Forecasting Demand for Neonatal Services

Juan Campanario

April 3, 2023

1 Introduction

Understanding how demand for services change over time is essential to policy makers. Whether it be polling stations, bus stops, or hospital services, predicting which locations anticipate the largest growth in demand allows policymakers to prioritize and direct funds to areas most in need. In this project, I focus on the demand for neonatal services in Peru. Demand for such services are a function of how many infants are born; thus, to forecast demand I use daily registered births to predict which districts will see the largest growth in births.

2 Data

To forecast demand, I use data on all live births registered in Peru between January 1st, 2005 and December 31st, 2018 as a proxy for usage of neonatal services. Peru's Live Birth Registy contains health records of newborns, including data on birth weights, gestational periods, and maternal ldemographic information such as birthplace, age, and level of education at the time of birth. Although Law N^{o} 26497 requires that all live births be registered, in 2015, the Peruvian National Institute of Statistics and Information estimated that 72% of live births were covered that year. Some areas such as the Amazon, have limited coverage due to a lack of public services.

In Figure 1, I plot the total number of births per day throughout the above timeframe. The number of registered births follows two main patterns: there are fewer births in the end of the year than the beginning of the year, and the number of births is increasing over the timeframe. In Figure 2c, I group births by month to document the former pattern. The median births in the months of January through September are similar but falls in the last three months of the year, with the fewest births in December. In Figure 2a, I highlight the latter pattern. There are more than one and a half times as many births registered in 2018 than 2005. Population growth alone does not explain the increase. Improved infrastucture, such as more hospitals and roads, results in better record keeping, partially explaining the growth. This feature of the data is important as it highlights that take-up of neo-natal services increases by two channels, which affect where funding for services and additional supporting infrastructure are needed.

In Figure 2d, I plot births grouped by the day of the week. There are notably fewer births registered on Sundays. As shown in Figure 3, there are significantly fewer births registered in the days leading up to New Years. While the day of birth may be misreported, it is unlikely that the report and actual birthdays vary significantly from each other. In addition, policymakers care about trends over a long period, as opposed to simply daily fluctuations. Thus, these particular features, while interesting, will not impact the general findings.

3 Methodology & Results

For my forecast I predict the total number of daily births in the last two years of available data, January 1st 2017 through December 31st 2018. I employ a supervised gradient-boosted decision tree model using the day of the week, month, quarter, and year to predict the total daily births. I train the model on the entire pre-period i.e., January 1st 2005 to December 31st 2016. In Figure 4, I plot the two periods together, separating the training and test periods by a dashed vertical line. In Figure 5a, I compare the true and predicted daily births in the test period. The model puts the greatest weight on the yearly variation and some weight towards the day of the week, day of the year, and monthly variations, as shown in Figure 7a.

Second, I build a model to forecast district-level registration. I calculate the total number of registered births within a year for 1,688 districts.[1] In Figure 8a, I map the log growth rates for each district from 2005 to 2018. The largest growth is found in the most inland regions.[2] I then train a decision tree model on the year, district, province, and department-level indicators, using 2017 and 2018 as testing data.

After calibration, I predict the total number of registered births for each district in 2025. In Figure 8b, I document the predicted log growth rates based on the observed registered births in 2018 and the predicted registered births in 2025. Compared to 2018, I predict that there will be more registered births in the highlands and the south, as indicated by the purple hues. In Table 1, I highlight the top 10 districts ranked by log growth rate with registered births above the median in both 2005 and 2018, to exclude districts with relatively few births.

