# Bayesian Analysis of the United States' Foreign Aid Allocation Patterns

Kevin M. Nguyen

University of Colorado Boulder

STAT 5630: Computational Bayesian Statistics

**Bayesian Analysis of the United States' Foreign Aid Allocation Patterns**

Since the end of World War II, one of the primary goals of the global community has been international development. According to the African Studies Center at Boston University, International development is "a benevolent flow of resources and expertise from developed nations to developing nations."[1] Here developed nations, or the global north, include wealthy countries such as the United States, Germany, the United Kingdom, Japan, and France. Developing nations, or the global south, constitute poorer countries in Latin America, sub-Saharan Africa, Southeast Asia, and the Middle East. In simple terms, international development seeks to bridge the gap in wealth between the global north and the global south by promoting economic growth and reducing poverty in developing nations. By United Nations estimates, the wealth disparity between countries has decreased since 1990. However, problems of poverty and inequality at the micro level continue to be severe. Thus, international development remains an ongoing concern. United Nations analysis has found that the average income of North America is 16 times higher than that of sub-Saharan Africa.[2] Despite macro-level improvements over the past quarter century, international development remains ongoing. Developed countries are continually searching for methods to alleviate wealth disparity between nations.

In the post-war era, the global community prioritized two methods of international development: foreign direct investment (FDI) and foreign aid. FDI is foreign investment in local economies by private actors. Corporate involvement in developing countries can stimulate economic growth and bring job opportunities to underemployed areas.[3] While FDI allows businesses to participate in international development, contemporary research debates whether the net effect of FDI is positive or negative for developing countries. Critics view globalization as a means to exploit developing countries while cutting costs and subverting labor laws and high wages in developed economies.[4] Foreign aid, on the other hand, is a much less controversial tool for international development. Foreign aid is money sent from wealthy nations directly to governments in developing countries or through non-governmental organizations (NGOs) and inter-governmental organizations (IGOs) for humanitarian and economic development in the global south.

International development involves not only financial aid but also human and social development. Aid donors aim to empower individuals and governments in the global south to create sustainable and equitable societies. International development programs often include education, healthcare, infrastructure, social services, and economic aid. Although international development has brought significant benefits to many countries, it has faced criticism. Some argue modern international development practices promote Westernization and cultural imperialism and only reinforce global inequality rather than bolster equity. Others criticize the effectiveness and accountability of development programs and the unequal power dynamics between donor and recipient countries.[5] The future of international development requires innovative and collaborative approaches that address the root causes of poverty and inequality while respecting local cultures and values. International development also requires addressing pressing issues such as climate change and environmental sustainability. Sustainable development practices need to be integrated into development programs to mitigate the impact of climate change on vulnerable communities. Collaboration is critical to the success of development aid. Cooperation between governments, NGOs, IGOs, and local communities can generate productive development efforts. Developing countries often lack the resources or expertise to address complex economic issues.[6] This deficit underlines the necessity of collaboration for sustainable development practices. International cooperation can also help facilitate knowledge sharing between developed and developing nations. Knowledge sharing enables local communities to take agency in their development. Collaboration is an essential component in the creation of long-term, sustainable development outcomes.

Implementing Bayesian data analysis techniques in international development can revolutionize development outcomes by improving the efficiency and effectiveness of aid programs. Through such analysis, aid organizations can analyze vast amounts of data to identify and address the most significant factors affecting poverty and inequality in the developing world. Moreover, Bayesian computation can ease the monitoring and evaluation of aid programs, allowing organizations and watchdogs to assess program impact and make data-driven decisions about resource allocation. Although Bayesian analysis cannot alleviate all the complex challenges of international development, it can be a powerful tool for advancing the goal of creating more fair and sustainable societies globally.

## Empirical Strategy

Contemporary economic research on the determinates of foreign aid allocation has primarily relied on frequentist regression methods. For this analysis, I sought to apply two Bayesian computational methods: Bayesian regression and Bayesian neural networks. While frequentist regression is prevalent in economic research papers, Bayesian methods provide additional benefits to the standard regression model, discussed later in the section. Additionally, the focus on regression in this line of research has left an opening for further investigation into the efficacy of classification methods. Furthermore, classification methods have potential in practical applications of foreign aid allocation. Classifying prospective recipient nations based on past aid allocation can uncover patterns that allow donor governments to address current allocation patterns, increase transparency, and determine how to allocate future aid based on political and economic goals.

