The Decision to Strike:

Quantifying Roman Mint Output and the Case for Roman Monetary Policy

Jacob Weber

April 2015

Advisor: Dr. Graham Oliver, Professor of Classics, Brown University

Second Reader: Dr. Peter van Dommelen, Joukowsky Family Professor in Archaeology and Professor of Anthropology, Brown University

Submitted in Partial Fulfillment of the Requirements of the Bachelors of Arts with Honors in Classical Archaeology, Brown University

Introduction: Motivation and Background

After the Second Punic war, the Romans completely overhauled their coinage, introducing the *denarius* system in 211 BC. Millions of silver *denarii* were minted nearly every year from 211 BC down to the end of the Republic in 31 BC (von Reden 2012, 273). But why did the Romans mint so many *denarii*? The traditional explanation is that the Roman government minted so many in order to meet the enormous expenditures required by military expansion (Harl 1996, 39-40; Pense 1992, 213; Crawford 1974, 694), but this is often accompanied by an uneasy awareness that such an explanation overly simplifies a complicated question (de Callataÿ 2011, 17-18).[1] This theory is particularly misleading for the Roman Republic in that it assumes that the Roman state did not have any other method of making payments rather than minting new coins, and it glosses over the varied reasons the Romans might have had for choosing to strike their own.

Although state expenditure, the majority of which was certainly military spending in the Roman Republic, may have been the primary means by which newly minted coins first entered circulation, this does not mean that state expenditure was always the *reason* for minting coins. Minting new coins was not the only method of making payments in the ancient world. Payment in bullion or "in kind" using other goods was also a possibility, as was reusing old coin acquired through trade, taxation, or booty (Harris 2006, 19; Howgego 1995, 35-38). Indeed, it was not always necessary for ancient states to mint new coins even to meet large expenditures. The largest indemnity ever imposed by the Romans, on Antiochus the IV and the Seleukids, famously

[1] I will use the following abbreviations: *RRC* for Crawford 1974, and OXREP for the Oxford Roman Economy Project.
[2] The title of Buttrey's famous 1993 address, "Calculating Ancient Coin Production: Facts and Fantasies," is a testament to exactly how heated such debates can become.
[3] It should be noted that the Roman state never borrowed money directly except during the darkest days of the Punic wars, although cities did so frequently (Howgego 1992, 13).

had no impact on the output of the Seleukid mint (Howgego 1995, 37) and was probably paid with other local coinages of Asia Minor that the Seleukids apparently countermarked, rather than reminted (Aperghis 2004, 232).

Given that minting coins was not the only way in which state expenditures might be met, other, more nuanced theories have been proposed. These acknowledge the importance of economic considerations such as the facilitation of exchange and taxation (Howgego 1990, 22-24), or even inflation and rudimentary monetary policy (Hitchner 2012, 283-285; Lo Cascio 1981) in addition to considering the political implications of minting coins (von Reden 2012, 272). At any given point in time, some or all of these considerations may have been factors in determining how many coins the Roman mint produced.

Given this fact, I will examine here the theory of rudimentary monetary policy as detailed by Elio Lo Cascio (1981) while making use of statistics and large sets of archaeological data. The application of statistics in archaeology is not always popular, even in a field as (relatively) quantifiable as numismatics, and has many difficulties that we will deal with below. As François de Callataÿ wrote in 2011, quoting Carmen Alfaro and Andrew Burnett:

> "…our impression is that (scientific studies) are fewer in number than before, partly because of expense, but also because the results they provide, though extremely important, have not been seen to revolutionize the subject, as was previously hoped. Much the same could be said of statistics…there has been a retreat from the optimism of previous years about the potential, for example, of using coins to make quantitative studies of past economies (de Callataÿ 2011, 14)."

Here, we will join those who have adopted the opinion that this pessimism regarding statistics and quantification is unjustified and engage in a statistical treatment of archaeological data as pioneered by other earlier scholars (e.g. Greene 1986). In particular, this pessimism seems all the more unfounded in light of new databases such as those provided by Kris Lockyear in the form of *Coin Hoards of the Roman Republic Online* and, for the economy in general, by the Oxford Roman Economy Project (OXREP). The structure of the argument shall be as follows:

In part one, I will verify Michael Crawford's output calculations for the Roman mint, which he arrived at through die studies, as good measures of the *relative* output of the Roman mint. I will do this using Lockyear's database, *Coin Hoards of the Roman Republic Online*, and multivariate linear regression, computed using the popular statistical computing software package Stata/SE (version 13.1), hereafter referred to as "Stata."

In part two, having confirmed that Crawford's output estimates are valid, I will examine the argument for Roman monetary policy (Lo Cascio 1981) under which the Roman mint is posited to have increased the money supply in response to the increasing demand for money. I will use Crawford's military expenditures and shipwrecks *per annum* from the OXREP database (Strauss 2013) to examine this, and will then show that both shipwrecks *per annum* and military expenditures have a statistically significant, positive impact on Roman mint output. In short, there is more than enough room for the existence of Roman monetary policy, and other variables, as drivers of Roman mint output. Thus, it is not possible to conclude from the data that military expenditures were the main driver of Roman mint output.

Part One: Checking Crawford's Output Estimates

1.1 Criticisms of Crawford's Approach

Although the debate regarding the influence of state expenditures, and in particular military expenditures, on the output of the Roman mint has been going on for many years, it became particularly heated[2] when Crawford, in 1974, attempted to calculate the yearly output of the Roman mint in his volume *Roman Republican Coinage.* He plotted his calculated mint output alongside Roman estimates for state expenditures, and observed how well the two graphs matched, stating that there was a "remarkable correlation between expenditure and volume of coinage" (Crawford 1974, 694).

In order to show this "remarkable correlation" (which will be examined in part two), Crawford first provided a comprehensive dating scheme for the entirety of the Roman Republican coinage, using hoard evidence, the classical sources, and his own judgment. He then estimated the number of coins produced by the Roman mint each year using dies. The process of such estimation is as follows: the Romans minted *denarii* by hammering a die into silver blanks. These dies usually broke after an average of 30,000 coins were made (Crawford 1974, 694; c.f. Howgego 1992, 3). By studying the extant samples of Roman coins, counting the number of different dies and multiplying by the average number of coins struck per die, we arrive at an estimate for the original size of each issue. Then, it is a simple matter to add up the amount of coins minted each year and compare this to Roman military expenditures, which Crawford promptly did for the period of 157-91 BC.

[2] The title of Buttrey's famous 1993 address, "Calculating Ancient Coin Production: Facts and Fantasies," is a testament to exactly how heated such debates can become.

As mentioned, Crawford's approach has received sharp criticism from the archaeological community (see reviews of Frier 1976 and Buttrey 1977 for a discussion of Crawford specifically; see Buttrey 1993 and Howgego 1992 for arguments against the method in general; see Lockyear 1999 and de Callataÿ 2005 for more optimistic views). In Crawford's case, the following arguments are usually made: first, that his number for coins per die is inaccurate, and second that his die estimates are also inaccurate. The latter criticism is due to the fact that Crawford did not, in fact, actually count the number of dies for each issue, but estimated many of them by multiplying an empirically determined constant times the number of specimens in his sample of 24 hoards (Crawford 1974, 672-673).

The first argument that is usually made is that the number of 30,000 coins per die, which Crawford used, isn't accurate. Crawford arrives at this figure from the fact that C. Annius took two legions with him to Spain in 81 and 82 BC, and paid them with the coins he minted from 100 dies, which he also took with him. Crawford took the estimated cost of maintaining two legions for a year (3,000,000) and divided that by 100 to reach the figure of 30,000 coins per die (Crawford 1974, 694).

There are many problems with this assumption. The first is that we do not know exactly how much the legions of C. Annius cost. Crawford estimated the cost by taking the number of legions times the number of soldiers in a legion times the amount each soldier received as a *stipendium*. However, we do not, in fact, even know how many men C. Annius took with him to Spain, because the size of a legion varied and was not standard under the republic: Livy and Sallust report that the senate dictated, year to year, what the size of a legion would be in response to Rome's need for manpower, fluctuating around a standard size (Roth 1994, 347). Even if we knew the exact number of men, and could estimate the cost of providing for the legions, we would not know exactly how much coin C. Annius would have needed to pay them. Although a soldiers pay was 3 *asses* per day, in reality very little of this money may have been paid out, as deductions would be made for food, clothing, and weapons; thus soldiers were effectively paid at least partially in kind (de Ligt 2007, 124; Pense 1992, 212). Moreover, assuming that new coin was the only way soldiers were paid is also incorrect, as soldiers could be paid from old coin or bullion that came from booty (de Ligt 2007, 124) or payment deferred. Soldiers were not always paid on time in ancient world, and the use of credit in the Roman world in particular is often underestimated: the Roman government might defer payments, as might individuals or entire cities, or use land to meet its obligations (Harris 2006, 9; Harl 1996, 43).[3] As a result, the classical sources disagree when it comes to the cost of a legion: another valid interpretation of the classical sources, for example, yields a cost of 1,000,000 *denarii* per legion and a result of 15,000 coins per die (Mattingly 1977, 207). In short, the number of 30,000 coins per die is liable to be extremely inaccurate, once all of these errors are taken into account (Buttrey 1993, 339-341).

Moreover, even if it were accurate, it is incredibly dangerous to assume from one sample that this number would be the accurate number of coins per die for all places and times in the Roman Empire. This mint in particular was exceptional, as the coins were produced in Spain, rather than at the central mint in Rome. Reapplying this average to *all* issues of Roman coin seems particularly dangerous then, in this respect, especially since comparative evidence from medieval times suggests that productivity per die can vary enormously: anywhere from between 2,000 pieces per die to 78,000 pieces per die year over year for silver pennies in 13th-14th century Britain (Buttrey 1993, 343-344). As a result, defining an exact average value for coins per die is

[3] It should be noted that the Roman state never borrowed money directly except during the darkest days of the Punic wars, although cities did so frequently (Howgego 1992, 13).

very difficult, and many potential estimates have been given whose variance reflects this fact (de Callataÿ 2011, 9; de Callataÿ 2005, 77).

Even if the number of coins per die were accurate, however, Crawford's die estimates themselves are also suspect due to the fact that they are just that: estimates. Rather than count out all the dies for the 12,554 coins that Crawford had in his sample of 24 hoards, he counted the dies for a few specimens and then obtained a die:specimen ratio, which he applied to the rest of the coins in his sample (de Callataÿ 1995, 290). In other words, Crawford counted the number of dies for 20 issues of coinage, and then made an educated guess at the rest (Crawford 1974, 672). Although the theoretical validity of extrapolation in general is now more accepted (de Callataÿ 2005, 75; Duncan-Jones 1994, 145), Crawford's particular method and multipliers have been attacked for being relatively statistically naïve compared to newer models, and in particular for not properly accounting for the wastage of older issues (de Callataÿ 1995, 307).

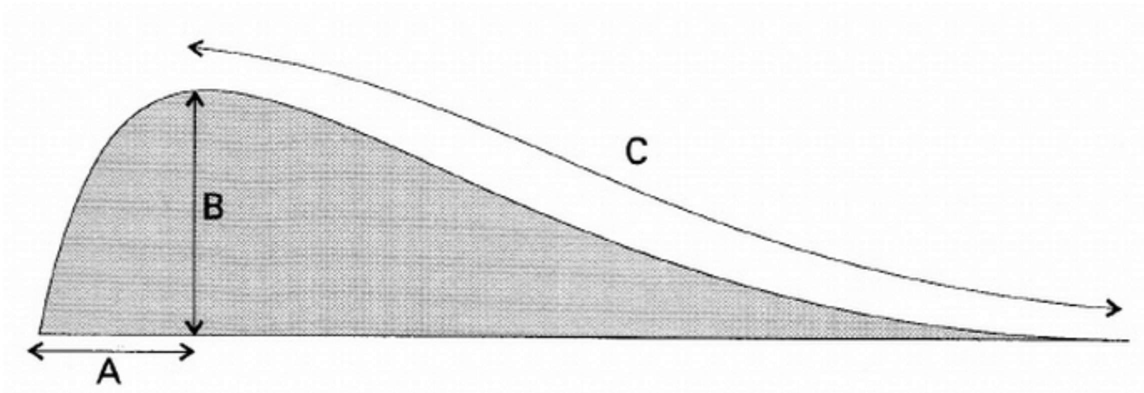
Of these two criticisms leveled against Crawford's methodology, the first (that his average number of 30,000 is of spurious accuracy) is by far the strongest: it seems very unlikely that 30,000 coins per die is accurate, given his methodology. But the second criticism is also valid: as Buttrey notes, Crawford's methodology for estimating dies can be proven incorrect in several cases from simply reading the legends on the coins themselves (Buttrey 1977, 152). Further, subsequent die studies have proven some of his estimates to be quite wrong (de Callataÿ 1995, 308).

