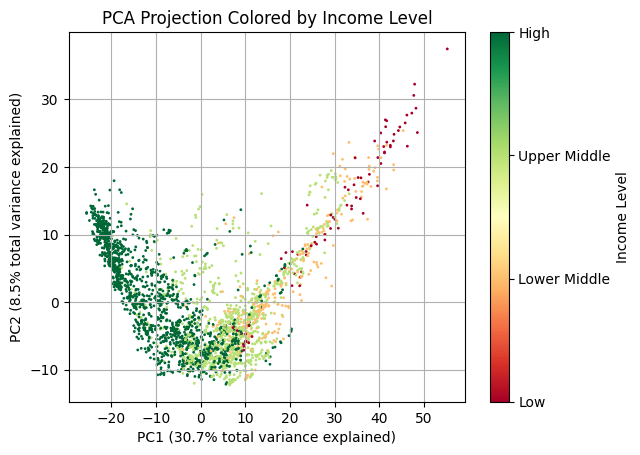
**Agenda:**

* Update about writing schedule:
  + Finished background and am nearly finished with the literature review. I will have both and likely the methods for review next week.
  + By mid-week 9 (in two weeks), I hope to have the first full draft
* Discussion of some of the diagrams to be included in the report
* Discussion of minority oversampling
* Discussion of possible poster

A graph of red and black dots

AI-generated content may be incorrect.

* Clusters of high-income countries are in a similar position to clusters of countries with low MMR estimates (as expected from the literature).
* Few samples of low-income countries.
* The low and lower-middle income countries appear to follow the clearest trends, with the upper-middle income samples everywhere in the plot.

A graph of colored dots

AI-generated content may be incorrect.

* No clear trend between year and income level.
* This could explain why the best performing split by year model had a relative error of 37% while the best performing split by country model had a relative error of 6%.
  + Split by year needs to predict the future (test set 2015-2018), but without clear clustering and trends that the model can learn, this would be difficult.

A graph of different colored dots

AI-generated content may be incorrect.

* This plot displays the MMR rate for a specific country year, colour coded by income level.
* It demonstrates that the high and upper-middle income countries are concentrated in the 0 to 500 part of the range, while the low and lower-middle countries are much more spread out.

A graph showing a graph of progress

AI-generated content may be incorrect.A graph of a bar graph

AI-generated content may be incorrect.

* This plot demonstrates that the model can predict most accurately for lower middle-income countries (smallest relative error with fewest outliers).
  + This indicates that the model was able to identify strong signals for lower-middle and high-income countries.
  + The high-income countries appear to have a clear signal on the PCA (up and to the left)
    - But outliers may be due to noise from the lower income countries, accounting for the better performance of the high-income sensitivity model.
  + The lower-middle income countries appear to also have a clear signal (up and to the right)
    - The lower error with few outliers indicates that the model has been able to identify a clear trend. Potentially, then, removing data obscures this trend, accounting for the reduced accuracy in the sensitivity analysis
      * This could be related to how the lower-middle income MMRs have a large range, and thus the model could learn to predict them from using trends seen for the other income levels. (also could be seen by how lower middle income countries’ dots overlap other income levels in the PCA)
        + This may be why the lower-middle income countries performed worse in the sensitivity analysis than in the original model.
  + While the low-income countries also have a clear-ish trend, there are very few datapoints, accounting for the higher error
    - More variation with fewer datapoints
      * This is shown in the plot below.
    - During sensitivity analysis, the model can concentrate on learning the patterns for the low-income countries versus the high income (which would be noise in this context), increasing sensitivity performance relative to the baseline.
  + The upper-middle countries are less clustered, making it more difficult for the model to determine a trend and thus explaining the larger error.
    - More variation, making it more difficult to identify a trend

A graph with blue dots

AI-generated content may be incorrect.

Demonstrates the greater disagreement among base estimators (inputs to the ensemble) when MMR is high. If MMR is high, it is more likely to correspond to data sparse countries, explaining the higher standard deviation of the base estimators.

A graph of numbers and a number of error

AI-generated content may be incorrect.

* This graph shows the relative error of the 20 base estimators most highly valued by the random forest ensemble model.
* However, some of these models definitely have higher predictive error than observed in the random forest base models (shown below).
* Additionally, the random forest base models did not have greatly different features than the most valued models.
* I do not know what else to investigate to try and explain why the random forest ensemble did not place high importance on any of the random forest models.