### Bayesian Linear Regression

Bayesian regression is an extension of the frequentist regression model that allows researchers to incorporate prior knowledge or beliefs about the relationship between the response and regressors in a model. Bayesian linear regression boasts many advantages over frequentist regression. These advantages include the inclusion of prior knowledge, flexibility in model specification, improved uncertainty quantification, and robustness to outliers. Incorporating prior information about model parameters can create models that generalize better to new data than frequentist regression models. These priors can come from previous studies, expert opinions, or a researcher's beliefs about the model parameters. Specifying prior distributions on parameters can lead to more accurate predictions and a better fit. Bayesian regression methods also benefit from better uncertainty quantification when compared to their frequentist counterpart. Bayesian credible intervals serve as a probability statement on the model parameter itself. Alternatively, frequentist confidence intervals are a statement of confidence in the interval construction process. This difference leads to better interpretability of the credible interval. Furthermore, Bayesian linear regression models can be more robust to outliers than frequentist regression methods since flexible prior distributions can be specified to account for extreme values.[7]

Bayesian linear regression is used in this paper to determine which economic, social, and political factors affect how the United States allocates foreign aid donations. Specifically, the regression model and equation are defined:

$$\boldsymbol{Y} \sim N(\boldsymbol{X\beta}, \sigma^2 \boldsymbol{I_n})$$

$$Y_i = \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \cdots + \beta_{16} x_{16,i} + \epsilon_i$$

where $Y$: $log$(foreign aid disbursements), $x_1$: log(net FDI), $x_2$: $log$(GDP), $x_3$: unemployment rate, $x_4$: $log$(refugee population), $x_5$: infant mortality rate, $x_6$: $log$(population), $x_7$:Freedom House political rights score, $x_8$:Freedom House civil liberties score, $x_9$: 2021 Corruption Perceptions Index, $x_{10}$: electoral democracy score, $x_{11}$: free speech score, $x_{12}$:Percent of adult population with sufferage, $x_{13}$:local government elections score, $x_{14}$:equal rights score, $x_{15}$: equal resources score, $x_{16}$:percent of territory owned by the state

The priors for the Bayesian regression are kept vague/uninformative. The specification of vague priors means that the estimates from the Bayesian linear regression will be close to the OLS estimates. However, the credible intervals resulting from the Bayesian regression are preferable to frequentist confidence intervals for the reasons explained above. As such, a multivariate Gaussian prior was selected for the vector of regression parameters so that the Bayesian estimates sit on the real number line. The inverse-Gamma prior is placed on the variance of the model so that the parameter estimate is strictly greater than zero.

$$\boldsymbol{\beta} \sim N(0, 100^2 \boldsymbol{I_{16}})$$

$$\sigma^2 \sim \text{Inv-Gamma}(1,1)$$

***No U-Turn Sampler (NUTS) in PyMC***

This project utilizes PyMC, "a probabilistic programming library in Python which users to build Bayesian models with a simple Python API and fit them using Markov chain Monte Carlo (MCMC) methods."[8] [9] PyMC, along with STAN, is one of the most popular probabilistic programming tools available. Bayesian linear regression in PyMC uses the No U-Turn Sampler (NUTS). NUTS is an MCMC algorithm introduced in 2011 that modifies the Hamiltonian Monte Carlo algorithm. Hoffman and Gelman designed NUTS to be more efficient than traditional MCMC algorithms by automatically tuning step size and total steps in each iteration, allowing for quicker exploration of the target distribution with fewer samples. Additionally, NUTS eliminates the need for manual tuning by the user. The algorithm terminates once it detects that the trajectory has looped back on itself, indicating sufficient exploration of the target distribution.[10]

**Bayesian Neural Networks**

Neural networks are a class of machine learning algorithms inspired by the structure and function of the human brain. Neural networks contain several interconnected processing units (called "neurons") that work together to produce an output. Each neuron in a neural network receives input from the previous layer, performs a computation on that input, then passes the result on the subsequent layer. The strength of connections between neurons is called weight. These weights are modified based on some gradient descent algorithm during the learning process.[11] Bayesian neural networks (BNNs) are an extension of the neural networks, incorporating Bayesian inference. While traditional neural networks treat the weights as parameters to optimize during gradient descent, BNNs treat the weights as random variables to model with Bayesian statistics. BNNs share many of the same benefits described for Bayesian regression, such as flexibility in model specification, improved uncertainty quantification, and robustness to outliers. Additionally, BNNs prevent overfitting to the data, a common issue with traditional neural networks. Since traditional neural networks do not have a mechanism for uncertainty quantification, BNNs are preferable to researchers who seek a probabilistic result associated with neural network estimates.[12]