Both of these criticisms are very concerning for Crawford's study, which relied on obtaining accurate numbers for the *absolute* production of the Roman mint. However, if we accept that his data is good in *relative* terms (i.e. all of his estimates are equally biased up or down, due to an incorrect number for the average coins/die, or an incorrect method of estimating the number of dies) then it might still be possible to show mint output is correlated with military expenditure, if not precisely equal to military expenditure. The idea is that there should be a strong correlation between the number of dies and the number of coins minted, as it is undeniable that a larger coinage will usually require more dies. Moreover, there is nothing stopping us from looking at correlations with other variables, as we will do in part two with shipwrecks, as long as we are cautious. But the question remains: how accurate *are* Crawford's estimates, even in relative terms?

1.2 Prior Work: Lockyear 1999

An attempt was made by Lockyear to examine the validity of Crawford’s die estimates as measures of the relative size of each issue, rather than an absolute measure (Lockyear 1999). In Lockyear’s experiment, he modeled the age profile of a “standard” hoard in each year using a battleship curve based on Crawford’s output estimates. The battleship curve is a standard model in archaeology used to model the frequency of a given object over time, assuming that the object is introduced or produced over a given initial period and then slowly disappears over time, as shown in the figure below from Lockyear’s paper:

Figure 1: A Battleship Curve



Source: Lockyear 1999, p. 225

The height, "B", was given by Crawford's output estimates as the total number of coins produced, while an attrition rate of 2% determined "C", and an "introduction delay" of one year or so determined "A" (Lockyear 1999, 225-226). The curves were used to predict the composition, as a percentage of each extant issue, of Italian hoards found in a given year. Given that Lockyear's models accurately matched what was actually found in the hoards, it stands to reason that Crawford's die estimates must be fairly accurate as a whole (Lockyear 1999, 240).

This elegant experiment has convinced even the most steadfast opponents of Crawford's methodology that something like a mean value for coin output per die must have existed (de Callataÿ 2005, 75), and thus that Crawford's output estimates have some validity as a whole, even if they can be individually shown to be wrong. However, if we were determined to criticize Lockyear's excellent paper, we might focus on a weakness which Lockyear himself is quite open about: the nature of the Kolmogorov-Smirnov test. The Kolmogorov-Smirnov test was used in Lockyear's experiment to determine whether or not each hoard profile could be shown to be different in a statistically significant way from his model. However, if a hoard could not be shown to be different by the Kolmogorov-Smirnov test, this does not necessarily mean that it comes from the population that Lockyear modeled, only that "we do not have cause to answer" (Lockyear 1999, 228-9). Hence, Lockyear's paper rests, in an abstract sense, on an *absence* of evidence.

This in and of itself is not a problem: basing an argument on an absence of statistical evidence may very well be quite appropriate when testing, for example, the efficacy of a new drug. What's more, if one is able to repeatedly show no difference between one's model and observed data, as Lockyear attempts (he measured his model against 14 different hoards in his paper, with the results presented in figure 2 below) then perhaps it can be safe to assume that the model does a good job of representing the true population.

However, Lockyear's model obtained rather lukewarm success at predicting the contents of actual hoards: less than *half* were "predicted" by the model (6 out of 14, as seen in figure 2). The expected error rate should have been around 5%, this being the interpretation of a p-value[4] of 5%. In other words, if Lockyear's model was accurately modeling the population, we would expect only 5% of the hoard to be rejected simply due to the chance inherent in random sampling — for example, he should have been able to predict 13 or 14 hoards out of 14. Instead, we observe a 57% rejection rate, which suggests that the model is largely failing to predict the contents of various hoards.

Lockyear's explanation for this is that the model does not account for regional variability in the coinage pool (note that he is modeling the coinage pool of *Italy*, but comparing it to specific hoards found in different regions), as well as the fact that larger samples means that it is

[4] Technically the definition is: the p-value is the probability of observing the data in question given that the null hypothesis is true (in this case, the probability of seeing the distribution of each hoard given that it is a random sample of the coins in the population which Lockyear modeled). See appendix. When this is small enough, we reject the null hypothesis and accept the alternative hypothesis. *Not* rejecting the null does not necessarily mean that the null is true; this is why one typically constructs a statistical test such that the thing we wish to prove is the *alternative* hypothesis, when possible. This is not what Lockyear has done here. To be fair, within the context of his model, I am not sure that he could have.

easier for the Kolmogorov-Smirnov test to reject the null hypothesis (the rejected hoards, or the hoards which weren't predicted by the model, have an average of 346.25 coins each, while the hoards that were not rejected — i.e. not different from what the model predicted — were 289.5 coins on average, so there may be some truth to this as the cause of the problem). That said, a model which only "predicts" (or, more appropriately, "cannot be shown to be statistically different from") around 43% of the data is not terribly comforting on its own. These results are summarized in figure 2 below:

Figure 2: Results of Lockyear's of Experiment

Lockyear's Results: Hoard Name	"Good Total"*	Rejected at .05 level (i.e. not modeled)
Ossero	460	No
Policoro	293	Yes
Tolfa	238	Yes
Villa Potenza	407	No
"Bahrfeldt" **	424	Yes
Brandosa	406	No
Broni	80	Yes
Casaleone	710	Yes
Carbonara	374 (324)***	No
Taranto	51	No
Cerignola	89	No
Imola	497	Yes
Olmeneta	383	Yes
Paterno	145	Yes
	Percent Rejected:	**57%**

* "Good Total" refers to the total number of well-identified *denarii* for years 157 onward.
** This hoard is unprovenanced, and may be subject to post-recovery bias/selection.
*** Some of the coins (50) came from a period not under consideration. They were dropped.

Source: Aggregated from Lockyear 1999

In the next section, I will conduct my own experiment using hoard data to test Crawford's output numbers, and determine their validity. This experiment will buttress Lockyear's results, and avoid the problems of regional variability and the sensitivity of the Kolmogorov-Smirnov test. Rather than show that there is no statistically significant difference from the values predicted by a simulation, I will use linear regression to show positive proof that Crawford's estimates *do* have predictive power, as described below. I will use linear regression to avoid the problems of rejection with the Kolmogorov-Smirnov test, as well as for its other agreeable statistical properties, such as its efficiency, unbiased nature, etc. Finally, since I will be looking at data for the entire Italian peninsula, I will also be able to avoid the problem of comparing "apples with oranges" (i.e. comparing specific hoards against an aggregate for the Italian peninsula, as Lockyear did) since I will only be looking at data for the whole of Italy. My results reinforce the belief that Crawford's estimates are, as a whole, more accurate than is commonly assumed, and thus can perhaps still be used to answer the question of whether or not coin output is correlated with Roman military expenditures, or other data.

In the following sections, I will first discuss the data. Then I will explain my experimental design, including problems with the data and how these dictated the way in which I structured and conducted my experiment. After this, I will discuss experimental methodology (i.e. the nuts and bolts of cleaning the data and running the regression). Finally, I will interpret the results of the experiment, and note that Crawford's calculations, regardless of the way in which they were generated, actually end up quite accurate as a whole.

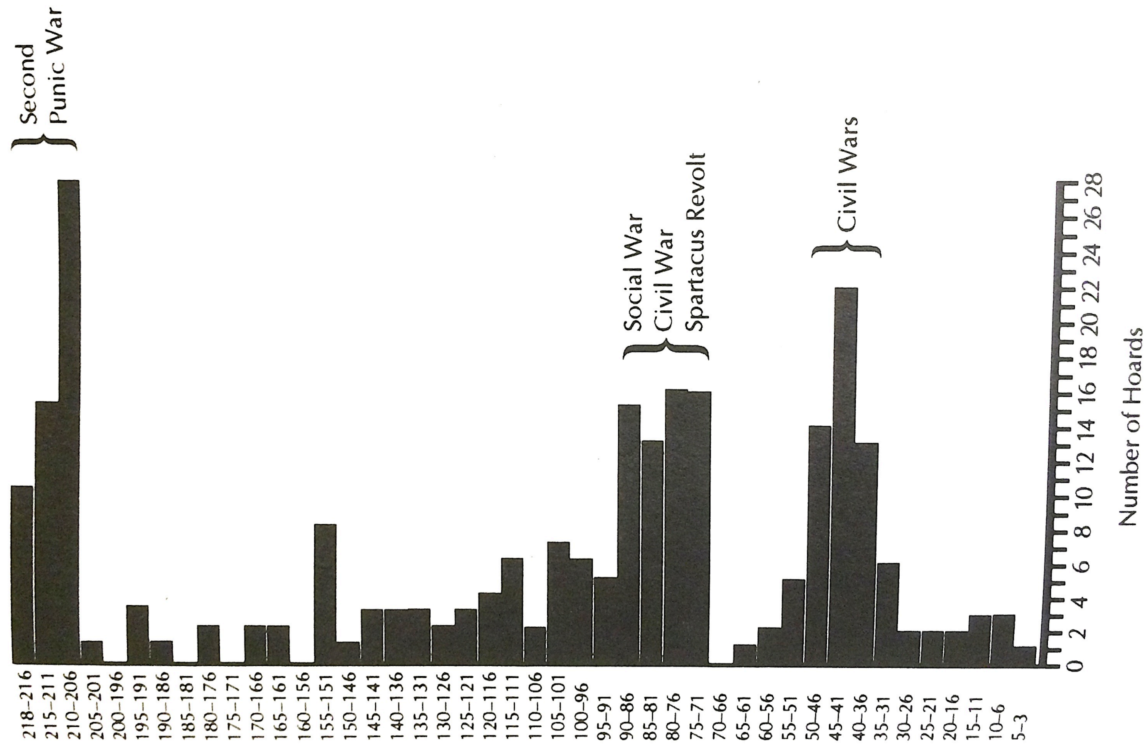
1.3 The Data: Coin Hoards

Crawford drew his 1974 die estimates from a sample of 24 coin hoards; my presentation here draws from a 100-hoard subset of Lockyear's online database of Republican *denarius* hoards, *Coin Hoards of the Roman Republic Online*, version 1 (beta). It should be noted that there are geographical biases in the data: hoards from Italy and Romania are heavily over-represented due to the original regional focus of Lockyear's database. The coverage of the database is also unfortunately only good for hoards of *denarii* as well, and not smaller bronze fractions.

1.4 Experimental Design

One might wonder why numismatists bother with die estimates at all when attempting to estimate the relative size of a coinage: why not simply look at the number of coins from each year that turn up in the archaeological record? In theory, if we find more coins from year A than from year B, then more coins were probably minted in year A. Unfortunately, survivorship bias in our coin finds makes this impossible: coin hoards are only found when they are not recovered, which is most common in times of war when the owners of said hoards would bury their coins and die, never returning to collect them (Duncan-Jones 1994, 77; Howgego 1992, 3). When looking at a distribution of hoards by year, there are indeed significant spikes during periods of warfare and unrest (Harl 1996, 12-13). Although these spikes may be somewhat exaggerated due to the use of closing dates to date hoards, the general trend almost certainly holds true (Lockyear 2012, 206-7). This means that, all other things equal, we will find more coins from each year if they were minted right before a period of war (say, in 90 BC before the Social Wars in Italy) than those minted after (70 BC).

