Humanities & Social Sciences Communications

ARTICLE

https://doi.org/10.1057/s41599-021-00938-z OPEN

Machine learning methods for "wicked" problems: exploring the complex drivers of modern slavery

Rosa Lavelle-Hill[1,2]✉, Gavin Smith[3], Anjali Mazumder[1], Todd Landman[4] & James Goulding[3]

Forty million people are estimated to be in some form of modern slavery across the globe. Understanding the factors that make any particular individual or geographical region vulnerable to such abuse is essential for the development of effective interventions and policy. Efforts to isolate and assess the importance of individual drivers statistically are impeded by two key challenges: data scarcity and high dimensionality, typical of many "wicked problems". The hidden nature of modern slavery restricts available data points; and the large number of candidate variables that are potentially predictive of slavery inflate the feature space exponentially. The result is a "small *n*, large *p*" setting, where overfitting and significant intercorrelation of explanatory variables can render more traditional statistical approaches problematic. Recent advances in non-parametric computational methods, however, offer scope to overcome such challenges and better capture the complex nature of modern slavery. We present an approach that combines non-linear machine-learning models and strict cross-validation methods with novel variable importance techniques, emphasising the importance of stability of model explanations via a Rashomon-set analysis. This approach is used to model the prevalence of slavery in 48 countries, with results bringing to light the importance of new predictive factors—such as a country's capacity to protect the physical security of women, which has been previously under-emphasised in quantitative models. Further analyses uncover that women are particularly vulnerable to exploitation in areas where there is poor access to resources. Our model was then leveraged to produce new out-of-sample estimates of slavery prevalence for countries where no survey data currently exists.

[1] The Alan Turing Institute, London, UK. [2] University of Tübingen, Tübingen, Germany. [3] N/LAB, University of Nottingham, Nottingham, UK. [4] Rights Lab, University of Nottingham, Nottingham, UK. ✉email: rosa.lavelle-hill@uni-tuebingen.de

Introduction

Slavery has existed in the world for over 4000 years, with the most common reference point being the period of transatlantic chattel slavery from the early 1600s to the late 1800s. The formal abolition of slavery was followed by the 1926 Slavery Convention, which is the first international legal instrument that defined and sought to prevent the practice of slavery thereafter. The twentieth century saw further effort to eradicate slavery formally as the issue became incorporated in a range of international human rights instruments, International Labour Organisation (ILO) laws, and domestic legislation. These efforts have eliminated many forms of slavery that involve the direct appeal to ownership of the other and the invocation of property rights claims by slave holders; however, since the 1990s, an increasing number of state and non-state actors have recognised that forms of slavery persist. These new forms of slavery, sometimes termed *modern slavery*, do not make direct appeals to ownership, but do involve the intentional denial of agency through the use of force, threat, and coercion such that enslaved peoples are treated "as if" they are property.

Today, modern slavery is a phenomena that continues to effect men, women and children in all countries in the world (ILO, 2017) and has been formally articulated by the United Nations as part of the Sustainable Development Goals (SDGs), in particular SDG 8.7, which commits states to take effective measures to end modern slavery by 2030. Recent estimates have put the number of enslaved people globally at over 40 million (WFF, 2018a), translating to 5.4 victims from every 1000 people world-wide. Modern slavery is highly challenging to identify and measure. Although several definitions of modern day slavery exist, here we use the definition set out in the Bellagio-Harvard Guidelines on the Legal Parameters of Slavery: "any situation of exploitation that a person cannot refuse or leave because of threats, violence, coercion, deception, and/or an abuse of power" (WFF, 2018a). This definition encapsulates a wide variety of exploitative and coercive practices such as forced labour, debt bondage, forced marriage, sexual exploitation, bonded labour, and human trafficking (ILO, 2017). These types of slavery can be broken down yet further, and alternative valid typologies exist (Cooper et al., 2017). Yet recent trends, perhaps driven by the quantification of the United Nations' SDGs, have increasingly considered this unifying umbrella definition of modern slavery, encapsulating as it does a wide range of exploitation—from coercive recruitment and grooming in the adult entertainment industry to bonded agricultural labour and state-sanctioned work camps.

Often perceived to be a Low-to-Middle-Income Country (LMIC) or poorer community issue, such an encapsulating definition has highlighted and demonstrated that no country is unscathed of exploitative and coercive practices. It has also turned the task of understanding the causes of modern slavery, encompassing many types of practices, into a "wicked problem" (Rittel and Webber, 1973; Pryshlakivsky and Searcy, 2013) with many potential predictors. Yet, if we are to address the global slavery challenge and meet SDG 8.7 (United Nations, 2021), it is crucial to not only find new ways to measure prevalence (WFF, 2018a; Bales et al., 2015; Larsen and Durgana, 2017), but to understand better the conditions, which allow its continued perpetuation. Thus, we must try to isolate and address the core, driving factors that are leaving particular individuals, regions, or countries at risk.

Estimating slavery prevalence. Estimating the prevalence of slavery remains a central part in helping to prevent it. National and regional estimates not only highlight the extent of the issue, but serve as dependent variables in analyses that attempt to model, and hence find explanations for, slavery's underlying root causes (Landman, 2020). However, the hidden nature of modern-slavery makes it intrinsically difficult to measure (Chan et al., 2020) and there remains much uncertainty around the *true* number of people enslaved (Silverman, 2018). In recent years, the WFF has made valuable progress in this area, providing estimates of slavery incidence across 48 countries in 2016 and 2018, based upon surveys from the Gallup World Poll (GWP). These estimates have been further extrapolated out-of-sample to countries where no GWP survey data existed using a theoretically driven risk or "Vulnerability Model" (WFF, 2018a; Diego-Rosell and Joudo Larsen, 2018). While this has proven a highly beneficial exercise (Gleason, 2019; Cockayne et al., 2019), the methods used to produce such estimates are not without their limitations (Silverman, 2018; Gleason, 2019; Cockayne et al., 2019; Datta et al., 2018; Gallagher, 2017; Guth et al., 2014; Weitzer, 2014).

A core criticism of the GSI prevalence estimates is the assumption that a country's vulnerability to modern slavery is equivalent to the actual prevalence of slavery within a country (Silverman, 2018). Owing to the hidden nature of modern slavery practices, and a lack of data in the field, this assumption is hard to validate. Moreover, despite recent advances in Multiple Systems Estimation (MSE) (Bales et al., 2015; Cruyff et al., 2017; Bales et al., 2020; Silverman, 2019), assessment of national slavery prevalence to any degree of accuracy remains an active research area. Other criticisms of the GSI include the poor geographic representativeness of the data (very few western countries have been surveyed) (Gleason, 2019), the loose or changing definition of modern slavery (Gallagher, 2017; Datta and Bales, 2013), the uncertainty of the prevalence estimates not being highlighted (Silverman, 2018), and a lack of transparency around the methodology and data (Silverman, 2018; Gallagher, 2017). Despite these documented short-commings, the GSI prevalence estimates remain the most ambitious and indeed only attempt thus far to estimate the national slavery prevalence in a large number of countries using survey data.

Identifying drivers of modern slavery. An issue with modelling a global problem like modern slavery (particularly when using the all-encompassing definition) is that there exists an extensive range of candidate independent variables to investigate, while at the same time a shortage of prevalence estimates to regress against (due to its hidden nature and aforementioned difficulties with measurement). Therefore, aggregated data used to model slavery prevalence in a population will likely be "small n, large p" in nature (Johnstone and Titterington, 2009). This, not unexpected, issue is symptomatic of many "wicked problems" (Head and Alford, 2015) facing computational social sciences, with few observational units, n, available to researchers in comparison to the vast number of potential driving factors, p, to be modelled. In such situations, dangers of overfitting and significant correlation between predictors can render traditional statistical approaches problematic—and as a result, investigation of the factors underlying modern-slavery remains a challenging statistical task.

