Multiple Systems Estimation for Hard-to-count Populations: Some Recent Developments

BERNARD W. SILVERMAN

UNIVERSITIES OF NOTTINGHAM AND OXFORD, U.K.

WWW.BERNARDSILVERMAN.CO.UK

Background: Modern Slavery

AN EXAMPLE OF A HARD-TO-REACH POPULATION

Modern Slavery

- Modern slavery encompasses many aspects of human trafficking and exploitation. Tens of millions of victims world-wide. Examples:
  - Individuals held in servitude within their own family or community
  - Debt bondage
  - Domestic servitude
  - Labour exploitation in mining, agriculture, fisheries, car washes, nail bars, cannabis farms, paving businesses
  - Sexual exploitation and forced prostitution
- No country is free from Modern Slavery.
- NOTE: The terms “Modern Slavery” and “Human Trafficking” are used synonymously: they are very closely related.

How many victims in the United Kingdom?

- From 2010 to 2017 I was Chief Scientific Advisor to the UK Home Office (Interior Ministry—responsible for issues to do with crime, counter-terrorism, borders, migration)
- The then Home Secretary Theresa May introduced a law against Modern Slavery and asked me to quantify the likely number of victims in the UK
- A government agency collates data from a range of sources to produce a "strategic assessment" of known victims: police forces, various government departments and agencies, local authorities, non-governmental organisations (NGOs) and many charitable and voluntary expert groups. In 2013 this identified 2744 unique victims.

The dark figure

- Despite all efforts, the known cases present a partial picture of the size of human trafficking. This is because:
  - Human trafficking is a hidden crime and some victims may be controlled or still in servitude
  - Victims may not come forward due to feelings of fear and shame.
  - Some individuals may not be identified as victims of human trafficking by professionals who encounter them.
  - Some victims may not view themselves as victims of exploitation.
  - Coverage is not complete: not all agencies respond to the call for information.
- **Multiple systems estimation (MSE)** can be used to estimate the "dark figure" of potential victims that do not come to attention.

Overlaps between lists in the Strategic Assessment

LA	Local Authorities	X					X	X	X								X	X	X	
NG	NGOs, charities		X				X			X	X	X				X	X	X	X	
PF	Police Forces			X				X		X			X	X		X	X		X	
GO	Government e.g. Border				X				X		X		X		X	X		X	X	
GP	General Public					X						X		X	X					
number		54	463	995	695	316	15	19	3	62	19	1	76	11	8	4	1	1	1	???

This table required weeks of work, going through thousands of individual confidential files. The “dark figure” would be the number in the last column—those which do not appear on any of the observed lists.

Results and comments

- Multiple systems estimation (an extension of classical mark-recapture) fits a statistical model to this table.
- The estimated confidence interval for the actual population size was from 10K to 13K, so this suggests that the Strategic Assessment was aware of 20% to 25% of all the potential victims of trafficking in the UK in 2013.
- This must be regarded as a tentative conclusion, because the model is based on assumptions that (while sensible) can't be easily verified, and the data have their own limitations.
- Nevertheless, for policy purposes even ball-park estimates are very important.
- Because MSE is used in many hard-to-count contexts, including the UK Census, it is important to develop its methodology.

UK is home to 13,000 slaves: Home Office says number is four times higher than previously thought

- The shocking figures include women trafficked for prostitution or service
- Others are brought in to work in factories, fields or even fishing boats
- The Home Office confirmed it underestimated the scale of the problem
- The most common victims are from Romania, Poland, Albania and Nigeria

By MAIL ON SUNDAY REPORTER

PUBLISHED: 01:55, 30 November 2014 | **UPDATED:** 15:56, 30 November 2014

Share 189 shares 442 View comments

Up to 13,000 people in Britain are being held in conditions of slavery, four times the number previously thought, the Home Office has said.

In what is said to be the first scientific estimate of the scale of modern slavery in the UK, the Home Office has said the number of victims last year was between 10,000 and 13,000.

They include women forced into prostitution, domestic staff and workers in fields, factories and fishing boats.

Scroll down for video

© EPA

Home Secretary Theresa May, pictured, described the scale of human trafficking as 'shocking'

Data from the National Crime Agency's Human Trafficking Centre had previously put the number of slavery victims in 2013 at 2,744.