Figure 3: Coin Hoards by Year (Harl 1996, 13)



Given this, the following experiment is suggested to test Crawford's output results for the years 150-100 BC: Rather than look at the total number of coins that have been found across hoards from the entire span of Roman history, we can look at how many coins from this period are found from hoards at an arbitrary point in time in the future. In other words, we look here at the number of coins from the period 150-100 BC that are found in the hoards of *later* periods. In particular, we will be examining the number of coins that are found from each year between 150-100 BC in hoards from the periods of 90-70 BC, which cover the Social Wars and ensuing unrest, and 50-27 BC, which includes the Civil Wars. In theory, the number of coins from each year between 150 and 100 BC that show up in these later hoards should be correlated with three things:

1) **Age**
   - Coins from, for example, 150 BC should show up less than coins from 100 BC, since coins fall out of circulation over time, all other things equal.

2) **Fineness**
   - Higher purity *denarii* are perhaps more likely to be hoarded, due to Gresham's Law.[5] So we should observe more *denarii* if they have higher silver content.

3) **Number of coins minted in that year**
   - This being calculated using Crawford's die estimates.

We don't have to consider fineness as a variable in this period, as the coins of the period 150-100 BC are all of extremely high purity (Walker 1980).[6] I assume the rate of attrition, or the rate at which coins disappear from circulation, is constant, or near constant, for this time period as demonstrated by Lockyear (1999, 236). Further, it is trivial to control for age by adding in year as an independent variable; I will use Crawford's dates given in the *RRC* for this, given that many of the suggested corrections to the *RRC* are disputed and mostly affect small issues of coinage (Lockyear 1999, 226; see Lockyear 2012, 196 for a more modern summary of these disputes and changes). If Crawford's output estimates are good, then after controlling for these other factors his output estimates should have significant explanatory value in predicting the frequency of coins by year (150-50 BC) that appear in later Italian hoards from the chosen periods of unrest. In other words, if we regress the frequency of coins found in later hoards on Crawford's output estimates, we would expect to see a close relationship between the two if the estimates had any validity.

[5] Gresham's Law: If coins of differing metal content trade at the same face value within a given monetary zone, then the purer coins get hoarded more often; "bad money chases out good." This law is notoriously difficult to apply and does not hold in all contexts (Sargent and Smith 1997, 5). Thus it is good to avoid it altogether by considering only coins of the same purity and denomination!

[6] Although his methodology has been challenged, Walker is reliable for coins of purity greater than 90%, which includes the entirety of the coinage from 150-100 BC (Howgego 2009, 272).

The time periods and hoards under consideration are careful, if somewhat arbitrary, choices on my part that attempt to tackle biases in the data, some of which I mentioned above. Although dropping data is never a good statistical practice when it can be avoided, I will show that it is necessary given the poor and biased quality of the data. Below, I will list five major problems and explain how my choices hopefully mitigate these problems. As a reminder to the reader, it is important to note that in linear regression, we care about measurement errors in our variables which do not affect the entirety of the data equally, rather than measurement error *per se*; the major assumption in simple linear regression is that the error terms are *uncorrelated* with the independent variables: i.e. that any error affects each year, or the issues in each year, equally. This means we must look for problems that would bias the observed number of coins in later hoards up or down *only for specific years*.

1.4.1 Lockyear's data is biased towards Romania and Italy

This might be a problem. For example, given that regional coin pools might have different makeups, we might observe more of one type of *denarius* in the material record simply because more of that issue was circulating in Romania, and not because it was a larger issue. To mitigate this, I examine only the hoards of Italy, since the coins of this period (150-100 BC) are without exception minted in Rome, thus controlling somewhat for the geographical bias in Lockyear's hoard data.

1.4.2 Different issues of coins might have had different functions: What if certain coins circulated outside of Italy more often than others?

This is a valid concern especially when examining coins of different denominations, or of different materials. In this case, since we are examining only Italian coins, it is a concern if particular issues were more likely to leave the Italian peninsula: They would be underrepresented relative to the other issues in our Italian hoards not because they were smaller issues, but because they had been exported or used in trade. To avoid this, I examine only the coins of 150-100 BC because the silver issues are, almost without exception, *denarii*:[7] coins of the same fineness and denomination. This frees us from dealing with the issue of the *victoriatus*, which was no longer minted around 170 BC (Harl 1996, 40) and which may have been exported or used for trade more commonly than the *denarius* and thus may be underrepresented in Italian hoards, according to an older theory (Burns 1927, 272). It is worth noting that Roman *denarii* did not circulate widely outside of Italy during this time period, as they competed with other local silver coinages in, for example, Spain and the Eastern Mediterranean (Kay 2014, 96-97), but Pliny writes specifically of the *victoriatus* that it was based on a coin "imported from Illyricum" (Pliny, *Hist. Nat.* XXXIII.viii.46).

1.4.3 Some coins are harder to identify than others: In particular, the earliest *denarius* types are often harder to distinguish and date than the later *denarius* types.

This is important because the misattribution of the various early *denarii* to the earliest *denarius* issue is a common mistake, and probably a partial cause of the observed spike in Punic War coin hoards dated after 211 BC (Lockyear 2007, 12; Harl 1996, 12-13). In this case, we might artificially observe more of, for example, the 211 issue *denarii* than of the 208 issue

[7] *RRC* 326/2: a *quinarius* issued in 101 BC (Crawford 1974).

denarii simply due to human error. Again, this is a problem because it affects some issues more than others: When misattribution affects all coins equally, it is not a problem for linear regression.

To avoid this problem, we again examine only the *denarii* of 150-100 BC because the *denarii* of 150-100 are easier to date. In this time period, the *denarii* begin to change types (Woytek 2012, 9; Harl 1996, 39-40) and thus become easier to identify.

1.4.4 The frequency of coins in hoards does not give an accurate picture of the coins actually in circulation, and therefore their relative frequencies do not give an accurate depiction of the size of each issue.

To mitigate this problem, we can look at hoards from periods of war and unrest in deference to those who doubt whether coins in hoards truly give an accurate picture of coins in circulation (for some potential problems with this assumption, see Duncan Jones 1994, Chapter 5; for the political selection of coins for their inclusion in hoards, see Howgego 1994, 12; c.f. Lockyear 1999, 219-20 who argues that hoards are random samples from the greater coin population, as we assume here). The advantage here is that a sample of coin hoards from a period of unrest is likely to contain a greater proportion of "modern" hoards, or hoards buried in haste, which give a better idea of what was actually in circulation.

This is a riskless statistical strategy; on the other hand, if there is no difference between hoards dating from periods of war and those dating from times of peace (Lockyear 1999, 219-220) we lose nothing by excluding a few years, and the hoards closing in those years, from our analysis, as the initial definition of "later" in the phrase "coins appearing in later hoards" is arbitrary. The logic should in theory hold for *any* fixed, future time period as long as there is data available for us to examine (i.e. coins of 150-100 BC still appearing in hoards).

1.4.5 By looking only at the output of *denarii*, we fail to account for the production of lower value bronze fractions.

This is indeed an unavoidable problem. Crawford is quite hesitant about the value of his die estimates for the bronze coinage, more so than for silver (Crawford 1974, 123), and as the coverage of Lockyear's database is only good for hoards of *denarii*, we cannot test them. However, it should be noted that in terms of value, *denarii* make up the bulk of the Roman money supply during this period (Crawford 1974, 696). Additionally, looking at the years from 150-100 BC also controls for this problem, as the production of bronze fractions drops precipitously after 150 BC and does not pick up again until the first century BC (Harl 1996, 47).

Before continuing, I will point out that there are a few things I did not attempt to control for. I did not exclude overly large or small hoards, despite the fact that small hoards can be shown to be less reliably dated than large hoards (due to small sample size and the use of closing dates to date hoards) and that large hoards tend to be less representative of the coins actually in circulation (Lockyear 1999, 220). This was due to the fact that defining "large" and "small" is difficult, arbitrary, and may introduce unnecessary bias into the data; both of these errors probably affect each year from 150-100 BC equally, and thus represent an instance of "classical measurement error" which should not threaten the validity of my results. Further, the imprecise dating of hoards is not a problem for this analysis, as being inaccurate by a year or two won't matter as long as the hoard is still within our cutoff range (i.e. 90-70 and 50-27).

I did have to go into the data and "cherry-pick" several hoards; namely, the hoards Crawford used to obtain his original die estimates. I ended up removing seven of his 24 hoards (after cutting out hoards not from time periods of war, outside of Italy, etc), which brought me to the final 100 hoards.[8] Since Crawford obtained most of his die estimates by multiplying an empirically determined constant by the number of extant coins in his sample of 24 hoards, if I did not remove these hoards it could be argued that linear regression would simply be rediscovering these constants.

Once his hoards are removed, and a regression run on the newer data from Lockyear's database, Crawford's estimates extend surprisingly well to the additional material.

1.4 Experiment Methodology and Problems Encountered

Crawford's output estimates were taken at face value. To obtain "frequency," or the number of coins that showed up from each year in later hoards, the following problems were encountered:

Problem 1: Some coins had a date range in Crawford's *RRC*, rather than a specific date.

Solution 1: Compute the *per annum* expected value for each year. For example, if a coin was minted in 101-100 BC, and we had 300 surviving examples of it in later hoards, assign 150 of the coins to each year (101 and 100 BC).

Problem 2: How does one deal with the lone *quinarius* 3issue?

Solution 2: Assign it a weight of ¼, according to its relative value, and add it to the other

[8] The roundness of this number was pure coincidence.

denarii.

1.5 The Experiment Results

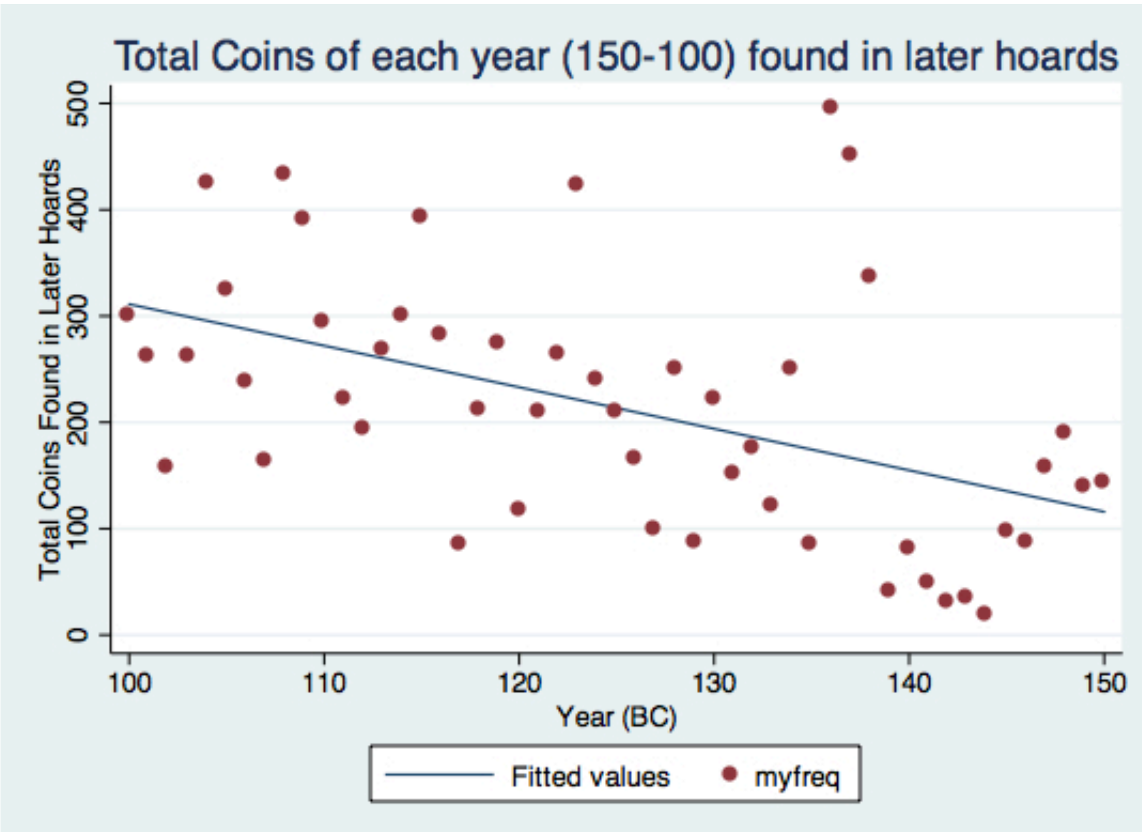
Before continuing, I will briefly review the qualitative interpretation of several numbers of interest commonly reported with linear regression. Although the following regression tables (in part one and part two) look very complicated, the reader unacquainted with statistics is encouraged to focus on the following: the R-squared value of the regression, reported in the top right, and both the p-value and coefficient for each independent variable. The R-squared value is a number between zero and one, and reflects the amount of variation explained by our regression model. In general, higher is better. The coefficient represents the effect of each independent variable (if there is more than one) on the dependent variable, and can be positive or negative. The p-value is interpreted as the "significance" of the independent variable in question, and it represents the probability of seeing data at least as extreme as that observed if the true effect of the variable is zero — in other words, if the true value of its coefficient is zero. For more information on this topic, consult the appendix at the end of this thesis.

