without a strong distinction between the parameters for these two. The parameter α is close to 1: lower than 1 for prose

(fatter tail), and higher for poetry (thinner tail). We can presumably relate this pattern to the fact that some bodies of literature typically resort to a broader range of vocabulary than others. Studies on contemporary languages have exhibited similar differences by genre, as well as across languages: see Grabska-Gradzińska et al. (2012) for example, who focused on English and Polish. A cost-benefit analysis may allow us to form a more precise understanding of the differences we can observe in the distribution fits for these various literary genres. Poetry, in ancient languages, is mostly defined by metric constraints, as Devine & Stephens (1976) illustrated in detail in the case of Homer: the rhythm of long and short vowels must follow a particular pattern. This constraint creates the need for more varied vocabulary, given a message to communicate, beyond the aesthetic search for rare words. It is possible for poets to sometimes deviate from the norm, for a particular literary effect, or because they could not otherwise express their meaning, but such a deviation has an aesthetic cost. Tragedy is written in verses (Saïd & Trédé [1990] 1999), but in spite of being subject to the same type of constraints as poetry in general, it has a steeper distribution (smaller α). Unlike poetry in general, tragedy must be able to express vivid dialog, and be clearly understandable, and it is generally considered to be closer to standard language (Vernant & Vidal-Naquet [1972] 2001). Hence, the difference in the fitted distribution parameter between the simple narrative and tragedy may capture the effect of metric constraints, while the difference between poetry and tragedy may capture the ability to use more flowery vocabulary as the coding/decoding constraints related to the message are less pregnant. Oratory texts, mostly Athenian judicial speeches, also need to strike the right balance between the need to be understood by the judges (an assembly of hundreds of citizens) and the necessity to use specialized vocabulary, in particular in financial affairs. Nevertheless, the potentially flowery language of the orators is significantly closer to simple narrative texts than to poetry. While it is difficult to construct an *a priori* model that could account for the magnitude of these differences, a formal model of optimal communication could allow us to scale the differences across genres, and potentially draw comparisons with genre-specific corpora in modern languages.

Table 3.7: *Summary Statistics on Distribution Fits Across Literary Genres*

| Statistic | Narrative | Oratory | Philosophy | Poetry | Tragedy |
|----------------|-----------|---------|------------|--------|---------|
| Lambda Exp | 0.001 | 0.001 | 0.001 | 0.003 | 0.003 |
| Alpha Pow | 0.792 | 0.852 | 0.820 | 1.137 | 1.091 |
| Alpha Trunc | 0.747 | 0.808 | 0.777 | 1.103 | 0.874 |
| Lambda Trunc | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Trunc vs Pow R | 1.574 | 1.552 | 1.824 | 0.991 | 1.479 |
| Trunc vs Pow p | 0.000 | 0.013 | 0.011 | 0.208 | 0.074 |
| Trunc vs Exp R | 4.814 | 5.171 | 6.085 | 4.966 | 3.510 |
| Trunc vs Exp p | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |
| Pow vs Exp R | 4.739 | 5.095 | 5.998 | 4.916 | 3.086 |
| Pow vs Exp p | 0.000 | 0.000 | 0.000 | 0.000 | 0.002 |

Note: The data excludes words with less than 100 occurrences in text.

3.2.2 Epigraphy and Constraints

The ancient Greeks used to inscribe a wealth of information on stone and on various artifacts, and these inscriptions are found on monuments, on steles, on vases and in many other places. These writings have reached us in their original form, albeit with some lacunae. This makes for a different situation from literary texts, which, with the exception of papyri, were copied many times over as manuscripts, and very often curated. The literary tradition has kept works that were deemed of value, but many written documents, such as accounts or contracts, have been lost. Inscriptions therefore provide unique evidence from ancient Greece, since they constitute true original documents.

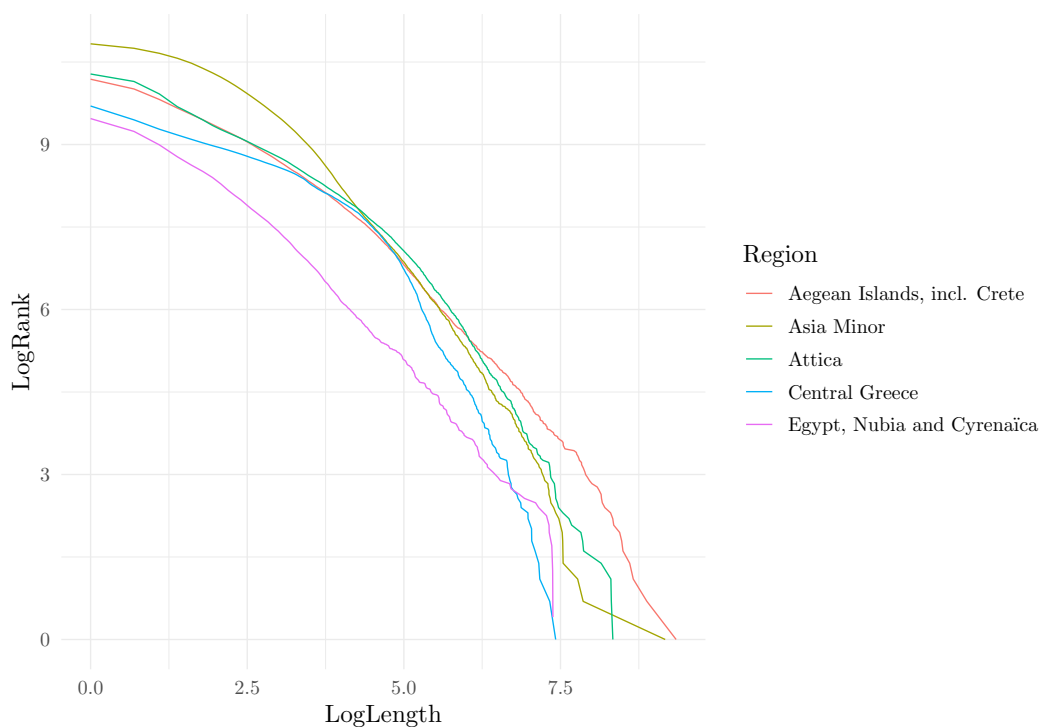
Epigraphists read and analyze these documents, and transcribing, editing and contextualizing the texts from archaeological material is a complex process. Many inscriptions are in a damaged or fragmentary state, and epigraphy is also concerned with forming the best hypotheses in order to fill in the missing information, based on some commonly observed recurrences (McLean 2002). Epigraphic sources have been collated into large volumes since the Renaissance, and there are few centralized editions for Greek or Roman Antiquity. Online, the largest and most comprehensive repository of Greek inscriptions is the Packard Humanities Institute (PHI)'s Searchable Greek Inscriptions website ("PHI Greek Inscriptions," n.d.). The PHI data available online is presented as one webpage for each inscription, that contains the inscription's text as well as some additional geographic information. Since there is no centralized database available containing the entirety of the PHI data, it must be processed page after page using an automated web browser. Since, for historians, inscriptions are usually analyzed

one by one, or in small groups, epigraphic data has not been, to date, considered and analyzed as a bulk, in a systematic fashion. Once all the inscriptions are electronically available in one place, it is possible to process them in order to map each recognizable word to a lemma, using the Classical Languages Toolkit (K. P. Johnson et al. 2019). Further, location information can be joined with the *polis*-level categorization data from the POLIS database.

With the centralized PHI data, one can carry out cross-sectional analysis which would be otherwise impossible. Figure 3.8 shows the distribution of the number of words in the PHI inscriptions on a logarithmic scale, for a few regions where there are the largest number of inscriptions. We can see that the message length distribution is different across regions. The shapes of the curves seem roughly consistent with an exponential distribution for some, and a power law for others. There are several factors that could affect the length of inscriptions. The material may certainly have been randomly broken, but in quite a few cases, fragments are paired back together, forming unique inscriptions. Mostly, it seems logical that the cost of putting messages to the stone and the nature of the messages should be the drivers affecting the length of these messages. The features that differentiate these distributions the most from straight power laws appear to be the tangents for very short lengths, almost flat in most cases. This indicates that single word or two-word inscriptions are much less frequent than a power law would imply, if it was to be consistent with the number of occurrences of longer messages. On the other hand, very long inscriptions are also shorter than a straight power law would imply. This cut-off, which may be captured by an exponential truncation, would seem to be necessary in order to reflect the constraint of material availability as well as the greater likelihood that longer inscriptions be broken and fragmentary. Short messages, in contemporary communication and advertising, are more effective, all else being equal (Baltas 2003). While ancient inscriptions are not the same thing as internet ads, they share the same aim of publicity. One may wonder, hence, what mechanisms lead to the distributions in Figure 3.8. While literary texts are essentially narrative, inscriptions sought to communicate specific information. They mostly recorded political decisions, treaties between cities, honors given, funeral rites, the freeing of slaves, or religious dedications (McLean 2002). Inscriptions therefore tend to express actions that some people take concerning other people; for example, a

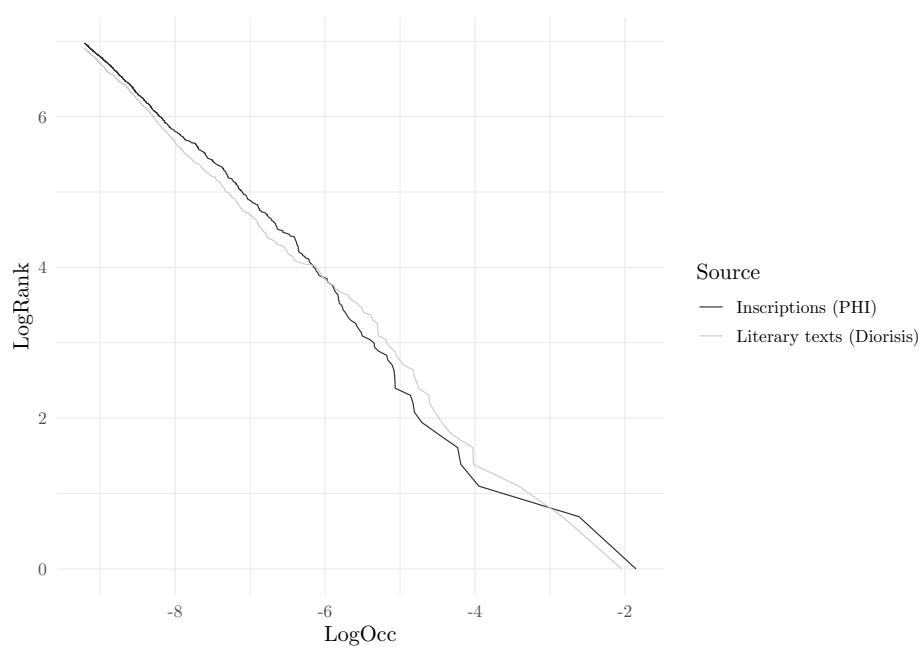
husband has a tombstone engraved in honor of his deceased wife, or the assembly honors a prominent citizen. As a result, the minimal information that can be communicated is of the form “A does B”, which necessarily uses at least three words. On the other hand, engraving is costly, and more or less so depending on the region and on how common the “epigraphic habit” is there. There are many instances of single-word funeral inscriptions such as $\chi\alpha\acute{\iota}\rho\epsilon$ (“be well”). The left-hand part of Figure 3.8 hence likely points to an optimization of the message in inscriptions, relative to production costs.

Figure 3.8: *Log/Log Cumulative Distributions of the Number of Words in Inscriptions for Selected Areas*



Interestingly, although inscriptions were generally short, the distribution of word lemma occurrences broadly matches that of a literary corpus. Figure 3.9 plots a comparison of the distribution of lemma occurrences in the Diorisis literary corpus and in the PHI corpus. They look very similar, which indicates that the language on inscriptions has comparable features to the standard literary Greek language.

Figure 3.9: *Log/Log Cumulative Distribution on Literary Texts and Inscriptions*



Note: The data includes the most common lemmas for each corpus, which account for more than 0.01% of occurrences. The horizontal axis is the logarithm of normalized frequency of each lemma, and the vertical axis is the logarithm of the lemma's rank.

Nevertheless, a more precise statistical fit, displayed in Table 3.10, shows that the use of words in inscriptions is less fat-tailed than in literary works: the inscriptions use rare words more often. This is presumably not related to a greater presence of poetry in these inscriptions, but rather to the more common appearance of random proper names, which would be much less likely in literature. In a nutshell, there is a greater amount of information content in inscriptions. In all the types of inscriptions we mentioned above, particular people are named, and these names are a substantial part of the information that is conveyed, making each inscription unique in that sense. The language of ancient Greek inscriptions, hence, may be seen as optimizing its informational content, given a certain cost of engraving. Econometric tests relating the vocabulary range in inscriptions with the size of the inscriptions and across regions could thus give a new perspective on the informational landscape of the ancient *polis*.

Table 3.10: *Summary Statistics on Distribution Fits*

| Statistic | PHI | Diorisis |
|----------------|-------|----------|
| Lambda Exp | 0.001 | 0.001 |
| Alpha Pow | 0.827 | 0.771 |
| Alpha Trunc | 0.789 | 0.743 |
| Lambda Trunc | 0.000 | 0.000 |
| Trunc vs Pow R | 1.351 | 1.968 |
| Trunc vs Pow p | 0.000 | 0.000 |
| Trunc vs Exp R | 4.522 | 7.291 |
| Trunc vs Exp p | 0.000 | 0.000 |
| Pow vs Exp R | 4.463 | 7.224 |
| Pow vs Exp p | 0.000 | 0.000 |

Note: The data excludes words with less than 100 occurrences in text.

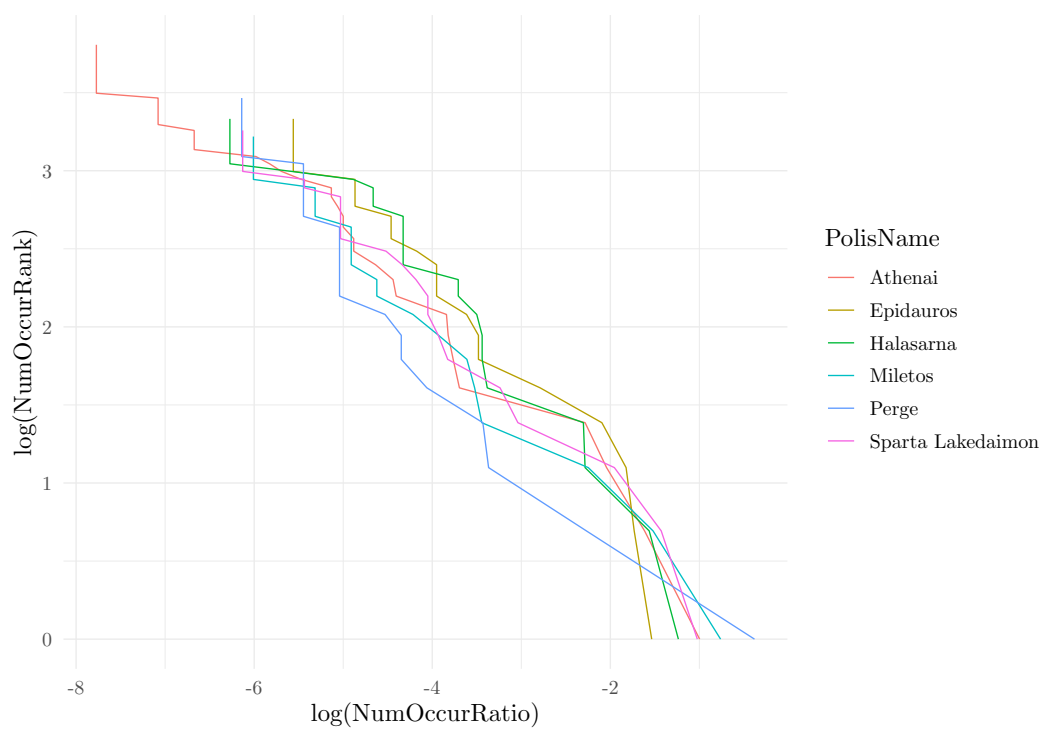
3.2.3 Gods and Market Shares

From the textual sources we have mentioned above, historians and classicists have created other, more complex corpora that are electronically available. We focus here on a particular example concerning religious worship. The study of ancient Greek religion and ritual is a significant part of historical research on the period. Certain inscriptions give us a glimpse of actual votive acts, and have been closely examined for this purpose. Epigraphic sources related to religious offerings or dedications typically contained the names of one or several gods along with some qualifications and reasons for the inscription. The centralization and analysis of such inscriptions have been carried out

for over a century, starting with thousands of paper files in the 19th century. Bonnet & Lebreton (2019) discuss the historiography of the study of divine invocations, and explain how methods have evolved over time, converging towards the use of databases. One instance of such a database is the Base de Donnée des Epiclèses Grecques (BDEG), a project initiated in the early 2000s (Lebreton et al. 2014; Brulé & Lebreton 2007). It contains the information from thousands of epigraphic inscriptions or literary references to votive acts, including location, the god or gods names and the nature of the inscription. More specifically, researchers gathered and made available data about votive acts from all sources, periods and region they may come from. The god name to whom the dedication was mainly addressed is recorded, typically in the first place in the invocation, with some additional qualification (commonly an adjective). The additional qualification, named cult-epithet, usually describes a particular feature of the god.

The data has been entered in over 11,000 forms, with in principle one for each observation of a divinity and epithets. The BDEG presents its data as separate webpages: it is designed to allow researchers to visualize one form at a time, and, here again, in order to analyze the data in bulk, it is necessary to programmatically collate, parse and compile all the data from all these webpages. Location information is provided in some cases with coordinates, and in other cases only with the gallicized Greek name. Once the names are manually mapped to the corresponding anglicized Greek names, location information can be joined with the data from the POLIS database. The geographic spread of the data covers the entire Greek world, and there appears to be a definite concentration on some particular gods. The divinities in the Greek pantheon were indeed far from equally treated by humans. The distribution of votive acts across individual gods for a few specific locations is shown in Figure 3.11.

Figure 3.11: *Log/Log Cumulative Distribution of the Number of Votive Acts Among Gods, for Selected Poleis*



Note: NumOccurRatio is the ratio of the number of votive acts for a given god relative to the total for the polis.

The somewhat straight lines are indicative of a power law, as confirmed by the statistical tests from Table 3.12, carried out on the entire dataset, not grouped by location. The estimate gives us approximately $\alpha = \frac{1}{2}$, so that the distribution has a fatter tail than Zipf's law: in each city there are a few gods that attract large numbers of dedications, but not the same gods are the most popular in every city. How can one make sense of these patterns? This is a question we have alluded to in Chapter 2, when introducing this example. Research in religion sciences based on economics has tended to focus on contemporary religion economics (Iannaccone 1998) on one hand, and, on the other hand, on the conditions leading to the existence of religion, or of certain features of ritual (Bulbulia 2010; Bulbulia & Frean 2010; Henrich 2009). In the economic modeling of the existence of religion, a cult tends to be considered as a club good, the benefit of which accrues as there are more participants. This type of model is not adapted to a polytheist world where worship is distributed. If one considers the benefits from the gods as a finite quantity, shared by worshipers, then a game equilibrium can consist of mixed strategies, hence potentially accounting for the fact that all gods get at least some worship. Given there are several "products", the fact that votive acts are distributed according to a power law could be envisioned as a "natural" pattern in goods consumption, as Kohli & Sah (2006) empirically showed that was the case in market shares. Other, more fundamental, explanations can be envisioned by considering a form of preferential attachment: the more popular gods attract more people, so that after a large number of cycles, one may observe a power law.

Table 3.12: *Summary Statistics on Distribution Fits for Votive Acts*

| Statistic | BDEG |
|----------------|-------|
| Lambda Exp | 0.013 |
| Alpha Pow | 0.482 |
| Alpha Trunc | 0.510 |
| Lambda Trunc | 0.000 |
| Trunc vs Pow R | 1.710 |
| Trunc vs Pow p | 0.060 |
| Trunc vs Exp R | 7.175 |
| Trunc vs Exp p | 0.000 |
| Pow vs Exp R | 7.106 |
| Pow vs Exp p | 0.000 |

In spite of the fact that the primary sources we have examined above were essentially unstructured categorical data, we have seen that a detailed analysis of their distributions and their drivers, using simple economic-driven paradigms, can shed light on many

aspects of ancient Greek society. The perspective we acquire thanks to the application of complexity methods to primary sources treated in bulk, in a cliometric context, gives us new insights into the data, even though these texts and inscriptions have been closely studied by humanists and classicists since the Renaissance.

3.3 Networks, Complexity, and Naming Gods and Men

The use of network theory by historians has expanded over the years (Lemercier 2012; Karila-Cohen et al. 2018), and has given rise to renewed epistemological debates around which kinds of historical data could be conceived of as networks. Network analysis in historiography is usually carried out in a descriptive way, to account for a set of relationships extracted from a given document corpus. The application of network analyses to literary texts, and to classical texts in particular, has also been researched, most often by physicists or computer scientists. For example, Kydros, Notopoulos, & Exarchos (2015) looked at the networks between characters in mythology, and in their book, Kenna, MacCarron, & MacCarron (2017) discuss the analysis of networks stemming from various mythological and historical texts. With a less quantitative but more illustrative logic, Rydberg-Cox (2011) was interested in networks in the specific context of the Greek tragedy as well as their visualization. Finally, the works of Waumans, Nicodème, & Bersini (2015), Rochat (2014) and Elson, Dames, & McKeown (2010) provide a general perspective on the use of networks in literature. In all these approaches, however, the precise nature of the links that put individuals (characters in a play, or historical characters in a narrative) in a relationship is not well determined: what exactly does it mean, in terms of a relationship, that two names appear in the same sentence? Networks are a frequent tool in complexity economics, and in particular in cliometrics: for example in order to account for international market integration through time (Bastidon et al. 2019; Bastidon & Parent 2022).

In this section, we concern ourselves with some ancient Greek data which naturally enough represents relations, and have therefore been tackled by historians as networks, in a few instances at least. These particular data are mostly compiled from inscriptions: they are the Lexikon of Greek Personal Names, and the Mapping Ancient Polytheisms database. Both rely on arduous work carried out by specialists, having sifted through

hundreds of thousands of inscriptions in order to produce the resulting electronic data.

3.3.1 The Economy of Names: Onomastic Networks

In ancient Greece, something as fundamental as naming a baby worked in a very different fashion from what we are familiar with today: one would only get a single anthroponym, a unique name. To this unique name could be associated a patronym, the name of one's father. In some cases, a person could acquire a nickname through their actions or their physical aspect, but the single name remained a core principle. The choice of a name by the parents carried meaning, and could reflect through etymological links the child's belonging to the broad family. Giving to a child the name of a grand-parent, paponymy, was also quite common. Names were often formed through derivational morphology: using one or two nouns or adjectives to make a name. In the study of ancient history, the analysis of names, onomastics, often goes hand in hand with prosopography, that is the gathering of all available historical information on particular individuals. For persons who have been considered as important in historiography, one can usually get sizable volumes of information. For less well-known people, it is a complex historical inquiry (Karila-Cohen 2016).

The available epigraphic sources provide anthroponyms, sometimes with a patronym. In order to go from lists of names to the notion of individuals, it is necessary to relate these inscriptions to each other. Prosopographers rely on instances of identical, or closely related names in a given geographic area, and are dated around the same time, so that they can transform these names into references to particular individuals. A systematic perspective on how to carry out this research was proposed in Bresson (1981), who suggested the use of certain family relationships in a network context, combined with the naming conventions that appeared to be most common, in order to build a family tree. This logic initially applied to Rhodes has been generalized (Karila-Cohen 2018). In his his seminal study, he made it a central hypothesis that the practices observed in Rhodes in the modern period and up to contemporary times obeyed similar rules to those practiced by the Ancients. He showed among other things the way in which paponymy was applied, with the names from the father and mother's sides alternating. Using this historical information, one can therefore create onomastic networks: someone with such name had a child with such name. Then, with

prosopographic work, one can create prosopographic networks: such person had such person as a child. There is more onomastic data available than prosopographic, and the prosopographic reconstruction relies on many assumptions.

Large volumes of onomastic and prosopographic data have been made electronically available through the Lexicon of Greek Personal Names (LGPN) housed by Oxford University (Parker, Yon, & Depauw 1996). The project started in the early 1970s, and has led to the publication of references to hundreds of thousands of names. The electronic interface to the LGPN is not designed for its data to be processed and analyzed in bulk, as is unfortunately often the case with digital sources in classics. It is designed as a tool to query a name or a name root, and observe its occurrences. It is impossible to simply download the entire data in a structured form in one batch; it has to be reconstructed from the data pertaining to each possible name. We created a centralized dataset containing all these elements, in a structured fashion in the sense that all the data is stacked together in unique data tables, and so that it may be joined with additional data sources, such as the POLIS database through geographic information. We found close to 40,000 unique names spread across about 350,000 individual entries. The relationship table includes approximately 250,000 links.

Before looking into the network aspect of naming relationships, we examine the distribution of men and women names. Women were essentially excluded from political life, and inscriptions very often reported official acts, so that the the number of observations is biased towards men. Indeed, the number of feminine names in the dataset is a tenth of that for men. Table 3.13 shows distribution fits for the names of men and women, across the entire dataset. The parameters for women indicate a flatter distribution: women names are less concentrated. The distribution fits in Table 3.13 also tell us that the distribution is more likely to be a truncated power law than an exponential, with a very small exponential parameter λ . Baek, Kiet, & Kim (2007) have shown that family names distributions across the world tended to follow power laws, but not always, and that the parameters were not constant from one country to the other. Focusing on the contrary on given names, specifically in the US and over time, Li (2012) showed that a simple power law could not fully account for the distribution of names. In general, Baek, Kiet, & Kim (2007) estimated that the power law coefficients in contemporary data on last names were generally around 2. We can see that the

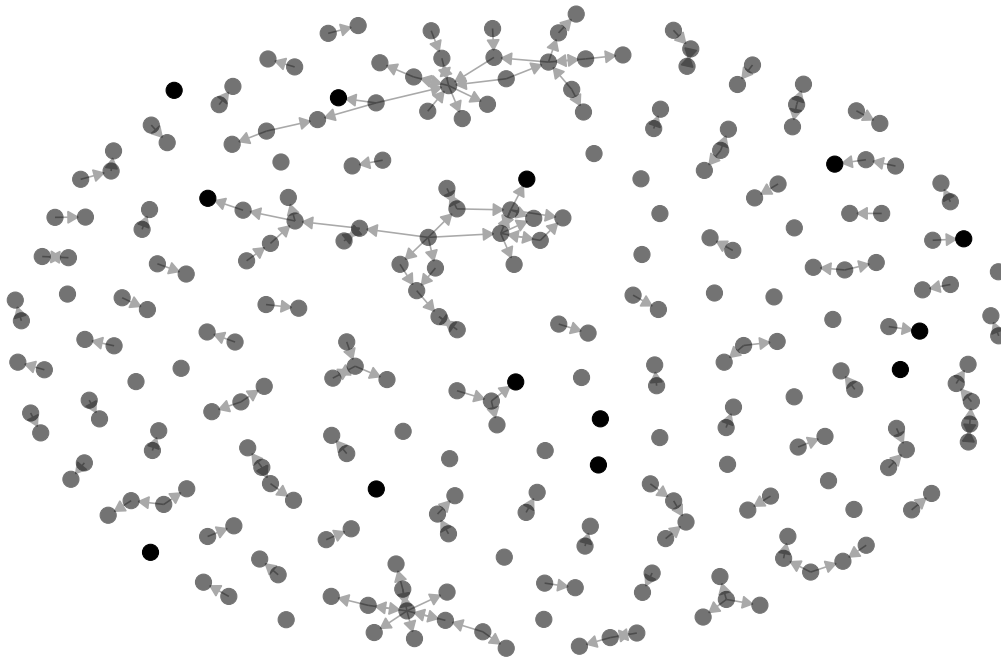
ancient Greek names exhibit a different behavior, with coefficients closer to 1. The lower α is indicative of a fatter distribution tail: names were more concentrated in ancient Greece than they are in the contemporary world. In modern times, names are family names, and are hence quite idiosyncratic, while given names are not used as a unique identification mechanism. The names in ancient Greece were given, or chosen, so that they were presumably selected from an existing stock, and affected by fashion and taste, which should translate into a higher degree of concentration.

Table 3.13: *Summary Statistics For Name Distributions Fits by Gender*

| Statistic | Man | Woman |
|----------------|--------|--------|
| Lambda Exp | 0.104 | 0.284 |
| Alpha Pow | 0.654 | 0.783 |
| Alpha Trunc | 0.747 | 0.946 |
| Lambda Trunc | 0.001 | 0.003 |
| Trunc vs Pow R | 18.134 | 12.879 |
| Trunc vs Pow p | 0.000 | 0.000 |
| Trunc vs Exp R | 33.355 | 22.523 |
| Trunc vs Exp p | 0.000 | 0.000 |
| Pow vs Exp R | 32.623 | 21.094 |
| Pow vs Exp p | 0.000 | 0.000 |

Using the parental relationships between names in a given location, we can build a network of these links. We implemented a comparable method as that laid out in Karila-Cohen (2018), where she focused on particular demes (suburbs) in Athens, but applied it to the entire LGPN data. We constitute a network where the edges are unique names, and the vertices are parental relationship entries in the LGPN: the resulting network is hence onomastic from the perspective of the edges, and prosopographic from the perspective of the vertices. Restricting the LGPN entries to the same demes as in Karila-Cohen (2018), we find approximately the same nodes and edges. As an illustration, Figure 3.14 displays the network thus obtained, reduced to the demes Oion Kerameikon and Oion Dekeleikon in Athens. We can see that a majority of nodes are linked, and even for this small network there is a wide range of components size. This network effectively displays the ancient Greek's child naming strategy in Oion, as a function of the father's name. The darker nodes, representing the most common names, are either isolated or are the leaves in this directed network. There are many observations of names for which there is no record of a child or a parent, so that, in the case of Oion at least, it may have been the case that the people with the most common names were, all else being equal, less likely to see their children appear in inscriptions.

Figure 3.14: Naming Network for the Oion Demes



Note: The arrows represent the naming relationship (\rightarrow), and the darker the nodes, the more frequent the name is.

Building all local networks at the *polis* level in the same manner, we can create a large network covering the entire dataset, for both men and women, and for the whole Greek world across time periods. Focusing on this large network, Table 3.15 displays the degree distribution fit, that is, the distribution of the number of different names that have been chosen by a person with a given name, in a given location. The number of distinct names chosen by each name appears to be a fat-tailed power law, with a coefficient around $\frac{1}{3}$. Let us assume there is a large network composed of individuals, nodes, each one with a name attribute, such that a new person appearing has no material effect on population counts. A naming relationship is a link created between a parent and the child, the new person. Such a prosopographic network can be collapsed into an onomastic network, where all the individuals are gathered in a single name node and the links grouped by the nodes from which and to which they go, as with the networks analyzed in Figure 3.14 and Table 3.15. If the individuals are distributed among attributes following a given distribution, and if the naming links are distributed uniformly, then the both the in- and out-degree distribution of the nodes

in the name network should follow that given distribution. We saw that names were distributed according to a power law, potentially truncated, according to Table 3.13, and the in-degree distribution for the name network is also a power law, according to Table 3.15. However, their parameters are quite different, the degrees being more concentrated than a random name selection would suggest. There is therefore a certain strategy at play in the selection of names, according to our data. This pattern could be accounted for by preferential attachment: if new links are more likely to attach to the nodes as a function of their existing number of links, then we would expect a greater concentration in the resulting degree distribution than in the underlying name distribution. The ancient Greeks' naming strategy, concerning people whose father name was mentioned along theirs in inscriptions, hence, would appear to reflect a certain conformism. We will closely examine the optimal choice of names in Chapter 4.

Table 3.15: *Summary Statistics on Distribution Fits for the Number of In-Degrees in the Naming Network*

| Statistic | LGPN |
|----------------|-------|
| Lambda Exp | 0.002 |
| Alpha Pow | 0.353 |
| Alpha Trunc | 0.341 |
| Lambda Trunc | 0.000 |
| Trunc vs Pow R | 3.646 |
| Trunc vs Pow p | 0.134 |
| Trunc vs Exp R | 7.472 |
| Trunc vs Exp p | 0.000 |
| Pow vs Exp R | 7.371 |
| Pow vs Exp p | 0.000 |

3.3.2 How to Speak to the Gods? The Economy of Divine Formulae

The BDEG data we discussed earlier represented information in an unstructured fashion on the way ancient invocations addressed the gods, in addition to the god's name: the forms in that database contained one or two epithets at most in each case. On a large scale, this data is useful in order to track the number of votive acts, but it is difficult to exploit if one wants to precisely understand the way in which the gods were invoked. The more recent large-scale MAP project (Bonnet 2017) offers a more precise and entirely revamped representation of the underlying data, mostly derived from inscriptions, covering the Greek and Semitic ancient worlds, and allows for the close study of the formulas the Ancients used to address divine beings. The MAP is

used by researchers for the philological study of dedications, and allows them to easily find series of invocations that mention a particular god or qualifier. To understand the relational logic which structures these divine powers, MAP takes into account a wide range of divine onomastic sequences, combinations of divine names or elements (names, epithets, titles, propositions), some shared by several gods, others specific to a particular god. Indeed, the simple act of performing a rite in ancient Greece involved addressing one or more gods with a complex series of qualifiers, the *epicleses*, presumably ordered and expressed in a deliberate manner. These qualifiers, which were sometimes common between certain gods, effectively created a network between all the gods; and the gods also created a system of relationships between the qualifiers. These onomastic sequences taken as a whole form a representation of the way in which the Greeks conceived of polytheism. The detailed study of onomastic sequences has been going on for many years, as we alluded to earlier, and recent research in this area includes Brulé (1998), Brulé (2005), Bonnet & Belayche (2017), who establish various assumptions implicit in how onomastic sequences are constructed. The idea of putting these sequences within the framework of network analysis is in fact at the basis of the MAP project, and another cornerstone is the extraction of onomastic formulas from the sources which, thanks to a particular syntax, can account for the great complexity of these inscriptions (Bonnet & Lebreton 2019).

The data that constitutes MAP, derived from the inscriptions by a team of specialists, is highly structured and, unlike what appears to be the norm in many ancient history electronic sources, accessible in bulk. Each material item, such as a stele or part of a monument for example, is an entry in a source data table, with detailed information such as its publication or location. A source contains one or several testimonies, addresses to the gods, each entered in a testimony table, so that each testimony, usually in the form of a sentence referring to divinities and qualifiers. These sentences are converted to formulaic expressions, akin to mathematical formulas, accounting for the links between the various elements in the testimony. These elements, mostly adjectives, nouns or god names, are also centralized in a specific element table. In order to understand how to construct a network that captures the semantic relationships between gods, it is necessary to examine the logic of onomastic formulas. One example of a testimony would be number 100, from Egypt, in the MAP database, from an inscription stating:

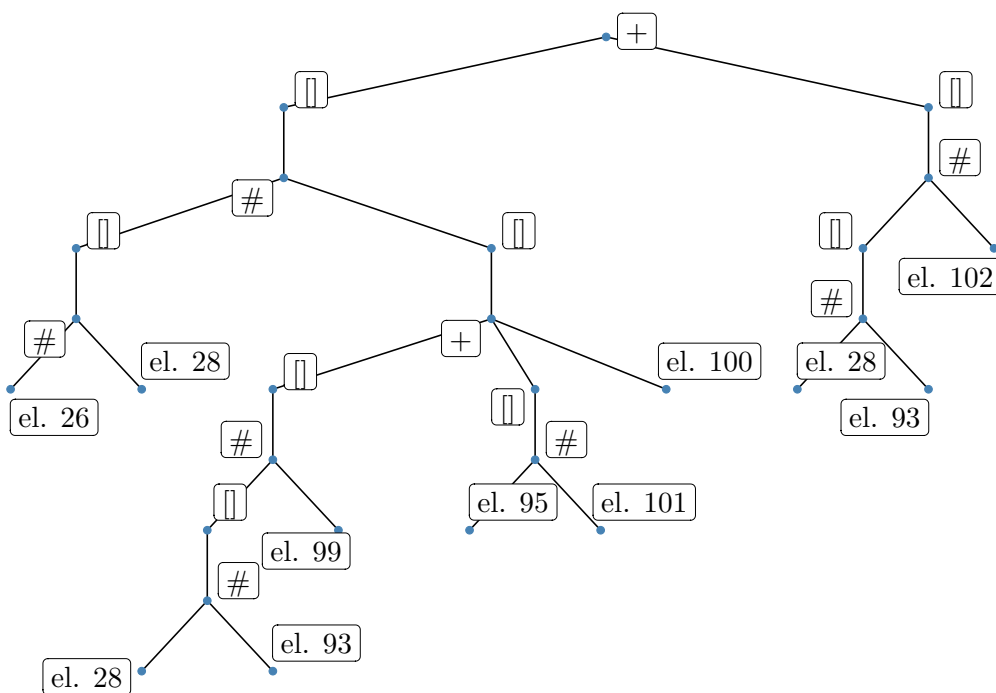
Beside the Lords Gods Priô the Greatest God and Horegebthis and Isis Rhesakemis and the Greatest Gods Who are with them²⁶.

This dedication is encoded as a formula referencing elements like the gods (Prion or Isis) or qualifications (Great), as well as the connections between them (such as “and”, for example). In this case, this is recorded as the following formula:

$$[[\{26\}\#\{28\}]\#\{99\}\#\{28\}\#\{93\}]]+\{100\}+[\{95\}\#\{101\}]]+\{102\}\#\{28\}\#\{93\}],$$

where the numbers are the identifications of particular elements, and the operators are the MAP researchers’ interpretation of the logic of the text. All these formulae can be systematically converted into tree network representations, as shown in Figure 3.16 for the case at hand²⁷.

Figure 3.16: *Syntactic Network Representation of Testimony #100*



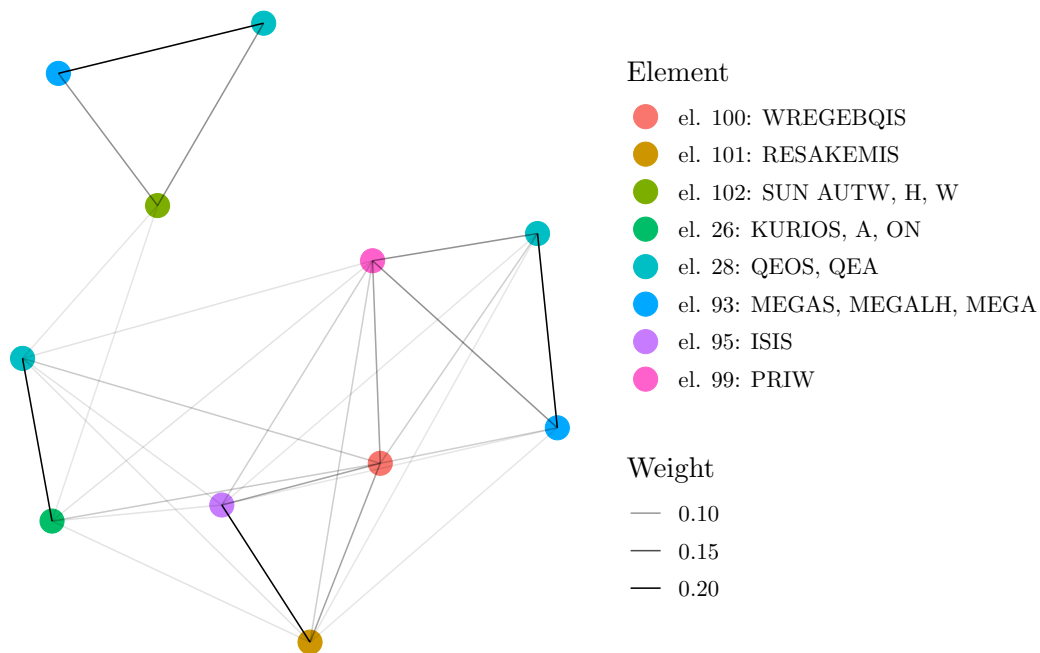
Note: Every node represents either an element or an operator and the links represent the application of an operator to one or several elements.

²⁶παρὰ τῶν κυρίων θεῶν Πριῶ τοῦ θεοῦ μεγίστου καὶ Ὁρεγέβθιος καὶ Ἴσιδος Ἰρσακέμεως καὶ οἱ σὺν αὐτοῖ θεῶν μεγίστων. MAP Translation.

²⁷Any operator-based expression could in fact be converted to a tree representation.

In order to better seize the nature of the links created between the elements and the gods, one approach is to allocate a certain distance to all the links in Figure 3.16 depending on the operators, so that the edge connecting two entities related by “and”, for example, would have a lesser distance than a simple apposition. Then, the total distance from any entity to any other entity in the formula can be computed, along the syntactic network. Finally, the strength of the connection between two entities, which can be used as a weight in the graph connecting all the entities appearing in a formula, can be chosen as the inverse of the distance. Figure 3.17 shows a representation of the resulting graph for Testimony 100, where only the links above a certain weight are shown. This way of approaching onomastic formulas through the syntactic networks they can generate is inspired from quantitative linguistics (Mehler et al. 2016).

Figure 3.17: *Distance Network Representation of Testimony #100*



Note: All the links between the nodes with a weight above 0.05 are represented, and they are weighed as a function of the formulaic distance: closer relationships are marked by denser lines. The nodes representing entities are in grey, and qualifications in blue. The elements are encoded in latinized Greek capitals.

One may build the large disconnected network generated by all the known onomastic formulas. Looking at all the gods in all testimonies, we can focus on the weighted degree distribution, as shown in Figure 3.18, which represents the strength

with which each occurrence of each god in invocations is related to other gods. This degree distribution does not appear as a straight line and it is therefore presumably not a simple power law, and not generated by a basic preferential attachment mechanism.

Table 3.19 shows that the degree distributions appears as an exponential, with a parameter λ equal to 2.5. We know that the dedications recorded in votive acts, organized as a function of the main divinity to which they were addressed, follow power laws, potentially truncated. It is hence somewhat surprising that the connections made between gods and other gods or qualifiers clearly follow exponential distributions. Indeed, if connections were random between power-law distributed acts towards gods, then the number of degrees would also have the same distribution, following an argument comparable to what we suggested about onomastic networks. Exponential degree distributions can emerge when nodes appear and get attached randomly among the existing nodes. In that case, Table 3.19 may indicate that invocations are expressed by merging in some new elements in the existing mix, without necessarily following a pre-existing pattern. In other words, invocations are in essence innovative, more than a simple reshuffling of known qualifiers and gods associations. In a broader context, this could be taken as a symptomatic feature of ancient polytheist religion, lacking the kind of dogma known in monotheist religions, and allowing worshipers to make official through an inscription any message they wanted.

One important question, potentially related to the network structure between gods and qualifiers, is whether the “language” used in these invocations possesses characteristics of a natural language. In the applications of economics and complexity to linguistics, many detailed analyses can indeed be carried out on grammar, in particular when reflecting the network structure of syntactic relationships (Čech, Mačutek, & Liu 2016). We will restrict ourselves here to a simple comparison, looking at the distribution of terms, for the sake of simplicity. Elements, which correspond to common names or proper names, are the natural equivalent of lemmas in POS tagging. Figure 3.20 plots the occurrence distribution for literary texts from the Diorisis corpus, and for the onomastic formulae in MAP, and they appear quite close to each other. In fact, distribution fits tell us they both follow power laws with the same parameter, to the second decimal. In spite of a very formulaic nature, the term distribution of divine invocations resembles that of natural language. While this may not hold if one were to look more closely

Figure 3.18: Degree Distribution of Divinities in Distance-Weighted Formulae Networks

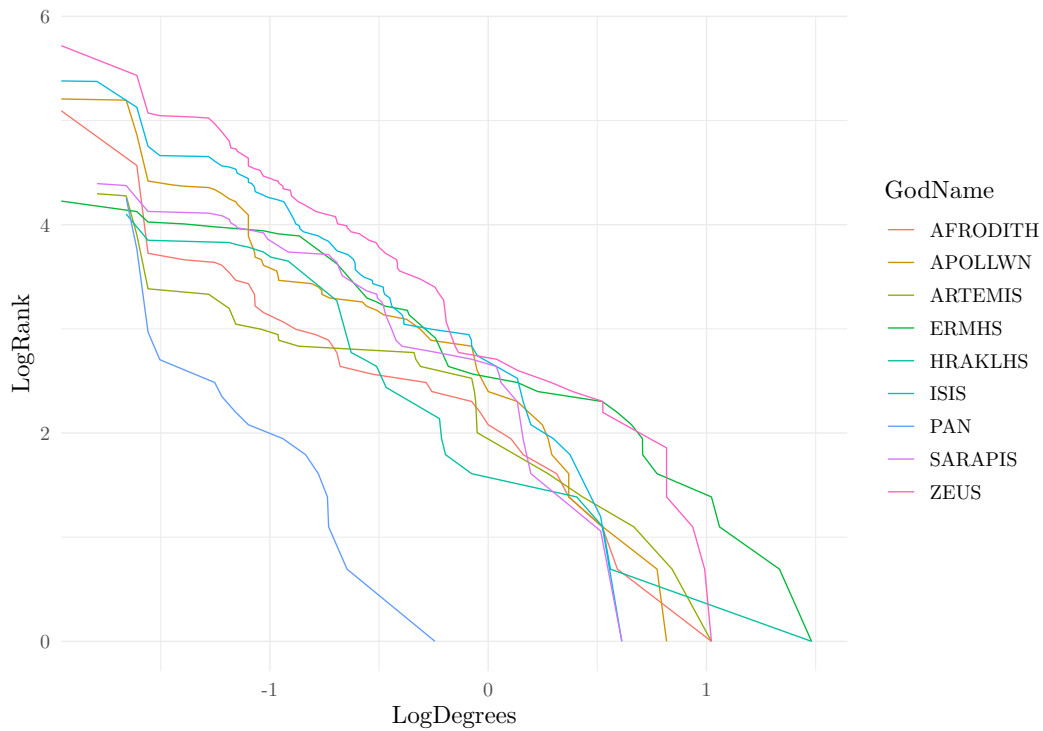
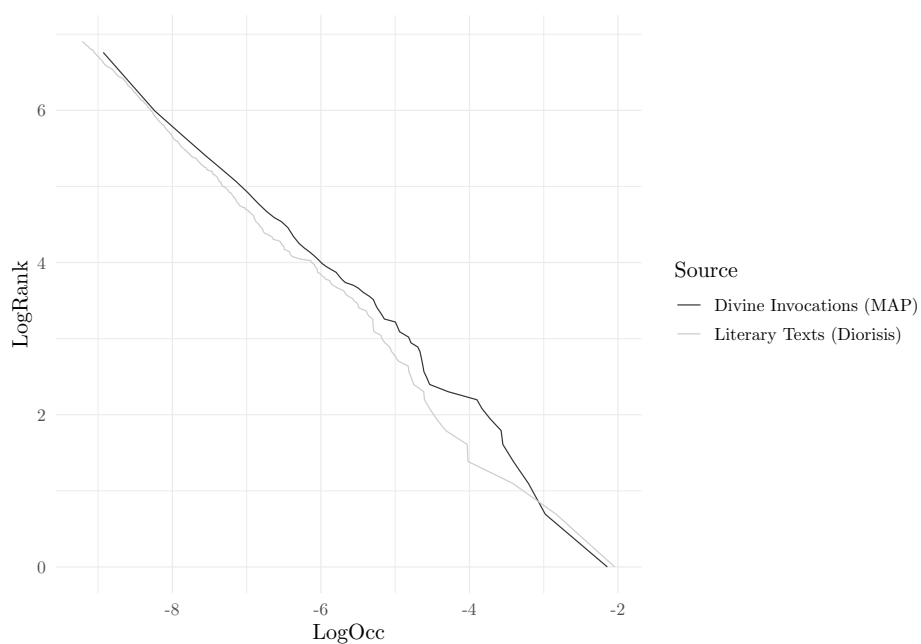


Table 3.19: Summary Statistics on Degree Distribution Fit

| Statistic | MAP |
|----------------|----------|
| Lambda Exp | 2.413 |
| Alpha Pow | 0.173 |
| Alpha Trunc | 0.000 |
| Lambda Trunc | 0.323 |
| Trunc vs Pow R | 337.738 |
| Trunc vs Pow p | 0.000 |
| Trunc vs Exp R | -70.928 |
| Trunc vs Exp p | 0.000 |
| Pow vs Exp R | -104.786 |
| Pow vs Exp p | 0.000 |

at the syntactic structure of both languages, it is consistent with a notion of random terms being merged into divine formulae, as we discussed above, since these terms would be pulled from the natural language. Further, as Ferrer i Cancho, Riordan, & Bollobás (2005) have shown, when a communication language exhibits Zipf's law, one can theoretically expect the emergence of syntactic patterns. Hence, the inclusion of qualifiers in divine formulae, as random as it may be, would not be inconsistent with the appearance of the logical syntax-like structure which we have observed.

Figure 3.20: *Log/Log Cumulative Distribution of Diorisis and MAP Corpora*



Note: The data includes the most common lemmas for each corpus, which account for more than 0.01% of occurrences. The horizontal axis is the logarithm of normalized frequency of each lemma, and the vertical axis is the logarithm of the lemma's rank.

The networks we have analyzed here represent very different things: how the ancient Greeks named their children on one hand, and how they addressed the gods in the dedications they inscribed on stone. In spite of these fundamental differences, the methodology for analyzing them remains the same, and involves a series of particular metrics, although we only considered degree distributions. These networks turned out to have fairly different characteristics, for which we offered some explanations, but fully developed theoretical models would be necessary to truly account for their formation.

3.4 Conclusion

Cliometrics have not been much concerned with ancient history, simply because there is little appropriate economic data from the period that one may analyze; in fact, whether there was an economy is debatable. Ancient history is indeed particular in the sense that there is little clean data of the form usually exploited in cliometrics. There are no clean time series of anything for the ancient times, maybe barring geological or astronomical information. Instead, there are fairly large volumes of unstructured data. This data is unstructured not only because it gathers texts and very diverse categorical information, but also because historians, who created all these information sources, tend to exploit it in this form.

On the other hand, one can apply economic principles to the analysis of potentially every aspect of human life in the ancient times, and that is what we have started to do here for ancient Greece. Cliometrics and complexity have extended the range of traditional cliometrics through a different approach to data, and we have followed in their path. By relying on complex systems analysis, we can develop different hermeneutics of data, and do not need them to be structured. This has allowed us to operate on the same primary sources that historians use, albeit with a different perspective.

The behavioral economics explanations, based on game theory or optimization arguments, that we have provided, effectively open new ways of inquiry into ancient Greece, which we will expand on in the following chapters, through Part II. Considering how genre affected texts, how costs and information were reflected in inscriptions, how worship could self-organize, how names could be picked in a fundamentally different manner than what we know, and how the language spoken to the gods was

built according to specific principles, we have not looked at economics through history, but we have instead looked at history through economics. This change of perspective, with significant help from complexity sciences, should largely open the distant past to cliometrics.

3.5 Chapter Bibliography

- Abry, Patrice, Cécile Bastidon, Pierre Borgnat, Pablo Jensen, & Antoine Parent.** 2022. "The Contours of a Cliophysics: How Can Econophysics Enrich Cliometrics? Case Studies in Debt Issues and Global Capital Markets." *Frontiers in Physics* 10: 894400.
- Alfarano, Simone, & Thomas Lux.** 2010. "Extreme Value Theory as a Theoretical Background for Power Law Behavior." Working Paper 1648. Kiel Working Papers. Kiel Institute for the World Economy. <https://ideas.repec.org/p/kiel/kieliw/1648.html>.
- Alstott, Jeff, Ed Bullmore, & Dietmar Plenz.** 2014. "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." *PLOS ONE* 9 (1): e85777.
- Aron, Raymond.** 1981. "Quelques remarques sur la compréhension et l'explication." *Revue européenne des sciences sociales* 19 (54-55): 71–82.
- Assael, Yannis, Thea Sommerschield, & Jonathan Prag.** 2019. "Restoring Ancient Text Using Deep Learning: A Case Study on Greek Epigraphy." In *Conference on Empirical Methods in Natural Language Processing*, edited by Jing Jiang, Vincent Ng, & Xiaojun Wan, 6369–6376. Stroudsburg: Association for Computational Linguistics.
- . 2020. "Sommerschield/Ancient-Text-Restoration." June 6, 2020. <https://github.com/sommerschield/ancient-text-restoration>.
- Baek, Seung Ki, Sebastian Bernhardsson, & Petter Minnhagen.** 2011. "Zipf's Law Unzipped." *New Journal of Physics* 13 (4): 043004.
- Baek, Seung Ki, Hoang Anh Tuan Kiet, & Beom Jun Kim.** 2007. "Family Name Distributions: Master Equation Approach." *Physical Review E* 76 (4): 046113.
- Baltas, George.** 2003. "Determinants of Internet Advertising Effectiveness: An Empirical Study." *International Journal of Market Research* 45 (4): 1–9.
- Barabási, Albert-László, & Réka Albert.** 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–512.
- Bastidon, Cécile, Michael Bordo, Antoine Parent, & Marc Weidenmier.** 2019. "Towards an Unstable Hook: The Evolution of Stock Market Integration Since 1913." Working Paper w26166. Cambridge: National Bureau of Economic Research. <https://doi.org/10.3386/w26166>.
- Bastidon, Cécile, & Antoine Parent.** 2022. "Cliometrics of World Stock Markets Evolving Networks." *Annals of Operations Research* To appear.
- Bastidon, Cécile, Antoine Parent, Pablo Jensen, Patrice Abry, & Pierre Borgnat.** 2020. "Graph-Based Era Segmentation of International Financial Integration." *Physica A: Statistical Mechanics and Its Applications* 539: 122877.
- Boccara, Nino.** 2010. *Modeling Complex Systems*. Graduate Texts in Physics. New York: Springer.
- Bonnet, Corinne.** 2017. "Mapping Ancient Polytheisms." 2017. <https://map-polytheisms.humanum.fr/>.
- Bonnet, Corinne, & Nicole Belayche, eds.** 2017. *Puissances divines à l'épreuve du comparatisme : constructions, variations et réseaux relationnels*. Bibliothèque de l'École des Hautes Études, sciences religieuses, volume 175. Turnhout: Brepols.

- Bonnet, Corinne, & Sylvain Lebreton.** 2019. "Mettre les polythéismes en formules ? À propos de la base de données Mapping Ancient Polytheisms." *Kernos* 32: 267–296.
- Borwein, David, Jonathan M. Borwein, & Pierre Maréchal.** 2000. "Surprise Maximization." *The American Mathematical Monthly* 107 (6): 517–527.
- Bresson, Alain.** 1981. "Règles de nomination dans la Rhodes antique." *Dialogues d'histoire ancienne* 7 (1): 345–362.
- Brulé, Pierre.** 1998. "Le langage des épiclèses dans le polythéisme hellénique (l'exemple de quelques divinités féminines). Quelques pistes de recherches." *Kernos* 11 (11): 13–34.
- . 2005. "Le polythéisme en transformation : les listes de dieux dans les serments internationaux en Grèce antique (V^e-II^e siècles)." In *Nommer les dieux : Théonymes, épithètes, épiclèses dans l'Antiquité*, edited by Nicole Belayche, Pierre Brulé, Gérard Freyburger, Yves Lehman, Laurent Pernot, & Antoine Prost, 143–173. *Recherches sur les rhétoriques religieuses*. Turnhout: Brepols.
- Brulé, Pierre, & Sylvain Lebreton.** 2007. "La Banque de données sur les épiclèses divines (BDDE) du Crescam : sa philosophie." *Kernos* 20: 217–228.
- Bulbulia, Joseph.** 2010. "Charismatic Signalling." *Journal for the Study of Religion, Nature and Culture* 3 (4): 518–551.
- Bulbulia, Joseph, & Marcus Frean.** 2010. "The Evolution of Charismatic Cultures." *Method & Theory in the Study of Religion* 22 (4): 254–271.
- Čech, Radek, Ján Mačutek, & Haitao Liu.** 2016. "Syntactic Complex Networks and Their Applications." In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, edited by Alexander Mehler, Andy Lücking, Sven Banisch, Philippe Blanchard, & Barbara Job, 167–186. *Understanding Complex Systems*. Berlin: Springer.
- Clauset, Aaron, Cosma Rohilla Shalizi, & M. E. J. Newman.** 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.
- Crane, Gregory R.** 2012. "Perseus Digital Library." April 2012. <http://www.perseus.tufts.edu>.
- Crettez, Bertrand, & Régis Deloche.** 2018. "An Analytic Narrative of Caesar's Death: Suicide or Not? That Is the Question." *Rationality and Society* 30 (3): 332–349.
- Deng, Weibing, Wei Li, Xu Cai, & Qiuping A. Wang.** 2011. "The Exponential Degree Distribution in Complex Networks: Non-Equilibrium Network Theory, Numerical Simulation and Empirical Data." *Physica A: Statistical Mechanics and Its Applications* 390 (8): 1481–1485.
- Devine, Andrew M., & Laurence Stephens.** 1976. "The Homeric Hexameter and a Basic Principle of Metrical Theory." *Classical Philology* 71 (2): 141–163.
- Diebolt, Claude, & Michael Hauptert, eds.** 2019. *Handbook of Cliometrics*. 2nd ed. Springer Reference. Cham: Springer International Publishing.
- . 2020. "How Cliometrics Has Infiltrated Economics – and Helped to Improve the Discipline." *Annals of the Fondazione Luigi Einaudi* 54 (1): 219–230.
- Elson, David, Nicholas Dames, & Kathleen McKeown.** 2010. "Extracting Social Networks from Literary Fiction." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 138–147. Uppsala: Association for Computational Linguistics.

- Ely, Jeffrey, Alexander Frankel, & Emir Kamenica.** 2015. "Suspense and Surprise." *Journal of Political Economy* 123 (1): 215–260.
- Fenoaltea, Stefano.** 2019. "Spleen: The Failures of the Cliometric School." *Annals of the Fondazione Luigi Einaudi* 53 (2): 5–24.
- Ferrer i Cancho, Ramon, Oliver Riordan, & Béla Bollobás.** 2005. "The Consequences of Zipf's Law for Syntax and Symbolic Reference." *Proceedings of the Royal Society B: Biological Sciences* 272 (1562): 561–565.
- Ferrer i Cancho, Ramon, & Ricard V. Solé.** 2003. "Least Effort and the Origins of Scaling in Human Language." *Proceedings of the National Academy of Sciences of the United States of America* 100 (3): 788–791.
- Fleck, Robert K., & F. Andrew Hanssen.** 2006. "The Origins of Democracy: A Model with Application to Ancient Greece." *The Journal of Law and Economics* 49 (1): 115–146.
- . 2013. "How Tyranny Paved the Way to Democracy: The Democratic Transition in Ancient Greece." *The Journal of Law and Economics* 56 (2): 389–416.
- . 2018. "What Can Data Drawn from the Hansen-Nielsen Inventory Tell Us about Political Transitions in Ancient Greece?" In *Ancient Greek History and Contemporary Social Science*, edited by Mirko Canevaro, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, 213–238. Edinburgh: Edinburgh University Press.
- Gabaix, Xavier.** 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics* 114 (3): 739–767.
- . 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30 (1): 185–206.
- Gonzales-Feliu, Jesus, & Antoine Parent.** 2016. "Clio-Combinatorics: A Novel Framework to Analyze Military Logistics Choices Using Operations Research Techniques." In *Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations*, edited by Alberto Ochoa-Zezzatti, Jöns Sánchez, Miguel Gastón Cedillo-Campos, & Margain de Lourdes, 79–101. Advances in Logistics, Operations, and Management Science. Hershey: IGI Global.
- Grabska-Gradzińska, Iwona, Andrzej Kulig, Jarosław Kwapięń, & Stanisław Drożdż.** 2012. "Complex Network Analysis of Literary and Scientific Texts." *International Journal of Modern Physics C* 23 (07): 1250051/1–15.
- Hansen, Mogens Herman, & Thomas Heine Nielsen.** 2004. *An Inventory of Archaic and Classical Poleis*. Oxford: Oxford University Press.
- Harrison, John, & Ju Yeong Kim.** 2020. "RSelenium: R Bindings for 'Selenium WebDriver'." February 3, 2020. <https://CRAN.R-project.org/package=RSelenium>.
- Henrich, Joseph.** 2009. "The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution." *Evolution and Human Behavior* 30 (4): 244–260.
- Hobson, Matthew S.** 2014. "A Historiography of the Study of the Roman Economy: Economic Growth, Development, and Neoliberalism." *Theoretical Roman Archaeology Journal* 2013: 11–26.
- Iannaccone, Laurence R.** 1998. "Introduction to the Economics of Religion." *Journal of Economic Literature* 36 (3): 1465–1495.