The new estimate is based on a statistical analysis by the Home Office chief scientific adviser, Professor Bernard Silverman, which aims for the first time to calculate the 'dark figure' of victim numbers who are not reported to the law enforcement agencies.

'Modern slavery is very often deeply hidden and so it is a great challenge to assess its scale,' he said.

'The data collected is inevitably incomplete and, in addition, has to be very carefully handled because of its sensitivity.'

Other approaches

Vulnerability models

- The 2018 Global Slavery Index (GSI) predicts individual country estimates for 167 countries using a vulnerability model
- Gallup World Polls gives estimates for 48 countries
- Regressing these on a number of vulnerability measures gives a prediction model then applied to all the countries in the world
- Some adjustments were then made, for example to average the predictions with multiple systems estimates
- None of the 48 countries are in North America, Western Europe or richer parts of Asia
- Even if you are happy with the idea of the extrapolation, the prediction intervals are extremely wide (e.g. from less than nothing to 4 million for the USA)
- The approach is arguably better at identifying risks and vulnerabilities than it is for estimating prevalence

Analysis of police documents

- Work with one UK police force (West Midlands); just published by Centre for Social Justice (London).
- Use machine learning/natural language processing (NLP) to analyse lengthy police documents
  1. Modern slavery events already tagged as such by the police, **312** individual victims identified
  2. Crime reports not so tagged, but where NLP identifies modern slavery elements, **374** victims identified, not individually
  3. Intelligence logs not tagged, but where NLP identifies modern slavery elements; numbers of individuals, but cannot distinguish between victims and suspects/offenders; estimate of actual victims **3511**.
- Scaling up to the whole UK by population gives an estimate of about 100,000 victims
- The victims identified through route 1 are the only ones likely to get on to the strategic assessment hitherto described; figures suggest that despite increased awareness, busy humans often miss signs of modern slavery. Also especially through route 3 there may be cases which are individually ambiguous but indicate a general underlying prevalence level.
- Many areas for future thought (e.g. confidence intervals, de-duplication, appropriateness of scale-up) but this is a fascinating "bottom-up" approach

Multiple systems estimation: the model

SOME RECENT WORK

MSE: The Poisson log-linear model

- Suppose there are t lists. A *capture history* is a collection $\boldsymbol{\omega}$ of lists, in other words a subset of $\{1, 2, \dots, t\}$. Define the *order* of $\boldsymbol{\omega}$ to be the number of elements in $\boldsymbol{\omega}$, the number of lists in the capture history. Capture histories appear twice in our formulation, as <u>capture counts</u>, and as <u>parameters in a model</u>.
- **Capture counts**: Let $N_{\boldsymbol{\omega}}$ be the number of cases that appear on the lists in $\boldsymbol{\omega}$ and no others. Our data are the $(2^t - 1)$ observed values $\{N_{\boldsymbol{\omega}}\colon \boldsymbol{\omega} \neq \emptyset\}$.
- **Parameters in a model**: We define parameters α_θ for all capture histories $\boldsymbol{\theta}$ in some set $\boldsymbol{\Theta}$ of all possible capture histories. We then model

$$N_{\boldsymbol{\omega}} \sim \mathrm{Poiss}\left[\exp\left(\sum_{\boldsymbol{\theta} \subseteq \boldsymbol{\omega}, \boldsymbol{\theta} \in \boldsymbol{\Theta}} \alpha_{\boldsymbol{\theta}}\right)\right]$$

MSE: The Poisson log-linear model

- For example, the number of cases on lists 3 and 4 but no others has expectation $\exp(\alpha_\emptyset + \alpha_3 +$

Choosing the model

Several approaches to model choice have been used. We concentrate on two:

- **BIC**: Choose among all hierarchical models using the BIC (Bayes information criterion).
- **Forward stepwise**: Start with the model containing just zeroth and first order parameters, and then add parameters stepwise. At each stage, choose the parameter with lowest p-value, and stop when that p-value reaches some threshold.
  - Originally restricted to effects of order 2 but now extended to allow effects of any order.
  - It is easy to constrain the search to consider only effects which preserve the hierarchical nature of the model.
- The standard software does not check for existence or identifiability of the estimates nor for fitting where some of the estimated parameters should be $-\infty$. The asymptotic justification of information theoretic methods won't hold, but they are used nevertheless.

