*Journal of Roman Studies* 109 (2019): The army and the spread of Roman citizenship

*by* Myles Lavan

ONLINE SUPPLEMENTARY APPENDIX 3

APPENDIX 3A: DESCRIPTION OF THE MODEL

The model was constructed in Microsoft Excel 2010, and is available online as Supplementary Appendix 2. The main user interface is on the first two sheets: OVERVIEW and INPUTS. MODEL contains the mathematical model itself. The other sheets are used in running the Monte Carlo simulation.

*Specifying inputs.*

The INPUTS sheet allows the user to specify beliefs about the input variables (minimum, most likely and maximum values; the type of probability density function to be used; any interdependencies) by entering data into the yellow cells.

- To fix an input value, specify the Most Likely value only.

- Note that the model allows for more variables than those described in the published study. In the interests of economy and simplicity, several minor variables were represented as fixed assumptions there. These are the grey cells in this sheet, which only contain a Most Likely value. These can be made variables by specifying Minimum and Maximum values and a Distribution Type.

- Note that the model will return an error if a Triangle distribution is selected and the Most Likely value is identical to either the Minimum or Maximum. This is because the Triangle distribution requires Minimum < Most Likely < Maximum.

- For ease of reference, the top left cell (A1) in INPUTS calculates the Most Likely value of the output given the specified Most Likely values of the inputs. This makes it possible to use the INPUT sheet and its Most Likely column to test discrete scenarios.

*Running a Monte Carlo simulation.*

The OVERVIEW sheet allows the user to initiate a Monte Carlo Simulation and provides an overview of results.

- You will need to have enabled Macros to run a MC simulation.

- Click the 'Run MC Simulation' button in the OVERVIEW sheet. This will initiate a Visual Basic Macro. While the Macro is running, Excel will suspend screen updating in its window (to accelerate the process) and will be unresponsive until the Macro has completed. To

interrupt the Macro at any time, press ESCAPE and select 'End' in the dialogue box that appears. This may not work on a Mac, in which case you would need to terminate Excel manually to interrupt the macro.

- When the Macro starts, a dialogue box will ask you how many THOUSAND iterations you want to run. One or two is generally sufficient to estimate the mean and 95 per cent credible interval; five or ten will allow greater precision in estimating the latter, if required. Note that the macro only accepts integer inputs (i.e. whole numbers).

- A second dialogue box gives you the option (by replacing the 1 with 0) of only simulating the main output (the total number of new citizens). Otherwise the model will capture c. 150 intermediate values (such as those in Tables 3 and 4). I recommend substituting 0 if running more than around 10,000 iterations, as the cumulative size of the data captured will slow the calculation and produce a large file. To simulate all values, just click OK.

- Excel will now run the simulation for the specified number of iterations. It reports its progress (the number of iterations completed) in the bottom left of the window.

- The OVERVIEW sheet aggregates the key results of the Monte Carlo simulation, reproducing the data in Tables 3 and 4 and also the histogram in Fig. 8.

*The model*

The mathematical model itself is constructed in the MODEL sheet. To minimise inter-sheet references, the first sixteen rows are used to reproduce the input values (taken from the INPUTS sheet via MC INPUTS) for use on this sheet.

The lines below Row 22 contain sections for the auxilia, fleet and legions. Within each section various quantities are calculated for each year from 15 to 212 C.E. inclusive, with years arranged horizontally and calculations laid out vertically. Calculations are made step-by-step to facilitate navigation and error checking. Explanatory notes are provided in purple font and in comments (indicated by a red triangle in the top right hand corner of a cell; hover over the triangle to read the comment).

*Other sheets*

The Monte Carlo simulation works by generating a new set of random numbers between 0 and 1 in MC INPUTS column C. These random numbers are used to assign each variable a value from its specified probability distribution. These values appear in column I. The mathematical model in MODEL, which draws its inputs from MC INPUTS column I, updates accordingly. So does OUTPUT column D, which draws key data from MODEL. Its sole purpose is to extract any quantities of interest from MODEL and arrange them in a single column. The macro then copies the

contents of OUTPUT column D to the first available row in SIMULATIONS (this sheet is cleared at the beginning of each simulation). (The column is transposed into a row to facilitate large simulations, since an Excel worksheet has a limit of 16,000 columns but 1,000,000 rows) The macro then generates a new set of random numbers in MC INPUTS and repeats these steps until it completes the specified number of iterations. At this point SIMULATIONS will contain as many rows of data as there were iterations. MONTE CARLO ANALYSIS calculates various descriptive statistics for the output data (mean and various percentiles). The most important are presented more accessibly in OVERVIEW. HISTOGRAM organises the data for the histogram of results in OVERVIEW.