### ***TensorFlow Probability and Variational Inference***

The BNNs constructed for this paper come from the TensorFlow Probability Python library, a probabilistic programming library built on top of TensorFlow. TensorFlow is a popular framework for artificial intelligence, deep learning, and neural networks. TensorFlow Probability allows for combining probabilistic models and deep learning algorithms, as present in BNNs.[13] While BNNs can implement MCMC in its backpropagation, computational complexity and sheer number of parameters to be estimated make variational inference a much more common choice. Variational inference is a technique used to approximate the posterior distribution over the network's weights. Variational inference approximates the posterior distribution with a simpler distribution, called the variational distribution. Variational inference is an optimization problem that minimizes Kullbeck-Leibler divergence between the variational distribution and posterior distribution with respect to the parameters of the variational distribution. Often, BNNs accomplish this optimization with stochastic gradient descent, like standard backpropagation.[14] Once the variational parameters are optimized, we can sample from the variational distribution for Bayesian inference.

## Data

The data for this project come from disparate sources. These sources include the World Bank, the V-dem Institute, Freedom House, Transparency International, and the United States Department of State. The primary goal of this research is to determine how the United States decides to allocate foreign aid. Moreover, this paper seeks to identify the exact factors that affect foreign aid allocation. I theorize that the amount of foreign aid sent to a developing country is a function of the donor's political interests, healthcare conditions, the refugee situation in the recipient country, and the recipient's level of economic need. As such, this project required collecting data that can serve as indicators for the above factors. First, data was pulled from the World Bank using the World Bank Indicators API. The world bank maintains a vast collection of economic data for each of the world's countries, dating back to the 1950s. As this paper aims to be a cross-sectional study, time series data available from the World Bank does not appear in this study. Additional research in this field may consider dynamically changing political and economic interests of donor nations with time series data. The World Bank data is limited to measurements (and estimates) from 2020 because of the completeness of the data in 2020 relative to other recent years. Six of the variables used for this analysis come from the World Bank Indicators Data. These variables include:

**'SL.UEM.TOTL.ZS'**: Unemployment Rate (%)

**'SM.POP.REFG.OR'**: Refugee Population

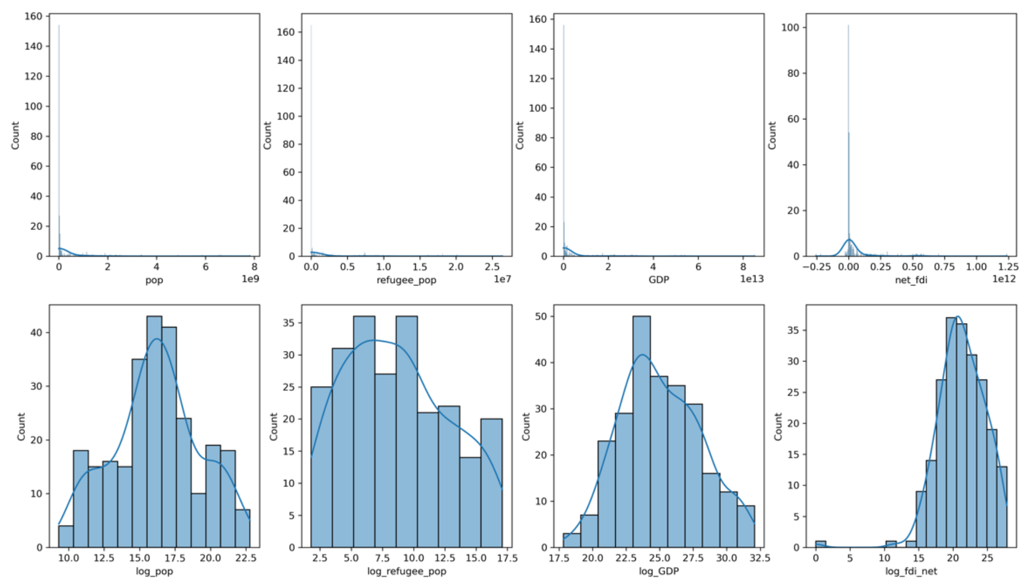
**'SP.POP.TOTL'**: Total Population

**'NY.GDP.MKTP.CD':** Nominal GDP (US Dollars)

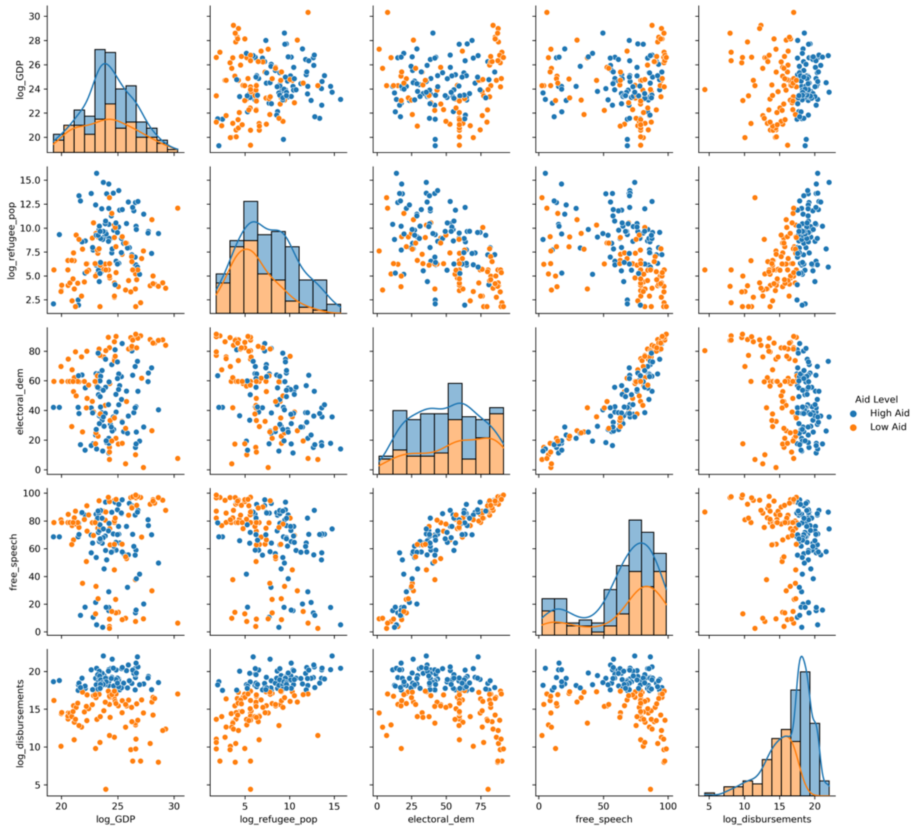
**'SP.DYN.IMRT.IN':** Infant Mortality Rate (%)

**'BX.KLT.DINV.CD.WD':** Net FDI

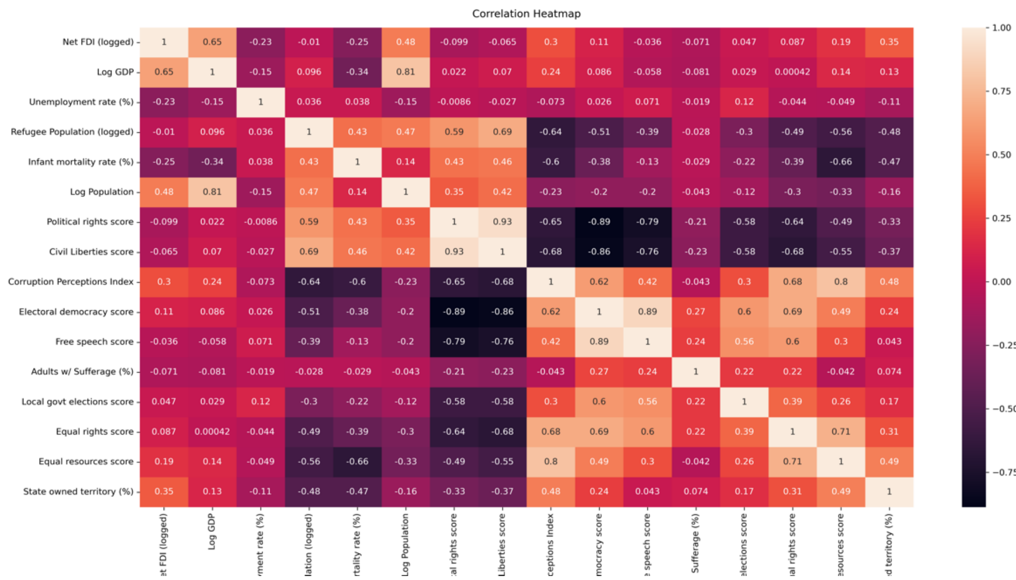
Preliminary exploratory data analysis found that four variables collected from the World Bank were highly skewed: total population, refugee population, Net FDI, and GDP. As such, log transformations were appropriate for the purposes of this research. This type of transformation affects the interpretation of the regression parameters, as noted in the results section.