Stepwise approach: more details

- A natural approach is to test the significance of any particular parameter α_{ω} by using the corresponding sufficient statistic $N_{\boldsymbol{\omega}}^{*}$ and assess this against a Poisson distribution with mean estimated using the current model with α_{ω} set to zero (hence omitted from the fit). This works equally whether or not $N_{\boldsymbol{\omega}}^{*}$ is zero
- Start with main effects then successively add capture histories to the model, searching over those which preserve the hierarchical structure, so that a capture history is only considered if all its sub-histories are already in the model.
- Among these, choose the effect α_{ω} which has the lowest p-value, but testing as well whether adding that effect passes the tests for existence and identifiability.
- Stop when no effect has p-value less than some specified threshold, e.g. 0.02; testing for sensitivity of the threshold is a good idea.

Bootstrapping to take account of model choice

Bootstrapping

- The standard approach is to choose the model and to find a confidence interval conditional on that choice. This ignores any possible error in the model choice stage.
- In order to take account of model selection, we adopt a bootstrap approach, for both the methods considered.
- The BC_a methodology of DiCiccio and Efron (1996) gives second-order accuracy and does not depend on any transformation of the scale on which the data are observed and the estimate of the total population made.

Constructing bootstrap replications

- Let M be the number of capture histories observed in the original data ($M = 2744$ for the UK data.) Typically we construct 1000 bootstrap replications. Each one is a sample of size M drawn with replacement from the observed capture histories.
- Equivalently, construct the capture counts $\{N_{\boldsymbol{\omega}}^{\text{boot}}: \boldsymbol{\omega} \neq \emptyset\}$ from a multinomial distribution with number of trials M, and probabilities proportional to $N_{\boldsymbol{\omega}}$.
- For each bootstrap sample, construct an estimate $\widehat{\mathbf{N}}_i^{\text{boot}}$ of total population size. We then use quantiles of $\{\widehat{\mathbf{N}}_i^{\text{boot}}\}$ to define the bootstrap confidence intervals. The BC_a approach corrects the levels of the quantiles by using jackknife estimates.
- Usually a jackknife involves leaving out each of the original data points in turn (2744 in the case of the UK data) but a trick makes this much quicker because the number of *distinct* values taken by the original data is much less, 18 for the UK data.

Recap of BC$_a$ and the jackknife trick

- Define the *bias-correction* $\hat{z}_0$ such that $\Phi(\hat{z}_0)$ is the proportion of the bootstrap replication estimates $\{\widehat{\mathbf{N}}_i^{\text{boot}}\}$ less than the original $\widehat{\mathbf{N}}$.

- For $N_{\boldsymbol{\omega}} > 0$, let $\widehat{\mathbf{N}}_{\boldsymbol{\omega}}^{(-1)}$ be the estimate of the population size from the original sample but with $N_{\boldsymbol{\omega}}$ replaced by $N_{\boldsymbol{\omega}} - 1$. The average of all the jackknife estimates is given by $\widehat{\Pi}_{(.)} = M^{-1} \sum_{\omega} N_{\boldsymbol{\omega}} \widehat{\mathbf{N}}_{\boldsymbol{\omega}}^{(-\mathbf{1})}$. Then define the *acceleration*

$$\hat{a} = \frac{1}{6}\left\{\sum_{\omega} N_{\boldsymbol{\omega}} \left(\widehat{\mathbf{N}}_{(.)} - \widehat{\mathbf{N}}_{\boldsymbol{\omega}}^{(-\mathbf{1})}\right)^{\mathbf{3}}\right\} \times \left\{\sum_{\omega} N_{\boldsymbol{\omega}} \left(\widehat{\mathbf{N}}_{(.)} - \widehat{\mathbf{N}}_{\boldsymbol{\omega}}^{(-\mathbf{1})}\right)^{\mathbf{2}}\right\}^{-3/2}$$

- The calculations of $\widehat{\mathbf{N}}_{(.)}$ and $\hat{a}$ require at most $2^t - 1$ estimates $\widehat{\mathbf{N}}_{\boldsymbol{\omega}}^{(-\mathbf{1})}$.

