When can I go? Predicting country open-to-tourism probabilities

Skyler Gray

December 9, 2021

1 Introduction

The pandemic has hit countries really hard. Economies have suffered and some once-popular forms of commerce are still at a standstill. In the travel and tourism industry, vacationers have been eyeing their next getaway location wondering, "When will I be able to travel there again?"

My family purchased tickets some time ago to travel to Taiwan at the end of April 2022. At the time we purchased the tickets, we reckoned the pandemic would have largely ended by then. Seeing that is not the case, we are unsure how likely it is we will get to take our trip to Taiwan four months down the line. The goal of this analysis is to estimate the probability that Taiwan is open to Tourists by April 2022. In this analysis, we consider a country with a generalized international travel policy of **Screening** or **No measures** to allow for tourist entry without initial quarantine requirements. As a secondary goal, we want to understand whether there exists a strong relationship between the number of new COVID cases or vaccination rate and international travel control levels.

2 The Data

The data for this analysis comes from the Oxford COVID-19 Government Response Tracker (https://doi.org/10.1038/s41562-021-01079-8) through the Global Change Data Lab's *Our World in Data* project website (https://ourworldindata.org/coronavirus). This dataset contains ordered data on the daily international travel control levels of around 180 different countries during the COVID-19 pandemic (see Figure 1). Each country's international travel control policies have been generalized by researchers into one of five categories, from least to most restrictive policy levels: **No measures**, **Screening**, **Quarantine from high-risk regions**, **Ban from high-risk regions**, and **Total border closure**. Additional data on monthly deaths or new cases and vaccination rate by country was also pulled from Our World in Data and is a compilation of data from multiple sources, including the United Nations Department of Economic and Social Affairs, the Food and Agriculture Organization, World Bank, the Center for Systems Science and Engineering (CSSE)

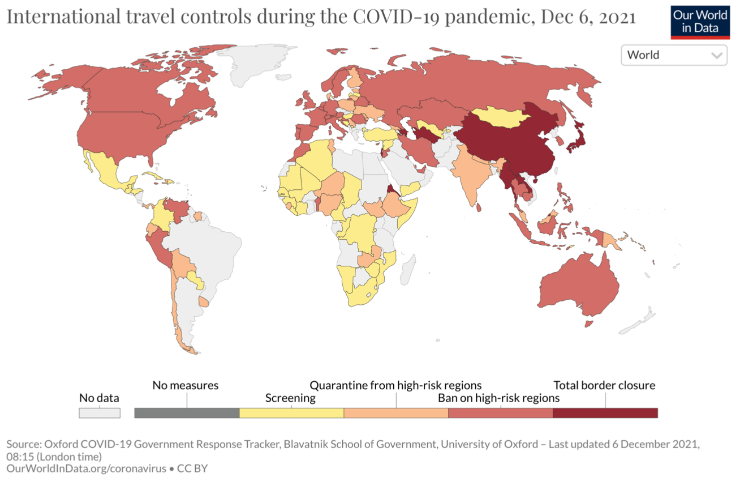

Figure 1: A visualization of international control policy status by country on December 6, 2021 from *Our World in Data*'s web page.

at Johns Hopkins University, and national government reports.

Because changes in travel policy levels are infrequent and my research questions do not directly pertain to daily travel policy levels, the data was transformed from daily observations of each country's travel control levels to monthly observations. This has the benefit of reducing the correlation between, making a simple modeling structure more appropriate for this analysis. Table 1 gives a brief description of the data used for modeling. Table 2 shows information on the dataset sample size per continent.

3 EDA

Figure 2 shows historically each country's international control policy level from January 2020 to October 2021. From about January to June, most countries were closed to international travel. From July onward, however, we see for the most part that Africa's travel policy levels relaxed to quarantining or screening policies. Most countries in Asia and Europe, however, required at least quarantining measures for international travel. Also, relative to most other countries in Asia, Taiwan appears to be one of the most conservative, having totally closed its borders on two separate occasions and banning countries from entering its borders most other months. Additional background on Taiwan is included in Figure 3.