After filtering Lockyear's data to only look at Italian hoards from the periods of unrest outlined above, I was left with 100 hoards containing 10,899 coins from the period under consideration (150-100 BC). I then ran a regression in Stata to see how well Crawford's output predicted the number of coins that remained in later hoards, controlling for age of each issue. There was significance at the .05 level[9] and an R^2 of 0.72 (which means that 72% of the variation in the data can be explained by Crawford's output estimates). Note that although "year" was not statistically significant, this was due to the inclusion of Crawford's output data.

[9] The somewhat arbitrary "gold standard" for most statistical analysis. This means that the p-value is less than .05.

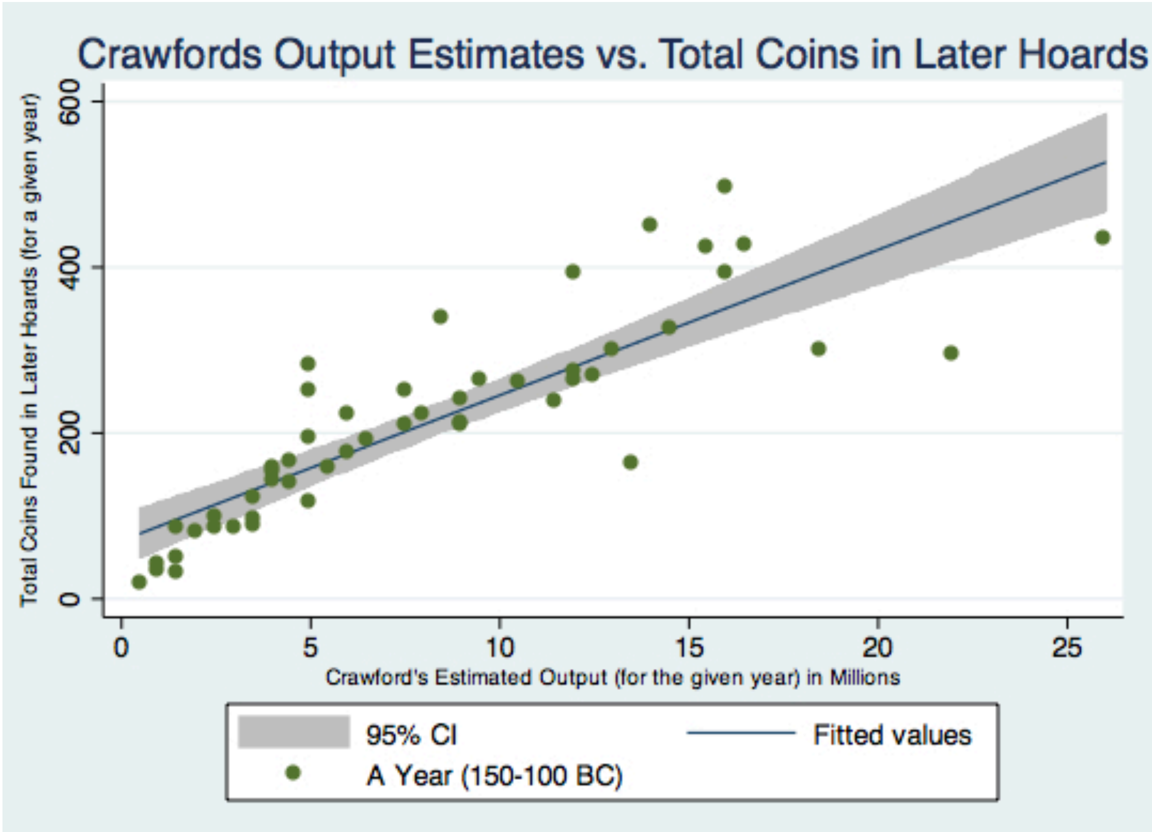
I conclude that Crawford's output numbers have some validity, as Lockyear notes, as a whole, even though they can be shown to be incorrect in individual instances (Buttrey 1976, 152). In short, because Crawford overestimates as much as he underestimates, he is simply introducing classical measurement error, which is no threat to the validity of linear regression.

Figure 4: Total Coins from 150-100 Found in Later Hoards by Year



Interpretation: There is a clear negative correlation; older coins show up less often, all other things equal (keep in mind that the years are in BC, so higher is older). This makes sense because they should fall out of circulation over time. However the correlation is weak, because other factors are stronger than coin age in determining how many we find in circulation later: namely, how much was originally minted.

Figure 5: Total Coins Found in Later Hoards vs. Estimated Output (by Year)



Interpretation: Each point is a year. For each year between 150 and 100, the higher Crawford's estimated output, the more coins we find in later hoards. This is what we would expect if Crawford's output estimates were at all valid as a totality. See figure 6 for the regression which produced the line of best fit.

Figure 6: Line of Best Fit for Coins Found in Later Hoards vs. Year

```
. reg  myfreq year,r

Linear regression                               Number of obs =         51
                                                F(  1,    49) =      18.13
                                                Prob > F      =     0.0001
                                                R-squared     =     0.2333
                                                Root MSE      =     106.33
```

myfreq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
year	-3.906244	.9173296	-4.26	0.000	-5.749687	-2.062801
_cons	701.7707	112.4371	6.24	0.000	475.82	927.7215

Interpretation: This is simply a regression of the number of coins found in later hoards by year, and gives the line of best fit in the first graph. Since age clearly matters — though other factors affect coin distribution as well — we should control for it in the next regression.

Figure 7: The Line of Best Fit for Coins Found in Later Hoards by Estimated Output, Controlling for Age

```
. reg myfreq craw_output year,r

Linear regression                               Number of obs =       51
                                                F(  2,    48) =    34.56
                                                Prob > F      =   0.0000
                                                R-squared     =   0.7257
                                                Root MSE      =   64.265
```

myfreq	Coef.	Robust Std. Err.	t	P>|t|	[95% Conf.	Interval]
craw_output	18.18859	2.997601	6.07	0.000	12.16151	24.21567
year	.417866	.8265118	0.51	0.615	-1.243947	2.079679
_cons	12.71681	119.1748	0.11	0.915	-226.9002	252.3338

Interpretation: Age no longer matters at all once the original size of each issue is accounted for. Crawford's output estimate (craw_output) is significant at the .001 level, and the model is able to predict 73% of the variance in the coin hoards of Roman Italy. The remaining variance is easily explained by random noise: misattribution of coin types, errors in data entry into the database,[10] imperfect recovery/reporting of the contents of hoards, some hoards being unrepresentative of the rest of the population, etc. Indeed, if the model explained 100% of the variance, I would be more skeptical of its validity! It should also be noted that robust standard errors were used (the "r" at the end of the command `reg myfreq craw_output year,r`) as the data is heteroskedastic: The variance is higher for the larger values, as is visible on the graph and provable using Stata's built in `hettest` command.

[10] Not that I have any proof of this, but it is always wise to assume some level of human error when working with a database that has thousands of hand-entered entries.

1.6 Conclusion for Part One: Checking Crawford's Output Estimates

As shown, Crawford's estimates as a whole have some validity and represent the relative output of the Roman mint much better than is commonly supposed. As such, it should be possible to compare Crawford's output estimates against other archaeological data, and see whether or not any interesting correlations occur which might suggest why the Romans minted coins.

In section two, I will compare Crawford's output estimates against military expenditures, also as calculated by Crawford (to determine the impact of military spending on Roman mint output) and shipwrecks *per annum* as reported on the OXREP Database (as a proxy for trade, economic growth, and hence economic considerations). The difficulties with the data will be examined, and I will argue that it is possible that one of the reasons behind the Roman's minting coins was a desire to keep pace with the demand for money in a growing economy.

Part Two: Why did the Romans Mint Coins?

2.1 Rudimentary Monetary Policy

Having demonstrated that Crawford's die estimates are fairly accurate for the period 150 to 100 BC, I will now use them to explore one of the theories mentioned above for why the Romans may have minted so many coins: as rudimentary monetary policy in response to a growing economy. The idea is not that the Romans had a sophisticated understanding of macroeconomics, but simply that the Romans may have minted and introduced new coins into circulation in response to a growing demand for coinage from an increasingly "monetized" populace and growing economy in Italy (Lo Cascio 1993; 1981). The mechanism for this explanation is as follows. Imagine that the Roman state needs to make a payment, and at some points during this time period (150-100 BC) faces a choice: With both bullion and old coin in the treasury, do they mint new coin from the bullion, or use old coin or the bullion itself instead? At the margin, the argument is that this decision would have been influenced by a perceived need by the common people for more coin due to a growing economy, and that as a result there should be a correlation between Roman mint output and real economic growth.

This argument is supported by the somewhat miraculous level of price stability that prevailed in the Roman economy during the republic (Temin 2013, 81-82; Lo Cascio 1981, 76; this is not to say that inflation was absent: c.f. Harl 1996, 43-4) given the large amount of coins being produced at the time, as the standard Liquidity Theory of Money dictates that an increase in money supply should result in higher prices (Mankiw 2015; c.f. Harris 2006, who applies the Quantity Theory of Money to the Roman world and argues that credit could have expanded to accommodate some, perhaps all, of this new “money demand”) unless money demand, which depends, among other things, on the overall level of economic activity, is increasing as well. Is it possible that the Romans deliberately increased the money supply in order to keep pace with a growing economy and keep inflation stable, as the monetary authorities of most countries do today?

The first step in considering this question seems to be to check whether or not Roman mint output and economic growth are correlated. However, the oft-repeated warning “correlation does not imply causation” must be sounded here. Just because we may be able to observe, empirically, that economic growth and Roman mint output moved together (and helped establish price stability) does not mean that the Romans deliberately minted coins in order to maintain stable prices. In fact, price stability may have been a happy accident for most of the Roman Empire and Republic, arising out of the fact that the state’s need for money was correlated with economic growth, since the primary expenditure of the state was imperial expansion (Temin 2013, 82). In other words, it is possible that military expansion drove both mint output and economic growth.

This argument is particularly important, since it is by no means clear that the economic sophistication of the Roman elites would have been sufficient to recognize the link between increased minting and higher price levels, even during the Empire (Temin 2013, 82; c.f. Vivenza 2012 and Duncan-Jones 1994, 25) let alone for the time period between 150 and 100 BC. It is, for example, debated whether or not the closure of the Macedonian silver mines in 167 could have been due to fears of inflation that a new influx of silver would cause since it is not clear that the senate would have understood the concepts of supply and demand (Kay 2014, 57; c.f. Lo Cascio 1981, 81). That said, it is possible that the Roman senate may have understood the causal link; Polybius wrote, woefully imprecisely, that “in his time,” a gold mine was found at Aquilea which was so productive that it caused the price of gold to drop by one-third in Italy (Polybius *Histories* 34.10.10-15). It has been pointed out that this event may very well have taught the Roman elite the basic price determination of supply and demand (Kay 2014, 57) even if they did not understand the concept before. It depends, of course, on exactly when the event Polybius alludes to took place. If the senate did not understand supply and demand, then the theory of rudimentary monetary policy would be unlikely to be true.