A graph of different colored bars

AI-generated content may be incorrect.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of Ensemble Model** | **MSE** | **RMSE** | **MAE** | **Relative Error** | **R2** |
| Voting ensemble | 7107 | 84.30 | 25.01 | 0.33 | 0.815 |
| Elastic Net ensemble | 1689 | 41.10 | 9.69 | 0.19 | 0.956 |
| Random Forest Ensemble | 2161 | 46.49 | 8.68 | 0.07 | 0.944 |
| SVM ensemble | 3441 | 58.66 | 25.16 | 0.48 | 0.910 |

* Despite the elastic net and random forest ensemble methods alternating between lower predictive error, I chose the random forest ensemble as the highest performing model due to its substantially lower relative error score.
* The random forest ensemble may have performed better than the elastic net ensemble on the relative error score because it performed better when predicting lower MMRs but worse when predicting higher MMR scores.
  + Predicting lower MMRs: would be a higher proportion of the ‘correct’ value, thus making relative error larger.
    - If elastic net was worse at predicting higher MMR scores, it would have a higher relative error.
  + Predicting higher MMR scores: would resemble ‘outliers’, and thus raise MSE and MAE
    - Potentially, the random forest ensemble performed worse on the less seen low-income, high MMR countries.
    - Given one of the aims is to predict better for these lower income countries, should I make the elastic net my highest performing model?
* I chose the random forest model because of its better performance on the dataset as a whole.

A graph of birth rate

AI-generated content may be incorrect.

* For the comparisons to the literature, I do not have ‘global’ data points. I could either:
  + Compare my results for a specific country against literature predictions for that same country.
  + Sum all the predicted MMRs for countries in the year for each method and compared these values (shown below) as an analogous ‘global’ measure. However, only country-years I have data for will be summed (which is why the sum is lower than the figure in the literature).
* Additionally, I am only comparing my model’s predictions on test data points to the literature’s predictions for the same points. Should I also be comparing my model’s predictions on the training data?
* The figure above shows the summed results
  + Generally, my estimates are smaller than those in the literature, demonstrating how my model was trained to predict underreported target data.
  + However, my estimates were among the highest in 2015, 2014, 2010, 2009, 2006, 2005, 2003, 2000, 1999, 1998, 1995. The input feature data had the smallest proportion of missing data in 2000, 2005, 2010, and 2015. Potentially, the two observations are loosely related to each other, where underpredicting decreased with the higher amount of data.

A graph of birth rate

AI-generated content may be incorrect.

* This diagram shows the MMR comparison for Germany, an example of a high-income country.
* Here, my estimates are still lower than the other models (and given the small scale of values, this difference could manifest as a high magnitude percentage difference). However, my results are clustered with the UN MMEIG and GBD Study results, whereas the Microsimulation results show a relatively more dramatic downward trend.
  + This difference may be due to incorrect initialisation of parameters in the microsimulation model
  + For example, the GMatH specification states that, for antenatal care (ANC) visit data:
    - ‘We used a two-part hierarchical Poisson model to estimate the number of ANC visits based on DHS data. We used upper middle income priors for high income countries due to lack of DHS data in high income countries.’
  + However, there is a study conducted in Germany saying that the global burden of disease study estimates underestimate true mortality. Thus, the microsimulation model may be closer to the truth.

**A graph of different colored dots

AI-generated content may be incorrect.**

* These are the results for an upper-middle income country, where UM countries; priors are sometimes used to inform the GMatH high-income country prior, potentially explaining why these results are closer to the other estimates.
* My results initially underestimate then are mixed among the other estimates.
  + However, the fluctuation looks like MMR is increasing rapidly, which may negatively influence policy decisions.
* Underestimating due to training on underreported data
* Maybe the fluctuation could be due was a change in socioeconomic conditions that caused an increase, or could be due to low-quality data.

A graph with numbers and colored dots

AI-generated content may be incorrect.

* My estimates were generally higher than the literature estimates for Rwanda (low-income country).
* Potentially, this can be due to the missing, low-quality data.
* However, given the underestimation of MMR values, my estimates could also be closer to the truth and are in some of the others’ uncertainty intervals.
  + Potentially, my model predicts higher MMRs when it sees missing data.