4 Methods

I chose to use a Bayesian ordered probit regression model for this analysis. Let Y_{ij} be the international travel control level of country i on month j, where $j = 1$ is the month of January 2020, and let Z_{ij} be an

Variable	**Description**
Continent	City council district of the request GPS coordinates
Reporting Full Vaccinations	A variable indicating whether the country has at least one fully vaccinated citizen in the middle of the month
Has New Cases	A variable indicating whether the country has had any new cases in the given month
Log(Population Density)	The log-scaled population density of the given country; this remains constant across time
Median Age	The median age of the given country
Log(GDP Per Capita)	The log-scaled GDP per capita of the given country
Month	A time variable from 1 to 22, where 1 represents January 2020 and 22 represents October 2021
Log(New Cases)	The log-scaled average number of new cases in the given month
Proportion Fully Vaccinated	The proportion of the given country's population fully vaccinated in the middle of the month
Month:Log(New Cases)	An interaction effect between time and the log-scaled number of new cases

Table 1: Brief descriptions of the data to be used in the final model

Continent	**Countries**	**n**
Africa	48	964
Asia	46	934
Europe	38	801
North America	17	335
Oceania	8	123
South America	12	246

Table 2: Number of countries and sample size by continent

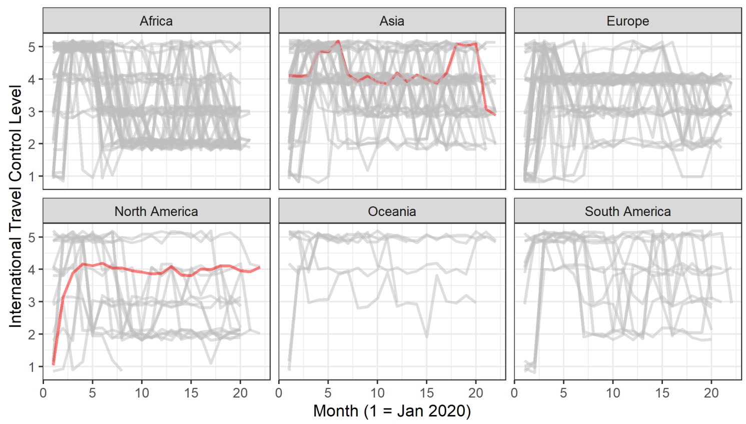

Figure 2: International control levels (jittered) by country and organized by continent. Taiwan is colored in red, as well as the United States for comparison.

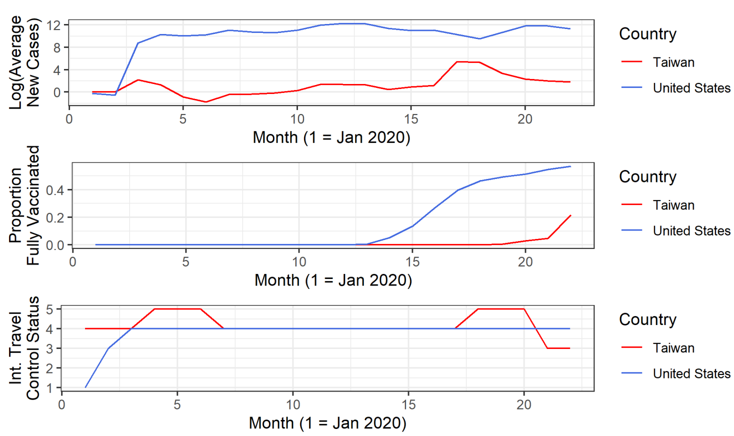

Figure 3: COVID-related measures in Taiwan (red) against time. The United States (blue) is added here for comparison. Until recently, Taiwan was considered by this policy measure more strict than the United States. The country prioritized giving first doses of the COVID vaccine to its citizens.

unobserved variable that determines Y_{ij}. Then

$$
\begin{aligned}
Y_{ij} &= k \quad \text{if} \quad \delta_{k-1} < Z_{ij} \leq \delta_k \\
Z_{ij}|\beta &\sim N(X_{ij}\beta, I) \\
\beta_\ell &\sim N(0, 10),
\end{aligned}
$$

where δ is an ordered vector of six real values used to "bucket" observations into one of the five travel control levels. To avoid identifiability issues, $\delta_0 = -\infty$, $\delta_1 = 1$, and $\delta_K = \infty$. For this application, then, $\delta_2, ..., \delta_4$ need to be estimated. Because obtaining convergence on posterior draws of these parameters is fairly difficult, I apply a transformation to these parameters to more quickly achieve convergence. The transformation is

$$
\begin{aligned}
\alpha_1 &= log(\delta_1) \\
\alpha_k &= log(\delta_k - \delta_{k-1}), \quad k \in \{2, ..., K-1\}.
\end{aligned}
$$

The following prior was used for each α_k:

$$\alpha_k \sim N(0, 10)$$

The mean parameter for Z_{ij} is

$$
\begin{aligned}
X_{ij}\beta = \beta_0 &+ ContNA_i\beta_1 + ContEU_i\beta_2 \\
&+ ContAS_i\beta_3 + ContOC_i\beta_4 \\
&+ ContSA_i\beta_5 + FullVaccineFlag_{ij}\beta_6 \\
&+ NewCasesFlag_{ij}\beta_7 + Log(PopDens_i)\beta_8 \\
&+ MedianAge_i\beta_9 + Log(GDPperCapita_i)\beta_{10} \\
&+ j\beta_{11} + Log(AvgNewCases_{ij})\beta_{12} \\
&+ PropFullyVaccinated_{ij}\beta_{13} + j \times Log(AvgNewCases_{ij})\beta_{14}
\end{aligned}
$$