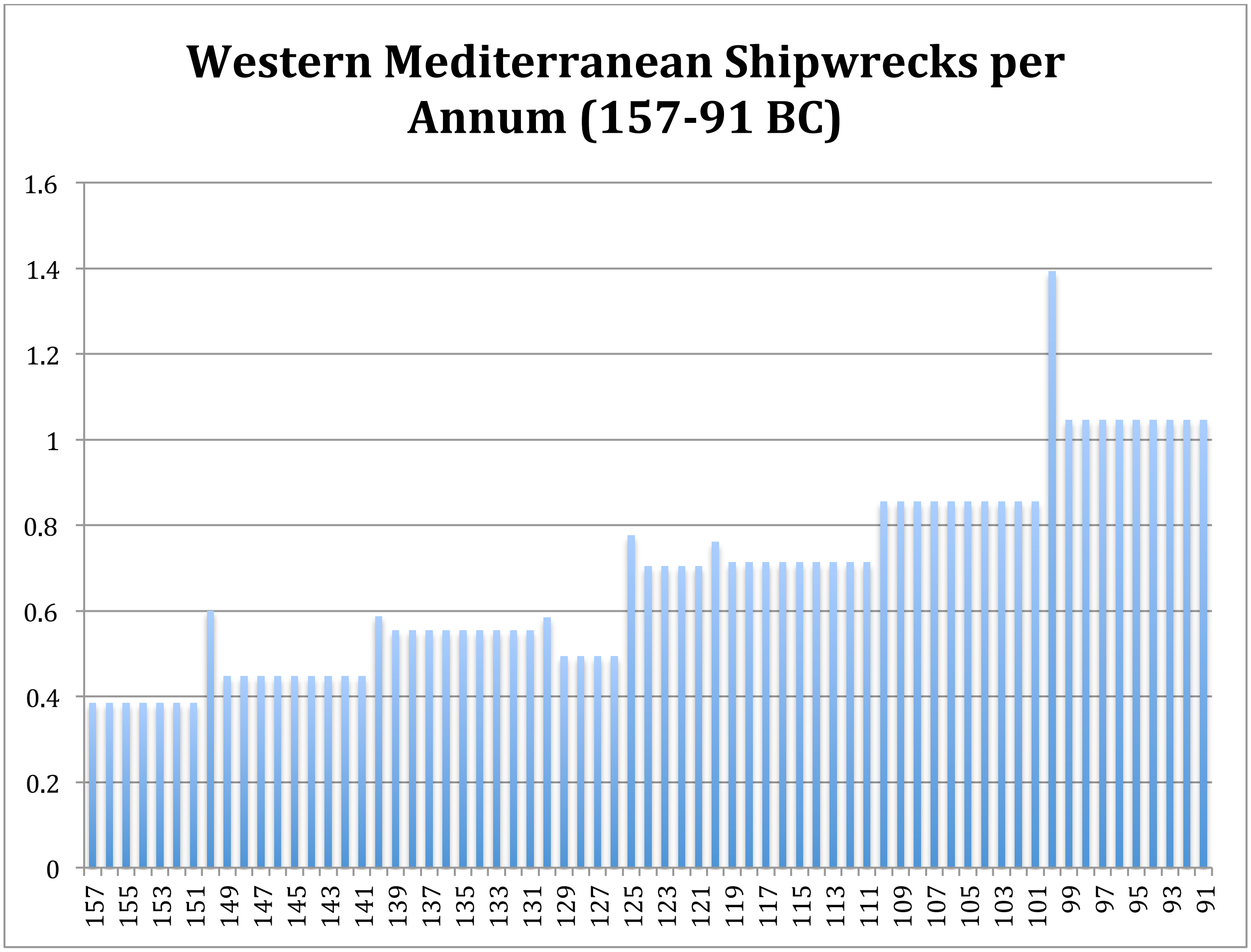
In order to examine whether or not the Romans may have minted coins in response to the increasing demand for money in Italy (at least to some degree) and not solely to meet state expenditures, we will look at a shipwrecks *per annum* —as a measure of economic growth —alongside both Military Expenditures and mint output, similar to the graph presented in Keith Hopkins’ famous paper (Hopkins 1980, 110). However, we must be cautious, as there are two major difficulties: problems with the data, and problems with statistics resulting from the nature of time series data.

In the following sections 2.2 and 2.3, I will discuss problems with shipwreck data as an index of trade and economic growth and Crawford’s military expenditures. In 2.4, I will then briefly comment on the problems with looking at graphs of “data over time” which can have surprisingly close correlations despite not having any real relationship, and must be analyzed very cautiously. In light of this, I will run a regression using first-differenced, log-transformed data in section 2.5, effectively looking at percent changes year-to-year in each variable in a manner similar to Hopkins’ 1980 treatment, and demonstrate that such a treatment reveals that both shipwrecks and military expenditures are equally good at explaining mint output.

2.2 Difficulties with "Shipwrecks as Trade"

In order to examine whether or not Roman mint output kept pace with economic growth, we need a proxy for economic activity. Here, we will use shipwrecks *per annum* from the OXREP database (Straus 2013) as has been suggested elsewhere (Paollili 2008; Hopkins 1980) as a proxy for trade. By *per annum*, I mean that I construct a histogram such that each shipwreck is divided up among every year in its range equally, including the endpoints, and the appropriate fraction is added to each bin of the histogram: For example, for a ship that could have sank in 101 BC or 102 BC, I would add .5 to each year (although in reality the ranges are often much larger). For the time period under consideration, 150-100 BC, there are 166 shipwrecks that might have sunk during this time. Out of these, there are 120 shipwrecks with endpoints in the range; which give the graph its "peaks and valleys" and yearly changes. Due to what I will here call "endpoint bias" (the tendency for archaeologists to round dates to the nearest century) there is an artificial peak at 100 BC: If we dropped that year from our analysis, we would lose 49 of our 120 wrecks that have endpoints in the range. With this graph, we have a rough measure of "shipwrecks," and hence shipping, which in theory reflects trade. The graph below includes a few additional years, for comparison, as this is the range of data which Crawford calculated in *Roman Republican Coinage*.

Figure 2.1 Shipwrecks from Libya, Spain, France, Italy, and the Adriatic



Although the direction of *causation* between trade and income/economic growth is complicated (Paollili 2008 claims trade caused an increase in growth during the Roman Empire through enabling regional specialization; for more modern arguments, see Frankel and Romer 1999), the *correlation* between the two is well established, and thus looking at changes in trade should give us some idea regarding changes in overall economic growth and overall income, which is what we are after here. However, much like hoard data, the use of shipwreck data in such a way has many potential problems and pitfalls and needs to be used cautiously. There are three main problems: geographic bias, survival bias, and inaccurate dating. I will list the problems here, and explain how we attempt to deal with them further below.

First, there is a large geographical bias in the OXREP database. 70% of all shipwrecks are found in the western Mediterranean (Gibbins 2001, 279) and this almost certainly reflects more intensive archaeological exploration, as well as local traditions of recreational diving in different countries. Further, we only find and date wrecks that sink in shallow waters, which might be a problem if more ships sailed in, and sank in, deeper waters in one period than in another (Wilson 2009, 229). This introduces another problem, which is that we often do not know the ports that a given ship would have been stopping in. Just because a ship was sailing past a region when it sank does not necessarily mean that it traded there. This is particularly concerning given that long distance trade might have been a very large part of Roman trade as a whole, in terms of value, even if only represented by a few ships, especially during the Empire (Wilson 2009, 217).

Further, we also deal with problems of survival bias. The assumption is that the number of shipwrecks is correlated with the number of ships sailing in a given year. However, this need not necessarily be the case, especially over long periods of time. Improvements in naval technology mean that larger and larger merchant fleets may leave behind fewer and fewer wrecks, as ships may sink less and less often. In particular, the introduction of ship pumps in the first century BC and the use of deadeyes[11] in the rigging of ships in the second century AD onward may have dramatically reduced the number of ships sinking, and thus the drop off in wrecks that occurs, for example, in the later Empire may simply be reflecting improved naval technology (Harris 2009, 260). Further, the only shipwrecks found are usually those carrying amphorae: Ships carrying sacks of grain, perishable cargo, or slaves, would not survive and vanish from the archaeological record (Wilson 2009, 228). Thus, our picture of Roman trade as

[11] A piece of wood with several holes used to guide and change the direction of lines in the rigging.

obtained from shipwrecks is incomplete and heavily biased towards durable goods and goods carried in amphorae.

Finally, the last problem is that, unlike with coin hoards, specific dates for shipwrecks are almost never available, even as "best guesses." The data in the OXREP database reflects this, giving only ranges for dates of each shipwreck, sometimes as large as a century or more. As a result, most histograms given in the literature use *per annum* estimates for shipwrecks per year as we do here. As a result, there is considerable noise and curve smoothing in this data set. This is the most severe problem, in my opinion, with the data, as will be discussed below.

2.2.1 Geographic Bias

We will tackle this set of problems in ascending order of difficulty. First, the problem of "only" discovering coastal wrecks is easily dismissed, as that is the place where most ships tend to wreck. It was common for ships to be blown by storms for many miles before finally foundering upon rocks, as is testified to both by the ancient sources and comparative evidence from Venetian shipping records, which report very few ships sinking in deep water (Gibbins 2001, 280-281). Second, there is the issue of long distance trade and geographical bias, which are both solved when the cargo of most wrecks is considered. Given the small size of most shipwrecks and their mixed cargos, it is probable that we are mostly uncovering wrecks engaged in "tramping," or the speculative and small scale contractual transport of goods along coastal routes, often within an established economic region (Dietler 2010, 132-138, Gibbins 2001, 294-295). As a result, we will restrict our analysis to regions in which archaeological exploration and recreational diving have been similarly prolific, and consider only wrecks from the Western

Mediterranean: the coastlines of Spain, Libya, France, and Italy including the Adriatic (and thus wrecks off modern day Croatia, etc.). We will also qualify our proxy variable as not standing in for "trade," but for regional small-scale tramping, which is linked to the level of activity in the larger nautical industry as often these smaller ships carried parts for larger ships (Gibbins 2001, 295). As a result, our proxy variable may not measure "trade" as a whole, but it will certainly be strongly *correlated* with trade as a whole, and hence with overall economic activity.

2.2.2 Survival Bias

Survival bias is indeed a large problem. We cannot pretend that by measuring shipwrecks we are measuring total trade. However, amphorae were widely used, and thus surviving shipwrecks represent local trade as described above in a variety of important goods, including *garum*, olive oil, and wine (Gibbins 2001, 271). As such, shipwrecks are still a reasonable measure of regional economic activity as described above.

Survival bias due to changing technology is a more severe concern. However, by focusing on the period from 150-100 BC, we restrict the time period under consideration to before the first century BC, when pumps started to be widely used on ships (Harris 2009, 260), and thus avoid excessively long-term comparison. The assumption that naval technology in the ancient world was fairly constant over most 50 year intervals seems reasonable for the purposes of this experiment.

2.2.3 Inaccurate Dating

Statistical noise due to inaccurate dating can only be solved with a larger sample size: in other words, discovering more shipwrecks. There is no way around this particular problem at the moment, since I must work with the data available. As a result, it may be difficult to obtain statistically significant results, since theory predicts that random noise both biases t-scores down (by increasing measured confidence intervals) and decreases R-squared values.