Owing to the "small n, large p" problem, prior research exploring the factors underlying slavery has predominantly been theoretically driven, with social science literature reporting individuals most vulnerable to exploitation as being: economic migrants; political asylum seekers; illiterate (Fitzgibbon, 2003) orphaned children (Rau, 2002); homeless people (ILO, 2001); and the jobless (Diego-Rosell and Joudo Larsen, 2018). Such categories pertain to those who either have poor well-being

(Diego-Rosell and Joudo Larsen, 2018) or acute poverty and/or significant debt, and hence seek a better way of life (Fitzgibbon, 2003). Children, estimated to make up a quarter of all victims (ILO, 2017), have also been identified as being particularly vulnerable to coercion through fake promises of work, food, or "western" lifestyles (Manzo, 2005). It is well understood that "trafficking thrives better on willingness", with traffickers targeting areas where individuals are most vulnerable, desperately poor and with few options available (Manzo, 2005).

At a national level, higher slavery incidences have been linked to both low GDP and higher levels of corruption (Bales, 2006). Yet poverty neither equates to slavery, nor makes it inevitable, and there are many poor areas where slavery is rare (Manokha, 2004). However, poverty can leave people in desperate circumstances, and is a well-established factor in making individuals easier for traffickers to coerce (Adesina, 2014). Slavery and human trafficking in turn undermines local and national economies, lowering GDP further (Datta and Bales, 2013), making cause and effect harder to discern. The impact of armed conflict on slavery is also well reported in the literature, with its effects being both direct and indirect. Men, women, and children have been abducted and forced to serve as soldiers, porters or smugglers of looted goods, forced labourers in military camps, and sex slaves for militia officers (Fitzgibbon, 2003; Van de Glind and Kooijmans, 2008). It is estimated that over 120,000 children have been used in armed conflicts in Africa alone (Van de Glind and Kooijmans, 2008), with battle deaths also leaving children orphaned and vulnerable (Fitzgibbon, 2003). Destruction of infrastructure and societal systems leaves populations isolated and unprotected, making conflict-torn areas easy targets, where criminal activity and criminal networks are rarely investigated.

At higher regional levels, areas such as Africa and Asia have been linked to higher prevalence of slavery (WFF, 2018a). This is thought to be caused by a number of factors including: geographic position with regards to trade routes and smuggling channels (Manzo, 2005); occurrence of natural disasters destroying livelihoods and displacing populations; diseases such as HIV and AIDS increasing the number of orphaned children (Rau, 2002; Roser and Ritchie, 2018); and desertification and rising sea levels causing famine (Brown et al., 2021). The sheer range of factors emphasises the complexity of slavery as a social problem, stemming from the multitude of possible causes interacting with one another.

Quantifying the effects of the drivers of modern slavery. Despite a rich literature considering the drivers of modern slavery, most findings stem from qualitative case data, victim/survivor interviews and small scale surveys. While a highly valuable depth of insight is obtainable from such methods, findings are necessarily drawn from small samples, limiting generalisability to specific regions, industries, and forms of slavery. Few of the predictors hypothesised, and detailed in the previous section, have been studied statistically, restricting our understanding of their impacts and interactions. As a consequence, not enough is known about the extent to which any individual driver engenders slavery incidence; nor how factors combine to allow the phenomena to perpetuate, limiting the formation of informed national policy.

There are a few recent exceptions; researchers have used national prevalence estimates from GWP surveys (WFF, 2018a) to explore relationships between slavery and specific phenomena statistically, such as fishing (Tickler et al., 2018) and globalisation (Landman and Silverman, 2019). The WFF have also sought to establish a theory-driven vulnerability model (WFF, 2018b), to summarise factors impacting national prevalence. In that framework, the WFF uses a combination of factor analysis and theory to group 23 national-level risk variables into five major "dimensions": Governance Issues; Lack of Basic Needs; Effects of Conflict; Inequality; and Disenfranchised Groups (WFF, 2018b), and allows a vulnerability score to be formulated for each country using its scoring on each dimension. However, the GSI Vulnerability Model has some notable limitations. The extent to which each individual factor predicts slavery prevalence is not reported—nor is there indication of how factors relate to one another. Importantly, the framework also currently drops variables when correlations occur between them (e.g., the Gender Inequality Index is removed (WFF, 2018b)); and the framework's combined vulnerability score (which considers all 5 dimensions together) shows only moderate correlation with country prevalence estimates ($r = 0.33$) (Diego-Rosell and Joudo Larsen, 2018).

Other quantitative studies have predominantly restricted themselves to linear hypothetico-deductive approaches (Tickler et al., 2018; Landman and Silverman, 2019), focusing on single or small sets of independent variables. While this is an understandable situation given limited data, it has left an open challenge of assessing the problem domain via an inductive and computational approach. Exploration of slavery drivers requires machinery that can accommodate the high-dimensional and non-linear nature of modern slavery, while also accounting for interactions and collinearities between explanatory variables. In this work, we extend WFF's valuable ground-work in the field, introducing a computational model able to directly map relationships between vulnerability indicators and national slavery prevalence. Our explicit target is to quantify the importance of individual factors in predicting slavery prevalence, in spite of the "small n, large p" context. Central to the approach is the first application of Rashomon-set analysis (Dong and Rudin, 2020) to the issue, a technique that allows assessment of the stability of model explanations and interactions across different individual variables.

Overcoming the limitations of traditional approaches. Studies capitalising on the national prevalence estimates represent an important step forward in quantifying and understanding the statistical relationships between variables that may affect slavery. Yet, the insights that can be gained from the linear regression/correlation analyses predominantly used thus far in the literature are constrained by three key factors to overcome: issues of (1) significant correlation between predictors; (2) ungrounded assumptions of linearity and (3) overfitting. These challenges respectively have direct impact on assessing the importance of explanatory variables, overall explanatory-power of different models, and the generalisability of resulting models.

Significant correlation of predictors. In a traditional regression model predictors should ideally be independent if β coefficients are to be used as reliable estimates of variable importance (Nathans et al., 2012). Owing to such inter-correlations, the WFF, for example, excludes several variables from its factor analyses despite "conceptual gaps that were potentially addressed by their inclusion" (WFF, 2018b). Omitted variables in this instance included: political/civil rights, wages, literacy, child mortality, corruption, GDP, government effectiveness, and gender inequality (WFF, 2018b), many of which have been highlighted as important in the literature. Omitting variables with partial collinearities can not only impact model accuracy, but can alter the explanations for slavery produced. Such variables can be involved in non-linear interactions, interpretation of which may offer benefits to policy. Recognising and understanding the effects of such collinearities in a model, rather than simply discarding them, is hence a key focus of this study.

Assumptions of linearity. Traditional regression analyses do not readily reveal non-linear dependencies, tipping-point thresholds, nor sub-population effects—despite such aspects being important issues when forming and implementing policy interventions. If addressing some key factor (e.g., education of women) predominantly impacts a particular context only (e.g., a specific age group), its importance can be lost within linear models—and insufficiently reflected via analysis of β coefficients. In domains such as slavery this is an important consideration, with the literature clearly identify variation in the relevance of particular drivers in different geographical regions. In Northern African countries, for example, relative AIDS prevalence is suggested as a strong candidate predictor of slavery, due to the number of children orphaned by the disease (Rau, 2002); in Western countries this relationship would be unanticipated, given the different and better resourced social and health services. Standard regression approaches often struggle to delineate non-linear contexts of this nature, despite their common occurrence in this domain. To overcome this limitation, the present study evaluates both non-linear and linear methods to model slavery prevalence.

Overfitting. Performing traditional regression analysis on small data carries significant risk of model over-fitting (Yarkoni and Westfall, 2017). This is a critical issue, with models that appear to provide "goodness of fit" to data, offering no guarantees of out-of-sample generalisability in reality. Such cases undermine the efficacy of any explanations emerging from models; and this adds risk to any real-world interventions formed from them. This issue is exacerbated in high-dimensional settings where multiple variables are modelled simultaneously (Yarkoni and Westfall, 2017). Partly, this may be why hypothetico-deductive approaches remain so prevalent in the social sciences, with studies tending to focus on a few key hypotheses in order to prevent the increase of family-wise error rates. While this does indeed help to prevent overfitting, the ability to uncover new, and perhaps unexpected, predictors is often lost. Furthermore, the ability to assess relative variable importances across a wide range of features is foregone, with interactions between variables left unmodelled—a situation that is particularly problematic to "wicked problems" such as modern slavery, where driving factors are unlikely to act in isolation.