## APPENDIX 3B: MODELLING ATTRITION

*A static model of attrition across service cohorts*

Let $N$ be the total strength of a formation, $t$ be the term of service in years and $a_t$ be the percentage loss through attrition over the $t$ years of service, i.e. to discharge. In other words,

$$a_t = \frac{R - V}{R}$$

where $R$ is the number of recruits and $V$ the number of discharged veterans, and hence $R$-$V$ the number of men lost during service.

The strength of the formation is distributed across $t$ service cohorts, i.e. cohorts of men in the same year of service. Due to attrition, the service cohorts become progressively smaller from the cohort of new recruits down to that of men in their final year of service as illustrated in Fig. 6 in the main text (reproduced for convenience below).
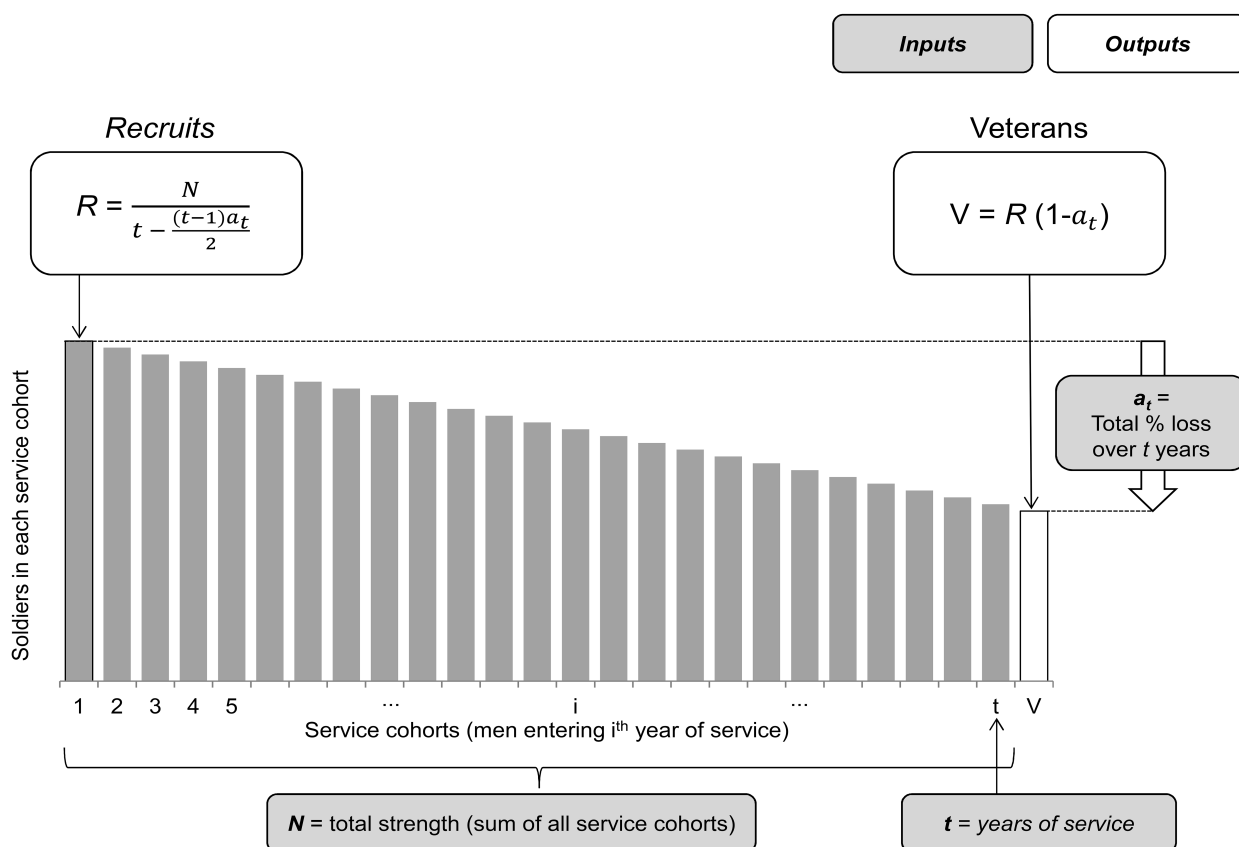


FIG. 6. A model of recruitment, attrition and discharge.

First, assume that soldiers are recruited and discharged at a fixed point each year. Once a year, a number of recruits ($R$) is required to restore the unit to strength $N$. They must replace both the veterans discharged at the end of service that year ($V$) and attrition over the past year (i.e. men who died or were discharged early).

Second, assume that attrition proceeds by fixed decrements. Because the mortality rate and risk of illness or injury leading to early discharge are likely to increase with age, attrition over the course of service is best approximated by assuming a fixed decrement (i.e. constant number of men lost each year) rather than a constant rate of loss (i.e percentage of men lost each year). On the rationale, see further Scheidel 1996: 123 n. 85.

