Team Control Number

52815

Problem Chosen

C

For office use only
T1 ________________
T2 ________________
T3 ________________
T4 ________________

For office use only
F1 ________________
F2 ________________
F3 ________________
F4 ________________

2016
MCM/ICM
Summary Sheet

(Your team's summary should be included as the first page of your electronic submission.)

Type a summary of your results on this page. Do not include the name of your school, advisor, or team members on this page.

We propose a Bayesian multi-stage model to maximize Return on Investment (RI) over the next five years.

First, a model was made to relate the distribution of students' majors at a school to the expected number of jobs that that school will add to to the economy. Here, we relate the unemployment outlook in each particular discipline to the percentage of students at each school pursuing that particular course of study. In the case that a school was going to produce a high number of graduates in a field that is already oversaturated (that is, has a high unemployment rate), it is actually assigned a negative contribution to the number of jobs in the economy.

Next, a model was fit to give the relationship between jobs added to the economy and change in real Gross Domestic Product (GDP). This was based on a linear relationship we observed between changes in logGDP and changes in unemployment, fit to data gathered beginning in 1948. This particular model has a surprisingly good fit considering the field.

Each school was given a Goodness rating based on added GDP and the median salary of its alumni. Salary was simply scaled and added to GDP in order to compute this metric.

The centerpiece of our proposal is a linear regression model relating Goodness to a combination of the school's tuition and enrollment. This model is meant to contain general information about how a school's resources impact its Goodness.

The posteriors of this overall model are used as priors for the parameters of a series of models (these priors are injected with additional variance so as not to overwhelm the data), each one fit to a particular school. Each of these models represents a predicted relationship between a school's available funds per student and its Goodness, as well as how a change in its funds would affect its Goodness, which we call a Goodness curve.

We model the investment of money in a school as increasing the "effective tuition" of that school, which will increase the Goodness of that school according to its Goodness curve. This change in Goodness is our proposal for measuring RoI.

Given this definition of RoI, we are presented with the question of where investments should be disbursed such that the sum of RoI is maximized: a simple optimization problem which we solve in order to obtain our results for proposal.

Such curves are calculated for each year from 2016 to 2020, based on econometrics forecasts in US GDP changes and population growth.

Each single stage of our model is simple, and the model in its entirety is complex rather than complicated. This leads to a model which, while far from trivial, is easy to interpret.

We present 73 schools to be funded based on our analysis. Among them are such well known schools as Princeton, Harvard, and MIT, along with a collection of other, lesser well known schools, which are expected to be as important of investment targets for the American economy according to our model. We also present the time periods during which we believe funding would be most effective.

Data manipulation was performed in R and Python. BUGS was used for Bayesian inference on the model parameters, and R was used to solve the concomitant optimization problem.

Mr. Alpha Chiang :

We are writing to you regarding the investment strategy of the Goodgrant Foundation. We have created a sufficient model for dispersing the donations to colleges in a way that would enhance the Return on Investment (RoI) for Goodgrant and thus the benefits for the impacted students. This model was developed due in large part to the analysis of data provided by the U.S. National Center on Education Statistics. Considering the lack of an official RoI measurement regarding the funding of schools, we opted to create our own before model creation. This measurement was a function of what we called Goodness obtained by and resources available to a given school. Goodness was assessed as a school's contribution to the U.S. Gross Domestic Product (GDP). Resources were defined as a school's annual income from tuition and external funds. This RoI was optimized by the model's simulations.

Jobs added by a given school were calculated by considering the unemployment rates for each major and the number of undergraduate degrees awarded by a school. This was compared to the unemployment rate of those with bachelor's degrees in the respective majors across the United States. A positive relationship was then found between the amount of jobs added and the change in GDP. Therefore, schools were assigned high goodness if they awarded a large amount of degrees in fields that were undersaturated in the U.S. economy.

The $500,000,000 that Goodgrant plans to donate is disbursed on a yearly basis in the analysis. The simulations determine which schools to invest in and how much to invest in each school. The time duration over which the money should be invested is also included. These durations were a result of forecasted GDP and population data over the next five years.

We first computed an overall model by looking at the relationship between resources and goodness for all the schools on which we had data. Then, using this first model as a starting point, we developed individual models for each school. That is to say, we have an individual relationship between how much funding each school receives and how much it's able to contribute to the economy based on that income, as well as a custom estimation for the rate at which decreasing marginal returns occur for each school.

Based on these data, we estimated RoI over time by comparing the change in resources to the change in Goodness as a result of each respective investment. Then, we ran a number of simulations to allocate the individual investments that maximized the RoI. These investments are allocated until the overall donation amount is exhausted, leaving us with a plan to best disburse the grants every year.

After optimizing RoI, 73 schools were given at least some sort of investment. Most were small, possibly underdeveloped schools that would benefit greatly from a private contribution. There are also a few highly developed schools that were recommended for investment such as Princeton. The yearly investments in these schools are provided by the model.

The recommended investments output from the model optimize an acceptable metric for gauging the added value to the invested schools. Multiple models were compared to one another to arrive at a concluded model for the recommended investments. We are confident that this model will realize its full potential.

Sincerely,

MCM Team Members

A Multilevel Bayesian Model based on Economic Contribution of University Alumni

February 1, 2016

Contents

1 Introduction

Starting in July 2016, the Goodgrant Foundation will provide an overall donation amount of $500,000,000 over the course of 5 years ($100,000,000 per year) to various schools. This donation amount will be allotted to certain schools based on the school's return on investment (RoI). A school's projected RoI will determine whether or not the school will receive any investment whatsoever, how much money the school will receive, and the time duration over which the school will receive the money. Solid investments will have a positive effect on overall student performance.