In response to issues of multicollinearity, non-linearity and overfitting, the field of machine learning has sought to develop alternative ways to guard against model overfitting, while accommodating both non-linear and multivariate data. This has produced a rich array of *cross-validation* techniques, methods, which support inductive experimental setups while defending against p-hacking[1] and "procedural overfitting" (Yarkoni and Westfall, 2017). Cross-validation aims to find model parameterisations that maximise generalisability, rather than minimising regression residuals (whilst additionally reducing the influence/biases of the researcher in model specification and model selection phases). Owing to their inductive nature, such frameworks permit discovery of potential new predictors—and crucially allows comparative assessment of the importance of variables, even in non-linear settings. Hence, a strict cross validation approach is taken in this study to help guard against model overfitting.

Cross-validation was, however, designed with large datasets in mind. If *stable* model interpretations are to be found using smaller datasets, we require additional considerations and alterations to a typical machine-learning methodology. The steps required to achieve this our outlined in detail in the Method section, and centre on integration of feature compression/selection directly into the modelling process, rather than being undertaken a priori as has been common in the field. Reduction of the variable space at some point is unavoidable, if the "large p" problem (Kim et al., 2015) is to be handled. A set of k, potentially latent, variables must therefore be identified, where $k < p$. Unlike methods such as *partial least squares*, which identify such latent variables via parametric modelling assumptions, importantly our new approach treats k as a hyper-parameter to be identified as part of model class exploration. To traverse the model space, $\mathcal{M}$, we employ full, leave-one-out cross validation (LOOCV), allowing model parameters to be optimised, whilst using all of the "small n" data. This further ensures that insights extracted from fitted models will generalise to the whole dataset as much as possible.

With a best performing compression strategy and model parameterisation, $\hat{M} \in \mathcal{M}$, having been identified via this method, a final issue remains. The potential for multicollinearlity in the data increases the likelihood that multiple *equally* well-performing models exist—models that have (almost) equivalent prediction accuracy to $\hat{M}$, but which leverage their independent variables in different ways. To interpret a single model, and assume its internal predictive mechanisms are fully representative of the phenomena being analysed, is not therefore credible in "small n" situations. To remedy this "instability", and to increase confidence in the driving factors isolated, we therefore employ a *Rashomon-set analysis* (Rudin, 2019; Fisher et al., 2019; Smith et al., 2020). This involves different well-performing model solutions (i.e., those within an "epsilon-threshold" of the best model's prediction accuracy) being fit to the full dataset. Variable importance methods can then be applied to all models in this Rashomon set, allowing analysis of variable importance volatility and shedding light on both masking and interaction effects across predictors. Together, these steps allow for meaningful, potentially novel insights to be extracted via a non-linear, inductive approach even in "small n, large p" settings.

Study overview

In this study, we apply the "small n, large p" workflow introduced in the previous section to model the prevalence of slavery over 70 country-year data points. To disassociate from some of the criticisms made of the GSI prevalence estimates (Silverman, 2018; Gleason, 2019; Gallagher, 2017), we strictly use 2016 and 2018 data, which have been derived directly from the Gallup World Poll surveys, and not extrapolated in any way (see Methods and Landman and Silverman (2019) for more detail). All data is drawn from published and freely available open source datasets, and is comprised of 106 independent variables (covering a range of economic, socio-demographic and contextual indicators) with a single dependent variable depicting slavery prevalence (derived from GWP surveyed data). The primary goal of the study is to explore the driving factors of modern slavery at the global level, whilst avoiding assumptions of traditional linear methodologies predominately used in the field to date. Use of this inductive methodology provides the opportunity for new candidate predictors to emerge. We evaluate the utility of this approach via comparison to a (i) pre-selection of features using domain theory, and (ii) use of full, raw feature sets with statistical methods such as Partial Least Squares and Lasso regression. As well as providing insight into the drivers of slavery, our final model is also used to generate new out-of-sample estimates of slavery prevalence for countries where no survey data currently exists.

Methods

Data

Selecting the dependent variable. A single dependent variable was used. This was the GSI's country level prevalence estimates derived from the Gallup World Poll (GWP) survey data

(provided by the ILO and WFF) and converted to a percentage of the population. This included prevalence estimates for 48 unique countries over 2016 and 2018[2]. For 22 countries there were estimates for both 2016 and 2018, giving a total of 70 data points over the two years[3]. It is noted that the countries in this sample did not include any countries from Western Europe or North America, and thus the generalisability of the findings to these regions are limited. The median value of prevalence was 0.46%, with the upper and lower quartiles as 0.26% and 0.77% respectively.

Selecting the independent variables. The independent variables were scraped from online sources such as the World Bank Development Indicators (2018), UNAIDS (2019), the WomanStats project (Caprioli and et al., 2009), the Early Warning Project, CIRI Human Rights Data Project (Cingranelli et al., 2014), Landman and Silverman (2019) and the UN's Sustainable Development Goal (SDG) indicators. A total of 106 features were selected and collated[4]. The features, feature descriptions, and sources can be found in the Supplementary Material, page 11, section Variable Descriptions. The data sources were verified to ensure that sufficient information on how the data was coded/collected was available, missing data was minimal for the 70 data points, and information existed for the time period that the data reflected. If the variable was a composite scale or score then the variables from which it was constructed from, and how, needed to be clear before being selected.

Data pre-processing. Information on the open source variables indicated that often data had been collected over multiple years. When this was the case, the most recent year in the collection time frame was recorded. The year the data was collected was an important part of the data processing as for 22 countries there were two dependent variable data points—slavery prevalence in 2016 and slavery prevalence in 2018. Therefore, data was sought to cover both these time periods. Overall, there was not enough data specific to the years 2016 and 2018 so the data was grouped by time period (2016 and before, and post 2016) and the most recent data from each group selected[5] As an example, the Physical Security of Women Scale had data only from the years 2014 and 2019, therefore the former went in group 1 to predict slavery in 2016 and the latter went into group 2 to predict slavery in 2018. Grouping the data in this manner reduced the missing data, but with the caveat that the features were assumed to be relatively stable between 2011 and 2016 for group 1, and between 2016 to 2019 for group 2.

After this, any missing data that remained was then dealt with in the following way. First, variables which had >50% of data missing were discarded. Then, remaining data were imputed in two steps. Where a country had either 2016 or 2018 data, the data from the year where it was available was used for both years (therefore, in a small number of cases the same feature value was used to predict two different dependent variable values—prevalence in 2016 and 2018). After this, the subsequent missing values were minimal, and can be viewed in the Supplementary Material, Fig. S6. The final step was to use a multivariate feature imputation method with regression trees for the remaining missing values[6]. Variables were then normalised to be between between 0 and 1.

Analysis. The methodology used in this study was exploratory, inductive and data driven. This involved evaluating a number of different approaches and model classes to assess what configuration of method and model produced the most accurate predictions. Firstly, the utility of using *all* the features versus a smaller pool of features selected from the literature (the approach used by the GSI (WFF, 2018b)) was tested. This theory driven selection process was undertaken using a literature review and by consulting with domain experts. The 106 features were reduced to 35 spanning poverty, globalisation and the country's wealth, education, politics, violence, and conflict (the full list of features and their sources can be found in the Supplementary Material, page 11, section Variable Descriptions). Secondly, the value in using feature decomposition prior to modelling verses inputting the raw variables into the model was assessed. Non-negative Matrix Factorisation (NMF) was chosen as it forces the matrices to be non-negative, making the emergent components easier to interpret than when using other methods such as Principal Component Analysis (PCA). Finally, three different model classes (also selected for their interpretability) linear regression, decision tree, and random forest, with optimised meta-parameters, were also evaluated. The resultant sixteen combinations of different methods and models types can be found in Table 1. Leave one out cross validation (LOOCV) was used to evaluate the model parameters selected to ensure that the model was generalisable to the full dataset. Given the lack of data and that fitting the model for an explanation, rather than pure prediction, was the goal of the study, no data was held back to create a separate test set.

To harness the model for out-of-sample prediction (the prediction of slavery in countries where no survey data exists) additional independent variable data were scraped for the 171 countries (in 2018) unseen by the model. Where independent variable data was not directly available, the same imputation method was used as described previously. For more information on the missing data prior to imputation for this sample see the Supplementary Material, Fig. S6. The model was then trained on the 70 data points for 2016 and 2018 where survey data was used in the estimates, and out-of-sample predictions made for the countries with no survey data.