4 Conclusion

In this essay, I forecast demand for neonatal sevices both at the national and district level using registered births as a proxy. I find that the areas in which registration grows the fastest in 2018 are not those that will require the most services in 2025. One major consideration is that I am using registered births as opposed to actual births. The government already prioritizes limited funds, and districts that receive greater resources between 2005 and 2018 would have inflated predictions for future registration. Second, my forecasts rely on relatively simple measures, which do not take into account other valuable data sources such as road

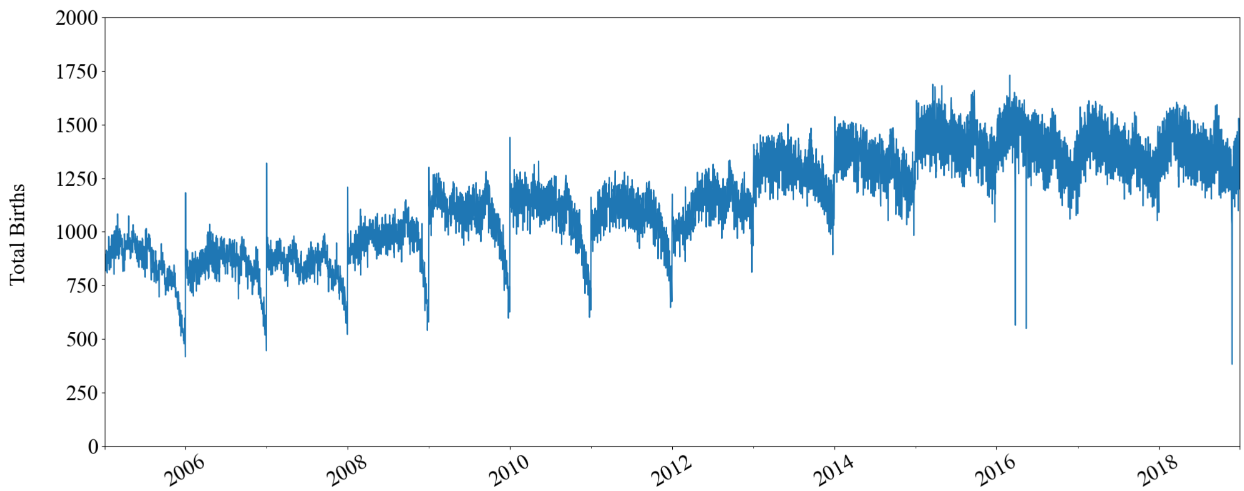
[1] There are 1,838 districts in Peru, but I only include those with at least one registered birth in both 2005 and 2018.

[2] I prefer to use log growth rates due to the right skew in typical percentage change as shown in Figure 6.The skew is due to districts with very few registered births in 2005, typically one, that come to have an order of magnitude larger in 2018 e.g., 40.

developments, census data on the number of people in districts, hospital and health outpost operating times, pollution that would all impact both the number of births and registered births. Third, the training data does not account for recent events that would lead to a less accurate prediction such as COVID19 or political conflict.

5 Figures & Table

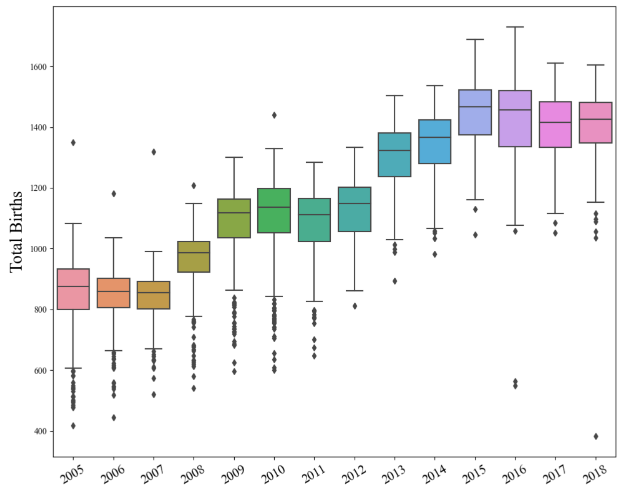
Figure 1: Total Daily Births, 2005 - 2018