- Ide, Nancy.** 2004. "Preparation and Analysis of Linguistic Corpora." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, & John Unsworth, 289–305. Malden: Blackwell Publishing.
- Jackson, Matthew O., & Brian W. Rogers.** 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review* 97 (3): 890–915.
- Johnson, Kyle P., Patrick J. Burns, John Stewart, & Todd Cook.** 2019. "The Classical Language Toolkit." January 2019. <http://cltk.org>.
- Johnson, Tim, & Josiah Ober.** 2014. "POLIS." 2014. <http://polis.stanford.edu>.
- Karila-Cohen, Karine.** 2016. "Prosopographia Attica 2.0 : base de données et raisonnement prosopographique." *Revue historique* 680 (4): 869–904.
- . 2017. "Quand les familles s'affichent sur les pierres (à partir de l'exemple athénien)." *Pallas Hors Série*: 91–117.
- . 2018. "Le graphe, la trace et les fragments : l'apport des méthodes quantitatives et des outils numériques à l'étude des élites civiques athéniennes." *Annales HSS* 73 (4): 785–815.
- . 2019. "La valeur du lien : relations extérieures et capital social des notables athéniens de la basse époque hellénistique." In *La cité interconnectée dans le monde gréco-romain (IV^e siècle a.C.-IV^e siècle p.C.) : transferts et réseaux institutionnels, religieux et culturels aux époques hellénistique et impériale*, edited by Madalina Dana & Ivana Savalli-Lestrade, 287–306. Scripta antiqua 118. Pessac: Ausonius Éditions.
- Karila-Cohen, Karine, Claire Lemerrier, Isabelle Rosé, & Claire Zalc.** 2018. "Nouvelles cuisines de l'histoire quantitative." *Annales HSS* 73 (4): 773–783.
- Kenna, Ralph, Máirín MacCarron, & Pádraig MacCarron, eds.** 2017. *Maths Meets Myths: Quantitative Approaches to Ancient Narratives*. Understanding Complex Systems. Cham: Springer International Publishing.
- Kohli, Rajeev, & Raaj Sah.** 2006. "Some Empirical Regularities in Market Shares." *Management Science* 52 (11): 1792–1798.
- Kydos, Dimitrios, Panagiotis Notopoulos, & Georgios Exarchos.** 2015. "Homer's Iliad – A Social Network Analytic Approach." *International Journal of Humanities and Arts Computing* 9 (1): 115–132.
- Le Riche, Antoine, Antoine Parent, & Lei Zhang.** 2022. "Introduction." In *Institutional Change and China Capitalism: Frontier of Cliometrics and Its Application to China*, edited by Antoine Le Riche, Antoine Parent, & Lei Zhang, xxvii–xliv. Transformations in Banking, Finance and Regulation 2. Hackensack: World Scientific.
- Lebreton, Sylvain.** 2019. "Des dieux et des personnes : une approche par les réseaux (mondes grecs et ouest-sémitiques, ca. 1000 av. n. è. – 400 d. n. è.)." Working Paper. Cinquième rencontre du groupe Res-Hist. Rennes: Université de Rennes. <https://reshist.hypotheses.org/1438>.
- Lebreton, Sylvain, Jean-Baptiste Barreau, Karine Karila-Cohen, & Pierre Brulé.** 2014. "Banque de Données des Épicleses Grecques (BDEG)." 2014. <https://epiclesesgrecques.univ-rennes1.fr/?lang=en>.
- Lemerrier, Claire.** 2012. "Formale Methoden der Netzwerkanalyse in den Geschichtswissenschaften: Warum und Wie?" *Österreichische Zeitschrift für Geschichtswissenschaften* 23 (1): 16–41.
- Li, Wentian.** 2012. "Analyses of Baby Name Popularity Distribution in U.S. For the Last 131 Years." *Complexity* 18 (1): 44–50.

- Limpert, Eckhard, Werner A. Stahel, & Markus Abbt.** 2001. "Log-Normal Distributions Across the Sciences: Keys and Clues." *BioScience* 51 (5): 341.
- Mandelbrot, Benoît.** 1953. "An Informational Theory of the Statistical Structure of Languages." In *Communication Theory*, edited by Willis Jackson, 486–502. Woburn: Butterworth.
- Marrou, Henri-Irénée.** (1954) 2016. *De la connaissance historique*. Histoire H21. Paris: Seuil.
- Maruvka, Yosi E., Nadav M. Shnerb, & David A. Kessler.** 2010. "Universal Features of Surname Distribution in a Subsample of a Growing Population." *Journal of Theoretical Biology* 262 (2): 245–256.
- McLean, Bradley H.** 2002. *An Introduction to Greek Epigraphy of the Hellenistic and Roman Periods from Alexander the Great down to the Reign of Constantine (323 B.C.-A.D. 337)*. Ann Arbor: University of Michigan Press.
- Mehler, Alexander, Andy Lücking, Sven Banisch, Philippe Blanchard, & Barbara Job, eds.** 2016. *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Understanding Complex Systems. Berlin: Springer.
- Mitzenmacher, Michael.** 2004. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2): 226–251.
- Mongin, Philippe.** 2018. "A Game-Theoretic Analysis of the Waterloo Campaign and Some Comments on the Analytic Narrative Project." *Cliometrica* 12 (3): 451–480.
- Ober, Josiah.** 2015. *The Rise and Fall of Classical Greece*. Princeton: Princeton University Press.
- Pantelia, Maria.** 2020. "Thesaurus Linguae Graecae." 2020. <http://stephanus.tlg.uci.edu>.
- Parker, Robert, Jean-Baptiste Yon, & Mark Depauw.** 1996. "Lexikon of Greek Personal Names." 1996. <http://www.lgpn.ox.ac.uk>.
- "**PHI Greek Inscriptions**". n.d. <https://inscriptions.packhum.org/>.
- Piantadosi, Steven T.** 2014. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychonomic Bulletin & Review* 21 (5): 1112–1130.
- "**R: The R Project for Statistical Computing**". 2021. The R Project for Statistical Computing. January 2021. <https://www.r-project.org/>.
- Reed, William J., & Barry D. Hughes.** 2003. "Power-Law Distributions from Exponential Processes: An Explanation for the Occurrence of Long-Tailed Distributions in Biology and Elsewhere." *Scientiae Mathematicae Japonicae* 8: 329–339.
- Rochat, Yannick.** 2014. "Character Networks and Centrality." PhD thesis, Lausanne: Université de Lausanne. <http://infoscience.epfl.ch/record/203889>.
- Rydberg-Cox, Jeff.** 2011. "Social Networks and the Language of Greek Tragedy." *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1 (3): 1–11.
- Saichev, Alex, Yannick Malevergne, & Didier Sornette.** 2010. *Theory of Zipf's Law and Beyond*. Lecture Notes in Economics and Mathematical Systems 632. Berlin: Springer.
- Saïd, Suzanne, & Monique Trédé.** (1990) 1999. *A Short History of Greek Literature*. Translated by Trista Selous. London: Routledge.
- Tauber, James.** 2020. "Working with the Diorisis Ancient Greek Corpus." J. K. Tauber. January 20, 2020. <https://jktauber.com/2020/01/20/working-with-the-diorisis-ancient-greek-corpus/>.

-
- . 2021. "Jtauber/Diorisis." Exploring the Diorisis Ancient Greek Corpus. 2021. <https://github.com/jtauber/diorisis>.
- Vatri, Alessandro, & Barbara McGillivray.** 2018a. "The Diorisis Ancient Greek Corpus." May 2, 2018. https://figshare.com/articles/dataset/The_Diorisis_Ancient_Greek_Corpus/6187256.
- . 2018b. "The Diorisis Ancient Greek Corpus: Linguistics and Literature." *Research Data Journal for the Humanities and Social Sciences* 3 (1): 55–65.
- Vernant, Jean-Pierre, & Pierre Vidal-Naquet.** (1972) 2001. *Mythe et tragédie en Grèce ancienne*. Vol. 1. 2 vols. Paris: La Découverte.
- Waumans, Michaël C., Thibaut Nicodème, & Hugues Bersini.** 2015. "Topology Analysis of Social Networks Extracted from Literature." *PLOS ONE* 10 (6): e0126470.
- Zanette, Damián H., & Susanna C. Manrubia.** 2001. "Vertical Transmission of Culture and the Distribution of Family Names." *Physica A: Statistical Mechanics and Its Applications*, Proceedings of the IUPAP International Conference on New Trends in the Fractal Aspects of Complex Systems, 295 (1): 1–8.
- Zipf, George K.** 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge: Addison-Wesley.

Part II

Cliometrics and Complexity Applications to Ancient Greek Society

Chapter 4

The Cliometrics of Name Choice in Ancient Greece

Some ancient Greek names have acquired worldwide fame, and are nowadays still given to children. Achilles, Hektor, Demosthenes, or Alexandros are firmly inscribed in European history and literature. These famous names were often given to children already in ancient times, and there are hundreds of epigraphic sources referring to some Achileus or Demosthenes across *Hellas*. Many other names, scarcer, were also given, of which we have only found a handful of instances in the epigraphic or literary record. As we have briefly mentioned in Chapter 3, in classical Greece, there were no family names as we know them today, but single names, sometimes associated to a political group (such as a deme in Athens), and as a result, for all matters and purposes, some people had the same name as many others, while others did not. Why was there such variability in the nature and effective function of the ancient Greeks' names? If unique names were helpful in identifying people, why did not everyone carry a different name, as is generally the case in the contemporary world? To understand how names were given, we will need to explore how they were used.

Onomastics, the study of names, and prosopography, the study of individual life histories are important facets of ancient Greek history. Names and individual identification has been used to more precisely follow the history of members of the elite, as Puech (2012), Karila-Cohen (2017) and Karila-Cohen (2019) have done. Studying names and how they were attributed has also allowed historians to better grasp the extent of families, as in Bresson (1981), Herman (1990) and Bresson (2019). Using

names from epigraphic sources may also help in measuring the reliability of some historical accounts, as Hornblower (2010) proposed to do in the case of Thucydides. As we indicated in Chapter 3, there has been a very substantial effort with the LGPN to collect, and later digitize, data from hundreds of thousands of inscriptions relating to names in the ancient Greek world. Thanks to the availability of this data, some research has focused on the network aspect of the family relationships based on the names in the inscriptions, such as Cline (2020), Karila-Cohen (2016) and Karila-Cohen (2018). Nevertheless, the historical approach has tended to focus on particular areas, or on particular families, rather than consider names at a much larger scale²⁸. Historical approaches have not either looked specifically into the reasons why some names may have been more common than others.

Our goal in this chapter is to apply an economic approach to the understanding of ancient Greek names, in order to address both the scale of data, as well as the logic of certain distributional features of this data. We will hence focus on accounting for the distribution of names, as well as naming strategies. More precise uses of names could be studied, in particular by focusing on their semantics²⁹, but we focus on the first step, analyzing these ancient Greek names in bulk, which has not been done to date, either from a theoretical nor empirical perspective.

We may wonder what economics could have to say about something as un-economic sounding as a name. As a matter of fact, the notion of identity in economics spans a wide range of related concepts. Berg (2019) gives an overview, and for our purposes we can determine that there are three connected axes in this body of research with applications or concepts that are useful for onomastics: group behavior (in the sense of club goods), identity *stricto sensu*, and conformity. Carr & Landa (1983) considered the benefits of belonging to a group in reducing transaction costs through the enforcement of certain rules, and imposing costs on breaching them. They showed with a simple model how one could account for clans, symbols or family and name as club goods. Depending on various conditions on the cost of informing other group members and

²⁸As we indicated in Chapter 3, obtaining the LGPN data in bulk is quite difficult, and this may explain the scarcity of research that could rely on the entire dataset.

²⁹Addressing methodological issues in the study of names, Motschenbacher (2020) has called for the study of names in the context of their use, that is from a linguistics perspective, more than simply studying their frequencies and differences.

the cost of contract breaches, they established an equilibrium for the optimal club size. While the way in which the groups are defined is related to notions of identity, they did not specifically look into defining identity from an economic standpoint. A precise definition of identity was given by Akerlof & Kranton (2000), who proposed to extend a basic utility function to account for their notion of identity: people benefit more from partaking in activities that are in line with the prescriptions of their chosen identity. They derived various equilibria, where varying shares of the population choose a particular identity. Not following the prescriptions of one's identity is costly to oneself, because, as they argue, rules are internalized psychologically, and breaching them creates anxiety. Breaching identity-prescribed rules is also costly to others: by observing someone's breaking the rules it arouses emotions that were suppressed in order to internalize the rules in the first place. Their model explains behaviors that could otherwise appear anti-economic. The notions of identity and group belonging are also strongly connected with conformity, which was studied by Bernheim (1994) and extended and formally clarified by Gillen (2015). In this approach, the utility functions of agents of various types include a term for status, which at equilibrium is itself a function of type. In certain conditions, when status plays an important role in utility, then there is an endogenous pooling equilibrium where many agents take the same action, while only those with extreme preferences have incentives to stray (and receive a lower status). Some experiments have been carried out and showed that conformity affected the outcome of public goods games, as reported by Carpenter (2004).

Names have also been extensively studied in statistical physics. Several publications following physical sciences-inspired methods have indeed examined the distribution of names, whether they be family names or given names. As with many social phenomena, the patterns of family name distributions match power laws. Zanette & Manrubia (2001) showed that these distributions exhibited an exponent of 1 (at the density level), and proposed a simple random growth model accounting for this behavior. Baek, Kiet, & Kim (2007) and Maruvka, Shnerb, & Kessler (2010) applied population dynamics models in order to account for the growth in number and distribution of family names. The first paper shows that the shape of family names distributions (generally power laws) can be explained by the way the rate of introduction of new names is expressed. Some other studies carried out purely empirical analyses of some specific aspects of

names distributions, such as dynamics or correlation, in particular Hahn & Bentley (2003), Mateos, Longley, & O'Sullivan (2011), Barucca et al. (2015) and Lee et al. (2016), who looked at given names. Following more of a psychological perspective, Busse & Seraydarian (1977), Young et al. (1993), and Rogerson (2016) focused on some of the implied identity in names. For example, the experiments conducted by Young et al. (1993) showed that people of a younger generation tended to associate first names from an older generation to lesser intelligence and popularity.

In order to tackle the analysis of ancient Greek names, we will first examine the historical context in which names were used at the time. We then develop a model for the optimal behavior that people will adopt with each other as a function of the unique information they have about their counterpart: their name. Considering different information frameworks, we justify the exchange of names in this context as the best system in order to promote cooperation. Then, we will be in a position, relying on that model, to study the optimal choice of names for children, given that equilibrium. With some simplifications, we show that the impact of strategic naming on the distribution of names works as an alteration of the typical mean-field approaches to name dynamics, which converge to a power law. Strategic naming adds a degree of freedom in the relationship between the observed number of names and the shape of the power law distribution. Finally, we confront these results with empirical data from the archaic and classical periods. We show that a form of conformist strategic naming could account for the particular shape of the names' power law distribution in Ancient Greece, which has a parameter around 1.3. This shape is in stark contrast with what has been observed in the large majority of empirical studies on contemporary name distributions, where the parameter was either close to 2 for most countries, or at 1 for Korea.

4.1 Ancient Greek Names in Context

We begin by addressing the issue of the effect of child mortality on the surviving literary and archaeological traces of names, and then discuss the manner in which names were used and exchanged in ancient Greece.

One potential issue in understanding name choices in ancient Greece is indeed whether *necronymy*, giving a child the name of a prior child who had passed, was common

or not. The very high child mortality compared to modern standards, depending on the practice of necronymy, could contribute to a randomization of name choices. The epigraphic sources almost exclusively refer to adults or adolescents, and as a result the names we get to observe are not an unbiased sample of those that were given, because of child mortality. There are historical and cliometric methodological issues in determining the shape of life tables applicable to ancient history, as Woods (2007) summarized. While there is no large amount of direct data that can be exploited, certain extrapolations across time and cultures allow for the linking of life expectancy with fairly stable life table shapes. In summary, one can estimate death rates of around 30% in the first year, and around 50% cumulatively for the first 5 years in ancient Greece (Woods 2007, 379–380.). Parkin (2013) finds numerous references in classical literature to the dangers of early childhood, and also points out that early infant mortality could reach up to 50% in high stress periods³⁰. Such high rates of mortality could naturally interact with name choices. There is no ancient testimony about the practice of necronymy, and whether this practice was common or not would substantially affect the distribution of observed names, since the choices of names that we observe on inscriptions about adults would become randomized. However, according to Hardie (1923), necronymy was common in certain conditions in Macedonia in the early 20th century, and she infers this was most likely inherited from ancient times. In Bresson (1981), comparable observations were made, also pointing towards a fairly systematic use of necronymy. As a result, it is reasonable to assume that despite child mortality, the names of adolescents or adults that we are able to observe today reflect the parents' initial choices, and not a randomized selection.

The extensive analysis of contemporary names we have mentioned earlier may not be relevant for ancient Greece. In addition to the use of a single anthroponym without a patronym in the modern sense, the choice of a name was also strongly affected by tradition and perceived prestige. For example, *paponymy*, the use of a grandparent's name for a child, was frequent. "Celebrity" names were sometimes selected; for example Broux & Clarysse (2016) examined the use of Ptolemaic queen names, such as Arsinoe, Kleopatra or Berenike, across large geographic areas. Honor and names are also closely related. By giving one's name in ancient Greece, a person would identify themselves,

³⁰See Parkin (2013), p. 47-48.

and providing this information was a meaningful act. The importance of a name can be shown to be deeply anchored in Greek literature; one good example may be Odysseus's introduction to Alcinous in Homer's *Odyssey* 9:

“But now I will tell you my name, so that you too may know it, and then when this pityless day is over I may be your host, although my home is far. I am Odysseus, son of Laertes, known among all men for my cunning, and my fame reaches heaven³¹.”

The importance of the name transpires from these verses, but they depict an important encounter between two kings. In contrast, when first meeting someone with whom one is not very well acquainted, names may not be fully given. Odysseus's most famous reply to the Cyclops's question, also in *Odyssey* 9, illustrates this rather well. Telling one's name is an important act, and precedes any form of interaction, and one is supposed to do it in good faith (which is not the case with the Cyclops):

“My name is Nobody; and my mother, my father and all my companions also call me Nobody³²”

While certain people may be known by others by reputation, it was not necessarily the case that they knew who the person was. The famous example of Aristides the Just makes this clear: As Plutarch reported³³, he was once confronted in the street by an illiterate man, asking him to write “Aristides” on an *ostrakon*³⁴ (Aristides obliged). Whether the story is true is not critically important, but it appears clearly that it was conceivable that a man could know of a name, without being able to recognize the person.

We can consider that the name of a person in classical Greece was used as an important piece of information in dealing with others. As the first step in a relationship, the exchange of names should indeed communicate a certain reputation in some cases, or a degree of prestige, and potentially some clues about the person's origin or family.

³¹νῦν δ' ὄνομα πρῶτον μυθήσομαι, ὄφρα καὶ ὑμεῖς / εἶδετ', ἐγὼ δ' ἂν ἔπειτα φυγὼν ὑπο νηλεὲς ἦμαρ / ὑμῖν ξεῖνος ἔω καὶ ἀπόπροθι δώματα ναίων. / εἴμ' Ὀδυσσεὺς Λαερτιάδης, ὃς πᾶσι δόλοισιν / ἀνθρώποισι μέλω, καὶ μευ κλέος οὐρανὸν ἴκει. Homer, *Odyssey*, 9, 16-20.

³²Οὔτις ἐμοί γ' ὄνομα: Οὔτιν δέ με κικλήσκουσι / μήτηρ ἠδὲ πατήρ ἠδ' ἄλλοι πάντες ἐταῖροι. Homer, *Odyssey*, 9, 366-367.

³³Plutarch, *Vitae Parallelae Aristides*, 7.

³⁴A pottery fragment used in an ostracism vote, in order to ban a person from Athens.

Cuesta et al. (2015) have shown with behavioral experiments on series of repeated games that reputation acquired in practice played a significant role in human cooperation. When no particular history of interactions is available, the name may serve as a proxy. The prestige and the reputation are then attached to the names, not to the persons themselves, and reflect the behavior of those who share that name. This prestige and perceived honor are hence a function of how the holders of the name behave with others, whether they act fairly or not in general. Therefore, if we can account for the use of names in interactions, then we may account for the ones that are the most beneficial or the most honorable, and, as a consequence, for the ones that would preferably given to children as a function of the existing name distribution. Giving the name of someone one wants to honor to a child could also be accounted for if more people bearing the name in question within the family increases some measure of honorability for that name.

In order to build a model appropriate to the context of ancient Greece, we combine features of several of the economic approaches we have mentioned earlier. We consider groups of people, which we call families, in the sense of clans or clienteles, rather than a small-size household. In our logic there should be a reasonably limited number of such families in a *polis*. Along with Carr & Landa (1983), we can assume that members of a same family need to behave well with each other, as they would be identified and punished otherwise, while when dealing with people who are not family members there would be more leeway for cheating. The interactions in which one might exchange names are those defined as *personalistic exchanges* by Carr & Landa (1983).

Although we do not need to specifically define identity (as a function of the name or of the family), some of the results from Akerlof & Kranton (2000) may be derived at equilibrium due to optimal behavior of the whole population: people may behave in a certain way with those with whom they share the same name, which is an identity-charged feature. Finally, the notion of conformity as defined and exploited in Bernheim (1994) does not either required to be formally expressed in our approach, but we will see that a conformist behavior for name giving rises up in certain conditions.

4.2 Modeling the Use of Names

We assume there is a continuous population of individuals $x \in \mathcal{P}$, and to each individual is associated a single family $\phi_x \in \mathcal{F}$, a finite set of cardinal I . We will consider a probability measure μ over \mathcal{P} that gives the relative frequency of groups of people. We will write Φ_f for the set of individuals with family f , so that $\mu(\Phi_f)$ is properly defined. We further assume that for any element $x \in \mathcal{P}$, $\mu(\{x\}) = 0$.

In our modeling, we will use independent random variables X and Y following a uniform distribution on \mathcal{P} , and define a probability \mathbb{P} , so that for all μ -measurable set $A \subset \mathcal{P}$, $\mathbb{P}[X \in A] = \mu(A)$. One simple way of representing the family sizes is with a series of positive values $(p_{f_i})_{1 \leq i \leq I}$ such that $\sum_i p_{f_i} = 1$ and for each element $f_i \in \mathcal{F}$, $\mu(\Phi_{f_i}) = p_{f_i}$.

All the people interact in pairs in the course of their life, and when they do, they can choose to act fairly, or to cheat. Incentives are in place to try and deter people from acting unfairly within the family: when they do, they receive a punishment (some reduction in utility, or some reparation that needs to be paid). When someone acts unfairly with respect to a member of their own family, they will be found individually. When people act unfairly with others outside of the family, they cannot be easily found and as a result no cost is directly assessed. Although there may be laws that punish crimes in general, the cheating we are describing here should not be a serious offence. As such, we consider that in practice some laws or rules that address the cheating could be enforced within a family, and much less so outside of a family.

When an individual $x \in \mathcal{P}$ interacts with $y \in \mathcal{P}$ and neither cheats, both gain g . If x cheats and y does not, then x gains $g + r$, but there is a cost: if x happens to be from the same family as y , ie $\phi_x = \phi_y$ then the cost is c_w . When both players cheat at the same time they have no gain and no loss. When a player does not cheat, but is cheated, there is a loss of r . We assume that $c_w > r$, so that the punishment always takes away at least as much as the ill-gotten gains. One could think of the punishment as a multiple of the potential gain, because such cheating reflects badly on the entire family. This setting is very simple, but refinements such as different costs schedules attached to participating, or when cheating outside of the family, do not change the overall shape of the results while at the same time obscuring the logic.

We will write the expectation operator \mathbb{E} with respect to \mathbb{P} , and use μ to denote the relative size of population groups. In particular, note that $\mathbb{P}[\phi_X = \phi_Y] = \mathbb{E}[\mu(\Phi_{\phi_X})]$. If we write the event $H_{x,y} \in \{0, 1\}$ when player x cheats on player y , then the gain for x in an interaction with y writes:

$$G_{x,y} = g \overline{H_{x,y}} \overline{H_{y,x}} + (g + r - c_w \mathbb{I}_{\phi_x = \phi_y}) H_{x,y} \overline{H_{y,x}} - r \overline{H_{x,y}} H_{y,x}.$$

The choice that an individual x can make is to cheat or not, which we will write H_x , conditioning on the information they may have (on their own family and name, or on the other person's family and name). The core setting of this problem is comparable to a prisoner's dilemma, where cheating corresponds to defecting, and not cheating to cooperation. Considering mixed strategies, the optimal probability that individual x cheats with individual y will be written $\eta_{x,y}$, and the expected gain for individual x will be denoted by π_x . Both measures depend on the information available to each individual. We will write F_x the information set that individual x can condition upon, with random variables X and Y representing the randomly selected individuals in an encounter.

From a general perspective, we can write the expected gain for player x conditioned by the fact that x is cheating or not, and conditioned by x 's available information:

$$\begin{aligned} \pi_x(H) &= \mathbb{E}[G_{X,Y} \mid \{H_{X,Y} = 1\} \cap F_x] \\ \pi_x(\overline{H}) &= \mathbb{E}[G_{X,Y} \mid \{H_{X,Y} = 0\} \cap F_x]. \end{aligned}$$

At the optimum, when all participants play the mixed strategy, we will have for all $x \in \mathcal{P}$:

$$\mathbb{E}[G_{X,Y} \mid \{H_{X,Y} = 1\} \cap F_x] = \mathbb{E}[G_{X,Y} \mid \{H_{X,Y} = 0\} \cap F_x].$$

As the most realistic setting, in line with what we can understand of relationships in ancient Greece, we will eventually assume that when people meet, they disclose their names but they do not disclose what family they belong to, because either they do not wish to, or they may not be able to prove what they would state. We consider that the name itself is necessarily given truthfully, the potential decision to cheat coming only after one knows the other person's name. As a result, there is no perfect knowledge of a

counterpart's family. Information on the distribution of names in families and on their sizes is known to everyone, on the other hand.

We begin by characterizing several possible equilibria, depending on the information available to the population: successively perfect information, no information (not even on one's family), knowledge of one's family, and finally knowledge of one's family and of the other person's name, which is the most representative case. Considering these various possible conditions will allow us to also consider the aggregate social optimum reached in each case.

4.2.1 Equilibria with Perfect Information

In the simplest case with full information, which we will denote with the superscript a , $F_x = F_y = \{\phi_X = \phi_x, \phi_Y = \phi_y\}$. We have the following result.

Proposition 4.1 (Interactions with Perfect Information). *When individual x encounters individual y with perfect information, the probability that x cheats is 1 if x and y are from a different family, and $1 - \frac{r}{c_w}$ otherwise. The aggregate probability of cheating is*

$$\mathbb{E}[\eta_{X,Y}^a(\phi_X, \phi_Y)] = 1 - \frac{r}{c_w} \sum_{1 \leq i \leq I} p_{f_i}^2,$$

and the aggregate gain is:

$$\mathbb{E}[\pi_X^a(\phi_X, \phi_Y)] = \frac{r}{c_w} (g + r - c_w) \sum_{1 \leq i \leq I} p_{f_i}^2.$$

Proof. The optimal probability of cheating depends on the information known by both, so we simply write:

$$\eta_{x,y}^a(\phi_x, \phi_y) = \mathbb{I}_{\phi_x = \phi_y} \eta_{x,y}^a(\phi_x, \phi_y)|_{\phi_x = \phi_y} + \mathbb{I}_{\phi_x \neq \phi_y} \eta_{x,y}^a(\phi_x, \phi_y)|_{\phi_x \neq \phi_y}.$$

The following normal form representation shows the game when the two participants are from different families. One can see that in this case, there is a pure strategy Nash equilibrium, when both cheat, that is $\eta_{x,y}^a(\phi_x, \phi_y)|_{\phi_x \neq \phi_y} = 1$, and the expected gain is null, $\pi_x^a(\phi_x, \phi_y)|_{\phi_x \neq \phi_y} = 0$.

| | | | |
|------------|-----------|--------------------------------|---------------|
| | | Player $y, \phi_x \neq \phi_y$ | |
| | | H | \bar{H} |
| Player x | H | $(0, 0)$ | $(g + r, -r)$ |
| | \bar{H} | $(-r, g + r)$ | (g, g) |

When people are from the same family, then the interaction can be represented as follows, and there are then two possible Nash equilibria in pure strategies: both cheat, or both cooperate.

| | | | |
|------------|-----------|-----------------------------|---------------------|
| | | Player $y, \phi_x = \phi_y$ | |
| | | H | \bar{H} |
| Player x | H | $(0, 0)$ | $(g + r - c_w, -r)$ |
| | \bar{H} | $(-r, g + r - c_w)$ | (g, g) |

Solving for the mixed strategy optimum in this case, we obtain:

$$\eta_{x,y}^a(\phi_x, \phi_y)|_{\phi_x=\phi_y} = \frac{c_w - r}{c_w} = 1 - \frac{r}{c_w},$$

which is between 0 and 1 with our assumptions.

With this mixed strategy, the expected gain for any player x writes:

$$\pi_x^a(\phi_x, \phi_y)|_{\phi_x=\phi_y} = \frac{r}{c_w}(g - c_w + r),$$

which is positive if $g > c_w - r$.

In these conditions, the overall expected gain from these interactions can be written:

$$\mathbb{E}[\pi_X^a(\phi_X, \phi_Y)] = \frac{r}{c_w}(g + r - c_w)\mathbb{E}[\mu(\Phi_{\phi_X})].$$

The aggregate probability of cheating is

$$\begin{aligned}\mathbb{E}[\eta_{X,Y}^a(\phi_X, \phi_Y)] &= \frac{c_w - r}{c_w} \mathbb{E}[\mu(\Phi_{\phi_X})] + (1 - \mathbb{E}[\mu(\Phi_{\phi_X})]) \\ &= 1 - \mathbb{E}[\mu(\Phi_{\phi_X})] \frac{r}{c_w} \\ &= 1 - \frac{r}{c_w} \sum_{1 \leq i \leq I} p_{f_i}^2,\end{aligned}$$

where we use the fact that: $\mathbb{E}[\mu(\Phi_{\phi_X})] = \sum_{1 \leq i \leq I} p_{f_i}^2$. □

When people from different families meet, the interaction is a straight prisoner's dilemma: the gain from defecting (cheating) is strictly superior to the gain when cooperating. When people from the same family interact, since $c_w > r$, this is not a typical prisoner's dilemma according to that definition, the gain from cheating is less than the gain attached to cooperation. Relying on a large set of empirical studies on prisoner's dilemmas, Mengel (2018) and Gaechter, Lee, & Sefton (2020) have shown that on non-repeated games, cooperation probability is still potentially high, typically in the 20%-60% range, in spite of the optimality of defection. It is however largely driven by the degree of temptation (defined in these papers) which in our case is equal to $\frac{r}{g}$. Assuming that r may be as large as g , this temptation would drive the likelihood of cooperation to very low levels, which would in turn justify the punishment c_w , so that at least in some cases there is high likelihood of cooperation.

With the cost c_w , in the case of interactions within the same family, we can see that as the punishment cost increases, the probability of cheating increases. This may seem paradoxical, but if there is a very large cost of cheating then the more likely choice is for everyone to cheat so that the probability of anyone cheating while the other is not cheating becomes very small. If the punishment cost can be optimally chosen, it should be only slightly larger than the extra gain obtained when cheating a non-cheating counterpart. Systematically cheating is a way of ensuring nobody gets caught and punished.

The aggregate outcome from these games with full information does not perform well socially, and, as we will see, can be improved when information on one's family is kept private.

4.2.2 Equilibria with No Information

At the other extreme, if individuals do not know, or do not factor in as part of their choices, either their own family nor their counterpart's family, the information set verifies $F_x = \emptyset$. We will denote this situation with the superscript \emptyset . The optimal strategy for x , η_x^\emptyset , does not depend on any input and is in fact constant as a function of x .

Proposition 4.2 (Interactions with No Information). *When individual x encounters individual y with no information, the probability that x cheats is constant and equal to:*

$$\eta_x^\emptyset = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \sum_{1 \leq i \leq I} p_{f_i}^2 \right)}.$$

The expected aggregate gain is:

$$\pi_x^\emptyset = g - (g + r) \left(1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \sum_{1 \leq i \leq I} p_{f_i}^2 \right)} \right).$$

Proof. In the expression for player x 's expected gain depending on their strategy, the only conditioning is that by $H_{X,Y} = 1$ and hence:

$$\pi_x^\emptyset(H) = \mathbb{E} [(1 - \eta_Y^\emptyset)(g + r - c_w \mathbb{I}_{\phi_Y = \phi_X})],$$

and

$$\pi_x^\emptyset(\bar{H}) = \mathbb{E} [-r\eta_Y^\emptyset + g(1 - \eta_Y^\emptyset)].$$

Expressing the expected payoffs of this game in normal form, we can see that when $c_w \mathbb{E} [\mu(\Phi_{\phi_Y})] > r$, the dominant strategy is for all players to cooperate.

| | | Player y | |
|------------|-----------|---|---|
| | | H | \bar{H} |
| Player x | H | $(0, 0)$ | $(g + r - c_w \mathbb{E} [\mu(\Phi_{\phi_Y})], -r)$ |
| | \bar{H} | $(-r, g + r - c_w \mathbb{E} [\mu(\Phi_{\phi_Y})])$ | (g, g) |

Equating $\pi_x^\emptyset(H)$ and $\pi_x^\emptyset(\bar{H})$ in the cases where there is no dominant strategy, and solving for η_Y^\emptyset (which does not depend on Y), we obtain:

$$\eta_x^\emptyset = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \mathbb{E}[\mu(\Phi_{\phi_Y})] \right)}.$$

Since $\mathbb{E}[\mu(\Phi_{\phi_Y})] = \sum_{1 \leq i \leq I} p_{f_i}^2$, we can also write:

$$\eta_x^\emptyset = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \sum_{1 \leq i \leq I} p_{f_i}^2 \right)}.$$

The expected profit writes

$$\begin{aligned} \pi_x^\emptyset &= -r\eta_Y^\emptyset + g(1 - \eta_Y^\emptyset) \\ &= g - (g + r) \left(1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \sum_{1 \leq i \leq I} p_{f_i}^2 \right)} \right). \end{aligned}$$

We can see how the aggregate probability of cheating enters the expression for the expected aggregate gain. \square

If $c_w < \frac{r}{\mathbb{E}[\mu(\Phi_{\phi_Y})]}$, then the probability of encountering someone from the same family is low enough that the potentially large cost of cheating a non-cheating counterpart is reduced and does not require to be eliminated by everyone choosing to cheat at the same time, and the optimal strategy is for all to cooperate. If $c_w > rI$, it implies the probability of cheating is strictly positive.

It is easy to see that $\pi_x^a(\phi_x, \phi_y) \leq \pi_x^\emptyset$, since $\sum_{1 \leq i \leq I} p_{f_i}^2 \leq 1$, and this situation with no information is more socially beneficial than when people have full knowledge.

4.2.3 Equilibria with Private Information

We now consider the case when in these interactions each person knows their own family, but not the other person's. The information set for x is therefore $F_x = \{\phi_X = \phi_x\}$, and we will denote this case with the superscript o . We will look for a mixed strategy optimum, expressed as a probability $\eta_x^o(\phi_x)$ of cheating, as a function of the participant x 's own family ϕ_x . The dependency of $\eta_x^o(\phi_x)$ to x is only through ϕ_x . Note that in this case, the distributions from which X and Y are drawn are the same.

Proposition 4.3 (Interactions with Private Information). *When individual x encounters individual y with private information, the probability that x cheats as function of their information set is:*

$$\eta_x^o(\phi_x) = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \mu(\Phi_{\phi_x}) \right)}.$$

The aggregate probability of cheating across all possible interactions is:

$$\mathbb{E}[\eta_Y^o(\phi_Y)] = 1 - \frac{r}{c_w} \sum_{1 \leq i \leq I} \frac{p_{f_i}}{\frac{r}{c_w} \vee p_{f_i}},$$

and the aggregate gain is:

$$\mathbb{E}[\pi_X^o(\phi_X, \phi_Y)] = g - (r + g) \left(1 - \frac{r}{c_w} \sum_{1 \leq i \leq I} \frac{p_{f_i}}{\frac{r}{c_w} \vee p_{f_i}} \right).$$

Proof. The expected gain when cheating, and when the other participant y plays the optimal strategy $\eta_y^o(\phi_y)$ can be written:

$$\pi_x^o(H) = \mathbb{E}[(1 - \eta_Y^o(\phi_Y))(g + r - c_w \mathbb{I}_{\phi_Y = \phi_X}) \mid \phi_X = \phi_x],$$

and the expected gain when not cheating is

$$\begin{aligned} \pi_x^o(\bar{H}) &= \mathbb{E}[-r\eta_Y^o(\phi_Y) + g(1 - \eta_Y^o(\phi_Y)) \mid \phi_X = \phi_x] \\ &= \mathbb{E}[-r\eta_Y^o(\phi_Y) + g(1 - \eta_Y^o(\phi_Y))]. \end{aligned}$$

At the optimum, both strategies must have the same expected return for x , so that $\pi_x^o(\bar{H}) = \pi_x^o(H)$. Noting that

$$\mathbb{E}[\eta_Y^o(\phi_Y) \mathbb{I}_{\phi_Y = \phi_x}] = \eta_x^o(\phi_x) \mu(\Phi_{\phi_x}),$$

we obtain the equation:

$$r = c_w \mu(\Phi_{\phi_x}) (1 - \eta_x^o(\phi_x)),$$

and therefore

$$\eta_x^o(\phi_x) = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \mu(\Phi_{\phi_x}) \right)}.$$

If there are very few people in x 's family, then there is a low probability of encountering people from the same family and needing to systematically cheat in order to optimally avoid the large punishment, and it is optimal to cooperate. Reciprocally, the more people there are in the family, the greater the cheating probability.

We can compute the average optimal probability $\mathbb{E}[\eta_{\phi_Y}]$ by integrating:

$$\begin{aligned}\mathbb{E}[\eta_Y^o(\phi_Y)] &= 1 - \mathbb{E}\left[\frac{r}{c_w \left(\frac{r}{c_w} \vee \mu(\Phi_{\phi_Y})\right)}\right] \\ &= 1 - \frac{r}{c_w} \sum_{1 \leq i \leq I} \frac{p_{f_i}}{\frac{r}{c_w} \vee p_{f_i}}.\end{aligned}$$

We can compute the expected gain across individuals:

$$\begin{aligned}\mathbb{E}[\pi_X^o(\phi_X, \phi_Y)] &= \mathbb{E}[-r\eta_Y^o(\phi_Y) + g(1 - \eta_Y^o(\phi_Y))] \\ &= g - (r + g) \left(1 - \frac{r}{c_w} \sum_{1 \leq i \leq I} \frac{p_{f_i}}{\frac{r}{c_w} \vee p_{f_i}}\right).\end{aligned}$$

Here again, we can directly see how the expected probability of cheating enters the aggregate gain expression. \square

We can recognize a form similar to the one for the gain when all information is known and the players are from the same family. The difference between the expected probability of cheating with private information and with no information at all writes:

$$\mathbb{E}[\eta_Y^o(\phi_Y)] - \mathbb{E}[\eta_Y^{\emptyset}(\phi_Y)] = \frac{r}{c_w \left(\frac{r}{c_w} \vee \sum_{1 \leq i \leq I} p_{f_i}^2\right)} - \frac{r}{c_w} \sum_{1 \leq i \leq I} \frac{p_{f_i}}{\frac{r}{c_w} \vee p_{f_i}}.$$

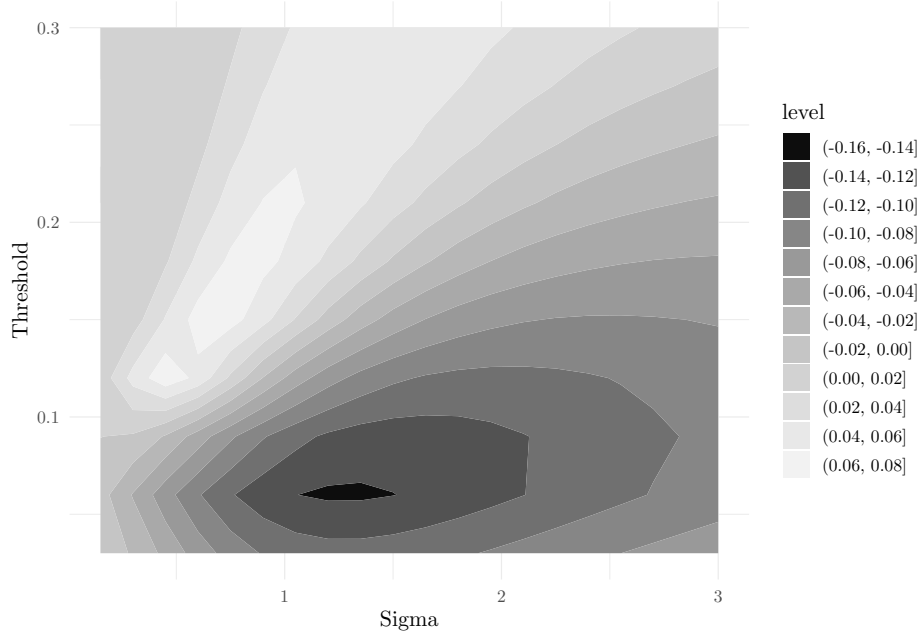
If the family size is regularly spread out and $p_{f_i} = \frac{1}{I}$, then

$$\mathbb{E}[\eta_Y^o(\phi_Y)] = \mathbb{E}[\eta_Y^{\emptyset}(\phi_Y)] = 1 - \frac{r}{c_w} \left(I \vee \frac{c_w}{r}\right).$$

Figure 4.1 shows the difference in the cheating probabilities between the private information case and the case with no information at all across a large range of simulated population distributions. The figure illustrates the fact that there is no systematic

relationship between the two probabilities. When the threshold $\frac{r}{c_w}$ is high, the cheating probability with private information appears generally higher.

Figure 4.1: *Difference in Cheating Probability with Private Information and with No Information*



Note: The count of each family is drawn randomly from a lognormal distribution with dispersion parameter σ . The numbers shown in the chart are the expected values across a large number of simulations. The threshold is $\frac{r}{c_w}$ and $I = 10$.

However, if $\frac{r}{c_w}$ is small and $\frac{r}{c_w} < \frac{1}{I}$, then $\frac{r}{c_w \left(\frac{r}{c_w} \sqrt{\sum_{1 \leq i \leq I} p_{f_i}^2} \right)} = \frac{r}{c_w \sum_{1 \leq i \leq I} p_{f_i}^2}$. In addition, if for all f , $p_f \geq \frac{r}{c_w}$, then $\sum_{1 \leq i \leq I} \frac{p_{f_i}}{\frac{r}{c_w} \sqrt{p_{f_i}}} = I$ and $\mathbb{E}[\eta_Y^o(\phi_Y)] = 1 - \frac{r}{c_w} I$. In these conditions, if $p_f \neq \frac{1}{I}$, then $\frac{r}{c_w \sum_{1 \leq i \leq I} p_{f_i}^2} < I \frac{r}{c_w}$, and $\mathbb{E}[\eta_Y^o(\phi_Y)] < \mathbb{E}[\eta_Y^\emptyset(\phi_Y)]$. When the punishment for cheating is costly relative to its potential gains, then having access to private information reduces the probability of cheating. In effect, if the distribution of people across families is spread out enough relative to $\frac{r}{c_w}$, then having private information is not detrimental to the social good.

4.2.4 Equilibria With Partially Common Information

We now address a generalization of the equilibrium with private information, considering that the distributions from which the participants are drawn are not the same. We assume that for each $x \in \mathcal{P}$, there exists μ_x , a measure giving the relative weight of the

population sharing the same characteristics as x , and associated probability \mathbb{P}_x such that for all measurable $A \subset \mathcal{P}$, $\mathbb{P}_x[X \in A] = \mu_x(A)$. All the μ_x represent the distribution of the population conditioned by some characteristic specific to x , for example x 's name. Naturally, we may have $\mu_x = \mu_z$ for some x and z in \mathcal{P} , for example two individuals with the same name.

We consider the case of interactions when individuals x and y know their own family, but not the other person's, and the measures μ_x and μ_y are known by both. This situation is denoted with the superscript n . The optimal cheating probability can be written $\eta_x^n(\phi_x, \mu_y)$, and it depends on the knowledge of x 's family ϕ_x and on the distributional information on y 's characteristics. In this context, μ_x and μ_y are given, just as μ was a given in the prior cases. The expression $\mu_y(\Phi_{\phi_x})$ captures the probability that x would be in the same family as y knowing x 's characteristics, based on the knowledge of y 's family.

If the μ_x represent names, then these μ_x can be represented by the mass of individuals with a given name in a given family. Set $\nu_x \in \mathcal{N}$, with \mathcal{N} finite, the name of individual x . We also define $p_n = \mu(\{x \in \mathcal{P} : \nu_x = n\}) = \sum_{f \in \mathcal{F}} p_{\nu_x, f}$ and as before, $p_f = \mu(\{x \in \mathcal{P} : \phi_x = f\}) = \mu(\Phi_f)$. Then $\mu_x(\Phi_f) = \frac{p_{\nu_x, f}}{p_{\nu_x}}$. All the $p_{n, f}$ can naturally be expressed as a function of the population ρ in a given family with a given name:

$$p_{n, f} = \frac{\rho_{n, f}}{\sum_{m \in \mathcal{N}, g \in \mathcal{F}} \rho_{m, g}}.$$

Proposition 4.4 (Interactions with Partially Common Information). *When individual x encounters individual y with partially common information, the probability that x cheats as function of their information set (which contains their own family and distributional information on y) is:*

$$\eta_x^n(\phi_x, \mu_y) = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \mu_y(\Phi_{\phi_x}) \right)} = 1 - \left(1 \wedge \frac{r}{c_w} \frac{1}{\mu_y(\Phi_{\phi_x})} \right).$$

In aggregate, the cheating probability across all possible encounters is:

$$\mathbb{E}[\eta_X^n(\phi_X, \mu_Y)] = 1 - \frac{r}{c_w} \sum_{n \in \mathcal{N}} p_n \sum_{f \in \mathcal{F}} \frac{p_f}{c_w \vee \frac{p_{n, f}}{p_n}}.$$

Proof. The expressions for x 's expected gain writes as follows:

$$\begin{aligned}\pi_x^n(H) &= \mathbb{E}_y [(1 - \eta_Y^n(\phi_Y, \mu_x))(g + r - c_w \mathbb{I}_{\phi_Y = \phi_x})] \\ \pi_x^n(\overline{H}) &= \mathbb{E}_y [-r\eta_Y^n(\phi_Y, \mu_x) + g(1 - \eta_Y^n(\phi_Y, \mu_x))].\end{aligned}$$

Swapping x and y , and equating the expected returns $\pi_x^n(H)$ and $\pi_x^n(\overline{H})$, we get:

$$\eta_x^n(\phi_x, \mu_y) = 1 - \frac{r}{c_w \left(\frac{r}{c_w} \vee \mu_y(\Phi_{\phi_x}) \right)}.$$

The probability that x cheats y is therefore increasing as a function of the chances that y would be in the same family as x conditioned on x 's name; this is a comparable effect to what we observed earlier in the probability of cheating.

The expected gain for x hence writes:

$$\mathbb{E}[\pi_x^n] = g - (g + r) \left(1 - \frac{r}{c_w} \mathbb{E} \left[\frac{1}{\frac{r}{c_w} \vee \mu_Y(\Phi_{\phi_x})} \right] \right).$$

The probability of cheating can be integrated relative to \mathbb{P}_x and \mathbb{P}_y to get the aggregate probability of cheating.

$$\mathbb{E}[\eta_X^n(\phi_X, \mu_Y)] = 1 - \mathbb{E} \left[\frac{r}{c_w \left(\frac{r}{c_w} \vee \mu_Y(\Phi_{\phi_X}) \right)} \right].$$

We can write for any measurable function h :

$$\begin{aligned}\mathbb{E}[h(\mu_Y(\Phi_{\phi_X}))] &= \int h(\mu_y(\Phi_f)) \mathbb{P}[Y \in dy] \mathbb{P}[f \in df] \\ &= \sum_{f \in \mathcal{F}} p_f \sum_{n \in \mathcal{N}} p_n h \left(\frac{p_{n,f}}{p_n} \right).\end{aligned}$$

Hence we have:

$$\mathbb{E} \left[\frac{r}{c_w \left(\frac{r}{c_w} \vee \mu_Y(\Phi_{\phi_X}) \right)} \right] = \frac{r}{c_w} \sum_{n \in \mathcal{N}} p_n \sum_{f \in \mathcal{F}} \frac{p_f}{c_w \vee \frac{p_{n,f}}{p_n}},$$

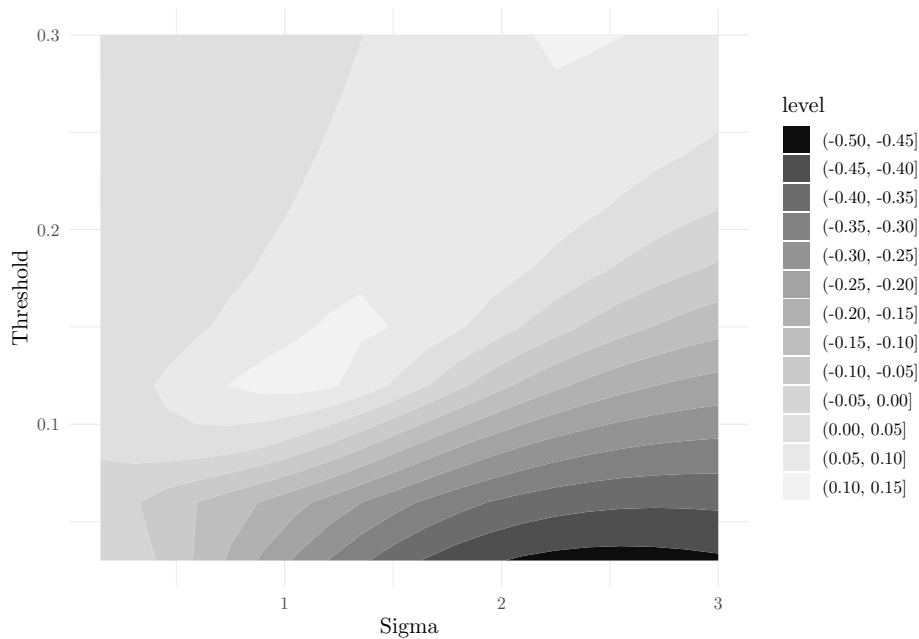
and

$$\mathbb{E}[\eta_X^n(\phi_X, \mu_Y)] = 1 - \frac{r}{c_w} \sum_{n \in \mathcal{N}} p_n \sum_{f \in \mathcal{F}} \frac{p_f}{\frac{r}{c_w} \sqrt{\frac{p_{n,f}}{p_n}}}$$

where we can recognize a form similar to that of $\mathbb{E}[\eta_Y^o(\phi_Y)]$. \square

Figure 4.2 shows the difference in the cheating probabilities between the partially private information case and the case with no information, and Figure 4.3 shows the difference with the private information case. While there appear to be some patterns, there is no systematic relationships between these quantities and no systematically better system. When the threshold $\frac{r}{c_w}$ is low (high cost of being caught), it appears based on Figure 4.2 that the cheating probability with partially common information is lower than with no information. There is a similar effect at play in Figure 4.3.

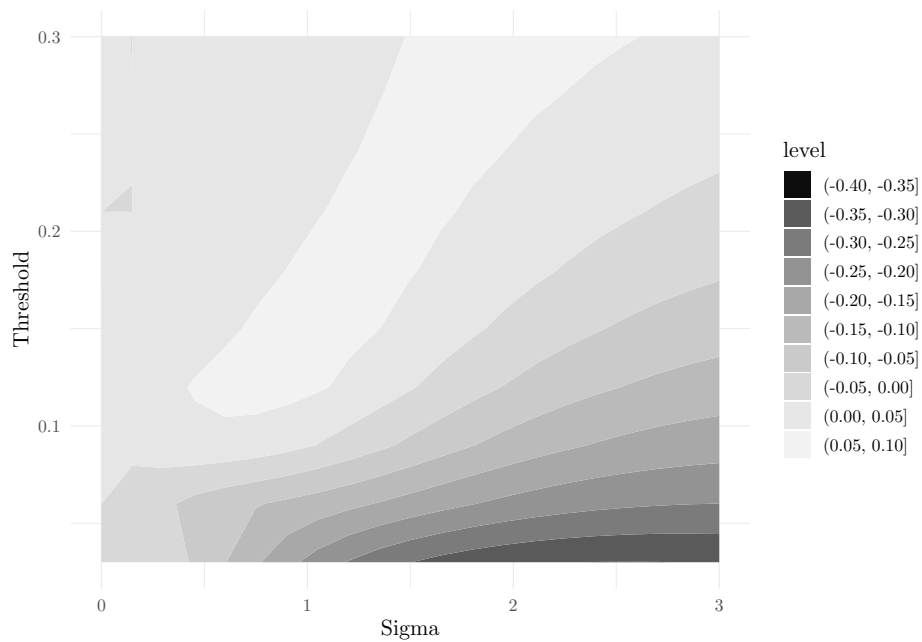
Figure 4.2: *Difference between the Cheating Probability with Partially Common Information and with No Information*



Note: The count of each name in each family is drawn randomly from a lognormal distribution with dispersion parameter σ . The numbers shown in the chart are the expected values across a large number of simulations. The threshold is $\frac{r}{c_w}$, $|\mathcal{N}| = 12$ and $I = 10$.

The expression for $\mathbb{E}[\eta_X^n(\phi_X)]$ shows that if and only if, for all f and n , $\frac{p_{n,f}}{p_n} > \frac{r}{c_w}$, then $\mathbb{E}[\eta_X^n(\phi_X)] = 1 - \frac{r}{c_w} I$. We know that if and only if for all f , $p_f > \frac{r}{c_w}$, then

Figure 4.3: *Difference between the Cheating Probability with Partially Common Information and with Private Information*



Note: The count of each name in each family is drawn randomly from a lognormal distribution with dispersion parameter σ . The numbers shown in the chart are the expected values across a large number of simulations. The threshold is $\frac{r}{c_w}$, $|\mathcal{N}| = 12$ and $I = 10$.

$\mathbb{E}[\eta_X^o(\phi_X)] = 1 - \frac{r}{c_w}I$. The first condition implies the second one, and as a result, if the second condition is verified, $\mathbb{E}[\eta_X^o(\phi_X)] \leq \mathbb{E}[\eta_X^r(\phi_X)]$.

We can see that if the distribution of names is well spread out across families, then this knowledge does not condition much, and as a result the aggregate gain is comparable to the one with only private information or the one with no information at all. In particular, if the punishment c_w is severe enough, then the conditional probabilities are likely greater than $\frac{r}{c_w}$ and the aggregate social gain is the same as if there was no information. The use of names as important information, consistent with the way we represent interactions in ancient Greece, does not lead to a lower social optimum in this case than in the no information case. The exchange of names can hence fulfill a social role, by allowing people to give each other something before entering in a more demanding relationship, which would not be necessary in the three alternate information frameworks we discussed earlier, without a significant social cost relative to these alternatives.

We can see that the names which would appear as the most attractive according to this logic are the ones that are most common outside of the family: one wants to maximize $\frac{\rho_{a_j}}{\rho_{a_j,f}}$.

4.3 The Dynamics of Naming Choices

We have concentrated on understanding a possible function of names, and on some reasons why names could be used in interpersonal exchanges. These names, being given to the children by the parents, had to be determined in some way, and we now examine the way in which they could be chosen within the framework of our simple model. We will consider several drivers of name choices, the strategic one based on the probability of cheating being just one of them.

In light of the historical context we mentioned earlier, we consider that the choice of the names observed in a Greek *polis* may come from three distinct sources: a strategic choice, traditional ponymy, or the random appearance of new names. We will consider that the stock of each name grows through time as a function of these three sources. Since the choice of names is applicable to newborns, it affects a fraction of the population being born over each instant dt . We consider that children belong to a unique family,

which could be that of the mother or the father. Individuals will pick the name with the greatest attractiveness, or of a parent, or from some external source at any point, without having to factor in the behavior of the others, since the change in population is assumed to be very small relative to the total population size.

In order to represent papyonymy within the framework of our model, we will transform it into a simple patronymy, so that to a certain extent, names simply grow in the population at a rate proportional to their current stock. In a continuous time setting, the alternating jump over a generation should not matter. In addition, as we will see, the data does not allow for a detailed prosopographic perspective (tracking each life history), but only an onomastic perspective (tracking the names).

4.3.1 Representing Strategic Name Choices

In the interactions between people according to the model we have developed so far, we have seen that the individual's characteristics affect both the expected gain, which depends on the probability of one being cheated, and the probability of one cheating on others. The greatest the probability that others cheat, the smallest the gain. Hence, the individuals for whom the characteristics μ_x are such that the expected gain is highest are better off. Varying the characteristics of x and y , the probability that x cheats writes $\eta_x^n(\phi_x, \mu_y)$, and depends on x 's family and on y 's name, written as a function of the population distribution ρ :

$$\eta_x^n(\phi_x, \mu_y) = 1 - 1 \wedge \frac{r}{c_w} \frac{\rho_{\nu_y}}{\rho_{\nu_y, \phi_x}}.$$

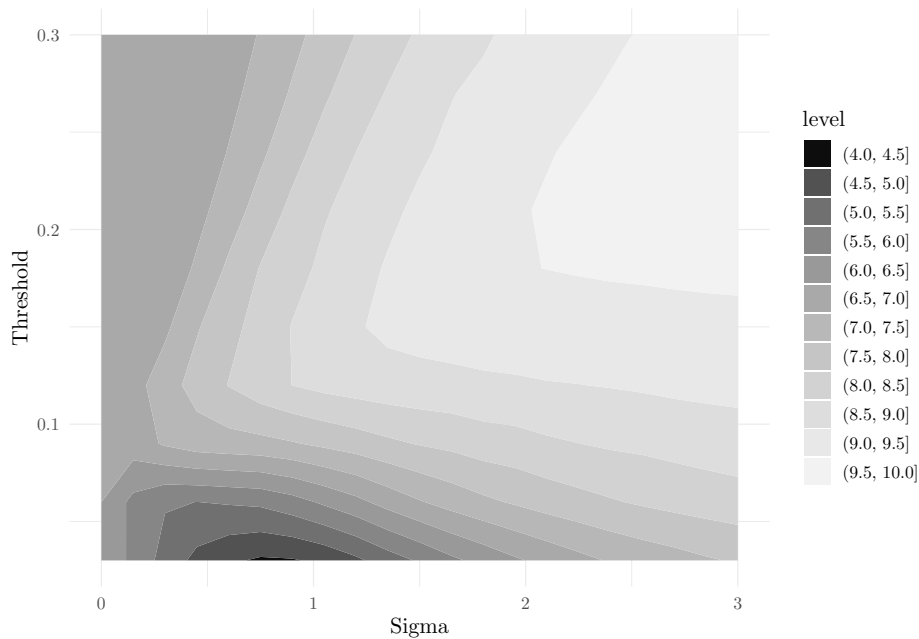
We can consider the desirability of a name to be related to the probability of not cheating, that is the probability that others do not cheat on a given person. The lowest the probability, the higher the expected outcome, and the better off the person. Also, from an honor standpoint, it should presumably be perceived as more honorable to not be cheated, in the cultural context of the ancient Greek *polis*. We can note nevertheless that when c_w is near r , then the choice does not have much of an impact, because the cheating probability is low anyway. If c_w is large, on the other hand, then picking one name or the other could have a significant impact on the amount of cheating the individual does or suffers. The attractivity of a name a may be defined as the probability

that people in general do not cheat on a , taking values in $[\frac{r}{c_w}, 1]$:

$$A(a, \rho) = \sum_f \frac{\rho_f}{\rho} \left(1 \wedge \frac{r}{c_w} \frac{\rho_a}{\rho_{a,f}} \right).$$

The optimal name choice depends on the manner in which all names and families are distributed. In studying the dynamics of name distributions, fully reflecting this complexity makes it impossible to obtain clear results. In order to illustrate some aspects of the optimal name choice, Figure 4.4 shows the average rank (in terms of popularity) of the name that gets picked across a large range of simulated population layouts, following the same logic as in the prior figures. We can see that there are some substantial variations in the nature of the names being picked, depending on each specific situation.

Figure 4.4: Rank by Number of Carriers of the Optimal Name that Minimizes the Probability of Being Cheated



Note: The count of each name in each family is drawn randomly from a lognormal distribution with dispersion parameter σ . The numbers shown in the chart are the expected values across a large number of simulations. The threshold is $\frac{r}{c_w}$, $|\mathcal{N}| = 12$ and $I = 10$.

Proposition 4.5 (Optimal Name Choice Logic). *If names are distributed in the same manner in each family, then the attractivity is constant. If names are concentrated in equally sized*

families, then:

- If the most common name is not very frequent, then it is optimal to chose the least common name;
- If the most common name is frequent, then it is optimal to chose the most common name.

Proof. We will first consider the case when names are distributed in the same way across families. We assume that for all name a , $\frac{\rho_{a,f}}{\rho_f} = \rho_a$. Applying this simplification we get $A(a, \rho) = \sum_f p_f \left(1 \wedge \frac{r}{c_w p_f}\right)$, which does not depend on a .