While it is acknowledged that, in general, investment in tertiary education has lower RoI than investment in primary education, there are still important returns to expanding higher education, especially in developed countries such as the United States [24]. In particular, we will show the impact that investment in universities can have on GDP. Investing money in the development of skilled workers is not only beneficial to the economy, it is necessary for the US to compete globally. We hope that this document will help to show how best that can be achieved.

Return on Investment was calculated as a function of the school's "goodness" and the school's resources. Goodness and resources were determined using the provided data sets and external data sets. Before review, the data required various preparation techniques.

2 Data Cleaning and Imputation Schema

Any serious statistical analysis begins with substantial work on the data. The data provided had a panoply of issues ranging from unnatural variables, structural and seemingly random missingness, and impossible observations. Their rectification was prerequisite to model fitting and analysis.

2.1 Data Gathering, Cleaning and Filtering

For our approach to evaluating RoI, the gathering of unemployment measurements for each major that was referred to in the given data sets was required. Multiple sources were referred to in collecting these measurements [6, 12, 19, 25, 26]. It was also necessary to obtain average starting salary for each major [8, 9, 13, 14, 15, 16, 17, 20].

Initial data cleaning was performed in Python, where we re-organized variables into a more natural structure (for example, there were two tuition variables, one for public and one for private schools, as opposed to one tuition variable and one private/public flag). Certain schools were found to have negative average tuition. We considered this to be measurement error, and removed from consideration for funding any school reporting negative average tuition.

Next, we analyzed the missingness structure in R. The MI package (short for Multiple Imputation) [23] has many useful tools for dealing with missing data problems. Among these is a heatmap of missing cases and observations clustered by missingness structure. Figure 1 shows the initial missingness structure.

In Figure 1, the x axis represents cases, the y variables (the names of which are not intended to be legible). There are a number of schools which are missing the majority of variables, shown at the top of the missingness image. These schools were not considered for funding due to the lack of information, and were removed from the analysis. There is a large block on the left, representing the lack of SAT or ACT scores for a majority of schools. These variables were not included in our analysis because of this lack of data for most schools. On the right, a strange missingness pattern is observed, but this is due to the factors applying

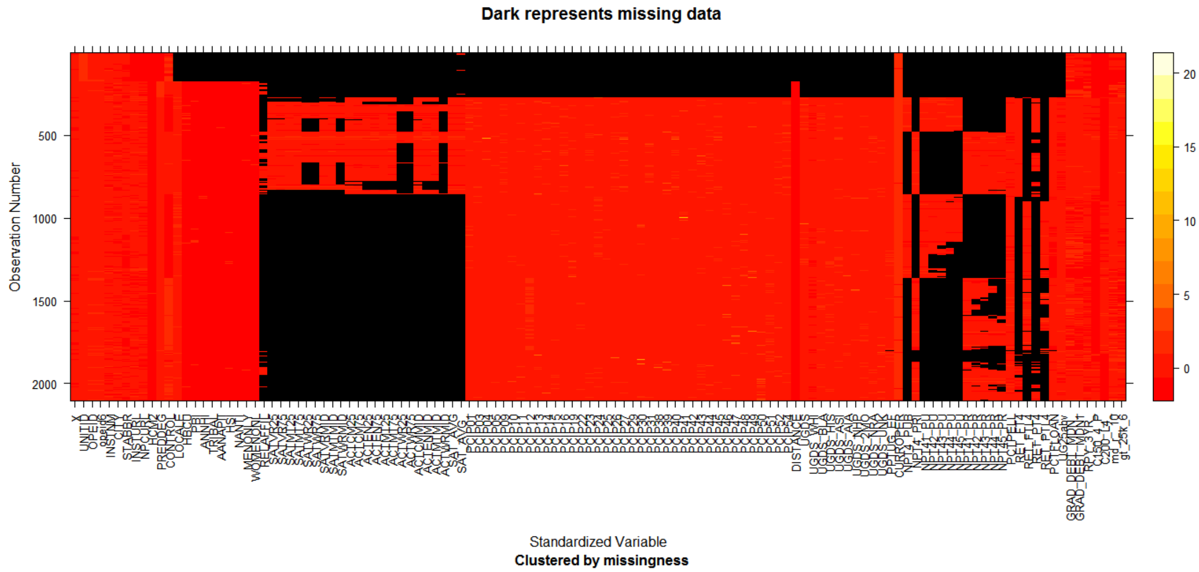

Figure 1: Initial Missingness Structure of Data

either to public or private schools and is rectified in the python script. Any further missingness was handled by our imputation schema.

2.2 Imputation Schema

We identified two distinct kinds of missingness in the data, and modeled it with two different mechanisms.

Firstly, data simply missing from the model were assumed to be Missing Completely at Random (MCAR). BUGS treats missing values as parameters and conducts inference on them automatically. As we are acting as Bayesians, we must specify priors for them, and choose to use a very disperse normal distribution centered at zero.

We were not comfortable making this same assumption with the "Privacy Suppressed" Data. In reporting the data, aggregators are required to censor data if they believe it may be possible to make reasonable inferences on personal data of any individual. Oftentimes, this is the case if the data represents an aggregation from very few people. We believe that schools with few people may have different characterisitcs than schools in general, and so these were not assumed to be MCAR, but Missing at Random (MAR), that is, whether or not they are missing depends on other variables in our model. These data were estimated using a multiple regression model depending on the complete data [21].

