

Team Control Number

For office use only

T1 _____

T2 _____

T3 _____

T4 _____

74048

For office use only

F1 _____

F2 _____

F3 _____

F4 _____

Problem Chosen

B

**2018
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.

In order to provide a recommendation for our company, we have been asked to model the distribution of native speakers over time. To do this, we consider the effects of differing growth rates across populations as well as migration patterns. Newborn people are assigned a native language in accordance with the predicted distribution of languages in a country at the year of their birth. While the country-level data used in this analysis only considers native languages, we advance an argument that the distribution of native languages is highly correlated with the general distribution of languages, and moreover that for the purposes of connecting well with potential clients, the company is most benefited from being able to speak in the client's native tongue. If country-level data on second and third languages within a country were to be provided, our model can be easily altered to fit the new information. This data is used to create an iterative function which represents the dynamical system for each year over the 50-year period.

Using our predictions for the distributions of languages and populations across countries, we develop a utility function which determines optimal office sites based on the preferences of the company. If the company prioritizes economic profit, the utility function is weighted more heavily towards an economic factor, which is derived from a Heckman two-stage formulation of the gravity model. Otherwise, if the company is more focused on diversity and internationalism, the utility function is more weighted towards maximizing the company's linguistic diversity according to a metric which rises with the number of people that speak a new language and falls with that languages correlation to other spoken languages. Under the supposition that the company is more interested in internationalism, we indicate that India, Indonesia, Malta, Russia, Singapore, and Thailand would be an optimal combination of sites in the long term. On the other hand, if the company envisions it will be more interested in economic profit 50 years from now, Malta, Singapore, India, Indonesia, Mauritius, and Russia would be a better option. These recommendations are made with the understanding the the company already has offices in the United States and China. To increase diversity, the company will place no more than one office in a particular country.

**A Discrete-Time Dynamics Model
to Determine Language
Distributions and Optimal Office
Placement**

Team #74048

Mathematical Contest in Modeling
Consortium for Mathematics and Its Applications
February 8, 2018

Summary

In order to provide a recommendation for our company, we have been asked to model the distribution of native speakers over time. To do this, we consider the effects of differing growth rates across populations as well as migration patterns. Newborn people are assigned a native language in accordance with the predicted distribution of languages in a country at the year of their birth. While the country-level data used in this analysis only considers native languages, we advance an argument that the distribution of native languages is highly correlated with the general distribution of languages, and moreover that for the purposes of connecting well with potential clients, the company is most benefited from being able to speak in the client's native tongue. If country-level data on second and third languages within a country were to be provided, our model can be easily altered to fit the new information. This data is used to create an iterative function which represents the dynamical system for each year over the 50-year period.

Using our predictions for the distributions of languages and populations across countries, we develop a utility function which determines optimal office sites based on the preferences of the company. If the company prioritizes economic profit, the utility function is weighted more heavily towards an economic factor, which is derived from a Heckman two-stage formulation of the gravity model. Otherwise, if the company is more focused on diversity and internationalism, the utility function is more weighted towards maximizing the company's linguistic diversity according to a metric which rises with the number of people that speak a new language and falls with that languages correlation to other spoken languages. Under the supposition that the company is more interested in internationalism, we indicate that India, Indonesia, Malta, Russia, Singapore, and Thailand would be an optimal combination of sites in the long term. On the other hand, if the company envisions it will be more interested in economic profit 50 years from now, Malta, Singapore, India, Indonesia, Mauritius, and Russia would be a better option. These recommendations are made with the understanding the the company already has offices in the United States and China. To increase diversity, the company will place no more than one office in a particular country.

Contents

1 Introduction	4
1.1 Previous Research	5
2 Mathematical Model	6
2.1 Assumptions	6
2.2 Dynamical system model	9
2.3 Economic value model	11
2.4 Language Diversity Metric	14
3 Results	15
4 Sensitivity analysis	18
5 Discussion	19
5.1 Strength of the dynamical model	20
5.2 Weakness of the dynamical model	20
Memorandum	22
References	24
Appendix A: Parameter, variable descriptions and list of languages	26
Appendix B: Marginal Effects of the Probit Estimation on 2016 Bilateral Trade involving the 25 largest countries in the world by GDP.	27
Appendix C: Ordinary Least Squares estimation of 2016 Bilateral Trade	

Flows among the 25 largest countries in the world by GDP.	28
--	-----------

Appendix D: Selected codes	29
-----------------------------------	-----------

1 Introduction

Although over 6000 languages are spoken in the world today [2], the majority of the world's population can communicate in one or more of a relatively small subset of those languages. The 23 largest languages, for example, allow communication with more than half of the world's people [16]. For businesses with a profit incentive to communicate well with potential clients, then, this diversity and overlap of languages presents something of an optimization problem, wherein the costs of learning a new language must be balanced against the potential profits of speaking with clients that might otherwise remain unreached. Additionally, a business might value the diversity of thought and cultural knowledge that comes with new languages, which provides benefits that are more difficult to measure but no less tangible.

Accurately solving this hypothetical optimization problem requires a knowledge of how the spoken language distribution evolves over time. Companies and organizations seeking to get ahead of competing businesses may be interested in this information in order to try to adopt prospective future major languages early on. However, intuiting which languages will be dominant in the long term is difficult, since numerous factors impact the rise and fall of languages. Large immigrant populations, for instance, can dramatically alter the language distribution of a country [20]. Furthermore, languages spoken in countries with higher growth rates may become more popular on a global scale. Finally, relative popularity of languages in a state may provide individuals with incentives to learn that language.

In this paper, we construct an iterative function for a discrete-time dynamical system which, given the current distribution of native speakers across countries in each of the 20 most spoken languages, determines that same distribution in the following year subject to estimates of natural population growth, migration, and language learning. Then, we use this system to predict the distribution of languages 50 years in the future.

Once these estimations are in place, they are assumed to be the true state of the world in 50 years, and we solve an optimization problem for where an international company should base their six new offices. Specifically, we estimate the value of an office placement i in terms of the additional linguistic diversity and economic trade potential it affords. Economic trade potential V_i is measured using a two-stage Gravity model of trade fitted under the Heckman method to the most recent available trade statistics from the UNComtrade database [8]. Then, the company is afforded a single per-capita profit share of the resulting trade for the country in which it is placed. Linguistic diversity D_i is measured using a novel metric which increases with the number of people who speak a particular language, and decreases with the correlation between a language and languages that are already used by existing offices. Our group allows the business to decide the relative weight it will give to economic and linguistic benefits in its decision making process through the

--- End of preview ---