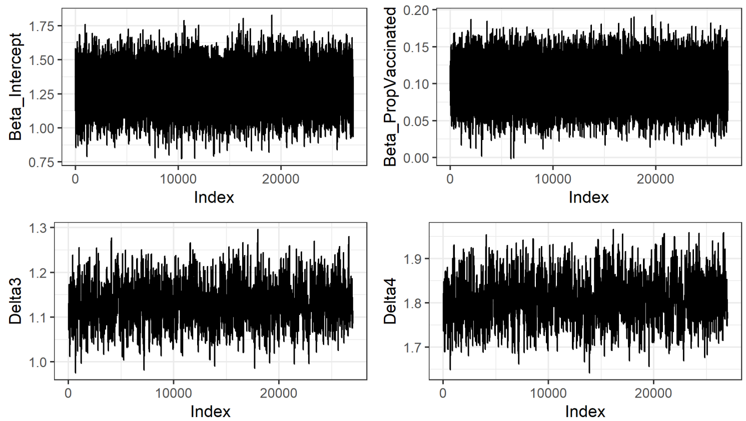

Figure 4: Trace plots for two coefficients and two cutoff values.

I wrote a Metropolis within Gibbs algorithm to sample from this distribution. The full conditional distributions of Z_{ij} and β are recognizable as distributions we can easily sample from, where

$$\beta|Z,\delta \sim N((X'X)^{-1}X'Z,(X'X)^{-1})$$
$$Z_{ij}|\beta,\delta \sim TN(X_{ij}'\beta,1,\delta_{Y_{ij}-1},\delta_{Y_{ij}}),$$

but the full conditional of α_k is not recognizable. Therefore, the posterior draws for the α parameters were derived using a metropolis random walk approach. Diagnostics and results are from 4 chains of 30,000 draws with 3,000 of those draws used for warm up. The samples were then thinned to every fourth draw, resulting in a total of 27,000 samples of the posterior used for analysis. As a minor detail to encourage convergence, I centered and scaled all the continuous covariates.

4.1 Diagnostics

Figure 4 shows trace plots from the posterior draws of two β coefficients as well as two δ cutoff values. The trace plots for all β coefficients look great, but the trace plots for the δ cutoff values could be better. This is also apparent when investigating posterior sample effective sample sizes (ESS) for δ (see Table 3), though I would consider the ESS large enough to use these posterior draws for analysis.

Coefficient	**Estimate**	**Lower**	**Upper**	**ESS**
(Intercept)	1.289	1.03	1.543	8553
Continent: Asia	0.286	0.166	0.407	26066
Continent: Europe	-0.333	-0.51	-0.153	27000
Continent: North America	-0.304	-0.453	-0.153	25348
Continent: Oceania	1.3	1.062	1.535	23383
Continent: South America	0.257	0.079	0.434	27000
Reporting Full Vaccinations	0.203	0.08	0.327	27000
Has New Cases	0.673	0.447	0.896	25155
Log(Population Density)	-0.045	-0.087	-0.003	27362
Median Age	0.068	-0.023	0.159	27000
Log(GDP Per Capita)	0.069	-0.001	0.14	27000
Month	-0.354	-0.431	-0.275	25368
Log(Average New Cases)	0.267	0.184	0.35	27000
Proportion Fully Vaccinated	0.101	0.053	0.148	27833
Time:Log(Average New Cases)	-0.218	-0.333	-0.102	27000
δ_2	1.131	1.046	1.221	797
δ_3	1.808	1.714	1.902	869
δ_4	2.782	2.682	2.883	793

Table 3: Estimates, 95% credible intervals, and ESS of the β coefficients and δ cutoff values.

5 Results

To estimate probability forecasts into April 2022, my model requires data on both the proportion of a country's population fully vaccinated and the average daily number of new COVID cases each month. From August to October, Taiwan's average daily number of new cases has hovered around 7, so assuming Taiwan doesn't experience another outbreak through to April, I assigned its new cases each month to be 7. As for vaccination proportions, I pulled information on the vaccination rate of Taiwan for November and December and roughly guessed the vaccination rate up until April; the estimated vaccination rates from November 2021 to April 2022 are 0.4, 0.6, 0.7, 0.75, 0.78, and 0.80.