If names are concentrated in families, this means we can write $p_{a,f} = p_f \mathbb{I}_{f \in F_a}$, where F_a is defined as the subset of families that carry name a . Also, for further simplification, we assume that the families are sized equally, and $p_f = \frac{1}{I}$. Hence, $p_a = \frac{|F_a|}{I}$. We can write:

$$\begin{aligned} A(a, \rho) &= \frac{1}{I} \sum_f \left(1 \wedge \frac{r}{c_w} \frac{I p_a}{\mathbb{I}_{f \in F_a}}\right) \\ &= \frac{1}{I} \sum_f (1 - \mathbb{I}_{f \in F_a}) + \frac{1}{I} \sum_f \mathbb{I}_{f \in F_a} \mathbb{I}_{I p_a \frac{r}{c_w} > 1} + \frac{1}{I} \sum_f \mathbb{I}_{f \in F_a} \frac{r}{c_w} p_a I \mathbb{I}_{I p_a \frac{r}{c_w} < 1} \\ &= 1 + \mathbb{I}_{I p_a \frac{r}{c_w} < 1} \left(p_a^2 I \frac{r}{c_w} - p_a\right). \end{aligned}$$

Hence, if $p_a > \frac{c_w}{2Ir}$, A is increasing as a function of p_a , and decreasing otherwise. Therefore, if the most frequent name's frequency is high enough it is optimal to pick it. If it is under the threshold, it is optimal to pick the least frequent name. If $p_a > \frac{c_w}{Ir}$ however, then the attractivity is maximal, and any name that satisfies this condition may be chosen. \square

The strategy of picking new names could hence appear as either leaning towards innovative fashion, by choosing an initially rare and new name that is picked by everybody, or towards a conformist choice, of picking the most common existing names. At the limit in either case, the informational content from communicating the name would become low.

In the simplified context of the Proposition 4.5, if it is optimal to pick the rarest name, then if new names sometimes appear (being invented or through immigration), then these names would naturally seem optimal, because they would be very rare at

first. Also, as the rarest names are picked, they become less rare until eventually they would catch up with other names. This situation could lead to a uniform distribution of names, where names would then not convey particular information. If on the contrary the most frequent name is picked, then it may become the only existing name, in which case the informational content in the names would also vanish.

More generally, we can see that if the cost c_w is not extreme and there are many families, then the optimal choice would appear to oscillate between the most frequent names, or just any name. Without being able to characterize the optimal choice more precisely, it is reasonable to approximate it by a random choice weighed by name frequency, as a mix between a selection of the most common names and a totally random choice.

4.3.2 A Model for the Distribution of Names

In the ancient Greek population for a *polis*, we consider that the natural population net growth rate is $\lambda - \mu$, where λ is the birth or immigration rate, and μ is the death or emigration rate. When people appear (whether they are born or immigrants), there are three possible manners in which their name is determined:

- The name may be given as a strategic choice according to the model we developed earlier, in which case it will be selected using a particular logic. The rate at which this may happen per unit of time, as a function of the number of people, is γ ;
- The name may be a new name that did not exist before. This could reflect the effect of immigration, or simply some creativity in naming. The rate at which new names may be acquired per unit of time, as a function of the number of people, is β ;
- Otherwise, the name is given using ponymy, or in our context patronymy, at the rate of $\lambda - \beta - \gamma$. We assume that $\lambda > \beta + \gamma$ so that only newly born people get a non-patronymic name.

With this approach, we follow certain aspects of the model from Baek, Kiet, & Kim (2007), however in their model they considered that names were changed after people were born, and did not have strategic name choices. We keep our notations ρ for

the count of people in the population, and N as the total number of names, but now depending on time since we are in a dynamic evolution context. The overall approach is also consistent with the model proposed by Reed & Hughes (2003b), who also relied on transition probabilities and on probability generating functions.

Proposition 4.6 (Evolution of the Number of Names). *The average number of names per people verifies*

$$\frac{N(t)}{\rho(t)} = \frac{\beta}{\lambda - \mu} \frac{e^{(\lambda-\mu)t} - 1}{e^{(\lambda-\mu)t}}$$

and converges to $\frac{\beta}{\lambda-\mu}$ as t goes to infinity.

Proof. The total population is not affected by how people are named, it only depends on the net growth rate α , and

$$\rho(t) = \rho(0)e^{(\lambda-\mu)t}.$$

We define $\pi(s)$ as the rate of names appearing at a point in time s , that is $\pi(s) = \beta\rho(s)$. Hence the total number of names at time t verifies $N(t) = \int_0^t \pi(s)ds$. As a result, we have:

$$\frac{N(t)}{\rho(t)} = \frac{\int_0^t \pi(s)ds}{\rho(0)e^{(\lambda-\mu)t}} = \frac{\int_0^t \beta e^{(\lambda-\mu)s} ds}{\rho(0)e^{(\lambda-\mu)t}} = \frac{\beta}{\lambda - \mu} \frac{e^{(\lambda-\mu)t} - 1}{e^{(\lambda-\mu)t}}.$$

Hence $\lim_{t \rightarrow \infty} \frac{N(t)}{\rho(t)} = \frac{\beta}{\lambda-\mu}$. □

We now define the conditional probability that there are k people carrying the name n at time t , given there were j people carrying it at time s :

$$p_n(j, k, s, t) = \mathbb{P}[\rho_n(t) = k \mid \rho_n(s) = j].$$

We can express the dynamics of the number of people carrying a given name over a short period of time dt as follows. We know that the part of births named by paponymy $\lambda - \beta - \gamma$ can explain an increase in the population carrying the name. The rate β , which refers to new names being created, can only take people away from existing names and does not add them to existing ones. The rate γ , concerning strategic naming, takes people away from a given name, but also redistributes them across all existing names. The overall dynamics for the population, in terms of the number of names or number of

people, need to be consistent with the way transitions in each cohort's population are expressed³⁵.

Within this framework, a cohort with $k + 1$ people could only transition to k if one of them died or emigrated, at the rate μ . Therefore, we should assume that over a short time period dt :

$$\mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = k + 1] = (k + 1)\mu dt.$$

A cohort with $k - 1$ people could transition to k if one person was born and named according to ponymy, or one person was strategically named to the current cohort. The probability for this last event may depend on the distribution of all other names, so for now we will write it as $\gamma(\rho, p, k - 1)$. Hence, we assume that:

$$\mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = k - 1] = (k - 1)(\lambda - \beta - \gamma)dt + \gamma(\rho, p, k - 1)dt.$$

Finally, the probability that a cohort with k people would remain with the same number of people is the probability that none of the possible changes we discussed above takes place. Hence, we would expect that:

$$\mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = k] = 1 - k(\lambda + \mu - \beta - \gamma)dt - \gamma(\rho, p, k - 1)dt.$$

Reflecting the strategic name choice, event with the simplest assumption such as picking a uniform distribution across names, makes the calculation of the limiting behavior impossible. In the case where there is no strategic choice, that is $\gamma = 0$, we can however derive some results following Baek, Kiet, & Kim (2007). Although even in this case our model setup is different from theirs (people are named at birth, do not change name afterwards), we find the same limit distribution.

³⁵The introduction of a non proportional increase in new names in Baek, Kiet, & Kim (2007) does not appear to be consistent with the transition dynamics they assume.

Proposition 4.7 (Transition Probability Expression). *For a small enough increment dt , the transition probability p_n verifies:*

$$\begin{aligned} \frac{dp_n(j, k, s, t)}{dt} &= -k(\lambda + \mu - \beta)p_n(j, k, s, t) \\ &\quad + (k - 1)(\lambda - \beta)p_n(j, k - 1, s, t) \\ &\quad + (k + 1)\mu p_n(j, k + 1, s, t). \end{aligned}$$

Proof. Considering an arbitrarily small time increment $dt > 0$, we have:

$$\begin{aligned} p_n(j, k, s, t + dt) &= \mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(s) = j] \\ &= \mathbb{E}[\mathbb{I}_{\rho_n(t+dt)=k} \mid \rho_n(s) = j] \\ &= \mathbb{E}[\mathbb{E}[\mathbb{I}_{\rho_n(t+dt)=k} \mid \rho_n(t)] \mid \rho_n(s) = j] \\ &= \sum_l \mathbb{P}[\rho_n(t) = l \mid \rho_n(s) = j] \mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = l \cap \rho_n(s) = j]. \end{aligned}$$

However, since dt is small, we will assume that ρ_n cannot change by more than 1, that is, for all k and l :

$$\mathbb{I}_{|k-l|>1} \mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = l \cap \rho_n(s) = j] = 0.$$

Hence, we have

$$\begin{aligned} p_n(j, k, s, t + dt) &= p_n(j, k, s, t) \mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = k \cap \rho_n(s) = j] \\ &\quad + p_n(j, k - 1, s, t) \mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = k - 1 \cap \rho_n(s) = j] \\ &\quad + p_n(j, k + 1, s, t) \mathbb{P}[\rho_n(t + dt) = k \mid \rho_n(t) = k + 1 \cap \rho_n(s) = j]. \end{aligned}$$

The particular expressions for the law of $\rho_n(t + dt)$ in our assumptions do not depend on $\rho_n(s)$. Replacing in the expression, we can write:

$$\begin{aligned} p_n(j, k, s, t + dt) - p_n(j, k, s, t) &= -k(\lambda + \mu - \beta)p_n(j, k, s, t)dt \\ &\quad + (k - 1)(\lambda - \beta)p_n(j, k - 1, s, t)dt \\ &\quad + (k + 1)\mu p_n(j, k + 1, s, t)dt. \end{aligned}$$

This gives us the statement in the proposition. □

Focusing on the case when $j = 1$, that is when we are conditioning by the point when there was only one person in the cohort, we want to calculate $p_n(1, k, s, t)$. In the case where $\gamma = 0$, this probability does not explicitly depend on n , as it is the same expression for all names.

4.3.3 Name Distribution Limiting Behavior

We are now interested in the behavior of the name distribution as t becomes large. We write $P_n(t, k)$ the probability that there are k people with name n at time t , without any conditioning. Also, we write A_n the time when a name n appeared. If we are not treating the names differently, as is the case when $\gamma = 0$, we do not need to keep the subscripts. We have $\mathbb{P}[A \in ds | A \leq t] = \frac{\pi(s)}{\int_0^t \pi(s) ds}$, since π tracks the rate of occurrence of new names. Therefore, conditioning the transition probabilities by the time of apparition of the name, we can write

$$P(t, k) = \int_0^t p(s, t, k) \mathbb{P}[A \in ds | A \leq t] = \frac{\int_0^t p(s, t, k) \pi(s) ds}{\int_0^t \pi(s) ds}.$$

We have the following result, which, being identical to that from Baek, Kiet, & Kim (2007) shows that the differences in the model set up did not have an influence.

Proposition 4.8 (Limiting Distribution without Strategic Naming). *If $\gamma = 0$, the limiting distribution when t goes to infinity P_∞ verifies $P_\infty(k) \sim k^{2 + \frac{\beta}{\lambda - \beta - \mu}}$.*

Proof. Using the explicit form for π , we have:

$$P(t, k) = \frac{\lambda - \mu}{1 - e^{-(\lambda - \mu)t}} \int_0^t e^{-(\lambda - \mu)(t-s)} p(s, t, k) ds.$$

The probability generating function Ψ for $z \in [0, 1)$ is $\Psi(t, z) = \sum_{k \geq 0} z^k p(k, s, t)$, where the dependency on s is implicit.

Note that $z \frac{\partial \Psi}{\partial z} = \sum_{k \geq 0} k z^k p(k, s, t)$, $\frac{\partial \Psi}{\partial z} = \sum_{k \geq 0} (k + 1) z^k p(k + 1, s, t)$ and $z^2 \frac{\partial \Psi}{\partial z} = \sum_{k \geq 0} (k - 1) z^k p(k - 1, s, t)$. Multiplying both sides of the equation in Proposition 4.7 by z^k and summing, we obtain after some simplifications, and writing $\nu = \lambda - \beta$:

$$\frac{\partial \Psi}{\partial t} = (\nu z^2 - (\nu + \mu) z + \mu) \frac{\partial \Psi}{\partial z}.$$

This first-order PDE has the boundary condition that at time $s \leq t$, there should be a single person in the cohort, so that $\Psi(s, z) = z$ since $p_n(1, 1, s, s) = 1$. This equation is similar to the one solved in Baek, Kiet, & Kim (2007), although the variables are not the same: what they wrote as $\lambda(z - 1)(z - \frac{\mu + \beta}{\lambda})$ corresponds here to $\nu(z - 1)(z - \frac{\mu}{\nu})$, where we have defined ν as $\nu = \lambda - \beta$. Their solution for the equation is nevertheless valid and their approximation as well, so that

$$\lim_{t \rightarrow \infty} P(t, k) \sim k^{-(1 + \frac{\lambda - \mu}{\nu - \mu})},$$

which gives the result. □

This result implies that with a proportional growth, the name distribution converges towards a power law of parameter $1 + \frac{\beta}{\lambda - \beta - \mu}$. Further, if we write the limit $r = \lim_{t \rightarrow \infty} \frac{N(t)}{\rho(t)} = \frac{\beta}{\lambda - \mu}$, then the power law to which the distribution converges is fully conditioned by r and has parameter $\frac{1}{1 - r}$. In this model's framework, when there is no strategic name choice, the distribution of names is strictly related to the ratio of the number of names relative to the number of people in the population.

The results derived so far do not account for any strategy in naming. As we mentioned earlier, a general term $\gamma(\rho, p, k - 1)$ that could be introduced to represent the strategic choices would make the model intractable, if it effectively depends on these parameters. However, we can account for these strategic choices in a simpler manner. Indeed, based on Proposition 4.5, we argued that, depending on conditions such as the punishment cost c_w and the number of families I , strategic naming could be captured through either a choice of the rarest name, or population-weighted random choice of names. We will then define γ as the intensity of strategic choices: choices of the least common name if $\gamma > 0$, and choice of the most common names (population-weighted) if $\gamma < 0$. We then have the following:

Proposition 4.9 (Limiting Distribution with Strategic Naming). *If $\gamma \neq 0$, the limiting distribution when t goes to infinity P_∞ verifies $P_\infty(k) \sim k^{2 + \frac{\beta + \gamma}{\lambda - \beta - \gamma - \mu}}$. However, the ratio $\frac{N(t)}{\rho(t)}$ still converges to $\frac{\beta}{\lambda - \mu}$.*

Proof. The optimal strategy of choosing the least common name can in fact be simply represented: at any point in time, given the expected continuous increase in the

population across all names, the least common name will be the latest invented name that has just been created. And if people strategically name their babies accordingly, the population with that name will increase, making the next new name the least common name. In a mean-field context, strategically choosing a name according to our model simply adds to the rate of name invention: by adding more people to the population carrying a name that was just invented, it makes the probability that this population increases by a given amount more likely.

If $\gamma < 0$, which represents the choice of the most common names, then in a mean-field context it is in fact not changing the distribution of the names with a positive population count, because the names that are redistributed are done so according to each name's share in the total population.

Hence, the parameter γ , whether positive or negative, which specifies the rate of strategic name choices, simply behaves like β : either it adds to it, reflecting the choice of brand new names, or it subtracts from it, reflecting the choice of existing names. The essential difference between the two is that γ does not add or take away new names, and therefore does not increase or decrease the name count: $\pi(s)$ does not depend on γ . In the proof for Proposition 4.9, the expression for $P(t, k)$ does not depend on β through the explicit form of $\pi(s)$, because it is cancelled out in the conditional probability $\mathbb{P}[A \in ds | A \leq t]$. As a result, replacing β with $\beta' = \beta + \gamma$ in Proposition 4.9 accounts for the impact of strategic name choices. Separately, β is not changed into β' in the expression of $pi(s)$, and the limiting behavior of $\frac{N(t)}{\rho(t)}$ is hence unchanged. \square

Thanks to this result, we can now see that there is an additional degree of freedom in the relationship between the number of names per people, and the limiting behavior of the name distribution. Recalling that $r = \lim_{t \rightarrow \infty} \frac{N(t)}{\rho(t)}$, the name distribution therefore follows a power law of parameter

$$\frac{\lambda - \mu}{\lambda - \mu - \beta - \gamma} = \frac{1}{1 - r - \frac{\gamma}{\lambda - \mu}},$$

where γ may be positive or negative depending on characteristics of the strategic name choice.

4.4 Data Analysis

In this section we examine the empirical evidence concerning names in Ancient Greece, and confront it to the models we have discussed so far. We use the data from the LGPN³⁶, which we addressed in Chapter 3. However, since the LGPN covers more than a millennium with very significant political and social evolution, it seems relevant to restrict it to the older archaic and classical periods, in order to retain a certain cultural unity, so we only retain observations anterior to 350 BC. Further, we know that the LGPN tracked names of both men and women, the latest representing only a fraction. In order to stay more strictly within the confines of our model, we hence only keep men names in our sample.

4.4.1 Number of Name Observations

As Baek, Kiet, & Kim (2007) reported, one typically observes a relationship between group size and the number of names observed in the group. The relationship is logarithmic in some cases (for contemporary China and Korea), or algebraic with an exponent close to 1 (for many other countries), where using our notations $N = \rho^b$ with $b \in [0, 1]$.

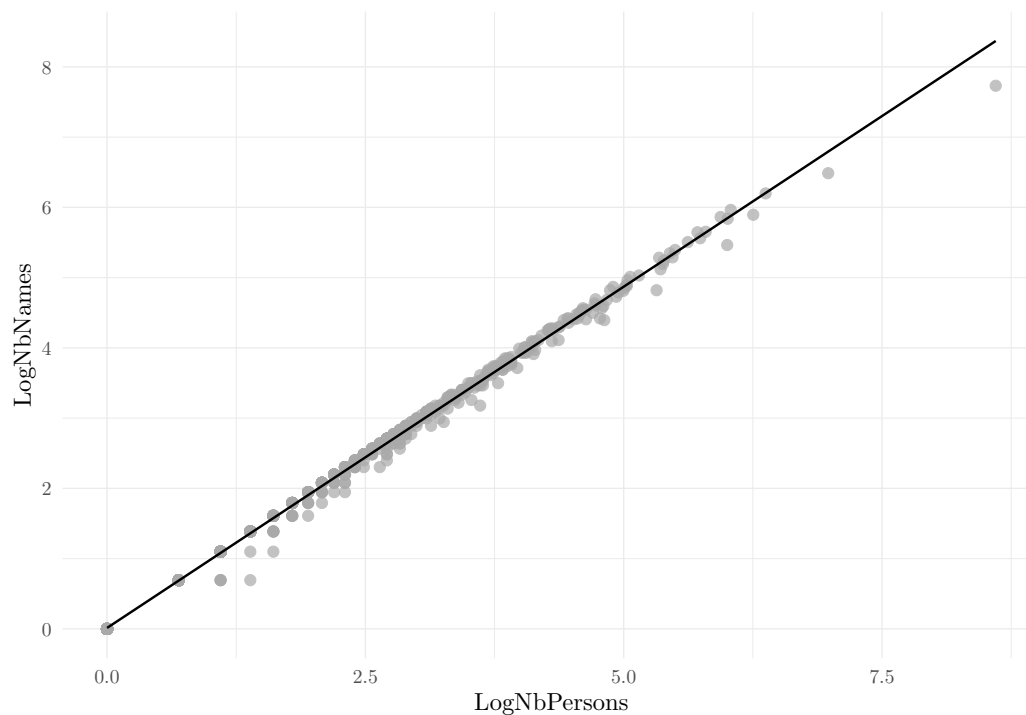
Figure 4.5 shows on a logarithmic scale the relationship between the number of names and the number of people across *poleis*. We can see that it is quite strong. On the logarithmic scale, the slope in the data is 0.97; on the linear scale it is 0.46. Our simple model was predicting a simple linear relationship; the presence of this exponent slightly below 1 could be explained by a reduction in the attractiveness on large places, so that the immigration or name innovation rate tapers slightly off as population increases. In any case, the ratio on the linear scale tells us that the immigration / name innovation rate amounts to roughly half of the population growth.

4.4.2 Empirical Name Distributions

Our formal model led us to expect a particular distribution shape, with particular parameters. Figure 4.6 shows the log/log plot of name frequency distributions across a large set of *poleis*. These distributions for the various places look quite similar, although

³⁶See Parker, Yon, & Depauw (1996).

Figure 4.5: *Log/log Plot of the Number of Distinct Men Names as a Function of Total Number of Mentions Across Poleis*

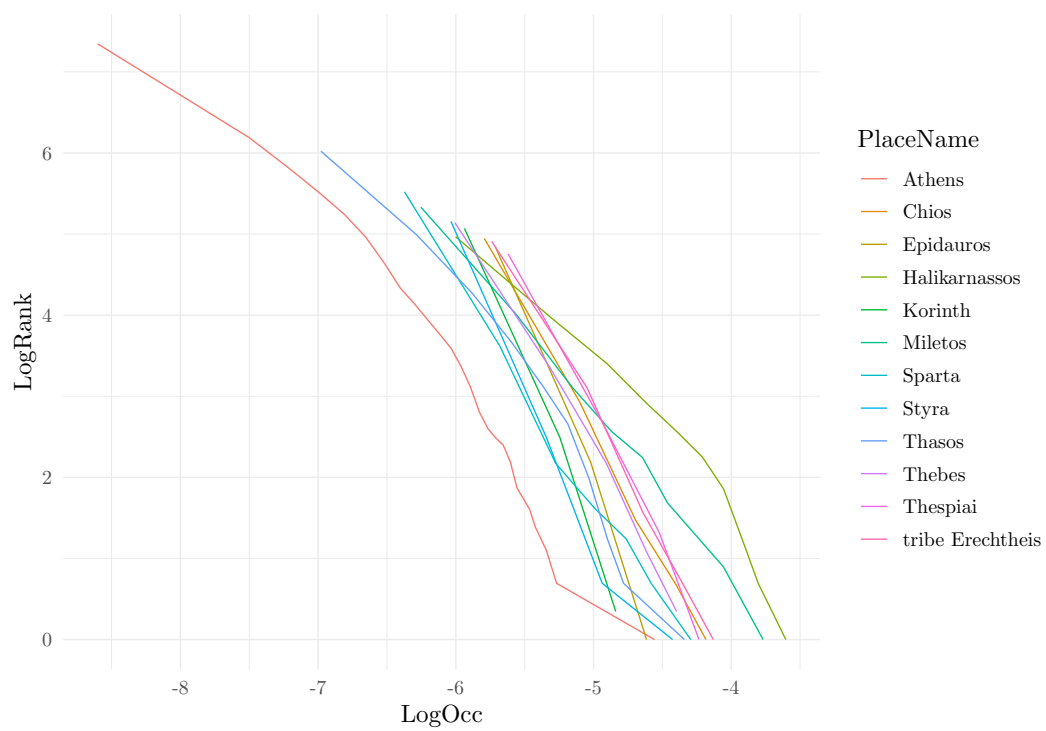


they are somewhat shifted due to the differences in total count. We can see that they appear to follow power laws, but have a bend that could reflect an exponential distribution as well.

We can test for the various distributions that may explain these shapes, following the maximum-likelihood approach from Clauset, Shalizi, & Newman (2009). Table 4.7 shows the distribution fits for all the locations, carried out using the implementation from Alstott, Bullmore, & Plenz (2014). In aggregate, these distributions are power laws or truncated power laws. At the individual *polis* level, they look more like truncated power laws or exponential distributions, but the dispersion of parameters for the power law fits is much less dispersed than for these other types of distributions.

The power law coefficients, around 1.3, are quite different from what has been observed on contemporary data. As Baek, Kiet, & Kim (2007) reported, for most large cities or countries in the world, the coefficients for the name distribution power law tends to cluster around 2. Their model accounted for this pattern, and for the relationship with the name ratio $r = \frac{\beta}{\lambda - \mu}$. In a few particular cases, such as Korea, the power law has a parameter of 1, and the number of names grows logarithmically as a function of total population. One study, focused on rural Sicily, found coefficients that were closer

Figure 4.6: *Log/log Plot of Name Distributions Across Poleis for Men*



Note: The data includes all locations with more than 2000 observations.

Table 4.7: Summary Statistics on Distribution Fits Across Places and in Aggregate

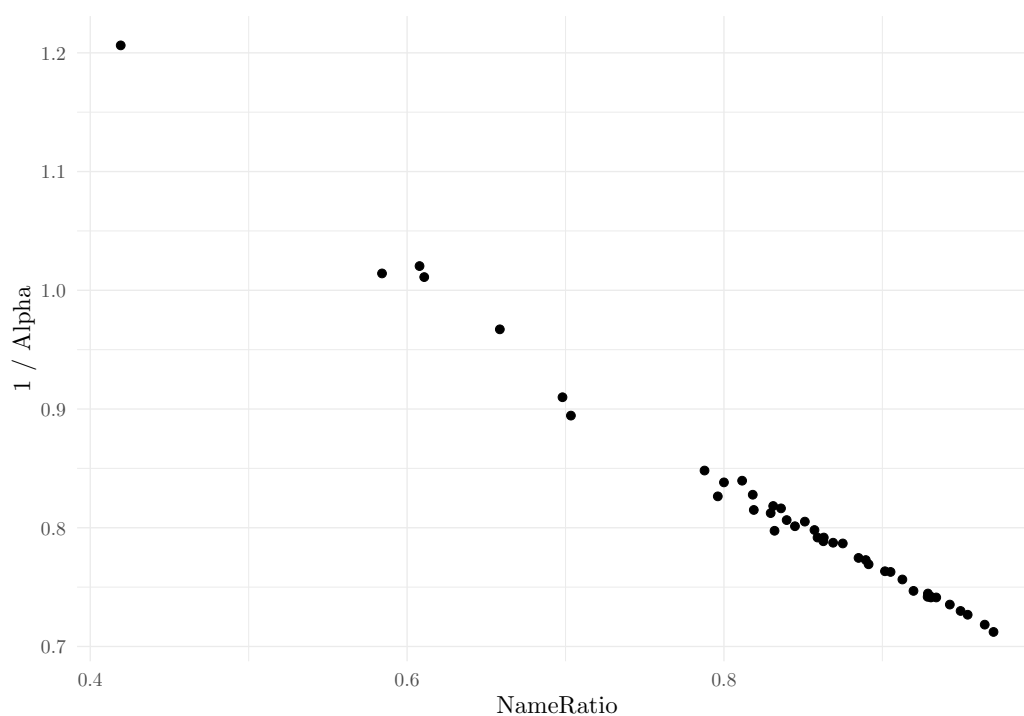
| Statistic | Median | St Dev | Aggregate |
|----------------|--------|--------|-----------|
| Lambda Exp | 1.959 | 0.669 | 0.444 |
| Alpha Pow | 1.263 | 0.124 | 0.832 |
| Alpha Trunc | 0.735 | 0.988 | 0.950 |
| Lambda Trunc | 0.801 | 0.725 | 0.017 |
| Trunc vs Pow R | 5.875 | 2.512 | 17.026 |
| Trunc vs Pow p | 0.000 | 0.000 | 0.000 |
| Trunc vs Exp R | 0.710 | 1.779 | 20.760 |
| Trunc vs Exp p | 0.141 | 0.274 | 0.000 |
| Pow vs Exp R | -4.958 | 3.074 | 17.546 |
| Pow vs Exp p | 0.000 | 0.187 | 0.000 |

Note: The distribution names in the tests are abbreviated as follows: *Exp* = exponential, *Pow* = (pure) power, *Trunc* = power law with exponential decay. *R*: ratio of goodness-of-fit; a positive number means that the first law of the two is preferred. *p*: significance level; the probability that the preference would be due to randomness. The same abbreviations are used in other comparable tables.

to 1 and in some cases even under 1: Pavesi et al. (2003). Sicily's ancient name, *Magna Graeca*, reminds us that a share of its population came from Greece in the archaic period, so it is interesting that this particular shape in name distribution, where the parameter is close to 1, would be found there as well.

The models for the limiting behavior of name distributions developed by Baek, Kiet, & Kim (2007), Maruvka, Shnerb, & Kessler (2010) or Reed & Hughes (2003b) effectively enforce a particular relationship between the name ratio and the power law parameter α , as we pointed out earlier: $\alpha = \frac{1}{1+r}$. Figure 4.8 plots the relationship in the case of Greece, across *poleis*, in logarithmic scale. We can see that it would not be possible to apply a relationship of the prescribed form. Adding the degree of freedom afforded by the strategic selection of names, we can express the relationship instead as $\alpha = \frac{1}{1 + \frac{\gamma}{\lambda - \mu} - r}$ where in this expression $\gamma > 0$ shows the intensity of name choices of the conformist sort (picking the most common names).

We find that $\frac{1}{\alpha} = 1.52 - 0.84 r$, and the coefficient for r is fairly close to 1. In addition, we can see that $\frac{\gamma}{\lambda - \mu} = -0.52$, which would suggest a fairly strong rate of conservative naming choices.

Figure 4.8: *Relationship Between Distribution and Number of Names*

4.5 Conclusion

We began our analysis by asking why one would have names in the context of ancient Greece, and proposed a model inspired from identity and clan economics. We showed that the exchange of names could be understood as a way of improving social cooperation, in comparison with a full information case in particular. Given this context, we argued that optimal name choices could be reduced to some simple alternatives, for simplicity, and showed that adding the possibility for strategic names lead to particular distributions at the limit. Then, we examined empirical data from the archaic and classical periods, and showed that this model could well account for the patterns we observed.

4.6 Chapter Bibliography

- Abry, Patrice, Cécile Bastidon, Pierre Borgnat, Pablo Jensen, & Antoine Parent.** 2022. "The Contours of a Cliophysics: How Can Econophysics Enrich Cliometrics? Case Studies in Debt Issues and Global Capital Markets." *Frontiers in Physics* 10: 894400.
- Akerlof, George A., & Rachel E. Kranton.** 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–753.
- Alstott, Jeff, Ed Bullmore, & Dietmar Plenz.** 2014. "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." *PLOS ONE* 9 (1): e85777.
- Assael, Yannis, Thea Sommerschild, & Jonathan Prag.** 2020. "Sommerschild/Ancient-Text-Restoration." June 6, 2020. <https://github.com/sommerschild/ancient-text-restoration>.
- Baek, Seung Ki, Hoang Anh Tuan Kiet, & Beom Jun Kim.** 2007. "Family Name Distributions: Master Equation Approach." *Physical Review E* 76 (4): 046113.
- Barucca, Paolo, Jacopo Rocchi, Enzo Marinari, Giorgio Parisi, & Federico Ricci-Tersenghi.** 2015. "Cross-Correlations of American Baby Names." *Proceedings of the National Academy of Sciences* 112 (26): 7943–7947.
- Bastidon, Cécile, & Antoine Parent.** 2022. "Cliometrics of World Stock Markets Evolving Networks." *Annals of Operations Research* To appear.
- Bastidon, Cécile, Antoine Parent, Pablo Jensen, Patrice Abry, & Pierre Borgnat.** 2020. "Graph-Based Era Segmentation of International Financial Integration." *Physica A: Statistical Mechanics and Its Applications* 539: 122877.
- Berg, Alastair.** 2019. "Identity in Economics: A Review." Working Paper ID 3421215. Rochester: Social Science Research Network. <https://doi.org/10.2139/ssrn.3421215>.
- Bernheim, B. Douglas.** 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–877.
- Bresson, Alain.** 1981. "Règles de nomination dans la Rhodes antique." *Dialogues d'histoire ancienne* 7 (1): 345–362.
- . 2019. "Chapitre IV. Une famille camiréenne de commerçants en blé." In *La cité marchande*, 95–99. Scripta Antiqua. Pessac: Ausonius Éditions.
- Broux, Yanne, & Willy Clarysse.** 2016. "Would You Name Your Child After a Celebrity? Arsinoe, Berenike, Kleopatra, Laodike and Stratonike in the Greco-Roman East." *Zeitschrift Für Papyrologie Und Epigraphik* 200: 347–362.
- Busse, Thomas V., & Louisa Seraydarian.** 1977. "Desirability of First Names, Ethnicity and Parental Education." *Psychological Reports* 40 (3): 739–742.
- Canevaro, Mirko, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, eds.** 2018. *Ancient Greek History and Contemporary Social Science*. Edinburgh Leventis Studies 9. Edinburgh: Edinburgh University Press.
- Carpenter, Jeffrey P.** 2004. "When in Rome: Conformity and the Provision of Public Goods." *The Journal of Socio-Economics* 33 (4): 395–408.
- Carr, Jack L., & Janet T. Landa.** 1983. "The Economics of Symbols, Clan Names, and Religion." *The Journal of Legal Studies* 12 (1): 135–156.

- Clauset, Aaron, Cosma Rohilla Shalizi, & M. E. J. Newman.** 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.
- Cline, Diane Harris.** 2020. "Athens as a Small World." *Journal of Historical Network Research* 4: 36–56.
- Cuesta, Jose A., Carlos Gracia-Lázaro, Alfredo Ferrer, Yamir Moreno, & Angel Sánchez.** 2015. "Reputation Drives Cooperative Behaviour and Network Formation in Human Groups." *Scientific Reports* 5 (1): 7843.
- Gaechter, Simon, Kyeongtae Lee, & Martin Sefton.** 2020. "Risk, Temptation, and Efficiency in Prisoner's Dilemmas." Working Paper 2020-15. Discussion Papers. University of Nottingham. <https://ideas.repec.org/p/not/notcdx/2020-15.html>.
- Gillen, Benjamin J.** 2015. "A Note on Endogenous Norms in a Theory of Conformity." *Economics Letters* 129: 57–61.
- Hahn, Matthew W, & R Alexander Bentley.** 2003. "Drift as a Mechanism for Cultural Change: An Example from Baby Names." *Proceedings of the Royal Society B: Biological Sciences* 270: S120–S123.
- Hardie, Margaret M.** 1923. "The Significance of Greek Personal Names." *Folklore* 34 (2): 149–154.
- Harrison, John, & Ju Yeong Kim.** 2020. "RSelenium: R Bindings for 'Selenium WebDriver'." February 3, 2020. <https://CRAN.R-project.org/package=RSelenium>.
- Herman, Gabriel.** 1990. "Patterns of Name Diffusion Within the Greek World and Beyond." *The Classical Quarterly* 40 (2): 349–363.
- Hornblower, Simon.** 2010. *Thucydidean Themes*. Oxford: Oxford University Press.
- Karila-Cohen, Karine.** 2016. "Prosopographia Attica 2.0 : base de données et raisonnement prosopographique." *Revue historique* 680 (4): 869–904.
- . 2017. "Quand les familles s'affichent sur les pierres (à partir de l'exemple athénien)." *Pallas Hors Série*: 91–117.
- . 2018. "Le graphe, la trace et les fragments : l'apport des méthodes quantitatives et des outils numériques à l'étude des élites civiques athéniennes." *Annales HSS* 73 (4): 785–815.
- . 2019. "La valeur du lien : relations extérieures et capital social des notables athéniens de la basse époque hellénistique." In *La cité interconnectée dans le monde gréco-romain (IV^e siècle a.C.-IV^e siècle p.C.) : transferts et réseaux institutionnels, religieux et culturels aux époques hellénistique et impériale*, edited by Madalina Dana & Ivana Savalli-Lestrade, 287–306. Scripta antiqua 118. Pessac: Ausonius Éditions.
- Karila-Cohen, Karine, Claire Lemercier, Isabelle Rosé, & Claire Zalc.** 2018. "Nouvelles cuisines de l'histoire quantitative." *Annales HSS* 73 (4): 773–783.
- Lee, Mi Jin, Woo Seong Jo, Il Gu Yi, Seung Ki Baek, & Beom Jun Kim.** 2016. "Evolution of Popularity in Given Names." *Physica A: Statistical Mechanics and Its Applications* 443: 415–422.
- Li, Wentian.** 2012. "Analyses of Baby Name Popularity Distribution in U.S. For the Last 131 Years." *Complexity* 18 (1): 44–50.
- Maruvka, Yosi E., Nadav M. Shnerb, & David A. Kessler.** 2010. "Universal Features of Surname Distribution in a Subsample of a Growing Population." *Journal of Theoretical Biology* 262 (2): 245–256.
- Mateos, Pablo, Paul A. Longley, & David O'Sullivan.** 2011. "Ethnicity and Population Structure in Personal Naming Networks." *PLOS ONE* 6 (9): e22943.

- Mengel, Friederike.** 2018. "Risk and Temptation: A Meta-study on Prisoner's Dilemma Games." *The Economic Journal* 128 (616): 3182–3209.
- Motschenbacher, Heiko.** 2020. "Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names." *Names* 68 (2): 88–103.
- Parker, Robert, Jean-Baptiste Yon, & Mark Depauw.** 1996. "Lexikon of Greek Personal Names." 1996. <http://www.lgpn.ox.ac.uk>.
- Parkin, Tim.** 2013. "The Demography of Infancy and Early Childhood in the Ancient World." In *The Oxford Handbook of Childhood and Education in the Classical World*, edited by Judith Evans Grubbs & Tim Parkin, 40–61. Oxford: Oxford University Press.
- Pavesi, Angelo, Paola Pizzetti, Enzo Siri, Enzo Lucchetti, & Franco Conterio.** 2003. "Brief Communication: Coexistence of Two Distinct Patterns in the Surname Structure of Sicily." *American Journal of Physical Anthropology* 120 (2): 195–199.
- "PHI Greek Inscriptions"**. n.d. <https://inscriptions.packhum.org/>.
- Puech, Vincent.** 2012. "La méthode prosopographique et l'histoire des élites dans l'Antiquité tardive." *Revue historique* 661 (1): 155–168.
- Reed, William J., & Barry D. Hughes.** 2003. "On the Distribution of Family Names." *Physica A: Statistical Mechanics and Its Applications* 319: 579–590.
- Rogerson, Peter A.** 2016. "The Effects of Months, Holidays, and the Birthdays of Presidents on Choice of Baby Names." *Names* 64 (4): 234–241.
- Woods, Robert.** 2007. "Ancient and Early Modern Mortality: Experience and Understanding." *The Economic History Review* 60 (2): 373.
- Young, Robert K., Ana H. Kennedy, Ann Newhouse, Pam Browne, & Del Thiessen.** 1993. "The Effects of Names on Perception of Intelligence, Popularity, and Competence." *Journal of Applied Social Psychology* 23 (21): 1770–1788.
- Zanette, Damián H., & Susanna C. Manrubia.** 2001. "Vertical Transmission of Culture and the Distribution of Family Names." *Physica A: Statistical Mechanics and Its Applications*, Proceedings of the IUPAP International Conference on New Trends in the Fractal Aspects of Complex Systems, 295 (1): 1–8.

Chapter 5

Laws of Divine Power in the Greek Pantheon: The Cliometrics of God Choice

It is well known that the ancient Greeks venerated many gods, and the pantheon of twelve gods is well known; the names Zeus, Apollo, Athena or Aphrodite are instantly recognizable for many. This familiarity hides in fact the extent to which these ancient Greek gods are still actively researched by historians and philologists. One particular angle through which the understanding of ancient religion has been furthered is the study of the ways in which the gods were addressed, which led to the creation of some of the electronic datasets we discussed in Chapter 3. As Brulé (1998), Brulé (2005) and Parker (2003) notably stressed, one cannot think of a god as a perfectly defined entity, because the divine notion attached to a name varies across *poleis*, and the qualifications with which they are addressed vary significantly. At the first order, one may wonder how the logic in which people worshipped multiple gods could optimally operate.

In this chapter we seek an endogenously optimal explanation for the choices by the ancient Greeks of which god(s) to address. We develop a simple economic model that relates the choices of gods to these gods' assumed performance in a competitive setting. The study of ancient Greek religion saw an opposition build up in the 1970s, according to Versnel (2011), between those seeing *kosmos* (order) into it and those seeing *chaos*. The proponents of *kosmos*, following Jean-Pierre Vernant's historical psychology approach, considered that logic in the Greek gods had to be found through structuralist study, and that each one had to be understood in relation with the others, and was defined "by the structure of relations that oppose and unite it to the other powers that constitute the

divine universe” (Versnel 2011, 27). The proponents of *chaos*, following Walter Burkert, considered the gods formed a “scattered and heterogeneous pantheon, a mythology of bits and pieces” (Versnel 2011, 28). These two perspectives have converged in some ways over the years, as Versnel argues, but the present work will clearly follow Vernant’s view, since we will explicitly assume a strong logic at play within the Greek pantheon.

Resorting to formal modeling to account for phenomena related to religion is not new. Indeed, in religion studies, game theory has been used, albeit not extensively, to explain the existence and salient characteristics of religion. Extending the field of religious studies, Bulbulia (2010) and Bulbulia & Frean (2010) proposed an evolutionary model for religious cooperation, and showed how sacred culture enhances cooperative prediction and response among anonymous partners³⁷. The evolution of costly displays is discussed in Henrich (2009). In his small book, Chwe (2001) discussed many aspects in which religious ritual can be analyzed as a rational action from the standpoint of simple game theory. Using a distinct approach from game theory, Kaše, Hampejs, & Pospíšil (2018) have considered the religious rituals of early Christianity from the perspective of agents-based modeling, in contrast with system dynamics.

More specifically from the standpoint of economics research, some early work by Carr & Landa (1983), the general framework on which we based our approach in Chapter 4, addressed religion as part of the study of identity and decision to belong to a group. Covering a wider ground, Iannaccone (1998) and Iyer (2016) present literature reviews for the economics of religion in all its facets. Economics research has focused on the structure of religious organizations, among other things. For example, Levy & Razin (2007) and Levy & Razin (2012) used symmetric Prisoner Dilemma games to represent religious organizations and showed that more demanding religious organizations attract less members, but these members are more cohesive. Iannaccone, Haight, & Rubin (2011) show how secular leaders influence the location of sacred places. The sacred places then gain widespread authority precisely because they lie beyond the centers of secular power. In McBride (2016), the focus is on showing how rituals create religious authority. Levy & Razin (2014) develop a model of social signaling of displays of religious behavior and cooperative acts in religious organizations. They show that ritual-based religions

³⁷In other words, a common sacred culture is shown to be one way of avoiding a lack of cohesion in the classical “stag hunt” problem, as originally defined by Jean-Jacques Rousseau.

require a costly involvement, which acts as a signal and also implies a higher level of coordination in social interactions and a higher incidence of mutual cooperation³⁸. In this body of literature, religions and some of their specific features are shown to optimally exist in an equilibrium. Applying economic theory to questions of beliefs has raised some debates and questions³⁹, as it has been argued that religious acts may not be equated to picking products on a shelf. From the standpoint of the theory of religious economy, one can consider religion as a social good more than simply a product.

Empirical tests have confirmed many economics-based theoretical models. In Ruffle & Sosis (2007), time-consuming religious rituals are explained as promoting cooperation, and the authors offered empirical tests on the cooperative behavior of members of religious and secular Israeli kibbutzim. Also, D. D. P. Johnson (2005), relying on the theoretical models that suggest human cooperation may be promoted if people believe in supernatural punishment, empirically tested across a large number of cultures⁴⁰ whether the likelihood of supernatural punishment was related to observable cooperation. In Power (2017), the author empirically showed that people are attending to the full suite of religious acts carried out by their peers, using these visible signals to discern multiple aspects of their character and intentions.

While the literature mentioned above, both in religion studies and in economics, has concentrated on fundamental explanations for the existence of religion or of some of its rituals, our approach will differ. Our initial focus is not on the reason why religion may optimally exist in a certain form, but rather, given that it exists, how it may be optimally practiced from the standpoint of both the population and the assumed characteristics of the gods. In particular, the specific aspects of polytheistic religion will play an important role in our analysis. There are few empirical studies in religion science that examined polytheism. Gries, Su, & Schak (2012) proposed specific methods of study and applied them to a survey in Taiwan.

³⁸This framework could have interesting implications for the study of ancient Greek rituals, which were precise and constraining; this is not our focus in the present work however.

³⁹See Bankston (2002) for example. There are also some discussions on how academics from a given religious background may be able to better apply this economic framework, see Iannaccone (2005).

⁴⁰Using the so-called Standard Cross-Cultural Sample (SCCS) of 186 human societies around the globe, spanning time and space.

There is also little research on polytheism within the framework of the theory of religious economy. Ferrero & Tridimas (2018) developed a model where people assumed certain behavior for the gods and made their own decisions as a consequence of this assumed decision logic. They argue that the overlap between different god's prerogatives created inefficiencies, and that could be an explanation for the convergence towards monotheism that was observed at the end of Antiquity. We will not in our approach distinguish between different gods' jurisdictions, but will instead consider their differences in terms of how easily it supposedly came to them, or how efficient they were, to give benefits back to their mortal worshippers.

In the next section of this chapter, we examine some aspects of ancient Greek religious thought, drawing some examples from Greek literature, and we identify important motivations in ritual behavior. Next, we develop a theoretical model that explains the distribution of choices of which god or gods to worship within a *polis* as a function of certain assumed qualities of the gods. We then relate such qualities to human performance in competitive sports or games, and infer that observations of which god or gods were addressed could follow a power law with exponential decay mixed with a uniform distribution. We also develop a random growth model whereby people update their perception of the gods' efficiency based on random life events, and show that in this case the distribution of votive acts would be expected to approximately follow a power law distribution as well. Then, in the third section, we combine data from two of the data sources we discussed in Chapter 3: the BDEG project on ancient Greek divine epithets and the extensive ancient Greek *poleis* survey, the POLIS data. We structure this data so that we can measure the distribution of votive acts across gods, and test the theoretical model from the first section. We find that the model's predictions are close to empirical observations, and also look for certain explanatory factors for the degree of efficiency that worshippers assumed for the gods.

Our approach fits within the methodological framework we developed in Chapter 2, and the example we gave at the end of that chapter is precisely what we focus on in this one. We are directly exploiting primary sources for the most part in order to better comprehend polytheist behavior, at a large scale. By rescaling the data in a particular way, we are not attempting to explain it, but rather to make its internal logic clearer.

5.1 Some Aspects of Ancient Greek Religion

We will argue that from the ancient Greek's perspective, the relationship between men and gods was fundamentally an exchange, and that we can set this exchange in economic terms. On one hand, people would carry out votive acts and rituals which were expected to please the gods⁴¹. Reciprocally, the gods were expected to give back thanks, *charis*, a form of *quid pro quo* through which the gods would effectively pay thanks to the worshippers through some actions⁴².

The nature of this exchange between men and gods is discussed and probed in a good amount of detail by Plato in *Euthyphro* throughout the book. In particular, the eponymous priest states:

“In summary, what is pious is to know what to say and do that is agreeable to the gods, either by praying or by doing sacrifices, and it ensures the safety of families and cities⁴³”

In this sentence, the term for “agreeable” is literally expressed as “bringing *charis*”. With an interestingly explicit comparison, in one of his replies, Socrates remarks further:

“Hence it seems to me piety is a kind of commercial traffic between men and gods.⁴⁴”

Aphrodite expresses the gods' need for being honored in Euripides's play *Hippolytus*, and this need actually causes the whole intrigue:

“Because this is essential to the race of the gods, they rejoice in being worshipped by men⁴⁵”

⁴¹For example, gods were supposed to enjoy and feast on the smoke raising from various meats cooked during sacrifice, see Ekroth (2011).

⁴²For example, in one inscription mentioned in Bonnet & Lebreton (2019), a mother invokes the goddess Artemis for the benefit of her children.

⁴³τόδε μέντοι σοι ἀπλῶς λέγω, ὅτι ἐὰν μὲν κεχαρισμένα τις ἐπίσθηται τοῖς θεοῖς λέγειν τε καὶ πράττειν εὐχόμενός τε καὶ θύων, ταῦτ' ἔστι τὰ ὅσια, καὶ σώζει τὰ τοιαῦτα τούς τε ἰδίους οἴκους καὶ τὰ κοινὰ τῶν πόλεων. Plato, *Euthyphro*, 14b. Our translation, unless otherwise specified.

⁴⁴Ἐμπορικὴ ἄρα τις ἂν εἴη, ὧς Εὐθύφρων, τέχνη ἢ ὁσιότης θεοῖς καὶ ἀνθρώποις παρ' ἀλλήλων. Plato, *Euthyphro*, 14e.

⁴⁵ἔνεστι γὰρ δὴ κὰν θεῶν γένει τόδε: / τιμώμενοι χαίρουσιν ἀνθρώπων ὑπο. Euripides, *Hippolytus*, 7-8.

If we look further back in time in Homer's work for example, the same notion is also clearly expressed by Telemachus in Homer's *Odyssey* 16, when he mistakes his father for a god, using the same term:

"Be good for us, so that we will give you agreeable sacrifices and golden, well-made gifts: spare us.⁴⁶"

These mentions of the relationship between men and gods only address the "amount" of effort going in both directions, and not the fact that the gods may have specific jurisdictions or specialties. The list of the skills typically associated with each god is well known, and Ferrero & Tridimas (2018) based their approach of polytheism on the particular fields for which each god is recognized. By projecting the multidimensional characteristics of the gods onto a single measure, our work hypothesis is strong.

However, as we will see in our discussion of the data, the gods that received the most dedications within various *poleis* were not always the same. At the same time, it would seem a reasonable assumption that the things people wanted help with across *Hellas* were generally the same. Hence, the roles of the gods were to some extent fungible, and one could not argue that choosing between one or the other only depended on the gods' particular domains⁴⁷.

Defining a god precisely, and in particular in opposition to other gods, is not straightforward, and simply stating that there are a certain number of gods is not innocuous. Should *Poseidon* and *Poseidon Phukios* ("of the algae") be considered the same⁴⁸? In Mantinea, one counted five different ways of addressing Zeus (Bruit Zaidman & Schmitt Pantel 1992, 212). The porous border lines between god denominations make the identification of each one non trivial, as well as their counting. We will nevertheless consider that some lines are drawn between all the acts of worship that one could consider, and that they can be grouped along the particular divine characteristics that our formal approach will need to use; so for our purposes, we will count a single *Poseidon* in the example above⁴⁹. The complex onomastic formulae that may have been used,

⁴⁶ἀλλ' ἴληθ', ἵνα τοι κεχαρισμένα δώομεν ἱρά / ἦδὲ χρύσεια δῶρα, τετυγμένα: φείδεο δ' ἡμέων. Homer, *Odyssey*, 16, 184-185.

⁴⁷We will consider some data that supports this argument in Table 5.3 further down.

⁴⁸Lebreton (2019) cites some examples of this kind.

⁴⁹There are some cases in the data where we actually separate the denominations, when the epithets are relative to a location.

which we briefly addressed in Chapter 3, are not accounted for in this approach.

Relying on these observations, we can ascribe basic motivations to worshippers, expressed in economic terms as utility maximization, which not only drive their own decisions but also the behaviors they assume the gods to follow. Since votive acts by the ancient Greeks were generally for a specific request, it seems reasonable that people would act rationally in order to maximize what they obtained in return.

The essential features that could underpin a formal model are as follows:

1. The different gods are assumed by the worshippers to offer a certain quantity of something that is beneficial to the worshippers. The quality or the nature of what is asked is also naturally important and would be carefully specified by the people participating in ritual; nevertheless we assume that this can be reduced down to a quantifiable metric, the *charis*.
2. The people who participate in worship share the total benefit. Note that in the models that generally explain religion itself, most approaches necessitate that the utility derived from religion would be increased by common participation. See Iannaccone (1998) or Iyer (2016) for descriptions of the models of religious structure relying on club goods. This is not in opposition with our approach, which takes the existence of religion for granted; we are only addressing the manner in which religion is practiced, but the participants could still withdraw large benefits from the fact that they all have a religious system in common. We consider that the total *charis* that a god can offer must be shared by all the worshippers; there is therefore no explicit advantage in everybody choosing the same god⁵⁰. The sharing may not be equally received: a king praying to a god among a crowd may expect (and may be expected) to receive much more than the other participants. We will nevertheless first consider the case of equally shared benefits and then generalize it.
3. Each god is assumed to be seeking to maximize the benefits they gain from the attendance of their worship, and that they derive a benefit equal to the number of people who worship them.

⁵⁰Although there may be in choosing the same religion, in accordance with the club goods approach.

4. The gods incur some internal cost in providing the *charis* from point 1, and that cost is a function of the total amount they give to worshippers. The worshippers assume that the gods need to spend time and effort to give them benefits, and hence it is not a free resource. For simplicity, we consider that this cost is proportional to the total amount of *charis* they give back. Different gods incur different proportional costs in giving blessings to humans, as they all have different types and degrees of power. It is generally accepted that the powers or domains of the gods are known to be far from fixed, they depend on the context, type of request, also on location. We will in a later section develop a model that accounts for some evolution in the assumptions about the gods. We will nevertheless assume that, locally, the common assumptions about the gods are stable enough.

5.2 Worship Decision Model

In this section, we develop a theoretical model that could account for the distribution of votive acts in a given polity. In the following subsection, we address the decisions of people worshipping, and after that, the assumed decisions of the gods who determine how much they would give back to the worshippers, as a function of their internal costs for doing so. Finally, we will look into the patterns that one may expect in the assumed behavior of the gods, first based on the regularity patterns in human performance, and then from the perspective of a random growth model.

5.2.1 People's Devotion Decisions

Focusing on people's decisions, we begin by holding the total amount of *charis* provided by each god as a constant in order to determine a first equilibrium; we will later allow it to vary when we consider the god's (assumed) decisions. We assume there are $K + 1$ gods that may be worshipped. We consider a population of n people, or groups of people. For our purposes we do not distinguish between individuals and groups of individuals. People expect to obtain a share of the *charis* from participating in the ritual pertaining to a specific god. So when people participate in ritual for god $k \in [0..K]$, there is a total volume χ_k of *charis*, for now considered a given.

When several people choose to worship the same god at the same time, we will first assume that the total benefit given back by the god is shared evenly. If we write h_k the number of people choosing god k , then each receives $\frac{\chi_k}{h_k}$. In our framework, the only metric that people consider in the practice of ritual is the amount of *charis* they expect to receive.

Each person (or group of persons) $i \in [1..n]$ is assumed to be risk neutral and seeks to maximize the share of *charis* they receive. As a result, and since we do not make any particular preference assumptions that would distinguish one person from another, the situation is represented by a symmetric game with perfect information. In particular, the problem of optimally accessing a resource shared in common is well known, a form of “El Farol Bar” or “Kolkata Paise” problem, as discussed in Arthur (1994) or A. S. Chakrabarti et al. (2009). We have the following result.

Proposition 5.1 (Optimal Worship Strategy). *The Nash equilibrium in the game for the worshippers is a symmetric mixed strategy with a probability distribution $(\pi_k)_{k \in [0..K]}$ across gods, and $\forall k \in [0..K]$:*

$$\lim_{n \rightarrow \infty} \pi_k = \frac{\chi_k}{\sum_{i=0}^K \chi_i}.$$

Proof. We assume a symmetrical Nash equilibrium in the game in the form of a mixed strategy: all players draw their choice for the god to choose from the same distribution $(\pi_k)_{k \in [0..K]}$ where $\pi_k > 0$. The equilibrium is characterized by the fact that the expected gain for any player is independent from the specific choice they make.

Set e_k the expected share in *charis* received from choosing god k when all others follow the optimal strategy, and can hence be expressed as a function of π_k . The number of other people choosing the same god follows a binomial distribution and:

$$e_k = \sum_{i=0}^{n-1} C_i^{n-1} \pi_k^i (1 - \pi_k)^{n-1-i} \frac{\chi_k}{i+1}.$$

This quantity is known, see Chao & Strawderman (1972) for example, and simplifies to:

$$e_k = \chi_k \frac{1 - (1 - \pi_k)^n}{n\pi_k}.$$

Since the Nash equilibrium is characterized by $e_k = e_0$ for all k , we have:

$$\frac{1 - (1 - \pi_0)^n}{n\pi_0} = \frac{\chi_k}{\chi_0} \frac{1 - (1 - \pi_k)^n}{n\pi_k}.$$

As n becomes large, both $(1 - \pi_0)^n$ and $(1 - \pi_k)^n$ go to zero and the equation becomes:

$$\pi_k = \frac{\chi_k}{\chi_0} \pi_0.$$

By definition $\sum_{k=0}^K \pi_k = 1$, so $1 = \pi_0 + \sum_{k=1}^K \pi_k$ and hence $1 = \pi_0 + \sum_{k=1}^K \frac{\chi_k}{\chi_0} \pi_0$. Solving for π_0 we obtain after simplification:

$$\pi_0 = \frac{\chi_0}{\sum_{i=0}^K \chi_i}.$$

Similarly, we also obtain that $\pi_k = \frac{\chi_k}{\sum_{i=0}^K \chi_i}$. □

Note that scaling the share received up or down by the same factor has no impact on the optimal probabilities. We can see that the higher the total amount of *charis* a god can provide, the greater the optimal probability of choosing that god for worshipping, although the optimal strategy randomizes across all the possible choices, and even a god who would offer very little *charis* sees a non zero probability of seeing some devotion acts. All gods hence see some worship all the time.

It is worth stressing that if the benefit of participating in worship was not shared but was simply unaffected by the choice of other people, or even was increased by other people participating, then there would be no mixing and the same god with the highest benefit would be worshipped all the time. Then in a polytheist context, it appears that by construction there needs to be a declining benefit in participating in the same rituals if one is to explain the multiplicity and spread of god choices. One could argue that the gods offered some specialization, and therefore there could be a need for a diversity of gods; but then why would not some of them be merged together? As we will see in the framework of the random growth model for the observed benefits associated to each god developed in a later section, the multiplicity of gods may be maintained simply due to chance.

As we have stressed, we made a particularly stringent assumption in that benefits would be shared equally among participants. We know this was likely not considered to be the case by the ancient Greeks. Even the most democratic polities had strong hierarchies, and honors during ritual were concentrated on a few participants. It makes sense, therefore, to assume that some people would collect a greater share of *charis* than others, relative to their number. In order to capture this situation, let us assume that there is a proportion r of privileged citizens⁵¹ of the *polis*. These people do not receive their fair equal of *charis*, but a multiple $\rho > 1$ of it. We write $(\pi_k^c)_{k \in [0..K]}$ for the probability that a normal citizen would choose a god, and $(\pi_k^h)_{k \in [0..K]}$ for that probability for a privileged citizen.

With these assumptions, we obtain the following proposition.

Proposition 5.2 (Optimal Worship Strategy with Unequal Shares). *The Nash equilibria in the game for the worshippers with unequal benefits consist of all mixed strategies where the probability distributions $(\pi_k^c)_{k \in [0..K]}$ and $(\pi_k^h)_{k \in [0..K]}$ verify at the limit as n goes to ∞ , $\forall k \in [0..K]$:*

$$\frac{1-r}{1-r+r\rho}\pi_j^c + \frac{r\rho}{1-r+r\rho}\pi_j^h = \frac{\chi_k}{\sum_{i=0}^K \chi_i}.$$

In addition, for each $k \in [0..K]$, if either π_k^c or π_k^h is equal to $\frac{\chi_k}{\sum_{i=0}^K \chi_i}$, then the other is also equal to that value.

Proof. As the number of people increases, we can consider directly as in the proof of Proposition 5.1 that the number of worshippers who choose a god with probability p behaves almost surely as np . With our assumptions, the total number of people weighted by their shares of *charis* worshipping god k is $n((1-r)\pi_k^c + r\rho\pi_k^h)$.

The expected gains from worshipping god k for the members of each class hence write:

$$e_k^c = \frac{\chi_k}{n((1-r)\pi_k^c + r\rho\pi_k^h)}$$

$$e_k^h = \frac{\rho\chi_k}{n((1-r)\pi_k^c + r\rho\pi_k^h)}.$$

Since all the high-class participants are identical, and all the normal citizens are identical, the equilibrium conditions are that their gains would not be dependent on

⁵¹Who may be kings, aristocrats, hierarchs, or any such relevant category.

their choice of god, and therefore $e_i^c = e_j^c$ and $e_i^h = e_j^h$ for i and j in $[0..K]$. These conditions result in the same thing:

$$\frac{\chi_i}{(1-r)\pi_i^c + r\rho\pi_i^h} = \frac{\chi_j}{(1-r)\pi_j^c + r\rho\pi_j^h}.$$

From this, we obtain by summing over i , and using the fact that $\sum_{i=0}^K \pi_i^c = \sum_{i=0}^K \pi_i^h = 1$:

$$\frac{(1-r)\pi_j^c + r\rho\pi_j^h}{(1-r) + r\rho} = \frac{\chi_j}{\sum_{i=0}^K \chi_i},$$

which is the first statement in the proposition. Set $a = \frac{1-r}{1-r+r\rho}$ for simplicity, then if we have $\pi_j^c = \frac{\chi_j}{\sum_{i=0}^K \chi_i}$, we obtain:

$$\pi_j^h = \frac{1}{1-a} \frac{\chi_j}{\sum_{i=0}^K \chi_i} - \frac{a}{1-a} \pi_j^c = \frac{\chi_j}{\sum_{i=0}^K \chi_i}.$$

The same calculation shows that the reciprocal relationship is also true. □

We can see that with these more general assumptions on the manner in which *charis* is shared, we obtain comparable results as with the case when it is shared equally.