In addition to the economic variables collected from the World Bank, this research suggests that political indicators influence aid allocation. The political variables in this research primarily come from the V-Dem Institute in CSV format. The V-dem Institute provides scores for world nations in many political areas, including these variables used in this research: electoral democracy, freedom of speech, local government elections, equal rights, and equal resources. These scores range from 0-1, with '1' being the "best." For the purposes of this analysis, these scores are scaled by a factor of 100. The V-dem dataset comes in a time series format, much like the World Bank data. Only scores from 2020 are present in this analysis.




Freedom House is a non-profit organization that scores countries based on political rights and civil liberties. Freedom House scores these countries on a scale of 1-7, with lower scores indicating more political rights/civil liberties. Similarly, Transparency International is a non-profit organization that published the annual Corruptions Perceptions Index (CPI). The CPI scores countries from 0 to 100 based on corruption levels within that country. Higher CPI scores indicate lower levels of corruption. Finally, the US Department of State records the United States' annual outbound aid by the recipient country. This aid dataset includes Presidential allocation, Congressional allocation, Obligations, and Disbursements. The log of the total disbursements serves as the target variable for the Bayesian linear regression. Furthermore, I discretized this variable into: "High Aid" and "Low Aid," split along the median of disbursements for BNN classification.

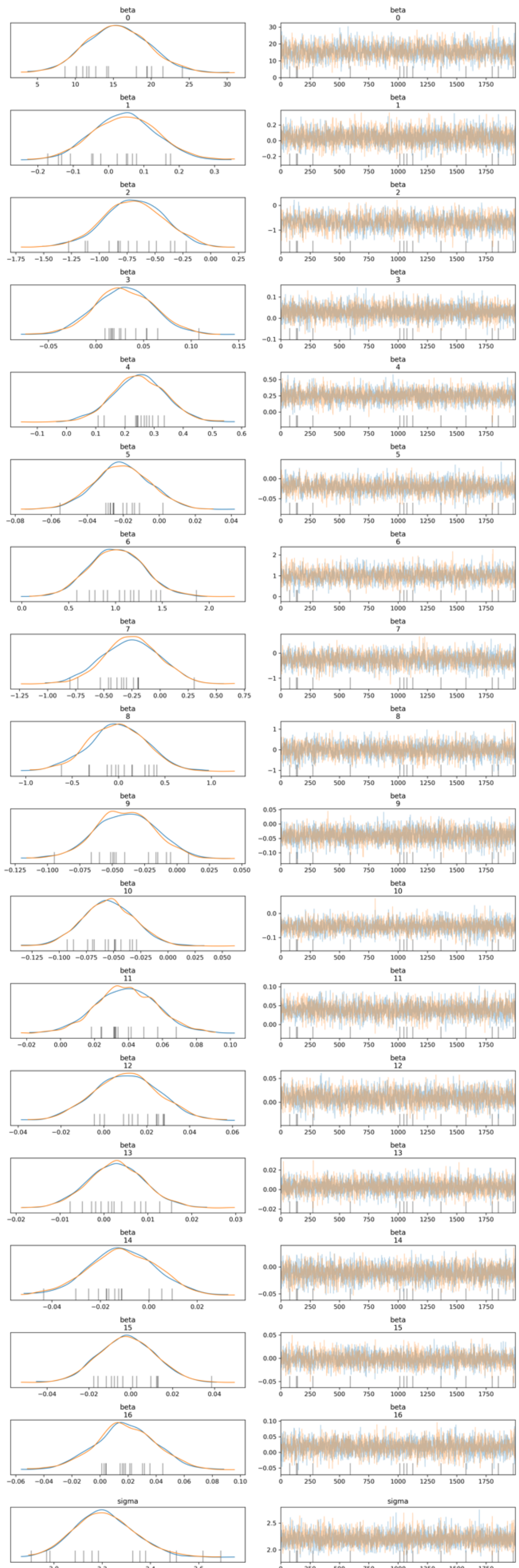



## Empirical Results

### Bayesian Linear Regression Results

The results from the Bayesian linear regression can be found in the table below.