Table 1 LOOCV performance of the different pipelines.

Feature selection	Feature compression	Model class	MAE
Full feature set	NMF	Linear Regression	0.266
Full feature set	NMF	Decision Tree	0.241
Full feature set	**NMF**	**Decision Tree (MF)**	**0.227**
Full feature set	NMF	Random Forest	0.237
Full feature set	Partial Least Squares (PLS) Regression		0.269
Full feature set	No compression	Lasso Regression	0.248
Full feature set	No compression	Decision Tree	0.322
Full feature set	No compression	Decision Tree (MF)	0.250
Full feature set	No compression	Random Forest	0.246
Theory-based	NMF	Linear Regression	0.291
Theory-based	NMF	Decision Tree	0.282
Theory-based	NMF	Decision Tree (MF)	0.232
Theory-based	NMF	Random Forest	0.248
Theory-based	Partial Least Squares (PLS) Regression		0.284
Theory-based	No compression	Lasso Regression	0.248
Theory-based	No compression	Decision Tree	0.293
Theory-based	No compression	Decision Tree (MF)	0.228
Theory-based	No compression	Random Forest	0.249

Note: The best performing pipeline is highlighted in bold. *MAE* mean absolute error. *NMF* Non-negative Matrix Factorisation. "Theory-based" refers to feature selection based on the literature ($P = 34$). MF refers to allowing the "max features" hyper-parameter to be tuned, so that the decision tree could not select from all features at each split. A Lasso regression (L1 norm regularisation) was used instead of a linear regression when no feature compression was used due to the the high number of independent variables. For a graphical visualisation of the MAEs see the Supplementary Material, Fig. S1.

Results

Model selection. The model selection phase of our analysis considered a range of candidate features spaces: the full feature set (all 106 national level variables), a subset of 34 variables reported in the literature as relating to modern-slavery, and a range of compressed feature spaces (expressing between 2 and 8 components). This meant that as well as testing different model classes, the utility of feature-selection based on theory and different feature compressions could also be tested. Non-negative Matrix Factorisation (NMF) was selected to identify latent variables due to ease of interpretation in comparison to other compression methods, such as PCA. NMF parameterisations (k) were explored as part of the grid search, meaning the interactions between feature compression and model class parameterisations were also captured. This allowed identification of optimal latent variable compression for each predictive mechanism. The model classes explored ranged from linear and regularised regression-based models, to non-linear approaches such as decision trees and random forest models (also chosen due to high interpretability). Partial least squares was also examined due to its internal identification of latent structures. The best performing parameterisation for each of these modelling strategies is detailed in Table 1, with Fig. 1 illustrating the distribution of performances for all models with MAE < 0.27.

The best performing model used the full feature pool ($p = 106$), compressed the feature space via latent NMF components ($k = 6$), then used a (non-linear) decision tree model with a restricted number of "max features" (MF) available at each split (to help prevent the tree from getting stuck at a local minima). For the full set of model parameters see the Supplementary Material, page 11, section Model Parameters. The best model's performance was a significant improvement on both the mean (MAE = 0.366, $p < 0.001$) and median (MAE = 0.349, $p < 0.001$) baseline predictions analysed using a Wilcoxon Signed-Rank t-test.

Predictions using leave-one-out analysis compared to the "actual" slavery prevalence estimations can be viewed in Fig. 2, accompanied by the distribution of 10,000 bootstrapped predictions to illustrate the associated uncertainty. The graph highlights that the best model, $\hat{M}$, was less effective at predicting a subset of countries, and in particular those with the highest prevalence of slavery, which it tended to underestimate[7].

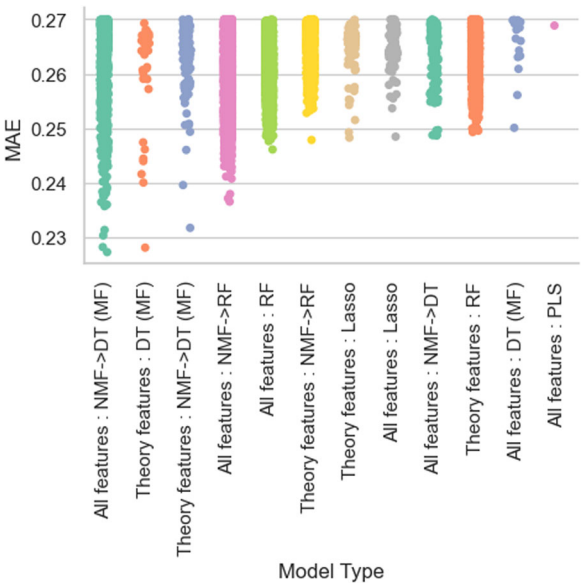

Fig. 1 The LOOCV performance of all models where MAE was < 0.27. Each point represents the performance of a distinct set of hyper-parameters within each model class.

Model interpretation. In the model interpretation phase, the best model, $\hat{M}$, was re-fit to the data, and the resulting NMF components and variable loadings analysed to determine component themes (see Supplementary Materials, Fig. S3 for a breakdown of the model). Components' relative importances were then compared using permutation importance (Altmann et al., 2010). Finally, a group of similarly performing models that constructed a *Rashomon set* (Rudin, 2019; Fisher et al., 2019; Smith et al., 2020) were utilised to investigate whether insights converged across multiple well performing models, or whether alternative explanations exist.

For $\hat{M}$, the NMF stage reduced the 106 different variables to 6 latent components, interpretable as: Democratic Rule, Armed Conflict, (lack of) Physical Security for Women, Social Inequality and Discrimination, Access to Resources, and Religious and Political Freedoms (specific variable loading's can be viewed in the Supplementary Material, Fig. S2). Interestingly, the component themes that emerged closely matched the vulnerability factors used by the WFF in their vulnerability model, with one notable exception: the addition of a component focusing specifically on the *Physical Security of Women*. The component permutation importance for $\hat{M}$ was also calculated, with Fig. 3 providing a comparison of the components' importance, and illustrating that access to resources was identified as the most important predictor of national slavery prevalence in the model.

This, however, only reflects the viewpoint of a single model. To test the stability of the insights produced, further analysis was performed using a Rashomon set approach (Rudin, 2019). Risk of model instability is relatively high in "small n, large p", and can be exacerbated when non-parametric approaches such as decision trees (Breiman, 2001) are employed to handle multicollinearity between the components. To deal with this it is therefore necessary to examine whether other well performing models gave converging or diverging explanations (referred to as the *Rashomon effect*(Semenova and Rudin, 2019)). In this case, the Rashomon set was defined as any model whose MAE performance was within 2.5% of the best performing model (< 0.233)[8]. Models retained in the Rashomon set are outlined in Table 2, and feature: three other models with the same pipeline as the best model (using NMF)→DT (MF) on all the features); a DT (MF) using just the theory features, and a final NMF→DT (MF), which used only theory-selected features. The disagreements in predictions made across the Rashomon set can be viewed in Fig. 4.

Despite NMF models utilising different random seeds for initialising coordinate descent and different coefficients for the regularisation terms (alpha), $K = 6$ emerged as optimal across all pipelines on the full feature set. Corresponding NMF themes remained predominantly stable across each model (see the Supplementary Material, Fig. S4), providing strong evidence for the applicability of the six latent variables identified.

Within model's sharing the same class as $\hat{M}$ (NMF→DT (MF), all features), the permutation importances of components were compared across models, in order to understand how other competing solutions compared to the explanation provided by $\hat{M}$. This comparison can be viewed in Fig. 5. The graph illustrates the different predictive strategies the models used following feature reduction. The variation across models highlights the risk of over-interpreting insights from a single model. For example, in the best performing model, $\hat{M}$, Democratic Rule and Armed Conflict are relatively unimportant predictors of slavery prevalence, yet are the most important in Rashomon models, M_2, M_3, and M_4.