3 Model Components

There were two means of assessing how contributions to a school would increase its goodness. The first was a measurement of how well the school is doing at satisfying the nation's needs with respect to the skills of its graduated students. The second was simply a function of the median salary for that school.

The resources of a given school were calculated by estimating the annual tuition income and the annual funds added externally to the school.

3.1 Model Overview

The data did not come with any kind of response; there was no measurement of how deserving a school was of money. As such, we had to create our own, which we will refer to as Goodness. Return on Investment was defined as a change in Goodness. We hypothesized that there would be a positive, concave down relationship between the amount of resources available to a school and its Goodness. The relationship was expected to be positive because, in general, a school with access to more money is expected to be able to produce better results, and concave down due to the law of diminishing marginal returns.

Specifically, resources are defined as the tuition per student, as well as the total enrollment for the school, and Goodness is a linear combination of a school's contribution to GDP (defined below) and the median salary of its alumni. We first computed an overall model:

$$Goodness = ln(Tuition) + Enrollment + \epsilon$$

$$\epsilon\ N(0, \tau I)$$

Next, we computed an individual model for each school. The posterior for the model described above was given extra uncertainty (variance), and was used as a prior for each school [1].

$$P(M_s|X) \propto L(M_s|X_s)P(M_s|X)$$

This gave us an estimated Goodness or RoI Curve for each school. Given these curves, the idea was to maximize RoI through a change in per capita dollar amount, theoretically added to each school's resources to change its effective tuition. A more precise overview of how schools were chose based on this model may be found in the optimization section.

3.2 Contribution to GDP

The model related the proportion of students in each course of study to GDP in a several step process. The first step was to identify how many jobs each school is expected to add to the economy. This was calculated based on the unemployment in the fields which a university was preparing students to work in. For example, if there is higher than usual unemployment in Psychology, we would rate colleges producing many Psychology majors as producing negative jobs (which would lead to an increase in unemployment). On the other hand, if there is lower than average unemployment in Computer Science, schools with many computer scientists would be rated as producing many jobs. The calculations were conducted as follows ($DegreesAwarded_\pi$ is the proportion of degrees awarded, $\overline{Unemployment}$ is the mean unemployment of all majors, $Unemployment_\sigma$ is the standard deviation of employment for all majors):

$$Students = DegreesAwarded_\pi(Enrollment)$$

$$Unemployment_{\delta} = \frac{Unemployment - \overline{Unemployment}}{Unemployment_{\sigma}}$$

$$JobsAdded = Students(DegreesAwarded_{\pi})(Unemployment_{\delta})$$

If a school was found to have a net decrease in jobs, it was removed from consideration for funding.

We next map jobs added to the economy to a change in the unemployment rate for each major. This is simply defined as follows:

$$CurrentJobs = LaborForce(1 - Unemployment_{initial})$$

$$Unemployment_{final} = 1 - \frac{CurrentJobs + JobsAdded}{LaborForce}$$

$$Unemployment_{\Delta} = Unemployment_{final} - Unemployment_{initial}$$

It is well known that there is a relationship between real gross domestic product (GDP) and unemployment [4], and there exist complicated economic models relating the two [22]. However, we chose to fit our own simple model to unemployment and GDP data spanning from 1948 to 2015 obtained from the Bureau of Economic Analysis [11] and the Bureau of Labor Statistics [2], respectively, for use in our greater model. We are not directly concerned with the relationship between GDP and unemployment, but with the relationship between a change in GDP and a change in unemployment. As such our model is as such:

$$ln(\Delta GDP) = \Delta Unemployment + \epsilon$$

$$\epsilon \ N(0, \sigma I)$$

We use the natural log of a change in GDP because while reasonable values for GDP have greatly increased since 1948, reasonable values for unemployment have remained about the same. So our model predicts not on an absolute change in GDP, but a percent change in GDP (when back-transformed from log space).

$$ln(\Delta GDP) = ln(GDP_{final}) - ln(GDP_{initial})$$

$$ln(\Delta GDP) = ln(GDP_{final}/GDP_{initial})$$

$$e^{ln(\Delta GDP)} = GDP_{final}/GDP_{initial}$$

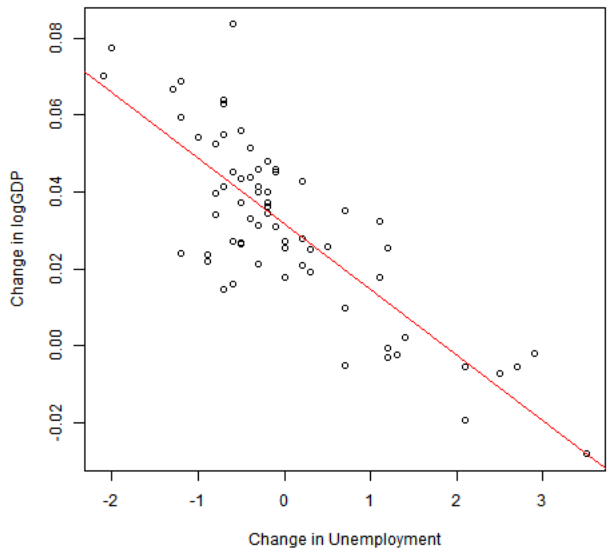

Figure 2: Regression of Δ logGDP vs Δ Unemployment

But in our model, we want an absolute change in GDP. We obtain this using the following formula:

$$\Delta GDP = e^{MLE} GDP_{initial} - GDP_{initial}$$

$$\Delta GDP = e^{ln(GDP_{final}/GDP_{initial})} GDP_{initial} - GDP_{initial}$$

$$\Delta GDP = \frac{GDP_{final}}{GDP_{initial}} GDP_{initial} - GDP_{initial}$$

$$\Delta GDP = GDP_{final} - GDP_{initial}$$

The model fit looks surprisingly good. See in Figure 2 a scatterplot of the chart with regression line plotted.