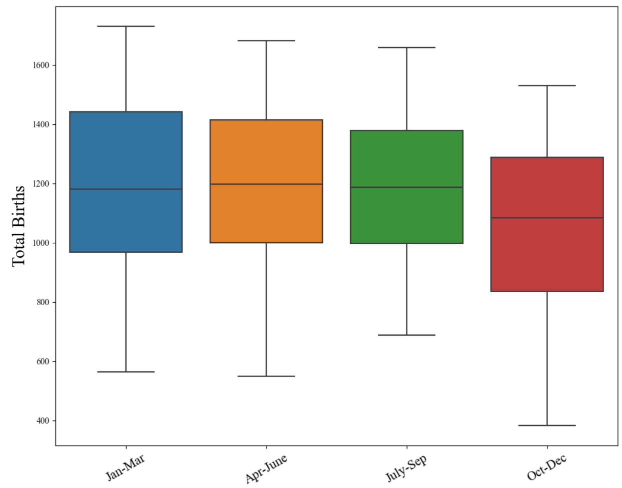
Note: All births recorded in the registry.

Figure 2: Variation in Total Births

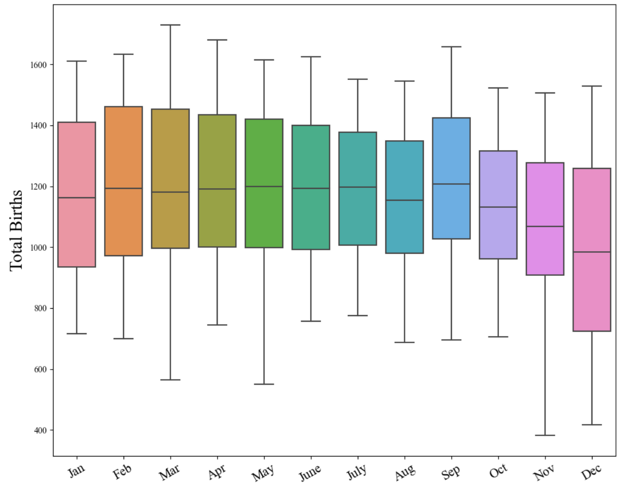
(a) Yearly



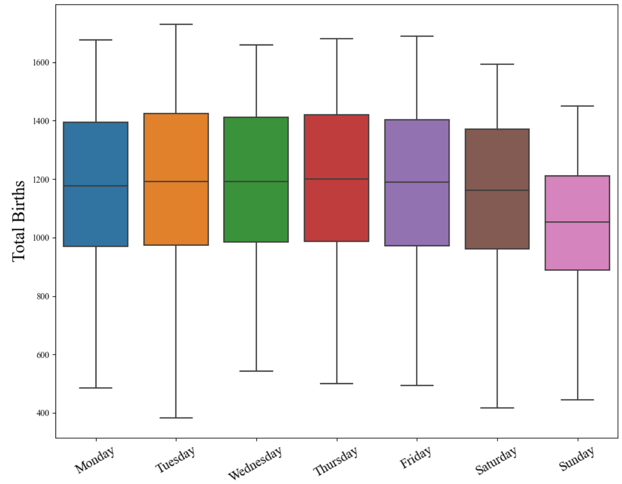
(b) Quarterly



(c) Monthly



(d) Day of Week



Note: All births recorded in the registry.