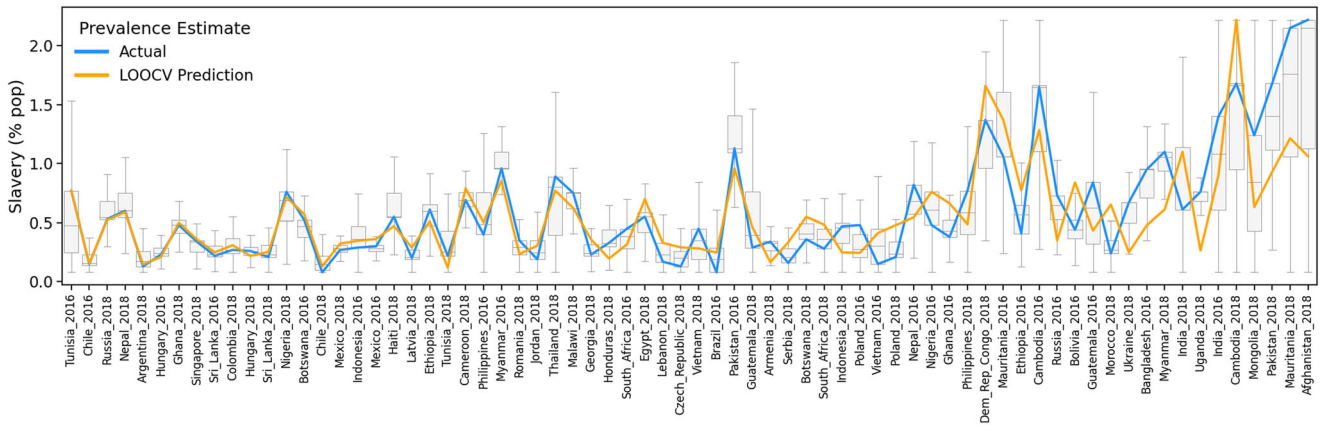

Fig. 2 The predictions of slavery prevalence (individuals enslaved as a % of the population) made by the best model using leave-one-out cross validation (LOOCV), compared to the "actual" prevalence as estimated using the Gallup World Poll (GWP) survey data. The grey box plots illustrate the distribution of 10,000 bootstrapped LOOCV predictions (using the full pipeline NMF→DT) to help illustrate the uncertainty associated with our model's predictions. The box shows the quartiles of the bootstrapped predictions while the whiskers extend to show the rest of the distribution, except for points that were determined to be "outliers" (using a function of the inter-quartile range), which are not plotted. The *x*-axis is ordered by the MAE.

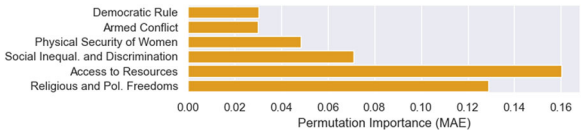

Fig. 3 The importance of the NMF components in the best performing model. Permutation importance (Altmann et al., 2010) was used.

Table 2 The Rashomon set.

Model name	MAE	Features	Model class	K components
Best Model	0.227	All features	NMF→DT (MF)	6
Rashomon 1	0.228	Theory-selected	DT (MF)	
Rashomon 2	0.228	All features	NMF→DT (MF)	6
Rashomon 3	0.230	All features	NMF→DT (MF)	6
Rashomon 4	0.231	All features	NMF→DT (MF)	6
Rashomon 5	0.232	Theory-selected	NMF→DT (MF)	5

Subsequent analysis of components indicate that, despite their stability, correlations remain between them (see Supplementary Material, Fig. S5). In particular, Religious and Political Freedoms is negatively correlated with Armed Conflict ($r = -0.62$); and (lack of) Physical Security for Women is negatively correlated with Access to Resources ($r = -0.44$). The effect of the latter relationship is particularly apparent in Fig. 5, with Rashomon models, M_2 and M_3 requiring only one of those components to perform well. This suggests that there may additionally exist non-linear relationships between the two components, which cannot be fully captured in the correlation coefficient.

To understand the unanticipated relationship between Access to Resources and the Physical Security of Women more fully, partial dependency plots were analysed to look for non-linear dependencies between the variables. Figure 6 illustrates that the Physical Security of Women is predictive of slavery in contexts where access to resources is low. In other words, women are particularly vulnerable to being exploited in areas where there is poor access to fuel, electricity, piped/clean water, sanitation, and education (see variable loadings of the Access to Resources component in the Supplementary Material, Fig. S2). To further highlight the non-linear interactions behind the Physical Security of Women component, Individual Conditional Expectation (ICE) graphs shown in Fig. 7 indicate that when a (lack of) physical security for women is extremely high, this dramatically increases the partial dependency of the prevalence prediction—more than any other component. The important role that Physical Security of Women can play in the prediction of slavery prevalence is made available here using a methodology that allows for non-linear interactions of this nature to be modelled.

Predicting prevalence for countries with no survey data. As a final stage of analysis we present new estimates for countries where no GWP survey data has been collected, projecting the best performing model out-of-sample in order to generate new estimates of prevalence. The model was fitted to the 70 country-year data points over the years 2016 and 2018 for which survey data were available. Predictions are output for the 172 countries in 2018 for which no survey data exists. Figure 8 shows the estimates produced compared to the estimates made by the model used in the 2018 edition of the Global Slavery Index (GSI) WFF (2018a). The GSI, which we compare to, used a hierarchical Bayesian linear model with additional adjustments (see WFF (2018a) and Diego-Rosell and Joudo Larsen (2018) for further methodological details).

These out of sample predictions constitute the best predictions within the parameters of our grid search, noting that that our parameter and model choices were influenced by the goals of the investigation—improved *understanding* of the predictors of slavery. It is well recognised that the goals of explanation and prediction do not always align (Yarkoni and Westfall, 2017; Breiman et al., 2001; Shmueli et al., 2010), and the model selection and parameter searches in our implementations were guided by a need for model interpretability (e.g., use of linear latent variable structures), not prediction. It therefore remains possible that predictions could be improved through the use of other models (i.e., non-linear compression and black box models) or data sampling techniques (such as taking the mean of the bootstrapped predictions). Nevertheless, considering the estimates produced by computational models with well-established and understood mechanisms, establishes not only (a) interpretable estimates, but also (b) whether different models with different features diverge or converge in their predictions, and (c) what further data and features might improve prediction.

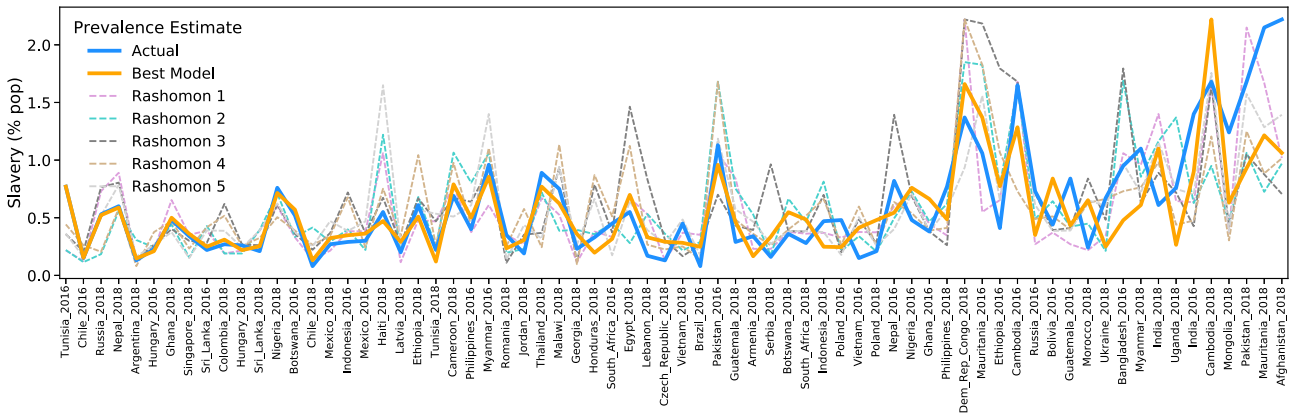

Fig. 4 The predictions of slavery prevalence (individuals enslaved as a % of the population) made by the best model and the five other well performing models in the Rashomon set (see Table 2). The *x*-axis is ordered by the MAE between our best model and the "actual" prevalence as estimated by the GWP survey data.