| | mean | sd | hdi_3% | hdi_97% | mcse_mean | mcse_sd | ess_bulk | ess_tail | r_hat |
|---|---|---|---|---|---|---|---|---|---|
| **beta[0]** | 14.902 | 3.938 | 7.574 | 22.604 | 0.085 | 0.061 | 2133.0 | 2565.0 | 1.0 |
| **beta[1]** | 0.045 | 0.093 | -0.129 | 0.219 | 0.001 | 0.001 | 4259.0 | 3215.0 | 1.0 |
| **beta[2]** | -0.665 | 0.271 | -1.161 | -0.141 | 0.007 | 0.005 | 1713.0 | 2308.0 | 1.0 |
| **beta[3]** | 0.031 | 0.031 | -0.032 | 0.086 | 0.000 | 0.000 | 4970.0 | 2723.0 | 1.0 |
| **beta[4]** | 0.254 | 0.091 | 0.092 | 0.433 | 0.002 | 0.001 | 3496.0 | 2732.0 | 1.0 |
| **beta[5]** | -0.019 | 0.015 | -0.045 | 0.011 | 0.000 | 0.000 | 2335.0 | 2667.0 | 1.0 |
| **beta[6]** | 0.995 | 0.289 | 0.447 | 1.537 | 0.007 | 0.005 | 1848.0 | 2670.0 | 1.0 |
| **beta[7]** | -0.259 | 0.259 | -0.733 | 0.249 | 0.005 | 0.004 | 2954.0 | 2724.0 | 1.0 |
| **beta[8]** | -0.001 | 0.315 | -0.588 | 0.596 | 0.005 | 0.005 | 3416.0 | 2914.0 | 1.0 |
| **beta[9]** | -0.039 | 0.023 | -0.083 | 0.004 | 0.000 | 0.000 | 2309.0 | 2666.0 | 1.0 |
| **beta[10]** | -0.056 | 0.025 | -0.101 | -0.010 | 0.001 | 0.000 | 2233.0 | 2710.0 | 1.0 |
| **beta[11]** | 0.040 | 0.018 | 0.005 | 0.071 | 0.000 | 0.000 | 2149.0 | 2490.0 | 1.0 |
| **beta[12]** | 0.012 | 0.016 | -0.017 | 0.041 | 0.000 | 0.000 | 3554.0 | 2680.0 | 1.0 |
| **beta[13]** | 0.003 | 0.007 | -0.009 | 0.015 | 0.000 | 0.000 | 4504.0 | 2991.0 | 1.0 |
| **beta[14]** | -0.011 | 0.013 | -0.037 | 0.014 | 0.000 | 0.000 | 3348.0 | 3077.0 | 1.0 |
| **beta[15]** | -0.002 | 0.013 | -0.026 | 0.023 | 0.000 | 0.000 | 2618.0 | 2832.0 | 1.0 |
| **beta[16]** | 0.019 | 0.022 | -0.022 | 0.060 | 0.000 | 0.000 | 3273.0 | 2909.0 | 1.0 |
| **sigma** | 2.211 | 0.123 | 1.986 | 2.443 | 0.002 | 0.001 | 3728.0 | 2457.0 | 1.0 |



The NUTS algorithm sampled from the posterior distribution for each parameter with two concurrent chains. As one can see from the trace plots, each of the 34 chains (17 parameters x 2 chains) appears to have converged to the correct posterior distribution. However, not all these estimates are "statistically significant" in the Bayesian context. In frequentist regression, the p-value produced from the linear regression determines the statistical significance of a parameter estimate. That p-value is critical for hypothesis testing, with the null hypothesis being that the true parameter value is zero (the variable has no effect on the response). In Bayesian regression, we can determine if a parameter estimate is "statistically significant" if the 95% credible interval does not contain zero. In this case, six of our 15 variables are deemed significant: log GDP, log refugee population, log population, electoral democracy score, and free speech. The interpretation of the Bayesian regression estimates for the parameters associated with these variables is as follows. Recall that the response, total aid disbursements, was transformed to a log-scale and the V-dem electoral democracy and free speech scores were multiplied by 100.