Figure 3: Births between December and January, 2005 - 2018

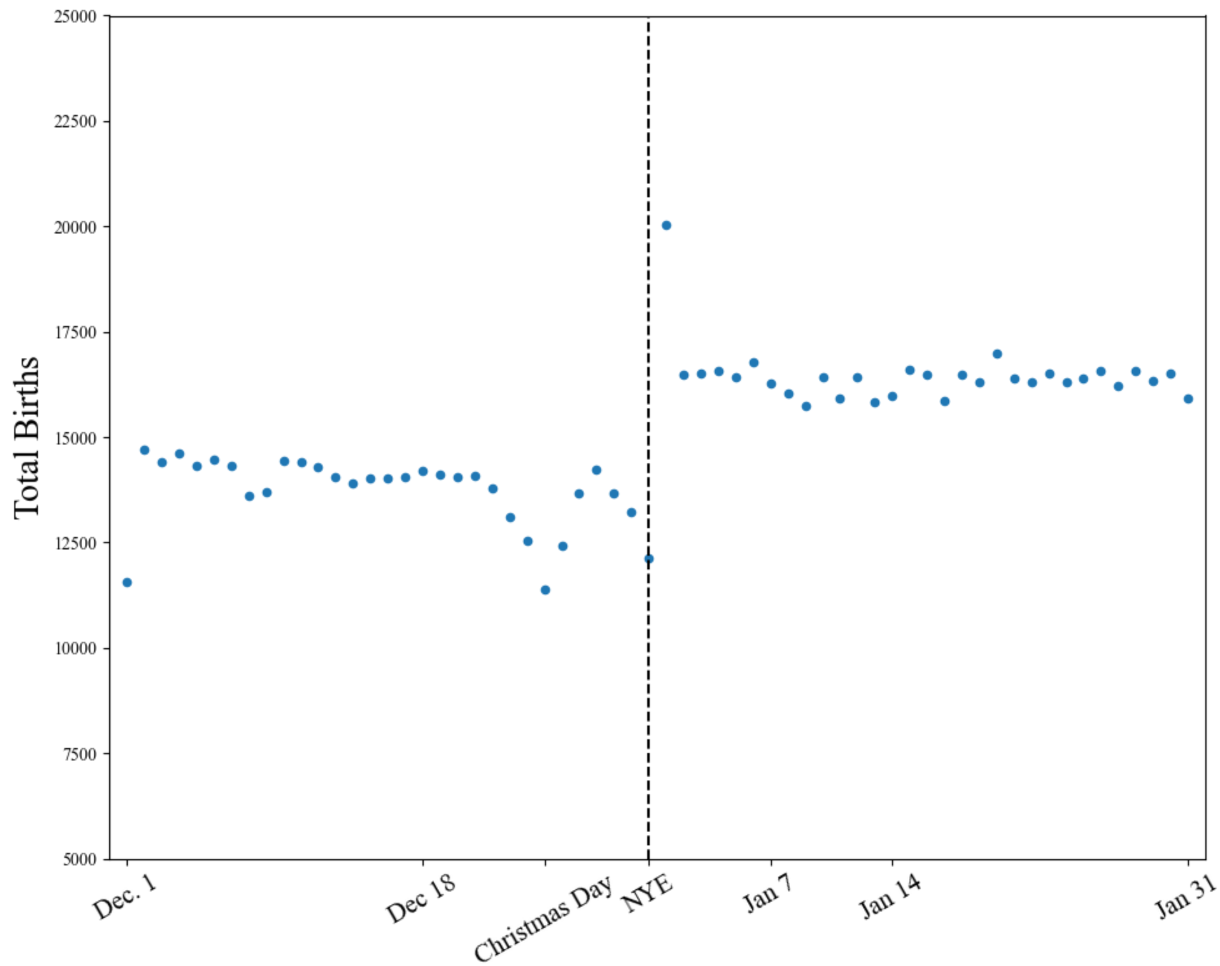
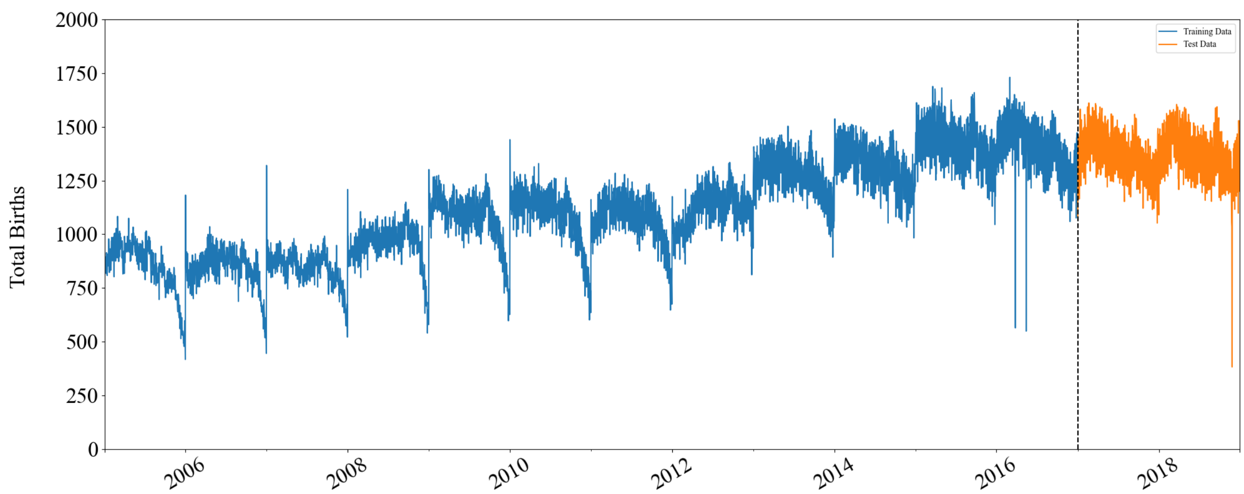