2.3 Crawford's Military Expenditure Estimates

Crawford's estimates for state expenditure are really estimates for *military* expenditure, since all he did to arrive at them was apply the standard military stipend to Brunt's calculations for troop estimates in *Italian Manpower* (Crawford 1974, 696-697; Brunt 1971, 432-433). In calculating the cost of a legion here, we no longer have to worry about difficulties posed by the fact that much of these costs could have been met by payments in kind, old coin, etc. If these things are true, then mint output should not correlate strongly with military expenditure. In other words we care only about cost, and not how the Romans *met* these costs. The difficulties regarding precise legion size still remain (see de Ligt 2007; Roth 1994) but inasmuch as these fluctuations are random from year to year, we do not need to overly concern ourselves with them. However, most concerning is the fact that the cost of a "legion" changed during this time period: Crawford's estimates rely on his raising the cost of a Roman legion from 600,000 *denarii* to 1,500,000 *denarii* in 123 BC, a change he attributes to C. Gracchus' reforms (Crawford 1974, 697) reflecting an increase in the size of the general's staff. These estimates are based on the classical sources, and may be inaccurate.

2.4 Problems With Time Series Data

Having discussed the problems with the data, we will now turn to the problems of looking at data that is recorded over time, known as time series data. As Hopkins noted, normal regression (as used in part one) and normal correlations can give erroneously high numbers for two sets of time series data, even if the two time series graphed against each other have no real relationship (Hopkins 1980, 111). This is particularly a problem when the two variables both trend up or down over time. For an example of this, see the graph in the appendix on linear regression, which displays the remarkable correlation between U.S. lemon imports and traffic fatalities for the late 1990s (Johnson 2008). This phenomenon is known as "spurious regression," and Hopkins attempted to control for it by looking at inter-annual percent changes rather than absolute values over time. Here, we will adopt a very similar method and run a natural log transformed first differenced regression, which removes the upward trends in our data (compare figures 2.3 and 2.4 below). We may interpret this more intuitively as the effect of a percentage change in X, or our independent variables, on the percentage change in Y since the difference between two logs is, especially for small differences, mathematically almost the same as the percentage difference.[12]

The conclusions are as follows: From analysis such as the correlations calculated by Hopkins (Hopkins 1980, 111) and the visual inspection of Crawford's graphs, which he claimed

[12] Regressing on inter-annual percentage changes gives residuals, or error terms, which are highly skewed (`sktest` and `swilk` in Stata both reject the null that they are normally distributed) and thus we use the analogous log transformations instead, which does not have this problem. This is important because without normally distributed error terms, normal hypothesis testing (used to ascertain significance) is invalid. Correlations for percentage changes as Hopkins calculated for this time period (150-100 BC) are .24 for military expenditures and a nearly negligible .05 for shipwrecks in our sample.

resulted in a "remarkable correlation" (Crawford 1974, 694), we *cannot* conclude that military expenditure is the only, or even the main, driver of mint output. Moreover, by using first differences regression with natural log transformations (presented below in figure 2.2), it becomes apparent that although both shipwrecks and military expenditures are correlated with mint output and significant at the .05 level, shipwrecks actually have greater impact on trade than military expenditures.[13]

2.4 Interpreting the Regression

The first differences regression of the natural log of mint output on the natural logs of military expenditure and shipwrecks is presented below in figure 2.2. The results indicate that a percentage change in shipwrecks has a *greater* effect on the percent change in output than does a percentage change in military expenditures, since the coefficient on that variable is 50% larger.[14]

---

[13] To diagnose the regression we check for autocorrelation in the residuals and non-stationarity in our time series data, both of which can lead to violations of ordinary OLS assumptions. Non-stationarity is a violation of the assumption that the characteristics of the data (mean, variance, etc.) do not change over time, while autocorrelation is the correlation of the residuals with each other (a frequent problem in time series data). To test for autocorrelation: the Durbin Watson test statistic was calculated using the `dwstat` command in Stata, and this returned 2.77, which suggests negative autocorrelation in the residuals at the .01 level. To account for this, Newey-West standard errors were used in the regression using the command `newey`, with a lag of 3 time periods. To test for stationarity: Augmented Dickey-Fuller (ADF) tests (implemented using the `dfuller` command in Stata) were run on the transformed data and on the residuals to check for stationarity using 1, 2 and 3 period time lags with and without trend and drift variables. We were able to reject the null of non-stationarity at the .01 level for all the data used here, once it was transformed, for all of these tests (except one, for which we could only reject at the .05 level: the p-value was .012 for log shipwrecks with two lags and a trend variable). Note that these ADF tests and Newey-West standard errors are based on papers published after 1980, and would not have been available to Hopkins.

[14] I also ran ordinary first difference regression with Newey-West standard errors, and used Feasible Generalized Least Squares methods on the difference in logs data above (both the Prais-Winsten and Cochrane-Orcutt methods using the `prais` command in Stata). All these gave the same results (significance at the .05 level and larger coefficients for shipwrecks) with the exception of the Cochrane-Orcutt method, which returned a p-value of .052 for shipwrecks (this being due, in my opinion, to the fact that this method requires dropping an observation and losing data, due to the way in which it is calculated). I report first difference in log transformations because it has the simplest interpretation (as percent change) and because Newey-West standard errors are preferable to Feasible Generalized Least Squares methods for large enough samples sizes due to the fact that Newey-West (see next page) does not assume that the errors follow any particular process (FGLS methods assume that the errors are correlated in only one particular way, which we do not know to be the case).

Note that the R-squared is .18 (there is no R-squared reported because Newey-West standard errors were used in this regression, see footnote number 14; the R-squared was obtained running a normal regression without using Newey-West standard errors, which should not impact the value of the R-squared), either suggesting the existence of other, unmeasured explanatory variables or significant measurement error and other noise, or both.

Figure 2.2: First-Differenced, Natural Log Transformed Regression of Mint Output on Shipwrecks *per annum* and Military Expenditures

```
. newey  diflogoutput diflogexp diflogships,lag(3)

Regression with Newey-West standard errors          Number of obs  =        51
maximum lag: 3                                      F(  2,     48)  =      6.39
                                                    Prob > F       =    0.0035

------------------------------------------------------------------------------
             |             Newey-West
diflogoutput |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   diflogexp |   .9634534   .3646512     2.64   0.011      .2302729    1.696634
 diflogships |   1.558914   .4436095     3.51   0.001      .6669769     2.45085
       _cons |  -.0243277   .0544755    -0.45   0.657     -.1338581    .0852027
------------------------------------------------------------------------------
```

Figure 2.3: Difference in Log Values (Percent Change Per Year) 150-100 BC

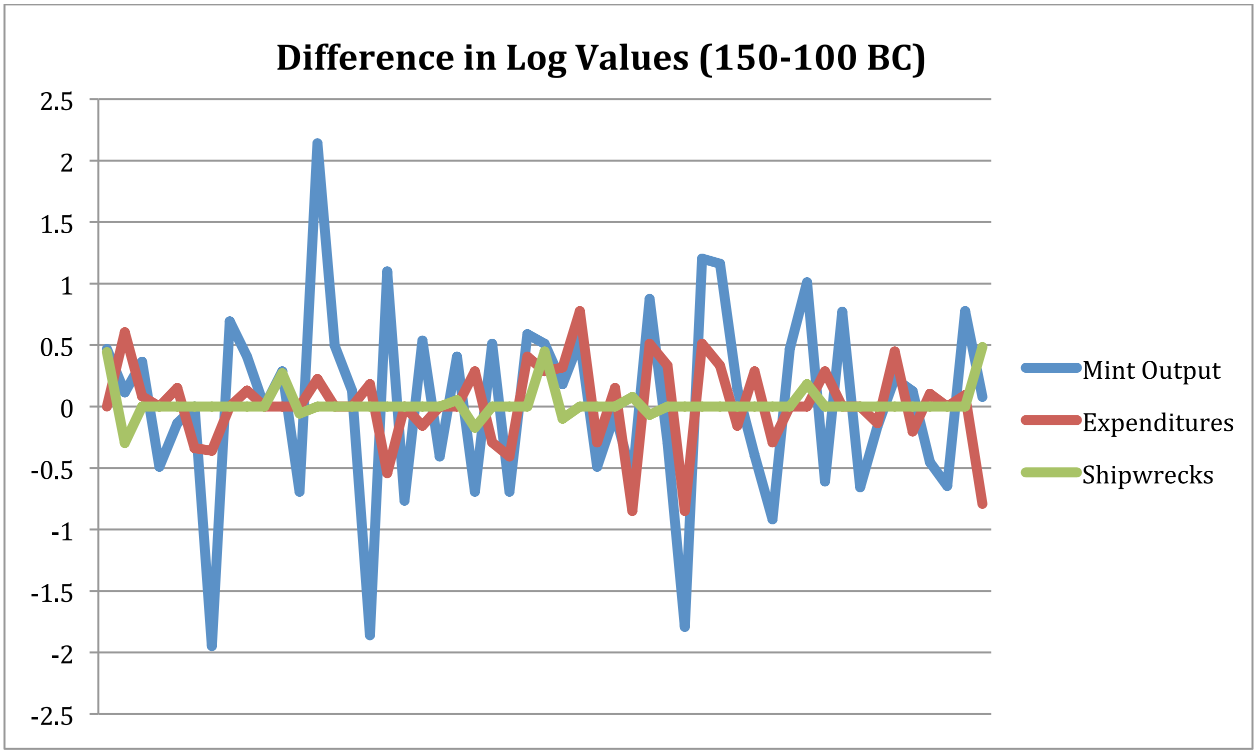
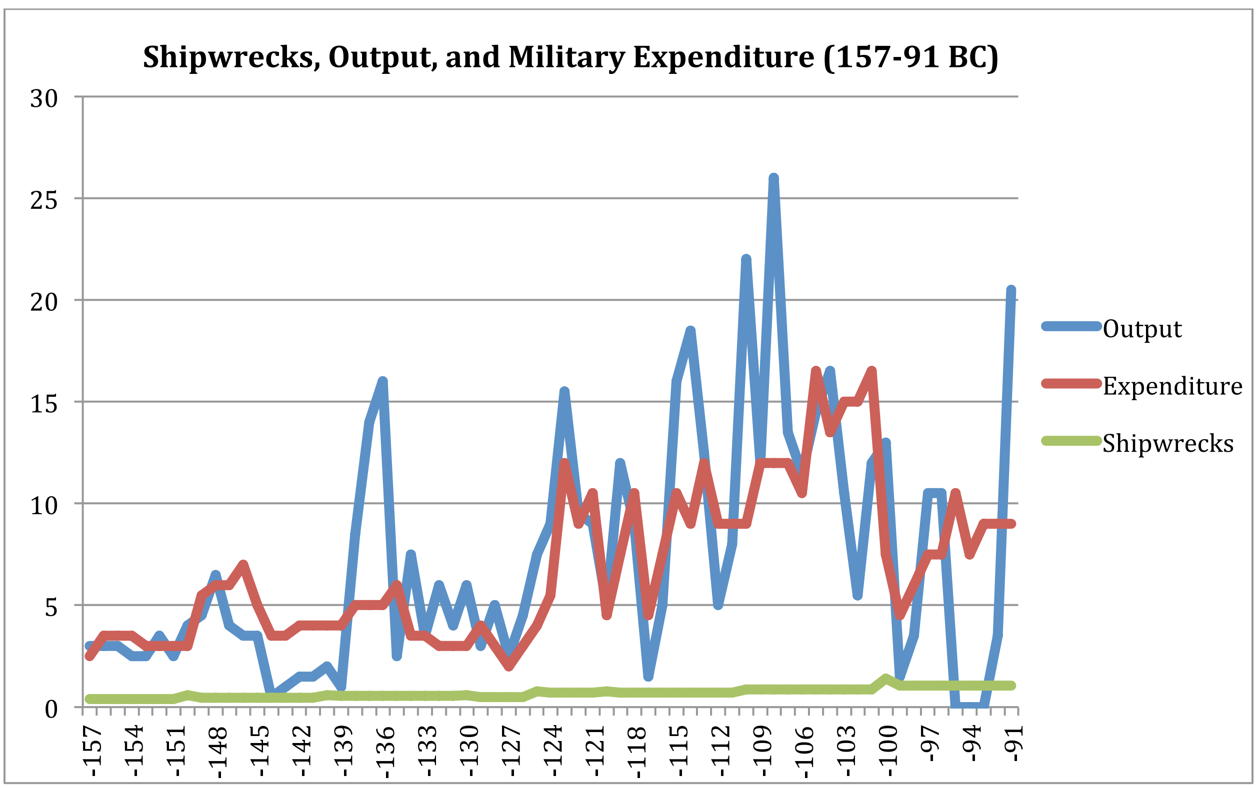


Figure 2.4: Hopkins (1980) Graph with Shipwrecks Overlaid 157-91 BC



As shown from this regression, for the years 150-100 BC[15] shipwrecks, and hence trade, have at least as much explanatory power as military expenditure, although both are significant, as long as our initial regression assumption that each X *causes* Y is accurate, and that the reverse does not hold true. While we may be relatively certain that Roman mint output was not a cause of Roman military expenditure, we may wonder if increases in Roman mint output might have *caused* short-term economic growth by reducing interest rates in accordance with the Liquidity Theory of Money. In this case, the observed significance may be due to the fact that Roman mint output increased economic activity, and hence trade.