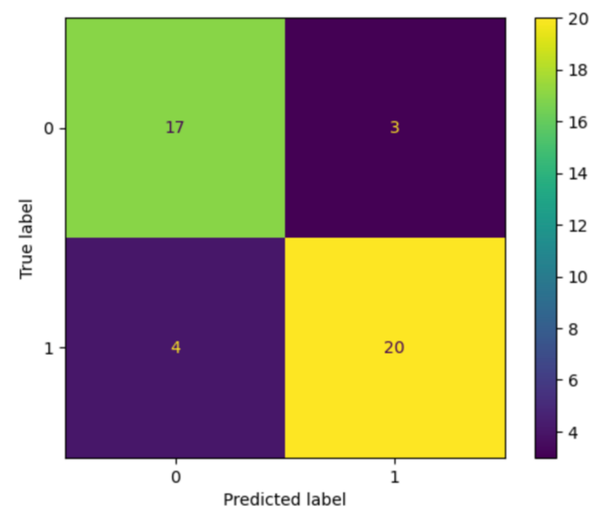
- A 1% increase in GDP is associated with a 69% decrease in aid disbursements.
- A 1% increase in population is associated with a 25.4% increase in aid disbursements.
- A 1% increase in the refugee population is associated with a 99.5% increase in aid disbursements.
- A 0.01-point increase in electoral democracy score is associated with a 5.6% decrease in aid disbursements.
- A 0.01-point increase in free speech score is associated with a 4% increase in aid disbursements.

**BNN Results**

For the Bayesian neural networks, an isotropic Gaussian prior is placed over the weights in the network, as is standard practice for modern BNNs.[15] Additionally, the BNN in this research consists of three hidden layers, each containing 24, 12, and 6 units. A batch normalization layer precedes these hidden layers to normalize the input data. The SoftMax function activates the final output layer to provide a probability associated with each class. The final prediction is the maximum of those class probabilities for each data point. The loss is set to 'Binary Cross entropy' with an Adam optimizer. With these specifications, the model contains 142,617 trainable parameters. The prediction accuracy of the testing dataset

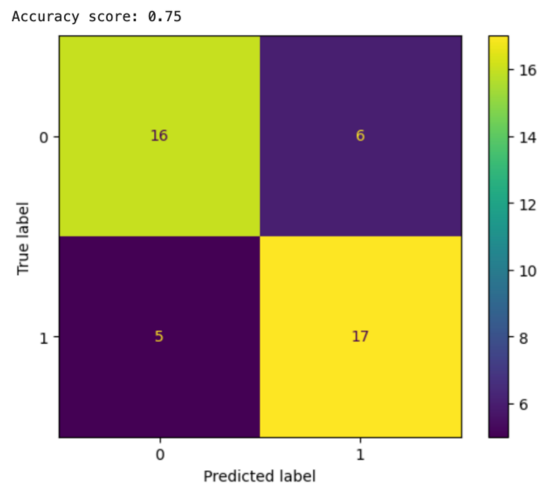
```
Model: "model_3"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 input_4 (InputLayer)        [(None, 17)]              0

 batch_normalization_3 (Batc  (None, 17)               68
 hNormalization)

 dense_variational_9 (DenseV  (None, 24)               93960
 ariational)

 dense_variational_10 (Dense  (None, 12)               45450
 Variational)

 dense_variational_11 (Dense  (None, 6)                3159
 Variational)

 dense_3 (Dense)             (None, 2)                 14

=================================================================
Total params: 142,651
Trainable params: 142,617
Non-trainable params: 34
_________________________________________________________________
2/2 [==============================] - 0s 41ms/step
```

Accuracy score: 0.84




For comparison, I constructed a traditional deep neural network with similar specifications. This neural network also contained three hidden layers with 24, 12, and 6 units apiece. These layers also come after a batch normalization layer. Furthermore, the artificial neural network (ANN) also had a loss set to Binary Cross entropy with an Adam optimizer. The main difference, of course, is the optimization of the ANN is performing stochastic gradient descent on the gradient of the loss function with respect to the weights rather than the gradient of the Kullbeck-Leibler Divergence with respect to the parameters of the variational distribution. Both neural networks trained on the testing set for 2000 epochs. As we can see below, the testing accuracy of the traditional ANN is only 75%, suggesting that the BNN is less prone to overfitting the training data. We can also note that the number of trainable parameters for the ANN is only 858, much less than that of the BNN.

```
Model: "sequential_6"
_________________________________________________________________
 Layer (type)                Output Shape              Param #
=================================================================
 batch_normalization_8 (Batc  (None, 17)               68
 hNormalization)

 dense_28 (Dense)            (None, 24)                432

 dense_29 (Dense)            (None, 12)                300

 dense_30 (Dense)            (None, 6)                 78

 dense_31 (Dense)            (None, 2)                 14

=================================================================
Total params: 892
Trainable params: 858
Non-trainable params: 34
_________________________________________________________________
2/2 [==============================] - 0s 29ms/step
```