This negative relationship was manipulated to evaluate the goodness that would be added by a possible Goodgrant contribution. The jobs that are added to a particular field decrease the overall unemployment rate and thus increase the GDP. While keeping resources constant, an increase in GDP leads to a positive ROI.

3.3 Final Cross Sectional Model Specification

Once the data were manipulated as desired in Python and R, BUGS was called from within R to sample from the posterior distribution. The model as specified to BUGS was as such:

$$Goodness = \beta_0 + \beta_1 ln(Price) + \beta_2 Enrollment + \beta_3 ln(Price) Enrollment + \epsilon$$

$$\epsilon \sim N(0, \tau I)$$

$$\beta \sim N(0, 0.00001)$$

$$\tau \sim \Gamma(0.00001, 0.00001)$$

For Missing Completely at Random Data:

$$Salary \sim N(0, 0.00001)$$

For Missing at Random Data:

$$Salary = \vec{\gamma} X$$

Where X is a matrix of other covariates and γ is a vector of coefficients relating the two.

3.4 Funding Over Time

In order to answer the question of *when* these funds should be dispersed, we simply recomputed our model posteriors for different values of economic data. We used forecasted econometric [10] and population data [7] in our calculations. We assumed that the percent of the US population making up the Labor Force would stay constant over the next five years for our Job Added calculations. We did not inject any other different information into the model when computing the coefficients at different times.

3.5 Analysis of Linear Model Assumptions

We will in this section perform an analysis of the residuals of the regression model in order to determine whether any of the assumptions of linear models have been validated.

We first consider a normal Quantile-Quantile Plot (Figure 3). It does not look perfectly normal, contrary to the assumptions which lead to our posterior estimates, but it is close enough.

This predicted versus residual plot is more worrisome (Figure 4). We actually see a decrease in variance as the predicted value increases. This plot shows that our model is not perfect.

The error of the prediction seems to increase as the predicted value increases (Figure 5). This plot again serves as evidence against this model the most correct.

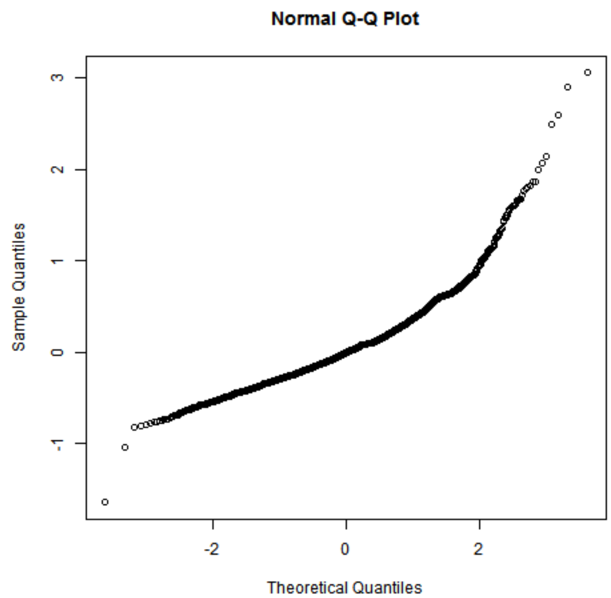

Figure 3: Q-Q Plot for Main Model

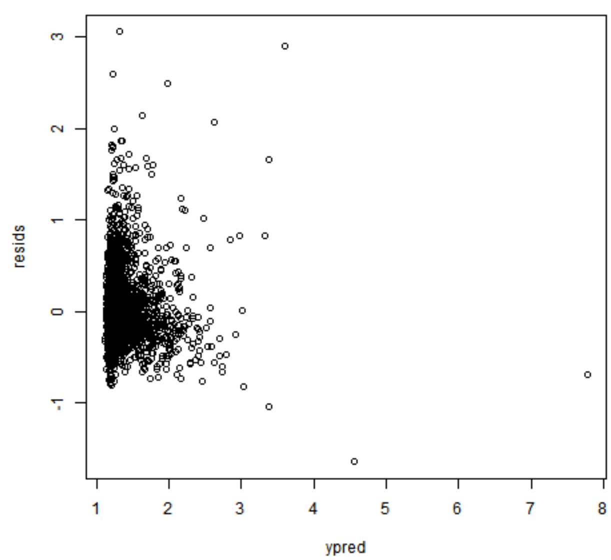

Figure 4: Residual vs Predicted Plot for Main Model

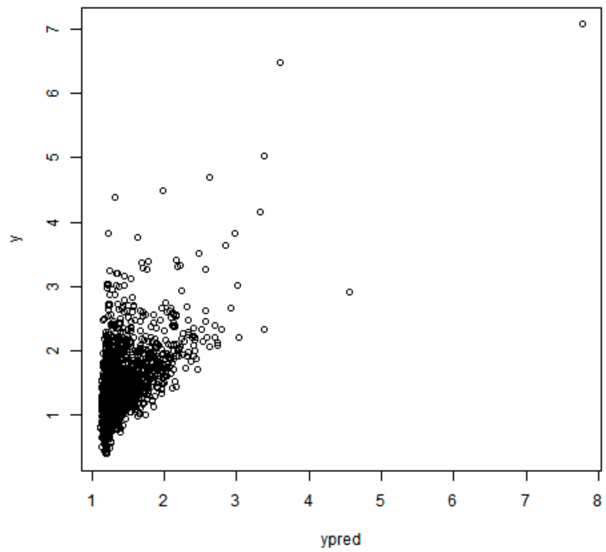

Figure 5: Actual vs Predicted Plot for Main Model

4 Solving the Optimization Problem

Our model as described above is only part of our solution. Given it, we are able to calculate the most beneficial distribution of available funds. How this is done will be described here.