Split by year:

A graph with different colored squares

AI-generated content may be incorrect.A graph with blue and orange bars

AI-generated content may be incorrect.

* These plots are for the random forest ensemble model, w/ data was split by year.
* In contrast to the previous plot, this shows high error for all income levels, likely due to the increased difficulty of predicting the future (test data is the ‘future’).
  + Strangely, the highest income countries have the highest predictive error
* This plot could explain the lack of large difference between the original and sensitivity models, as the original does not appear to do amazingly on any one income level.
  + This could be due to the lack of clustering for specific years, and how none of the clusters strongly overlap with the clusters per income level.
  + Additionally, as demonstrated by the following plot, there is not a strong trend in MMR over time for any of the income levels.

A graph of different colored dots

AI-generated content may be incorrect.

A diagram of a number of dots

AI-generated content may be incorrect.

* Similar to split by year data, the standard deviation of base estimators’ predictions increased as mean predicted MMR increased, again showing higher uncertainty about data sparse, high MMR regions.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Type of Ensemble Model** | **MSE** | **RMSE** | **MAE** | **Relative Error** | **R2** |
| Voting ensemble | 6436 | 80.22 | 26.87 | 0.41 | 0.796 |
| Elastic Net ensemble | 5408 | 73.54 | 25.03 | 0.41 | 0.829 |
| Random Forest Ensemble | 5134 | 71.65 | 23.65 | 0.37 | 0.838 |
| SVM ensemble | 8008 | 89.49 | 37.55 | 0.56 | 0.747 |

* Random forest ensemble was the best performing in all circumstances, but with a substantially greater relative error than when the data was split by country (6% versus 36%).
  + Again, this is likely due to lack of easily learnable trends over the years.

A graph with numbers and dots

AI-generated content may be incorrect.

A graph with numbers and dots

AI-generated content may be incorrect.

* As above, with trend of underestimating for high income countries

A graph with numbers and dots

AI-generated content may be incorrect.

Use of synthetic minority over-sampling:

* During the conference, someone asked about this technique during one of the Q&As. Should I use this to generate more samples from low-income countries?
* I’ve done a brief literature review, and all the methods I found were only related to applying this technique to non-missing data.
  + Is it worth it for me to keep looking, or mention this as a possible extension of the work?



Merged data

Collected between 1985 and 2018 and stored in the World Bank’s Gender Data Portal and the WHO’s Health Inequality Data Repository. The final merged dataset contains 2789 observations and 721 features and covers 172 countries.

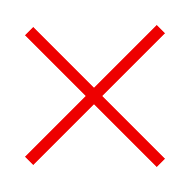
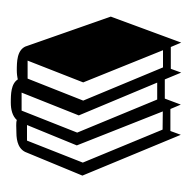
This single dataset will be modified and replicated many times to test different pre-processing techniques.

Feature selection

Created 5 versions of the dataset:

No feature selection, hand-picked features based on literature review, and features with pairwise correlation coefficient magnitude >0.6, >0.7, >0.8 respectively.







Split data into train, validation, test sets

Split each version of the dataset into train: test sets (90:10) and then used 5-fold cross validation to further split the train sets into train: validation (80:20) by splitting for predictive or missing data analysis.





**Predictive analysis:**

Training/validation 1985 -> 2014 Test: 2015 -> 2018.

**Missing data analysis:**

Split such that all data from the same country is in train, validation or test.

A row or column is removed if it contains greater than [85%, 90%, 95%, 100%] missing data. This creates 4 versions of each data fold.

Missing data removal

Training base estimators

Each of a Random Forest, LightGBM, and XGBoost model is trained on each fold, with their hyperparameters finetuned over 1,000 Optuna trials.

Ensemble models trained to combine 300 predictions from each model type (x3) trained on each feature subset (x5), each fold (x5), & each missing data threshold (x4).

Training ensemble models

The model’s performance was evaluated when it was only trained/tested on data from a specific income level as part of a sensitivity analysis.

The best model’s estimates were compared to estimates from the UN MMEIG, Global Burden of Disease Study, and GMatH microsimulation model.

Sensitivity analysis and comparison to literature

Tested voting ensemble, Elastic Net stacking ensemble, Random Forest ensemble, and Support Vector Machine ensemble.