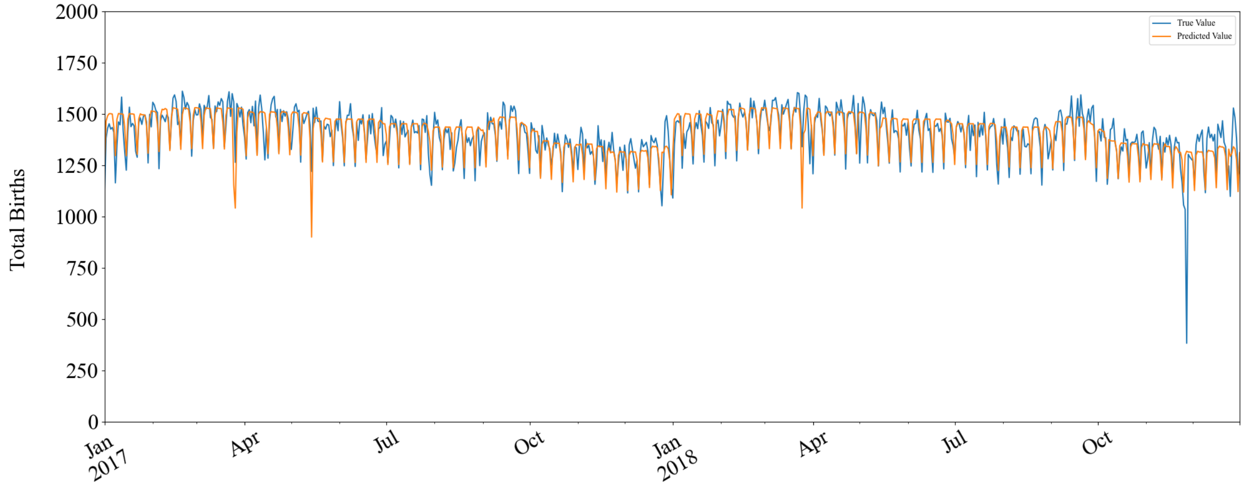
Figure 4: Training and Test Data



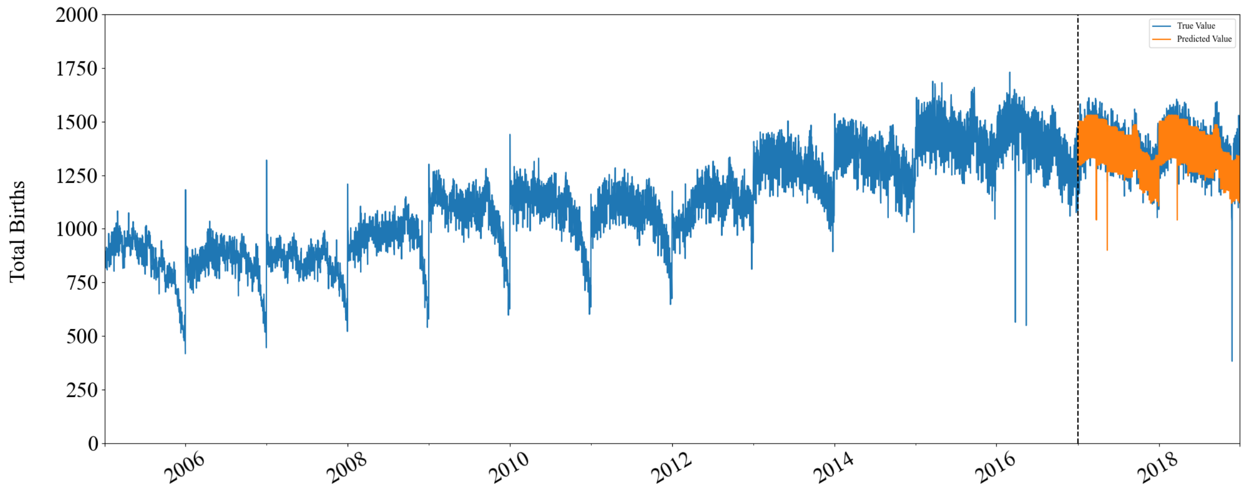
Note: All births recorded in the registry.

Figure 5: Forecasting Daily Total Births

(a) Forecasting Births, 2017 -2018



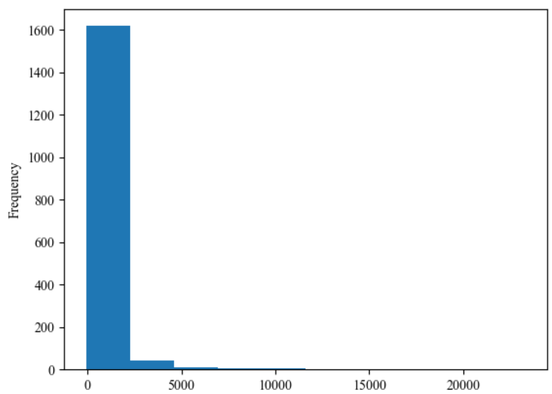
(b) Forecast with Training Data