In our analysis so far we have considered a constant number of gods $K + 1$, but this is a strong simplification of the actual dynamics of any *polis*-level pantheon. In the ancient Greek polytheist context, foreign gods were not forbidden and could be adopted and adapted. While in the Roman Republic and Empire the adoption of new gods or rituals was regulated⁵², this was not the case in the classical Greek world. However, there is a minimal size for any cult to be carried out, so that enough people can participate in the sacrifices and libations; there is also a minimal group size for the justification for shrines and temples to be built. Hence, we could expect that gods with extremely few followers could not give rise to votive acts simply because the organization of the rituals would be problematic. As a result of this threshold, we would expect to see a finite number of gods, the maximum number being a function of the *polis*'s resources.

⁵²The scandal of the Bacchanalia in 186 BC, as related in Livy's *Ab urbe condita*, 39, 15, led the Roman Senate to legislate specifically to address problems caused by Dionysiac rituals, as was confirmed through epigraphic sources.

5.2.2 Optimal Divine Strategy

So far we have considered that the total amount of *charis* χ_k was a given for every god. Now, we reflect the fact that it is provided by the gods at some internal cost, and they are assumed to choose the optimal amount. We assume that each god wants to maximize the proportion of people who come and worship them. Classical Greek literature, such as epics or tragedy, is filled with examples of the gods taking punitive actions against mortals when they did not properly respect them. As we pointed out earlier, we know too that the gods were assumed to enjoy some of the output from sacrifices, which would be all the more important, the more people partake in it. We accounted earlier for the fact that certain persons would receive a greater share of *charis*; it also appears logical to consider that from the standpoint of the gods, they count for more, so that it is the properly weighted number of people that is relevant.

We assume that the the gods are held to be risk neutral, and know how the people will make their optimal choices, and seek to maximize their gains in the form of the weighted probability of being chosen⁵³ minus a proportional cost attached to producing the *charis*. The gain for god k , as a function of the *charis* χ_k can therefore be written $g_k(\chi_k) = \pi_k(\chi_k) - c_k\chi_k$, where c_k is the exogenously given proportional cost. All gods will seek to maximize their gains at the same time, reflecting their knowledge of each other's strategy.

The following proposition characterizes the endogenously determined *charis*, the probability of being chosen and expected gain at the equilibrium where all gods and worshippers act optimally.

Proposition 5.3 (Optimal Divine Strategy). *The optimal strategy for all gods is for the charis they offer to verify for all k in $[0..K]$ and with $M = \{i : c_i = \max_{k \in [0..K]} \{c_k\}\}$ and $\gamma_k = c_M - c_k$, such that for $k \neq M$, $\gamma_k > 0$ and $\gamma_M = 0$:*

$$\chi_k^* = \frac{\gamma_k(K + 1) + \sum_{j=0}^K \gamma_j}{4(K + 1) \left(\sum_{j=0}^K \gamma_j \right)^2}.$$

⁵³They reason as a function of the population size: we are implicitly assuming that the citizens of a *polis* reason from the perspective of that *polis*, not accounting for the fact that other *poleis* may have different sizes and compete for the gods' attention.

At this optimum, the probability of being chosen is

$$\pi_k^* = \frac{\gamma_k}{2 \sum_{j=0}^K \gamma_j} + \frac{1}{2(K+1)},$$

and the gain for the gods is

$$g_k^* = \pi_k^* \left(1 - \frac{c_k}{2 \sum_{j=0}^K \gamma_j} \right).$$

Proof. Since the gods are measuring their attractiveness in weighted terms $(1-r)\pi_k^c + r\rho\pi_k^h$, then according to Propositions 5.1 and 5.2, we know that this is equal to $\pi_k = \frac{\chi_k}{\sum_{i=0}^K \chi_i}$. There is effectively no difference with the case where benefits are equally shared.

We first note that:

$$\frac{\partial \pi_k}{\partial \chi_k}(\chi_k) = \frac{\sum_{i=0}^K \chi_i - \chi_k}{\left(\sum_{i=0}^K \chi_i \right)^2}.$$

Further, $\sum_{i=0}^K \chi_i - \chi_k$ does not depend on χ_k so that $\frac{\partial \pi_k}{\partial \chi_k}(\chi_k)$ is decreasing, and the second-order condition is verified for the existence of a maximum, as $\frac{\partial^2 \pi_k}{\partial \chi_k^2}(\chi_k) = -\frac{2(\sum_{i=0}^K \chi_i - \chi_k)}{\left(\sum_{i=0}^K \chi_i \right)^2}$.

The first order conditions for each god k 's gain optimization writes: $\frac{\partial \pi_k}{\partial \chi_k}(\chi_k) - c_k = 0$; which based on Proposition 5.1 simplifies into:

$$\sum_{i=0}^K \chi_i - \chi_k = c_k \left(\sum_{i=0}^K \chi_i \right)^2.$$

Since the costs c_k are arbitrary, without loss of generality we choose to rank them so that c_0 is the highest and c_K the lowest. Taking differences for all l and k in $[0..K]$ we hence have

$$\chi_k - \chi_l = (c_l - c_k) \left(\sum_{i=0}^K \chi_i \right)^2,$$

and in particular, at the optimum,

$$\chi_k = (c_0 - c_k) \left(\sum_{i=0}^K \chi_i \right)^2 + \chi_0.$$

Using that expression we can write:

$$\begin{aligned}\sum_{j=0}^K \chi_j &= \sum_{j=0}^K \left((c_0 - c_j) \left(\sum_{i=0}^K \chi_i \right)^2 + \chi_0 \right) \\ &= (K + 1)\chi_0 + \left(\sum_{i=0}^K \chi_i \right)^2 \sum_{j=0}^K (c_0 - c_j).\end{aligned}$$

Hence, taking the square and writing $X = \left(\sum_{i=0}^K \chi_i \right)^2$ we obtain:

$$X = (K + 1)^2 \chi_0^2 + 2(K + 1)\chi_0 \sum_{j=0}^K (c_0 - c_j)X + \left(\sum_{i=0}^K (c_0 - c_j) \right)^2 X^2.$$

Since all the (χ_k) are determined up to a scalar, god 0 can chose χ_0 arbitrarily: there is a continuum of possible optimal equilibria depending on χ_0 . We consider that they always choose the highest χ_0 , that is the highest volume of *charis* compatible with the existence of optimal solutions. Hence χ_0 is set so that the determinant of the equation is null, that is:

$$1 - 4(K + 1)\chi_0 \sum_{j=0}^K (c_0 - c_j) = 0.$$

In that case we obtain the solution

$$X = \left(\sum_{i=0}^K \chi_i \right)^2 = \frac{1}{4 \left(\sum_{j=0}^K (c_0 - c_j) \right)^2}.$$

Replacing in the equation for the optimality we can write after some simplifications:

$$\chi_k^* = \frac{(c_0 - c_k)(K + 1) + \sum_{j=0}^K (c_0 - c_j)}{4(K + 1) \left(\sum_{j=0}^K (c_0 - c_j) \right)^2},$$

and

$$\frac{\chi_k^*}{\chi_0^*} = \frac{(c_0 - c_k)(K + 1) + \sum_{j=0}^K (c_0 - c_j)}{\sum_{j=0}^K (c_0 - c_j)}.$$

Replacing into the expression for π_k , we obtain at the optimum

$$\pi_k^* = \frac{c_0 - c_k}{2 \sum_{j=0}^K (c_0 - c_j)} + \frac{1}{2(K + 1)}.$$

Finally, the optimal expected gain for gods writes:

$$\begin{aligned}
 g_k^* &= \pi_k^* - c_k \chi_k^* \\
 &= \frac{c_0 - c_k}{2 \sum_{j=0}^K (c_0 - c_j)} + \frac{1}{2(K+1)} - c_k \frac{(c_0 - c_k)(K+1) + \sum_{j=0}^K (c_0 - c_j)}{4(K+1) \left(\sum_{j=0}^K (c_0 - c_j) \right)^2} \\
 &= \left(\frac{c_0 - c_k}{2 \sum_{j=0}^K (c_0 - c_j)} + \frac{1}{2(K+1)} \right) \left(1 - \frac{c_k}{2 \sum_{j=0}^K (c_0 - c_j)} \right) \\
 &= \pi_k^* \left(1 - \frac{c_k}{2 \sum_{j=0}^K (c_0 - c_j)} \right).
 \end{aligned}$$

We have specified the ranking of the gods as a function of costs (c_k) with the only effective requirement that c_0 should be the largest. As a result, all the expressions at the optimum can be written by replacing the index 0 with $M = \{i : c_i = \max_{k \in [0..K]} \{c_k\}\}$, so that $c_M = \max_{k \in [0..K]} \{c_k\}$. Setting $\gamma_k = c_M - c_k$ we obtain the expressions in the proposition. \square

The following consequences are worth stressing:

- The expected gain for worshippers goes to zero as n goes to infinity. Since $\frac{1-(1-\pi_k)^n}{n\pi_k} = \frac{1}{n\pi_k} + o(n)$, then the equilibrium gain expectation for the worshippers is

$$\bar{e} = \frac{1}{2n \sum_{j=0}^K \gamma_j} + o(n).$$

One could make the costs all converge to zero, arguing that they are uniformly reduced as a function of the total population for all gods, so that γ_k is instead written $\frac{\gamma_k}{n}$; in this case, the worshippers' expected gain is not null at the limit.

- The derivative of the optimal probability relative to efficiency is always positive: for all $k \neq M$,

$$\frac{\partial \pi_k^*}{\partial \gamma_k} = \frac{\sum_{j=0}^K \gamma_j - \gamma_k}{2 \left(\sum_{j=0}^K \gamma_j \right)^2} > 0.$$

Also, the derivative of the optimal probability for one god relative to the other gods' efficiency is always negative, for $l \neq k$ and both different from M :

$$\frac{\partial \pi_k^*}{\partial \gamma_l} = -\frac{\gamma_k}{2 \left(\sum_{j=0}^K \gamma_j \right)^2},$$

and one can verify that $\sum_{k=0}^K \sum_{i=0}^K \frac{\partial \pi_k^*}{\partial c_i} = 0$.

- The optimal probability π_M^* , which is the lowest among all the π_k^* , is not affected by changes in the efficiency $(\gamma_k)_{k \in [0..K]}$ and is constant with $\pi_M^* = \frac{1}{2(K+1)}$. The model therefore predicts a minimal probability of a god being worshipped that only depends on the total number of gods. The optimal probabilities can also be understood as worshippers choosing, with a probability of $\frac{1}{2}$, a god randomly with a uniform probability across all the gods, and with a probability of $\frac{1}{2}$, a god k following a probability of $\frac{\gamma_k}{\sum_{j=0}^K \gamma_j}$.
- The simple expression $\gamma_k = c_M - c_k$, and the contribution of $\frac{\gamma_k}{\sum_{j=0}^K \gamma_j}$ in the expression for the amount of worship look interestingly like the mathematical expressions of the structuralist angle that Vernant followed in looking at Greek religion. Indeed, the full quote from Vernant, with added emphasis, is that:

A god is a power that represents a type of action, a kind of force. Within the framework of a pantheon, each of these powers is *defined not in itself* as an isolated object but by virtue of its *relative position in the aggregate of forces*, by the structure of relations that oppose and unite it to the other powers that constitute the divine universe⁵⁴.

Since the gods are believed to place themselves on a uni-dimensional space though the choice of *charis* they give back to worshippers, we may relate this model to the approach to competition initially developed by Hotelling (1929). However, the gods' efficiency, which we consider as a given at this stage, condition their ability to distinguish and define themselves by their effort in giving *charis*. As a result, the optimal position is driven by their characteristics rather than by strategy, in which the outcome differs from the competitive setting envisioned by Hotelling.

⁵⁴Versnel (2011), p. 27.

At this stage, with our assumptions, we have established a relationship between differences in the god's efficiency at carrying out acts in favor of their worshippers, and the probability that they are worshipped. We will now turn to two different approaches that attempt to explain the distribution of votive acts that one should expect to observe.

5.2.3 Human Drivers of Divine Performance

Why would the gods be expected to possess different degrees of ability in offering *charis* to humans? It seems reasonable to consider that could be the case because the gods are assumed by people to be to some extent in competition between each other, and as a result would have acquired different levels of expertise. The ancient Greeks were familiar with competition between each other, especially in sporting events. In fact, exceptional winners at sporting events were revered through poetry, in particular in Pindar's *epinicia* (odes to victors and victory). By winning and displaying rare skill or strength, they differentiated themselves from the rest, and were often qualified as divine (Miller 2018). In the *Iliad*, the heroes, whose prowess at war was extraordinary, are also commonly qualified as divine. The literary perspective clearly establishes a link between extreme performance and divinity. Hence, it is not inconceivable that, since the attributes of extreme performance are associated to gods, then the attributes of gods be associated to extreme performance.

Human physical and cognitive performance has largely been documented to follow power law distributions. Donner & Hardy (2015) show how power laws, or combinations thereof, represent human learning curves for a variety of tasks. Record running times as a function of distance have been shown to be well captured by power laws as well, as discussed by Vandewalle (2018). Of greater importance for our purposes, the rankings of human performance have been documented to follow various types of power laws. For example, Smith (2015) documents the presence of power laws in several psycho-motor actions. More specifically for sports, Deng et al. (2012) showed that across a large set of sports (tennis, soccer, snooker, and many others), performance as a function of ranking followed power laws with exponential decays. Focusing on fluctuations of rankings in sports and games over time, Morales et al. (2016) showed that power laws with an exponential decay also accounted well for rankings in sports and competitive games

such as poker. A generative model for a power law with exponential decay was proposed in Baek, Bernhardsson, & Minnhagen (2011), based on random group formation.

Note that these considerations would exclude the case where the assumed characteristics of the gods would be distributed with some variations around an average. The shape of the rankings obtained by drawing outcomes from a Gaussian distribution, for example, and sorting them, would differ from what one gets with power laws or variations thereof.

Power laws have been applied to ranking modeling in many areas of economics and for most human activities. Gabaix (2016) offers an overview in economics: the most salient examples include the distribution of city sizes, firm sizes, stock market shocks, or individual wealth. Focusing on the ways in which the emergence of power laws may be accounted for through various theoretical arguments, Mitzenmacher (2004) lists a large section of human experience where rankings have been documented to follow these distributions, including the size of computer files, number of connections in networks, or word frequencies in texts. In particular, market shares in very diverse sectors (from cereals to rifles) show noticeable regularity and are better represented by power laws than by exponential relationships, as discussed in Kohli & Sah (2006)⁵⁵.

Let us rank the gods by their efficiency $(\gamma_k)_{k \in [1..K]}$ with $\gamma_1 = \max\{\gamma_j : j \in [0..K]\}$ setting by convention $\gamma_0 = \gamma_M = 0$. We consider that people give them anthropomorphic qualities reflective of competitive activities. The terms $\frac{\gamma_k}{\sum_{j=0}^K \gamma_j}$ represent the gods' efficiency scaled to their total efficiency, which, thanks to this scaling, can be understood as a dimensionless measure, akin to a scaled number of points in a competitive sport or game. Hence, following Deng et al. (2012) or Morales et al. (2016), we would expect for all $k \in [1..K]$ that the scaled efficiency follows a power law with exponential decay, or as we explained in Chapter 3, a truncated power law, and to have an expression of the form

$$\frac{\gamma_k}{\sum_{j=0}^K \gamma_j} = ak^{-1-\alpha}e^{-\lambda k},$$

⁵⁵One could directly consider the choices of gods to interact with as a similar situation to market shares for various brands. This could naturally lead us to expect some form of power law distribution for votive acts; however this would not fundamentally explain why it should be the case.

for some positive numbers a , α and λ . The exponential decay appears to be a necessary feature in order to account for a clear concave shape in log-log plots of competitive sports ranking data. In the analysis of sports rankings, the power exponent α has been associated with the degree of competitiveness. Stated simply, the steeper the drop in gains as a function of rank, the more the activity is deemed to be competitive.

With this definition, the probabilities with which people are expected to worship gods become, for $k \in [1..K]$:

$$\pi_k^* = ak^{-1-\alpha}e^{-\lambda k} + \frac{1}{2(K+1)}.$$

The differences $\pi_k^* - \pi_0^*$ would therefore be expected to follow power laws. The distribution for π_k^* itself is therefore a mixed power law with a uniform distribution.

5.2.4 Random Growth and Divine Performance

An alternative approach to explaining the shape of the assumed god efficiency distribution is to account for how it evolved. We can indeed formalize the distribution π_k by considering it would be based on an evolution of beliefs over time. Since people ask the gods for various things, we will model the fact that some of the wishes are granted (through sheer randomness), and that based on this observation people will update their beliefs on the gods' efficiency. In effect, we will consider that the observed benefits from worshipping each god follow a random growth model (Gabaix 2016.).

We have considered so far that the efficiency of the gods formed a uni-dimensional continuum: their specific jurisdiction did not enter the analysis and only their efficiency at giving *charis* back was important. Now, while we keep the same broad logic, we will also consider that people who sacrifice to a particular god will have framed their requests in the context of that god's specific skills. Hence, two people asking for the same fundamental thing of two separate gods would phrase it in different terms; as a result they would not measure whether their requests were satisfied in the same manner going forward. Random life events hence affect the people who worshipped different gods in different manners. We will also consider there are systemic life events affecting everybody the same way. The more people had started worshipping a god, the more a positive outcome of life events affecting these worshippers would be noticed. In effect,

the total amount of perceived positive outcomes for a given god at some future time is the share of population initially worshipping this god compounded by all the positive outcomes through time.

Let us assume that worship is organized at discrete times $t \in \mathbb{N}$. At time $t = 0$, the gods' efficiency $(\gamma_{k,0})_{k \in [1..K+1]}$ are allocated in some unspecified manner, leading to optimal probabilities $\pi_{k,0}^*$ with a minimum $\pi_{M_0,0}^*$. We consider that the strategies that we described earlier take place without participants planning through time, their horizon is always just that of the current time. At each round, however, the collective assumptions about the pantheon are informed by some events. The events are all fundamentally random, but they are partially colored by people's perception as a function of the gods they most recently worshipped.

For the $K + 1$ gods, systemic life events are represented across the board through time by i.i.d. random variables $S_t \stackrel{(d)}{=} S$ while the perception of god-specific events are $G_{k,t} \stackrel{(d)}{=} G$, for some distributions G and S , where G is positive. These variables capture what people measure (in any unit they deem appropriate) as beneficial life events that they associate to each god. We define a measure of positive life events $p_{k,t}$ associated to each god k at each period t . The systemic events S_t represent things that can not be associated to any particular god specifically and are therefore spread out, while $G_{k,t}$ are associated to each god and therefore compounded by the current level of cumulative positive life events associated to the god. We therefore express the dynamics of that measure as a function of $G_{k,t}$ and S_t as follows:

$$p_{k,t+1} = G_{k,t}p_{k,t} + S_t.$$

At the first period, not knowing anything yet, the "goodness" of each god is simply its relative efficiency, which is equal to the share of people worshipping it (in excess of the least favored god). Hence

$$p_{k,0} = (\pi_{k,0}^* - \pi_{M_0,0}^*) = \frac{\gamma_{k,0}}{2 \sum_{j \in [1..K+1]} \gamma_{j,0}}.$$

Observing how everybody's life is affected as a function of which gods they have worshipped, people set the assumed divine efficiency accordingly, so that simply

$\gamma_{k,t} = p_{k,t}$ at each period⁵⁶. We know that the probabilities $\pi_{k,t}^*$ are then set by normalizing $(\gamma_{k,t})_{k \in [1..K+1]}$ at each step. After time has advanced and the random distribution of the $p_{k,t}$ for each k converges, we consider that the perceptions of the gods' efficiency have been set, the local myths have been expressed and hence these perceptions should be held constant going forward.

With these assumptions, we have the following result.

Proposition 5.4 (Convergence to Power Law). *Assume there exists α such that $\mathbb{E}[G^\alpha] = 1$, $\mathbb{E}[G^\alpha(\ln(G) \vee 0)] < \infty$, and $0 < \mathbb{E}[|S|^\alpha] < \infty$. Also assume that $\frac{S}{1-G}$ is not degenerate and there exists no λ such that the support of $\ln(G)$ conditioned on $G \neq 0$ would be included in $\lambda\mathbb{Z}$. Then:*

- i. *As t goes to infinity, for each $k \in [1..K + 1]$, $\lim_{t \rightarrow \infty} p_{k,t} \stackrel{(p)}{=} \Upsilon$, and Υ asymptotically follows a power law with parameter⁵⁷ α .*
- ii. *As a result, as t goes to infinity, for each $k \in [1..K + 1]$, $(\pi_{k,t}^* - \pi_{M_t,t}^*)$ converges in probability towards a normalized random draw from an approximate power law of coefficient α .*

Proof. A general description of generative approaches to power laws, with applications to economics, can be found in Gabaix (2008) and in Gabaix (1999), and we follow the outline of random growth models presented in that paper.

We use Kesten's results: Theorem B, p. 210 as well as the particular application of Theorem 4, p. 235 in Kesten (1973). The simpler presentation in Goldie (1991), Theorem 2.3, p. 129 and Theorem 4.1 p. 135, makes the theorem's conditions more explicit. The conditions of Proposition 5.4 are particular cases of the theorem's condition for positive random variables G and S . The general result is that the random growth recurrence equation has a steady-state solution and converges towards a distribution that is asymptotically a power law.

⁵⁶Since the distributions G and S are quite arbitrary, they can in fact reflect in any particular relationship one would want to specify between γ and p .

⁵⁷Note that the coefficient is meant as the exponent of the counter cumulative distribution, not of the density, in which we follow the more common convention, unlike that of Clauset, Shalizi, & Newman (2009) or Alstott, Bullmore, & Plenz (2014).

According to the theorem, as t goes to infinity, for each $k \in [1..K + 1]$, $p_{k,t}$ converges in probability to the solution Υ of $\Upsilon \stackrel{(d)}{=} G\Upsilon + S$. In addition, this random variable Υ verifies particular asymptotic conditions. There exists $a > 0$ such that $\lim_{x \rightarrow \infty} x^\alpha \mathbb{P}[\Upsilon > x] = a$. This effectively states that the tail of the distribution of Υ follows a power law of parameter α , that is corresponding to a power law of density $\mathbb{P}[\Upsilon \in dx] = \left(\frac{\alpha}{l_0}\right) \left(\frac{x}{l_0}\right)^{-(1+\alpha)} dx$, if it followed that law entirely and not just on the tail, where l_0 is the strictly positive minimum value the random variable can have.

At the limit when t goes to infinity, each $p_{k,t}$ for all k follows the same distribution with a power law tail. As a result, at the limit, the efficiency $\gamma_{k,t} = p_{k,t}$ for each k is drawn from that distribution Υ . The terms

$$\pi_{k,t}^* - \pi_{M_0,t}^* = \frac{\gamma_{k,t}}{2 \sum_{j \in [1..K+1]} \gamma_{j,t}}$$

therefore simply represent K normalized draws from an approximate power law of parameter α . The probability distribution $(\pi_{k,t}^*)_k$ that is used by people to determine their random choice of which god to worship is therefore itself drawn from the approximate power law. This is the second statement in the Proposition. \square

Note that the proposition does not state that the distribution that people choose to determine which god to worship is a power law, but rather that some numbers are drawn from a power law, then normalized, and then the people randomly choose a god according to that normalized distribution. Proposition 5.1 hence tells us that given any starting conditions, and given a wide range of potential life events distributions affecting people's perception of what the gods pay them as *charis*, we may empirically observe that votive acts appear to be distributed across gods following simple power laws. If a distribution follows a power law, then a sample from that distribution, plotted cumulatively on a log/log scale, is expected to form a straight line.

The parameter for the power law that describes the asymptotic behavior of the god's perceived efficiency is defined by $\mathbb{E}[G^\alpha] = 1$, so that the degree of dispersion of the random events that affect the worshippers of each god will drive the steepness of the power law distribution. For example, if we assume that G follows a uniform distribution over $[0, a]$, then the proposition's conditions are verified if there exists α

such that $a = (1 + \alpha)^{\frac{1}{\alpha}}$, which is decreasing as a function of α . Inverting this expression gives us the parameter α so that the proposition's condition is verified, and we can see that for $a = 2$ we have $\alpha = 1$, and if we reduce a below 2 then α increases: a lower dispersion in the magnitude of the god-specific events corresponds to a steeper limit power law distribution.

5.3 Empirical Analysis of Veneration Acts

Relying on a wealth of data that has been painstakingly gathered by historians over the years, we now turn to the empirical analysis of the model developed above. We will not only directly test the model, but also look for potential explanatory factors behind the specific values driving the differences in divine efficiency.

We discuss our sources in the first subsection, where we explain how the epigraphic and literary data on votive acts can be shaped for our analysis, and merged further with geographic data. Next, we show some of the patterns directly observable in the data and in particular focus on the number of gods worshipped in each *polis*. The third subsection consists of the statistical tests of the model, and in the last subsection we will examine the empirical drivers of the worship act distribution.

As we have stressed earlier, the ancient Greek world was not a unified polity, but a constellation of independent political organizations, the *poleis*. Even though all shared polytheism, they did not necessarily worship the same gods. As we look into worship practices, it is therefore not appropriate to commingle them across *poleis*, but rather to consider each *polis* separately: comparable patterns may be apparent across various *poleis*. In our approach of the data we will therefore always keep this general framework, of not considering Greece as a whole, but rather a set of separate entities. We are effectively considering the information on gods and worship to be structured at the polity level, rather than across the entire geographic space of ancient Greece.

5.3.1 Epigraphic and Literary Sources Combined with *Polis*-level Data

The empirical analysis of ancient Greek religious practice is made difficult by the relative paucity of historical and material sources. There are no systematic records of religious acts, and the historical texts did not consistently keep track of the data we would need

to test the models developed above. One reasonable approach is to focus on epigraphic sources; while they do not represent every single worship act, they can constitute a reasonably consistent set of observations. For our analysis, we use the BDEG data discussed in Chapter 3, in Section 3.2.3. The nature of the data is such that we cannot precisely date the observations. They need to be considered as a whole, spanning a long time period through Antiquity. In the context of the model we developed in the prior section, this effectively corresponds to many effective choices of which gods to worship cumulatively through time.

As we have alluded to before, the names of divinities are not something that is trivially defined. The BDEG tracks in its entries the epicleses associated with the gods, which implicitly defines the “name” of the god as what comes before the sequence of qualifications. In some cases, the name of the god itself is considered to include a qualification, such as *Apollo Heraklea*, for example. We keep that same logic, considering that this manner of addressing the god effectively defines it as a separate entity from simply *Apollo*. Further, gods are sometimes addressed as a list: for example “Zeus and Hera”. In these instances, we split this as two entries into each one of the separate gods. Indeed, we are implicitly assuming that in the cases where rituals were carried out for more than one god at the same time, each god would have obtained their due share of these rituals.

After the processing operations we described above, the underlying BDEG data can be represented as a series of vectors

$$\Delta_i = (\delta_i^{polis}, \delta_i^{NbObs}, \delta_i^{deity}, \dots, \delta_i^N),$$

each corresponding to one of the I entries in the BDEG, with $i \in [1..I]$ and N the number of numeric or categorical characteristics available. The set of unique god names observed across all the data is D , and we will write D_p for the set of unique god names observed for *polis* p . In the dataset, the number of gods in each *polis*, that is $|D_p|$, is called “NumGods”. The set of all *poleis* on which data is available is Π .

With each entry Δ_i , δ_i^{NbObs} represents the number of occurrences of the same votive inscription mentioned in the form. The data does not in fact precisely say how many observations were made, but specifies a range such as “2 to 5”, “>10”, etc. We convert

these categories to numerical approximations. About 2,500 of the 11,000 entries in the data are given such a range.

We write the total number of acts for a given *polis* as $T_p = \sum_{\{i:\delta_i^{polis}=p\}} \delta_i^{NbObs}$. From this, we derive an empirical distribution of votive acts across deities for each *polis* $\pi_{p,d}$ for a *polis* p and a god d as:

$$\pi_{p,d}^e = \frac{1}{T_p} \sum_{\{i:\delta_i^{polis}=p \cap \delta_i^{deity}=d\}} \delta_i^{NbObs}.$$

The gods in D_p can be ranked so that for each *polis* p , they are in descending order. The difference between each god k and the smallest probability (indexed at 0) is $\pi_{p,k}^e - \pi_{p,0}^e$. We call this quantity “NumOccurRatioAdj” in the dataset, and the god d ’s rank in *polis* p , $k(d,p)$, is called “NumOccurRank”. The resulting aggregated data therefore contains $\sum_{p \in \Pi} D_p$ rows.

In addition, some inscription-level or inscription group-level data such as the presence of a cult or the source can be aggregated at the *polis* level. They could not be used in a relevant manner at the deity level, since we need to examine the data in distribution form. For example, the aggregate percentage of epigraphic observations for *polis* p can be computed as follows, where $\delta_i^{epi} \in \{0, 1\}$ is a binary flag representing whether the source is marked as epigraphic:

$$a_p^{epi} = \frac{\sum_{\{i:\delta_i^{polis}=p\}} \delta_i^{epi} \delta_i^{NbObs}}{\sum_{\{i:\delta_i^{polis}=p\}} \delta_i^{NbObs}}.$$

The same calculation can also be carried out for the evidence of a cult where the inscription was found or referenced, a_p^{cult} .

Summary statistics for the BDEG after preprocessing are shown in Table 5.1, across all the data. One can see that many *poleis* only have a handful of observations. The same statistics are shown for the subset of *poleis* for which there were more observations in Table 5.2. We can see that we have some religious information on a little less than 400 different *poleis*, and filtering only for those with a minimal number of observations for very basic distributional analysis, we have about 100.

Table 5.1: Summary Statistics on BDEG Data Across All Poleis

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|-------------------|-----|--------|----------|--------|--------|-------|
| NumGods | 396 | 5.553 | 6.294 | 1 | 3 | 45 |
| NumOccurPolis | 396 | 39.321 | 144.466 | 1 | 5 | 2,374 |
| PctEpi | 396 | 0.804 | 0.323 | 0.000 | 1.000 | 1.000 |
| PctCult | 396 | 0.833 | 0.261 | 0.000 | 0.955 | 1.000 |
| ShareHighestPolis | 396 | 0.625 | 0.290 | 0.135 | 0.562 | 1.000 |
| MinOccurRatio | 396 | 0.404 | 0.409 | 0.0004 | 0.200 | 1.000 |

Note: NumGods: $|D_p|$, number of separate gods observed in the location.
 NumOccurPolis: T_p , number of votive inscriptions in the location.
 PctEpi: a_p^{epi} , average percentage of epigraphic sources in the location.
 PctCult: a_p^{cult} , average percentage of cult signs in the location.
 ShareHighestPolis: $\pi_{p,|D_p|}^e$, percentage of acts towards the most favored god in the location.
 MinOccurRatio: $\pi_{p,0}^e$, percentage of acts towards the least favored god in the location.

Table 5.2: Summary Statistics on BDEG Data On Poleis with more than 25 Observations

| Statistic | N | Mean | St. Dev. | Min | Median | Max |
|-------------------|----|---------|----------|--------|--------|-------|
| NumGods | 93 | 14.301 | 7.200 | 2 | 13 | 45 |
| NumOccurPolis | 93 | 148.806 | 271.402 | 26 | 60 | 2,374 |
| PctEpi | 93 | 0.870 | 0.174 | 0.286 | 0.954 | 1.000 |
| PctCult | 93 | 0.847 | 0.130 | 0.238 | 0.882 | 1.000 |
| ShareHighestPolis | 93 | 0.430 | 0.208 | 0.135 | 0.369 | 0.975 |
| MinOccurRatio | 93 | 0.018 | 0.018 | 0.0004 | 0.017 | 0.148 |

Note: Metrics as defined before.

As we noted earlier, the names of the gods that received the most attention within each *polis* were fairly diverse. For each *polis*, we extracted the name of the most commonly mentioned god, as well as the second most commonly mentioned, and from that data we computed their frequencies across all *poleis*. Table 5.3 shows the twelve most common. We can see that there is substantial variety in which gods were worshipped the most in each *polis*. This further stresses that the particular powers each god was linked to were presumably not the main reason why they were selected.

Table 5.3: Ranking of the First and Second Gods Most Commonly Mentioned in each Polis

| First God | Nb Cities First | Nb Obs First | Second God | Nb Cities Second | Nb Obs Second |
|-----------|-----------------|--------------|------------|------------------|---------------|
| Zeus | 142 | 5624 | Zeus | 52 | 594 |
| Apollon | 51 | 1007 | Apollon | 34 | 1032 |
| Artemis | 41 | 720 | Artemis | 30 | 236 |
| Athena | 29 | 366 | Athena | 23 | 945 |
| Theoi | 23 | 117 | Theoi | 23 | 167 |
| Dionysos | 13 | 165 | Theos | 15 | 57 |
| Theos | 13 | 59 | Aphrodite | 10 | 37 |
| Aphrodite | 9 | 30 | Demeter | 8 | 23 |
| Hermes | 9 | 54 | Dionysos | 8 | 44 |
| Asclepios | 7 | 206 | Hermes | 7 | 14 |
| Demeter | 7 | 15 | Poseidon | 7 | 21 |
| Poseidon | 7 | 37 | Thea | 7 | 27 |

In order to complement the data from the BDEG, we merge it with the POLIS database (T. Johnson & Ober 2014), the computerized and augmented version of Hansen & Nielsen (2004). The majority of *poleis* listed in the POLIS database have geographic coordinates. Hence the mapping of cult epithet observations to a particular *polis* is preferentially done using these coordinates, and when they were not available using the location names⁵⁸.

5.3.2 Analyzing Cult Data

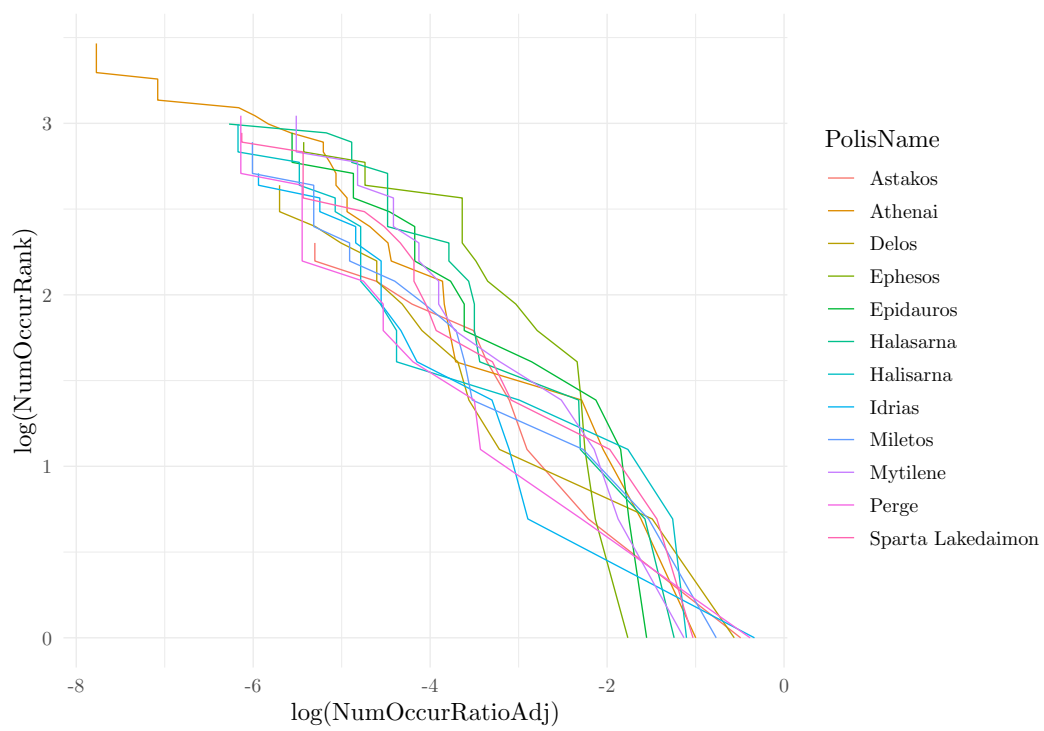
As we have discussed in the first section, we have various reasons to expect some form of power law in the data. The first reason is that people would directly assume this type of distribution for the gods' efficiency, because it is what one observes when looking at the outcomes of high level competitive activities. The second reason is that people would be inferring that some gods are more efficient than others based on some random life events taking place, which would cumulatively lead to a power law-based random draw of votive choices at each period. The first framework would call for a constant power law distribution over time, the second one for a randomly selected distribution over time, following a power law. Since we cannot distinguish time periods precisely, and consider data as a whole, through time, we could not make a difference between these two approaches.

In either case, we would expect some power law relationship, and the most direct way to visualize whether that is the case is illustrated in Figure 5.4, which shows the log/log plot of $\ln(\pi_{p,k}^e - \pi_{p,0}^e)$ as a function of $\ln(k)$, for *poleis* for which there were over 100 observations. All the curves do not have the same length, as the number of gods for which data has been found is not the same for all *poleis*, and the number of gods people worshipped may not be the same either. The curves appear to be fairly straight, which bodes well for our ability to represent the distributions as power laws.

As we mentioned earlier, the number of gods worshipped by people in a given location may depend on a variety of factors, if we consider that there is some threshold under which an ill-favored god would not receive votive acts. Table 5.5 shows several regressions for the number of gods in each *polis*, for which there was data in the sources.

⁵⁸This required a substantial amount of manual mapping of the names, as they are not expressed in the same fashion: typically Greek or English for POLIS, and Greek or French for the BDEG.

Figure 5.4: *Log/log Plot of God Worship Distribution For Various Poleis*



Note: NumOccurAdjRatio is the term $\pi_{p,k}^e - \pi_{p,0}^e$. NumOccurRank is the rank k .

Selecting the significant factors, one can account for a reasonable fraction of this number. Interestingly, the size of the city does not appear to play a very strong role, but the size captured in the POLIS data is an approximation of the *polis*'s surface, not its population (although the two may be related). We know there were substantial variations in the density of *poleis* in classical Greece (Ober 2015). The factors that seem both significant and impactful are “fame” and the share of epigraphic sources for the BDEG data. A greater use of epigraphic sources in a particular place could explain that we have found more remains of votive acts than we would have otherwise. The metric for fame refers to the importance of a *polis* in literary sources, and the greater its importance, the greater number of people might have transited through the place, hence potentially leaving a trace of worship.

Table 5.5: *Empirical Drivers of the Number of Gods For Which References Were Found*

| | Dependent variable: | | |
|-------------------------|---------------------|-------------------------|-----------------------|
| | NumGods | | |
| | Size (1) | OLS All Terms (2) | Selected Terms (3) |
| SizeProxy | 0.015*** (0.004) | -0.0001 (0.002) | 0.003** (0.001) |
| Size | -0.873 (0.759) | | |
| PctEpi | | 5.167*** (1.923) | 3.044*** (0.932) |
| PctCult | | 2.159 (2.335) | |
| Fame | | 1.725*** (0.277) | 1.523*** (0.149) |
| Democracy | | 0.701 (1.357) | |
| Colonies | | 0.051 (0.135) | |
| DelianLeague | | 2.251* (1.206) | |
| Greek | | 1.906 (2.407) | |
| Constant | 4.756*** (1.351) | -7.360** (3.591) | -0.809 (0.864) |
| Observations | 280 | 97 | 280 |
| R ² | 0.171 | 0.605 | 0.408 |
| Adjusted R ² | 0.165 | 0.569 | 0.401 |
| Residual Std. Error | 5.962 | 5.352 | 5.049 |
| F Statistic | 28.606*** | 16.829*** | 63.375*** |

Note:

*p<0.1; **p<0.05; ***p<0.01

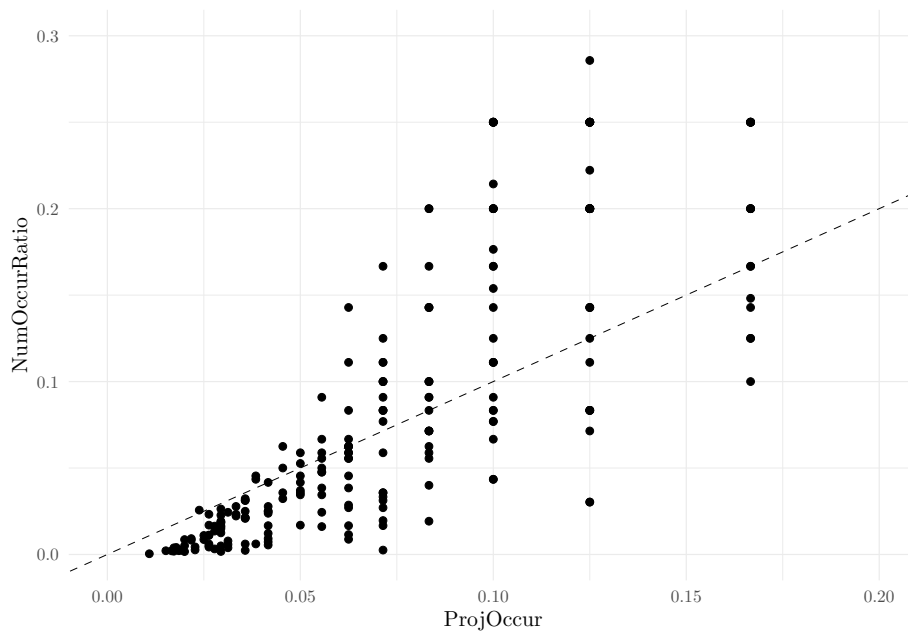
5.3.3 Model Test

The theoretical model we presented earlier offers predictions for two things: the frequency of worship for the least worshipped god (which should be $\frac{1}{2}$ of a random uniform selection), and the difference between the other gods' frequency and that least worshipped god (which is expected to follow a power law).

Figure 5.6 plots for each *polis* p the actual relative frequency $\pi_{p,0}^e$ versus the predicted probability according to the model $\frac{1}{2D_p}$, for all *poleis* with more than 1 god observed.

The actual frequencies appear somewhat lower than the predictions, and the shape of the relationship points to some heteroscedasticity.

Figure 5.6: *Predicted vs. Actual Probability of Worship for the Least Worshipped God*



Note: $\text{ProjOccur} = \frac{1}{2D_p}$, $\text{NumOccurRatio} = \pi_{p,0}^e$.

The regressions of predicted vs. actual values shown in Table 5.7 confirm the visual perception. Since the smaller values for the observations and projections also correspond to the cases where there are many datapoints, it seems appropriate to use the total number of observations for each *polis* as weights to compensate for it at the first order. Then, in order to really weight the data by the number of observations in each *polis*, the weights need to be squared. The coefficients, actually larger than 1, show that there definitely is mixing, whereby the least favored gods still receive a minimal amount of attention. Figure 5.6 shows however that in cases when there were many gods (and therefore a low projected probability), the actual values were lower than projected. This overestimation of the data could be explained by the fact that the inscriptions that have been found are not necessarily comprehensive for the gods which do not appear often, and as a result the number of separate gods derived from the data is likely underestimated.

Table 5.7: Empirical Regressions for Predicted vs. Actual Probability of Worship for the Least Worshipped God

| | Dependent variable: | | |
|-------------------------|----------------------------|--------------------------|--------------------------|
| | NumOccurRatio | | |
| | OLS | | |
| | All data | More than 5 gods | All with weights |
| | (1) | (2) | (3) |
| ProjOccur | 1.922*** (0.054) | 0.955*** (0.046) | 1.146*** (0.051) |
| Observations | 279 | 150 | 279 |
| R ² | 0.817 | 0.744 | 0.640 |
| Adjusted R ² | 0.817 | 0.743 | 0.639 |
| Residual Std. Error | 0.095 (df = 278) | 0.031 (df = 149) | 0.354 (df = 278) |
| F Statistic | 1,244.893*** (df = 1; 278) | 433.630*** (df = 1; 149) | 495.202*** (df = 1; 278) |

Note:

*p<0.1; **p<0.05; ***p<0.01

Now shifting our attention to the other gods who received more favors from the worshippers, we consider several possible models to account for the relationship between $\ln(\pi_{p,k}^e - \pi_{p,0}^e)$ and k . We test a pure power model, a pure exponential, and a power law with exponential decay, since these are the kinds of models that have been used in human performance modeling, and we also know that pure power laws as well as with an exponential decay would be consistent with the random growth model of beliefs.

We use the methodology developed by Clauset, Shalizi, & Newman (2009), but we are facing a particular issue here in that we are not considering a single distribution, but a large number of them, with varying amounts of data. This is addressed by carrying out as many fits as there are *poleis*. For each *polis* p , we will obtain some parameters, for example α_p in the case of a pure power law, fitted to the BDEG data for that *polis*. This is comparable to the way in which we analyzed simulated samples in Chapter 3.

The data consists of the empirical differences (expressed in observations counts, not as relative frequencies) $(\pi_{p,k}^e - \pi_{p,0}^e) T_p$ with the $\pi_{p,k}^e$ which we defined earlier, for each *polis* p . We select only those *poleis* (a little less than a hundred of them) for which there are at least 25 observations in total⁵⁹. Since there are not many observations for each *polis*, we first look at the data across all the *poleis*; we apply the recommended methodology in order to estimate the parameters of various potential distribution choices: exponential⁶⁰, straight power law, and power law with exponential decay. The truncated power law

⁵⁹The descriptive statistics for this subset were shown in Table 5.2.

⁶⁰Which is the upper bound effectively defining a fat tail distribution.

has two parameters, versus only one for the other two, and would hence be likely to better fit the data. We run goodness-of-fit comparison tests between these distributions. For power law and truncated power law fits, the minimum value l_0 is set to 1 (it is not determined by the fit). Table 5.8 shows that the data matches a power law or a truncated power law significantly better than an exponential.

Table 5.8: *Summary Statistics on Distribution Fits Across All Poleis*

| Statistic | Values |
|----------------|--------|
| Lambda Exp | 0.069 |
| Alpha Pow | 0.468 |
| Alpha Trunc | 0.280 |
| Lambda Trunc | 0.007 |
| Trunc vs Pow R | 6.655 |
| Trunc vs Pow p | 0.000 |
| Trunc vs Exp R | 6.895 |
| Trunc vs Exp p | 0.000 |
| Pow vs Exp R | 5.601 |
| Pow vs Exp p | 0.000 |

Note: The distribution names are abbreviated as follows: *Exp* = exponential, *Pow* = (pure) power, *Trunc* = power law with exponential decay. *R*: ratio of goodness-of-fit; a positive number means that the first law of the two is preferred. *p*: significance level; the probability that the preference would be due to randomness.

Now looking at all the *poleis* separately, we carried out similar tests for each one, and the results are summarized in Table 5.9. We can see that the power law coefficient estimates (both for the pure and truncated power laws) are rather tightly concentrated. Although the significance levels for the comparisons across *poleis* are low, in fact the relative dispersion of the parameters is lowest for the pure power law. This could indicate generating mechanisms being quite similar across a large set of different polities. The research on sports and competition rankings such as Deng et al. (2012) and Morales et al. (2016) pointed to some substantial variations in the parameterization of power laws across activities. Therefore, if anthropomorphizing the god's qualities drives their perceived efficiency, one may have expected some degree of variability across space, as driven by differences in political and social environment across cities.

As we mentioned earlier, we have had to apply an approximation to capture the number of votive acts related to entries into the BDEG that referred to multiple

Table 5.9: *Summary Statistics on Distribution Fits across Poleis*

| Statistic | Mean | St. Dev. | Min | Median | Max |
|----------------|--------|----------|---------|--------|--------|
| Lambda Exp | 0.177 | 0.133 | 0.010 | 0.151 | 0.511 |
| Alpha Pow | 0.498 | 0.117 | 0.261 | 0.489 | 0.782 |
| Alpha Trunc | 0.095 | 0.186 | 0.000 | 0.000 | 0.869 |
| Lambda Trunc | 0.065 | 0.072 | 0.000 | 0.040 | 0.392 |
| Trunc vs Pow R | 3.774 | 8.276 | -0.476 | 1.975 | 74.397 |
| Trunc vs Pow p | 0.195 | 0.194 | 0.000 | 0.133 | 0.814 |
| Trunc vs Exp R | -0.089 | 3.664 | -28.140 | 0.344 | 5.238 |
| Trunc vs Exp p | 0.314 | 0.301 | 0.000 | 0.215 | 0.993 |
| Pow vs Exp R | -1.174 | 4.856 | -37.952 | -0.443 | 5.365 |
| Pow vs Exp p | 0.295 | 0.319 | 0.000 | 0.129 | 0.957 |

Note: Metrics as defined before.

observations. Table 5.10 shows the distribution fits on the data compiled excluding these observations, that is only including the entries that referred to a unique occurrence. The slope of the power fit increases in this case, showing that entries in the data that only reflected a single observation tended to be associated with the most commonly worshipped gods. This could be a bias in the manner in which the data was entered, as more attention was potentially dedicated to the inscriptions relating to the most important gods. Nevertheless, the dispersion around this mean slope, across all *poleis*, remains quite low and, in relative terms, it is also minimal for the pure power law.

Table 5.10: *Summary Statistics on Distribution Fits across Poleis Excluding Observations of Multiple BDEG Entries*

| Statistic | Mean | St. Dev. | Min | Median | Max |
|----------------|--------|----------|--------|--------|-------|
| Lambda Exp | 0.428 | 0.208 | 0.114 | 0.394 | 1.099 |
| Alpha Pow | 0.698 | 0.123 | 0.499 | 0.686 | 1.049 |
| Alpha Trunc | 0.142 | 0.243 | 0.000 | 0.000 | 0.977 |
| Lambda Trunc | 0.171 | 0.125 | 0.007 | 0.124 | 0.470 |
| Trunc vs Pow R | 2.629 | 1.697 | -0.411 | 2.381 | 7.771 |
| Trunc vs Pow p | 0.147 | 0.166 | 0.002 | 0.093 | 0.814 |
| Trunc vs Exp R | 0.234 | 1.073 | -2.689 | 0.401 | 2.585 |
| Trunc vs Exp p | 0.458 | 0.288 | 0.007 | 0.408 | 0.964 |
| Pow vs Exp R | -0.729 | 1.380 | -4.499 | -0.635 | 1.978 |
| Pow vs Exp p | 0.398 | 0.300 | 0.000 | 0.367 | 0.986 |

Note: The distribution names are abbreviated as follows:
Exp = exponential, *Pow* = (pure) power, *Trunc* = power law with exponential decay.
R: ratio of goodness-of-fit; a positive number means that the first law of the two is preferred.
p: significance level; the probability that the preference would be due to randomness.

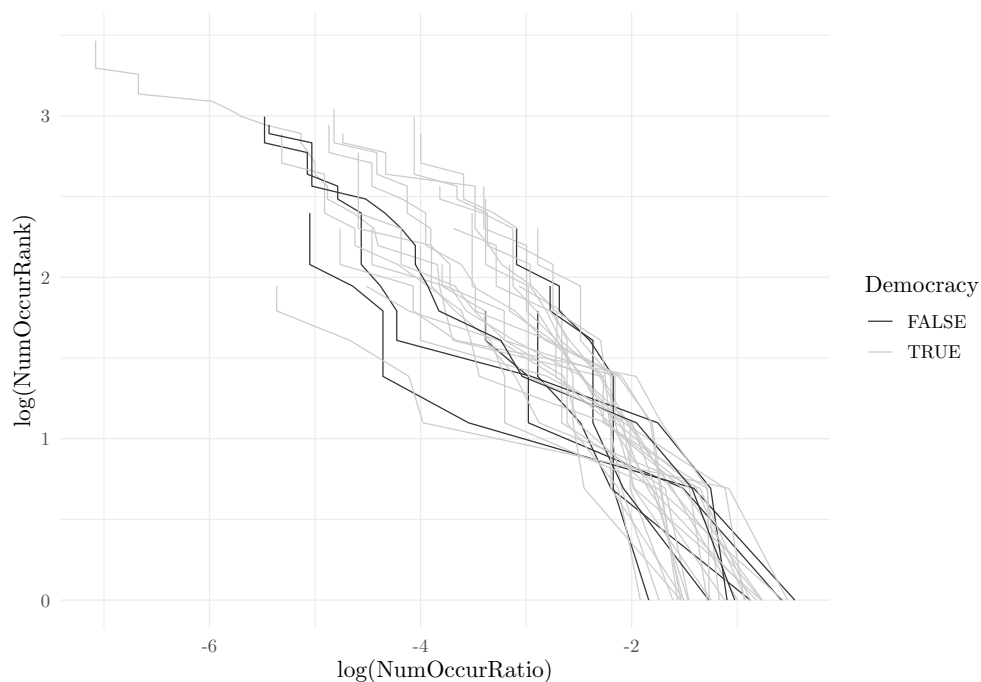
5.3.4 Explaining Differences Across *Poleis*

While we may account for the overall shape of the distribution of votive acts across gods, an interesting question is to determine what in each *polis*'s characteristics may drive

differences in these distributions. Since for each *polis* we possess a series of categorical or numerical variables, we may be able to find to what extent the perceived differences between the gods is impacted by certain features of the *polis*.

Figures 5.11, 5.12 and 5.13 show some illustrations of such a potential relationship, and plot the distributions in a log/log scale respectively as a function of the political regime, the degree of hellenicity, and the size of the *polis*, as categorized by the POLIS data. The curves for non democratic *poleis* may seem flatter than those for democratic regimes, but that is not much more than an impression. *Poleis* for which the population was considered more Hellenic and less varied appear to show steeper curves than these categorized as including more people of barbarian origin. *Polis* size on the other hand does not appear to have much of an impact, on the basis of casual observation.

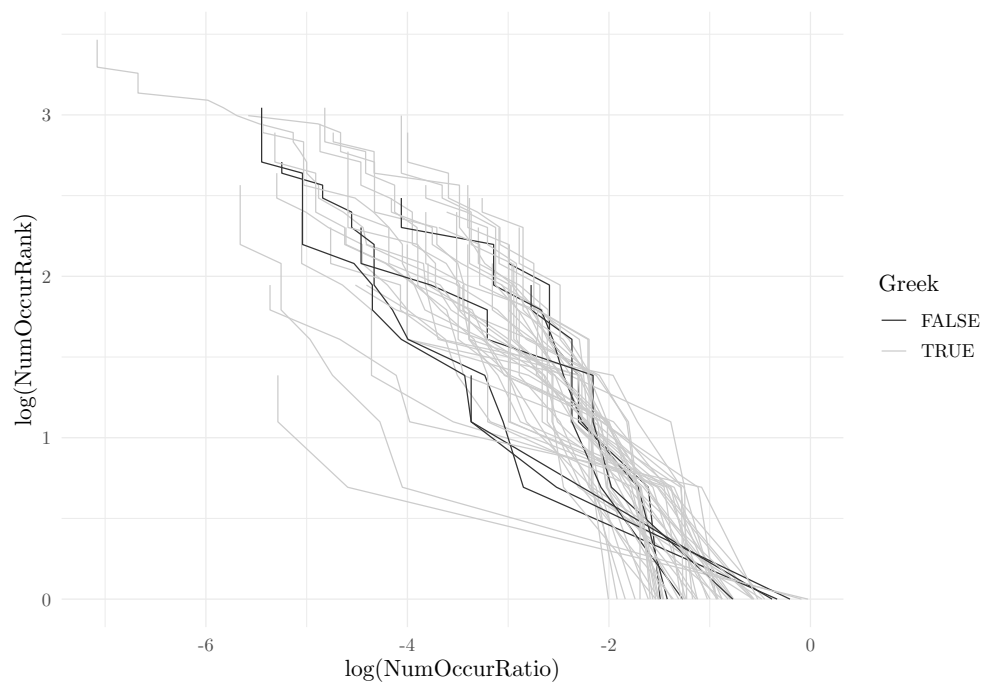
Figure 5.11: *Log/log Plot of God Worship Distribution Grouped by Political Regime*



Note: NumOccurAdjRatio is the term $\pi_{p,k}^e - \pi_{p,0}^e$. NumOccurRank is the rank k . The data only includes *poleis* with more than 25 observations.

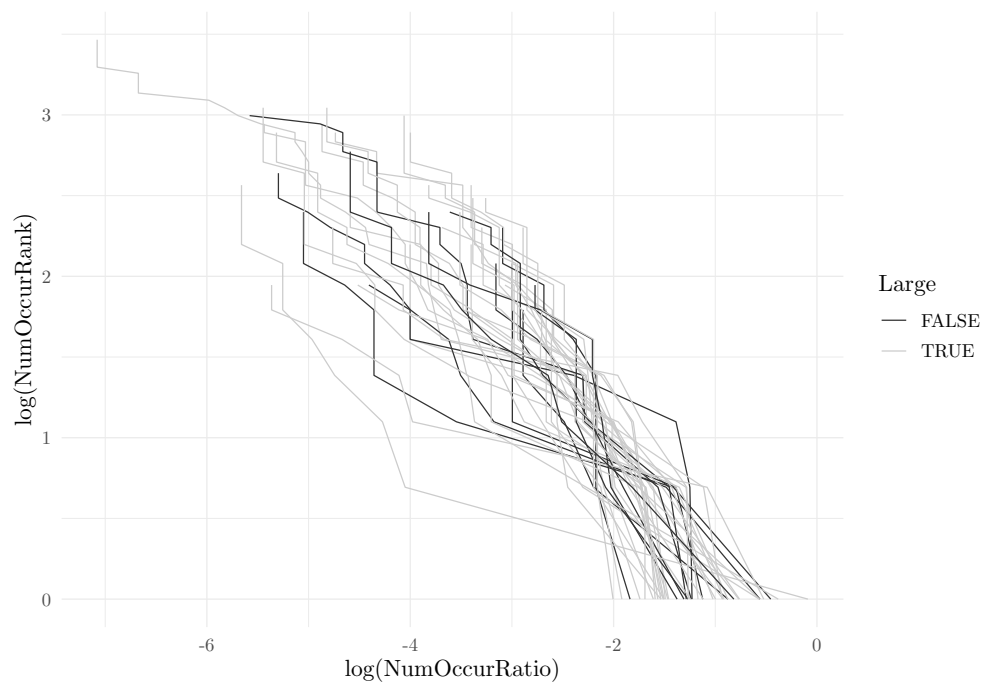
There are many more *polis*-level characteristics that may or may not play a role in empirically explaining the distribution of votive acts. We can use the results from the distribution fits, and try to explain the estimated parameters for various types of distributions across *poleis* as a function of the *poleis*' characteristics. Table 5.14 shows

Figure 5.12: *Log/log Plot of God Worship Distribution Grouped by Hellenicity*



Note: NumOccurAdjRatio is the term $\pi_{p,k}^e - \pi_{p,0}^e$. NumOccurRank is the rank k . The data only includes *poleis* with more than 25 observations.

Figure 5.13: Log/log Plot of God Worship Distribution Grouped by Polis Size



Note: NumOccurAdjRatio is the term $\pi_{p,k}^e - \pi_{p,0}^e$. NumOccurRank is the rank k . The data only includes *poleis* with more than 25 observations.

two of these regressions for the parameter α , for the power law and for the truncated power law. The parameters for these fits are narrowly distributed, and the regression coefficients are small. The R-squares are high, but essentially no variable has a really strong or significant effect.

Table 5.14: *Regressions of Distribution Parameters with All Variables*

| | Dependent variable: | |
|-------------------------|-------------------------|-------------------------------|
| | 'Alpha Pow' | 'Alpha Trunc' |
| | OLS Power law (1) | OLS Truncated Power (2) |
| PctEpi | -0.137 (0.131) | 0.108 (0.159) |
| PctCult | -0.167 (0.180) | -0.336 (0.219) |
| Fame | 0.011 (0.010) | -0.024* (0.012) |
| Democracy | -0.014 (0.065) | -0.264*** (0.079) |
| Size | -0.038* (0.021) | -0.028 (0.026) |
| Colonies | -0.0001 (0.003) | 0.009** (0.004) |
| DelianLeague | -0.037 (0.047) | 0.063 (0.057) |
| Greek | -0.152 (0.204) | 0.479* (0.247) |
| HasVictories | -0.038 (0.056) | 0.004 (0.067) |
| GaveProxeny | -0.037 (0.047) | -0.079 (0.058) |
| HasWalls | 0.125* (0.063) | 0.020 (0.077) |
| Elevation_m | 0.00003 (0.0001) | 0.0003** (0.0001) |
| NumOccurPolis | -0.0001** (0.0001) | 0.0001 (0.0001) |
| Constant | 0.951*** (0.287) | 0.215 (0.348) |
| Observations | 36 | 36 |
| R ² | 0.656 | 0.723 |
| Adjusted R ² | 0.453 | 0.559 |
| Residual Std. Error | 1.006 | 1.221 |
| F Statistic | 3.228*** | 4.410*** |

Note:

*p<0.1; **p<0.05; ***p<0.01

By including so many variables, and with so few observations, we cannot expect a strong robustness in the regressions in Table 5.14. Selecting less explanatory variables, focusing on the ones that do seem to have some significant impact we obtain Table 5.15. With fewer variables, the coefficients are generally significant, and the R-squares remain high. In addition, the signs and magnitudes of the coefficients are somewhat consistent across the pure power fit and the truncated fit. Not all these coefficients appear reasonably robust however. The larger the *polis*, the flatter (more fat-tailed) the distribution; *poleis* considered more Greek also have a flatter curve, and the total number of observations also tends to flatten the curve. The magnitude of the impact of these coefficients is in the order of 0.1-0.2 at the maximum, hence flattening the curve from a steep of about 0.7 down to about 0.4 when the effects are combined on Greek democratic large cities, as opposed to smaller non-democratic *poleis* from the margins.