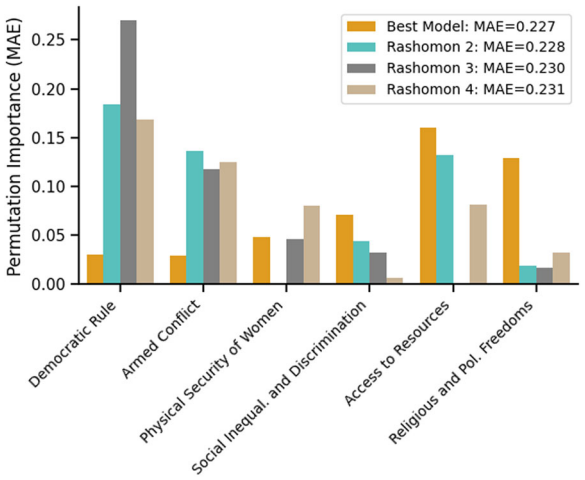

Fig. 5 A comparison of variable importance ratings to other Rashomon set models. Comparisons were only made between models in the same model class as the best model (NMF→DT (MF)).

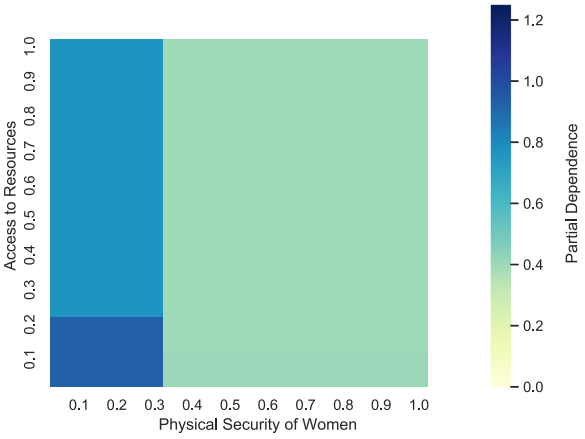

Fig. 6 A heat map illustrating that the partial dependency of the prevalence prediction of the best model is especially high when both Access to Resources and Physical Security of Women are low. Here, a lower score for Physical Security of Women indicates less security.

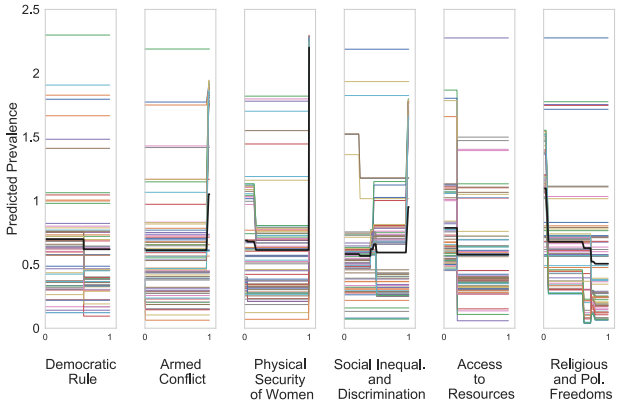

Fig. 7 Individual Conditional Expectation (ICE) plots to illustrate the non-linear effects of components on the best model's predictions. Each coloured line is a country-year data point with some jitter applied. The thick black line is the average partial dependency of the component on the prevalence prediction.

Discussion

This study, for the first time, presents an inductive machine-learning methodology and contemporary variable importance analysis to explore the complex predictors of modern slavery in the real world. By evaluating multiple different pipelines the utility of theory-driven feature selection versus guided feature compression (to help deal with the "large p" problem) has been tested, as well as comparing the performance of non-linear models versus their more traditional linear counterparts. Generally, and with but few exceptions, models accessing all features performed better than those using features selected from the qualitative literature. Given the generous selection of variables entered into the theory-selected feature pool this highlights the potential of an inductive methodology to uncover novel predictors, even in "small *n*, large *p*" contexts. Model class comparisons showed that non-linear models gave better predictions than their linear counterparts, supporting the notion that our analysis is capturing non-linearities that have not been previously modelled quantitatively before. Finally, allowing guided feature compression, parameterisable as part of a grid search, outperformed traditional approaches that incorporated latent structures, such as partial least squares.

The majority of the models in our Rashomon set benefited from identification of latent components. This broadly validates

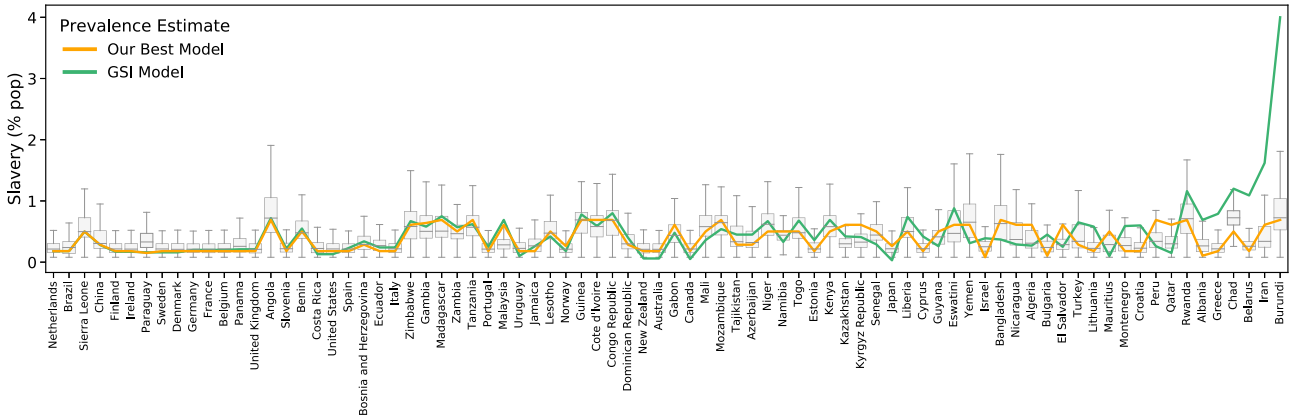

Fig. 8 The predictions made for the 2018 prevalence of slavery (individuals enslaved as a % of the population) in countries where no GWP survey data exists. The estimates made by the best performing model are compared to those made by the GSI. The grey box plots illustrate the distribution of 10,000 bootstrapped LOOCV predictions (using the full pipeline NMF→DT) to help illustrate the uncertainty associated with our model's predictions. The box shows the quartiles of the bootstrapped predictions while the whiskers extend to show the rest of the distribution, except for points that were determined to be "outliers" (using a function of the inter-quartile range), which are not plotted. The *x*-axis is ordered by the disagreement between our model's predictions and the predictions made by the GSI model. The countries displayed are those which had < 10% missing independent variable data. For all 172 out-of-sample predictions and information on missing data see the Supplementary Material, Table S1.

the utility of summarising predictors of slavery into *k* latent factors (first done by the WFF in the construction of the Vulnerability Model (WFF, 2018b)). Whilst the final model decided on by the WFF consists of 5 factors, the naturally occurring solution (based on eigenvalues greater than 1) actually consisted of 6 factors (WFF, 2018b), corroborating our bottom-up findings. The present study extends the WFF's initial work by constructing components directly shaped by the outcome variable, slavery prevalence, and thus a step towards overcoming a core criticism of the GSI—that vulnerability or risk of slavery is not the same as prevalence (Silverman, 2018).

Importantly, our analysis highlights the additional contribution of a new component, *Physical Security of Women*, reflecting the onus on a country's law enforcement to protect woman from domestic violence, rape and sexual assault, marital rape, shame/honour killings or femicides (Caprioli and et al., 2009). The Physical Security of Women has been previously been overlooked in quantitative models (WFF, 2018b), perhaps due to the non-linear effect it has on the prediction of slavery, in combination with its interactions with other components. Most notable is the finding that not only is there is a greater vulnerability to women's physical security in areas lacking access to resources, but this in turn is an indicator of slavery occurrence (including the sexual exploitation and forced marriage of women).

In previous models, gender inequality features have had to be removed because of issues with multicollinearity (WFF, 2018b). Our methodology, which does not require the removal of correlated variables, demonstrates that gender inequality is likely a core piece of the puzzle in predicting national slavery figures. In addition to the Physical Security of Women being a predictive component, variables depicting either the reporting or prevalence of rape also loaded highly onto two other components (see the variable loadings in the Supplementary Material, Fig. S2). This highlights that when explanation is the goal, it can be important for the researcher not to remove features that are correlated a priori, and instead navigate issues of multicollinearity within the modelling so that a more complete explanation can be achieved.