However, this is unlikely for two reasons. First, coins are not the only determinant of the ancient money supply (de Callataÿ 2011, 15-16; Harris 2006), which is a problem because it is changes in money supply, and not mint output, that should influence economic growth in the short run according to the Liquidity Theory of Money. Second, the effect of monetary policy in modern economies is notoriously "lagged" by an amount that varies from six months (Mankiw 2015, 764) to two years (Lindbeck 2015). In ancient economies, with their far more rudimentary capital markets, it would take longer for interest rates (and hence output) to adjust. So even if changes in mint output *did* represent the overall money supply, or more reasonably was heavily correlated with changes in the Roman money supply, we would not expect to see an effect in the same year that the coins were minted. Thus, the causality probably goes the other way.

However, we must be hesitant regarding any conclusions drawn from such a simple regression (which only includes two independent variables!) on such a complex system as the Roman economy. It is always possible that some other variable, correlated with both mint output

[15] For the sake of thoroughness, it should be noted that although I restricted my analysis here to the years for which I tested the accuracy of Crawford's output estimates in part one, if one is willing to accept *all*'s output estimates as valid and includes the years from 157 down to 91 BC for which the difference in log values is calculable (natural log of zero is undefined) the results do not change for the regression in figure 2.2.

and shipwrecks, is actually driving both. For example, one might consider large public works projects or other forms of government spending not captured by Crawford's military expenditures, which the Roman state would have both minted coins to fund and also might have increased local economic activity. We cannot control in any systematic fashion for this particular variable at this time.

2.5 Conclusion

The data and regression presented above suggest that it is possible that Lo Cascio's theory of rudimentary monetary policy may have been a driving factor in the determining Roman mint output, alongside military expenditures. We conclude the following:

1) Rudimentary monetary policy is one possible explanation for the correlation observed in the data.
2) There may be other variables causing the perceived correlation (for example, other unmeasured state expenditures).
3) Military expenditures are *not* the only factor in determining Roman mint output.
4) There is still room for other explanations for minting, such as political concerns (von Reden 2012, 272).

In short, we must rely on the arguments from the classical sources and our own judgment to determine whether or not Roman monetary policy is a feasible explanation for the correlation observed in the data.

Part Three: Conclusions and Suggestions for Future Research

Statistics can be quite useful for helping archaeologists make sense of large amounts of data, and hard numbers in particular can force us to confront preconceived notions regarding the behavior of ancient governments (de Callataÿ 2011, 17). As is the case here, careful statistical analysis provides further evidence that two commonly held beliefs should be questioned: that Crawford's output estimates are too flawed to be of any use, and that the output of the Roman mint was overwhelmingly determined by the need for military expenditures to be met.

As demonstrated in part one, linear regression with hoard data reveals that Crawford's output estimates (at least for the period tested here, 150-100 BC) are quite good as relative measures of the amount of *denarii* the Roman mint produced from year-to-year, regardless of the manner in which he arrived at them. These results buttress some of the weaknesses in prior work (Lockyear 1999) and together provide further confidence moving forward in the use of such estimates as measures of Roman mint output.

However, as examined in part two, the data that Crawford collected does not support the hypothesis that military expenditures are the only, or even the most important, determination of the output of the Roman mint. Using slightly more modern techniques, but with an interpretation similar to Hopkins (Hopkins 1980), linear regression reveals that shipwrecks *per annum* (interpreted here as changes in real economic output and trade) are equally correlated with the output of the Roman mint, providing further evidence for Lo Cascio's theory of rudimentary monetary policy.

Going forward, whether or not the correlation of mint output and trade also extends to other measures of trade or economic growth (perhaps measured from pottery, although this comes with its own pitfalls) in the Italian peninsula. If it did, this might make us more confident in Lo Cascio's theory. Additionally, as mentioned above, currently we cannot control for other state expenditures, which might explain the correlation between mint output and economic growth. Including data on government expenditures, such as they may be, perhaps in addition with looking at the level of inscriptions or dedications *per annum* might be able to give us some idea of the level of non-military government spending. If such a measure were included in the above regression, and it decreased the significance of our variable for shipwrecks, then we might cast out Lo Cascio's theory.

Hopefully, the reader will now be convinced that the pessimism embodied in the quote at the beginning of this thesis is understandable, but unjustified. The statistical treatment of quantifiable archaeological data can still be quite profitable. Moreover, as the amount of data that is being published in easily accessible, and easily analyzable, form continues to increase, there will only be more room for analysis of the style here and an increasing role for statistics in archaeology. The implications are daunting, as the potential for error will always remain a concern, but also exciting as they allow us to revitalize or disprove old theories with new evidence.

Appendix: Linear Regression

Describing a Regression: Interpreting Coefficients, P-Values, and R-Squared

This paper makes heavy use of linear regression, which computes the line of best fit between two (or more) sets of data. In other words, we solve for a line *assuming* changes in one or more independent variables (Xs) can explain changes in a dependent variable (Y). Then, several numbers are commonly reported: the coefficient(s) for the independent variable(s), the p-value of each of these coefficients, and the R-squared, described as follows:

1) *Coefficient(s) for the Independent Variable(s):* The size of this indicates the magnitude of the effect a one-unit change in each X will have on Y. This can be either positive or negative, indicating a positive or negative effect.

2) *P – Value:* For each coefficient, the probability of obtaining such a coefficient (or a more extreme/larger coefficient) given that the real value of the coefficient is zero. If the probability is less than 5% (p-value of .05 or less) it is usually assumed that the given independent variable does have an impact on the dependent variable. In the tables here, very small p-values are rounded to zero.

Note that an insignificant result (p-value greater than .05) does not mean that the true value of the coefficient is zero (i.e. that X does not affect Y) only that we cannot say that it *isn't* zero. The true value could still be anything.

3) *R – Squared:* The fraction of the variation in the dependent variable that is explained by the independent variables (a number between 0 and 1). Note that this number can only go up as more independent variables are added. A low R–squared can either reflect noise in the data (i.e. inaccuracy in measurement) or the existence of other explanatory variables which aren't included in the regression.

Note that it is possible to have a very significant variable that has a small coefficient and a very small contribution to R-squared. Imagine a system that is determined by many, many variables: All will be significant, but each will only be able to explain a part of the variation of the dependent variable.

Limitations of Linear Regression

Looking at the three numbers described above, however, is not sufficient. The researcher must also examine the graphs of the data and use judgment regarding the initial regression assumption (i.e. does it make *sense* that X effects Y?). If it does not make sense, we might consider the results to be a statistical fluke or due to some third variable which drives both the dependent and the independent variable (imagine I notice that "trade" and "mint output" track together closely; I might question this relationship because it is probable that both trade and mint output are positively correlated with military expansion and I would want to include a variable representing military expansion in my regression if possible). Good statistics is not a replacement for good research, or for good judgment, in any endeavor.

It is important to also look at the graphs of any regression whenever possible, as demonstrated by Anscombe's Quartet (Anscombe 1973, 19-20) in figures 1-4 reproduced below. Notably, each of these graphs represents a dataset with *exactly* the same coefficients (.5) and R-squared (.67) but each graph tells a very different story:

Figure A: Anscombe's Quartet

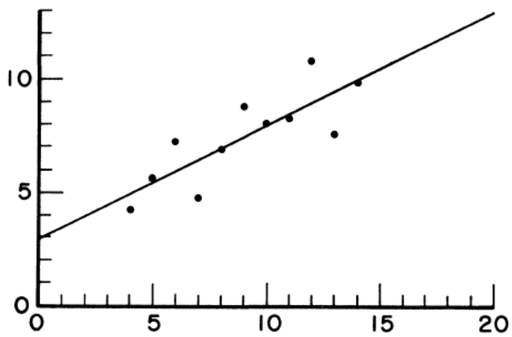


Figure 1

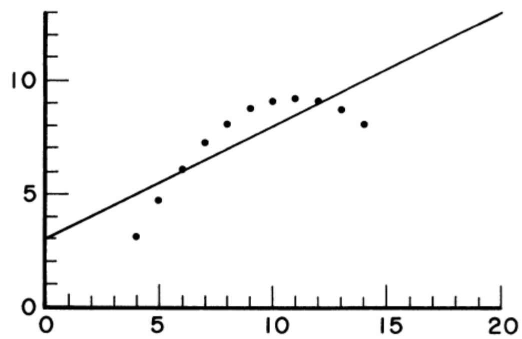


Figure 2

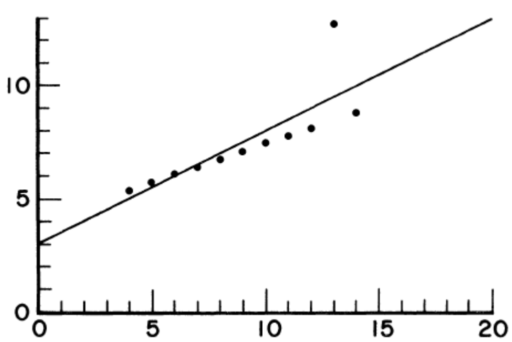


Figure 3

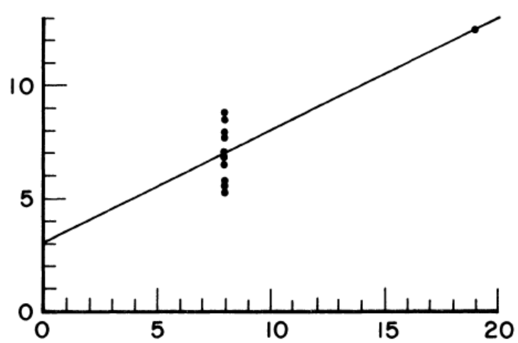


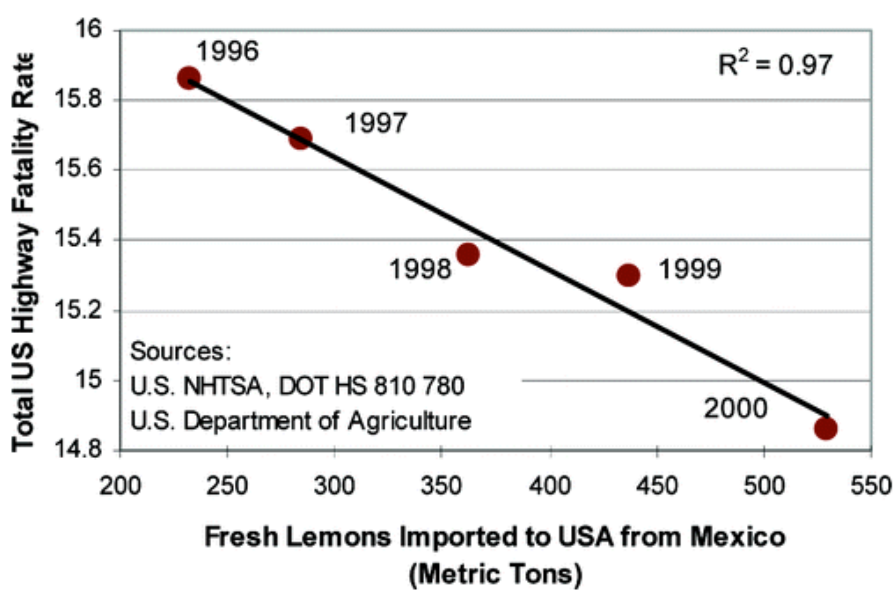
Figure 4

Figure 1 is what we hope to see in our regressions: a relationship between X and Y, and, since the R-squared is .67, some "noise," representing either measurement error in Y or room for another explanatory variable which is not correlated with X (if it *were* correlated with X, then we might see points on one part of the X axis be more spread out around the line than on another part). In figure 2, X and Y have a relationship that is definitely not linear, so a different (non-linear) regression model would be more appropriate. In Figure 3 and Figure 4, the regression is being heavily affected by one observation, an outlier, which is important to consider: Is this a measurement error, or due to some idiosyncrasy at that point?