- To find the upper end-point of a one-sided β-confidence interval, say, use the $\tilde{\beta}$ quantile of the $\{\widehat{\mathbf{N}}_i^{\text{boot}}\}$, where

$$\Phi^{-1}(\tilde{\beta}) = \hat{z}_0 + \{\hat{z}_0 + \Phi^{-1}(\beta)\}/[1 - \hat{a}\{\hat{z}_0 + \Phi^{-1}(\beta)\}].$$

Stepwise method

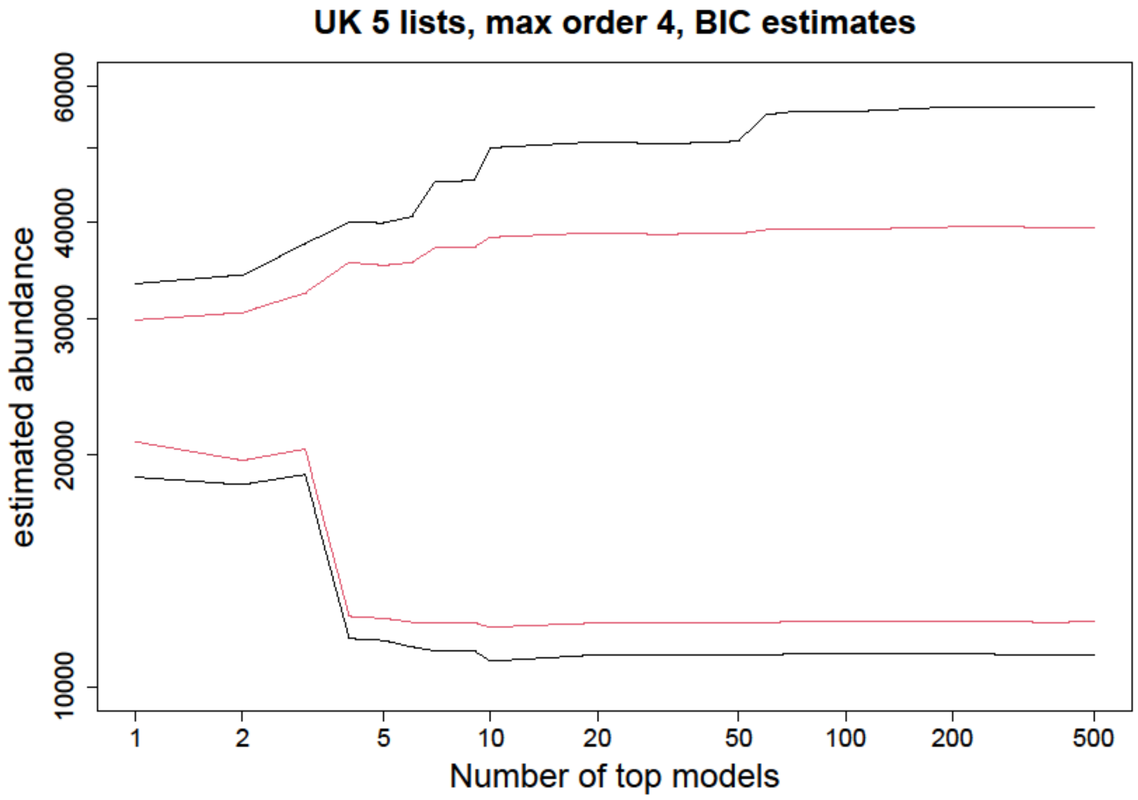
- The model fitted to the UK data by the stepwise method includes the parameters indexed by the pairwise histories 12, 13, 24, 25, 35 and 45, and no higher order effects
- The point estimate is 11313
- The standard approach, profile likelihood conditional on the chosen model, gives 95% CI (9900, 13100). This is the value that was reported for the UK data.
- The bootstrap approach taking account of the model selection gives (8800, 14700). So in round terms that would be 9 to 15 thousand rather than the 10 to 13 thousand that was reported.