4.1 Minimal Equivalent Example

Consider an example (see Figure 6) in which we have only two schools competing for Goodgrant's funds. School A has a coefficient of 1 to relate log of funding with Goodness, while School B has a coefficient of 0.2. Both graphs above show what the change in Goodness is for a disbursement of 1.3 thousand dollars to each school would be.

Obviously, the disbursement has a larger effect on the school on the left (School A), so we would choose to fund it over the school on the right (School B). This is essentially how our optimization problem works; we iteratively check which school would most benefit from a small disbursement until we have exhausted our funds.

The above example is just a simple case of our investments. When considering our real model, the problem becomes how much overall disbursement each school should be given from the Goodgrant Foundation to maximize RoI.

4.2 Full Optimization Scheme

Our approach for the full optimization problem is simply to iteratively calculate the *Goodness/Resource* slope for each school, and choosing to fund the school which resulted in the largest increase in Goodness on

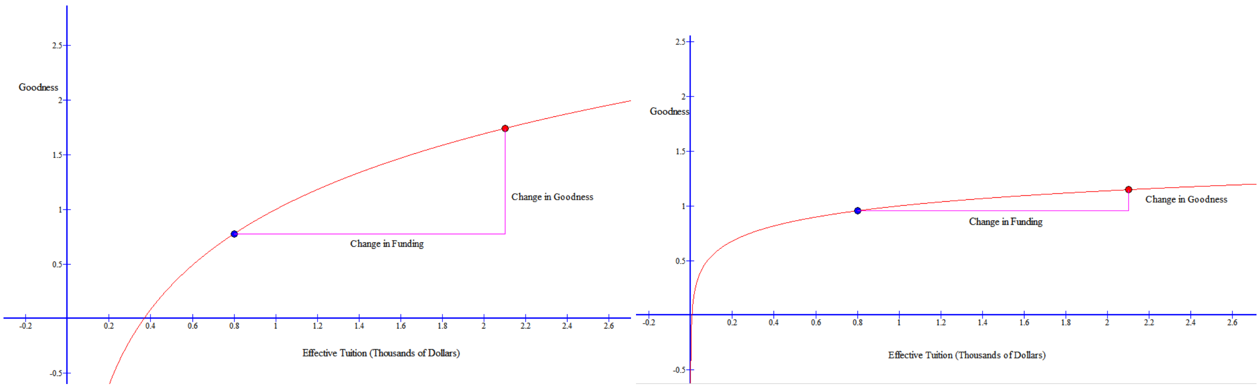

Figure 6: 2D Example of Optimizing Funding Distribution

each iteration until all our funds are depleted.

4.3 Algorithm

Below is the algorithm we have created for the optimization process. We denotes **r** as resources, **s** as number of students enrolled in a school, and $\boldsymbol{\beta}$ as the parameter matrix (recall that this is matrix as opposed to a vector because each school has its own model). The variable of interest, Goodness, is represented by **g**, and constant money distributed denotes as m

---

Algorithm 1

Initiate **r**, **s**, and m_0
Compute $g_{init} = \boldsymbol{\beta} * \mathbf{r}$
while $m_0 > 0$
——Compute $g_{next} = \boldsymbol{\beta} * (\mathbf{r} + \mathbf{1})$
——Compute $\boldsymbol{\Delta} = g_{next} - g_{init}$
——Find $where.max(\boldsymbol{\delta}) = i$
——Updates $m_k = m_0 - \mathbf{s}_i$
——Updates $g_{init} = g_{next}$
end

---

5 Problem Statement and Special Considerations

The concern of this paper is an allocation of resources problem. The goal is to maximize the impact of 100 million dollars injected into the US tertiary education system per year for the next five years.

We are to make a recommendation to the Goodgrant foundation for how to evaluate the impact of disbursed money on a school, to which schools to give money, how much to give to each school, and the expected return for each school.

Outside of analysis, various characteristics were considered when evaluating a school's worthiness of investment. These characteristics impacted our approaches to analysis. Among these characteristics are the external funds added. In general, public schools are given more external funds than are private schools. An estimated amount of twice that of private schools was used when calculating a public school's resources [5]. Further, community colleges' funding was multiplied by five. Justification for this is based on results from such models.

6 Model Analysis

6.1 Why Bayes?

Many times, Bayesian approaches can be more difficult computationally than frequentist ones. Further, the frequentist approach is the standard approach in statistics. As such, we feel it necessary to defend our use of full Bayesian methods. We chose to implement a Bayesian model for two main reasons.

Firstly, imputation of missing data is natural in a Bayesian model, both in theory and in implementation.

Second, it is important to maintain full accountability for uncertainty when we are applying functions with non-zero concavity to data in which we are not absolutely certain [27]. For instance, consider the exponential function relating Jobs Added an GDP. Because exponential functions are concave up, we would underestimate change in GDP if we did not account for our uncertainty.

6.2 Model Limitations

Because our gathered unemployment data relates only to undergraduate degrees, we only considered universities which primarily granted undergraduate degrees.