Note: Panel (a) contains the actual and predicted data between January 1st 2017 and December 31st 2018, and Panel (b) includes the data for all dates in the sample.

Figure 6: Percent Change and Log Growth Rate

(a) Percent Change



(b) Log Growth Rate

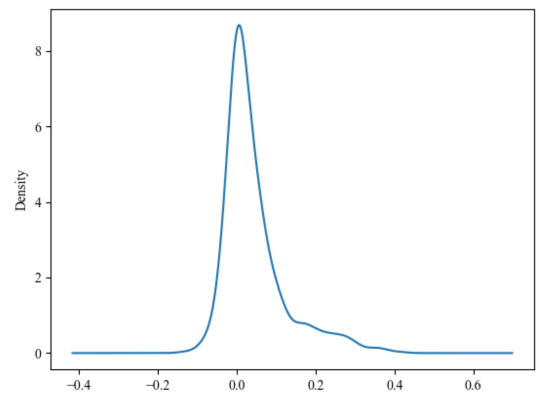
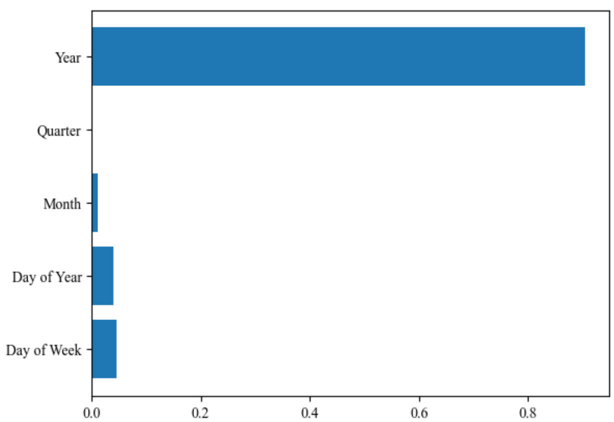