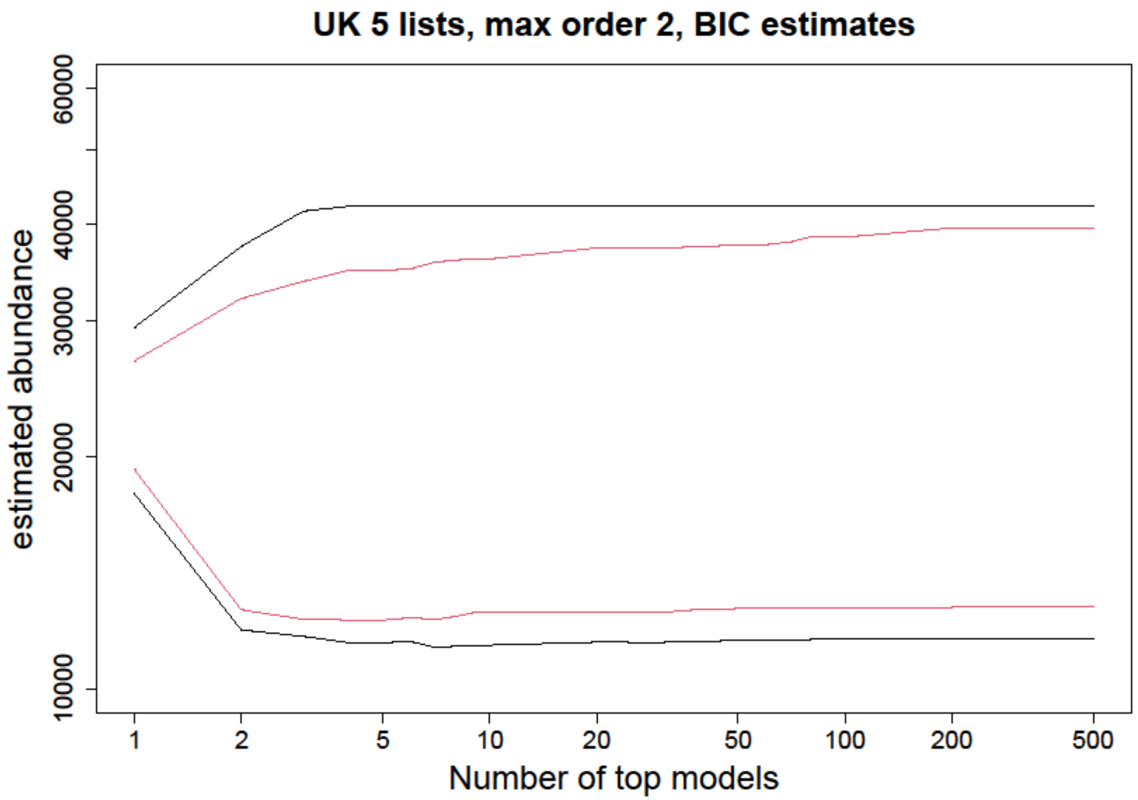
Bootstrapping the BIC method?

- The model choice and estimation will have to be carried out for each bootstrap replication (the bootstrap step) plus the number of non-zero $N_{\boldsymbol{\omega}}$ (the jackknife step) plus, of course, the original data.
- For five lists and 1000 bootstrap replications, this means carrying out the full estimation procedure up to 1032 times. (Less the number of $N_{\boldsymbol{\omega}}$ that are zero.)
- Each estimation requires an estimate for every possible hierarchical model. For five lists, there are 6893 models if all hierarchical models are considered. So to carry out the full bootstrap procedure would take the fitting of the order of 7 million generalised linear models. Can this be speeded up?

Considering top models only

- Carry out the BIC procedure on the original data finding the BIC for every hierarchical model up to a given maximum order.
- Choose a suitable number N_{top} and let $\mathcal{M}$ be the set of the N_{top} best models.
- For each bootstrap replication, find the population estimate and BIC only for the models in $\mathcal{M}$ and minimise the BIC over these models, rather than over all possible models. The rationale is that the best BIC model for a bootstrap replication is likely to be among the best few models for the original data.
- This approach can speed things up enormously. For example for five list data there are 6893 models, so putting $N_{\text{top}} = 100$ speeds up computation (for both the bootstrap and jackknife replications) by a factor of nearly 70.
- This also facilitates simulations to assess coverage probabilities: in progress.





Confidence intervals (80% and 95%) for different numbers of top models in the BIC approach, allowing all models and only those up to order 2 (pairwise effects).

- The intervals for 1 top model correspond to conditioning on the chosen model, the current standard approach.
- Considering even 20 top models goes nearly all the way to accounting for the model choice step in the BIC method.