One of the model's most serious weaknesses is the lack of any time series data with respect to most of the model's inputs. As mentioned in Section 3.3, all data directly related to schools, as well as field-specific unemployment, is entirely cross-sectional, and time series estimates are a result of varying only national economic data.

Much data on school funding was estimated, oftentimes in an ad-hoc manner. Should reality differ significantly from our estimates, the model will be wrong.

6.3 Model Assumptions

As this is a linear model, it assumes that the relationships as described by the model equation are accurate. That is, that goodness is approximately a linear combination of the log of its tuition and its enrollment. The Section 3.5 shows that there are some issues with this assumption. As discussed in Model Limitations, this model assumes that much of its inputs stay constant over the next five years.

Unemployment data for different majors were not all found for the same year; as such, we make the assumption that unemployment rates have been fairly constant between and within fields during the last few years, and will remain so for the next five years.

Because there are so many missing variables, the correctness of the model depends strongly on the assumption that there is no significant structure in the missing data which cannot be accounted for in other known variables, that is to say, we must assume the data are not missing based on unobserved predictors[3].

6.4 Model Strengths

While there are many steps to go from inputs to outputs, each step individually is easy to understand, meaning that the model at each stage, as well as as a whole, is easily interpretable. In the same vein, mapping from a subset of the variables to a change in GDP is very easily interpretable as a form of Return on Investment.

Because we take a Bayesian approach, we maintain full accountability for uncertainty, as described in Section 6.1.

6.5 Potential Improvements

We believe there are many improvements which could be made to this model under different time constraints. The first would be a change from what is essentially a linear regression model to a dynamic linear model. It is possible to download historical data similar to those provided with the problem, which would allow such a model to be fit. However, significant data manipulation efforts would need to be undertaken.

The accuracy of the model would potentially increase with more resource factors to act as "controls", such as faculty:student ratios or more precise state and alumni funding estimates.

The model would also benefit from additional factors included in Goodness. In particular, we hypothesize that the inclusion of some measure of how much research is produced by a university would be very beneficial.

7 Results from Model

Table 1 shows the result from Bayesian Time-Series Model with constraints on the maximum disbursement to a particular school. Another table 2 is generated from Bayesian model with no constrains from how much a school should receive.

Table 1: Schools Investment Strategy with Max Investment Constraints

Schools	**Year One**	**Year Two**	**Year Three**	**Year Four**	**Year Five**	**Total**	**ROI**
Georgia Perimeter College	50017570	0	0	0	0	50017570	145.520
CUNY Bernard M Baruch College	0	14465088	21889404	13656906	0	50011398	92.565
University of Wisconsin Colleges	0	13181345	27852585	8976693	0	50010623	89.411
Harvard University	7903908	22066896	11198661	8835492	0	50004957	86.323
Princeton University	23615407	13446146	6746626	6191822	0	50000001	85.440
Bergen Community College	0	0	0	8254656	17071199	25325855	81.962
Moorpark College	0	0	3668301	8734050	8330940	20733291	81.962
United States Merchant Marine Academy	11182734	3809966	1880554	1816368	1008774	19698396	81.962
MCPHS University	0	0	0	9588544	5712000	15300544	81.962
Richland College	0	0	0	0	14668722	14668722	81.962
University of Puerto Rico-Arecibo	0	0	0	0	14623308	14623308	81.961
Massachusetts Institute of Technology	0	0	1573990	7328750	4167240	13069980	81.961
Shorter College	908490	0	10073910	0	0	10982400	81.961
University of Puerto Rico-Aguadilla	0	9492054	0	0	1292928	10784982	81.961
Stanford University	0	0	0	5702660	4446260	10148920	81.961
Tallahassee Community College	0	0	0	0	9725706	9725706	81.961
Leeward Community College	0	1358428	3096504	2645672	2331276	9431880	81.961
New England College of Business and Finance	0	8653068	0	0	0	8653068	81.961
University of Puerto Rico-Humacao	0	0	0	0	8166562	8166562	81.961
Baptist Memorial College of Health Sciences	1074576	2287878	1156050	1105404	642984	6266892	81.960
Albany College of Pharmacy and Health Sciences	0	0	2160750	2478950	1394275	6033975	81.960