A Rashomon set analysis highlighted that despite the latent components found to be stable, there existed a high degree of variability in the importance ratings the models assigned to each component. All components were used to differing degrees in each Rashomon model, emphasising the care that must be employed when using machine learning not just for prediction, but to understand underlying factors. There is danger, particularly in the social sciences, in focusing on a single model for interpretation. Even if the model performs well, misleading explanations can be drawn, particularly in the presence of significant correlations (Fisher et al., 2019), and when there is a degree of uncertainty around the accuracy of the data (see limitations below).

Finally, this study leveraged machine learning to generate new estimates of slavery prevalence. These estimates are of course useful unto themselves in understanding both the extent of slavery, and surrounding uncertainty. Additional insights can also be obtained by reflecting on the differences between the estimates produced by different models/methodologies. It is highlighted here that we are not claiming these estimates to be an improvement over the GSI's, instead merely being transparent about what our model would predict given the data available (which is still subject to data representation issues (Gleason, 2019)), so that insights from comparing the estimates can be gleaned. Our model was not optimised for prediction (instead explanations), and we did not have access to the raw GWP data used by the WFF. However, we make an effort to use a transparent data driven methodology, as well as estimating the uncertainty of our estimates (two criticisms of the GSI methodology (Silverman, 2018; Gallagher, 2017)).

In contrast, the WFF's approach we considered in comparison, involved multiple adjustments and does not solely reflect the information held in independent variable data (WFF, 2018a). This prompts the question of what additional information or intuition experts believe is not currently being captured in data and what would be valuable data for researchers to have. Understanding these data gaps, as well as when and why certain estimates don't align with experts' expectations, will be crucial for advancing the measurement of modern-day slavery forward. Future efforts might focus on how human judgement and computational modelling can be combined to build prediction models—harnessing the advantages of a data driven approach (objectivity, quantification, and out-of-sample predictions) combined with expert human judgement and additional contextual information.

In this study, while key limitations of traditional regression approaches were mitigated against, several limitations remain inherent to the data. The dependent variable, although derived from survey data collected using a representative sampling

methodology (Gallup World Poll, 2020), nonetheless represents an estimate rather than a direct measure of slavery. Consequently, prevalence estimates must be considered more a reflection of slavery risk across the given the population, rather then corresponding to formal incidence numbers. Further, the estimates we produce are not able to offer indication of the breakdown in the typology of slavery occurring within a country. Recent work has focused on establishing proxy indicators for specific types of vulnerability/exploitation, or exploitation within specific industries, using small scale surveys married to digital traces such as mobile phone records (Engelmann et al., 2018), satellite imagery (Boyd et al., 2018; Jackson et al., 2018; 2020; Foody et al., 2019), and vessel data (Nakamura et al., 2018). Such approaches can help to bolster the data in this domain without putting vulnerable individuals at further risk.

This study applied machine-learning methods to the context of modern slavery to better understand the global drivers of exploitative practices. Using rigorous cross validation approaches, careful consideration of model stability, and an in-depth variable importance analysis, stable predictive components emerged using data characterised as "small *n*, large *p*". Notably, a novel predictive component was found in the *Physical Security of Women*, which was shown to predict prevalence non-linearly, and in association with other components (in particular Access to Resources). The limitations our methodology overcame, namely assumptions of linearity, significant correlations between variables, and overfitting, allowed for the discovery of the predictive capabilities of this component. Although the importance of this feature may be unsurprising to many working with exploited women and girls, it has previously been under-emphasised in global quantitative models and research studying modern slavery as an umbrella term for all coercive activities (e.g., WFF (2018b)). This study also investigated under what contexts the Physical Security of Women is most predictive—where access to resources is poor. These findings make the case for further exploration of data-driven, inductive approaches and out-of-sample prediction to complement existing methodologies being used to study complex social, or "wicked", problems such as modern-day slavery.

Data availability

The datasets generated during and/or analysed during the current study, along with the code used to analyse them, are available for download in the ml-slavery GitHub repository: https://github.com/ml-slavery/ml-slavery

Received: 13 April 2021; Accepted: 4 October 2021;
Published online: 17 November 2021

Notes

1 Where researchers collect or select data or statistical analyses based on either the conscious or unconscious motivation to obtain a statistically significant result (Head et al., 2015).
2 The 2018 GSI in actually includes estimates of prevalence for a total of 167 countries. However, the majority of these were produced using a risk model, and not estimated from survey data. The reliability and practical usefulness of these extrapolated estimates have been queried (Silverman, 2018; Landman and Silverman, 2019). Therefore, this analysis only uses estimates that were derived directly from the GWP survey data (as in Landman and Silverman (2019)).
3 Given that the sampling procedure for both the 2016 and 2018 Gallup World Poll was random (Gallup World Poll, 2020), there was no known reason to assume that the 2016 survey would have directly influenced the 2018 survey; and thus the country prevalence estimates for 2016 and 2018 were treated as independent.
4 Features were selected based on past literature, the authors' intuition, and conversations with domain experts. Initially there were 137 variables in the feature pool, before 31 were discarded due to missing data, duplication, or measuring a phenomena too similar to the dependent variable.
5 For ease, the 2018 SDG data was used for 2016 and 2018 due to the multiple sources and dates from which the data were collected.
6 The CART regression tree method was chosen due to features being too highly correlated to perform regression-based imputation methods, such as predictive mean matching (causing singularities in the matrix). Further, as linear models such as NMF and linear regressions were being utilised in this analysis, a non-linear imputation method was preferable to avoid the variables' collinearity being inflated by the imputation process.
7 In this particular analysis we chose not to add a weighting to the high estimates, but recognise that this might be appropriate if using the model as a tool to identify areas at high risk.
8 While selecting an epsilon-values of this nature when generate Rashomon set's remains a subjective task, Fig. 1 illustrates the visible gap in performance between the set of models included in comparison to competitors.

References

Adesina OS (2014) Modern day slavery: poverty and child trafficking in nigeria. African Iden 12:165–179
Altmann A, Toloşi L, Sander O, Lengauer T (2010) Permutation importance: a corrected feature importance measure. Bioinformatics 26:1340–1347
Bales K, Hesketh O, Silverman B (2015) Modern slavery in the UK: How many victims? Significance 12:16–21
Bales K (2006) Testing a theory of modern slavery. Free the Slaves. https://glc.yale.edu/sites/default/files/files/events/cbss/Bales.pdf
Bales K, Murphy LT, Silverman BW (2020) How many trafficked people are there in greater new orleans? lessons in measurement. J Human Traffick **6**, 375–387 (2020). https://doi.org/10.1080/23322705.2019.1634936
Boyd DS et al. (2018) Slavery from space: demonstrating the role for satellite remote sensing to inform evidence-based action related to un sdg number 8. ISPRS J Photogramm Remote Sens 142:380–388
Breiman L (2001) Random forests. Mach Learn 45:5–32
Breiman L et al. (2001) Statistical modeling: the two cultures (with comments and a rejoinder by the author). Stat Sci 16:199–231
Brown D et al. (2021) Modern slavery, environmental degradation and climate change: fisheries, field, forests and factories. Environ Plann E: Nat Space 4:191–207
Caprioli M et al. (2009) The womanstats project database: Advancing an empirical research agenda. J Peace Res 46:839–851
Chan L, Silverman BW, Vincent K (2020) Multiple systems estimation for sparse capture data: Inferential challenges when there are nonoverlapping lists. J Am Stat Assoc 116, 1297–1306. https://doi.org/10.1080/01621459.2019.1708748
Cingranelli DL, David LR, & Clay KC (2014). The CIRI Human Rights Dataset. http://www.humanrightsdata.com.
Cockayne J et al. (2019) Symposium: Modelling modern slavery risk. Delta 8.7 . http://collections.unu.edu/eserv/UNU:6722/Symposium_ModellingModernSlaveryRisk_Final.pdf
Cooper C, Hesketh O, Ellis N, Fair, A (2017) A typology of modern slavery offences in the UK. Home Office
Cruyff M, van Dijk J, van der Heijden PG (2017) The challenge of counting victims of human trafficking: not on the record: a multiple systems estimation of the numbers of human trafficking victims in the netherlands in 2010–2015 by year, age, gender, and type of exploitation. Chance 30:41–49
Datta MN, Bales K (2013) Slavery is bad for business: analyzing the impact of slavery on national economies. Brown J World Affair 19:205–223
Datta MN, Gustafson O, Lubin C, Kelleher G, Berg R (2018) Assessing the global slavery index. The SAGE Handbook of Human Trafficking and Modern Day Slavery 38. Sage
Diego-Rosell P, Joudo Larsen J (2018) Modelling the risk of modern slavery. Available at SSRN 3215368. https://doi.org/10.2139/ssrn.3215368
Dong J, Rudin C (2020) Exploring the cloud of variable importance for the set of all good models. Nat Mach Intell 2:810–824
Engelmann G, Smith G, Goulding J (2018). The unbanked and poverty: predicting area-level socio-economic vulnerability from m-money transactions. In 2018 IEEE International Conference on Big Data (Big Data), 1357–1366. IEEE
Fisher A, Rudin C, Dominici F (2019) All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously. J Mach Learn Res 20:1–81
Fitzgibbon K (2003) Modern-day slavery? The scope of trafficking in persons in Africa. Afr Securit Stud 12:81–89
Foody GM, Ling F, Boyd DS, Li X, Wardlaw J (2019) Earth observation and machine learning to meet sustainable development goal 8.7: mapping sites associated with slavery from space. Remote Sens 11:266
Gallagher AT (2017) What's wrong with the global slavery index? Anti-Traffic Rev . https://doi.org/10.14197/atr.20121786