## Conclusion

The Bayesian linear regression results confirm my initial hypothesis, donor governments (specifically the United States) consider political and economic interests in addition to need when deciding how to allocate foreign aid. Bayesian regression produced some expected results, such as the negative effect of increased wealth (GDP) and the positive impact of the refugee population on foreign aid disbursements. Additionally, the Bayesian regression also provided some fascinating results. Notably, United States aid allocation favors nations with more free speech and disfavors nations with democratic electoral systems.

The Bayesian neural network demonstrated the advantages of Bayesian inference in deep learning and classification problems generally. One key consideration is the time complexity of BNNs. Traditional neural networks are already computationally expensive relative to other classification methods, but the added Bayesian inference in backpropagation further complicates the training process. Although BNNs outperformed traditional ANNs in this paper, researchers should only consider BNNs as an alternative to standard machine learning classification when computation costs are a non-factor.

In conclusion, analysis of foreign aid allocation patterns remains a fascinating area of research, held back by the prominence of frequentist methods in economic and political science research. The results of this paper demonstrate the efficacy and interpretability of Bayesian methods in this field. Moreover, Bayesian methods have a place outside of academia. Aid organizations, donor governments, and recipient governments can leverage Bayesian linear regression and BNNs to uncover aid allocation patterns and determine the factors that can stimulate aid donation.

## References

[1] African American Studies Center, Pardee School of Global Studies, Boston University. *International Development.* *https://www.bu.edu/africa/outreach/teachingresources/international-development/*

[2] United Nations. *Inequality – Bridging the Divide.* *https://www.un.org/en/un75/inequality-bridging-divide*

[3] Xinxin Wang, Zeshui Xu, Yong Qin & Marinko Skare (2022) Foreign direct investment and economic growth: a dynamic study of measurement approaches and results, *Economic Research-Ekonomska Istraživanja*, 35:1, 1011-1034.

[4] Kim, M. (2015). Re-conceptualising exploitative and explorative FDI: A balancing-process approach to firm internationalisation. *European journal of international management : EJIM, 9*(5).

[5] Hasselskog, M. (2022). What happened to the focus on the aid relationship in the ownership discussion?. *World Development*, *155*, 105896.

[6] United Nations. *Economic and Social Challenges and Opportunities*. https://www.un.org/development/desa/en/wp-content/uploads/2020/07/RECOVER_BETTER_0722-1.pdf

[7] Bishop, C., & Tipping, M. (2003). Bayesian regression and classification. *Nato Science Series sub Series III Computer And Systems Sciences* 190: 267-288.

[8] Patil, A., Huard, D., & Fonnesbeck, C. J. (2010). PyMC: Bayesian stochastic modelling in Python. *Journal of statistical software*, *35*(4), 1.

[9] PyMC. https://www.pymc.io/welcome.html

[10] Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: adaptively setting path lengths in Hamiltonian Monte Carlo. *J. Mach. Learn. Res.*, *15*(1), 1593-1623.

[11] Lawrence, J. (1993). *Introduction to neural networks*. California Scientific Software.

[12] Jospin, L. V., Laga, H., Boussaid, F., Buntine, W., & Bennamoun, M. (2022). Hands-on Bayesian neural networks—A tutorial for deep learning users. *IEEE Computational Intelligence Magazine*, *17*(2), 29-48.

[13] TensorFlow Probability. https://www.tensorflow.org/probability

[14] Blei, D. M., Kucukelbir, A., & McAuliffe, J. D. (2017). Variational inference: A review for statisticians. *Journal of the American statistical Association*, *112*(518), 859-877.

[15] Fortuin, V., Garriga-Alonso, A., Ober, S. W., Wenzel, F., Rätsch, G., Turner, R. E., ... & Aitchison, L. (2021). Bayesian neural network priors revisited. *arXiv preprint arXiv:2102.06571*.