Figure 5 shows the model's estimation of the probability that Taiwan will open to tourism From January 2020 to April 2022. There are several comments I could make on this model's ability to predict in-sample time periods, but I will focus here only on forecasting future probabilities. I like that the model prediction uncertainty increases with time. However, I suspect that this model is providing too conservative of an estimate on the probability that Taiwan will open in April 2022. This model estimates, against my intuition, that an increasing vaccination rate is *positively* correlated with a more strict international travel control level (see Table 3). Thus, with a dramatic increase in the proportion of fully vaccinated individuals, there is a temporary forecast of a decrease in the probability that Taiwan will be open to tourism in future months.

The β coefficient mean estimate and 95% credible intervals are shown in Figure 6 below. The ordered probit regression model identified a positive correlation of both the average daily cases and population

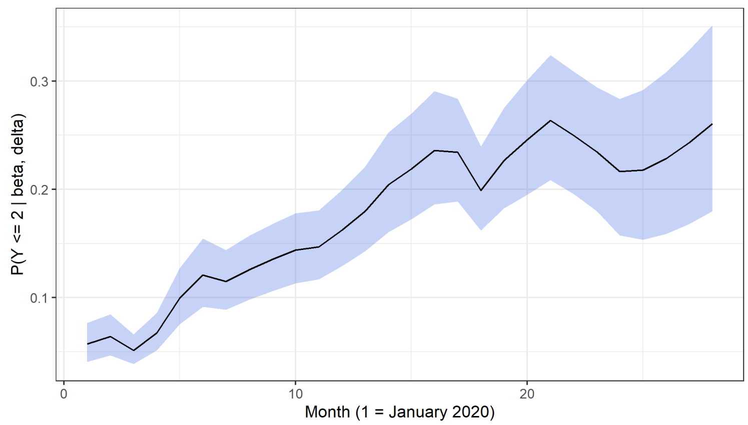

Figure 5: Model predictions for the probability that Taiwan will be open to tourism on the given month. Equivalently, this is the probability Taiwan's international policy control level is **Screening** or less. The black line is the mean predicted probability while the blue bound represents the 95% credible interval of said probability.

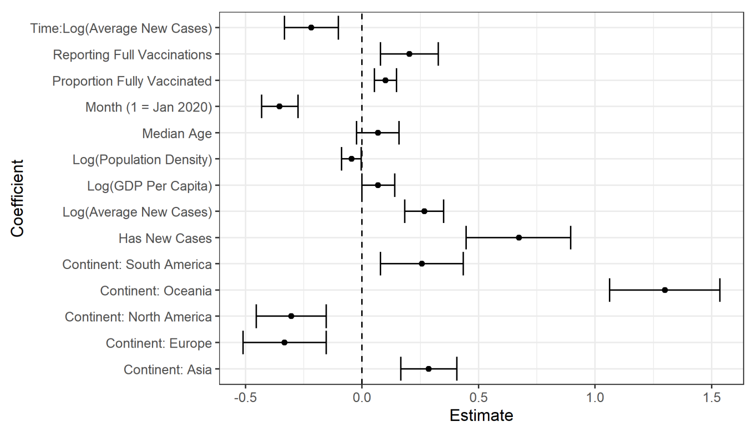

Figure 6: Estimates and 95% credible intervals of the model β coefficients.

proportion fully vaccinated with international travel policy strictness. While the positive correlation for daily cases makes sense, I am not sure why there is a strong positive correlation between the population proportion fully vaccinated and the strictness of a country's international travel policy. One potential explanation for this relationship may be that countries with higher vaccination rates are inherently more conservative than those with low rates. However, there may be confounding variables not included in the model, such as general government structure, a measure of freedoms granted to news media in a given country, or a measure on whether a new COVID variant is making headlines.

6 Conclusion

In this analysis, I used a Bayesian ordered probit regression model to predict the probability of Taiwan opening to tourists in April of next year as well as explore effects of COVID-related statistics on travel policy strictness. While we cannot claim causation, both vaccination rates and new COVID case counts appear to have an intensifying effect on a country's policy strictness.

This model may be of use to other vacationers out there (Got any currently-closed-to-tourist destinations in mind, Dr. Berrett?). However, given the unexpected positive effect of vaccination rates on travel control levels, I think this model needs further investigation and revision before it will be ready for use by an audience larger than just myself.

The next step for this analysis is to estimate just how good or bad this model is at predicting future observations. From there, research into alternative models will be easier using this predictive metric for comparisons. I am interested in including additional features into the model and understanding whether including them leads to a more intuitive understanding of the relationship between country vaccination rates and country travel policy strictness.