Gallup World Poll (2020). How does the gallup world poll work? Measures the attitudes and behaviors of the world's residents. https://www.gallup.com/178667/gallup-world-poll-work.aspx
Gleason KA (2019) Facing choices when modelling modern slavery risk. Symposium: Modelling Modern Slavery Risk 8–10 (2019). http://collections.unu.edu/eserv/UNU:6722/Symposium_ModellingModernSlaveryRisk_Final.pdf
Van de Glind H, Kooijmans J (2008) Modern-day child slavery 1. Child Soc 22:150–166
Guth A, Anderson R, Kinnard K, Tran H (2014) Proper methodology and methods of collecting and analyzing slavery data: an examination of the global slavery index. Soc Inclus (ISSN: 2183-2803) 2:14–22. https://mars.gmu.edu/jspui/bitstream/handle/1920/9895/2014-11-17-Guth-Article.pdf
Head BW, Alford J (2015) Wicked problems: implications for public policy and management. Admin Soc 47:711–739
Head ML, Holman L, Lanfear R, Kahn AT, Jennions MD (2015) The extent and consequences of p-hacking in science. PLoS Biol 13:e1002106
ILO (2017) Global estimates of modern slavery: Forced labour and forced marriage. https://www.ilo.org/wcmsp5/groups/public/@dgreports/@dcomm/documents/publication/wcms_575479.pdf
ILO (2001) Children in prostitution–a rapid assessment. ILO Tanzania (ISBN: 92-2-112832-6. https://www.ilo.org/ipec/Informationresources/WCMS_IPEC_PUB_2445/lang-en/index.htm
Jackson B, Boyd DS, Ives CD, Sparks JLD, Foody GM, Marsh S, Bales K .(2020) Remote sensing of fish-processing in the Sundarbans Reserve Forest, Bangladesh: an insight into the modern slavery-environment nexus in the coastal fringe. Maritime Stud 19(4):429–444
Jackson B, Bales K, Owen S, Wardlaw J, & Boyd DS (2019) Analysing slavery through satellite technology: How remote sensing could revolutionise data collection to help end modern slavery. J Modern Slavery, 4(2):169–200
Johnstone IM, Titterington DM (2009) Statistical challenges of high-dimensional data. Phil Trans R Soc A 4237–4253. https://doi.org/10.1098/rsta.2009.0159
Kim B, Patel K, Rostamizadeh A, Shah J (2015). Scalable and interpretable data representation for high-dimensional, complex data. In Proceedings of the AAAI Conference on Artificial Intelligence, vol. 29, 1763–1769
Landman T (2020) Measuring modern slavery: Law, human rights, and new forms of data. Human Rights Quarterly 42:303–331
Landman T, Silverman BW (2019) Globalization and modern slavery. Politics and Governance 7:275–290
Larsen JJ, Durgana DP (2017) Measuring vulnerability and estimating prevalence of modern slavery. Chance 30:21–29
Manokha I (2004) Modern slavery and fair trade products: Buy one and set someone free. In The Political Economy of New Slavery, 217–234. Springer
Manzo K (2005) Exploiting West Africa's children: trafficking, slavery and uneven development. Area 37:393–401
Nakamura K et al. (2018) Seeing slavery in seafood supply chains. Sci Adv 4:e1701833
Nathans LL, Oswald FL, Nimon K (2012) Interpreting multiple linear regression: a guidebook of variable importance. Pract Assess Res Eval 17:9
Pryshlakivsky J, Searcy C (2013) Sustainable development as a wicked problem. In Managing and engineering in complex situations, 109–128 (Springer, 2013)
Rau, B. (2002) Combating child labour and HIV/AIDS in sub-Saharan Africa (International Labour Office (ISBN 92-2-113288-9). https://www.ilo.org/wcmsp5/groups/public/-ed_protect/-protrav/-ilo_aids/documents/publication/wcms_119161.pdf
Rittel HW, Webber MM (1973) Dilemmas in a general theory of planning. Policy Sciences 4:155–169
Roser M, Ritchie H (2018) Hiv/aids. Our World in Data. https://ourworldindata.org/hiv-aids
Rudin C (2019) Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nat Mach Intell 1:206–215
Semenova L, Rudin C (2019) A study in rashomon curves and volumes: a new perspective on generalization and model simplicity in machine learning. Preprint at https://arxiv.org/abs/1908.01755 (2019)
Shmueli G et al. (2010) To explain or to predict? Stat Sci 25:289–310
Silverman BW (2018) Demonstrating risks is not the same as estimating prevalence. Paper presented at Delta 8.7 Modelling the Risk of Modern Slavery Symposium (2018). https://delta87.org/2018/12/demonstrating-risk-not-same-estimating-prevalence/
Silverman BW (2020) Multiple-systems analysis for the quantification of modern slavery: classical and Bayesian approaches. J Royal Stat Soc: Series A (Statistics in Society) 183(3):691–736
Smith G, Mansilla R, & Goulding J (2020). Model Class Reliance for Random Forests. Advances in Neural Information Processing Systems, 33.
Tickler D et al. (2018) Modern slavery and the race to fish. Nature communications 9:4643
United Nations (2021) UN Sustainable Development Goal 8.7. https://sdgs.un.org/goals/goal8
Weitzer R (2014) Miscounting human trafficking and slavery. Open Democracy . https://www.opendemocracy.net/en/beyond-trafficking-and-slavery/miscounting-human-trafficking-and-slavery/
WFF GSI (2018b) Methodology, Vulnerability Model. https://www.globalslaveryindex.org/2018/methodology/vulnerability/
WFF GSI (2018a) Methodology, Prevalence. Section: Data Limitations. https://www.globalslaveryindex.org/2018/methodology/prevalence/
Yarkoni T, Westfall J (2017) Choosing prediction over explanation in psychology: lessons from machine learning. Perspect Psychol Sci 12:1100–1122

Acknowledgements

The authors would like to thank Sir Bernard Silverman for his constructive comments and feedback. This work was supported by Wave 1 of The UKRI Strategic Priorities Fund under the EPSRC Grant EP/T001569/1 and EPSRC Grant EP/W006022/1, particularly the "Data Science for Addressing Modern Slavery" project of the "Criminal Justice System" theme within those grants & The Alan Turing Institute. It was also supported by EPSRC Grant, EP/T003928/1, "Risk prediction for Women's Health and Rights."

Funding

Open access funding enabled and organized by Projekt DEAL.

Competing interests

The authors declare no competing interests.

Ethical approval

Not applicable as this article does not contain any studies with human participants performed by any of the authors.

Informed consent

Not applicable as this article does not contain any studies with human participants performed by any of the authors.

Additional information

Supplementary information The online version contains supplementary material available at https://doi.org/10.1057/s41599-021-00938-z.

Correspondence and requests for materials should be addressed to Rosa Lavelle-Hill.

Reprints and permission information is available at http://www.nature.com/reprints

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.