We can associate the power coefficient to the assumed degree of competitiveness between the gods, in the context of the anthropomorphic perspective that would

Table 5.15: *Regressions of Distribution Parameters with Selected Variables*

| | <i>Dependent variable:</i> | |
|-------------------------|--|--|
| | 'Alpha Pow' OLS Power law (1) | 'Alpha Trunc' OLS Truncated Power (2) |
| Greek | -0.195 (0.178) | 0.345 (0.277) |
| Size | -0.011 (0.014) | -0.062*** (0.021) |
| NumOccurPolis | -0.0001*** (0.00001) | 0.00002 (0.00002) |
| Democracy | -0.013 (0.036) | -0.130** (0.057) |
| Constant | 0.717*** (0.176) | 0.124 (0.274) |
| Observations | 36 | 36 |
| R ² | 0.534 | 0.384 |
| Adjusted R ² | 0.474 | 0.305 |
| Residual Std. Error | 0.986 | 1.532 |
| F Statistic | 8.895*** | 4.838*** |

Note:

*p<0.1; **p<0.05; ***p<0.01

derive the god's perceived efficiency from human performance observation. If we place ourselves in the context of a random growth model explaining the eventual god efficiency, then there would not be any particular reason for some gods to be assumed to be better than other gods, other than pure chance, but the steepness of the power law curve would be related to the degree of dispersion of the random events affecting the perception of each god's efficiency. In light of these considerations, the random growth hypothesis appears the most appropriate to explain the empirical distribution of epigraphic and literary references to cult due to a lack of relationship with the characteristics of the local environment.

5.4 Conclusion

We began this chapter by looking into the way the ancient Greeks considered their relationships with divine beings. Based on several salient elements, we postulated a few principles off of which we could build a simple model relating some particular characteristics of the gods and the propensity that people may have to worship them. We related the number of people worshipping each god to a measure of that god's efficiency. We proposed that this divine efficiency could have been estimated by people based on their observation of human performance. Another potential for the gods' efficiency to evolve was based on random life occurrences in a differential growth model. Using data from the BDEG and from POLIS, we then examined the empirical distribution of votive acts, as captured mainly by epigraphic sources.

As we have seen, the model based on optimal god choice would have called for a somewhat greater mixing of the power law with a uniform than what we observed empirically. Nevertheless, the data did show a substantial enough amount of random mixing in the choice of deities: even the least favored gods did see quite a few offerings or dedications. This is consistent with the intuition of the model, that it is not optimal to overcrowd what is deemed as a relatively limited resource.

The distribution of choices itself across the full range of gods fitted the power law (with or without exponential decay) quite well. The fitted exponent parameters α_p were also surprisingly stable across *poleis*. Searching for explanatory factors, it appeared that not only few of them happened to be significant, but also their magnitude was extremely small. This observation seems consistent with the random growth model, for which there would be no fundamental difference from one *polis* to another. In contrast, one could expect that the particular environment the *polis* experiences could affect the collective perception of human performance, and hence the shape of the empirically observed cult occurrences.

5.5 Chapter Bibliography

- Alstott, Jeff, Ed Bullmore, & Dietmar Plenz.** 2014. "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." *PLOS ONE* 9 (1): e85777.
- Arthur, W. Brian.** 1994. "Inductive Reasoning and Bounded Rationality." *The American Economic Review* 84 (2): 406–411.
- Baek, Seung Ki, Sebastian Bernhardsson, & Petter Minnhagen.** 2011. "Zipf's Law Unzipped." *New Journal of Physics* 13 (4): 043004.
- Bankston, Carl L.** 2002. "Rationality, Choice and the Religious Economy: The Problem of Belief." *Review of Religious Research* 43 (4): 311–325.
- Bonnet, Corinne, ed.** 2021. *Noms de Dieux : Portraits de Divinités Antiques*. Essais - Histoire. Toulouse: Anacharsis.
- Bonnet, Corinne, & Nicole Belayche, eds.** 2017. *Puissances divines à l'épreuve du comparatisme : constructions, variations et réseaux relationnels*. Bibliothèque de l'École des Hautes Études, sciences religieuses, volume 175. Turnhout: Brepols.
- Bonnet, Corinne, & Sylvain Lebreton.** 2019. "Mettre les polythéismes en formules ? À propos de la base de données Mapping Ancient Polytheisms." *Kernos* 32: 267–296.
- Bruit Zaidman, Louise, & Pauline Schmitt Pantel.** 1992. *Religion in the Ancient Greek City*. Cambridge: Cambridge University Press.
- Brulé, Pierre.** 1998. "Le langage des épiclèses dans le polythéisme hellénique (l'exemple de quelques divinités féminines). Quelques pistes de recherches." *Kernos* 11 (11): 13–34.
- . 2005. "Le polythéisme en transformation : les listes de dieux dans les serments internationaux en Grèce antique (V^e-II^e siècles)." In *Nommer les dieux : Théonymes, épithètes, épiclèses dans l'Antiquité*, edited by Nicole Belayche, Pierre Brulé, Gérard Freyburger, Yves Lehman, Laurent Pernot, & Antoine Prost, 143–173. Recherches sur les rhétoriques religieuses. Turnhout: Brepols.
- Brulé, Pierre, & Sylvain Lebreton.** 2007. "La Banque de données sur les épiclèses divines (BDDE) du Crescam : sa philosophie." *Kernos* 20: 217–228.
- Bulbulia, Joseph.** 2010. "Charismatic Signalling." *Journal for the Study of Religion, Nature and Culture* 3 (4): 518–551.
- Bulbulia, Joseph, & Marcus Frean.** 2010. "The Evolution of Charismatic Cultures." *Method & Theory in the Study of Religion* 22 (4): 254–271.
- Canevaro, Mirko, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, eds.** 2018. *Ancient Greek History and Contemporary Social Science*. Edinburgh Leventis Studies 9. Edinburgh: Edinburgh University Press.
- Carr, Jack L., & Janet T. Landa.** 1983. "The Economics of Symbols, Clan Names, and Religion." *The Journal of Legal Studies* 12 (1): 135–156.
- Chakrabarti, Anindya Sundar, Bikas K. Chakrabarti, Arnab Chatterjee, & Manipushpak Mitra.** 2009. "The Kolkata Paise Restaurant Problem and Resource Utilization." *Physica A: Statistical Mechanics and Its Applications* 388 (12): 2420–2426.

- Chakrabarti, Bikas K.** 2007. "Kolkata Restaurant Problem as a Generalised El Farol Bar Problem." In *Econophysics of Markets and Business Networks: Proceedings of the Econophysics-Kolkata III*, edited by Arnab Chatterjee & Bikas K. Chakrabarti, 239–246. New Economic Windows. Milano: Springer.
- Chao, M. T., & W. E. Strawderman.** 1972. "Negative Moments of Positive Random Variables." *Journal of the American Statistical Association* 67 (338): 429–431.
- Chwe, Michael Suk-Young.** 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton: Princeton University Press.
- Clauset, Aaron, Cosma Rohilla Shalizi, & M. E. J. Newman.** 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.
- Deng, Weibing, Wei Li, Xu Cai, Alain Bulou, & Qiuping A Wang.** 2012. "Universal Scaling in Sports Ranking." *New Journal of Physics* 14 (9): 093038.
- Donner, Yoni, & Joseph L. Hardy.** 2015. "Piecewise Power Laws in Individual Learning Curves." *Psychonomic Bulletin & Review* 22 (5): 1308–1319.
- Ekroth, Gunnel.** 2011. "Meat for the Gods." In *Nourrir les dieux ?*, edited by Vinciane Pirenne-Delforge & Francesca Prescendi, 15–41. Liège: Presses Universitaires de Liège.
- Ferrero, Mario, & George Tridimas.** 2018. "Divine Competition in Greco–Roman Polytheism." *Homo Oeconomicus* 35 (3): 143–166.
- Gabaix, Xavier.** 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics* 114 (3): 739–767.
- . 2008. "Power Laws in Economics and Finance." Working Paper 14299. Working Paper Series. Washington: National Bureau of Economic Research. <https://doi.org/10.3386/w14299>.
- . 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30 (1): 185–206.
- Goldie, Charles M.** 1991. "Implicit Renewal Theory and Tails of Solutions of Random Equations." *Annals of Applied Probability* 1 (1): 126–166.
- Gonzales-Feliu, Jesus, & Antoine Parent.** 2016. "Clio-Combinatorics: A Novel Framework to Analyze Military Logistics Choices Using Operations Research Techniques." In *Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations*, edited by Alberto Ochoa-Zezzatti, Jöns Sánchez, Miguel Gastón Cedillo-Campos, & Margain de Lourdes, 79–101. Advances in Logistics, Operations, and Management Science. Hershey: IGI Global.
- Gries, Peter, Jenny Su, & David Schak.** 2012. "Toward the Scientific Study of Polytheism: Beyond Forced-Choice Measures of Religious Belief." *Journal for the Scientific Study of Religion* 51 (4): 623.
- Hansen, Mogens Herman, & Thomas Heine Nielsen.** 2004. *An Inventory of Archaic and Classical Poleis*. Oxford: Oxford University Press.
- Henrich, Joseph.** 2009. "The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution." *Evolution and Human Behavior* 30 (4): 244–260.
- Hotelling, Harold.** 1929. "Stability in Competition." *The Economic Journal* 39 (153): 41–57.
- Iannaccone, Laurence R.** 1998. "Introduction to the Economics of Religion." *Journal of Economic Literature* 36 (3): 1465–1495.

- . 2005. "Economics of Religion. Debating the Costs and Benefits of a New Field." *Faith and Economics* 46: 1–9.
- Iannaccone, Laurence R., Colleen E. Haight, & Jared Rubin.** 2011. "Lessons from Delphi: Religious Markets and Spiritual Capitals." *Journal of Economic Behavior & Organization* 77 (3): 326–338.
- Iyer, Sriya.** 2016. "The New Economics of Religion." *Journal of Economic Literature* 54 (2): 395–441.
- Johnson, Dominic D. P.** 2005. "God's Punishment and Public Goods: A Test of the Supernatural Punishment Hypothesis in 186 World Cultures." *Human Nature* 16 (4): 410–446.
- Johnson, Tim, & Josiah Ober.** 2014. "POLIS." 2014. <http://polis.stanford.edu>.
- Kaše, Vojtěch, Tomáš Hampejs, & Zdeněk Pospíšil.** 2018. "Modeling Cultural Transmission of Rituals in Silico: The Advantages and Pitfalls of Agent-Based Vs. System Dynamics Models." *Journal of Cognition and Culture* 18 (5): 483–507.
- Kesten, Harry.** 1973. "Random Difference Equations and Renewal Theory for Products of Random Matrices." *Acta Mathematica* 131 (0): 207–248.
- Kohli, Rajeev, & Raaj Sah.** 2006. "Some Empirical Regularities in Market Shares." *Management Science* 52 (11): 1792–1798.
- Lebreton, Sylvain.** 2019. "Des dieux et des personnes : une approche par les réseaux (mondes grecs et ouest-sémitiques, ca. 1000 av. n. è. – 400 d. n. è.)." Working Paper. Cinquième rencontre du groupe Res-Hist. Rennes: Université de Rennes. <https://reshist.hypotheses.org/1438>.
- Lebreton, Sylvain, Jean-Baptiste Barreau, Karine Karila-Cohen, & Pierre Brulé.** 2014. "Banque de Données des Épicleses Grecques (BDEG)." 2014. <https://epiclesesgrecques.univ-rennes1.fr/?lang=en>.
- Levy, Gilat, & Ronny Razin.** 2007. "A Theory of Religion: Linking Individual Beliefs." Working Paper 19. 2007 Meeting Papers. 2007 Meeting Papers. London: Society for Economic Dynamics. <https://ideas.repec.org/p/red/sed007/19.html>.
- . 2012. "Religious Beliefs, Religious Participation, and Cooperation." *American Economic Journal: Microeconomics* 4 (3): 121–151.
- . 2014. "Rituals or Good Works: Social Signaling in Religious Organizations." *Journal of the European Economic Association* 12 (5): 1317–1360.
- McBride, Michael.** 2016. "A Rational Choice Theory of Religious Authority." *Rationality and Society* 28 (4): 410–438.
- Miller, Peter J.** 2018. "In the Shadow of Praise: Epinician Losers and Epinician Poetics." *Bulletin of the Institute of Classical Studies* 61 (1): 21–41.
- Mitzenmacher, Michael.** 2004. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2): 226–251.
- Morales, José A., Sergio Sánchez, Jorge Flores, Carlos Pineda, Carlos Gershenson, Germinal Cocho, Jerónimo Zizumbo, Rosalío F. Rodríguez, & Gerardo Iñiguez.** 2016. "Generic Temporal Features of Performance Rankings in Sports and Games." *EPJ Data Science* 5 (1): 33.
- Ober, Josiah.** 2015. *The Rise and Fall of Classical Greece*. Princeton: Princeton University Press.
- Parker, Robert.** 2003. "The Problem of the Greek Cult Epithet." *Opuscula Atheniensia* 28 (2003): 173–183.

- Power, Eleanor A.** 2017. "Discerning Devotion: Testing the Signaling Theory of Religion." *Evolution and Human Behavior* 38 (1): 82–91.
- Ruffle, Bradley J., & Richard Sosis.** 2007. "Does It Pay To Pray? Costly Ritual and Cooperation." *The B.E. Journal of Economic Analysis & Policy* 7 (1).
- Saichev, Alex, Yannick Malevergne, & Didier Sornette.** 2010. *Theory of Zipf's Law and Beyond*. Lecture Notes in Economics and Mathematical Systems 632. Berlin: Springer.
- Smith, Thomas Jay.** 2015. *Variability in Human Performance*. Human Factors and Ergonomics. Boca Raton: CRC Press.
- Vandewalle, Henry.** 2018. "Modelling of Running Performances: Comparisons of Power-Law, Hyperbolic, Logarithmic, and Exponential Models in Elite Endurance Runners." *BioMed Research International* 2018: 1–23.
- Versnel, Henk S.** 2011. *Coping with the Gods: Wayward Readings in Greek Theology*. Religions in the Graeco-Roman World, v. 173. Leiden: Brill.

Conclusion: What Can History Bring to Cliometrics?

The epistemological framework we have proposed in this work effectively took a step away from the kind of political economy research on ancient Greece that has been conducted so far, which we surveyed in Chapter 1, and which we found to be less data-driven than typical cliometrics. Cliometrics *per se* are more focused on the contemporary period, for which more data is available. We argued, in Chapter 2, that cliometrics could expand its range by effectively becoming more historical, by getting closer to the sources with which historians work, thanks to complexity-based approaches. We pointed out that this could allow cliometrics to tackle ancient history. After walking through some of the existing primary data sources for ancient Greece, in Chapter 3, we effectively applied this framework to a selection of particular examples.

However, if we examine what we have effectively done, in practice, it seems like a lot of mathematics and data modeling. What is the difference, then, between the new cliometrics we advocate, those done with a much stronger historical perspective, and the old? Given some data on social or economic issue which we were set to examine, the method we have followed does not differ from what one would do in a contemporary setting. As a matter of fact, delving into names and the choice of names in Chapter 4, we used a fairly generic game-theoretical setup, and the mean-field approach we relied on in order to analyze the underlying data was designed for contemporary name distribution analysis. In the case of polytheist cult analysis, which we addressed in Chapter 5, there was not much pre-existing research to lean on: a natural consequence of the fact that in the contemporary world, polytheist cults not a very actively studied field for which there may be large amounts of data.

The nature of the tools we used, in empirical analyses and in finding a theoretical framework for the problems we tackled, are very similar to the ones we would use in general economics. The very shape of the data, in fact, is also quite comparable: onomastic databases, just like linguistic corpora, work the same way whether they pertain to ancient Greece or today's world. We have also heavily lent on complexity sciences, including dynamic systems and networks, in order to tackle the ancient data. While this allowed us a certain freedom away from straight stylized facts, it is not in and by itself strictly related to history. If the tools we use, at every level of an economics-based gnosiology, are the same for ancient times and for the current world, then, what do historical insights really bring to cliometrics?

Based on this attempt at extending cliometrics towards history, and if we assume that we have effectively offered some new insights into various aspects of ancient Greek society, then we can see that it was not through methodology, or a particular form of data, that this was attained. Therefore, it is clear that we have not advocated to somehow import the historical method into cliometrics; this would presumably simply have turned cliometrics into economic history, as we defined it in Chapter 2. Painting the background for the questions we addressed was important, in order to set the stage properly, and clarify the motivations of all the actors, and it does require some historical knowledge; nevertheless it is not strictly necessary. Where, then, is the fundamentally historical aspect of this work? We argue it resides in the identification of the question to address. While, from an economic perspective, one may want to look for possible applications, illustrations, or confirmations of some theory through historical data, a historically-minded perspective works the other way around: given some historical situation, or simply given some data, what insights could economics provide? Tackling historical material from this angle does require to have an understanding of the historical context, and this is the precise point through which history may help cliometrics expand their range.

Complexity sciences are an indispensable complement to cliometrics in order to efficiently tackle historical sources, as we have seen: an essential tool in handling the "metrics" in cliometrics. The "clio" part of cliometrics, on the other hand, would be best served by a deep dive into the historical material, away from stylized facts, so that new questions can be explored.

Appendix

Data Extraction and Analysis Algorithms

In this appendix, we provide the main computer programs used to generate the data exploited in this work, along with some comments. All these programs are written in R (“R: The R Project for Statistical Computing” 2021).

Diorisis

The main datasets are available from Vatri & McGillivray (2018a). Tauber (2020) offered a critique of the Diorisis data structure as well as some advice on processing it. The data is available as a large list of files, one for each work, in XML format. Each entry represents a word in a sentence, and is associated with lemmatization information. We process all these files in parallel using the following code (which takes about 12 hours on a server), and add some categorization information.

```

1  # Extract Ancient Greek Diorisis textual data
2  library(dplyr)
3  library(tidyverse)
4  library(tidytext)
5  library(stringi)
6  library(xml2)
7  library(TEIdytext)
8  library(parallel)
9  library(reticulate)
10 use_condaenv("r-reticulate")
11 Sys.setenv(KMP_DUPLICATE_LIB_OK = "True")
12 betacodeconv <- import("betacode.conv")
13 diorepath <- "../Data/ExtractDiorisis/Diorisis/"
14 diotempdatadir <- "../Data/ExtractDiorisis/DiorisisRdata/"
15 diodatacleanpath <- "../Data/ExtractDiorisis/alldiorisisclean.Rdata"
16 alldiofl <- list.files(diorepath)
17 alldiofl_done <- as.numeric(gsub('.Rdata', '', list.files(diotempdatadir)))
18 options(fork.allowed = F)
19 numCores <- detectCores() - 5
20 vec_beta_uni <- function(texvec) {
21   ret <- rep(NA, length(texvec))
22   for (ki in 1:length(texvec)) {
23     ret[ki] <- betacodeconv$beta_to_uni(texvec[ki])
24   }
25   return(ret)
26 }
27 process_dior_text <- function(fli) {
28   # Exclude the ones already computed if any

```

```

29 if (!(fli %in% alldiofl_done)) {
30   # List files in directory to identify the one required
31   fle <- alldiofl[fli]
32   # Parse XML
33   texttmp <- TEIdytext::TEIdy(paste0(diorepath, fle))
34   ttl <- texttmp$.text[1]
35   auth <- texttmp$.text[2]
36   # Clean XML
37   cleantxt <- texttmp %>%
38     transmute(
39       Title = ttl,
40       Author = auth,
41       SentenceID = sentence.id,
42       Word = ifelse(!is.na(word.form), vec_beta_uni(word.form), NA),
43       WordNumber = word,
44       WordID = word.id,
45       Lemma = lemma.entry,
46       LemmaPOS = lemma.POS,
47       LemmaID = lemma,
48       Role = analysis.morph,
49       Punctuation = punct.mark,
50       PunctID = punct,
51       Disambiguation = as.numeric(lemma.disambiguated)
52     ) %>%
53     filter(!is.na(Punctuation) | !is.na(Word))
54     save(cleantxt, file = paste0(diotempdatadir, fli, ".Rdata"))
55   }
56   return(0)
57 }
58 # Process all XML files
59 ignore <- mclapply(sample(1:length(alldiofl)), process_dior_text, mc.cores = numCores)
60 # Load and combines tibbles
61 alldiotibb <- list.files(diotempdatadir)
62 alldiorisisraw <- list()
63 for (ffi in alldiotibb) {
64   load(paste0(diotempdatadir, ffi))
65   alldiorisisraw[[ffi]] <- cleantxt
66 }
67 alldiorisis <- bind_rows(alldiorisisraw)
68 # collapse Diorisis for ambiguous words
69 alldiorisisclean <- alldiorisis %>%
70   mutate(
71     Lemma = ifelse(!is.na(Word) & is.na(Lemma), 'unknown', Lemma),
72     LemmaPOS = ifelse(!is.na(Word) & is.na(LemmaPOS), 'unknown', LemmaPOS),
73     Role = ifelse(!is.na(Word) & is.na(Role), 'unknown', Role)) %>%
74   group_by(Title, Author, SentenceID, WordID) %>%
75   mutate(AllRoles = paste(unique(Role), collapse = " | "),
76     NbRoles = n(),
77     RowNb = row_number()) %>%
78   filter(RowNb == 1 | !is.na(Punctuation)) %>%
79   ungroup() %>%
80   mutate(AllRoles = ifelse(!is.na(Punctuation), NA, AllRoles),
81     NbRoles = ifelse(!is.na(Punctuation), NA, NbRoles),
82     SentenceID = as.numeric(SentenceID),
83     WordID = as.numeric(WordID),
84     PunctID = as.numeric(PunctID)) %>%
85   select(-LemmaID, -Role, -RowNb)
86 save(alldiorisisclean, file = diodatacleanpath, compress = "gzip")

```

PHI

The PHI data is represented with one entry for each inscription, that contains the inscription's text as well as some additional geographic information ("PHI Greek Inscriptions," n.d.). Figure 7.16 shows an example of the interface, in this case for the

inscription *IG IX, 2, 56* (part of the *Inscriptiones Graecae*), numbered 147867 in the PHI. We can see here a fairly vague datation, “Roman period”, and some geographic information (Ainis, Hypata).

Figure 7.16: Query for a Particular Entry in the PHI

Regions : Central Greece (IG VII-IX)

IG IX,2 56 ← IG IX,2 1357 IG IX,2 57 →

[] Ainis — Hypata — Roman period — cf. Corrigenda p. IX
See also: SEG 3:457,d.

[-----]

1 πόλεως τοῖς γυμνασιάρχοις Ἀ[ν]-
τάνδρῳ Ὀλυμπίχου δηνάρια χεῖλ[ια]
πεντακόσια, Κασσίῳ Ἐπαφρᾶ δη-
νάρια χεῖλια πεντακόσια εἰς τὴν

5 ἐπισκευὴν τοῦ γυμνασίου, Ἀγα-
θόποδι Ἀγαθόποδος καὶ Λευκίῳ
Τειμοκράτους τοῖς ταμίαι<ς> δην[ά]-
ρια ὀκτακόσια τεσσαράκοντα ἐ-
πτὰ ἥμισυ.

PH147867

αβγ All Regions Search Concordance Browse

The extraction’s method is inspired from the Python code from Assael, Sommerschield, & Prag (2020), who used the PHI for a different purpose, as their goal was to automatically fill in missing text using machine learning (Assael, Sommerschield, & Prag 2019). The extraction of every single inscription web page follows a simple, but brutish, approach, since it just looks for every possible inscription reference number between 1 and 400,000.

```

1 # Extract PHI data
2 library(dplyr)
3 library(parallel)
4 options(fork.allowed = F)
5 locpath <- "./phi_raw_data/"
6 url_text_pattern <- 'https://epigraphy.packhum.org/text/'
7 mxnb_id <- 400000
8 numCores <- detectCores()
9 get_phi <- function(phid) {
10   try({
11     download.file(url = paste0(url_text_pattern, phid), destfile = paste0(locpath, phid))
12   })

```

```

13 }
14 mclapply(1:mxnb_id, get_phi, mc.cores = numCores)

```

The following code shows the parallel processing of the data, once all the raw HTML code for each page has been downloaded. The dates and geographic data are cleaned up and parsed to the extent possible.

```

1  # Process PHI raw html downloads
2  library(rvest)
3  library(dplyr)
4  library(parallel)
5  locpath <- "./phi_raw_data/"
6  locsave <- "./allphi.Rdata"
7  options(fork.allowed = F)
8  numCores <- detectCores() - 1
9  allf <- list.files(locpath)
10 mxnb_id <- max(as.numeric(allf))
11 # Processes the PHI data into tibbles
12 process_phi <- function(fli, path = locpath) {
13   ret <- NULL
14   try({
15     file <- paste0(path, fli)
16     if (file.exists(file)) {
17       pge <- read_html(file)
18       dataref <- pge %>% html_nodes("title") %>% html_text()
19       dataref_clean <- gsub(" - PHI Greek Inscriptions", "", dataref, fixed = T)
20       region_info_raw <- (pge %>% html_nodes("span") %>%
21         html_nodes("[class='name']") %>% html_text())[1]
22       region_detail <- region_info_raw[1]
23       region_broad <- region_info_raw[2]
24       extra_info <- paste(pge %>%
25         html_nodes("[class='small light note']") %>%
26         html_text(), collapse = "|")
27       date_info <- pge %>% html_nodes("[class='ti']") %>% html_text()
28       text_info <- (pge %>% html_nodes("table") %>% html_table(trim = T))[[1]]
29       names(text_info) <- c("line", "text")
30       text_info$Index <- fli
31       text_info$Title <- dataref_clean
32       text_info$BroadRegion <- region_broad
33       text_info$DetailRegion <- region_detail
34       text_info$ExtraInformation <- extra_info
35       text_info$Date <- date_info
36       text_info$line <- as.character(text_info$line)
37       text_info$text <- as.character(text_info$text)
38       ret <- as_tibble(text_info)
39     } else {
40       ret <- NULL
41     }
42   })
43   return(ret)
44 }
45 allphiraw <- mclapply(1:mxnb_id, process_phi, mc.cores = numCores)
46 allphi <- bind_rows(allphiraw) %>%
47   group_by(Index) %>%
48   mutate(
49     LineNb = row_number(),
50     StartLineRef = ifelse(line == "", NA, line),
51     EndLineRef = ifelse(line == "", NA, line)) %>%
52   tidyr::fill(StartLineRef) %>%
53   tidyr::fill(EndLineRef, .direction = "up") %>%
54   ungroup()
55 save(allphi, file = locsave, compress = "gzip")

```

Lemmatization, using the Classical Language ToolKit (K. P. Johnson et al. 2019), is done separately, in a manner that produces an output comparable to that of the Diorisis.

```

1 library(dplyr)
2 library(stringr)
3 library(stringi)
4 library(tidyr)
5 library(tidyverse)
6 library(reticulate)
7 use_condaenv("r-reticulate")
8 Sys.setenv(KMP_DUPLICATE_LIB_OK = "True")
9 scipy <- import("scipy")
10 numpy <- import("numpy")
11 cltk <- import("cltk")
12 greek_accentuation <- import("greek_accentuation")
13 pywarn <- import("warnings")
14 pywarn$filterwarnings('ignore')
15 cltkNormalize <- cltk$corpus$utils$formatter$cltk_normalize
16 BackoffGreekLemmatizer <- cltk$lemmatize$greek$backoff$BackoffGreekLemmatizer
17 lemmatizer_2 <- BackoffGreekLemmatizer(verbose = TRUE)
18 grkbreath <- greek_accentuation$syllabify$rebreath
19 getact <- greek_accentuation$characters$accent
20 stripact <- greek_accentuation$characters$strip_accents
21 locsave <- "./allphi.Rdata"
22 load(locsave)
23 newsave <- "./allphiclean.Rdata"
24 wordssave <- "./allphiwords.Rdata"
25 phidiosave <- "./allphidio.Rdata"
26 catalogsave <- "./phicatalog.Rdata"
27 ## Lemmatize PHI Using CLTK
28 # Vectorize normalize function
29 vec_normalize <- function(textvec) {
30   ret <- rep(NA, length(textvec))
31   for (rvec in 1:length(textvec)) {
32     if (!is.na(textvec[rvec])) {
33       ret[rvec] <- cltkNormalize(textvec[rvec])
34     }
35   }
36   return(ret)
37 }
38 # Logic to clean up second accents in words for lemmatizer
39 clean_encli_accents <- function(somword) {
40   wdchars <- str_extract_all(somword, boundary("character"))[[1]]
41   accts <- rep(F, length(wdchars) )
42   for (ji in 1:length(wdchars)) {
43     axm <- getact(wdchars[ji])
44     accts[ji] <- !is.null(axm)
45   }
46   if (sum(accts) > 1) {
47     wdchars[max(which(accts))] <- stripact(wdchars[max(which(accts))])
48   }
49   return(paste(wdchars, collapse = ""))
50 }
51 clean_encli_accents_vec <- function(textvec) {
52   ret <- rep(NA, length(textvec))
53   for (rvec in 1:length(textvec)) {
54     if (!is.na(textvec[rvec])) {
55       ret[rvec] <- clean_encli_accents(textvec[rvec])
56     }
57   }
58   return(ret)
59 }
60 if (!file.exists(newsave)) {
61   # Clean up text for better processing
62   # The inscriptions are collapsed into one text for each, rather than one entry per line
63   allphiclean <- allphi %>%
64     mutate(
65       GreekText = text) %>%
66     mutate(
67       GreekText = gsub("[...]", "", GreekText)) %>%
68     mutate(
69       GreekText = gsub("[-]", "", GreekText)) %>%
70     mutate(
71       GreekText = gsub("#?", "", GreekText)) %>%
72   # Clean up various special characters generally associated with spacing

```

```

73 mutate(
74   GreekText = gsub("[ ]", " ", GreekText)) %>%
75 mutate(
76   GreekText = gsub("[.]+", " ", GreekText)) %>%
77 mutate(
78   GreekText = gsub("[ ]+", " ", GreekText)) %>%
79 mutate(
80   GreekText = gsub("[{.*}]", "", GreekText)) %>%
81 mutate(
82   GreekText = gsub("[[]]", "", GreekText)) %>%
83 mutate(
84   GreekText = gsub("[ ]", "", GreekText)) %>%
85 mutate(
86   GreekText = gsub("><", "", GreekText)) %>%
87 mutate(
88   GreekText = gsub("[ ]", "", GreekText)) %>%
89 mutate(
90   GreekText = gsub("[\x10140-\x10200]", "", GreekText)) %>%
91 mutate(
92   GreekText = gsub("[%]", " ", GreekText)) %>%
93 mutate(
94   GreekText = gsub("[^ :.]*[0-9][^ :.]*", "", GreekText)) %>%
95 filter(!grepl("vac", GreekText) & !grepl("lacu", GreekText)) %>%
96 # Special case of columns and eta to be replaced with a full stop
97 mutate(
98   GreekText = gsub("\xe2\x88\xb6\xce\x97(\xe2\x88\xb6|$|[ ])", ".", GreekText)) %>%
99 group_by(Index) %>%
100 summarise(
101   Title = first(Title),
102   BroadRegion = first(BroadRegion),
103   DetailRegion = first(DetailRegion),
104   ExtraInformation = first(ExtraInformation),
105   Date = first(Date),
106   NbLines = n(),
107   AllGreekText = paste(GreekText, collapse = " ")) %>%
108 mutate(
109   AllGreekText = gsub("[-] ", "", AllGreekText)) %>%
110 mutate(
111   AllGreekText = gsub("[{.*}]", "", AllGreekText)) %>%
112 mutate(
113   AllGreekText = gsub("[a-gi-zA-Z]", "", AllGreekText)) %>%
114 mutate(
115   AllGreekText = gsub("[²³¹]", "", AllGreekText)) %>%
116 # Dashes
117 mutate(
118   AllGreekText = gsub("\xe2\x80(\x93|\x94)", "", AllGreekText)) %>%
119 mutate(
120   AllGreekText = gsub("[ ]+", " ", AllGreekText)) %>%
121 mutate(
122   AllGreekText = gsub("^[ ]*", "", AllGreekText)) %>%
123 mutate(
124   AllGreekText = gsub("[ ]", "h", AllGreekText)) %>%
125 mutate(
126   AllGreekText = gsub("[ ]", "", AllGreekText)) %>%
127 mutate(
128   AllGreekText = gsub("[ ]+", "", AllGreekText)) %>%
129 mutate(
130   AllGreekText = gsub("[.]+", ".", AllGreekText))
131 save(allphiclean, file = newsave, compress = "gzip")
132 } else {
133   load(newsave)
134 }
135 if (!file.exists(wordssave)) {
136   # Create table of words
137   wrdseps = "(\xe2\x88\xb6|\xc2\xb7|\xce\x87|\xe2\x80\x93|\xcd\xbe|[: ',.;])"
138   allphiwords <- allphiclean %>%
139     select(Index, Title, AllGreekText) %>%
140     separate_rows("AllGreekText", sep = wrdseps) %>%
141     mutate(GreekWords = AllGreekText) %>%
142     select(-AllGreekText) %>%
143     # Clean up special characters
144     mutate(GreekWords = gsub("[^[:graph:]]", "", GreekWords)) %>%
145     mutate(GreekWords = gsub("[.]", "", GreekWords)) %>%

```

```

146   filter(GreekWords != "" & GreekWords != "h") %>%
147   mutate(GreekWords = vec_normalize(tolower(GreekWords)))
148   tpwr<= gsub(".", "", allphiwords %>% select(GreekWords) %>% pull(), fixed = T)
149   tpwr<= rep(NA, length(tpwr))
150   for (ki in 1:length(tpwr)) {
151     if (grepl("^h", tpwr[ki])) {
152       tpwr<= grkbreath(tpwr[ki])
153     } else {
154       tpwr<= tpwr[ki]
155     }
156   }
157   allphiwords <- allphiwords %>%
158   mutate(GreekWords = tpwr<=)
159   tpwr<= lemmatizer_2$lemmatize(clean_encli_accents_vec(tpwr<=))
160   tpwr<= unlist(lapply(lapply(tpwr<=, `[`, i = 3),
161                       function(x) ifelse(is.null(x), NA, x)))
162   tpwr<= unlist(lapply(lapply(tpwr<=, `[`, i = 2),
163                       function(x) ifelse(is.null(x), NA, x)))
164   allphiwords <- allphiwords %>%
165   mutate(Lemma = tpwr<=,
166          Method = ifelse(tpwr<= == "<IdentityLemmatizer>", NA, tpwr<=)) %
167 >%
168   mutate(Method = as.factor(Method))
169   save(allphiwords, file = wordssave, compress = "gzip")
170 } else {
171   load(wordssave)
172 }
173 # Structure the lemmatized corpus in the same way as the Diorisis
174 if (!file.exists(phidiosave)) {
175   # Apostrophes considered as part of the word
176   wrdseps = "((\xe2\x88\xb6|[ ])+)"
177   propseps = "([,])+)"
178   puncseps = "((\xc2\xb7|\xc3\x87|\xcd\xbe|[.:.;])+)"
179   target_punc = "."
180   target_prop = ","
181   gsub8 <- function(pattern, repl, stringvec) {
182     Encoding(stringvec) <- "UTF-8"
183     gsub(pattern, repl, stringvec)
184   }
185   allphibse <- allphiclean %>%
186   select(Index, Title, AllGreekText) %>%
187   # Ensure there are spaces after apostrophes
188   mutate(AllGreekText = gsub("'", " ", AllGreekText, fixed = T)) %>%
189   # Reduce double spaces
190   mutate(AllGreekText = gsub("[ ]+", " ", AllGreekText)) %>%
191   mutate(AllGreekText = gsub8(puncseps, target_punc, AllGreekText)) %>%
192   mutate(AllGreekText = gsub(propseps, target_prop, AllGreekText)) %>%
193   separate_rows("AllGreekText", sep = paste0("[", target_punc, "]")) %>%
194   group_by(Index) %>%
195   mutate(SentenceID = row_number()) %>%
196   ungroup()
197   allphibse <- allphibse %>%
198   bind_rows(allphibse %>%
199     mutate(
200       Punctuation = target_punc,
201       AllGreekText = NA)) %>%
202   arrange(Index, SentenceID)
203   allphibse <- allphibse %>%
204   separate_rows("AllGreekText", sep = paste0("[", target_prop, "]")) %>%
205   group_by(Index, SentenceID) %>%
206   mutate(PropID = row_number()) %>%
207   ungroup()
208   allphibse <- allphibse %>%
209   bind_rows(allphibse %>%
210     filter(is.na(Punctuation)) %>%
211     mutate(
212       Punctuation = target_prop,
213       AllGreekText = NA)) %>%
214   group_by(Index, SentenceID) %>%
215   mutate(PropIDmax = max(PropID)) %>%
216   ungroup() %>%
217   filter(PropID < PropIDmax) %>%
218   dplyr::select(-PropIDmax) %>%

```

```

219   arrange(Index, SentenceID, PropID)
220   allphibsecln <- allphibse %>%
221     separate_rows("AllGreekText", sep = wrdseps) %>%
222     # Clean up non alpha characters after word separation
223     mutate(AllGreekText = gsub("[^[:graph:]]", "", AllGreekText)) %>%
224     mutate(AllGreekText = gsub("[.]", "", AllGreekText)) %>%
225     filter(is.na(AllGreekText) | (AllGreekText != "" & AllGreekText != "h")) %>%
226     rename(Word = AllGreekText) %>%
227     group_by(Index, SentenceID) %>%
228     mutate(WordID = cumsum(!is.na(Word))) %>%
229     ungroup() %>%
230     group_by(Index) %>%
231     mutate(
232       WordNumber = cumsum(!is.na(Word)),
233       PunctID = cumsum(!is.na(Punctuation))) %>%
234     ungroup() %>%
235     mutate(
236       PunctID = ifelse(is.na(Punctuation), NA, PunctID),
237       WordID = ifelse(is.na(Word), NA, WordID),
238       WordNumber = ifelse(is.na(Word), NA, WordNumber))
239   allphibsenrm <- allphibsecln %>%
240     mutate(Word = vec_normalize(Word))
241   tpwr <- gsub(".", "", allphibsenrm %>% select(Word) %>% pull(), fixed = T)
242   tpwr_br <- rep(NA, length(tpwr))
243   for (ki in 1:length(tpwr)) {
244     if (!is.na(tpwr[ki])) {
245       if (grepl("^h", tpwr[ki])) {
246         tpwr_br[ki] <- grkbreath(tpwr[ki])
247       } else {
248         tpwr_br[ki] <- tpwr[ki]
249       }
250     }
251   }
252   allphidio <- allphibsenrm %>%
253     mutate(Word = tpwr_br)
254   tpwr_lem_all <- lemmatizer_2$lemmatize(clean_encli_accents_vec(tolower(tpwr_br)))
255   tpwr_lem_tool <- unlist(lapply(lapply(tpwr_lem_all, `[`, i = 3),
256     function(x) ifelse(is.null(x), NA, x)))
257   tpwr_lem_root <- unlist(lapply(lapply(tpwr_lem_all, `[`, i = 2),
258     function(x) ifelse(is.null(x), NA, x)))
259   allphidio <- allphidio %>%
260     mutate(Lemma = tpwr_lem_root,
261       Method = ifelse(tpwr_lem_tool == "<IdentityLemmatizer>", NA, tpwr_lem_tool)) %
262 >%
263     mutate(Method = as.factor(Method))
264   allphidio <- allphidio %>%
265     mutate(
266       LemmaPOS = NA,
267       AllRoles = NA,
268       # NA Location until fixed
269       Location = NA,
270       LatinizedLemma = stri_trans_general(
271         stri_trans_general(Lemma, id = "any-latin"), id = "Latin-ASCII"),
272       LatinizedWord = stri_trans_general(
273         stri_trans_general(Word, id = "any-latin"), id = "Latin-ASCII")
274     save(allphidio, file = phidiosave, compress = "gzip")
275   } else {
276     load(phidiosave)
277   }
278   # Processes PHI region and date information
279   if (!file.exists(catalogsave)) {
280     dtrg <- allphi %>%
281       distinct(Index, BroadRegion, DetailRegion, Date)
282     regs <- allphi %>%
283       distinct(BroadRegion) %>%
284       mutate(
285         RegionClean = gsub("[ ]*[(.]*)[ ]*", "", BroadRegion) %>%
286         mutate(
287           RegionClean = gsub("[ ]*[[.]*)[ ]*", "", RegionClean) %>%
288         mutate(
289           RegionClean = gsub("^[ ]*", "", RegionClean) %>%
290         mutate(
291           RegionClean = gsub("[ ]*$", "", RegionClean))

```

```

292 regsdet <- allphi %>%
293   distinct(DetailRegion) %>%
294   mutate(
295     AreaClean = gsub("[ ]*([.]*[ ])*", "", DetailRegion)) %>%
296   mutate(
297     AreaClean = gsub("^[ ]*", "", AreaClean)) %>%
298   mutate(
299     AreaClean = gsub("[ ]*$", "", AreaClean))
300 dtrs <- allphi %>% distinct(Date) %>%
301   # Process text at the end of the string
302   mutate(
303     DateExtractBC = str_extract(Date, "[0-9]{1,3}([?]|)[-][0-9]{1,3}([?]|)([ ]|[aA][.][ ]|
304 [bB][cC]|$)")) %>%
305   mutate(
306     DateExtractBC = ifelse(is.na(DateExtractBC),
307                           str_extract(Date, "([ ]|[c][.][ ]|[ ])[0-9]{1,3}([?]|)([/][0-9]|([?]|)([ ]
308 [aA][.][ ]|[ ]|[bB][cC]))"),
309                           DateExtractBC)) %>%
310   mutate(
311     DateExtractBC = ifelse(is.na(DateExtractBC),
312                           str_extract(Date, "(1st|2nd|3rd|4th|5th|6th|7th|8th|9th)(|[/](1st|2nd|3
313 rd|4th|5th|6th|7th|8th|9th))[ ]|[c][.][ ]|[ ]|[aA][.][ ]|[bB][cC]")),
314                           DateExtractBC)) %>%
315   mutate(
316     DateExtractAD = str_extract(Date, "[0-9]{1,3}([?]|)[-][0-9]{1,3}([?]|)[ ]|[Aa][Dd]([
317 .][,][ ]|$|[ ])")) %>%
318   mutate(
319     DateExtractAD = ifelse(is.na(DateExtractAD),
320                           str_extract(Date, "[0-9]{1,3}([?]|)[ ]|[Aa][Dd]([.][,][ ]|$|[ ])"),
321                           DateExtractAD)) %>%
322   mutate(
323     DateExtractAD = ifelse(is.na(DateExtractAD),
324                           str_extract(Date, "(1st|2nd|3rd|4th|5th|6th|7th|8th|9th)(|[/](1st|2nd|3
325 rd|4th|5th|6th|7th|8th|9th))[ ]|[c][.][ ]|[ ]|[aA][dD]([.][,][ ]|$|[ ])"),
326                           DateExtractAD)) %>%
327   mutate(
328     DateExtractBC = gsub("1st", "50", DateExtractBC, fixed = T)) %>%
329   mutate(
330     DateExtractBC = gsub("2nd", "150", DateExtractBC, fixed = T)) %>%
331   mutate(
332     DateExtractBC = gsub("3rd", "250", DateExtractBC, fixed = T)) %>%
333   mutate(
334     DateExtractBC = gsub("4th", "350", DateExtractBC, fixed = T)) %>%
335   mutate(
336     DateExtractBC = gsub("5th", "450", DateExtractBC, fixed = T)) %>%
337   mutate(
338     DateExtractBC = gsub("6th", "550", DateExtractBC, fixed = T)) %>%
339   mutate(
340     DateExtractBC = gsub("7th", "650", DateExtractBC, fixed = T)) %>%
341   mutate(
342     DateExtractBC = gsub("8th", "750", DateExtractBC, fixed = T)) %>%
343   mutate(
344     DateExtractBC = gsub("9th", "850", DateExtractBC, fixed = T)) %>%
345   mutate(
346     DateExtractAD = gsub("1st", "50", DateExtractAD, fixed = T)) %>%
347   mutate(
348     DateExtractAD = gsub("2nd", "150", DateExtractAD, fixed = T)) %>%
349   mutate(
350     DateExtractAD = gsub("3rd", "250", DateExtractAD, fixed = T)) %>%
351   mutate(
352     DateExtractAD = gsub("4th", "350", DateExtractAD, fixed = T)) %>%
353   mutate(
354     DateExtractAD = gsub("5th", "450", DateExtractAD, fixed = T)) %>%
355   mutate(
356     DateExtractAD = gsub("6th", "550", DateExtractAD, fixed = T)) %>%
357   mutate(
358     DateExtractAD = gsub("7th", "650", DateExtractAD, fixed = T)) %>%
359   mutate(
360     DateExtractAD = gsub("8th", "750", DateExtractAD, fixed = T)) %>%
361   mutate(
362     DateExtractAD = gsub("9th", "850", DateExtractAD, fixed = T)) %>%
363   mutate(
364     DateExtractAD = ifelse(is.na(DateExtractAD),

```

```

365         ifelse(grepl("late Roman period", Date), "350-550 AD", NA),
366         DateExtractAD))
367   dtrsclean <- dtrs %>%
368     mutate(DateExtractBC = gsub("[/][0-9]{1,3}", "", DateExtractBC)) %>%
369     mutate(DateExtractBC = gsub("[^0-9]", " ", DateExtractBC)) %>%
370     # Separate raw data into columns
371     separate(DateExtractBC, sep = "[ ]", remove = F, convert = T, into = c("BCFrom", "BCTo"
372 )) %>%
373     mutate(
374       BCFrom = ifelse(is.na(BCFrom), as.numeric(DateExtractBC), BCFrom),
375       BCTo = ifelse(is.na(BCTo), as.numeric(DateExtractBC), BCTo)) %>%
376     mutate(DateExtractAD = gsub("[^0-9]", " ", DateExtractAD)) %>%
377     # Separate raw data into columns
378     separate(DateExtractAD, sep = "[ ]", remove = F, convert = T, into = c("ADFrom", "ADTo"
379 )) %>%
380     mutate(
381       ADFrom = ifelse(is.na(ADFrom), as.numeric(DateExtractAD), ADFrom),
382       ADTo = ifelse(is.na(ADTo), as.numeric(DateExtractAD), ADTo)) %>%
383     mutate(
384       BCFrom = -BCFrom,
385       BCTo = -BCTo) %>%
386     mutate(
387       DateFrom = pmin(BCFrom, BCTo, ADFrom, ADTo, na.rm = T),
388       DateTo = pmax(BCFrom, BCTo, ADFrom, ADTo, na.rm = T)) %>%
389     dplyr::select(Date, DateFrom, DateTo)
390   phicatalog <- dtrg %>%
391     left_join(regs, by = "BroadRegion") %>%
392     left_join(regsdet, by = "DetailRegion") %>%
393     left_join(dtrsclean, by = "Date")
394   save(phicatalog, file = catalogsave, compress = "gzip")
395 } else {
396   load(catalogsave)
397 }

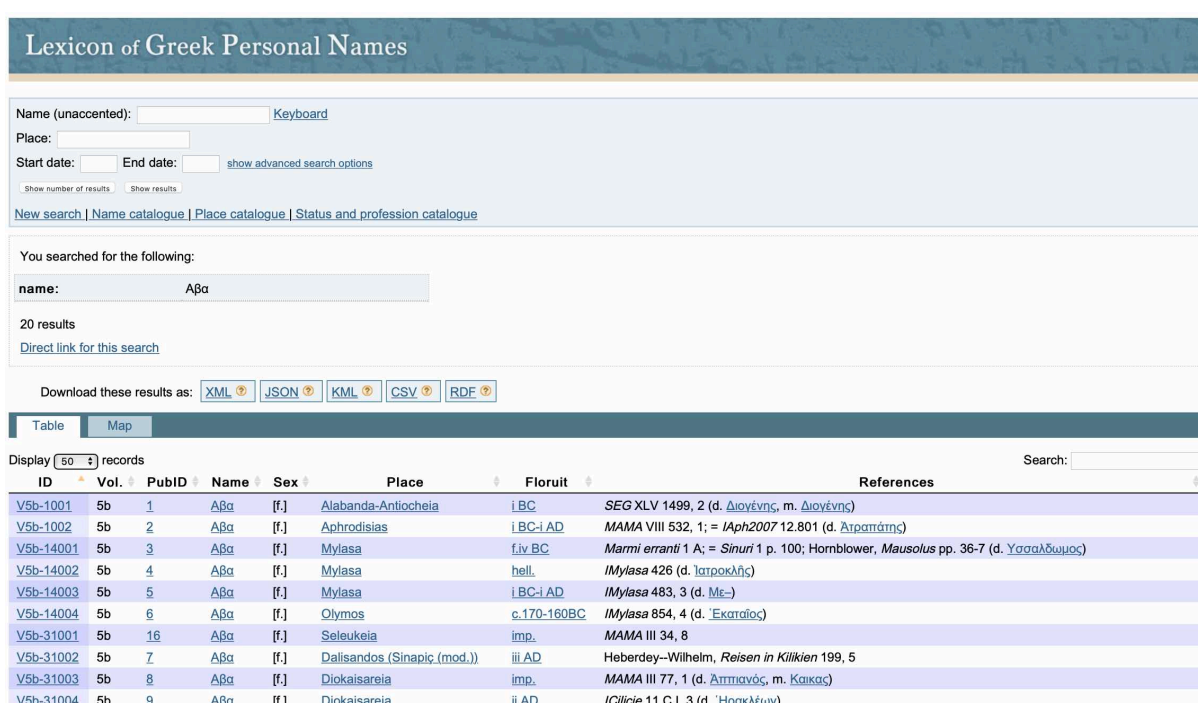
```

LGPN

The electronic interface to the LGPN is not designed for its data to be processed and analyzed in bulk (Parker, Yon, & Depauw 1996). It is designed as a tool to query a name or a name root, and observe its occurrences. Figure 7.17 shows the results of query for a name. This feminine name, $A\beta\alpha$, appears 20 times and for each one of those there is a separate identification, the ID, because each entry may be considered a separate individual. Information is provided on the volume and publication from which this observation is pulled. Some chronological information is given (on the first line, for example, “i BC-i AD” means “within a century before or after CE”), itself typically derived from the epigraphic analysis of the inscription where the name was observed. The reference gives an inscription and a line number. On the first line, “SEG XLV 1499, 2” refers to the *Supplementum Epigraphicum Graecum* volume 45:1499. Finally, the reference field shows the name of the mother and the daughter of this person, based on prosopographic analysis.

One can also see from Figure 7.17 that although it is possible to obtain the data in various formats, this data only pertains to the multiple occurrences of a single name. It is hence impossible to simply download the entire data in a structured form in one batch; it has to be reconstructed from the data pertaining to each possible name. In addition, the various file formats that are illustrated in the figure do not all contain the same information. In particular, the file in CSV format contains the core individual and name data, while the XML file contains data on relationships, locations, and bibliographical references. Both data files hence need to be extracted each time.

Figure 7.17: Query for a Name in the LGPN



Lexicon of Greek Personal Names

Name (unaccented): Keyboard

Place:

Start date: End date: [show advanced search options](#)

[Show number of results](#) [Show results](#)

[New search](#) | [Name catalogue](#) | [Place catalogue](#) | [Status and profession catalogue](#)

You searched for the following:

name: Αβρα

20 results

[Direct link for this search](#)

Download these results as: [XML](#) [JSON](#) [KML](#) [CSV](#) [RDF](#)

Table Map

Display records

| ID | Vol. | PubID | Name | Sex | Place | Floruit | References |
|-----------|------|-------|------|------|-----------------------------|-------------|--|
| V5b-1001 | 5b | 1 | Αβρα | [f.] | Alabanda-Antiocheia | i BC | SEG XLV 1499, 2 (d. Διονένης, m. Διονένης) |
| V5b-1002 | 5b | 2 | Αβρα | [f.] | Aphrodisias | i BC-i AD | MAMA VIII 532, 1; = <i>IAph2007</i> 12.801 (d. Ατραπάρτης) |
| V5b-14001 | 5b | 3 | Αβρα | [f.] | Mylasa | f.iv BC | <i>Marmi erranti</i> 1 A; = <i>Sinuri</i> 1 p. 100; Hornblower, <i>Mausolus</i> pp. 36-7 (d. Υσσαλδωμος) |
| V5b-14002 | 5b | 4 | Αβρα | [f.] | Mylasa | hell. | <i>IMylasa</i> 426 (d. Ιατροκλήτης) |
| V5b-14003 | 5b | 5 | Αβρα | [f.] | Mylasa | i BC-i AD | <i>IMylasa</i> 483, 3 (d. Με-) |
| V5b-14004 | 5b | 6 | Αβρα | [f.] | Olymos | c.170-160BC | <i>IMylasa</i> 854, 4 (d. Έκκραίος) |
| V5b-31001 | 5b | 16 | Αβρα | [f.] | Seleukeia | imp. | MAMA III 34, 8 |
| V5b-31002 | 5b | 7 | Αβρα | [f.] | Dalisandos (Sinapic (mod.)) | iii AD | Heberdey-Wilhelm, <i>Reisen in Kilikien</i> 199, 5 |
| V5b-31003 | 5b | 8 | Αβρα | [f.] | Diokaisareia | imp. | MAMA III 77, 1 (d. Αρριανός, m. Καϊκας) |
| V5b-31004 | 5b | 9 | Αβρα | [f.] | Diokaisareia | ii AD | <i>ICilicie</i> 11 C I, 3 (d. Ηρακλέων) |

In order to programmatically flow through the steps of a manual search, we use the package RSelenium (Harrison & Kim 2020). First, the function “extract_name_list” runs searches for the names starting with every possible letter, in order to generate an up-to-date and comprehensive list of names accessible in the database. Then the functions “extract_all_names_csv” and “extract_all_names_xml” loop through all these names and download the corresponding files (which takes about a day on a fast server).

The specialized function “extract_xml_spec”, given a file name, parses and flattens the XML tree and extracts the information on family relationships, name aliases, bibliographical references, and geographical details. The CSV data is easier to process and does not require specialized treatment. The function “create_single_dt” treats all

the files in parallel processing (this still takes another day on a fast server).

```

1  ## Extraction of data
2  library(dplyr)
3  library(RSelenium)
4  library(rvest)
5  library(wdman)
6  library(stringr)
7  # Store
8  csv_path <- "../Data/ExtractLGPN/CSV_downloads/"
9  xml_path <- "../Data/ExtractLGPN/XML_downloads/"
10 allst_path <- "../Data/ExtractLGPN/allst_names.Rdata"
11 namesdata_path <- "../Data/ExtractLGPN/namesdata.Rdata"
12 relvarsdata_path <- "../Data/ExtractLGPN/relvarsdata.Rdata"
13 namevarsdata_path <- "../Data/ExtractLGPN/namevarsdata.Rdata"
14 placevarsdata_path <- "../Data/ExtractLGPN/placevarsdata.Rdata"
15 biblivarsdata_path <- "../Data/ExtractLGPN/biblivarsdata.Rdata"
16 dwlink_csv <- ";style=csv"
17 dwlink_xml <- ";style=xml"
18 ## General data on Greek names
19 extract_name_list <- function() {
20   # Overall setup
21   lgpn_start_link <- "http://clas-lgpn5.classics.ox.ac.uk:8080/exist/apps/lgpn1-search/index.ht
22 ml"
23   partial_link <- "http://clas-lgpn2.classics.ox.ac.uk/cgi-bin/lgpn_search.cgi?name="
24   # Start browser phantom
25   pJS <- phantomjs()
26   pJS$process
27   remDr <- remoteDriver(browserName = 'phantomjs', port = 4567)
28   remDr$open()
29   remDr$navigate(lgpn_start_link)
30   alllt <- paste0(letters, "")
31   allnms <- list()
32   for (ix in 1:length(alllt)) {
33     try({
34       Sys.sleep(0.5)
35       # Find entry for search
36       myel <- remDr$findElement("id", "namequery")
37       # Send letter wildcard
38       myel$clearElement()
39       myel$sendKeysToElement(list(alllt[ix], key = "enter"))
40       # Get list of names
41       srcnm <- remDr$getPageSource()
42       linklist <- gsub("[?]"$, "", gsub("<", "", gsub(">", "",
43         gsub("&\"", "", gsub("&'", ""),
44         str_extract_all(srcnm, "\\http[:] [/] [/].*\"[>][<]")[1]), fixed = T),
45         fixed = T), fixed = T), fixed = T), fixed = F)
46       linklist <- linklist[grepl(partial_link, linklist, fixed = T)]
47       # Create storage
48       dttmp <- data.frame(Filter = alllt[ix],
49         Name = gsub(partial_link, "", linklist, fixed = T),
50         Link = linklist,
51         stringsAsFactors = F)
52       allnms[[alllt[ix]]] <- dttmp
53     })
54   }
55   remDr$close()
56   pJS$stop()
57   allnames <- bind_rows(allnms)
58   save(allnames, file = allst_path)
59 }
60 # Saving all CSVs
61 extract_all_names_csv <- function() {
62   load(allst_path)
63   cleanallnames <- as_tibble(allnames) %>% distinct(Filter, Name, Link) %>% data.frame()
64   llks <- cleanallnames$Link
65   dirnm <- substr(cleanallnames$Filter, 1, 1)
66   for (flnm in unique(dirnm)) {
67     dir.create(file.path(csv_path, flnm), showWarnings = FALSE)
68   }
69   for (nmi in 1:length(llks)) {

```

```

70   lkdwl <- paste0(llks[nmi], dwlink_csv)
71   try(download.file(lkdwl, destfile = paste0(csv_path, dirnm[nmi], "/", nmi, ".csv"), method
72 = "wget"))
73   if (nmi %% 100 == 0) {
74     Sys.sleep(2)
75   }
76 }
77 }
78 # Saving all XMLs
79 extract_all_names_xml <- function() {
80   load(allst_path)
81   cleanallnames <- as_tibble(allnames) %>% distinct(Filter, Name, Link) %>% data.frame()
82   llks <- cleanallnames$Link
83   dirnm <- substr(cleanallnames$Filter, 1, 1)
84   for (flnm in unique(dirnm)) {
85     dir.create(file.path(xml_path, flnm), showWarnings = FALSE)
86   }
87   for (nmi in 1:length(llks)) {
88     lkdwl <- paste0(llks[nmi], dwlink_xml)
89     try(download.file(lkdwl, destfile = paste0(xml_path, dirnm[nmi], "/", nmi, ".xml"), method
90 = "wget"))
91     if (nmi %% 100 == 0) {
92       Sys.sleep(2)
93     }
94   }
95 }
96 # XML extraction adhoc function
97 extract_xml_spec <- function(flenm, tbtype = NULL) {
98   dta <- data.table(fxml_importXMLFlat(flenm))
99   ix_nms <- dta[elem. == 'person' & attr. == 'id' & level4 == 'listPerson', which = T]
100  if (length(ix_nms) > 0) {
101    ix_nms <- c(ix_nms, dim(dta)[1])
102    for (cnti in 1:(length(ix_nms) - 1)) {
103      dta[ix_nms[cnti]:(ix_nms[cnti + 1] - 1), PersonId := dta[ix_nms[cnti], value.]]
104    }
105    if (ix_nms[length(ix_nms) - 1] < dim(dta)[1]) {
106      dta[dim(dta)[1], PersonId := dta[dim(dta)[1] - 1, PersonId]]
107    }
108    if (tbtype == 'namevars') {
109      # Extract names and variants
110      dta[, PersonName := ifelse(elem. == 'persName' & is.na(attr.) &
111                                level6 == 'persName' & is.na(level7), value., NA)]
112      dta[, PersonNameType := ifelse(elem. == 'persName' & is.na(attr.) &
113                                     level6 == 'persName' & is.na(level7), shift(value., type =
114 'lead'), NA)]
115      namevars <- dta[!is.na(PersonName), list(PersonId, PersonName, PersonNameType)]
116      return(namevars)
117    }
118    if (tbtype == 'biblivars') {
119      # Extract Bib references
120      dta[, Title := ifelse(elem. == 'title' & level6 == 'bibl' & level7 == 'title', value.,
121 NA)]
122      dta[, Biblio := ifelse(elem. == 'title' & level6 == 'bibl' &
123                             level7 == 'title', shift(value., type = 'lag'), NA)]
124      biblivars <- dta[!is.na(Title), list(PersonId, Title, Biblio)]
125      return(biblivars)
126    }
127    if (tbtype == 'placevars') {
128      # Extract place references
129      dta[, PlaceName := ifelse(elem. == 'placeName' & is.na(attr.) & level6 == 'birth', valu
130 e., NA)]
131      dta[, PlaceCode := ifelse(elem. == 'placeName' & is.na(attr.) &
132                                level6 == 'birth', ifelse(shift(attr., n = 1, type = 'lead') == 'key',
133                                                            shift(value., n = 1, type = 'lead'),
134                                                            ifelse(shift(attr., n = 2, type = 'lead') == 'key',
135                                                                    shift(value., n = 2, type = 'lead'), NA)), NA)]
136      placevars <- dta[!is.na(PlaceName), list(PersonId, PlaceName, PlaceCode)]
137      return(placevars)
138    }
139    if (tbtype == 'relvars') {
140      # Extract relationship data
141      dta[, Relationship := ifelse(elem. == 'persName' &
142                                   attr. == 'type' & value. == 'relationship' & level7 == 'label',

```

```

143     shift(value., n = 2, type = 'lag'), NA]]
144     dta[, RelName := ifelse(elem. == 'persName' &
145       attr. == 'type' & value.=='relationship' & level7 == 'label',
146       shift(value., type = 'lag'), NA)]
147     relvars <- dta[!is.na(Relationship), list(PersonId, Relationship, RelName)]
148     return(relvars)
149   }
150 }
151 return(NULL)
152 }
153 ## Loading all files into a single one
154 create_single_dt <- function() {
155   library(readr)
156   library(purrr)
157   library(flatxml)
158   library(furrr)
159   library(data.table)
160   future::plan(multisession)
161   flas <- list.files(csv_path)
162   alldt <- list()
163   for (fla in flas) {
164     print(paste("Doing directory", fla))
165     dirpath <- paste0(csv_path, "/", fla, "/")
166     csv_nms <- list.files(dirpath)
167     alldt[[fla]] <- paste0(dirpath, csv_nms) %>%
168       future_map_dfr(read_csv, col_types = cols("volume" = col_character()))
169   }
170   namesdata <- bind_rows(alldt)
171   save(namesdata, file = namesdata_path)
172   flbs <- list.files(xml_path)
173   alldt_relvars <- list()
174   alldt_namevars <- list()
175   alldt_placevars <- list()
176   alldt_biblivars <- list()
177   for (flb in flbs) {
178     print(paste("Doing directory", flb))
179     dirpath <- paste0(xml_path, "/", flb, "/")
180     xml_nms <- list.files(dirpath)
181     alldt_relvars[[flb]] <- paste0(dirpath, xml_nms) %>%
182       future_map_dfr(extract_xml_spec, tbtype = 'relvars')
183     print("done relvars")
184     alldt_namevars[[flb]] <- paste0(dirpath, xml_nms) %>%
185       future_map_dfr(extract_xml_spec, tbtype = 'namevars')
186     print("done namevars")
187     alldt_placevars[[flb]] <- paste0(dirpath, xml_nms) %>%
188       future_map_dfr(extract_xml_spec, tbtype = 'placevars')
189     print("done placevars")
190     alldt_biblivars[[flb]] <- paste0(dirpath, xml_nms) %>%
191       future_map_dfr(extract_xml_spec, tbtype = 'biblivars')
192     print("done biblivars")
193   }
194   relvarsdata <- bind_rows(alldt_relvars)
195   namevarsdata <- bind_rows(alldt_namevars)
196   biblivarsdata <- bind_rows(alldt_biblivars)
197   placevarsdata <- bind_rows(alldt_placevars)
198   save(relvarsdata, file = relvarsdata_path)
199   save(namevarsdata, file = namevarsdata_path)
200   save(biblivarsdata, file = biblivarsdata_path)
201   save(placevarsdata, file = placevarsdata_path)
202 }

```

The resulting data is structured, in the sense that, for example, the entirety of the names in the LGPN is in the “namesdata” dataframe, and the entirety of all family relationships are in the “relvarsdata” dataframe. However, it still requires further processing in order to be more precisely categorized, and also in order to be made compatible with other data.