As a final note, the major regression assumption made at the beginning was that X effects Y, but this assumption is made at the discretion of the researcher. Regression does *not* tell us anything about the direction of an effect. It only reflects *correlations* between variables. For example, in the graph in figure 1, I can say that either X causes Y, Y causes X, or both. The choice is up to the judgment of the researcher.

It is also important to remember that, given enough data, it is always possible to find correlations somewhere:

Figure B: Traffic Fatalities vs. Fresh Lemon Imports



(Reproduced from Johnson 2008).

The relationship here is clearly a statistical fluke, unless one is willing to posit a causal link between lemonade consumption and traffic safety. This graph is either simply due to chance, or another variable driving both X and Y (imagine that income in the united states grew between 1996 and 2000, causing the citizens there to both import more lemons and spend more on safety features in cars). This sort of accidental correlation is particularly common in time series data, and is usually why time series data needs to be carefully scrutinized and possibly transformed before use, as done in part two.

In short, statistical methods are not a replacement for qualitative research and good judgment. The two must inform each other.

Works Cited

Anscombe, F. J. (1973) "Graphs in Statistical Analysis." *The American Statistician* 27, no. 1 (February 1, 1973): 17–21.

Aperghis, G. G. (2004) *The Seleukid Royal Economy : The Finances and Financial Administration of the Seleukid Empire*. London: Cambridge University Press.

Brunt, P. A. (1971) *Italian Manpower, 225 B.C.-A.D. 14*. Oxford: Oxford University Press.

Burns, A. R. (1927) *Money and Monetary Policy in Early Times*. New York: Alfred A. Knopf.

Buttrey, T. V. (1993) "Calculating Ancient Coin Production: Facts and Fantasies." *The Numismatic Chronicle (1966-)* 153 (January 1, 1993): 335–51.

———. (1977) "Review." *The Classical World* 71, no. 2 (October 1, 1977): 151–53.

Crawford, Michael H. (1974) *Roman Republican Coinage*. Cambridge: Cambridge University Press.

De Callataÿ, François. (2011) "Quantifying Monetary Production in Greco-Roman Times: A General Frame." In *Quantifying Monetary Production in Greco-Roman Times*, edited by Francois de Callataÿ. Bari: Edipuglia.

———. (2005) "A Quantitative Survey of Hellenistic Coinages: Recent Achievements." In *Making, Moving and Managing: The New World of Ancient Economies, 323-31 BC*, edited by Zosia H. Archibald, John Kenyon Davies, and Vincent Gabrielsen, 73–91. Oxford: Oxbow Books.

———. (1995) "Calculating Ancient Coin Production: Seeking a Balance." *The Numismatic Chronicle (1966-)* 155 (January 1, 1995): 289–311.

De Ligt, Luuk. (2007) "Roman Manpower and Recruitment." In *A Companion to the Roman Army*, edited by Paul Erdkamp, 114–31. Blackwell Companions to the Ancient World. Malden, MA ; Oxford: Blackwell.

Dietler, Michael. (2010) *Archaeologies of Colonialism*. Berkeley: University of California Press.

Duncan-Jones, Richard (1994) *Money and Government in the Roman Empire*. Cambridge: Cambridge University Press.

Frankel, J. and Romer, D. (1999) "Does Trade Cause Growth?" *American Economic Review* 89, no. 3 (1999): 379–99.

Frier, Bruce W. (1976) "Michael H. Crawford: 'Roman Republican Coinage' (Book Review)." *Phoenix* 30, no. 4 (January 1, 1976): 375.

Gibbins, David. (2001) “Shipwrecks and Hellenistic Trade.” In *Hellenistic Economies*, edited by Z Archibald and et al. London: Routledge.

Greene, Kevin. (1986) *The Archaeology of the Roman Economy*. Great Britain: University of California Press.

Harl, Kenneth W. (1996) *Coinage in the Roman Economy, 300 B.C. to A.D. 700*. Ancient Society and History. Baltimore: Johns Hopkins University Press.

Harris, W. V. (2009)“A Comment on Andrew Wilson: ‘Approaches to Quantifying Roman Trade.’” In *Quantifying the Roman Economy: Methods and Problems*, 259–65. Oxford: Oxford University Press.

———. (2006) “A Revisionist View of Roman Money.” *The Journal of Roman Studies* 96 (January 1, 2006): 1–24.

Hitchner, Bruce. (2009) “Coinage and Metal Supply.” In *Quantifying the Roman Economy: Methods and Problems*, edited by Andrew Wilson and Alan Bowman, 281–86. Oxford: Oxford University Press.

Hopkins, Keith. (1980) “Taxes and Trade in the Roman Empire (200 B.C.-A.D. 400).” *The Journal of Roman Studies* 70 (January 1, 1980): 101–25.

“How Does Monetary Policy Affect the U.S. Economy?” *Federal Reserve Bank of San Francisco*. Accessed April 16th, 2015. http://www.frbsf.org/us-monetary-policy-introduction/real-interest-rates-economy/.

Howgego, Christopher. (2009) “Some Numismatic Approaches to Quantifying the Roman Economy.” In *Quantifying the Roman Economy: Methods and Problems*, 287–95. Oxford: Oxford University Press.

———. (1995) *Ancient History from Coins*. Approaching the Ancient World. London: Routledge, 1995.

———. (1994) “Coin Circulation and the Integration of the Roman Economy.” *The Journal of Roman Archaeology* 7 (1994): 5–21.

———. (1992) “The Supply and Use of Money in the Roman World 200 B. C. to A. D. 300.” *The Journal of Roman Studies* 82 (1992): 1.

———. (1990) “Why Did Ancient States Strike Coins?” *The Numismatic Chronicle (1966-)* 150 (January 1, 1990): 1–25.

Johnson, Stephen R. (2008) “The Trouble with QSAR (or How I Learned To Stop Worrying and Embrace Fallacy).” *Journal of Chemical Information and Modeling* 48, no. 1 (January 1, 2008): 25–26.

Kay, Philip. (2014) *Rome's Economic Revolution*. Oxford Studies on the Roman Economy. Oxford: Oxford University Press.

Lindbeck, Assar. (2015) "Government Economic Policy." *Encyclopedia Britannica*. Accessed April 16, 2015. http://www.britannica.com/EBchecked/topic/240167/government-economic-policy/26312/The-problem-of-time-lags.

Lo Cascio, Elio. (1993) "How Did the Romans View Their Coinage and Its Function?" In *Coin Finds and Coin Use in the Roman World: The Thirteenth Oxford Symposium on Coinage and Monetary History, 25.-27.3.1993, a NATO Advanced Research Workshop*, edited by Cathy E. King and David G. Wigg. Studien Zu Fundmünzen Der Antike, Bd. 10. Berlin: G. Mann.

———. (1981) "State and Coinage in the Late Republic and Early Empire." *The Journal of Roman Studies* 71 (January 1, 1981): 76–86.

Lockyear, Kris. (2013) "Coin Hoards of the Roman Republic Online." Version 1 (beta). Accessed November 2, 2014. http://numismatics.org/chrr/.

———. (2012) "Dating Coins, Dating with Coins." *Oxford Journal of Archaeology* 31, no. 2 (May 1, 2012): 191–211.

———. (2007) *Patterns and Process in Late Roman Republican Coin Hoards, 157-2 BC*. BAR International Series. Oxford: Archaeopress.

———. (1999) "Hoard Structure and Coin Production in Antiquity — an Empirical Investigation." *The Numismatic Chronicle (1966-)* 159 (January 1, 1999): 215–43.

Mankiw, N. Gregory. (2015) *Principles of Economics*. Eighth edition. Mason, OH: South Western Cengage Learning.

Mattingly, Harold. (1977) "Coinage and the Roman State." *The Numismatic Chronicle (1966-)*, Seventh Series, 17 (137) (January 1, 1977): 199–215.

Paolilli, Antonio Luigi. (2008) "Development and Crisis in Ancient Rome: The Role of Mediterranean Trade." *Historical Social Research / Historische Sozialforschung* 33, no. 4 (126) (January 1, 2008): 274–88.

Pense, Alan. (1992) "The Decline and Fall of the Roman *Denarius*." *Materials Characterization* 29 (1992): 213–22.

Pliny the Elder. *The Natural History*. Edited by Bill Thayer. Translated by H. Rackham. Loeb Classical Library, 2006. Accessed April 16, 2015. http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Pliny_the_Elder/home.html#tech_details.

Polybius. *The Histories of Polybius*. Edited by Bill Thayer. Translated by Heinemann. Loeb Classical Library, 2012. Accessed April 16, 2015. http://penelope.uchicago.edu/Thayer/E/Roman/Texts/Polybius/34*.html.

"Regression With Stata Chapter Two: Regression Diagnostics." *UCLA: Statistical Consulting Group*, 2015. Accessed April 16, 2015. http://www.ats.ucla.edu/stat/stata/webbooks/reg/chapter2/statareg2.htm.

Romer, David H., and Jeffrey A. Frankel. (1999) "Does Trade Cause Growth?" *American Economic Review* 89, no. 3 (1999): 379–99.

Roth, Jonathan. (1994) "The Size and Organization of the Roman Imperial Legion." *Historia* 43, no. 3 (1994): 346-362.

Sargent, Thomas J., and Bruce D. Smith. (1997) "Coinage, Debasements, and Gresham's Laws." *Economic Theory* 10, no. 2 (August 1, 1997): 197–226.

Strauss, J. (2013) "Shipwrecks Database." *The Oxford Roman Economy Project*. Accessed May 20th, 2014. http://oxrep.classics.ox.ac.uk/databases/shipwrecks_database/.

Temin, Peter. (2013) *The Roman Market Economy*. Princeton: Princeton University Press.

Torres-Reyna, Oscar. "Data Analysis Notes: Links and General Guidelines." *Princeton Data and Statistical Services*. Accessed April 18, 2015. http://www.princeton.edu/~otorres/Stata/statnotes.

Vivenza, Gloria (2012) "Roman Economic Thought." In *The Cambridge Companion to the Roman Economy*, edited by Walter Scheidel, 25–44. Cambridge: Cambridge University Press.

Von Reden, Sitta. (2012) "Money and Finance." In *The Cambridge Companion to the Roman Economy*, edited by Walter Scheidel, 266–86. Cambridge Companions to the Ancient World. Cambridge: Cambridge University Press.

Walker, D.R. (1980) "The Silver Contents of the Roman Republican Coinage." In *Metallurgy in Numismatics*, edited by D. M Metcalf and W. A Oddy. Special Publication (Royal Numismatic Society (Great Britain)) ; London: Royal Numismatic Society.

Wilson, Andrew. (2009) "Approaches to Quantifying Roman Trade." In *Quantifying the Roman Economy: Methods and Problems*, edited by Andrew Wilson and Alan Bowman, 213–49. Oxford: Oxford University Press.

Woytek, Bernhard. (2012) "The Denarius Coinage of the Roman Republic." In *The Oxford Handbook of Greek and Roman Coinage*, edited by William E. Metcalf, 315–34. Oxford ; New York: Oxford University Press.