Berea College	2031360	961722	477687	461817	260268	4192854	81.960
Universidad Pentecostal Mizpa	0	0	0	4042477	0	4042477	81.960
Christian Brothers University	337552	1513204	750244	723656	405756	3730412	81.960
Belanger School of Nursing	0	0	0	3687048	0	3687048	81.959
St Paul's School of Nursing-Staten Island	0	0	1220598	1282599	721602	3224799	81.959
Yeshivat Mikdash Melech	282112	0	497192	1673216	0	2452520	81.959
Machzikei Hadath Rabbinical College	0	2286008	0	0	0	2286008	81.959
Eastern West Virginia Community and Technical College	0	727872	1323008	0	0	2050880	81.959
Talmudical Seminary of Bobov	0	382522	221040	991303	0	1594865	81.958
Lincoln Trail College	805363	343252	172159	164697	94874	1580345	81.958
Doane College-Lincoln Grand Island and Master	0	110313	451346	431879	248024	1241562	81.958
Jewish Theological Seminary of America	0	0	0	33327	1190273	1223600	81.957
Crouse Hospital College of Nursing	0	287616	358176	345744	193200	1184736	81.957
Divine Word College	0	467072	675866	0	0	1142938	81.957
Curtis Institute of Music	757376	361730	0	0	0	1119106	81.957
Saint John Vianney College Seminary	335951	0	0	144298	359282	839531	81.956
Hobe Sound Bible College	0	0	0	0	677448	677448	81.955
Uta Mesivta of Kiryas Joel	0	0	675913	0	0	675913	81.955
Soka University of America	0	596576	0	0	0	596576	81.953
Olney Central College	0	115971	179800	169012	105183	569966	81.337
Leech Lake Tribal College	0	543660	0	0	0	543660	80.123
Bethesda University of California	0	0	0	0	525511	525511	78.548
College of the Marshall Islands	0	519688	0	0	0	519688	77.618
Cleary University	0	0	0	301500	213300	514800	68.546
Universidad Teologica del Caribe	0	273625	207756	0	0	481381	62.110
Yeshiva Derech Chaim	0	251576	0	0	225368	476944	59.620
Rabbinical Academy Mesivta Rabbi Chaim Berlin	137104	209428	0	114472	0	461004	56.118
Carolina Christian College	346140	0	95436	0	0	441576	51.497
Maple Springs Baptist Bible College and Seminary	0	0	438520	0	0	438520	51.485
Bais Medrash Toras Chesed	0	165072	0	0	179322	344394	45.115
Sh'or Yoshuv Rabbinical College	0	0	0	0	332310	332310	44.451
San Juan Bautista School of Medicine	13257	77382	197694	38394	0	326727	43.746
National Institute of Massotherapy	0	306234	0	0	0	306234	38.073
St Vincent's College	0	0	0	0	287956	287956	35.602
Rabbinical Seminary of America	0	0	0	0	281643	281643	31.962
Yeshiva College of the Nations Capital	90620	148419	0	25415	0	264454	30.920
Allegheny Wesleyan College	0	0	263588	0	0	263588	29.886
Boise Bible College	0	263337	0	0	0	263337	29.779
University of the West	0	232114	0	0	0	232114	23.546
Talmudical Academy-New Jersey	160480	0	0	0	10064	170544	22.952
Telshe Yeshiva-Chicago	0	0	155909	0	0	155909	19.602
Talmudical Institute of Upstate New York	0	0	110177	0	0	110177	17.663
Rabbinical College Telshe	0	0	99848	0	0	99848	15.885
University of the Potomac-VA Campus	0	13598	31590	30602	16848	92638	15.295
White Earth Tribal and Community College	0	0	20938	0	69658	90596	15.105
Heritage Bible College	0	0	74638	0	0	74638	14.640
Southeastern Baptist College	0	69190	0	0	0	69190	14.216
South Florida Bible College and Theological Seminary	0	0	0	22577	0	22577	11.547
Central Maine Medical Center College of Nursing and Health Professions	0	0	0	0	19936	19936	9.632
Huntsville Bible College	0	11950	0	0	0	11950	2.642
American Indian College of the Assemblies of God Inc	0	0	3542	0	0	3542	2.585

Table 2: Schools Investment Strategy without Constraints

Schools	**Year One**	**Year Two**	**Year Three**	**Year Four**	**Year Five**	**Total**	**ROI**
Georgia Perimeter College	58544804	53548222	41285850	34604186	32971734	220954796	92.5870
University of Wisconsin Colleges	0	0	13530241	19085225	17326555	49942021	92.5870
Princeton University	21103488	8369166	5155490	4867620	3250314	42746078	92.5866
CUNY Bernard M Baruch College	0	0	13807584	15081498	13259664	42148746	92.5864
Harvard University	3813672	13782453	8573484	8056746	5480334	39706689	92.5859
United States Merchant Marine Academy	10464234	2357638	1431252	1363234	882318	16498676	92.5856
Moorpark College	0	0	0	2472408	7336602	9809010	92.5850
Shorter College	815100	0	8850600	0	0	9665700	92.5849
University of Puerto Rico-Aguadilla	0	7442994	0	0	683982	8126976	92.5848