In particular, the geographic information obtained from the LGPN comes both in the form of coordinates and as place names. We had to address some issues raised with geographic location with the PHI, about inscriptions, but in the case of the LGPN it is more complex, as there are several layers of interpretation. First, there is a distinction between where an inscription is found and where the people that it refers to used to live, a piece of information that may not be explicit in the source. Further, when ancient locations are referred to in a source, it is not necessarily clear where exactly that location is.

Geographic coordinates provided in an electronic source may hence give an artificial sense of precision, although their semantics are imprecise: on one extreme they could point to the exact location where an inscription was found (and the meaning of that location actually remains to be understood), on the other they could point to the center of a modern city according to Google Maps, because that city is presumably the same as the location referenced in a source. Given these considerations, our pre-processing maps the location information from the LGPN to that of the POLIS catalogue, first based on names and then based on coordinates (using a distance threshold in terms of degrees of latitude and longitude). The names are manually identified, as the code below illustrates.

```

1  ## Process / analysis of LGPN data
2  library(data.table)
3  library(dplyr)
4  library(readxl)
5  library(stringdist)
6  library(fields)
7  source("../Data/ExtractPOLIS/clean_polis.R")
8  namesdata_path <- "../Data/ExtractLGPN/namesdata.Rdata"
9  nameaugdata_path <- "../Data/ExtractLGPN/nameaugdata.Rdata"
10 # Load names data "namesdata"
11 load(namesdata_path)
12 namesdata <- namesdata %>%
13   mutate(CleanSettlement = gsub("-", " ", gsub("[ ]([.]*", "", settlement), fixed = T))
14 # Exclude poleis that do not have Hanssen/ data
15 all_polis <- all_polis[!is.na(all_polis$Polisity), ]
16 # Settlement / Polis clean name mapping
17 # CleanSettlementLGPN / PolisNameMap
18 manual_map <- list(
19   c("Thessalonike", "Thessalonica"),
20   c("Athens", "Athenai"),
21   c("Delphi", "Delphoi"),
22   c("Taras Tarentum", "Tarentum"),
23   c("Chytroi", "Chyton"),
24   c("Gorgippia", "Gorgippeia"),
25   c("Poseidonia Paestum", "Poseidonia"),
26   c("Makyneia", "Makynea"),
27   c("Rhithymna", "Rhithymnos"),
28   c("Syracuse", "Syrakousai"),
29   c("Chersonesos", "Chersonasos"),
30   c("Metropolis", "Matropolis"),
31   c("Polurrhenia", "Polyrhen"),
32   c("Aigai", "Aigeai"),

```

```

33 c("Korinth", "Korinthos"),
34 c("Thisbe", "Thisbai"),
35 c("Alalkomenia", "Alalkomenai"),
36 c("Mantineia Antigoneia", "Mantineia"),
37 c("Hermione", "Hermion"),
38 c("Sybaris Thourioi Copiae", "Thourioi"),
39 c("Myrmekion", "Myrmekeion"),
40 c("Bouthrotos", "Bouthroton"),
41 c("Nysa", "Nyssa"),
42 c("Tralles Seleukeia", "Tralleis"),
43 c("Stratonikeia", "Stratonicea"),
44 c("Knossos", "Knosos"),
45 c("Hierapytna", "Hierpytna"),
46 c("Hyele Velia", "Hyele Elea"),
47 c("Besbykos", "Bysbikos"),
48 c("Hierapolis Kastabala", "Hierapolis"),
49 c("Telmessos", "Telemessos"),
50 c("Hipponion Vibo Valentia", "Hipponion"),
51 c("Gonnoi", "Gonnos"),
52 c("Salymbria", "Selymbria"),
53 c("Pargasa Bargasa", "Bargasa"),
54 c("Taucheira Arsinoe", "Taucheira"),
55 c("Angeiai", "Angeia"),
56 c("Bisanthe Panion", "Bisanthe"),
57 c("Myrleia Apameia", "Myrleia"),
58 c("Phanagoreia", "Phanagoria"),
59 c("Brentesion Brundisium", "Brentesion"),
60 c("Nikaia", "Nicaea"),
61 c("Halaisa", "Alaisa"),
62 c("Koliorgeis", "Koliyrgeis"),
63 c("Kasthanaia", "Kasthanaie"),
64 c("Methana Arsinoe", "Methana"),
65 c("Kytinion", "Kytinion"),
66 c("Megalopolis", "Megale polis"),
67 c("Pythion", "Pythoion"),
68 c("Olympe", "Olympa"),
69 c("Kassope", "Kassopa"),
70 c("Trichonion", "Trichoneion"),
71 c("Euesperides Berenike", "Euhesperides"),
72 c("Thebai", "Thebe"),
73 c("Pyrgoi", "Pyrgos"),
74 c("Phellos", "Phelloe"),
75 c("Pithekoussai Aenaria", "Pithekoussai"),
76 c("Alyzia", "Alyzeia"),
77 c("Bylliones", "Byllis"),
78 c("Isioi", "Issioi"),
79 c("Laodikeia", "Laodicea"),
80 c("Charadros", "Charadrous"),
81 c("Adramyteion", "Adramyttion")
82 )
83 manual_map_df <- as.data.frame(do.call(rbind, manual_map), stringsAsFactors = F)
84 names(manual_map_df) <- c("CleanSettlementLGPN", "PolisNameMap")
85 lgpn_locs <- unique(namesdata$CleanSettlement)
86 lgpn_locs <- lgpn_locs[!is.na(lgpn_locs)]
87 polis_locs <- all_polis[, c("polis_id", "PolisName")]
88 multi_names <- polis_locs %>%
89   group_by(PolisName) %>%
90   summarise(Nb = n()) %>%
91   filter(Nb > 1) %>% as.data.frame()
92 polis_locs <- polis_locs[!(polis_locs$PolisName %in% multi_names$PolisName), ]
93 names(polis_locs) <- c("polis_id_map", "PolisNameMap")
94 lgpn_locs <- lgpn_locs[gsub(" ", "", lgpn_locs) != ""]
95 polis_locs <- polis_locs[gsub(" ", "", polis_locs$PolisNameMap) != "", ]
96 # Compute name string distance
97 matdist <- adist(lgpn_locs, polis_locs$PolisNameMap)
98 min_dist <- apply(matdist, 1, min)
99 pot_maps_nm <- list()
100 for (bdi in 1:length(lgpn_locs)) {
101   pots <- polis_locs[(1:dim(polis_locs)[1])[matdist[bdi, ] == min_dist[bdi]], ]
102   pottmp <- data.frame(CleanSettlementLGPN = lgpn_locs[bdi], MinDistMap = min_dist[bdi], pots)
103   pot_maps_nm[[bdi]] <- pottmp
104 }
105 name_map <- bind_rows(pot_maps_nm)

```

```

106 zero_match <- name_map %>% filter(MinDistMap == 0) %>%
107   select(polis_id_map, PolisNameMap, CleanSettlementLGNP)
108 manual_match <- manual_map_df %>% left_join(polis_locs)
109 lay1_match <- bind_rows(zero_match, manual_match)
110 names(lay1_match) <- c("name_polis_id_map", "name_PolisNameMap", "CleanSettlement")
111 namesdata <- namesdata %>% left_join(lay1_match)
112 # Mapping by geo coordinates and verification
113 ## Map by geo coordinates
114 polis_coord <- all_polis[, c("polis_id", "PolisName", "xcoord", "ycoord")]
115 polis_coord <- polis_coord[!is.na(polis_coord$xcoord), ]
116 names(polis_coord) <- c("geo_polis_id_map", "geo_PolisNameMap", "LongPolis", "LatPolis")
117 matdist <- rdist(namesdata[, c("lat", "long")], polis_coord[, c("LatPolis", "LongPolis")])
118 min_dist <- apply(matdist, 1, min)
119 min_dix <- apply(matdist, 1, which.min)
120 geo_thres <- 0.05
121 namesdata <- namesdata %>%
122   mutate(
123     geo_polis_id_map = ifelse(min_dist < geo_thres, polis_coord$geo_polis_id_map[min_dix], NA),
124     geo_PolisNameMap = ifelse(min_dist < geo_thres, polis_coord$geo_PolisNameMap[min_dix], NA),
125     geo_MinDist = ifelse(min_dist < geo_thres, min_dist, NA))
126 # Synthesizes mapping
127 nameaugdata <- namesdata %>%
128   rename(
129     PersonId = id,
130     PersonName = name) %>%
131   mutate(
132     polis_map_id = ifelse(!is.na(name_polis_id_map), name_polis_id_map, geo_polis_id_map))
133 save(nameaugdata, file = nameaugdata_path)

```

BDEG Data

The BDEG, accessible online, offers a search interface, with which one can identify a particular inscription for a votive act (Lebreton et al. 2014). Figure 7.18 shows an example of the output from an entry in the BDEG database, for Zeus Heliopolites. We can see that there is some identification information, about the location and region, but no particular dating in this case. A source is provided, pointing to the epigraphic publication from which this inscription was obtained.


Since the data is not available in bulk, one needs to access it page by page. The following R script extracts the raw data from the BDEG database, by programmatically running a web browser, in a similar fashion to the way the LGPN data was extracted.

```

1  ## Extraction of data
2  library(RSelenium)
3  library(rvest)
4  library(wdman)
5  # Basic helper
6  unaccent <- function(txt) {
7    txt <- iconv(txt, from="UTF-8", to="ASCII//TRANSLIT//IGNORE")
8    return(gsub("[`~]", "", txt))
9  }
10 unaccent_extra <- function(txt) {
11   txt <- gsub("\\\\", "", txt)
12   gsub("[^]", "", txt)
13 }
14 # Store

```

Figure 7.18: Query for a Particular Entry in the BDEG







Centre de Recherche en
Archéologie, Archéosciences, Histoire

| | |
|---|---|
| <p>Deity Zeus Ζεὺς</p> | <p>Location Place : Hawara (Al-Humayma) Region : Arabie</p> |
| <p>Epithet1 Héliopolitès Ἡλιοπολίτης «Héliopolitain»</p> | |
| <p>Epithet2 Megistos, Kapetôlinos Μεγίστος, Καπετωλίνος «très grand, Capitolin»</p> | |
| Source | |
| J.P. Oleson et alii, ADAJ 43 (1999) p. 411-450. | |
| Source type : Epigraphique | |
| Source date : | |
| <p>Number of occurrences : 1 Evidence for a cult : YES Largely recovered : YES Associated deities : Commentary : autel+dédicace</p> | |
| <p>Fiche importée du fichier : fp_export061106_utf16.csvligne : 5002importé le : 2006-11-07 16:31:21 Fiche n° 5002 création : 2004-06-01 [SL] dernière modification : 2007-05-15 [SL]</p> | |

```

15 allst_path <- "../Data/ExtractBDEG/allst_epiclese.Rdata"
16 allplaces_tb_path <- "../Data/ExtractBDEG/allplaces_tb_epiclese.Rdata"
17 allst_clean_path <- "../Data/ExtractBDEG/allst_clean_epiclese.Rdata"
18 ## General data on epicleses
19 extract_all <- F
20 process_all <- F
21 if (extract_all) {
22   # Overall setup
23   bdeg_start_link <- "https://epiclesesgrecques.univ-rennes1.fr/recherche-fiche_num.php?lang=fr"
24   "
25   # Start browser phantom
26   pJS <- phantomjs()
27   pJS$process
28   remDr <- remoteDriver(browserName = 'phantomjs', port = 4567)
29   remDr$open()
30   remDr$navigate(bdeg_start_link)
31   allst <- list()
32   nb_tot <- 11550
33   nb_start <- 1
34   for (ix in nb_start:nb_tot) {
35     try({
36       Sys.sleep(0.5)
37       # Search the file
38       myel <- remDr$findElement("name", "id")
39       myel$sendKeysToElement(list(as.character(ix), key = "enter"))
40       # Identify and select file
41       myel <- remDr$findElement("link text", as.character(ix))
42       myel$clickElement()
43       allwin <- myel$getWindowHandles()
44       lastwin <- allwin[[length(allwin)]]
45       myel$switchToWindow(lastwin)
46       pgsrc <- myel$getPageSource()[[1]]
47       myel$switchToWindow(lastwin)
48       myel$closeWindow()
49       remDr$switchToWindow(allwin[[1]])
50       allst[[length(allst) + 1]] <- pgsrc
51       if (ix %% 15 == 0) {
52         save(allst, file = allst_path)

```

```

53     }
54   })
55 }
56 save(allst, file = allst_path)
57 remDr$close()
58 pJS$stop()
59 } else {
60   load(allst_path)
61 }
62 ## Extraction of geolocations
63 if (extract_all) {
64   places_list <- "https://epiclesesgrecques.univ-rennes1.fr/liste_lieux.php?lang=fr"
65   allplaces_tb <- read_html(places_list, encoding = "ISO-8859-1") %>% html_table() %>% `[[(1)
66   names(allplaces_tb) <- c("Region", "Location", "CoordRaw")
67   reghd <- (1:dim(allplaces_tb)[1])[allplaces_tb$Region == allplaces_tb$Location &
68     allplaces_tb$Region == allplaces_tb$CoordRaw]
69   reghd[length(reghd) + 1] <- dim(allplaces_tb)[1] + 1
70   for (kix in 1:(length(reghd) - 1)) {
71     allplaces_tb$Region[(reghd[kix]):(reghd[kix + 1] - 1)] <- allplaces_tb$Region[reghd[kix]]
72   }
73   allplaces_tb <- allplaces_tb[allplaces_tb$Region != allplaces_tb$Location &
74     allplaces_tb$Region != allplaces_tb$CoordRaw, ]
75   allplaces_tb$Lat <- as.numeric(gsub("lat: ", "", gsub("[/].*", "", allplaces_tb$CoordRaw)))
76   allplaces_tb$Long <- as.numeric(gsub(" long: ", "", gsub("[/].*", "", allplaces_tb$CoordRaw)
77 ))
78   allplaces_tb$Region <- trimws(gsub("[[:punct:]]", "", unaccent(allplaces_tb$Region)))
79   allplaces_tb$Location <- unaccent(allplaces_tb$Location)
80   allplaces_tb$LocationClean <- trimws(gsub("[[:punct:]]", " ",
81     gsub("[,].*", "", gsub("[[:punct:]]", "", allplaces_tb$Location))))
82   save(allplaces_tb, file = allplaces_tb_path)
83 } else {
84   load(allplaces_tb_path)
85 }
86 ## Data processing
87 if (process_all) {
88   allst_clean <- list()
89   cnt <- 1
90   for (pgi in allst) {
91     pgic <- unaccent(pgi)
92     data_divinite <- gsub("Divinite", "",
93       unaccent_extra(pgic %>% read_html() %>% html_nodes(".divinite") %>%
94         html_text() %>% trimws()))
95     data_location_region <- trimws(gsub(".*[:]", "",
96       unaccent_extra(pgic %>% read_html() %>%
97         html_nodes(".localisation") %>% html_text()))
98     data_location_lieu <- trimws(sub("Region", "",
99       sub("[:].*", "", sub(".*?[:]", "", unaccent_extra(pgic %>%
100         read_html() %>% html_nodes(".localisation") %>% html_text()))))
101     data_epicleses <- trimws(gsub("Epiclese1", "", pgic %>%
102       read_html() %>% html_nodes(".epicleses") %>% html_text()))
103     data_fiche <- as.numeric(gsub("Fiche", "", pgi %>%
104       read_html() %>% html_nodes("title") %>% html_text()))
105     src_raw <- (pgi %>% read_html() %>% html_table()[[1]]
106     src_raw[3, ] <- trimws(gsub("Nature de la source :", "",
107       unaccent_extra(unaccent(src_raw[3, ]))))
108     src_raw[4, ] <- trimws(gsub("Date de la source :", "",
109       unaccent_extra(unaccent(src_raw[4, ]))))
110     data_nature <- paste(src_raw[3, ], collapse = "|")
111     data_date <- paste(src_raw[4, ], collapse = "|")
112     src_raw <- (pgi %>% read_html() %>% html_table()[[2]]
113     data_nb_occure <- trimws(gsub(".*[:]", "", src_raw[1, 1]))
114     data_culte <- grepl("OUI", src_raw[2, 1])
115     # Further cleaning on some
116     data_divinite <- trimws(gsub("[[:punct:]]", "", data_divinite))
117     data_location_region <- trimws(gsub("[[:punct:]]", "", data_location_region))
118     rowtmp <- data.frame(Divinite = data_divinite,
119       LocationRegion = data_location_region,
120       LocationLieu = data_location_lieu,
121       Epicleses = data_epicleses,
122       Fiche = data_fiche,
123       NbOccure = data_nb_occure,
124       Culte = data_culte,
125       Nature = data_nature,

```

```

126         Date = data_date)
127     allst_clean[[length(allst_clean) + 1]] <- rowtmp
128     cnt <- cnt + 1
129 }
130 save(allst_clean, file = allst_clean_path)
131 } else {
132     load(allst_clean_path)
133 }

```

The pre-processing of the data is carried out by the following code, and is mostly concerned with properly mapping place names, and aggregating some data fields:

```

1  ## Preprocessing of data
2  library(data.table)
3  library(dplyr)
4  library(readxl)
5  library(stringdist)
6  library(fields)
7  library(ggplot2)
8  library(stringr)
9  library(tidyr)
10 source("../Data/ExtractPOLIS/clean_polis.R")
11 # Store
12 allst_clean_path <- "../Data/ExtractBDEG/allst_clean_epiclese.Rdata"
13 allplaces_tb_path <- "../Data/ExtractBDEG/allplaces_tb_epiclese.Rdata"
14 load(allst_clean_path)
15 load(allplaces_tb_path)
16 local_bdeg_all <- "../MAP-Gods/ActsofGods/local_bdeg_all.Rdata"
17 all_clean_data <- function() {
18     if (!file.exists(local_bdeg_all)) {
19         ### Combine data
20         # Data on geo loc
21         all_data <- rbindlist(allst_clean)
22         all_data$LocationLieuClean <- trimws(gsub("[:punct:]", " ",
23           gsub("[,].*", "", gsub("[().*]", "", all_data$LocationLieu))))
24         all_data$LocationLieuClean <- gsub("[ ]de[ ].*", "", all_data$LocationLieuClean)
25         all_data$LocationLieuClean <- gsub("[ ]du[ ].*", "", all_data$LocationLieuClean)
26         all_data <- all_data[gsub(" ", "", all_data$LocationRegion) != ""]
27         all_data <- all_data[!grepl("test 2", all_data$LocationRegion)]
28         all_data$NumericalOccurIns <- case_when(
29           all_data$NbOccur == "<5" | all_data$NbOccur == "1<5" ~ 2.5,
30           all_data$NbOccur == "5>10" | all_data$NbOccur == "5<10" ~ 7,
31           all_data$NbOccur == "= 5" ~ 5,
32           all_data$NbOccur == "<10" ~ 7,
33           all_data$NbOccur == ">10" ~ 20,
34           T ~ 1)
35         # Exclude Byzantine period
36         all_data <- all_data[!grepl("yz", Date)]
37         all_places <- data.table(allplaces_tb)
38         # Aggregate to eliminate double entries
39         all_places <- all_places[!is.na(Lat)]
40         all_places <- data.table(all_places %>% group_by(Region, LocationClean) %>%
41           summarise(Lat = mean(Lat), Long = mean(Long)))
42         # Unmapped
43         unmp <- all_data[!(LocationRegion %in% all_places$Region &
44           LocationLieuClean %in% all_places$LocationClean)]
45         # Data on polis
46         all_polis <- data.table(all_polis)
47         # Further transform to match French names
48         all_polis$PolisNameFr <- all_polis$PolisName
49         all_polis$PolisNameFr <- gsub("Syrakousai$", "Syracuse", all_polis$PolisNameFr)
50         all_polis$PolisNameFr <- gsub("Kythera$", "Cythere", all_polis$PolisNameFr)
51         all_polis$PolisNameFr <- gsub("Gytheion$", "Gythion", all_polis$PolisNameFr)
52         all_polis$PolisNameFr <- gsub("Megale polis$", "Megalopolis", all_polis$PolisNameFr)
53         all_polis$PolisNameFr <- gsub("Metapontion$", "Metaponte", all_polis$PolisNameFr)
54         all_polis$PolisNameFr <- gsub("Kromna$", "Kremna", all_polis$PolisNameFr)
55         all_polis$PolisNameFr <- gsub("Olympa$", "Olympie", all_polis$PolisNameFr)
56         all_polis$PolisNameFr <- gsub("Knosos$", "Knossos", all_polis$PolisNameFr)

```

```

57 all_polis$PolisNameFr <- gsub("Phanagoria$", "Phanagoreia", all_polis$PolisNameFr)
58 all_polis$PolisNameFr <- gsub("Delphoi$", "Delphes", all_polis$PolisNameFr)
59 all_polis$PolisNameFr <- gsub("Korinthos$", "Corinthe", all_polis$PolisNameFr)
60 all_polis$PolisNameFr <- gsub("Miletos$", "Milet", all_polis$PolisNameFr)
61 all_polis$PolisNameFr <- gsub("Kalyrna$", "Calymna", all_polis$PolisNameFr)
62 all_polis$PolisNameFr <- gsub("Teuthrania$", "Teuthrone", all_polis$PolisNameFr)
63 all_polis$PolisNameFr <- gsub("Arkades$", "Arcades", all_polis$PolisNameFr)
64 all_polis$PolisNameFr <- gsub("Boias$", "Boiai", all_polis$PolisNameFr)
65 all_polis$PolisNameFr <- gsub("Alalkomenai$", "Alalcomenes", all_polis$PolisNameFr)
66 all_polis$PolisNameFr <- gsub("Plateae$", "Platees", all_polis$PolisNameFr)
67 all_polis$PolisNameFr <- gsub("Karpathos$", "Carpathos", all_polis$PolisNameFr)
68 all_polis$PolisNameFr <- gsub("Korone$", "Coronee", all_polis$PolisNameFr)
69 all_polis$PolisNameFr <- gsub("Poteidaia$", "Potidee", all_polis$PolisNameFr)
70 all_polis$PolisNameFr <- gsub("Telemessos$", "Telmessos", all_polis$PolisNameFr)
71 all_polis$PolisNameFr <- gsub("Potniaia$", "Potnies", all_polis$PolisNameFr)
72 all_polis$PolisNameFr <- gsub("Oenoanda$", "Oinoanda", all_polis$PolisNameFr)
73 all_polis$PolisNameFr <- gsub("Zankle Messana$", "Zankle", all_polis$PolisNameFr)
74 all_polis$PolisNameFr <- gsub("gea$", "gee", all_polis$PolisNameFr)
75 all_polis$PolisNameFr <- gsub("caea$", "cee", all_polis$PolisNameFr)
76 all_polis$PolisNameFr <- gsub("icea$", "icee", all_polis$PolisNameFr)
77 all_polis$PolisNameFr <- gsub("nea$", "nee", all_polis$PolisNameFr)
78 all_polis$PolisNameFr <- gsub("karnassos$", "carnasse", all_polis$PolisNameFr)
79 all_polis$PolisNameFr <- gsub("acus$", "aque", all_polis$PolisNameFr)
80 all_polis$PolisNameFr <- gsub("icus$", "ique", all_polis$PolisNameFr)
81 all_polis$PolisNameFr <- gsub("idus$", "ide", all_polis$PolisNameFr)
82 all_polis$PolisNameFr <- gsub("aleia$", "alie", all_polis$PolisNameFr)
83 all_polis$PolisNameFr <- gsub("inthos$", "inthe", all_polis$PolisNameFr)
84 all_polis$PolisNameFr <- gsub("onica$", "onique", all_polis$PolisNameFr)
85 all_polis$PolisNameFr <- gsub("auros$", "aure", all_polis$PolisNameFr)
86 all_polis$PolisNameFr <- gsub("alos$", "ale", all_polis$PolisNameFr)
87 all_polis$PolisNameFr <- gsub("esia$", "esie", all_polis$PolisNameFr)
88 all_polis$PolisNameFr <- gsub("phesos$", "phese", all_polis$PolisNameFr)
89 all_polis$PolisNameFr <- gsub("amea$", "amee", all_polis$PolisNameFr)
90 all_polis$PolisNameFr <- gsub("alaia$", "alee", all_polis$PolisNameFr)
91 all_polis$PolisNameFr <- gsub("artos$", "arte", all_polis$PolisNameFr)
92 ## Map by geo coordinates
93 polis_coord <- all_polis[, c("polis_id", "PolisName", "xcoord", "ycoord")]
94 polis_coord <- polis_coord[!is.na(polis_coord$xcoord)]
95 names(polis_coord) <- c("polis_id_geo", "PolisNameGeo", "LongPolis", "LatPolis")
96 matdist <- rdist(all_places[, c("Lat", "Long")],
97               polis_coord[, c("LatPolis", "LongPolis")])
98 min_dist <- apply(matdist, 1, min)
99 pot_maps_geo <- NULL
100 for (bdi in 1:dim(all_places)[1]) {
101   pots <- polis_coord[(1:dim(polis_coord)[1])[matdist[bdi, ] == min_dist[bdi]], ][1, ]
102   pottmp <- cbind(all_places[bdi, ], MinDistGeo = min_dist[bdi], pots)
103   pot_maps_geo <- rbind(pot_maps_geo, pottmp)
104 }
105 all_data_merged <- merge(all_data, pot_maps_geo,
106   by.x = c("LocationRegion", "LocationLieuClean"),
107   by.y = c("Region", "LocationClean"), all.x = T)
108 ## Mapping location names between Polis and BDEG
109 bdeg_locs <- unique(unmp$LocationLieuClean)
110 polis_locs <- all_polis[, c("polis_id", "PolisNameFr")]
111 multi_names <- polis_locs %>% group_by(PolisNameFr) %>%
112   summarise(Nb = n()) %>% filter(Nb > 1) %>% as.data.frame()
113 polis_locs <- polis_locs[!(PolisNameFr %in% multi_names$PolisNameFr)]
114 names(polis_locs) <- c("polis_id_map", "PolisNameMap")
115 bdeg_locs <- bdeg_locs[gsub(" ", "", bdeg_locs) != ""]
116 polis_locs <- polis_locs[gsub(" ", "", polis_locs$PolisNameMap) != ""]
117 # Compute name string distance
118 matdist <- adist(bdeg_locs, polis_locs$PolisNameMap)
119 min_dist <- apply(matdist, 1, min)
120 pot_maps_nm <- NULL
121 for (bdi in 1:length(bdeg_locs)) {
122   pots <- polis_locs[(1:dim(polis_locs)[1])[matdist[bdi, ] == min_dist[bdi]], ]
123   pottmp <- data.frame(Bdeg = bdeg_locs[bdi], MinDistMap = min_dist[bdi], pots)
124   pot_maps_nm <- rbind(pot_maps_nm, pottmp)
125 }
126 name_map <- pot_maps_nm[pot_maps_nm$MinDistMap == 0, ]
127 all_data_merged <- merge(all_data_merged, name_map,
128   by.x = c("LocationLieuClean"), by.y = c("Bdeg"), all.x = T)
129 all_data_merged[, PolisId := ifelse(!is.na(polis_id_geo), polis_id_geo, polis_id_map)]

```

```

130 # Clean some god names
131 all_data_merged <- all_data_merged[, Divinite := gsub(' kai ', ' et ', Divinite, fixed =
132 T)]
133 # Split observations when two gods mentioned
134 andstr <- " et "
135 dbles <- all_data_merged %>% filter(grepl(andstr, Divinite, fixed = T)) %>% as_tibble()
136 all_data_merged_nodbl <- all_data_merged[!grepl(andstr, Divinite, fixed = T)]
137 all_data_merged_nodbl[, NbMultipleNames := 0]
138 dbles <- dbles %>%
139   mutate(NbMultipleNames = 1 + str_count(Divinite, andstr)) %>%
140   uncount(NbMultipleNames, .id = "inrank", .remove = F) %>%
141   rowwise() %>%
142   mutate(newDivinite = str_split_fixed(as.character(Divinite), " et ", n = 10)[inrank]) %
143 >%
144   mutate(Divinite = newDivinite) %>%
145   dplyr::select(-newDivinite, -inrank) %>%
146   data.table()
147 all_data_merged_expand <- rbindlist(list(all_data_merged_nodbl, dbles))
148 ## Aggregate by gods
149 # Create aggregated god/polis data
150 all_data_bdeg <- all_data_merged_expand %>% group_by(Divinite, PolisId) %>%
151   summarize(
152     NumericalOccur = ceiling(sum(NumericalOccurIns)),
153     NumericalOccurC1 = sum(ifelse(NumericalOccurIns <= 1, NumericalOccurIns, 0)),
154     NumericalOccurC5 = sum(ifelse(NumericalOccurIns <= 5, NumericalOccurIns, 0)),
155     NumericalOccurC10 = sum(ifelse(NumericalOccurIns <= 10, NumericalOccurIns, 0)),
156     NumericalOccurC50 = sum(ifelse(NumericalOccurIns <= 50, NumericalOccurIns, 0)),
157     NbOccur = n(),
158     LongPolis = mean(LongPolis, na.rm = T),
159     LatPolis = mean(LatPolis, na.rm = T),
160     PctCulte = sum(Culte) / n(),
161     PctEpigraphique = sum(grepl("Epigraphique", Nature) * NumericalOccurIns) /
162       sum(NumericalOccurIns),
163     PctLitteraire = sum(grepl("Litteraire", Nature) * NumericalOccurIns) /
164       sum(NumericalOccurIns)
165   ) %>% data.table()
166 # Create distributions for each Polis
167 all_data_bdeg <- all_data_bdeg[Divinite != '']
168 all_data_bdeg[, NbOccurPolis := sum(NbOccur), by = PolisId]
169 all_data_bdeg[, NumOccurC1Polis := sum(NumericalOccurC1), by = PolisId]
170 all_data_bdeg[, NumOccurC5Polis := sum(NumericalOccurC5), by = PolisId]
171 all_data_bdeg[, NumOccurC10Polis := sum(NumericalOccurC10), by = PolisId]
172 all_data_bdeg[, NumOccurC50Polis := sum(NumericalOccurC50), by = PolisId]
173 all_data_bdeg[, NumOccurPolis := sum(NumericalOccur), by = PolisId]
174 all_data_bdeg[, NumGods := length(unique(Divinite)), by = PolisId]
175 all_data_bdeg[, NbOccurRatio := NbOccur / NbOccurPolis]
176 all_data_bdeg[, NumOccurRatio := NumericalOccur / NumOccurPolis]
177 all_data_bdeg[, NbOccurRank := 1 + NumGods - frank(NbOccur, ties.method = "first"),
178   by = PolisId]
179 all_data_bdeg[, NumOccurRank := 1 + NumGods - frank(NumericalOccur, ties.method = "first")
180 ],
181   by = PolisId]
182 all_data_bdeg[, PctEpi := weighted.mean(PctEpigraphique, NbOccur, na.rm = T), by = PolisId]
183 d]
184 all_data_bdeg[, PctCult := weighted.mean(PctCulte, NbOccur, na.rm = T), by = PolisId]
185 all_data_bdeg[, ShareHighestPolis := sum(NumOccurRatio * (NumOccurRank == 1)), by = Polis
186 Id]
187 all_data_bdeg[, MinOccurRatio := min(NumOccurRatio, na.rm = T), by = PolisId]
188 all_data_bdeg[, MinNumOccur := min(NumericalOccur, na.rm = T), by = PolisId]
189 all_data_bdeg[, MinNumOccurC1 := min(NumericalOccurC1, na.rm = T), by = PolisId]
190 all_data_bdeg[, MinNumOccurC5 := min(NumericalOccurC5, na.rm = T), by = PolisId]
191 all_data_bdeg[, MinNumOccurC10 := min(NumericalOccurC10, na.rm = T), by = PolisId]
192 all_data_bdeg[, MinNumOccurC50 := min(NumericalOccurC50, na.rm = T), by = PolisId]
193 all_data_bdeg[, NumOccurRatioAdj := NumOccurRatio - MinOccurRatio]
194 all_flat_bdeg = all_data_bdeg[NumOccurRank == 1]
195 # Merge with All Polis data
196 all_polis_merged <- merge(all_data_bdeg, all_polis,
197   by.x = c("PolisId"),
198   by.y = c("polis_id"), all.x = T)
199 ## Filters
200 # Exclude empty data and clean up
201 all_polis_merged <- all_polis_merged[!is.na(PolisId)]
202 retbd <- list(

```

```
203     all_polis_merged = all_polis_merged,
204     all_polis = all_polis,
205     all_data_bdeg = all_data_bdeg,
206     all_flat_bdeg = all_flat_bdeg
207 )
208     save(retbd, file = local_bdeg_all)
209 } else {
210     load(local_bdeg_all)
211 }
212     return(retbd)
213 }
```


Résumé long

Cette thèse propose de relier l'analyse économique formelle au détail des sources historiques anciennes, prolongeant ainsi la cliométrie traditionnelle, grâce aux développements récents qui ont associé la cliométrie et l'analyse de la complexité, sous le nom de cliométrie & complexité. Nous structurons notre approche en deux grandes parties.

La première partie, composée de trois chapitres, s'intéresse au cadre épistémologique général dans lequel on peut appliquer la cliométrie à l'histoire ancienne. Dans le premier chapitre, nous nous intéressons aux recherches existantes en économie qui se sont penchées sur l'histoire ancienne, en particulier sur la Grèce antique. Nous passons en revue le développement de la *polis* et de la démocratie grecque antique. Nous examinons la recherche en économie politique qui a fourni un arrière-plan formel à ce développement historique, et soulignons qu'elle appartient aux *analytic narratives* plutôt qu'à la cliométrie. Nous proposons ensuite, dans le chapitre 2, d'étendre la cliométrie, afin qu'elle puisse englober l'histoire ancienne. La cliométrie traditionnelle se concentre généralement sur les données économiques des périodes modernes et contemporaines et n'a pas grand-chose à voir avec l'histoire ancienne, principalement en raison d'un manque de données pertinentes. Dans le chapitre 3, nous prenons le cadre du chapitre 2 et illustrons son application à la culture grecque antique, en passant en revue diverses sources de données importantes dans le processus. En se concentrant sur les sources primaires non économiques, une approche cliométrique permet de cadrer en termes économiques le matériau culturel volumineux et complexe que nous ont légué les Anciens.

La seconde partie, composée des deux chapitres suivants, s'appuie sur la nouvelle épistémologie de la cliométrie ancienne développée dans la première partie, et se

concentre sur l'analyse détaillée économique et cliométrique de facettes particulières de la vie grecque antique. Dans la Grèce antique, les gens portaient un seul nom, et certains noms étaient assez courants, tandis que d'autres étaient très rares. Si ces noms étaient utilisés pour distinguer les gens, pourquoi tout le monde n'avait-il pas un nom différent ? Dans le chapitre 4, en examinant la manière dont les anciens Grecs choisissaient les noms, nous développons un modèle économique de l'existence des noms, comme moyen d'échanger des informations d'identification. Enfin, le panthéon grec est un aspect bien connu de la culture antique ; dans le polythéisme grec ancien, les fidèles pouvaient en effet choisir à quels dieux ils s'adressaient et, ce faisant, ils s'attendaient à une forme de bénéfice dans une relation de contrepartie. Dans le chapitre 5, nous examinons le choix optimal du dieu à adorer, en fonction de la stratégie présumée des dieux pour rendre des faveurs aux fidèles, et le rapportons à une forme de mesure de l'efficacité divine.

8.1 L'économie politique de la Grèce ancienne, cadre historique et revue de la littérature

L'histoire de la Grèce archaïque (VIII^e-VI^e siècle av. J.-C.) et classique (V^e-IV^e siècle av. J.-C.) peut, à bien des égards, s'articuler autour de deux innovations critiques : la création de la *polis*, la cité, et l'émergence subséquente de la *dēmokratia*, la forme particulière de démocratie participative des anciens Grecs. La création de la *polis* est en effet essentielle pour expliquer les développements de la vie sociale et intellectuelle dans la Grèce antique, qui ont conduit à la démocratie athénienne radicale, évolution unique à l'époque. De nombreux travaux de recherche ont cherché à replacer ce processus historique dans un cadre d'économie politique, afin que tant l'émergence d'un État que celle de la démocratie puissent être appréhendées en termes économiques. Puisqu'ils s'efforcent d'analyser la matière historique dans une perspective économique, ces efforts doivent-ils être qualifiés de cliométriques ?

Pour répondre à cette question, nous commençons par définir la *polis*, examinons les conditions qui semblent avoir favorisé son émergence dans la Grèce archaïque, et passons en revue les aspects particuliers sur lesquels l'économie politique s'est concentrée dans son analyse de la *polis* antique. Ensuite, nous nous tournons vers l'émergence de la démocratie et discutons de la recherche économique qui vise à expliquer ses nombreuses

caractéristiques idiosyncratiques dans l'Athènes archaïque et classique. Enfin, nous qualifierons ces approches économiques par opposition à la cliométrie.

8.1.1 La naissance de la *polis*

La cité grecque, la *polis*, peut être définie comme un ensemble de pratiques politiques collectives, portées par des citoyens libres, qui les engagent, et qu'ils respectent tous. En tant que système politique, il définit comment choisir ceux qui décident pour les autres, et tous doivent alors obéir aux décideurs. A quelques exceptions près, les membres de la *polis* qui participent à son fonctionnement politique sont des hommes libres nés de parents eux-mêmes citoyens. Bien que les définitions de ce qu'est la *polis* puissent varier quelque peu, il est généralement admis que son émergence n'est pas liée à une date précise et qu'elle peut être principalement caractérisée comme un processus imperceptible. Pour comprendre et nuancer cette émergence, il faut d'abord s'intéresser aux conditions sociales durant les « âges obscurs » de la Grèce antique, vers le X^e siècle av. J.-C. On peut constater que ces conditions tendaient généralement vers une plus grande égalité, contrairement à l'organisation sociale qui avait prévalu à l'époque mycénienne antérieure.

Entre le XIII^e et le XI^e siècle av. J.-C., près de 90 % des sites mycéniens ont été abandonnés. La destruction des palais mycéniens ou la dissolution de leur pouvoir politique stoppa leur activité de collecte de la production agricole, ce qui entraîna automatiquement de plus grands surplus individuels pour la population. Cela associe aussi plus directement les surplus que chacun obtient à la quantité et à la qualité de la production de ses terres, et transforme une partie de la population en petits propriétaires terriens, soucieux de l'exploitation à long terme de leurs terres. La baisse de la population et la perte de richesse au début de la période ont également contribué à une plus grande égalité sociale et économique. En effet, même s'il y avait des « big men », chefs, aristocrates ou rois, ils n'avaient pas l'apanage d'un grand luxe, ni d'une grande richesse, d'une grande influence ou de la suprématie militaire. Les *basileis*, les rois ou chefs aristocratiques décrits dans les épopées, ne reposaient pas sur une structure politique très solide, comme l'illustrent les difficultés de Télémaque dans l'*Odyssée* : puisque son père, le « big man », n'est pas là, sa puissance n'est pas automatiquement transmise. De plus, Ulysse, qui est décrit dans l'*Odyssée* comme un *basileus* aux ressources

infinies, possède au total une trentaine de troupeaux, un tout autre ordre de grandeur par rapport aux richesses des palais mycéniens. Cette tendance a vraisemblablement été renforcée par la disparition de l'écriture, qui a dû contribuer à une plus grande égalité culturelle et à la réduction d'éventuels attributs de différenciation entre les aristocrates et le reste de la population.

La disparition de la plupart des réseaux commerciaux a d'abord entraîné une pénurie de cuivre et d'étain, qui ont dû être importés. Cela a conduit au développement de la métallurgie du fer, plus difficile mais utilisant un matériau beaucoup plus répandu. La proportion d'objets en bronze dans les sépultures entre les périodes sub-mycénienne et géométrique tardive a diminué (Whitley 1991, 358). Le fer, plus largement disponible que le bronze, facilitait également l'armement des populations : selon la même source, le nombre d'armes par tombe a augmenté au cours de la période. L'accès aux armes n'est donc plus l'apanage d'une classe guerrière. Les outils de travail pouvaient également être améliorés, et avec de meilleurs outils, les rendements agricoles augmentent. Du fait de l'évolution spécifique du milieu social dans la Grèce antique entre le VIII^e et le X^e siècle av. J.-C., on constate que la *polis* n'a pas été créée *ex nihilo*, mais a bénéficié d'un vivier particulier : un nivellement important suivi d'une ascension commune.

Au VIII^e siècle, l'accélération du commerce, la croissance de la population et l'expansion de nombreuses colonies sont souvent considérées comme des marqueurs de l'émergence de la *polis*. La *polis* n'est pas une entité matérielle, mais elle est composée de ses citoyens. C'est une différence fondamentale avec les cités contemporaines autour du monde grec et aussi avec les forteresses qui existaient autrefois. En conséquence, la cité est réellement mobile. Assimiler la *polis* à tous ses citoyens permet aux colonies de jouer un rôle important dans la régulation de la croissance, en plus de permettre l'accès à des ressources éloignées. Souvent, pendant une période difficile, certains habitants de la cité se sont déplacés ailleurs, sans que cela traduise une exclusion ou une punition, et ils ont l'opportunité de créer leur propre cité. En période de troubles civils cette même approche a aussi permis de désamorcer des guerres civiles fratricides.

L'écriture a été redécouverte par les Grecs à travers l'alphabet phénicien au VIII^e siècle. Dans le contexte de l'émergence de la *polis*, où les différences de richesse étaient moins prononcées qu'auparavant, la capacité de lire et d'écrire a pu être plus largement

répartie au sein de la population. Cela se reflète, entre autres, dans le large éventail de textes produits dans la Grèce archaïque. Tous illustrent la large diffusion de l'usage de l'écrit par opposition à l'usage centralisateur du linéaire B, à des fins purement comptables. En particulier, les lois écrites ont rapidement émergé : selon Strabon, les premières lois écrites seraient dues au législateur Zaleucos de Locres au début du VII^e siècle. Écrire des lois suppose qu'une partie suffisante de la population soit capable de les lire, et réduit également l'influence des classes supérieures, qui étaient auparavant les seules capables d'établir ce qui était légitime ou non.

Dans la Grèce archaïque, de nombreux conflits entre cités émergentes ont eu lieu : Hansen & Nielsen (2004) ont identifié des dizaines de destructions totales de cités. Tous les hommes qui pouvaient s'équiper participaient à ces affrontements, grâce à un accès relativement facile aux armes, combattant côte à côte en phalange hoplitique. Ces hommes ne constituaient pas une classe guerrière et leur technique de combat ne nécessitait pas d'entraînement continu. Il s'agissait d'une méthode de combat égalitaire, reposant sur le courage individuel et la cohésion de groupe, par opposition aux duels héroïques, forme de combat affichée dans les textes homériques. Cette égalité effective dans le combat, qui ne se traduit pas nécessairement par une participation politique, entraîne des tensions entre une partie de la population et les aristocrates. La plupart des cités grecques ont généralement évolué dans la seconde moitié de la période archaïque de régimes aristocratiques vers la tyrannie (Hansen & Nielsen 2004), promue par la classe hoplitique. Les tyrans, cependant, obéissaient généralement aux lois. Par exemple, selon Thucydide, le gouvernement d'Hipparque et d'Hippias à Athènes à la fin du VI^e siècle était raisonnable et respectait la multitude, et la cité conservait le plein bénéfice de ses lois existantes.

On peut ainsi clairement identifier des aspects idiosyncratiques dans l'essence de la *polis*, qui sont propres à cette période, et qui permettent de qualifier son émergence de rupture. Ces aspects étaient cependant profondément liés aux conditions sous-jacentes discutées ci-dessus, de sorte que cette rupture n'a pas été soudaine. Il existe un cadre général, en économie politique, pour l'analyse de l'émergence et de la formation de l'État, comme cela est décrit, par exemple, par Laffont (2001), et concernant le fondement de la capacité de l'État, par Besley & Persson (2009) ; la question de l'établissement d'une autocratie a également été examinée par Myerson (2008). L'émergence de la

polis grecque a été examinée à la lumière des principes de l'économie institutionnelle à la fois par des historiens inspirés par les approches économiques (Ober 2015), et par des économistes s'intéressant à l'histoire ancienne (Lyttkens 2013), axés à la fois sur le contexte général et sur les aspects spécifiques de la société grecque antique. Des conditions particulières sont requises pour l'émergence de l'état de droit, selon Fleck & Hanssen (2019) : par exemple, un compromis est nécessaire, afin que chacun s'engage à suivre une règle commune, et un processus de décision majoritaire minimise le coût de ce compromis. En particulier, Pitsoulis (2011) propose un modèle qui saisit l'impact du rôle accru des hoplites sur la naissance d'un système égalitaire. L'émergence d'un nouveau paysage politique, tourné vers l'égalité, peut aussi être liée aux idéaux communs liés à l'athlétisme et à la guerre (Kyriazis & Economou 2015; Economou & Kyriazis 2017). Le rôle et le coût de la guerre dans l'Athènes classique étaient effectivement tous deux très importants (Lyttkens & Gerding 2022), et le fait de devoir prendre des décisions relatives à la guerre peut justifier l'émergence d'un processus administrant le choix social (Economou & Kyriazis 2016). Tridimas (2015) voit une relation entre la guerre et la démocratie : en s'appuyant sur les pauvres pour équiper la flotte de guerre, Athènes a effectivement redistribué des ressources à leur avantage, ce qui induit une préférence du régime démocratique pour la guerre.

L'évolution de l'organisation politique dans la Grèce antique a été associée à la croissance économique et au potentiel de croissance. Fleck & Hanssen (2006) ont utilisé l'économie institutionnelle pour analyser l'émergence de la démocratie dans diverses cités grecques, comme moyen de favoriser l'investissement personnel dans l'agriculture. Avec un cadre comparable, ils ont également considéré la transition vers la tyrannie et l'ont liée au potentiel de croissance (Fleck & Hanssen 2013, 2018). Soulignant une forme de cercle vertueux, ils ont montré dans Fleck & Hanssen (2015) que l'ancien système démocratique grec soutenait efficacement la création de richesse. Avec une autre perspective sur la création de la *polis* démocratique, McCannon (2012) a considéré que c'était un moyen d'atténuer l'augmentation de la volatilité de la richesse individuelle à travers les générations.

8.1.2 L'invention de la *dēmokratia*

La démocratie athénienne, selon Ismard (2005), peut être définie comme « le régime politique sans régime » ou encore « l'accès de ceux dont l'accès ne va pas de soi », qualificatifs qui soulignent l'aspect exceptionnel de cette organisation politique. De manière plus conventionnelle, on peut définir ce régime politique comme une démocratie participative par tirage au sort, fondée d'abord sur l'*isonomia*, l'égalité des droits, et, plus tard, sur l'*isēgoria*, l'égalité de parole entre tous les citoyens. En 507 av. J.-C., Clisthène introduit les premières réformes devant l'assemblée qui définit cette *dēmokratia* athénienne, mais il s'appuyait sur le fait qu'il y avait une assemblée et qu'il existait des lois écrites, qui étaient respectées. En ce sens, ce nouveau système politique était largement redevable à Dracon, qui fut le premier à rédiger ses fameuses lois en 621 av. J.-C. à Athènes, et à Solon, qui constitua l'assemblée des citoyens et les classes oligarchiques censitaires en 584 av. J.-C. Clisthène était donc également redevable à Pisistrate et à Hippias, les tyrans qui avaient continué à promouvoir le fonctionnement du système politique et le respect de la loi, comme nous l'avons souligné plus haut.

Le système démocratique promu par Clisthène s'est avéré totalement différent de ce qui avait été observé auparavant. Il ne précisait pas comment choisir quelqu'un pour diriger, comme le faisaient les organisations oligarchiques ou aristocratiques antérieures des *poleis*, mais, au contraire, il précisait comment diriger, par l'agrégation des connaissances et des compétences de la population. Ce fut une rupture substantielle en termes de fonctionnement politique. En effet, Vernant (1965) a caractérisé la démocratie comme le passage de la question du choix à celle de comment choisir. Auparavant, en temps de crise, la *polis* faisait appel à un législateur ou à un tyran. Avec l'*isonomia*, le système a été conçu pour fonctionner et résoudre les crises par lui-même. Grâce à l'*isēgora*, la confrontation des idées a permis d'exercer cette rationalité et de rechercher les meilleures solutions. Cependant, si chacun peut donner son avis, sans nécessairement avoir une connaissance approfondie du domaine en question, comment peut-on bien gérer la cité ? Selon Vidal-Naquet ([1981] 2005), les hoplites ne se distinguaient pas des autres combattants dans les inscriptions, et comme le précise dans son Oraison funèbre Périclès, ce n'est pas la préparation à la guerre qui distingue les Athéniens, mais la bravoure. Cela souligne que ce n'était pas une technique ou un savoir-faire particulier qui prime, mais l'effort et la volonté. Ober (2008) a examiné

certaines aspects du fonctionnement de la démocratie athénienne, et en particulier sa capacité à agréger efficacement les informations. Pour Ober, le gouvernement d'Athènes peut être considéré comme une machine complexe et efficace, conçue pour identifier et rassembler les connaissances et le savoir-faire de chacun (Ober 2008, 118–167). En rompant les liens traditionnels entre les membres de chaque tribu et en introduisant l'aléa dans la répartition des responsabilités, les réformes de Clisthène ont forcé les Athéniens à collaborer en partageant les informations privées que chacun possédait. Ismard (2005) montre que la description du système démocratique par ceux qui lui sont hostiles, comme Platon, confirme que son fonctionnement repose avant tout sur le brassage des citoyens et la rupture de leurs relations privilégiées.

La recherche économique s'est penchée sur la logique du vote et sur les principes de base de la démocratie (Davis, Hinich, & Ordeshook 1970 ; Nurmi 1986). La conception des systèmes démocratiques athéniens était cependant innovante à plusieurs niveaux et a suscité un intérêt particulier pour l'économie institutionnelle, avec des travaux de recherche portant sur son fonctionnement général ainsi que sur nombre de ses particularités. Si certaines charges étaient attribuées par des élections, beaucoup étaient tirées au sort, et Tridimas (2011) a montré l'optimalité de cette procédure pour à la fois assurer la représentativité et réduire le risque de corruption, mais aussi réduire l'effort nécessaire pour choisir (Tridimas 2012). Tangian (2008) a analysé la configuration spécifique de l'assemblée athénienne, la *boulê* de 500 et son comité tournant de 50 en particulier, et a montré que la taille de ces instances représentatives était bien adaptée à Athènes. La fréquence des votes dans la démocratie participative athénienne était élevée, et Tridimas (2017) a montré comment, en augmentant cette fréquence, les citoyens obtenaient une plus grande utilité. Le célèbre principe d'ostracisme, utilisé jusqu'en 415 av. J.-C., était un dispositif politique qui permettait aux citoyens de bannir celui qui recevrait le plus de votes contre lui lors d'une session particulière de l'assemblée, votes qui étaient inscrits sur des fragments de poteries. Tridimas (2016) s'est penché sur le fonctionnement de ces votes d'ostracisme et l'a analysé comme un référendum négatif sur les politiciens les plus visibles. Une autre institution particulière, mise en place après 415 av. J.-C., était le *graphê paranomon* : le droit qu'avait chaque citoyen de poursuivre quiconque argumentait quoi que ce soit à l'assemblée, en déclarant qu'il s'agissait d'une proposition anticonstitutionnelle. Si l'accusation ne réunissait pas un

certain nombre de suffrages, elle était passible de lourdes sanctions. Lyttkens, Tridimas, & Lindgren (2018) ont analysé les résultats de quelques-uns de ces votes et montrent que de telles procédures pourraient en fait être utilisées par les politiciens pour mesurer l'étendue de leur soutien.

Dans le cadre du processus démocratique, les décisions judiciaires étaient traitées par de grands jurys populaires, de tailles diverses, où les jurés étaient tirés au sort. Les spécificités de ce dispositif juridique ont été examinées d'un point de vue économique. Par exemple, dans les procès pour homicide, après la présentation des preuves, l'accusé pouvait choisir de s'exiler et de ne plus rien risquer, ou de poursuivre le procès. McCannon (2010a) a examiné, dans le cadre de la théorie des jeux, les conditions dans lesquelles une personne choisirait une option ou l'autre. Le célèbre procès de Socrate a également été scruté d'un point de vue économique : McCannon (2010b) l'a analysé du point de vue de la décision médiane du juré, et Guha (2011) a examiné la rationalité de la stratégie du philosophe à son procès. McCannon (2011) se concentre sur l'optimalité de la taille des jurys à Athènes, qui pouvaient atteindre 2000 personnes.

Les liturgies, et en particulier les triérarchies, étaient des engagements semi-volontaires de citoyens riches pour couvrir des dépenses particulières liées à la *polis*, comme des représentations théâtrales, ou la défense navale. Kaiser (2007) a étudié ce mécanisme particulier et a établi les conditions dans lesquelles les citoyens pourraient volontairement s'engager ou au contraire chercher à esquiver leurs obligations. Toutes les dépenses financières de la cité n'étaient pas prises en charge par des citoyens riches, car certains revenus pouvaient être directement affectés à des besoins spécifiques. Le processus par lequel les décisions financières étaient votées a été étudié par Tridimas (2013), dans le cas particulier de l'attribution d'argent nouvellement découvert dans les mines attiques à la construction d'une grande flotte de combat en 483 av. J.-C.

8.1.3 *Analytic narratives* et cliométrie

La grande majorité de la recherche économique que nous avons mentionnée ci-dessus, que nous avons qualifiée d'économie politique, peut être plus précisément cantonnée à la microéconomie, à l'économie institutionnelle ou même simplement à la théorie des jeux. Néanmoins, puisque les travaux en question ont à voir avec l'histoire, et avec certaines

transitions historiques importantes, on peut se demander si elles peuvent être qualifiées de cliométriques. Étymologiquement, la cliométrie désigne la mesure de l'histoire, et plus précisément, elle concerne l'application de méthodes économétriques à des données historiques (Diebolt & Parent 2011), ce qui souligne l'importance d'utiliser effectivement un certain volume de données. Bien que la nature économique des analyses que nous avons discutées ci-dessus ne puisse être contestée, la mesure dans laquelle elles utilisent des données est assez limitée. En fait, une majorité de la modélisation de l'économie politique qui a été appliquée à la Grèce antique ne repose que sur quelques éléments résumés tirés de matériel historique, et il apparaît par conséquent qu'il n'est peut-être pas approprié de qualifier ce travail de cliométrique.

L'application d'un cadre économique à un contexte historique en dehors d'une perspective économétrique, et sans s'appuyer sur un important volume de données avec lequel on peut tester des hypothèses, pourrait, de manière plus appropriée, être qualifiée comme faisant partie des *analytic narratives*. Ces *analytic narratives* ont été introduites par Bates et al. (1998), et plus récemment développées par Mongin (2016) ; elles cherchent à expliquer des situations ou des événements historiques en utilisant la modélisation économique (Alexandrova 2009). Alors que la cliométrie a tendance à fonctionner à partir des données historiques sous-jacentes et construit une compréhension économique de ces données historiques, les *analytic narratives* fonctionnent dans l'autre sens et projettent une explication sur les événements historiques, basée sur des modèles économiques.

8.2 L'extension de la cliométrie à l'histoire ancienne par la complexité

La cliométrie est aujourd'hui devenue synonyme d'« économie historique quantitative ». Par conséquent, là où il n'y a pas de données économiques, il n'y a pas de cliométrie ; en particulier, il y a peu de recherches en cliométrie concernant l'histoire ancienne : l'ouvrage de référence dans le domaine, Diebolt & Hauptert (2019), ne mentionne même pas l'Antiquité. L'économie en général semble pourtant se faufiler dans de nombreux aspects des sciences sociales, de l'étiquette du siège de toilette (Choi 2011) à l'utilisation d'éponges contraceptives dans la série *Seinfeld* (Dixit 2012), allant bien au-delà d'une focalisation stricte sur l'économie. Cette gamme très variée de questions que l'économie peut examiner est souvent posée à travers de simples faits stylisés, comme dans les deux

exemples ci-dessus. En cliométrie, cependant, il faut disposer des données historiques pertinentes, sous la bonne forme, pour effectuer toute analyse. Cette nécessité peut expliquer le fait que la cliométrie n'ait pas cherché à couvrir un plus large champ en sciences sociales.

Cependant, de nouvelles perspectives sur les données, en particulier les données historiques, peuvent constituer une opportunité de réévaluer ce qu'est la cliométrie et les questions qu'elle peut poser. L'utilisation de l'analyse des systèmes complexes et la cliométrie, fusionnées dans le nouveau domaine de cliométrie & complexité, ont effectivement permis une nouvelle herméneutique des données, par rapport à la cliométrie traditionnelle (Durlauf et al. 2023 ; Abry et al. 2022). En s'appuyant sur des analyses de complexité, on peut laisser toute la granularité des données s'infiltrer dans les modèles économiques, sans avoir à la réduire à des faits agrégés ou stylisés. La recherche dans le domaine est toutefois restée largement concentrée sur l'histoire de l'économie.

Nous tentons ici de comprendre comment la cliométrie couvre effectivement beaucoup moins de terrain que l'économie en général, et, sur la base de l'épistémologie différentielle entre économie et histoire, et nous proposons une version élargie de la cliométrie & complexité qui pourrait étendre ces limites.