Sparsity, existence and identifiability

Sparsity

- Data collected in the context of modern slavery and human trafficking typically have the property that some or many of the observed counts are zero. In general, for any capture history, define $N^*_{\boldsymbol{\omega}} = \sum_{\boldsymbol{\psi} \supseteq \boldsymbol{\omega}, \boldsymbol{\psi} \neq \emptyset} N_{\boldsymbol{\psi}}$. Then $N^*_{\boldsymbol{\omega}}$ is the number of cases which appear on all the lists in $\boldsymbol{\omega}$ and possibly some others. For example, if there are four lists, then $N^*_{12} = N_{12} + N_{123} + N_{124} + N_{1234}$.
- Call $\boldsymbol{\omega}$ a *non-overlapping* capture history if $N^*_{\boldsymbol{\omega}} = 0$. The reason is that no cases fall in the intersection of the lists in $\boldsymbol{\omega}$.
- If a non-overlapping history $\boldsymbol{\omega}$ is in the model $\boldsymbol{\Theta}$ then the maximum likelihood estimate of $\alpha_{\boldsymbol{\omega}}$ is $-\infty$, and any higher-order parameters including all the lists in $\boldsymbol{\omega}$ can't be estimated and don't affect the distribution of the data anyway. If you use the standard packages an error will be generated. A better approach is to record the maximum likelihood estimate of $\alpha_{\boldsymbol{\omega}}$ as $-\infty$ and then to fit the remaining parameters by the standard glm approach but leaving out all the data points $N_{\boldsymbol{\psi}}$ for $\boldsymbol{\psi} \supseteq \boldsymbol{\omega}$. See Chan *et al.*, (2020). This is called an **extended maximum likelihood estimator**, allowing parameters to be in $[-\infty, \infty)$.

Existence and identifiability of the estimates

- Let $\Theta^{\dagger}$ be the parameter set excluding non-overlapping histories
- For each non-overlapping capture history $\boldsymbol{\theta}$ in the model Θ, remove $\boldsymbol{\theta}$ and all capture histories containing $\boldsymbol{\theta}$ from the set of observable capture histories Ω. Let $\Omega^{\dagger}$ be the set of capture histories remaining.
- Let $\boldsymbol{A}$ be the incidence matrix that maps the parameters in $\Theta^{\dagger}$ to the logarithm of the expected values of the counts of capture histories in $\Omega^{\dagger}$, so that $\boldsymbol{A}_{\boldsymbol{\omega\theta}} = 1$ if $\boldsymbol{\theta} \subseteq \boldsymbol{\omega}$ and 0 otherwise. Let $\mathbf{t}$ be the vector of sufficient statistics $\left(N^{*}_{\boldsymbol{\theta}} \middle| \boldsymbol{\theta} \in \Theta^{\dagger}\right)$.
- Solve the linear programming problem of maximising *s* over all *s* and all real vectors $\boldsymbol{x} = \left(x_{\boldsymbol{\omega}} \middle| \boldsymbol{\omega} \in \Omega^{\dagger}\right)$ satisfying the constraints $\boldsymbol{A}^{\mathrm{T}}\boldsymbol{x} = \mathbf{t}$ and $x_{\boldsymbol{\omega}} \geq s \; \forall \; \boldsymbol{\omega} \in \Omega^{\dagger}$. Let s^{*}be the maximising value.
- Then we require $\boldsymbol{A}$ to be of full column rank and $s^{*} > 0$. Unless both these conditions are satisfied the estimate of the population size may be infinite or not specified.
- See Fienberg and Rinaldo (2012, supplementary material), and Chan, Silverman & Vincent (2020).

Conclusion

- This is an area where innovative statistics are really important for public policy
- Unfortunately the methodology is not really standard yet and not surprisingly it's getting more diverse!
- There may be different tools to use, but it's important to hone each one as much as we can
- If you choose from a range of models, you need to take that into account when assessing the accuracy of estimation
- Even with modern computing power, it really pays to think carefully how to do that

References

THANK YOU!!

- Ongoing work, with Lax Chan and Kyle Vincent, including software package building on R package **SparseMSE**

- Lax Chan, Bernard W. Silverman & Kyle Vincent. (2020) Multiple Systems Estimation for Sparse Capture Data: Inferential Challenges when there are Non-Overlapping Lists. *Journal of the American Statistical Association*. DOI: 10.1080/01621459.2019.1708748

- Bernard W. Silverman. (2020) Multiple Systems Analysis for the quantification of Modern Slavery: Classical and Bayesian approaches. (with Discussion) *J. Roy. Stat. Soc. Ser A*. 183, 691-736. DOI: 10.1111/rssa.12505