University of Puerto Rico-Arecibo	0	0	0	0	7642122	7642122	92.5847
New England College of Business and Finance	0	6862050	0	0	0	6862050	92.5846
Leeward Community College	0	0	1738076	2088064	2052472	5878612	92.5843
Baptist Memorial College of Health Sciences	649590	1426896	885204	832356	562611	4356657	92.5840
Berea College	1852029	596712	365010	345966	228528	3388245	92.5840
Universidad Pentecostal Mizpa	0	0	0	3340905	0	3340905	92.5840
Belanger School of Nursing	0	0	0	3151014	0	3151014	92.5837
University of Puerto Rico-Humacao	0	0	0	0	2855252	2855252	92.5837
Christian Brothers University	53176	938672	571064	543320	356048	2462280	92.5830
Yeshivat Mikdash Melech	264936	0	414352	1472424	0	2151712	92.5825
Machzikei Hadath Rabbinical College	0	1948716	0	0	0	1948716	92.5820
Albany College of Pharmacy and Health Sciences	0	0	0	467625	1219050	1686675	92.5817
Eastern West Virginia Community and Technical College	0	343808	1140224	0	0	1484032	92.5812
Talmudical Seminary of Bobov	0	273844	185121	864512	0	1323477	92.5805
Lincoln Trail College	741403	213200	131651	124189	83148	1293591	92.5803
Divine Word College	0	410557	598614	0	0	1009171	92.5802
Curtis Institute of Music	717482	267302	0	0	0	984784	92.5801
St Paul's School of Nursing-Staten Island	0	0	0	343620	630468	974088	92.5800
Jewish Theological Seminary of America	0	0	0	0	952637	952637	92.5789
Saint John Vianney College Seminary	317625	0	0	89859	317702	725186	92.5758
Crouse Hospital College of Nursing	0	0	149184	259392	169008	577584	92.5723
Doane College-Lincoln Grand Island and Master	0	0	0	273259	217021	490280	89.7398
Hobe Sound Bible College	0	0	0	0	378106	378106	89.4726
Carolina Christian College	325656	0	48240	0	0	373896	88.6888
Yeshiva Derech Chaim	0	155584	0	0	177840	333424	87.2778
Rabbinical Academy Mesivta Rabbi Chaim Berlin	97416	129560	0	74784	0	301760	86.6832
Olney Central College	0	0	66526	128557	91698	286781	76.5440
Universidad Teologica del Caribe	0	117211	167757	0	0	284968	68.5279
Maple Springs Baptist Bible College and Seminary	0	0	280440	0	0	280440	67.3181
Massachusetts Institute of Technology	0	0	0	0	279620	279620	62.6718
San Juan Bautista School of Medicine	8910	63504	174879	30159	0	277452	57.5081
Yeshiva College of the Nations Capital	84755	125557	0	16146	0	226458	50.1864
Bethesda University of California	0	0	0	0	226366	226366	49.7724
Bais Medrash Toras Chesed	0	84246	0	0	140448	224694	49.3984
Allegheny Wesleyan College	0	0	215377	0	0	215377	42.9887
National Institute of Massotherapy	0	211572	0	0	0	211572	35.6938
Sh'or Yoshuv Rabbinical College	0	0	0	0	172140	172140	34.1147
Talmudical Academy-New Jersey	145724	0	0	0	0	145724	32.9731
University of the West	0	143792	0	0	0	143792	25.6321
Soka University of America	0	137608	0	0	0	137608	21.8912
Telshe Yeshiva-Chicago	0	0	87770	0	0	87770	19.4881
Talmudical Institute of Upstate New York	0	0	87431	0	0	87431	17.7403
Rabbinical College Telshe	0	0	55216	0	0	55216	16.8754
Southeastern Baptist College	0	49136	0	0	0	49136	16.0518
White Earth Tribal and Community College	0	0	0	0	41470	41470	13.0379
University of the Potomac-VA Campus	0	0	884	22932	14742	38558	10.7562
Heritage Bible College	0	0	2479	0	0	2479	2.9507

8 Conclusions

8.1 Unconstrained Model

Looking through the unconstrained data, we are pleased to see well-known, highly-respected schools such as Princeton, Harvard and the Massachusetts Institute of Technology. We believe this provides on its own some legitimacy to the model. We are also pleased to some lesser known universities, even including colleges in Puerto Rico, which is going through a kind of economic crisis at present [18]. However, we are investing almost half of all our total to Georgia Perimeter College. In order to rectify this, we introduce the constrained model.

8.2 Constrained Model

Here, we cap the total investment in any one school to 50 million dollars. This has the effect of increasing the number of schools to which we disburse any funds, and obviously of decreasing the mean investment to schools. We believe this model to be better because it better protects the potential RoI against unforeseen issues at any single school. This constrained model serves as our final recommendation to the charity for how to disburse their funds.

8.3 Evaluation of Assumptions

Overall, we are satisfied with the model's conformity to its assumptions. To the best of our knowledge, missing data do not pose an issue. There may be some issues with the linear model assumptions, but we do not believe them to be so serious as to seriously threaten the model's power. Given this, we are confident in the conclusions of our model.

References

[1] Bayesian linear models: Gory details.

[2] Economy at a glance.

[3] Missing data imputation.

[4] Okun's law.

[5] State funding trends and policies on affordability.

[6] Unemployment by major - some degrees pay while others leave you paying. 2011.

[7] 2014 national population projections, 2014.

[8] The college degrees with the highest starting salaries in 2015, 2014.

[9] College graduate salaries by major, 2014.

[10] Eiu economic and commodity forecast, december 2015, 2015.

[11] National economic accounts, 2015.

[12] Accomodation and food services: Naics 72, 2016.

[13] Architect 1 salaries, biologist 1 salaries, equipment engineer 1 salaries, information security analyst 1 salaries, mechanic technician 1 salaries, precision assembler 1 salaries, public relations specialist 1 salaries, production assistant salaries, 2016.

[14] Average family consumer science salaries, average entry level personal assistant salaries, 2016.

[15] Entry-level environmental consultant salary, entry-level transportation salary, 2016.

[16] Entry-level translator salary, 2016.

[17] Entry-level translator salary, entry-level librarian salary, entry-level naval architect salary, religious education director salary, entry level chaplain salary, 2016.

[18] For richer, for poorer, 2016.

[19] Rate of unemployment by major, 2016.

[20] Stem grads projected to earn class of 2016's highest average starting salaries, 2016.

[21] Charles DiMaggio. Bayesian analysis for epidemiologists part v: Special topics.

[22] Jonathan Eaton, Samuel Kortum, and Brent Neiman. On deficits and unemployment. *Revue économique*, 64(3):405–420, 2013.

[23] Andrew Gelman and Jennifer Hill. Opening windows to the black box. *Journal of Statistical Software*, 40, 2011.

[24] Claudio E Montenegro and Harry Anthony Patrinos. Returns to schooling around the world. *Background Paper for the World Development Report*, pages 8258024–132095074719, 2013.

[25] Lynn O'Shaughnessy. 25 college majors with the highest unemployment rates.

[26] Anna Swanson. The college majors with the highest unemployment rates, 2015.

[27] Nassim Nicholas Taleb. *Antifragile: Things That Gain From Disorder*. Random House, 2012.