8.2.1 L'importance de la distinction épistémologique entre histoire et économie

La distinction que l'on peut faire entre cliométrie et histoire économique renvoie à celle entre l'économie et l'histoire, plus large et plus fondamentale. L'espace entre l'histoire et l'économie, dans lequel résident la cliométrie et l'économie historique, n'est en effet pas un continuum. Des différences épistémiques fondamentales empêchent une transition en douceur de l'un à l'autre. Les causes les plus fondamentales sont, dans notre analyse, l'utilisation de faits stylisés par opposition aux données brutes réelles, et la tendance à tirer des conclusions générales par opposition à des conclusions spécifiques.

L'économie utilise fréquemment des faits stylisés : des résumés de la réalité, qui capturent vraisemblablement les composants de cette réalité jugés précieux pour la modélisation, afin de tester des idées ou de s'en inspirer. Pour Hirschman (2016), les faits stylisés sont en effet des « régularités empiriques à la recherche d'explications

théoriques et causales », et ce sont des outils heuristiques utiles dans les sciences sociales en général. Dans leur étude méthodologique des faits stylisés, Arroyo Abad & Khalifa (2015) les distinguent de ce qu'ils définissent comme des faits de base. Les faits de base et les faits stylisés prétendent décrire des phénomènes, et les deux sont censés être expliqués par une théorie économique. Cependant, alors que les faits de base doivent être « validement déduits de données fiables », les faits stylisés ne le sont pas et sont le résultat d'un certain degré d'interprétation des faits. Selon ces définitions, il apparaît que ni les faits stylisés ni les faits de base ne consistent réellement en l'information brute sous-jacente, ils sont, par définition, interprétés.

L'économie historique, en revanche, consiste à écrire des éléments d'histoire qui se rapportent à l'économie, en utilisant essentiellement des sources primaires. Les historiens produisent des connaissances sur le passé humain, selon une méthode qui se veut scientifique, basée sur un matériau historique. La notion trompeusement simple d'un fait, d'un point de vue historique, est en soi discutable. Ce n'est pas quelque chose d'enfoui que l'on vient découvrir, prêt à l'emploi, pour des théoriciens de l'histoire comme Febvre ([1952] 1992), Aron ([1938] 1991) ou Marrou ([1954] 2016). Par conséquent, le fondement de toute théorie rigoureuse de l'histoire doit être la manière dont les documents sont décrits, suivie d'une méthode pour relier cette description à la justification qui l'établit comme fait historique.

Ainsi, en n'opérant pas directement sur les sources primaires, la cliométrie est quelque peu déconnectée de l'histoire. Le recours à des informations stylisées, ou des faits au sens économique, élimine la quantité massive de méta-informations qui accompagne l'établissement de tout « fait » historique. En conséquence, du point de vue de l'historien, ses découvertes ne sont pas historiquement valables car elles peuvent être considérées comme une reformulation des affirmations antérieures de certains historiens. D'un point de vue historien, nous pouvons voir que la question n'est pas tant celle de la gamme de données qui ont pu être recueillies par les économistes, mais si cela a été fait selon la méthode historique. Si l'association de la cliométrie à la complexité peut apparaître comme une simple distinction praxéologique, elle a donc des conséquences épistémologiques profondes. L'utilisation d'approches systémiques complexes permet d'opérer directement sur des données très brutes, et donc typiquement sur des sources primaires, qui n'ont pas à être prétraitées par des historiens.

Si la relation aux sources primaires contribue à distinguer l'économie de l'histoire, et à son tour la cliométrie de l'économie historique, ce n'est pas la seule distinction de ce type. En fait, la cliométrie recherche des règles universelles, soit de manière déductive, en commençant par la théorie économique générale, soit de manière plus inductive, en recherchant le modèle formel qui correspond le mieux aux données. Elle recherche des ensembles de principes plus ou moins universels, ou tente de les vérifier. Cette recherche d'un ensemble de règles s'effectue à travers l'utilisation d'une modélisation formelle, qui s'appuie sur des données organisées de manière à illustrer les liens recherchés. Ainsi, une distinction fondamentale entre l'histoire et la cliométrie, qui peut rendre leurs approches inintelligibles l'une à l'autre, est celle entre les visions nomothétiques et idiographiques de l'histoire. Cette distinction repose sur la nature des relations avec les données : la recherche de l'universalité en économie nécessite l'utilisation de données gérables, et la cliométrie doit opérer sur des entités prédéfinies en quelque sorte, et par rapport à une théorie, et opérer sur des faits de base ou stylisés. Les faits de base ou stylisés en économie ne sont dérivés que de données sous-jacentes, et ils sont tous deux conçus dans le périmètre d'une théorie. Ce que les économistes peuvent appeler des faits basiques ou stylisés peuvent en fait être appelés « lois générales » par les historiens. Ces faits stylisés ne sont pas mécaniquement dérivés de données ou d'informations brutes, ils sont nécessairement le résultat d'une analyse et d'une interprétation, lorsqu'ils sont immergés dans un appareil théorique.

Cependant, il convient de construire des structures, plutôt que de les présumer, en l'histoire (Marrou [1954] 2016), une perspective qui rejette la plupart du temps toute « grande explication » ainsi que l'utilisation de notions génériques. En effet, on ne peut pas supposer qu'un modèle du monde, très large, fournirait une explication satisfaisante de tous les aspects de la réalité. Puisque l'histoire se définit comme l'étude de phénomènes distinctifs et singuliers, Marrou estime qu'il est inutile de développer des « lois historiques ». Les parallèles et les analogies en revanche sont simplement des exemples qui s'appuient sur quelques caractéristiques communes. Le but des sciences sociales, et surtout de l'économie, est de réduire une situation complexe à un modèle simple. Selon Ober, un historien de la Grèce antique : « l'objectif du spécialiste des sciences sociales de tester la théorie, visant à une compréhension plus générale du comportement humain, peut être strictement hors de propos pour l'historien qui reste

intensément concentré sur le passé grec » (Canevaro et al. 2018). La pensée économique moderne est utilisée par les économistes comme cadre d'analyse, ce qui va à l'encontre de la culture historiographique de la majorité des historiens.

Les différences épistémiques entre l'économie et l'histoire sont suffisamment substantielles et fondamentales pour qu'il soit difficile d'imaginer de la recherche qui puisse vraiment faire avancer les deux domaines en même temps. Cette discontinuité, notamment par rapport à l'aspect nomothétique de la modélisation économique, ne peut être simplement franchie. Néanmoins, elle peut offrir à la cliométrie un angle nouveau, et une marge d'expansion, si l'on considère cette distinction épistémique comme une direction vers laquelle progresser, notamment grâce à l'association de la cliométrie et de la complexité.

8.2.2 Une nouvelle cliométrie associée à la complexité

Nous proposons d'élargir la définition de la cliométrie, en étendant son rapport à l'économie : tout d'abord en ne se limitant pas aux questions purement économiques, et ensuite en se concentrant sur le même matériau historique que les historiens. Puisque certaines sources primaires sont disponibles en masse sous forme de données électroniques, alors par une approche plus systématique et programmatique la cliométrie & complexité peut rendre l'écriture de l'histoire plus scientifique.

Les liens entre la cliométrie et l'étude de l'économie sont forts, au point que même ceux qui cherchent à critiquer le domaine les prennent pour acquis et ne les remettent pas en cause (Boldizzoni 2011 ; Rawski et al. 1996). Cette focalisation sur l'économie n'est cependant pas un reflet fidèle de l'éventail des domaines dans lesquels les modèles économiques, et en particulier les modèles microéconomiques, ont apporté de nouvelles théories ou méthodes dans les sciences sociales. On peut évoquer les travaux de Boyd & Richerson (2005) dans l'étude de l'émergence de la société humaine, une perspective ensuite appliquée à l'écologie comportementale, à la psychologie de l'évolution et à la notion de coopération (Axelrod & Hamilton 1981 ; Axelrod 2006). Au sein des sciences sociales, les approches fondées sur l'économie ont atteint toutes les strates. Par exemple, les questions liées à la religion ont été explorées sous un angle microéconomique, à la fois par des économistes (Iannaccone 1998), et par des spécialistes

des études religieuses, notamment en développant des modèles de signaux coûteux pour les rituels (Henrich 2009 ; Bulbulia & Frean 2010). La théorie des jeux a également été appliquée à l'étude de la littérature : Brams (1994) a proposé une revue sur le sujet, et de Ley (1988) a développé quelques exemples d'applications à la littérature française contemporaine. Comme en témoigne Posner ([1973] 2014), l'analyse économique du droit, d'abord utilisée par Coase (1960), a également apporté un éclairage nouveau sur la compréhension de l'environnement juridique et de son évolution. Des outils microéconomiques ont également été utilisés pour étudier la structure des arguments et la rhétorique (Beigman Klebanov & Beigman 2010 ; Lipman & Seppi 1995 ; Matthews 1989).

La cliométrie, tout comme l'économie en général, n'a ainsi pas à nécessairement porter sur l'économie. Elle peut utiliser l'économie pour aborder de nombreuses questions historiques qui ne sont pas, en soi, économiques, mais que l'on peut éclairer par l'application de modèles microéconomiques. Par conséquent, lors de l'analyse de documents historiques, on peut utiliser ces approches afin d'améliorer la compréhension des documents. Le principe fondamental est que d'appliquer les méthodes microéconomiques aux matériaux historiques qui sont aussi proches que possible des sources originales, et non au produit final des interprétations des historiens. En n'étant pas liée à des modèles spécifiques sur l'économie, la gamme de données qui peuvent être analysées en cliométrie devient plus variée et peut concerner tous les aspects de la vie humaine. En conséquence, les mêmes documents sur lesquels travaillent les historiens peuvent être utilisés, mais sous un angle différent. La cliométrie & complexité peut être utilisée pour identifier les mécanismes générateurs qui peuvent avoir produit les distributions que l'on observe dans les données. Considérer ces mécanismes générateurs relie fondamentalement la cliométrie à l'analyse des systèmes complexes, qui se concentrent sur l'explication de l'émergence de comportements au sens large (Boccaro 2010). La cliométrie, s'appuyant sur la modélisation de la complexité associée à des modèles microéconomiques, peut donc aider à déterminer selon quel processus les données ont été engendrées. La prise en compte de ce processus permet une meilleure compréhension tant du point de vue de la modélisation économique que du point de vue de l'historien.

Comme la cliométrie ne doit pas nécessairement être fongible avec l'histoire, l'extension du domaine à des questions non économiques et à l'utilisation des sources primaires que permet une approche par les systèmes complexes peut, dans certaines conditions, constituer une passerelle épistémologique. En effet, l'utilisation d'approches économiques atemporelles sur des phénomènes historiques est compatible avec l'établissement de structures, plutôt que leur simple postulat, à condition que les approches économiques en question soient construites à partir de données historiques plutôt que d'être posées comme des applications d'une théorie économique particulière. Cela correspond à voir la microéconomie du point de vue d'un historien comme un instrument auxiliaire plutôt que comme une application de théories économiques particulières, puisque l'historien doit de toute façon faire une hypothèse fondamentale de rationalité pour comprendre les acteurs.

Méthodologie cliométrique et histoire ancienne ne semblent pas, à première vue, avoir grand-chose en commun. Nous avons fait valoir qu'en fait, ils le peuvent, grâce à l'analyse des systèmes complexes. L'étude et la lecture attentive des sources anciennes est une prérogative des historiens, mais les outils numériques qui ont été développés à ces fins sont souvent sous-utilisés du point de vue de l'analyse des données, et les nouvelles approches développées dans le domaine de la cliométrie & complexité sont particulièrement bien adaptées à ces données. Si l'on considère que la cliométrie n'a pas à porter sur l'économie *stricto sensu*, alors de larges pans de sources primaires, en particulier dans l'histoire ancienne, peuvent être analysées à la lumière de modèles microéconomiques. Les progrès des humanités numériques ont amené une grande quantité de documents historiques sous forme électronique. Cela étend la notion d'analyse cliométrique de deux manières: premièrement, la cliométrie pourrait s'intéresser à tout type de comportement humain, dans la mesure où il y a des raisons de croire qu'il y a une certaine optimalité ; et en outre, en n'ayant pas à être spécifiques à l'économie, les données peuvent être directement tirées de sources primaires, grâce aux sciences de la complexité. Plus la distance aux sources est faible, plus la distance épistémologique à l'histoire est faible, ce qui rend la cliométrie plus intelligible aux historiens. Une approche selon la cliométrie & complexité examine cependant le matériel historique d'une manière fondamentalement différente de celle des historiens, apportant potentiellement un nouvel éclairage sur une large gamme de questions historiques. Cet

accent mis sur le traitement des données peut également améliorer considérablement la manière dont les corpus de documents historiques sont traités, en rendant systématique et explicite de nombreux aspects de l'analyse des données qui sont généralement cachés par les historiens, renforçant ainsi la validité scientifique de ce que l'on affirme.

8.3 Une Perspective sur la Grèce ancienne fondée sur la cliométrie & complexité

Restant dans la voie de la cliométrie & complexité, nous proposons dans ce chapitre une analyse d'inspiration économique des diverses facettes de la culture grecque antique, servant d'illustration de l'utilisation de nouveaux angles méthodologiques dans l'étude du passé à travers l'économie. Nous considérons que les méthodes économétriques avancées ne doivent pas se limiter à l'étude de l'économie, mais au contraire que de grandes quantités de données peuvent toujours bénéficier d'un examen quantitatif et être mises en perspective à travers des modèles économiques. Nous appliquons des analyses économétriques et économiques, inspirées des modèles de complexité, à plusieurs ensembles de données importants relatifs à la Grèce antique, directement issus de la production littéraire ou épigraphique des Anciens. Se concentrer sur les sources primaires, qui sont de grands ensembles de données, recoupe bien l'approche de la cliométrie & complexité, conçue pour analyser l'émergence de mécanismes générateurs. Les historiens de la Grèce antique considèrent généralement les sources primaires au niveau atomique. Dans l'histoire ancienne, l'analyse philologique d'une seule inscription, ou de l'occurrence d'un nom particulier dans un texte, peut souvent donner lieu à un commentaire détaillé, fondé sur la connaissance que possède l'historien du contexte dans lequel les documents historiques ont été produits. L'analyse économique d'une situation particulière, généralement à l'aide de la théorie des jeux, peut également éclairer une telle analyse historique : c'est l'objet des *analytic narratives*, comme nous l'avons vu précédemment. Cependant, la cliométrie peut offrir une nouvelle perspective en se concentrant sur les sources primaires à travers le prisme de l'économie de la complexité, basée sur des herméneutiques fondamentalement différentes. Compte tenu de la nature catégorique des données, on peut en effet les examiner dans une perspective distributionnelle, et se demander quel mécanisme générateur a pu les

produire. L'analyse des mécanismes générateurs, et l'identification des conditions de leur optimalité en termes de comportement humain, relèvent néanmoins de l'analyse économique.

8.3.1 Approche empirique et données

La plupart des sources que nous discutons ici contiennent de grandes quantités de données de nature catégorielle. Avant de commencer à les approfondir et à essayer de comprendre divers aspects de la culture grecque antique sous cet éclairage, il est utile d'établir le type de modèles que l'on peut naturellement attendre de ces données. Il convient de noter que l'analyse des distributions n'est pas particulièrement courante en cliométrie, ni en relation avec l'économie de la complexité, car les données sont rarement de nature catégorielle. L'analyse empirique des distributions de données non structurées représentant des phénomènes sociaux ou physiques complexes fait pourtant partie intégrante des sciences de la complexité (Boccaro 2010).

Nous nous intéressons principalement aux lois de puissance, aux lois exponentielles, et aux lois log-normales. Les lois de puissance sont couramment utilisées en sciences de la complexité, car elles apparaissent fréquemment dans la nature et dans de nombreux phénomènes humains, et peuvent être produites par de nombreux types de processus sous-jacents (Boccaro 2010 ; Mitzenmacher 2004). Les lois de puissance ont en effet été appliquées dans de nombreuses approches en économie (Gabaix 2016). Les mécanismes générateurs de ces distributions sont typiquement associés à un attachement préférentiel dans la formation de réseau (Barabási & Albert 1999), ou à des processus de croissance aléatoires (Gabaix 1999). On peut également montrer que les lois de puissance correspondent au comportement limite des fluctuations extrêmes (Alfarano & Lux 2010). Un cas particulier de loi de puissance est la loi de Zipf en linguistique, avec un paramètre de 1. Zipf a observé que la relation entre la fréquence d'occurrence des mots dans le langage naturel et le rang de leur fréquence suit cette distribution particulière (Zipf 1949). Ce modèle a été largement étudié en linguistique computationnelle, et de nombreux modèles expliquant son émergence dans le langage ont été proposés. Les lois de puissance, y compris la loi de Zipf, se recoupent fondamentalement avec l'économie, dans le sens où il peut être démontré qu'elles émergent à la suite d'une forme d'optimisation. Cela est

apparu pour la première fois lorsque Mandelbrot (1953) a montré que la loi de Zipf était optimale pour la communication. La loi de puissance tronquée, plus générale, avec une queue exponentielle, peut être associée à la formation de groupes aléatoires. En effet, Baek, Bernhardsson, & Minnhagen (2011) montre que la formation de groupes de tout type de données catégorielles afin de minimiser le coût d'identification d'un élément donné entraîne une distribution de loi de puissance tronquée pour la taille du groupe. Les distributions exponentielles sont généralement associées à des temps aléatoires, ou à l'occurrence de surprises, en raison de leur « manque de mémoire » par conditionnement. Ely, Frankel, & Kamenica (2015) modélisent la notion de maximisation de la surprise d'un point de vue économique, mais ne caractérise pas la distribution optimale des événements en général. Borwein, Borwein, & Maréchal (2000) montrent que sur de longues périodes, le maximum de surprise est atteint avec une distribution partiellement exponentielle. Des distributions exponentielles peuvent également émerger dans les réseaux : dans un réseau hors équilibre, en croissance continue, si les nouveaux nœuds qui apparaissent se lient aux nœuds existants suivant une distribution uniforme, alors la distribution des degrés du réseau est exponentielle, avec un paramètre dépendant du degré moyen dans le réseau (Deng et al. 2011). Enfin, Limpert, Stahel, & Abbt (2001) ont noté l'occurrence de la distribution log-normale dans les phénomènes biologiques et les mécanismes de croissance, et soulignent qu'il s'agit d'un bon candidat pour les situations où la taille minimale ne peut pas être la plus commune. Certaines données économiques que l'on pensait initialement suivre la loi de Zipf se sont d'ailleurs avérées suivre des distributions log-normales, telles que la taille des entreprises et des villes (Saichev, Malevergne, & Sornette 2010).

Nous nous intéressons à cinq sources de données primaires sur la Grèce ancienne, réparties en deux grandes catégories. Tout d'abord, les textes : plus que tout autre objet du passé, ils sont la matière première par excellence de l'histoire. La seconde catégorie est celle de données qui peuvent être naturellement structurées sous la forme de réseaux.

La source primaire la plus naturelle pour l'étude de la Grèce antique est le corpus de toute la littérature qui nous a été transmise, couvrant la période archaïque jusqu'à la fin de l'Antiquité. Ces textes nous sont parvenus, pour la plupart, par le biais de manuscrits qui ont été copiés au fil du temps ; il n'y a pratiquement aucun cas où nous possédons des écrits littéraires originaux. Il existe de nombreuses ressources

disponibles électroniquement compilant des textes grecs anciens, telles que le Thesaurus Linguae Graeca, ou TLG (Pantelia 2020), ou le site web Perseus hébergé par l'Université Tufts (Crane 2012). Ces ressources ne sont néanmoins généralement pas absolument exhaustives, ne permettent pas aux utilisateurs de télécharger les données comme un ensemble unique et ne contiennent que le texte brut. Afin d'associer toute instance d'un mot à son lemme (sa racine) et donc d'identifier le rôle de chaque mot dans une phrase, les mots doivent être catégorisés (Ide 2004). Le Diorisis (Vatri & McGillivray 2018b) est un corpus centralisé et complet de texte grec ancien, et rassemble 820 œuvres différentes ; il a été composé à l'origine pour analyser l'évolution sémantique du grec ancien au fil du temps, avec des méthodes de linguistique computationnelle, et s'adresse aux historiens et aux hellénistes. Les données de Diorisis contiennent une ligne par mot ou signe de ponctuation, pour un total de plus de 14 millions de lignes.

Les anciens Grecs avaient l'habitude d'inscrire de nombreuses informations sur des monuments, sur des stèles, sur des vases et dans de nombreux autres endroits. Ces écrits nous sont parvenus sous leur forme originale, bien qu'avec quelques lacunes. Cela crée une situation différente des textes littéraires, qui, à l'exception des papyrus, ont été copiés de nombreuses fois sous forme de manuscrits. La tradition littéraire a conservé des œuvres jugées précieuses, mais de nombreux documents écrits, comme des comptes ou des contrats, ont été perdus. Les inscriptions constituent donc un témoignage unique de la Grèce antique, puisqu'elles constituent de véritables documents originaux. Les épigraphistes lisent et analysent ces documents, et transcrire, éditer et contextualiser les textes à partir du matériel archéologique est un processus complexe. De nombreuses inscriptions sont dans un état dégradé ou fragmentaire, et l'épigraphie s'attache aussi à formuler les meilleures hypothèses pour combler les informations manquantes, à partir de quelques récurrences couramment observées (McLean 2002). En ligne, le dépôt le plus vaste et le plus complet d'inscriptions grecques est le site web *Searchable Greek Inscriptions* du Packard Humanities Institute (PHI) ("PHI Greek Inscriptions," n.d.). Les données PHI disponibles en ligne sont présentées sous la forme d'une page web pour chaque inscription, qui contient le texte de l'inscription ainsi que des informations géographiques supplémentaires. Comme il n'y a pas de base de données centralisée disponible contenant l'intégralité des données PHI, elles doivent être traitées page après page à l'aide d'un navigateur web automatisé. Comme, pour les historiens,

les inscriptions sont le plus souvent analysées une par une, ou par petits groupes, les données épigraphiques n'ont pas été, à ce jour, considérées et analysées comme un tout, de façon systématique. Une fois que toutes les inscriptions sont disponibles électroniquement en un seul endroit, il est possible de les traiter afin de relier chaque mot reconnaissable à un lemme, en utilisant le *Classical Languages Tool Kit* (K. P. Johnson et al. 2019). Avec les données PHI centralisées, on peut ainsi effectuer une analyse transversale qui serait autrement impossible.

À partir des sources textuelles que nous avons mentionnées ci-dessus, les historiens et les hellénistes ont créé d'autres corpus plus complexes qui sont disponibles électroniquement. L'étude de la religion et des rituels grecs anciens est une partie importante de la recherche historique sur la période. Les sources épigraphiques liées aux offrandes ou aux dédicaces religieuses contenaient généralement les noms d'un ou plusieurs dieux ainsi que certaines qualifications et raisons de l'inscription. La Base de Donnée des Epiclèses Grecques (BDEG) est un projet initié au début des années 2000 (Lebreton et al. 2014), contenant les informations de milliers d'inscriptions épigraphiques ou de références littéraires à des actes votifs, avec notamment l'emplacement, le nom du ou des dieux et la nature de l'inscription. Les données ont été saisies dans plus de 11 000 formulaires, un pour chaque observation de la mention d'une divinité. La BDEG présente ses données sous forme de pages web distinctes : elle est conçue pour permettre aux chercheurs de visualiser un formulaire à la fois, et, là encore, afin d'analyser les données en masse, il est nécessaire de rassembler, d'analyser et de compiler toutes les données de ces pages web.

Nous nous intéressons à certaines données issues de la Grèce ancienne, qui représentent assez naturellement des relations, et ont donc été abordées par les historiens comme des réseaux, dans quelques cas au moins. Ces données particulières sont principalement compilées à partir d'inscriptions : il s'agit du *Lexikon of Greek Personal Names* (LGPN) et de la base de données *Mapping Ancient Polytheisms* (MAP). Tous deux s'appuient sur un travail ardu effectué par des spécialistes, ayant passé au crible des centaines de milliers d'inscriptions afin de produire les données électroniques qui en résultent.

Dans la Grèce antique, nommer un enfant se faisait d'une manière très différente de ce que nous connaissons aujourd'hui : on n'avait qu'un seul anthroponyme, un nom unique. A ce nom unique pouvait être associé un patronyme, le nom de son père. Les sources épigraphiques disponibles fournissent des anthroponymes, parfois accompagnés d'un patronyme. Pour passer des listes de noms à la notion d'individus, il faut relier ces inscriptions les unes aux autres. Les prosopographes s'appuient sur des instances de noms identiques ou étroitement liés dans une zone géographique donnée, et sont datées à peu près au même moment, de sorte qu'ils peuvent transformer ces noms en références à des individus particuliers. A partir de ces informations historiques, on peut donc créer des réseaux onomastiques : quelqu'un portant tel nom a eu un enfant portant tel nom. Ensuite, avec le travail prosopographique, on peut créer des réseaux prosopographiques : telle personne a eu telle personne comme enfant. Il existe plus de données onomastiques disponibles que prosopographiques, et la reconstruction prosopographique repose sur de nombreuses hypothèses. De grands volumes de données onomastiques et prosopographiques ont été mis à disposition par voie électronique grâce au LGPN (Parker, Yon, & Depauw 1996). Le projet a débuté au début des années 1970 et a conduit à la publication de références à des centaines de milliers de noms. L'interface électronique du LGPN n'est pas conçue pour que ses données soient traitées et analysées en masse, comme c'est malheureusement souvent le cas des sources numériques en histoire ancienne. Il est conçu comme un outil pour interroger un nom ou une racine de nom et observer ses occurrences. Il est impossible de télécharger l'intégralité des données sous une forme structurée en une seule fois ; tout doit être reconstruit à partir des données relatives à chaque nom possible. Nous avons créé une base de données centralisée contenant tous ces éléments, de manière structurée dans le sens où toutes les données sont empilées dans des tableaux de données uniques, et afin qu'elles puissent être jointes à des sources de données supplémentaires. Nous avons obtenu près de 40 000 noms uniques répartis sur environ 350 000 entrées individuelles. La table de relations comprend environ 250 000 liens. En utilisant les relations parentales entre les noms dans un lieu donné, nous pouvons construire un réseau de ces liens, à partir des données du LGPN. Nous avons mis en œuvre une méthode comparable à celle présentée dans Karila-Cohen (2018), qui s'est concentrée sur des quartiers particuliers d'Athènes, mais l'avons appliquée à l'ensemble des données

du LGPN. Nous constituons un réseau où les sommets sont des noms uniques, et les arêtes sont les relation parentale dans le LGPN.

Le projet MAP (Bonnet 2017) offre une représentation précise et entièrement remaniée de données sous-jacentes comparables à celles de la BDEG, principalement issues d'inscriptions, couvrant les mondes antiques grecs et sémitiques, et permet d'étudier de près les formules utilisées par les Anciens pour s'adresser aux dieux. Pour comprendre la logique relationnelle qui structure ces puissances divines, le MAP prend en compte un large éventail de séquences onomastiques divines, combinaisons de noms ou d'éléments divins (noms, épithètes, titres, propositions), certaines communes à plusieurs dieux, d'autres propres à un dieu. En effet, le simple fait d'accomplir un rite dans la Grèce antique impliquait de s'adresser à un ou plusieurs dieux avec une série complexe de qualificatifs, les épiclèses, ordonnés et exprimés de manière délibérée. Ces qualificatifs, parfois communs à certains dieux, créaient effectivement un réseau entre tous les dieux ; et les dieux créaient également un système de relations entre les qualificatifs. Ces séquences onomastiques prises dans leur ensemble forment une représentation de la manière dont les Grecs concevaient le polythéisme. L'idée de placer ces séquences dans le cadre de l'analyse de réseau est en effet à la base du projet MAP, et une autre pierre angulaire est l'extraction de formules onomastiques des sources qui, grâce à une syntaxe particulière, peuvent rendre compte de la grande complexité de ces inscriptions (Bonnet & Lebreton 2019). Les données qui constituent le MAP, dérivées des inscriptions par une équipe de spécialistes, sont très structurées et accessibles en masse. Chaque élément matériel, comme une stèle ou une partie de monument par exemple, est une entrée dans une table de données source, avec des informations détaillées telles que sa publication ou sa localisation. Une source contient un ou plusieurs témoignages, adresses aux dieux, inscrits chacun dans une table de témoignages, ainsi que chaque témoignage, généralement sous la forme d'une phrase faisant référence à des divinités et des qualificatifs. Ces phrases sont converties en formules, apparentées à des formules mathématiques, rendant compte des liens entre les différents éléments du témoignage. Ces éléments, principalement des adjectifs, des noms ou des noms de dieux, sont également centralisés dans une table d'éléments spécifique.

8.3.2 Résultats

Nous nous sommes tout d'abord intéressés, sur la base du corpus Diorisis, à l'économie du genre littéraire. Des propriétés linguistiques fondamentales peuvent être observées sur les données, comme la loi de Zipf, en fonction de la catégorisation des œuvres selon leur genre (tragédie, prose, philosophie, etc.). La partie gauche des courbes liant rang et fréquence, représentant la queue de la distribution, est bien droite et de pente proche de 1. Il existe cependant des variations dans les coefficients de la loi de puissance dans les données. La poésie, la comédie et la tragédie semblent toutes présenter des queues plus fines que les autres genres, leurs mots les plus fréquents étant moins courants que pour les autres genres. Nous pouvons vraisemblablement relier ces observations au fait que certains types de littérature recourent généralement à un éventail de vocabulaire plus large que d'autres. Une analyse coûts-bénéfices peut nous permettre de mieux comprendre les différences que nous pouvons observer dans les variations de la forme de ces distributions pour ces différents genres littéraires. La poésie, dans les langues anciennes, est principalement définie par des contraintes métriques : le rythme des voyelles longues et courtes doit suivre un schéma particulier. Cette contrainte crée le besoin d'un vocabulaire plus varié, étant donné un message à communiquer, au-delà de la recherche esthétique de mots rares. Il est possible que des poètes s'écartent parfois de la norme, pour un effet littéraire particulier, ou parce qu'ils ne pourraient autrement exprimer le sens, mais un tel écart a un coût esthétique. La tragédie est écrite en vers, mais bien qu'elle soit soumise au même type de contraintes que la poésie en général, elle a une distribution plus pentue, indiquant un coefficient plus faible. Contrairement à la poésie en général, la tragédie doit pouvoir exprimer un dialogue vivant et être clairement compréhensible, et elle est généralement considérée comme plus proche du langage courant. Par conséquent, la différence dans le paramètre de distribution entre le simple récit et la tragédie peut saisir l'effet des contraintes métriques, tandis que la différence entre la poésie et la tragédie peut capturer la capacité d'utiliser un vocabulaire plus fleuri car les contraintes de codage/décodage liées au message sont moins importantes. Les textes oratoires, pour la plupart des discours judiciaires athéniens, doivent également trouver le juste équilibre entre le besoin d'être compris par les juges (une assemblée de centaines de citoyens) et la nécessité d'utiliser un vocabulaire spécialisé, en particulier dans les affaires financières. Néanmoins, le langage potentiellement fleuri des orateurs

est sensiblement plus proche des simples textes narratifs que de la poésie. Bien qu'il soit difficile de construire un modèle *a priori* qui pourrait rendre compte de l'ampleur de ces différences, un modèle formel de communication optimale pourrait nous permettre rendre compte de l'échelle des différences entre les genres et potentiellement d'établir des comparaisons avec des corpus spécifiques à un genre dans les langues modernes.

L'étude des inscriptions peut s'aborder d'une manière similaire. Nous établissons tout d'abord que les distributions de la longueur des textes inscrits est différente selon les régions. Ces distributions sont à peu près cohérentes avec une distribution exponentielle pour certaines, et une loi de puissance pour d'autres. Plusieurs facteurs peuvent affecter la longueur des inscriptions. Le matériau peut certainement avoir été brisé au hasard, mais dans de nombreux cas, des fragments sont regroupés par les épigraphistes, formant des inscriptions uniques. Dans la plupart des cas, il semble logique que le coût de la mise au point des messages et la nature des messages soient les facteurs déterminants de la longueur de ces messages. L'aspect qui différencie le plus ces distributions de lois de puissance semblent être leur comportement pour des longueurs de textes très courtes. Cela indique que les inscriptions d'un seul mot ou de deux mots sont beaucoup moins fréquentes qu'une loi de puissance ne l'impliquerait, si elle devait être cohérente avec le nombre d'occurrences de messages plus longs. D'autre part, les inscriptions très longues sont également plus courtes qu'une loi de puissance ne l'impliquerait. Cette coupure, qui peut être capturée par une troncature exponentielle, semblerait nécessaire pour refléter la contrainte de disponibilité matérielle ainsi que la plus grande probabilité que des inscriptions plus longues soient brisées et fragmentaires. Les messages courts, dans la communication et la publicité contemporaines, sont plus efficaces, toutes choses égales par ailleurs (Baltas 2003). Bien que les inscriptions anciennes ne soient pas la même chose que des publicités sur internet, elles partagent le même objectif de rendre un message public. Alors que les textes littéraires sont essentiellement narratifs, les inscriptions cherchaient à communiquer des informations spécifiques. Ils ont surtout enregistré des décisions politiques, des traités entre villes, des honneurs rendus, des rites funéraires, la libération d'esclaves ou des dédicaces religieuses (McLean 2002). Les inscriptions tendent donc à exprimer des actions que certaines personnes entreprennent à l'égard d'autres personnes ; par exemple, un mari fait graver une pierre tombale en l'honneur de sa femme décédée, ou l'assemblée honore un citoyen éminent. Par

conséquent, l'information minimale qui peut être communiquée est de la forme « A fait B », qui utilise nécessairement au moins trois mots. En revanche, la gravure est coûteuse, et plus ou moins selon les régions et selon l'habitude culturelle de recourir aux inscriptions. L'aspect particulier de la distribution de la longueur des messages indique donc vraisemblablement une optimisation du message dans les inscriptions, par rapport aux coûts de sa production. Par ailleurs, nous montrons que l'utilisation des mots dans les inscriptions est plus étalée que dans les œuvres littéraires : les inscriptions utilisent plus souvent des mots rares. Cela n'est vraisemblablement pas lié à une plus grande présence de poésie dans ces inscriptions, mais plutôt à l'apparition plus courante de noms propres aléatoires, ce qui est beaucoup moins probable dans la littérature. En fait, il y a une plus grande quantité d'information dans les inscriptions. Dans tous les types d'inscriptions, des personnes particulières ou des cités sont nommées, et ces noms constituent une partie substantielle de l'information transmise, ce qui rend chaque inscription unique en ce sens. La langue des inscriptions grecques anciennes peut donc être considérée comme optimisant son contenu informationnel, compte tenu d'un certain coût de gravure.

Nous avons également abordé l'étude de la distribution des actes votifs à partir de la BDEG. Nous établissons que le nombre d'actes relevés pour chaque dieu suit une loi de puissance. L'estimation des paramètres indique que cette distribution a une queue plus grosse que la loi de Zipf : dans chaque *polis*, il y a quelques dieux qui attirent un grand nombre de dédicaces, mais des dieux différents sont les plus populaires dans chaque ville. Comment donner un sens à ces distributions ? Si l'on considère les bénéfices offerts par les dieux comme une quantité finie, partagée par les gens qui les prient, alors un équilibre, dans le cadre de la théorie des jeux, peut consister en des stratégies mixtes, expliquant ainsi potentiellement le fait que tous les dieux bénéficient d'une popularité minimale. Étant donné qu'il existe plusieurs « produits », le fait que les actes votifs soient distribués selon une loi de puissance pourrait être envisagé comme un occurrence naturelle de consommation de biens : Kohli & Sah (2006) ont montré empiriquement que c'était le cas des parts de marché. D'autres explications, plus fondamentales, peuvent être envisagées en considérant une forme d'attachement préférentiel, par exemple.

Notre analyse des données qui peuvent être représentées en réseaux commence avec l'onomastique, sur la base du LGPN. Nous construisons tous les réseaux liant les

noms entre eux au niveau des *poieis*, et pouvons ainsi créer un vaste réseau couvrant l'ensemble des données, pour l'ensemble du monde grec à travers les périodes. Nous pouvons mesurer la forme de la distribution des degrés à l'échelle de ce grand réseau, c'est-à-dire la distribution du nombre de noms différents qui ont été choisis par une personne portant un nom donné, dans un lieu donné. Le nombre de noms distincts choisis par chaque nom paraît être une loi de puissance avec une queue de distribution épaisse. Supposons qu'il existe un grand réseau composé d'individus, des nœuds, chacun avec un attribut de nom, de sorte qu'une nouvelle personne apparaissant n'a aucun effet matériel sur la taille de la population. Une relation de nom est un lien créé entre un parent et son enfant (la nouvelle personne apparaissant sur le réseau). Un tel réseau prosopographique peut se condenser en un réseau onomastique, où tous les individus sont regroupés en un seul nœud nominatif et les liens agrégés en fonction des nœuds d'où et vers lesquels ils vont. Si les individus ont certains attributs suivant une distribution donnée, et si les liens de dénomination sont distribués uniformément, alors la distribution des degrés entrants et sortants des nœuds dans le réseau de noms devrait suivre cette distribution donnée. Nous établissons alors que les noms sont distribués selon une loi de puissance. Cependant, les paramètres de la distribution des degrés et de la distribution des noms sont assez différents, les degrés étant plus concentrés que ne le suggère une sélection aléatoire de noms. Il y a donc une certaine stratégie en jeu dans la sélection des noms, selon nos données. Ce schéma pourrait s'expliquer par un attachement préférentiel : si de nouveaux liens sont plus susceptibles de s'attacher aux nœuds en fonction de leur nombre de liens existants, nous nous attendrions à une plus grande concentration dans la distribution de degrés résultante que dans la distribution de noms sous-jacente. La stratégie de dénomination des Grecs anciens semble donc refléter un certain conformisme.

Enfin, nous avons analysé les formules onomastiques utilisées par les Grecs anciens pour s'adresser aux dieux, en nous appuyant sur les données du MAP. Une question importante, potentiellement liée à la structure du réseau entre les dieux et leurs qualificatifs, est de savoir si la « langue » utilisée dans ces invocations possède les caractéristiques d'une langue naturelle. Dans les applications de l'économie et des sciences de la complexité à la linguistique, de nombreuses analyses détaillées peuvent être menées sur la grammaire, en particulier lorsqu'elle reflète la structure en réseau

des relations syntaxiques (Čech, Mačutek, & Liu 2016). Nous nous limitons ici à une simple comparaison, en regardant la distribution des termes, par souci de simplicité. Les éléments, qui correspondent à des noms communs ou à des noms propres, sont l'équivalent naturel des lemmes dans l'analyse quantitative du langage. La distribution des occurrences pour les textes littéraires du corpus Diorisis et pour les formules onomastiques dans le MAP apparaissent assez proches les unes des autres. En fait, nous observons qu'elles suivent des lois de puissance avec le même paramètre, à la deuxième décimale près. Malgré leur caractère très formulaique, la distribution des termes utilisés dans les invocations divines ressemble à celle du langage naturel. Bien que cela puisse ne pas tenir si l'on devait regarder de plus près la structure syntaxique des deux « langues », c'est cohérent avec l'idée que des termes aléatoires sont sélectionnés dans des formules divines, puisque ces termes seraient extraits de la langue naturelle. De plus, comme l'ont montré Ferrer i Cancho, Riordan, & Bollobás (2005), lorsqu'un langage de communication se comporte en suivant la loi de Zipf, on peut théoriquement s'attendre à l'émergence de relations syntaxiques. Ainsi, l'inclusion de qualificatifs dans les formules divines, aussi aléatoire soit-elle, ne serait pas incompatible avec l'apparition de la structure syntaxique logique sur ces formules onomastiques.

8.4 Analyse cliométrique du choix du nom en Grèce ancienne

Certains noms grecs anciens ont acquis une renommée mondiale et sont encore aujourd'hui donnés aux enfants. Achille, Hector, Démosthène ou Alexandre sont fermement inscrits dans l'histoire et la littérature européennes. Ces noms célèbres étaient souvent donnés aux enfants déjà dans les temps anciens, et il existe des centaines de sources épigraphiques faisant référence à quelque Achille ou Démosthène à travers la Grèce ancienne. De nombreux autres noms, plus rares, ont également été donnés, dont nous n'avons trouvé qu'une poignée d'exemples dans les archives épigraphiques ou littéraires. Pourquoi y avait-il une telle variabilité dans la nature et la fonction effective des noms des anciens Grecs ? Si les noms uniques étaient utiles pour identifier les personnes, pourquoi tout le monde ne portait-il pas un nom différent, comme c'est généralement le cas dans le monde contemporain ? Pour comprendre comment les noms ont été donnés, nous devons explorer comment ils ont été utilisés.

L'onomastique, l'étude des noms, et la prosopographie, l'étude des histoires de vie individuelles, sont des facettes importantes de l'histoire de la Grèce antique. Les noms et l'identification individuelle ont été utilisés pour suivre plus précisément l'histoire des membres de l'élite, comme l'ont fait Puech (2012), Karila-Cohen (2017) et Karila-Cohen (2019). L'étude des noms et de leur attribution a également permis aux historiens de mieux appréhender l'étendue des familles, comme dans Bresson (1981) et Bresson (2019). L'utilisation de noms provenant de sources épigraphiques peut également aider à mesurer la fiabilité de certains récits historiques, comme Hornblower (2010) a proposé de le faire dans le cas de Thucydide. Il y a eu un effort très substantiel avec le LGPN pour collecter, puis numériser, les données de centaines de milliers d'inscriptions relatives aux noms dans le monde grec ancien. Grâce à la disponibilité de ces données, certaines recherches se sont concentrées sur l'aspect de réseau des relations familiales basées sur les noms dans les inscriptions, telles que Cline (2020), Karila-Cohen (2016) et Karila-Cohen (2018). Néanmoins, l'approche historique a eu tendance à se concentrer sur des zones particulières, ou sur des familles particulières, plutôt que de considérer les noms à une échelle beaucoup plus grande. Les approches historiques n'ont pas non plus examiné spécifiquement les raisons pour lesquelles certains noms peuvent avoir été plus courants que d'autres. Notre objectif ici est d'appliquer une approche économique à la compréhension des noms grecs anciens, afin d'aborder les données à grande échelle, et la logique de certaines caractéristiques distributionnelles de ces données. Nous nous concentrons donc sur la forme de la distribution des noms, ainsi que sur les stratégies de choix de ces noms.

Nous pouvons nous demander ce que l'économie pourrait avoir à dire sur quelque chose d'*a priori* aussi peu économique qu'un nom. En fait, la notion d'identité en économie recouvre un large éventail de concepts connexes. Berg (2019) donne un aperçu, et pour nos besoins, nous pouvons déterminer qu'il existe trois axes connexes dans ce corpus de recherche avec des applications ou des concepts utiles pour l'onomastique : le comportement de groupe (au sens des *club goods*), l'identité *stricto sensu*, et le conformisme. Carr & Landa (1983) ont examiné l'avantage d'appartenir à un groupe en réduisant les coûts de transaction grâce à l'application de certaines règles et en imposant des coûts en cas de violation de celles-ci. Ils ont montré avec un modèle simple comment on pouvait rendre compte de la création de clans, de symboles ou d'une

famille et du nom comme d'un *club good*. En fonction de diverses hypothèses sur le coût d'information des autres membres du groupe et le coût des ruptures de contrat, ils ont établi un équilibre pour la taille optimale du groupe. Une définition précise de l'identité a été donnée par Akerlof & Kranton (2000), qui a proposé d'étendre la fonction d'utilité pour rendre compte de leur notion d'identité : les personnes bénéficient davantage de la participation à des activités conformes aux prescriptions de leur identité choisie. Ils ont dérivé divers équilibres, où des parts variables de la population choisissent une identité particulière. Ne pas suivre les prescriptions de son identité est coûteux pour soi, car, selon eux, les règles sont intériorisées psychologiquement, et les enfreindre crée de l'anxiété. Leur modèle explique des comportements qui pourraient autrement apparaître anti-économiques. Les notions d'identité et d'appartenance à un groupe sont également fortement liées au conformisme, qui a été étudiée par Bernheim (1994) et étendue et formellement clarifiée par Gillen (2015). Dans cette approche, les fonctions d'utilité des agents de divers types incluent un terme pour le statut, qui à l'équilibre est lui-même une fonction du type des agents. Dans certaines conditions, lorsque le statut joue un rôle important dans l'utilité, il existe alors un équilibre de type *pooling* endogène où de nombreux agents entreprennent la même action, tandis que seuls ceux qui ont des préférences extrêmes sont incités à s'en éloigner (et reçoivent un statut inférieur).

Les noms ont également été largement étudiés en physique statistique. Plusieurs publications suivant des méthodes inspirées des sciences physiques se sont en effet penchées sur la distribution des noms, qu'il s'agisse de noms de famille ou de prénoms. Comme pour de nombreux phénomènes sociaux, les modèles de distribution des noms de famille suivent des lois de puissance. Zanette & Manrubia (2001) ont proposé un modèle de croissance aléatoire simple rendant compte de ce comportement. Baek, Kiet, & Kim (2007) et Maruvka, Shnerb, & Kessler (2010) ont appliqué des modèles de dynamique des populations afin de tenir compte de la croissance du nombre et de la distribution des noms de famille. Le premier article montre que la forme des distributions des noms de famille peut s'expliquer par la façon dont le taux d'introduction de nouveaux noms est exprimé. D'autres études ont réalisé des analyses purement empiriques de certains aspects spécifiques des distributions de noms, comme la dynamique ou la corrélation, notamment Hahn & Bentley (2003), Mateos, Longley, & O'Sullivan (2011), Barucca et al. (2015) et Lee et al. (2016), qui se sont penchés sur les prénoms.

8.4.1 Modèle

Afin d'aborder l'analyse des noms grecs anciens, nous développons dans ce chapitre un modèle du comportement optimal que les individus adopteront entre eux en fonction de l'information dont ils disposent sur leurs homologues : leur nom. Considérant différents degrés d'information, nous justifions l'échange de noms dans ce contexte comme le meilleur système pour promouvoir la coopération. Ensuite, en nous appuyant sur ce modèle, nous étudions le choix optimal des prénoms pour les enfants, compte tenu de cet équilibre. Avec quelques simplifications, nous montrons que l'impact de la dénomination stratégique sur la distribution des noms fonctionne comme une modification des approches typiques de champ moyen de la dynamique des noms, qui convergent vers une loi de puissance. La dénomination stratégique ajoute un degré de liberté dans la relation entre le nombre observé de noms et la forme de la distribution de la loi de puissance.

Nous considérons que toutes les personnes interagissent par paires au cours de leur vie, et lorsqu'elles le font, elles peuvent choisir d'agir équitablement ou de tricher. Des incitations sont en place pour essayer de dissuader les gens d'agir injustement au sein de la famille : lorsqu'ils le font, ils reçoivent une punition (une réduction de l'utilité ou une réparation qui doit être payée). Lorsqu'une personne agit injustement à l'égard d'un membre de sa propre famille, elle sera retrouvée individuellement. Il s'agit d'une famille au sens large, comprenant toutes les personnes liées par des relations de clientèles. Lorsque des personnes agissent de manière injuste avec des personnes extérieures à la famille, elles ne peuvent pas être facilement retrouvées et, par conséquent, elles ne subissent pas de réduction d'utilité. Bien qu'il puisse y avoir des lois qui punissent les crimes en général, la tricherie que nous décrivons ici n'est pas envisagée comme une infraction grave. Conformément à ce que nous pouvons comprendre des relations interpersonnelles dans la Grèce antique, nous supposons que lorsque les gens se rencontrent, ils divulguent leurs noms mais ils ne divulguent pas à quelle famille ils appartiennent, car soit ils ne le souhaitent pas, soit ils peuvent ne pas être en mesure de prouver ce qu'ils affirment. Nous considérons que le nom lui-même est nécessairement donné de manière véridique, la décision potentielle de tricher ne venant qu'après que l'un connaît le nom de l'autre. Par conséquent, il n'y a pas d'information parfaite sur la famille d'une personne que l'on rencontre. Les informations sur la répartition des noms

dans les familles et sur leurs tailles sont connues de tous, en revanche.

On peut caractériser plusieurs équilibres possibles, selon l'information dont dispose la population : information parfaite, aucune information (même pas sur sa propre famille), connaissance de sa famille, et enfin connaissance de sa famille et du nom de l'autre, qui est le cas le plus représentatif. On peut mesurer l'optimum social global atteint dans chaque cas. Lorsque l'information est parfaite, l'incitation à tricher est élevée, et le gain total n'est pas très élevé. Si aucune information n'est disponible, alors la situation est plus optimale socialement, il y a moins d'incitation à tricher. Lorsque la punition pour ceux qui trichent est coûteuse par rapport au gain potentiel, alors l'accès à des informations privées réduit la probabilité de tricher. En effet, si la répartition des personnes dans les familles est suffisamment étalée, alors le fait d'avoir des informations privées n'est pas préjudiciable au bien social, par rapport au cas sans information. Enfin, dans le cas où les informations privées sont connues, et certaines informations sur la contrepartie sont également connues, si la distribution des noms est bien répartie entre les familles, alors cette connaissance ne conditionne pas grand-chose, et par conséquent le gain agrégé est comparable à celui qui n'a que des informations privées ou à celui qui n'a aucune information du tout. En particulier, si la punition est suffisamment sévère, alors le gain social est le même que s'il n'y avait aucune information. L'utilisation des noms comme information importante, cohérente avec la façon dont nous nous représentons les interactions dans la Grèce antique, ne conduit pas à un optimum social plus faible dans ce cas que dans le cas sans information. L'échange de noms peut donc remplir un rôle social, en permettant aux personnes d'échanger quelque chose avant d'entrer dans une relation plus exigeante, ce qui ne serait pas nécessaire dans les trois cadres d'information alternatifs dont nous avons parlé plus haut, sans un coût social important par rapport à ceux-ci. On montre également que les noms qui apparaîtraient comme les plus attractifs selon cette logique sont ceux qui sont les plus répandus en dehors de la famille.

À partir de ce cadre, nous pouvons examiner le choix optimal d'un nom. À la lumière du contexte historique que nous avons évoqué précédemment, nous considérons que le choix des noms observé dans une *polis* grecque peut provenir de trois sources distinctes : un choix stratégique, le choix traditionnel du nom du père ou du grand-père, ou l'émergence aléatoire de nouveaux noms. Nous considérons que le stock de chaque

nom croît dans le temps en fonction de ces trois sources et que les enfants appartiennent à une famille unique, qui peut être celle de la mère ou celle du père. Les individus choisissent le nom le plus attractif, ou d'un parent, ou d'une source externe à tout moment, sans avoir à prendre en compte le comportement des autres, puisque le changement de population est supposé être très faible par rapport au total taille de la population.

S'il est optimal de choisir le nom le plus rare, et si de nouveaux noms apparaissent parfois (par invention ou par immigration), alors ces noms sembleraient naturellement optimaux, car ils seraient très rares au début. De plus, au fur et à mesure que les noms les plus rares sont choisis, ils deviennent moins rares jusqu'à ce qu'ils finissent par rattraper d'autres noms. Cette situation peut conduire à une répartition uniforme des noms, les noms ne transmettant alors pas d'informations particulières. Si au contraire le nom le plus fréquent est choisi, alors il peut devenir le seul nom existant, auquel cas le contenu informationnel des noms disparaîtrait également. Plus généralement, si le coût lorsque l'on triche n'est pas extrême et qu'il y a beaucoup de familles, alors le choix optimal semblerait osciller entre les noms les plus fréquents, ou n'importe quel nom. Sans pouvoir caractériser plus précisément le choix optimal, il est raisonnable de l'approximer par un choix aléatoire pondéré par la fréquence des noms, comme un mélange entre une sélection des noms les plus courants et un choix totalement aléatoire.

À partir de ces principes, et en utilisant certains aspects du modèle dynamique d'évolution des noms de Baek, Kiet, & Kim (2007), nous montrons que les aspects particuliers du cas grec conduisent à des propriétés différentes du modèle résultant. Nous montrons qu'il existe un degré de liberté supplémentaire dans la relation entre le nombre de noms par personne et le comportement limite de la distribution des noms.

8.4.2 Résultats

Nous examinons les résultats empiriques concernant les noms dans la Grèce antique et les confrontons aux modèles que nous avons discuté jusqu'ici. Nous utilisons les données du LGPN. Cependant, comme le LGPN couvre plus d'un millénaire avec des évolutions politiques et sociales très importantes, il semble pertinent de le restreindre aux périodes archaïques et classiques, plus anciennes, afin de conserver une certaine

unité culturelle, on ne retient donc que les observations antérieures à 350 av. J.-C. De plus, nous savons que bien que le LGPN suive les noms des hommes et des femmes, ces derniers ne représentent qu'une fraction du total. Afin de rester plus strictement dans le cadre de notre modèle, nous ne retenons donc que les noms d'hommes dans notre échantillon.

Comme Baek, Kiet, & Kim (2007) l'ont signalé, on observe généralement une relation entre la taille du groupe et le nombre de noms observés dans le groupe. La relation est logarithmique dans certains cas (pour la Chine et la Corée contemporaines), ou algébrique avec un exposant proche de 1 (pour de nombreux autres pays). Notre modèle prédit une relation linéaire, ce qui est le cas dans les observations ; la présence d'un exposant légèrement inférieur à 1 pourrait s'expliquer par une diminution de l'attractivité des grandes cités, de sorte que le taux d'immigration ou d'innovation de nom diminue légèrement lorsque la population augmente. La pente nous indique que le taux d'immigration ou d'innovation des noms s'élève à environ la moitié de la croissance démographique.

La distribution des noms en agrégé quant à elle suit une loi de puissance ou une loi de puissance tronquée. Au niveau de chaque *polis*, ces distributions ressemblent davantage à des lois de puissance tronquées ou à des distributions exponentielles, mais la dispersion des paramètres pour les lois de puissance est beaucoup moins dispersée que pour ces autres types de distributions. Les coefficients des lois de puissance, autour de 1,3, sont assez différents de ce qui a été observé sur les données contemporaines. Comme Baek, Kiet, & Kim (2007) l'ont souligné, pour la plupart des grandes villes ou pays du monde, les coefficients de la loi de puissance pour la distribution des noms ont tendance à se regrouper autour de 2. Dans quelques cas particuliers, comme la Corée, la loi de puissance a un paramètre proche de 1, et le nombre de noms croît de manière logarithmique en fonction de la population totale. Une étude, centrée sur la Sicile rurale, a trouvé des coefficients plus proches de 1 et dans certains cas même inférieurs à 1 (Pavesi et al. 2003). L'ancien nom de la Sicile, *Magna Graeca*, nous rappelle qu'une partie de sa population venait de Grèce à l'époque archaïque, il est donc intéressant que cette forme particulière dans la distribution du nom, où le paramètre est proche de 1, s'y retrouve.

8.5 Analyse cliométrique du choix des dieux en Grèce ancienne

Les anciens Grecs vénéraient de nombreux dieux, et les noms du panthéon des douze dieux sont bien connus : Zeus, Apollon et Poséidon sont des noms familiers. Cette familiarité cache en fait à quel point ces anciens dieux grecs sont encore activement examinés par les historiens et les philologues. Un angle particulier par lequel la compréhension de la religion antique a été approfondie est l'étude de la manière dont on s'adressait aux dieux. Comme l'ont notamment souligné Brulé (1998), Brulé (2005) et Parker (2003), on ne peut pas envisager un dieu comme une entité parfaitement définie, car la notion divine attachée à un nom varie selon les *poleis*, et les qualifications avec lesquelles on s'adresse à eux varient considérablement. Au premier ordre, on peut se demander comment la logique selon laquelle les gens adoraient plusieurs dieux pouvait fonctionner de manière optimale. Dans ce chapitre, nous cherchons une explication endogène optimale pour les choix par les anciens Grecs de quel(s) dieu(x) adorer. Nous développons un modèle économique simple qui relie les choix des dieux à la performance supposée de ces dieux dans un cadre concurrentiel.

Le recours à la modélisation formelle pour rendre compte des phénomènes liés à la religion n'est pas nouveau. En effet, dans les études sur la religion, la théorie des jeux a été utilisée, quoique de façon peu extensive, pour expliquer l'existence et les caractéristiques saillantes de la religion en général. Étendant le domaine des études religieuses, Bulbulia (2010) et Bulbulia & Frean (2010) ont proposé un modèle évolutif de coopération religieuse et ont montré comment la culture sacrée améliore la coopération entre partenaires anonymes. Dans son petit livre, Chwe (2001) a discuté de nombreux aspects dans lesquels le rituel religieux peut être analysé comme une action rationnelle du point de vue de la simple théorie des jeux.

Plus précisément du point de vue de la recherche en économie, Carr & Landa (1983) ont abordé la religion comme faisant partie de l'étude de l'identité et de la décision d'appartenance à un groupe. Couvrant un terrain plus large, Iannaccone (1998) et Iyer (2016) présentent des revues de littérature sur l'économie de la religion sous toutes ses facettes. La recherche économique s'est concentrée sur la structure des organisations religieuses, entre autres. Par exemple, Levy & Razin (2007) et Levy & Razin (2012) ont utilisé des jeux symétriques de dilemme du prisonnier pour représenter

les organisations religieuses et ont montré que les plus exigeantes attirent moins de membres, mais qu'elles sont plus cohésives. Iannaccone, Haight, & Rubin (2011) montrent comment les dirigeants laïques influencent l'emplacement des lieux sacrés. Les lieux sacrés acquièrent alors une autorité généralisée précisément parce qu'ils se situent au-delà des centres du pouvoir séculier. Dans McBride (2016), l'accent est mis sur la façon dont les rituels créent l'autorité religieuse. Levy & Razin (2014) développent un modèle de signal pour les manifestations de comportement religieux et les actes de coopération dans les organisations religieuses. Ils montrent que les religions basées sur les rituels nécessitent une implication coûteuse, qui agit comme un signal et implique également un niveau plus élevé de coordination dans les interactions sociales et une incidence plus élevée de la coopération mutuelle.

Les tests empiriques ont confirmé de nombreux modèles théoriques de la religion basés sur l'économie. Dans Ruffle & Sosis (2007), les rituels religieux chronophages sont expliqués comme favorisant la coopération, et les auteurs ont proposé des tests empiriques sur le comportement coopératif des membres des kibboutz israéliens religieux et laïcs. Power (2017) a montré de manière empirique que les gens sont attentifs aux actes religieux accomplis par leurs pairs, en utilisant ces signaux visibles pour discerner de multiples aspects de leur caractère et de leurs intentions.

Alors que la littérature mentionnée ci-dessus, tant dans l'étude de la religion qu'en économie, s'est concentrée sur les explications fondamentales de l'existence de la religion ou de certains de ses rituels, notre approche sera différente. Notre objectif initial n'est pas la raison pour laquelle la religion peut exister de manière optimale sous une certaine forme, mais plutôt, étant donné qu'elle existe, comment elle peut être pratiquée de manière optimale du point de vue à la fois de la population et des caractéristiques supposées des dieux. En particulier, les aspects spécifiques de la religion polythéiste joueront un rôle important dans notre analyse. Il existe peu d'études empiriques en sciences de la religion qui ont examiné le polythéisme. Il existe également peu de recherches sur le polythéisme dans le cadre de la théorie de l'économie religieuse. Ferrero & Tridimas (2018) ont développé un modèle où les gens supposaient un certain comportement des dieux et prenaient leurs propres décisions en conséquence de cette logique de décision supposée. Ils arguent que le chevauchement entre les différentes prérogatives des dieux a créé des inefficacités, et cela pourrait expliquer la convergence

vers le monothéisme observée à la fin de l'Antiquité. Dans notre approche, nous ne ferons pas de distinction entre les juridictions des différents dieux, mais nous considérerons plutôt leurs différences en termes d'efficacité pour offrir des avantages à leurs adorateurs mortels.

8.5.1 Modèle

Nous développons un modèle théorique qui explique la distribution des choix du ou des dieux à vénérer au sein d'une *polis* en fonction de certaines qualités supposées des dieux. Nous relierons ensuite ces qualités à la performance humaine dans les sports ou les jeux de compétition, et en déduisons que les observations d'actes votifs peuvent suivre une loi de puissance tronquée, associée à une distribution uniforme. Nous développons également un modèle de croissance aléatoire dans lequel les gens mettent à jour leur perception de l'efficacité des dieux en fonction d'événements aléatoires, et montrons que dans ce cas, la distribution des actes votifs devrait également suivre approximativement une loi de puissance.

Lorsque plusieurs personnes choisissent d'adorer le même dieu en même temps, nous supposons que le bénéfice total rendu par le dieu est partagé équitablement. Plus le montant total de bénéfice qu'un dieu peut fournir à ses adorateurs est élevé, plus la probabilité optimale de choisir ce dieu est grande, bien que la stratégie optimale rende aléatoires tous les choix possibles, et même un dieu qui offrirait très peu voit une probabilité non nulle d'être adoré. Tous les dieux reçoivent donc un culte. Sur la base de ce premier modèle, où nous avons considéré que le montant total de bénéfice pour ses adorateurs était donné pour chaque dieu, nous ajoutons la possibilité supposée qu'ont les dieux de choisir ce qu'ils veulent offrir, en fonction d'un coût qui leur est propre. Nous supposons que chaque dieu veut maximiser la proportion de personnes qui viennent l'adorer. La littérature grecque classique, comme les épopées ou la tragédie, regorge d'exemples de dieux prenant des mesures punitives contre les mortels lorsqu'ils ne les respectaient pas correctement. Nous savons aussi que les dieux étaient supposés bénéficier d'une partie du produit des sacrifices, qui serait d'autant plus importante si les gens y participent davantage. Cette approche nous permet alors d'établir une relation entre les différences d'efficacité des dieux pour accomplir des actes en faveur de leurs fidèles et la probabilité qu'ils soient adorés.