Hence, the fixed decrement (i.e. the annual loss in each cohort) is:

$\frac{R.a_t}{t}$    *i.e. total loss over t years distributed evenly across t years*      [1]

Immediately after recruitment, the strengths of the individual service cohorts sum to strength $N$:

$$\sum_{i=1}^{t} c_i = N \qquad\qquad [2]$$

where $c_i$ is the number of men entering the $i^{\text{th}}$ year of service.

The size of the $i^{\text{th}}$ service cohort can be represented as a function of the number of recruits, $R$:

$$c_i = R - (i-1)\frac{R.a_t}{t} \qquad\qquad [3]$$

*i.e. the number of recruits less the fixed decrement (from [1]) repeated over i-1 years.*

Note that, immediately after recruitment, $c_1 = R$ and $c_{25}$ (the cohort of men entering the final year of service) is one decrement larger than V, the cohort of men just discharged.

It is now possible to derive the formulae in Fig. 6.

$N = c_1 + c_2 + c_3 + \cdots + c_t$      *from [2]*

$\Rightarrow N = (R - 0\frac{R.a_t}{t}) + (R - 1\frac{R.a_t}{t}) + (R - 2\frac{R.a_t}{t}) + \cdots + (R - (t-1)\frac{R.a_t}{t})$      *from [3]*

$\Rightarrow N = t.R - \left(1 + 2 + \cdots + (t-1)\right)\frac{R.a_t}{t}$

$\Rightarrow N = t.R - \frac{(t-1)t}{2}\frac{R.a_t}{t}$      *from the rule $\sum_{k=1}^{n} k = \frac{n(n+1)}{2}$*

$\Rightarrow N = t.R\left(1 - \frac{(t-1)a_t}{2t}\right)$

$\Rightarrow N = R\left(t - \frac{(t-1)a_t}{2}\right)$

$\Rightarrow \boldsymbol{R} = \frac{N}{t - \frac{(t-1)a_t}{2}}$      [4]

The number of veterans, *V*, can then be calculated from *R* and the rate of attrition over *t* years:

$$V = R(1 - a_t) \qquad\qquad\qquad\qquad [5]$$

Hence the number of recruits required to maintain a formation at strength *N* given *t* years of service and attrition $a_t$ can be calculated for any specified *N*, *t* and $a_t$

*Example*

Take for example, a simplified representation of the auxilia in the second century with N=200,000, t=25, $a_{25}$=50 per cent.

The model estimates the annual number of recruits required to maintain strength N at:

$$R = \frac{N}{t - \frac{(t-1)a_t}{2}} \qquad\qquad \textit{from [4]}$$

$$= \frac{200,000}{25 - \frac{(24)0.5}{2}}$$

$$= 10,526$$

The annual number of veterans, i.e. men who have survived twenty-five years from recruitment is estimated at:

$$V = R(1 - a_t) \qquad\qquad \textit{from [5]}$$

$$= 10,526(1 - 0.5)$$

$$= 5,263$$

*Note*

This model generalises the approach in Scheidel 1996: 121–7. Note that there is a minor computational error in Scheidel's calculations that works slightly to over-estimate attrition in all cases. His calculations appear to confuse the cohort of veterans with the cohort of men in their final year of service. Regarding his Table 3.14, for example, a legion with 25 service cohorts, 264 new recruits and 120 men starting their twenty-fifth year would indeed have a total strength of exactly 4,800 men (assuming linear attrition between service cohorts), but those 120 men would have to survive another year to be discharged. The corresponding number of veterans would be 114 not 120, implying 56.8 per cent attrition. (This can be calculated from [5] and [6] with N=4,800, t=25 and $a_t$=56.8 per cent.) To produce 120 veterans, recruits would have to fall to 259 and attrition drop to 53.6, not 54.5 per cent (N=4,800, t=25 and $a_t$=53.6 per cent).

*Another useful formula*

Rearranged, equation 4 makes it possible to infer the attrition rate from total strength and the number of recruits.

$$R = \frac{N}{t - \frac{(t-1)a_t}{2}}$$

$$\Rightarrow t - \frac{(t-1)a_t}{2} = \frac{N}{R}$$

$$\Rightarrow a_t = \frac{2(t - \frac{N}{R})}{t-1}$$

This formula underlies the calculation of implied attrition rates in footnote 171.

*Extrapolating to other attrition rates*

The overall model works with a single assumption about attrition, the attrition rate over twenty-five years, i.e $a_{25}$. This is appropriate for the auxilia and legions in the second century, but other attrition rates are needed for the fleets and the auxilia in the first century. Again assuming fixed-decrement attrition, the model scales the twenty-five-year rate up or down proportionately:

The attrition over x years, $a_x = \frac{x}{25} a_{25}$

This is a crude simplification, but it has very limited effect on the result since most relevant periods are close to twenty-five years.

Scheidel, W. 1996: *Measuring Sex, Age and Death in the Roman Empire: Explorations in Ancient Demography*, Journal of Roman Archaeology Supplementary Series 21, Ann Arbor.