Figure 7: Feature Importance

(a) Features for Daily Total Births



(b) Feature for Yearly District-Level Births

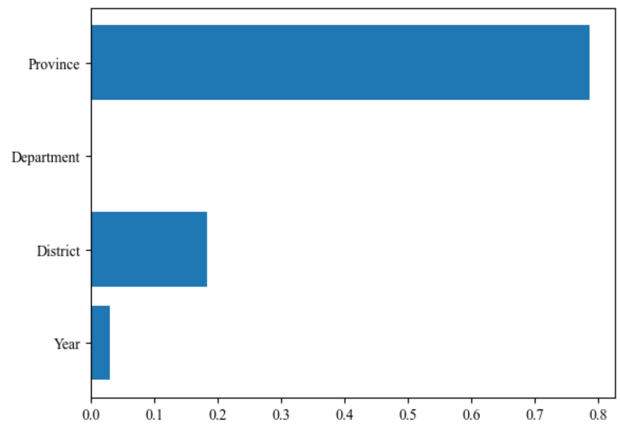
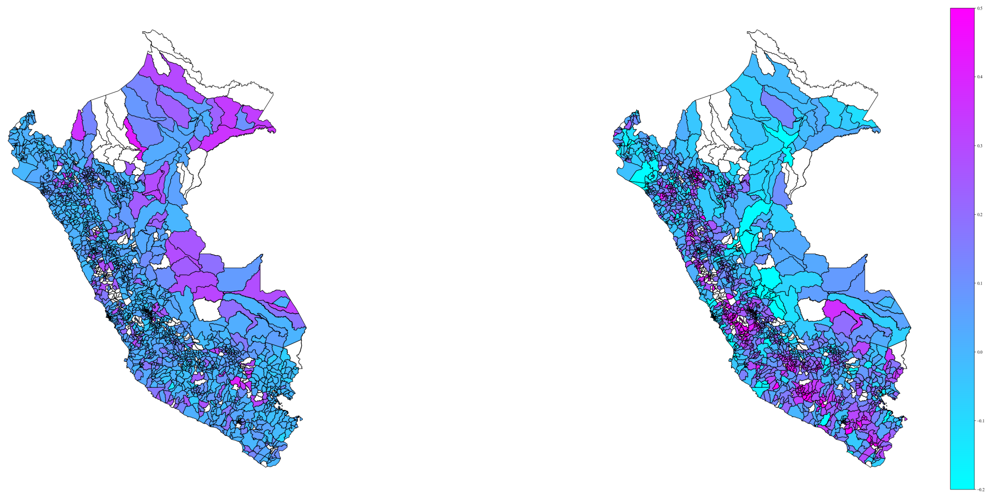


Figure 8: Results

(a) Log Growth Rate, 2005 to 2018 (b) Log Growth Rate, 2018 to 2025



Note: Panel (a) I compute the log growth rate between 2005 and 2018 for each district that had at least one birth registered in both years, and Panel (b) I compute the log growth rate between 2018 and the predicted registered births in 2025.

Table 1: Top 10 Districts in 2025

	UBIGEO	Department	Province	District	Predicted Log Growth Rate
1.	021802	Ancash	Santa	Caceres del Perú	0.254
2.	022003	Ancash	Yungay	Mancos	0.180
3.	211202	Puno	Sandia	Cuyocuyo	0.180
4.	211209	Puno	Sandia	Alto Inambari	0.178
5.	131102	La Libertad	Gran Chimu	Lucma	0.173
6.	131003	La Libertad	Santiago de Chuco	Cachicadan	0.169
7.	081303	Cusco	Urubamba	Huayllabamba	0.168
8.	060810	Cajamarca	Jaen	San Felipe	0.163
9.	080903	Cusco	La Convencion	Hauyopata	0.161
10.	040706	Arequipa	Islay	Punta de Bombon	0.161

UBIGEO is the geographic identifier used for districts.