Nous tournons alors vers deux approches différentes qui peuvent expliquer la répartition des actes votifs que l'on doit s'attendre à observer, en rendant compte de la manière dont l'efficacité des dieux peut être conçue par les humains. Pourquoi les dieux devraient-ils posséder différents degrés de capacité à offrir des bénéfices aux humains ? Il semble raisonnable de considérer que cela pourrait être le cas parce que les dieux sont supposés par la population être dans une certaine mesure en concurrence les uns avec les autres et, par conséquent, auraient acquis différents niveaux d'expertise. Les anciens Grecs connaissaient la compétition entre eux, en particulier dans les événements sportifs. En fait, les vainqueurs exceptionnels des événements sportifs étaient vénérés par la poésie. En gagnant et en faisant preuve d'une compétence ou d'une force rare, ils se différenciaient des autres et étaient souvent qualifiés de divins. Il a été établi que les performances physiques et cognitives humaines suivent des distributions de loi de puissance. Donner & Hardy (2015) montrent comment des lois de puissance peuvent représenter les courbes d'apprentissage humain pour une variété de tâches, par exemple. Les temps de course record en fonction de la distance se sont également avérés suivre des lois de puissance (Vandewalle 2018). Les classements de la performance humaine suivent également divers types de lois de puissance. Par exemple, Smith (2015) documente la présence de lois de puissance dans plusieurs actions psychomotrices. Plus spécifiquement pour le sport, Deng et al. (2012) a montré que dans un large éventail de sports (tennis, football, snooker et bien d'autres), les performances en fonction du classement suivaient des lois de puissance tronquées.

Une approche alternative pour expliquer la forme de la distribution supposée de l'efficacité des dieux consiste à tenter d'expliquer comment elle a pu évoluer. On peut en effet formaliser la distribution des choix des actes votifs en considérant qu'elle serait basée sur une évolution des croyances dans le temps. Puisque les gens demandent diverses choses aux dieux, nous modélisons le fait que certains des souhaits sont exaucés (par pur hasard), et que sur la base de cette observation, les gens mettront à jour leurs croyances sur l'efficacité des dieux. Les bénéfices observés de l'adoration de chaque dieu suivent alors un modèle de croissance aléatoire (Gabaix 2016.). Nous avons considéré jusqu'ici que l'efficacité des dieux formait un continuum unidimensionnel : leur juridiction spécifique n'entrait pas en ligne de compte dans l'analyse et seule leur efficacité à rendre des bénéfices importait. Cependant, tout en gardant la même

logique générale, nous considérons également que les personnes qui sacrifient à un dieu particulier auront formulé leurs demandes dans le contexte des compétences spécifiques de ce dieu. Par conséquent, deux personnes demandant la même chose fondamentale à deux dieux distincts la formuleraient en des termes différents ; par conséquent, ils n'observeraient pas si leurs demandes ont été satisfaites de la même manière à l'avenir. Les événements aléatoires de la vie affectent donc les personnes qui adorent différents dieux de différentes manières. Nous considérerons également qu'il existe des événements systémiques de la vie qui affectent tout le monde de la même manière. Plus les gens ont commencé à adorer un dieu, plus le résultat positif des événements de la vie affectant ces adorateurs est remarqué, vraisemblablement. En effet, le montant total des résultats positifs perçus pour un dieu donné à un moment futur est la part de la population adorant initialement ce dieu, démultipliée par tous les résultats positifs au fil du temps. Avec ces hypothèses, nous montrons que, étant donné des conditions de départ aléatoires et compte tenu d'un large éventail de distributions potentielles d'événements de la vie affectant la perception des gens de ce que les dieux leur paient comme bénéfiques, nous pourrions observer empiriquement que les actes votifs sont répartis entre les dieux selon de simples lois de puissance.

8.5.2 Résultats

Nous structurons les données disponibles issues de la BDEG afin de pouvoir mesurer la répartition des actes votifs entre les dieux et testons le modèle théorique. Les données sur les actes votifs de la BDEG sont croisées avec les données de la base POLIS de Stanford, qui contient certaines informations agrégées au niveau de chaque cité, comme notamment sa taille, son régime politique (bien que celui-ci ait souvent changé au cours du temps), ou encore l'importance de sa population grecque.

Le nombre de dieux vénérés par la population dans un lieu donné peut dépendre de divers facteurs, si l'on considère qu'il existe un seuil en dessous duquel un dieu peu aimé ne recevrait pas d'actes votifs. En sélectionnant les facteurs significatifs dans une régression du nombre de dieux par *polis*, on peut expliquer une certaine fraction de ce nombre. Les facteurs qui semblent à la fois significatifs et importants sont la « notoriété » et la part des sources épigraphiques pour les données BDEG. Une plus grande utilisation des sources épigraphiques dans un lieu particulier pourrait expliquer

que nous avons trouvé plus de restes d'actes votifs que nous n'en aurions autrement. La renommée fait référence à l'importance d'une *polis* dans les sources littéraires, et plus son importance est grande, plus le nombre de personnes qui ont transité par le lieu est important, laissant ainsi potentiellement une trace de culte. De manière générale, le modèle prédit un certain rapport entre le nombre de dieux et la « part de marché » du dieu le moins adoré. Les régressions des valeurs prédites par rapport aux valeurs réelles nous indiquent que le modèle est proche de la réalité, mais surestime en partie le nombre d'actes votifs. Cette surestimation des données pourrait s'expliquer par le fait que les inscriptions trouvées ne sont pas nécessairement complètes pour les dieux qui n'apparaissent pas souvent, et par conséquent le nombre de dieux distincts dérivés des données est probablement sous-estimé.

Si l'on se penche maintenant sur toutes les *poleis* séparément, on peut observer que les distributions des actes votifs parmi les dieux suivent bien des lois de puissance. Nous pouvons voir que les coefficients (à la fois pour les lois de puissance pures et tronquées) sont assez concentrés. La dispersion relative des paramètres est la plus faible pour la loi de puissance pure. Cela indique que les mécanismes générateurs sont assez similaires dans un grand nombre de cités différentes. Bien que nous puissions rendre compte de la forme globale de la distribution des actes votifs entre les dieux, une question intéressante est de déterminer ce qui, dans les caractéristiques de chaque *polis*, peut entraîner des différences dans ces distributions. Puisque pour chaque *polis* nous possédons une série de variables catégorielles ou numériques, nous pouvons observer dans quelle mesure les différences perçues entre les dieux sont impactées par certaines caractéristiques de la *polis*. En testant nombre de tels facteurs dans des régressions, nous relevons que les paramètres de ces lois de puissance sont étroitement distribués, et les coefficients de régression sont faibles. En fait, pratiquement aucune variable n'a un effet vraiment fort ou significatif.

On peut lier le coefficient des lois de puissance au degré supposé de compétitivité entre les dieux, dans le cadre de la perspective anthropomorphe qui dériverait l'efficacité perçue du dieu de l'observation des performances humaines. Si nous nous plaçons dans le contexte d'un modèle de croissance aléatoire expliquant l'efficacité des dieux, alors il n'y aurait pas de raison particulière pour que certains dieux soient supposés être meilleurs que d'autres dieux, autre que le pur hasard, mais la pente de

la distribution serait liée au degré de dispersion des événements aléatoires affectant la perception de l'efficacité de chaque dieu. À la lumière de ces considérations, l'hypothèse de croissance aléatoire apparaît la plus appropriée pour expliquer la distribution empirique des références épigraphiques et littéraires au culte en raison d'un manque de relation avec les caractéristiques de l'environnement local.

Full Bibliography

- Abry, Patrice, Cécile Bastidon, Pierre Borgnat, Pablo Jensen, & Antoine Parent.** 2022. "The Contours of a Cliophysics: How Can Econophysics Enrich Cliometrics? Case Studies in Debt Issues and Global Capital Markets." *Frontiers in Physics* 10: 894400.
- Akerlof, George A., & Rachel E. Kranton.** 2000. "Economics and Identity." *The Quarterly Journal of Economics* 115 (3): 715–753.
- Alexandrova, Anna.** 2009. "When Analytic Narratives Explain." *Journal of the Philosophy of History* 3 (1): 1–24.
- Alfarano, Simone, & Thomas Lux.** 2010. "Extreme Value Theory as a Theoretical Background for Power Law Behavior." Working Paper 1648. Kiel Working Papers. Kiel Institute for the World Economy. <https://ideas.repec.org/p/kie/kieliw/1648.html>.
- Alstott, Jeff, Ed Bullmore, & Dietmar Plenz.** 2014. "Powerlaw: A Python Package for Analysis of Heavy-Tailed Distributions." *PLOS ONE* 9 (1): e85777.
- Aron, Raymond.** 1981. "Quelques remarques sur la compréhension et l'explication." *Revue européenne des sciences sociales* 19 (54-55): 71–82.
- . (1938) 1991. *Introduction à la philosophie de l'histoire : essai sur les limites de l'objectivité historique*. Tel 58. Paris: Gallimard.
- Arroyo Abad, Leticia, & Kareem Khalifa.** 2015. "What Are Stylized Facts?" *Journal of Economic Methodology* 22 (2): 143–156.
- Arthur, W. Brian.** 1994. "Inductive Reasoning and Bounded Rationality." *The American Economic Review* 84 (2): 406–411.
- . 2021. "Foundations of Complexity Economics." *Nature Reviews Physics* 3 (2): 136–145.
- Assael, Yannis, Thea Sommerschild, & Jonathan Prag.** 2019. "Restoring Ancient Text Using Deep Learning: A Case Study on Greek Epigraphy." In *Conference on Empirical Methods in Natural Language Processing*, edited by Jing Jiang, Vincent Ng, & Xiaojun Wan, 6369–6376. Stroudsburg: Association for Computational Linguistics.
- . 2020. "Sommerschild/Ancient-Text-Restoration." June 6, 2020. <https://github.com/sommerschild/ancient-text-restoration>.

- Axelrod, Robert M.** 2006. *The Evolution of Cooperation*. 2nd ed. New York: Basic Books.
- Axelrod, Robert M., & William D. Hamilton.** 1981. "The Evolution of Cooperation." *Science, New Series*, 211 (4489): 1390–1396.
- Baek, Seung Ki, Sebastian Bernhardtsson, & Petter Minnhagen.** 2011. "Zipf's Law Unzipped." *New Journal of Physics* 13 (4): 043004.
- Baek, Seung Ki, Hoang Anh Tuan Kiet, & Beom Jun Kim.** 2007. "Family Name Distributions: Master Equation Approach." *Physical Review E* 76 (4): 046113.
- Baltas, George.** 2003. "Determinants of Internet Advertising Effectiveness: An Empirical Study." *International Journal of Market Research* 45 (4): 1–9.
- Bankston, Carl L.** 2002. "Rationality, Choice and the Religious Economy: The Problem of Belief." *Review of Religious Research* 43 (4): 311–325.
- Barabási, Albert-László, & Réka Albert.** 1999. "Emergence of Scaling in Random Networks." *Science* 286 (5439): 509–512.
- Barker, Elton, & Melissa Terras.** 2016. "Greek Literature, the Digital Humanities, and the Shifting Technologies of Reading." *Oxford Handbooks Online* 2016: 1–25.
- Barucca, Paolo, Jacopo Rocchi, Enzo Marinari, Giorgio Parisi, & Federico Ricci-Tersenghi.** 2015. "Cross-Correlations of American Baby Names." *Proceedings of the National Academy of Sciences* 112 (26): 7943–7947.
- Bastidon, Cécile, Michael Bordo, Antoine Parent, & Marc Weidenmier.** 2019. "Towards an Unstable Hook: The Evolution of Stock Market Integration Since 1913." Working Paper w26166. Cambridge: National Bureau of Economic Research. <https://doi.org/10.3386/w26166>.
- Bastidon, Cécile, & Antoine Parent.** 2022. "Cliometrics of World Stock Markets Evolving Networks." *Annals of Operations Research* To appear.
- Bastidon, Cécile, Antoine Parent, Pablo Jensen, Patrice Abry, & Pierre Borgnat.** 2020. "Graph-Based Era Segmentation of International Financial Integration." *Physica A: Statistical Mechanics and Its Applications* 539: 122877.
- Bates, Robert H., Avner Greif, Margaret Levi, Jean-Laurent Rosenthal, & Barry R. Weingast.** 1998. *Analytic Narratives*. Princeton: Princeton University Press.
- Beigman Klebanov, Beata, & Eyal Beigman.** 2010. "A Game-Theoretic Model of Metaphorical Bargaining." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 698–709. Uppsala: Association for Computational Linguistics.
- Berg, Alastair.** 2019. "Identity in Economics: A Review." Working Paper ID 3421215. Rochester: Social Science Research Network. <https://doi.org/10.2139/ssrn.3421215>.

- Bermejo Barrera, Jose Carlos.** 2001. "Making History, Talking about History." *History and Theory* 40 (2): 190–205.
- Bernheim, B. Douglas.** 1994. "A Theory of Conformity." *Journal of Political Economy* 102 (5): 841–877.
- Besley, Timothy, & Torsten Persson.** 2009. "The Origins of State Capacity: Property Rights, Taxation, and Politics." *American Economic Review* 99 (4): 1218–1244.
- Bitros, George, Emmanouil M. L. Economou, & Nicholas C. Kyriazis.** 2021. *Democracy and Money: Lessons for Today from Athens in Classical Times*. Routledge International Studies in Money and Banking. London: Routledge.
- Boccaro, Nino.** 2010. *Modeling Complex Systems*. Graduate Texts in Physics. New York: Springer.
- Boldizzoni, Francesco.** 2011. *The Poverty of Clio: Resurrecting Economic History*. Princeton: Princeton University Press.
- Bonnet, Corinne.** 2017. "Mapping Ancient Polytheisms." 2017. <https://map-polytheisms.huma-num.fr/>.
- , ed. 2021. *Noms de Dieux : Portraits de Divinités Antiques*. Essais - Histoire. Toulouse: Anacharsis.
- Bonnet, Corinne, & Nicole Belayche, eds.** 2017. *Puissances divines à l'épreuve du comparatisme : constructions, variations et réseaux relationnels*. Bibliothèque de l'École des Hautes Études, sciences religieuses, volume 175. Turnhout: Brepols.
- Bonnet, Corinne, & Sylvain Lebreton.** 2019. "Mettre les polythéismes en formules ? À propos de la base de données Mapping Ancient Polytheisms." *Kernos* 32: 267–296.
- Borwein, David, Jonathan M. Borwein, & Pierre Maréchal.** 2000. "Surprise Maximization." *The American Mathematical Monthly* 107 (6): 517–527.
- Boulding, Kenneth E.** 1948. "Samuelson's Foundations: The Role of Mathematics in Economics." *Journal of Political Economy* 56 (3): 187–199.
- Bourdieu, Pierre.** 2017. *Anthropologie économique : cours au Collège de France, 1992-1993*. Raisons d'agir. Paris: Seuil.
- Boutier, Jean.** 2014. "L'usage historique des archives." In *Corpus, sources et archives*, by Jean Boutier, Jean-Louis Fabiani, & Jean-Pierre Olivier de Sardan, 9–22. Études et travaux de l'IRMC. Tunis: Institut de recherche sur le Maghreb contemporain.
- Boyd, Robert, & Peter J. Richerson.** 2005. *The Origin and Evolution of Cultures*. Evolution and Cognition. Oxford: Oxford University Press.
- Brams, Steven J.** 1994. "Game Theory and Literature." *Games and Economic Behavior* 6 (1): 32–54.

- Bresson, Alain.** 1981. "Règles de nomination dans la Rhodes antique." *Dialogues d'histoire ancienne* 7 (1): 345–362.
- . 2019. "Chapitre IV. Une famille camiréenne de commerçants en blé." In *La cité marchande*, 95–99. Scripta Antiqua. Pessac: Ausonius Éditions.
- Broux, Yanne, & Willy Clarysse.** 2016. "Would You Name Your Child After a Celebrity? Arsinoe, Berenike, Kleopatra, Laodike and Stratonike in the Greco-Roman East." *Zeitschrift Für Papyrologie Und Epigraphik* 200: 347–362.
- Brownlow, Graham A.** 2010. "Structure and Change: Douglass North's Economics." *Journal of Economic Methodology* 17 (3): 301–316.
- Bruit Zaidman, Louise, & Pauline Schmitt Pantel.** 1992. *Religion in the Ancient Greek City*. Cambridge: Cambridge University Press.
- Brulé, Pierre.** 1998. "Le langage des épiclèses dans le polythéisme hellénique (l'exemple de quelques divinités féminines). Quelques pistes de recherches." *Kernos* 11 (11): 13–34.
- . 2005. "Le polythéisme en transformation : les listes de dieux dans les serments internationaux en Grèce antique (V^e-II^e siècles)." In *Nommer les dieux : Théonymes, épithètes, épiclèses dans l'Antiquité*, edited by Nicole Belayche, Pierre Brulé, Gérard Freyburger, Yves Lehman, Laurent Pernot, & Antoine Prost, 143–173. Recherches sur les rhétoriques religieuses. Turnhout: Brepols.
- Brulé, Pierre, & Sylvain Lebreton.** 2007. "La Banque de données sur les épiclèses divines (BDDE) du Crescam : sa philosophie." *Kernos* 20: 217–228.
- Bulbulia, Joseph.** 2010. "Charismatic Signalling." *Journal for the Study of Religion, Nature and Culture* 3 (4): 518–551.
- Bulbulia, Joseph, & Marcus Frean.** 2010. "The Evolution of Charismatic Cultures." *Method & Theory in the Study of Religion* 22 (4): 254–271.
- Busse, Thomas V., & Louisa Seraydarian.** 1977. "Desirability of First Names, Ethnicity and Parental Education." *Psychological Reports* 40 (3): 739–742.
- Calero Valdez, André.** 2020. "Making Reproducible Research Simple Using RMarkdown and the OSF." In *Social Computing and Social Media. Design, Ethics, User Behavior, and Social Network Analysis*, edited by Gabriele Meiselwitz, 27–44. Lecture Notes in Computer Science 12194. Cham: Springer International Publishing.
- Canevaro, Mirko.** 2018. "What Was the Law of Leptines' Really About? Reflections on Athenian Public Economy and Legislation in the Fourth Century BCE." *Constitutional Political Economy* 29 (4): 440–464.

- Canevaro, Mirko, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, eds.** 2018. *Ancient Greek History and Contemporary Social Science*. Edinburgh Leventis Studies 9. Edinburgh: Edinburgh University Press.
- Carpenter, Jeffrey P.** 2004. "When in Rome: Conformity and the Provision of Public Goods." *The Journal of Socio-Economics* 33 (4): 395–408.
- Carr, Jack L., & Janet T. Landa.** 1983. "The Economics of Symbols, Clan Names, and Religion." *The Journal of Legal Studies* 12 (1): 135–156.
- Carroll, Stuart.** 2017. "Thinking with Violence." *History and Theory* 56 (4): 23–43.
- Carugati, Federica, Josiah Ober, & Barry R. Weingast.** 2016. "Development and Political Theory in Classical Athens." *Polis: The Journal for Ancient Greek and Roman Political Thought* 33 (1): 71–91.
- Čech, Radek, Ján Mačutek, & Haitao Liu.** 2016. "Syntactic Complex Networks and Their Applications." In *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*, edited by Alexander Mehler, Andy Lüicking, Sven Banisch, Philippe Blanchard, & Barbara Job, 167–186. Understanding Complex Systems. Berlin: Springer.
- Cesarano, Filippo.** 2006. "Economic History and Economic Theory." *Journal of Economic Methodology* 13 (4): 447–467.
- Chakrabarti, Anindya Sundar, Bikas K. Chakrabarti, Arnab Chatterjee, & Manipushpak Mitra.** 2009. "The Kolkata Paise Restaurant Problem and Resource Utilization." *Physica A: Statistical Mechanics and Its Applications* 388 (12): 2420–2426.
- Chakrabarti, Bikas K.** 2007. "Kolkata Restaurant Problem as a Generalised El Farol Bar Problem." In *Econophysics of Markets and Business Networks: Proceedings of the Econophys-Kolkata III*, edited by Arnab Chatterjee & Bikas K. Chakrabarti, 239–246. New Economic Windows. Milano: Springer.
- Chankowski, Véronique.** 2020. "Juste prix, prix libres et prix fixés dans les cités du monde grec antique : histoire d'un malentendu." In *Les infortunes du juste prix : marchés, justice sociale et bien commun de l'Antiquité à nos jours*, edited by Véronique Chankowski, Clément Lenoble, & Jérôme Maucourant, 25–68. Documents. Lormont: Le Bord de l'eau.
- Chao, M. T., & W. E. Strawderman.** 1972. "Negative Moments of Positive Random Variables." *Journal of the American Statistical Association* 67 (338): 429–431.
- Choi, Jay P.** 2011. "Up or Down? A Male Economist's Manifesto on the Toilet Seat Etiquette." *Economic Inquiry* 49 (1): 303–309.
- Chwe, Michael Suk-Young.** 2001. *Rational Ritual: Culture, Coordination, and Common Knowledge*. Princeton: Princeton University Press.

- Clauset, Aaron, Cosma Rohilla Shalizi, & M. E. J. Newman.** 2009. "Power-Law Distributions in Empirical Data." *SIAM Review* 51 (4): 661–703.
- Cline, Diane Harris.** 2020. "Athens as a Small World." *Journal of Historical Network Research* 4: 36–56.
- Coase, Ronald H.** 1960. "The Problem of Social Cost." *The Journal of Law and Economics* 3: 1–44.
- Crane, Gregory R.** 2012. "Perseus Digital Library." April 2012. <http://www.perseus.tufts.edu>.
- Crettez, Bertrand, & Régis Deloche.** 2018. "An Analytic Narrative of Caesar's Death: Suicide or Not? That Is the Question." *Rationality and Society* 30 (3): 332–349.
- Cuesta, Jose A., Carlos Gracia-Lázaro, Alfredo Ferrer, Yamir Moreno, & Angel Sánchez.** 2015. "Reputation Drives Cooperative Behaviour and Network Formation in Human Groups." *Scientific Reports* 5 (1): 7843.
- Davis, Otto A., Melvin J. Hinich, & Peter C. Ordeshook.** 1970. "An Expository Development of a Mathematical Model of the Electoral Process." *American Political Science Review* 64 (2): 426–448.
- Deng, Weibing, Wei Li, Xu Cai, Alain Bulou, & Qiuping A Wang.** 2012. "Universal Scaling in Sports Ranking." *New Journal of Physics* 14 (9): 093038.
- Deng, Weibing, Wei Li, Xu Cai, & Qiuping A. Wang.** 2011. "The Exponential Degree Distribution in Complex Networks: Non-Equilibrium Network Theory, Numerical Simulation and Empirical Data." *Physica A: Statistical Mechanics and Its Applications* 390 (8): 1481–1485.
- Devine, Andrew M., & Laurence Stephens.** 1976. "The Homeric Hexameter and a Basic Principle of Metrical Theory." *Classical Philology* 71 (2): 141–163.
- Diebolt, Claude.** 2011. "The Stakes of Cliometrics in Ancient History." *Historical Social Research / Historische Sozialforschung* 36 (3): 350–361.
- Diebolt, Claude, & Michael Hauptert, eds.** 2019. *Handbook of Cliometrics*. 2nd ed. Springer Reference. Cham: Springer International Publishing.
- . 2020. "How Cliometrics Has Infiltrated Economics – and Helped to Improve the Discipline." *Annals of the Fondazione Luigi Einaudi* 54 (1): 219–230.
- Diebolt, Claude, & Antoine Parent.** 2011. *Essais cliométriques : capital humain, monnaie et finance, croissance et cycles*. Bern: Lang.
- Dixit, Avinash.** 2012. "An Option Value Problem from *Seinfeld*." *Economic Inquiry* 50 (2): 563–565.
- Donner, Yoni, & Joseph L. Hardy.** 2015. "Piecewise Power Laws in Individual Learning Curves." *Psychonomic Bulletin & Review* 22 (5): 1308–1319.

- Durlauf, Steven, Antoine Parent, Cécile Bastidon, Patrice Abry, Pierre Borgnat, & Pablo Jensen.** 2023. *Complexity and Cliometrics*. Edited by Steven Durlauf & Antoine Parent. Vol. forthcoming. Elements in Complexity and Agent-Based Economics 1. Cambridge: Cambridge University Press.
- Economou, Emmanouil M. L., & Nicholas C. Kyriazis.** 2016. "Choosing Peace Instead of War: A Lesson from Athenian Democracy." *Peace Economics, Peace Science and Public Policy* 22 (2): 191–212.
- . 2017. "The Emergence and the Evolution of Property Rights in Ancient Greece." *Journal of Institutional Economics* 13 (1): 53–77.
- . 2019. *Democracy and Economy: An Inseparable Relationship from Ancient Times to Today*. Newcastle upon Tyne: Cambridge Scholars Publishing.
- . 2022. "Asymmetric Information, Social Choices, and Democracy." In *Democracy in Times of Crises: Challenges, Problems and Policy Proposals*, edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis, & Athanasios Platias, 117–136. Cham: Springer International Publishing.
- Ekroth, Gunnel.** 2011. "Meat for the Gods." In *Nourrir les dieux ?*, edited by Vinciane Pirenne-Delforge & Francesca Prescendi, 15–41. Liège: Presses Universitaires de Liège.
- Elson, David, Nicholas Dames, & Kathleen McKeown.** 2010. "Extracting Social Networks from Literary Fiction." In *Proceedings of the 48th Annual Meeting of the Association for Computational Linguistics*, 138–147. Uppsala: Association for Computational Linguistics.
- Ely, Jeffrey, Alexander Frankel, & Emir Kamenica.** 2015. "Suspense and Surprise." *Journal of Political Economy* 123 (1): 215–260.
- Febvre, Lucien.** 1934. "Comptabilité et Chambre des Comptes." *Annales d'histoire économique et sociale* 6 (26): 148–153.
- . (1952) 1992. *Combats pour l'histoire. L'ancien et le nouveau* 12. Paris: Armand Colin.
- Fenoaltea, Stefano.** 2019. "Spleen: The Failures of the Cliometric School." *Annals of the Fondazione Luigi Einaudi* 53 (2): 5–24.
- Ferrer i Cancho, Ramon, Oliver Riordan, & Béla Bollobás.** 2005. "The Consequences of Zipf's Law for Syntax and Symbolic Reference." *Proceedings of the Royal Society B: Biological Sciences* 272 (1562): 561–565.
- Ferrer i Cancho, Ramon, & Ricard V. Solé.** 2003. "Least Effort and the Origins of Scaling in Human Language." *Proceedings of the National Academy of Sciences of the United States of America* 100 (3): 788–791.

- Ferrero, Mario, & George Tridimas.** 2018. "Divine Competition in Greco-Roman Polytheism." *Homo Oeconomicus* 35 (3): 143–166.
- Fine, Ben, & Dimitris Milonakis.** 2009. *From Economics Imperialism to Freakonomics: The Shifting Boundaries Between Economics and Other Social Sciences*. London: Routledge.
- Fleck, Robert K., & F. Andrew Hanssen.** 2006. "The Origins of Democracy: A Model with Application to Ancient Greece." *The Journal of Law and Economics* 49 (1): 115–146.
- . 2013. "How Tyranny Paved the Way to Democracy: The Democratic Transition in Ancient Greece." *The Journal of Law and Economics* 56 (2): 389–416.
- . 2015. "The Foundations of Wealth-Enhancing Democracy: Aristotle, Lindahl, and Institutional Design in Ancient Greece." Working Paper. Clemson: Clemson University. <https://www.aeaweb.org/conference/2016/retrieve.php?pdfid=13727&tk=KZyieYK3>.
- . 2018. "What Can Data Drawn from the Hansen-Nielsen Inventory Tell Us about Political Transitions in Ancient Greece?" In *Ancient Greek History and Contemporary Social Science*, edited by Mirko Canevaro, Andrew Erskine, Benjamin D. Gray, & Josiah Ober, 213–238. Edinburgh: Edinburgh University Press.
- . 2019. "Engineering the Rule of Law in Ancient Athens." *The Journal of Legal Studies* 48 (2): 441–473.
- Foley, Robert.** 1985. "Optimality Theory in Anthropology." *Man* 20 (2): 222–242.
- Foucault, Michel.** 1969. *L'archéologie du savoir*. Paris: Gallimard.
- Fowler, Robert L., ed.** 2004. *The Cambridge Companion to Homer*. Cambridge Companions to Literature. Cambridge: Cambridge University Press.
- Gabaix, Xavier.** 1999. "Zipf's Law for Cities: An Explanation." *The Quarterly Journal of Economics* 114 (3): 739–767.
- . 2008. "Power Laws in Economics and Finance." Working Paper 14299. Working Paper Series. Washington: National Bureau of Economic Research. <https://doi.org/10.3386/w14299>.
- . 2016. "Power Laws in Economics: An Introduction." *Journal of Economic Perspectives* 30 (1): 185–206.
- Gaechter, Simon, Kyeongtae Lee, & Martin Sefton.** 2020. "Risk, Temptation, and Efficiency in Prisoner's Dilemmas." Working Paper 2020-15. Discussion Papers. University of Nottingham. <https://ideas.repec.org/p/not/notcdx/2020-15.html>.
- Gibbs, Fred, & Trevor Owens.** 2013. "The Hermeneutics of Data and Historical Writing." In *Writing History in the Digital Age*, edited by Jack Dougherty & Kristen Nawrotzki, 159–170. Digital Humanities. Ann Arbor: University of Michigan Press.

- Gillen, Benjamin J.** 2015. "A Note on Endogenous Norms in a Theory of Conformity." *Economics Letters* 129: 57–61.
- Goldie, Charles M.** 1991. "Implicit Renewal Theory and Tails of Solutions of Random Equations." *Annals of Applied Probability* 1 (1): 126–166.
- Gonzales-Feliu, Jesus, & Antoine Parent.** 2016. "Clio-Combinatorics: A Novel Framework to Analyze Military Logistics Choices Using Operations Research Techniques." In *Handbook of Research on Military, Aeronautical, and Maritime Logistics and Operations*, edited by Alberto Ochoa-Zezzatti, Jöns Sánchez, Miguel Gastón Cedillo-Campos, & Margain de Lourdes, 79–101. Advances in Logistics, Operations, and Management Science. Hershey: IGI Global.
- Grabska-Gradzińska, Iwona, Andrzej Kulig, Jarosław Kwapien, & Stanisław Drożdż.** 2012. "Complex Network Analysis of Literary and Scientific Texts." *International Journal of Modern Physics C* 23 (07): 1250051/1–15.
- Gries, Peter, Jenny Su, & David Schak.** 2012. "Toward the Scientific Study of Polytheism: Beyond Forced-Choice Measures of Religious Belief." *Journal for the Scientific Study of Religion* 51 (4): 623.
- Guha, Brishti.** 2011. "Preferences, Prisoners and Private Information: Was Socrates Rational at His Trial?" *European Journal of Law and Economics* 31 (3): 249–264.
- Hahn, Matthew W, & R Alexander Bentley.** 2003. "Drift as a Mechanism for Cultural Change: An Example from Baby Names." *Proceedings of the Royal Society B: Biological Sciences* 270: S120–S123.
- Halkos, George E., Emmanouil M. L. Economou, & Nicholas C. Kyriazis.** 2022. "Tracing the Optimal Level of Political and Social Change Under Risks and Uncertainties: Some Lessons from Ancient Sparta and Athens." *Journal of Risk and Financial Management* 15 (9): 416.
- Hansen, Mogens Herman, & Thomas Heine Nielsen.** 2004. *An Inventory of Archaic and Classical Poleis*. Oxford: Oxford University Press.
- Hardie, Margaret M.** 1923. "The Significance of Greek Personal Names." *Folklore* 34 (2): 149–154.
- Harrison, John, & Ju Yeong Kim.** 2020. "RSelenium: R Bindings for 'Selenium WebDriver'." February 3, 2020. <https://CRAN.R-project.org/package=RSelenium>.
- Hauptert, Michael.** 2019. "History of Cliometrics." In *Handbook of Cliometrics*, edited by Claude Diebolt & Michael Hauptert, 3–32. Cham: Springer International Publishing.
- Henrich, Joseph.** 2009. "The Evolution of Costly Displays, Cooperation and Religion: Credibility Enhancing Displays and Their Implications for Cultural Evolution." *Evolution and Human Behavior* 30 (4): 244–260.

- Herman, Gabriel.** 1990. "Patterns of Name Diffusion Within the Greek World and Beyond." *The Classical Quarterly* 40 (2): 349–363.
- Hirschman, Daniel.** 2016. "Stylized Facts in the Social Sciences." *Sociological Science* 3: 604–626.
- Hobson, Matthew S.** 2014. "A Historiography of the Study of the Roman Economy: Economic Growth, Development, and Neoliberalism." *Theoretical Roman Archaeology Journal* 2013: 11–26.
- Hoekstra, Rik, & Marijn Koolen.** 2019. "Data Scopes for Digital History Research." *Historical Methods: A Journal of Quantitative and Interdisciplinary History* 52 (2): 79–94.
- Hornblower, Simon.** 2010. *Thucydidean Themes*. Oxford: Oxford University Press.
- Hotelling, Harold.** 1929. "Stability in Competition." *The Economic Journal* 39 (153): 41–57.
- Iannaccone, Laurence R.** 1998. "Introduction to the Economics of Religion." *Journal of Economic Literature* 36 (3): 1465–1495.
- . 2005. "Economics of Religion. Debating the Costs and Benefits of a New Field." *Faith and Economics* 46: 1–9.
- Iannaccone, Laurence R., Colleen E. Haight, & Jared Rubin.** 2011. "Lessons from Delphi: Religious Markets and Spiritual Capitals." *Journal of Economic Behavior & Organization* 77 (3): 326–338.
- Ide, Nancy.** 2004. "Preparation and Analysis of Linguistic Corpora." In *A Companion to Digital Humanities*, edited by Susan Schreibman, Ray Siemens, & John Unsworth, 289–305. Malden: Blackwell Publishing.
- Ionesco, Eugène.** 1954. *La cantatrice chauve : anti-pièce ; suivi de La leçon ; drame comique*. Collection folio 236. Paris: Gallimard.
- Ismard, Paulin.** 2005. "Le désassemblage démocratique ou la cité antiplatonicienne." *Gestes*, 2005. <https://www.revue-geste.fr/articles/geste2/GESTE%202%20-%20Assembler%20-%20Ismard.pdf>.
- Iyer, Sriya.** 2016. "The New Economics of Religion." *Journal of Economic Literature* 54 (2): 395–441.
- Jackson, Matthew O., & Brian W. Rogers.** 2007. "Meeting Strangers and Friends of Friends: How Random Are Social Networks?" *American Economic Review* 97 (3): 890–915.
- Jockers, Matthew L., & Rosamond Thalken.** 2020. *Text Analysis with R: For Students of Literature*. Quantitative Methods in the Humanities and Social Sciences. Cham: Springer International Publishing.
- Johnson, Dominic D. P.** 2005. "God's Punishment and Public Goods: A Test of the Supernatural Punishment Hypothesis in 186 World Cultures." *Human Nature* 16 (4): 410–446.

- Johnson, Kyle P., Patrick J. Burns, John Stewart, & Todd Cook.** 2019. "The Classical Language Toolkit." January 2019. <http://cltk.org>.
- Johnson, Tim, & Josiah Ober.** 2014. "POLIS." 2014. <http://polis.stanford.edu>.
- Kaiser, Brooks A.** 2007. "The Athenian Trierarchy: Mechanism Design for the Private Provision of Public Goods." *The Journal of Economic History* 67 (02).
- Karila-Cohen, Karine.** 2016. "Prosopographia Attica 2.0 : base de données et raisonnement prosopographique." *Revue historique* 680 (4): 869–904.
- . 2017. "Quand les familles s'affichent sur les pierres (à partir de l'exemple athénien)." *Pallas Hors Série*: 91–117.
- . 2018. "Le graphe, la trace et les fragments : l'apport des méthodes quantitatives et des outils numériques à l'étude des élites civiques athéniennes." *Annales HSS* 73 (4): 785–815.
- . 2019. "La valeur du lien : relations extérieures et capital social des notables athéniens de la basse époque hellénistique." In *La cité interconnectée dans le monde gréco-romain (IV^e siècle a.C.-IV^e siècle p.C.) : transferts et réseaux institutionnels, religieux et culturels aux époques hellénistique et impériale*, edited by Madalina Dana & Ivana Savalli-Lestrade, 287–306. Scripta antiqua 118. Pessac: Ausonius Éditions.
- Karila-Cohen, Karine, Claire Lemerrier, Isabelle Rosé, & Claire Zalc.** 2018. "Nouvelles cuisines de l'histoire quantitative." *Annales HSS* 73 (4): 773–783.
- Kaše, Vojtěch, Tomáš Hampejs, & Zdeněk Pospíšil.** 2018. "Modeling Cultural Transmission of Rituals in Silico: The Advantages and Pitfalls of Agent-Based Vs. System Dynamics Models." *Journal of Cognition and Culture* 18 (5): 483–507.
- Kenna, Ralph, Máirín MacCarron, & Pádraig MacCarron, eds.** 2017. *Maths Meets Myths: Quantitative Approaches to Ancient Narratives. Understanding Complex Systems.* Cham: Springer International Publishing.
- Kesten, Harry.** 1973. "Random Difference Equations and Renewal Theory for Products of Random Matrices." *Acta Mathematica* 131 (0): 207–248.
- Kohli, Rajeev, & Raaj Sah.** 2006. "Some Empirical Regularities in Market Shares." *Management Science* 52 (11): 1792–1798.
- Krugman, Paul R., & Robin Wells.** 2015. *Economics.* 4th ed. New York: Worth Publishers.
- Kydros, Dimitrios, Panagiotis Notopoulos, & Georgios Exarchos.** 2015. "Homer's Iliad – A Social Network Analytic Approach." *International Journal of Humanities and Arts Computing* 9 (1): 115–132.
- Kyriazis, Nicholas C., & Emmanouil M. L. Economou.** 2015. "Macroculture, Sports and Democracy in Classical Greece." *European Journal of Law and Economics* 40 (3): 431–455.

- . 2019. "The Ancient Athenian Economy: A Review." In *The Influence of Hellenic Philosophy on the Contemporary World*, edited by John G. Dellis & Stephanos A. Païpetēs, 56–67. Newcastle upon Tyne: Cambridge Scholars Publishing.
- Laffont, Jean-Jacques.** 2001. *Incentives and Political Economy*. Clarendon Lectures in Economics. Oxford: Oxford University Press.
- Langlois, Charles-Victor, & Charles Seignobos.** (1898) 2014. *Introduction aux études historiques*. Bibliothèque idéale des sciences sociales. Lyon: ENS Éditions.
- Le Riche, Antoine, Antoine Parent, & Lei Zhang.** 2022. "Introduction." In *Institutional Change and China Capitalism: Frontier of Cliometrics and Its Application to China*, edited by Antoine Le Riche, Antoine Parent, & Lei Zhang, xxvii–xlv. Transformations in Banking, Finance and Regulation 2. Hackensack: World Scientific.
- Lebreton, Sylvain.** 2019. "Des dieux et des personnes : une approche par les réseaux (mondes grecs et ouest-sémitiques, ca. 1000 av. n. è. – 400 d. n. è.)." Working Paper. Cinquième rencontre du groupe Res-Hist. Rennes: Université de Rennes. <https://reshist.hypotheses.org/1438>.
- Lebreton, Sylvain, Jean-Baptiste Barreau, Karine Karila-Cohen, & Pierre Brulé.** 2014. "Banque de Données des Épicleses Grecques (BDEG)." 2014. <https://epiclesesgrecques.univ-rennes1.fr/?lang=en>.
- Lee, Mi Jin, Woo Seong Jo, Il Gu Yi, Seung Ki Baek, & Beom Jun Kim.** 2016. "Evolution of Popularity in Given Names." *Physica A: Statistical Mechanics and Its Applications* 443: 415–422.
- Lemercier, Claire.** 2012. "Formale Methoden der Netzwerkanalyse in den Geschichtswissenschaften: Warum und Wie?" *Österreichische Zeitschrift für Geschichtswissenschaften* 23 (1): 16–41.
- Levitt, Steven D., & Stephen J. Dubner.** 2009. *Freakonomics: A Rogue Economist Explores the Hidden Side of Everything*. New York: Harper Perennial.
- Levy, Gilat, & Ronny Razin.** 2007. "A Theory of Religion: Linking Individual Beliefs." Working Paper 19. 2007 Meeting Papers. 2007 Meeting Papers. London: Society for Economic Dynamics. <https://ideas.repec.org/p/red/sed007/19.html>.
- . 2012. "Religious Beliefs, Religious Participation, and Cooperation." *American Economic Journal: Microeconomics* 4 (3): 121–151.
- . 2014. "Rituals or Good Works: Social Signaling in Religious Organizations." *Journal of the European Economic Association* 12 (5): 1317–1360.
- Ley, Herbert de.** 1988. "The Name of the Game: Applying Game Theory in Literature." *SubStance* 17 (1): 33–46.

-
- Li, Wentian.** 2012. "Analyses of Baby Name Popularity Distribution in U.S. For the Last 131 Years." *Complexity* 18 (1): 44–50.
- Limpert, Eckhard, Werner A. Stahel, & Markus Abbt.** 2001. "Log-Normal Distributions Across the Sciences: Keys and Clues." *BioScience* 51 (5): 341.
- Lipman, Barton L., & Duane J. Seppi.** 1995. "Robust Inference in Communication Games with Partial Provability." *Journal of Economic Theory* 66 (2): 370–405.
- Lyttkens, Carl Hampus.** 2006. "Reflections on the Origins of the Polis: An Economic Perspective on Institutional Change in Ancient Greece." *Constitutional Political Economy* 17 (1): 31–48.
- . 2013. *Economic Analysis of Institutional Change in Ancient Greece: Politics, Taxation and Rational Behaviour*. Routledge Explorations in Economic History 58. London: Routledge.
- Lyttkens, Carl Hampus, & Henrik Gerding.** 2022. "Coping with Chronic Warfare: The Athenian Experience." In *Democracy and Salamis: 2500 Years After the Battle That Saved Greece and the Western World*, edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis, & Athanasios Platias, 181–200. Cham: Springer International Publishing.
- Lyttkens, Carl Hampus, George Tridimas, & Anna Lindgren.** 2018. "Making Direct Democracy Work: A Rational-Actor Perspective on the *Graphe Paranomon* in Ancient Athens." *Constitutional Political Economy* 29 (4): 389–412.
- Malkin, Irad.** 2011. *A Small Greek World: Networks in the Ancient Mediterranean*. Greeks Overseas. New York: Oxford University Press.
- Mandelbrot, Benoît.** 1953. "An Informational Theory of the Statistical Structure of Languages." In *Communication Theory*, edited by Willis Jackson, 486–502. Woburn: Butterworth.
- Margo, Robert A.** 2018. "The Integration of Economic History into Economics." *Cliometrica* 12 (3): 377–406.
- Marrou, Henri-Irénée.** (1954) 2016. *De la connaissance historique*. Histoire H21. Paris: Seuil.
- Maruvka, Yosi E., Nadav M. Shnerb, & David A. Kessler.** 2010. "Universal Features of Surname Distribution in a Subsample of a Growing Population." *Journal of Theoretical Biology* 262 (2): 245–256.
- Mateos, Pablo, Paul A. Longley, & David O'Sullivan.** 2011. "Ethnicity and Population Structure in Personal Naming Networks." *PLOS ONE* 6 (9): e22943.
- Matthews, Steven A.** 1989. "Veto Threats: Rhetoric in a Bargaining Game." *The Quarterly Journal of Economics* 104 (2): 347.
- Maynard Smith, John.** 1972. *On Evolution*. Edinburgh: Edinburgh University Press.
- . 1982. *Evolution and the Theory of Games*. 1st ed. Cambridge University Press.

- McBride, Michael.** 2016. "A Rational Choice Theory of Religious Authority." *Rationality and Society* 28 (4): 410–438.
- McCannon, Bryan C.** 2010a. "Homicide Trials in Classical Athens." *International Review of Law and Economics* 30 (1): 46–51.
- . 2010b. "The Median Juror and the Trial of Socrates." *European Journal of Political Economy* 26 (4): 533–540.
- . 2011. "Jury Size in Classical Athens: An Application of the Condorcet Jury Theorem." *Kyklos* 64 (1): 106–121.
- . 2012. "The Origin of Democracy in Athens." *Review of Law & Economics* 8 (2).
- McGillivray, Barbara, Jon Wilson, & Tobias Blanke.** 2019. "Towards a Quantitative Research Framework for Historical Disciplines." In *Proceedings of the Workshop on Computational Methods in the Humanities 2018*, 2314:53–58. Aachen: CEUR Workshop Proceedings.
- McLean, Bradley H.** 2002. *An Introduction to Greek Epigraphy of the Hellenistic and Roman Periods from Alexander the Great down to the Reign of Constantine (323 B.C.-A.D. 337)*. Ann Arbor: University of Michigan Press.
- Mehler, Alexander, Andy Lücking, Sven Banisch, Philippe Blanchard, & Barbara Job, eds.** 2016. *Towards a Theoretical Framework for Analyzing Complex Linguistic Networks*. Understanding Complex Systems. Berlin: Springer.
- Mengel, Friederike.** 2018. "Risk and Temptation: A Meta-study on Prisoner's Dilemma Games." *The Economic Journal* 128 (616): 3182–3209.
- Miller, Peter J.** 2018. "In the Shadow of Praise: Epinician Losers and Epinician Poetics." *Bulletin of the Institute of Classical Studies* 61 (1): 21–41.
- Mitzenmacher, Michael.** 2004. "A Brief History of Generative Models for Power Law and Lognormal Distributions." *Internet Mathematics* 1 (2): 226–251.
- Mongin, Philippe.** 2016. "What Are Analytic Narratives?" Working Paper ID 2796567. Rochester: Social Science Research Network. <https://papers.ssrn.com/abstract=2796567>.
- . 2018. "A Game-Theoretic Analysis of the Waterloo Campaign and Some Comments on the Analytic Narrative Project." *Cliometrica* 12 (3): 451–480.
- Morales, José A., Sergio Sánchez, Jorge Flores, Carlos Pineda, Carlos Gershenson, Germinal Cocho, Jerónimo Zizumbo, Rosalío F. Rodríguez, & Gerardo Iñiguez.** 2016. "Generic Temporal Features of Performance Rankings in Sports and Games." *EPJ Data Science* 5 (1): 33.
- Motschenbacher, Heiko.** 2020. "Corpus Linguistic Onomastics: A Plea for a Corpus-Based Investigation of Names." *Names* 68 (2): 88–103.

- Murray, Sarah.** 2017. *The Collapse of the Mycenaean Economy: Imports, Trade, and Institutions 1300–700 BCE*. Cambridge: Cambridge University Press.
- Myerson, Roger B.** 2008. "The Autocrat's Credibility Problem and Foundations of the Constitutional State." *American Political Science Review* 102 (1): 125–139.
- Narang, Neil, & Brad L. LeVeck.** 2020. "Stylized Facts and Comparative Statics in (Social) Science Inquiry." *International Studies Quarterly* 64 (2): 482–485.
- Nurmi, Hannu.** 1986. "Mathematical Models of Elections and Their Relevance for Institutional Design." *Electoral Studies* 5 (2): 167–181.
- Ober, Josiah.** 1993. "Thucydides' Criticism of Democratic Knowledge." In *Nomodeiktes: Greek Studies in Honor of Martin Ostwald*, 81–98. Ann Harbor: University of Michigan Press.
- . 2008. *Democracy and Knowledge: Innovation and Learning in Classical Athens*. Princeton: Princeton University Press.
- . 2015. *The Rise and Fall of Classical Greece*. Princeton: Princeton University Press.
- Offenstadt, Nicolas.** 2011. "Archives, documents, sources." In *Historiographies : concepts et débats*, edited by Christian Delacroix, François Dosse, Patrick Garcia, & Nicolas Offenstadt, 1:220–231. Folio Histoire 179. Paris: Gallimard.
- Osborne, Robin.** 1996. "Pots, Trade and the Archaic Greek Economy." *Antiquity* 70 (267): 31–44.
- Pantelia, Maria.** 2020. "Thesaurus Linguae Graecae." 2020. <http://stephanus.tlg.uci.edu>.
- Parent, Antoine.** 2004. "L'analyse économique est-elle intemporelle ?" *Temporalités* 1: 98–106.
- Parker, Robert.** 2003. "The Problem of the Greek Cult Epithet." *Opuscula Atheniensi* 28 (2003): 173–183.
- Parker, Robert, Jean-Baptiste Yon, & Mark Depauw.** 1996. "Lexikon of Greek Personal Names." 1996. <http://www.lgpn.ox.ac.uk>.
- Parkin, Tim.** 2013. "The Demography of Infancy and Early Childhood in the Ancient World." In *The Oxford Handbook of Childhood and Education in the Classical World*, edited by Judith Evans Grubbs & Tim Parkin, 40–61. Oxford: Oxford University Press.
- Pavesi, Angelo, Paola Pizzetti, Enzo Siri, Enzo Lucchetti, & Franco Conterio.** 2003. "Brief Communication: Coexistence of Two Distinct Patterns in the Surname Structure of Sicily." *American Journal of Physical Anthropology* 120 (2): 195–199.
- Pflug, Gunther.** 1971. "The Development of Historical Method in the Eighteenth Century [1954]." *History and Theory* 11: 1–23.
- "PHI Greek Inscriptions". n.d. <https://inscriptions.packhum.org/>.
- Piantadosi, Steven T.** 2014. "Zipf's Word Frequency Law in Natural Language: A Critical Review and Future Directions." *Psychonomic Bulletin & Review* 21 (5): 1112–1130.

- Pitsoulis, Athanassios.** 2011. "The Egalitarian Battlefield: Reflections on the Origins of Majority Rule in Archaic Greece." *European Journal of Political Economy* 27 (1): 87–103.
- Popper, Karl R.** (1934) 1992. *The Logic of Scientific Discovery*. 2nd ed. London: Routledge.
- Posner, Richard A.** (1973) 2014. *Economic Analysis of Law*. 9th ed. Aspen Casebook Series. New York: Wolters Kluwer Law & Business.
- Power, Eleanor A.** 2017. "Discerning Devotion: Testing the Signaling Theory of Religion." *Evolution and Human Behavior* 38 (1): 82–91.
- Puech, Vincent.** 2012. "La méthode prosopographique et l'histoire des élites dans l'Antiquité tardive." *Revue historique* 661 (1): 155–168.
- "R: The R Project for Statistical Computing"**. 2021. The R Project for Statistical Computing. January 2021. <https://www.r-project.org/>.
- Rawski, Thomas G., Susan B. Carter, Jon S. Cohen, Stephen Cullenberg, Peter H. Lindert, Donald N. McCloskey, Hugh Rockoff, & Richard Sutch.** 1996. *Economics and the Historian*. Berkeley: University of California Press.
- Reed, William J., & Barry D. Hughes.** 2003a. "Power-Law Distributions from Exponential Processes: An Explanation for the Occurrence of Long-Tailed Distributions in Biology and Elsewhere." *Scientiae Mathematicae Japonicae* 8: 329–339.
- . 2003b. "On the Distribution of Family Names." *Physica A: Statistical Mechanics and Its Applications* 319: 579–590.
- Rochat, Yannick.** 2014. "Character Networks and Centrality." PhD thesis, Lausanne: Université de Lausanne. <http://infoscience.epfl.ch/record/203889>.
- Rogerson, Peter A.** 2016. "The Effects of Months, Holidays, and the Birthdays of Presidents on Choice of Baby Names." *Names* 64 (4): 234–241.
- Roman, Sabin, & Erika Palmer.** 2019. "The Growth and Decline of the Western Roman Empire: Quantifying the Dynamics of Army Size, Territory, and Coinage." *Cliodynamics* 10 (2): 1–23.
- Ruffle, Bradley J., & Richard Sosis.** 2007. "Does It Pay To Pray? Costly Ritual and Cooperation." *The B.E. Journal of Economic Analysis & Policy* 7 (1).
- Rydberg-Cox, Jeff.** 2011. "Social Networks and the Language of Greek Tragedy." *Journal of the Chicago Colloquium on Digital Humanities and Computer Science* 1 (3): 1–11.
- Rygiel, Philippe.** 2011. "L'enquête historique à l'ère numérique." *Revue d'histoire moderne et contemporaine* 58-4 bis (5): 30–40.
- Saichev, Alex, Yannick Malevergne, & Didier Sornette.** 2010. *Theory of Zipf's Law and Beyond*. Lecture Notes in Economics and Mathematical Systems 632. Berlin: Springer.

- Säid, Suzanne, & Monique Trédé.** (1990) 1999. *A Short History of Greek Literature*. Translated by Trista Selous. London: Routledge.
- Schwuchow, Soeren C., & George Tridimas.** 2022. "The Political Economy of Solon's Law Against Neutrality in Civil Wars." *Public Choice* 192 (3-4): 249–272.
- Smith, Thomas Jay.** 2015. *Variability in Human Performance*. Human Factors and Ergonomics. Boca Raton: CRC Press.
- Stahl, Michael, & Uwe Walter.** 2009. "Athens." In *A Companion to Archaic Greece*, edited by Kurt A. Raaflaub & Hans van Wees, 138–161. Blackwell Companions to the Ancient World. Malden: Blackwell Publishing.
- Stoddart, Simon, & James Whitley.** 1988. "The Social Context of Literacy in Archaic Greece and Etruria." *Antiquity* 62 (237): 761–772.
- Tangian, Andranik.** 2008. "A Mathematical Model of Athenian Democracy." *Social Choice and Welfare* 31 (4): 537–572.
- Tauber, James.** 2020. "Working with the Diorisis Ancient Greek Corpus." J. K. Tauber. January 20, 2020. <https://jktauber.com/2020/01/20/working-with-the-diorisis-ancient-greek-corpus/>.
- . 2021. "Jtauber/Diorisis." Exploring the Diorisis Ancient Greek Corpus. 2021. <https://github.com/jtauber/diorisis>.
- Toynbee, Arnold J.** (1954) 1987. *A Study of History*. Vol. 2. 2 vols. New York: Oxford University Press.
- . (1934) 1987. *A Study of History*. Vol. 1. 2 vols. Oxford: Oxford University Press.
- Tridimas, George.** 2011. "A Political Economy Perspective of Direct Democracy in Ancient Athens." *Constitutional Political Economy* 22 (1): 58–82.
- . 2012. "Constitutional Choice in Ancient Athens: The Rationality of Selection to Office by Lot." *Constitutional Political Economy* 23 (1): 1–21.
- . 2013. "Homo Oeconomicus in Ancient Athens: Silver Bonanza and the Choice to Build a Navy." *Homo Oeconomicus* 30 (4): 435–458.
- . 2015. "War, Disenfranchisement and the Fall of the Ancient Athenian Democracy." *European Journal of Political Economy* 38: 102–117.
- . 2016. "Conflict, Democracy and Voter Choice: A Public Choice Analysis of the Athenian Ostracism." *Public Choice* 169 (1-2): 137–159.
- . 2017. "Constitutional Choice in Ancient Athens: The Evolution of the Frequency of Decision Making." *Constitutional Political Economy* 28 (3): 209–230.
- . 2020. "Modelling the Quest for Status in Ancient Greece: Paying for Liturgies." *Homo Oeconomicus* 37 (3-4): 213–236.

- . 2022. "Policy Making by Randomly Selected Citizens: The Perspective of Elected Politicians." In *Democracy in Times of Crises: Challenges, Problems and Policy Proposals*, edited by Emmanouil M. L. Economou, Nicholas C. Kyriazis, & Athanasios Platias, 117–136. Cham: Springer International Publishing.
- Turchin, Peter.** (2003) 2018. *Historical Dynamics: Why States Rise and Fall*. Princeton Studies in Complexity. Princeton: Princeton University Press.
- Turchin, Peter, Thomas E. Currie, Edward A. L. Turner, & Sergey Gavrilets.** 2013. "War, Space, and the Evolution of Old World Complex Societies." *Proceedings of the National Academy of Sciences* 110 (41): 16384–16389.
- Vandewalle, Henry.** 2018. "Modelling of Running Performances: Comparisons of Power-Law, Hyperbolic, Logarithmic, and Exponential Models in Elite Endurance Runners." *BioMed Research International* 2018: 1–23.
- Vatri, Alessandro, & Barbara McGillivray.** 2018a. "The Diorisis Ancient Greek Corpus." May 2, 2018. https://figshare.com/articles/dataset/The_Diorisis_Ancient_Greek_Corpus/6187256.
- . 2018b. "The Diorisis Ancient Greek Corpus: Linguistics and Literature." *Research Data Journal for the Humanities and Social Sciences* 3 (1): 55–65.
- Vernant, Jean-Pierre.** 1965. "Espace et organisation politique en Grèce ancienne." *Annales ESC* 20 (3): 576–595.
- Vernant, Jean-Pierre, & Pierre Vidal-Naquet.** (1972) 2001. *Mythe et tragédie en Grèce ancienne*. Vol. 1. 2 vols. Paris: La Découverte.
- Versnel, Henk S.** 2011. *Coping with the Gods: Wayward Readings in Greek Theology*. Religions in the Graeco-Roman World, v. 173. Leiden: Brill.
- Vidal-Naquet, Pierre.** (1981) 2005. *Le chasseur noir : Formes de pensée et formes de société dans le monde grec*. Poche 194. Paris: La Découverte.
- Waumans, Michaël C., Thibaut Nicodème, & Hugues Bersini.** 2015. "Topology Analysis of Social Networks Extracted from Literature." *PLOS ONE* 10 (6): e0126470.
- Weintraub, E. Roy.** 2002. *How Economics Became a Mathematical Science*. Science and Cultural Theory. Durham: Duke University Press.
- Whitley, James.** 1991. "Social Diversity in Dark Age Greece." *The Annual of the British School at Athens* 86: 341–365.
- Wickham, Hadley, & Garrett Grolemund.** 2017. *R for Data Science*. Boston: O'Reilly.
- Woods, Robert.** 2007. "Ancient and Early Modern Mortality: Experience and Understanding." *The Economic History Review* 60 (2): 373.

-
- Young, Robert K., Ana H. Kennedy, Ann Newhouse, Pam Browne, & Del Thiessen.** 1993. "The Effects of Names on Perception of Intelligence, Popularity, and Competence." *Journal of Applied Social Psychology* 23 (21): 1770–1788.
- Zanette, Damián H., & Susanna C. Manrubia.** 2001. "Vertical Transmission of Culture and the Distribution of Family Names." *Physica A: Statistical Mechanics and Its Applications*, Proceedings of the IUPAP International Conference on New Trends in the Fractal Aspects of Complex Systems, 295 (1): 1–8.
- Zipf, George K.** 1949. *Human Behavior and the Principle of Least Effort: An Introduction to Human Ecology*. Cambridge: Addison-Wesley.
- Zurbach, Julien.** 2013. "The Formation of Greek City-States: Status, Class, and Land Tenure Systems." *Annales HSS* 68 (04): 615–657.

New cliometrics for Ancient History

Abstract:

This thesis proposes to link formal economic analysis with the full detail of ancient historical sources, thereby extending traditional cliometrics, thanks to the recent developments that have associated cliometrics and complexity. Most economic approaches into ancient history have remained in the field of analytic narratives, rather than cliometrics, because of a lack of data. We propose a new epistemological framework for cliometrics, relying on complexity: we argue they do not have to be about the economy, but can operate on primary historical sources, and could address a much broader set of periods and societies, leaning on microeconomic models. We focus on the case of ancient Greek history, and show we can frame into economic terms the massive and complex cultural material now available in electronic form. We then drill further down and concentrate on two particular aspects of ancient Greek society: the choice of names, and which gods to pray to. We first develop an economic model for the existence of names, as a way of exchanging identification information. We study the optimal choice of names in these conditions and show that the impact of strategic naming on the distribution of names works as an alteration of existing mean-field approaches to name dynamics. Confronting theoretical results with empirical data from the archaic and classical periods, we show conformist strategic naming may account for the particular shape of the name distribution in Ancient Greece. Finally, we look into the optimal choice of which god to worship as a function of the presumed strategy of the gods for returning favors to the worshippers, and relate it to a form of divine efficiency measure. Data from a large volume of epigraphic and literary sources on actual acts of worships from the ancient Greeks shows that the distributions of these acts effectively follow power laws with a high degree of regularity, as predicted by the model.

Keywords:

Ancient Greece, game theory, historical economics, cliometrics, complexity