Business Analytics Using Forecasting:

Forecasting quantity of newborns to better allocate cram schools/day care centers and their resources

Team 2

Di Chien, Lo
Guerman Shion (Gao Fei, Hsiao)
Jieng Wuen, Cai (Kate)

Summary

Cram Schools (CS) are specialized schools that train their students to meet particular goals, most commonly to pass the entrance examinations of high schools or universities. People in Taiwan are aware about the importance of early childhood education and development. Also, Day Care Center (DCC) is an ongoing service during specific periods, such as the parents' time at work. Recently large numbers of women workers are entering or returning to the workforce. The growing economy is encouraging parents to spend more on these child development services.

Taiwan has the lowest birth rate in the world, it's a serious national security threat and definitely a big demographic problem. It's also a big shock for cram schools and day care centers, facing the plummeting birth rate, they need to figure out solutions to use their resources more efficiently.

The profitability of a CS/DCC is largely determined by the quantity of students. For all providers, the largest expense is labor. Hence, it can be argued that a sustainable competitive advantage can be achieved, if market size of new potential students could be accurately forecasted, it could allow a better management, and ultimately lead an increase in performance. Our basic business objective is to develop a more targeted strategy of resources allocation of CS/DCC in Taiwan, staff (full-time and par-time teachers) and site (size and location).

In this project we are trying to accurately forecast two horizons for the demand of newborns, short-term and long-term, with respective strategies 1 year for staffing purpose and 3 years for allocating branches. By leveraging the power of forecasting and predict the demand of newborns as to "when" and "where" should be opened or closed a branch we aim to obtain an optimized equilibrium of demand and supply.

Some main features to note within the data are: chosen series have two different sizes, collected between 1981-2013 for newborn babies and crude married rate (33 periods, yearly) and 1999-2013 for bank deposit rate, female work rate and female education. (15 periods, yearly). The data shows a twelve year seasonality based on Chinese Zodiac and a downward trend overall. This forecast has ongoing capability, since the data is updated yearly its implementation don't require automation. In this report, we used data from the Ministry of the Interior of Taiwan.

In the first step we forecast using naïve as a benchmark to compare other models. Test results in model based forecasts proved that the influence of married rate series are statistically significant when predicting the newborn series. Among the data-driven models not one outperformed the benchmark.

The store can increase its profits by preventing overcapacity or under-capacity. Different models work best and are more reliable depending on the horizon, for short-term forecast (one year) we used linear regression and for long-term (3 years) we used naïve as they showed better performance among other models. One should also keep in mind the fact that external data will be available only after a certain lag of at least a year. Based on our research, we recommend a model using a combination of two forecasting techniques that are simple, yet robust and flexible enough to create a reliable prediction.

Problem description

(1) **Business goal**

Allocate CS/DCC in strategic cities of Taiwan. If the cram school or day care centers can reach the school capacity in the chosen locations, it would be considered a success.

(2) Forecasting Goal:

Determine the baby population with a horizon of 1 year (staffing purpose) and 3 years (allocation purpose).

Data description & data preparations

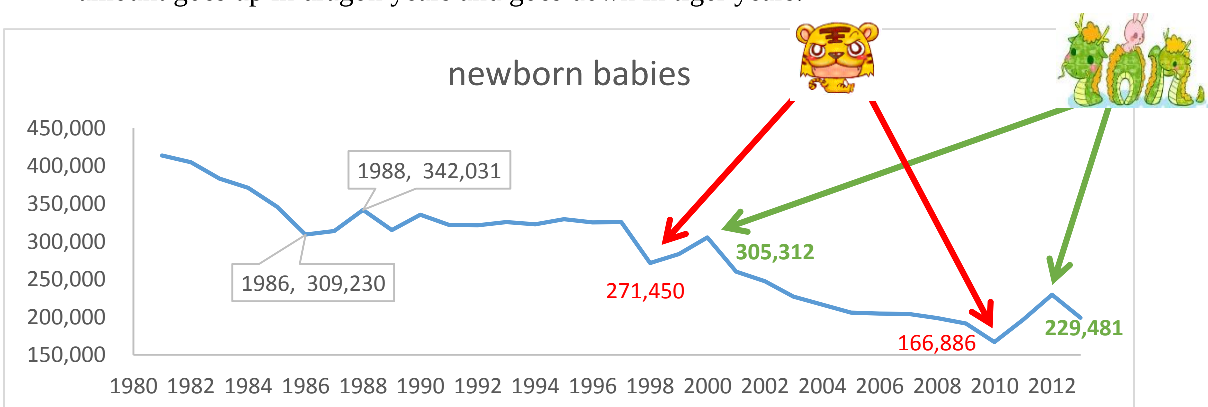
(1) Brief description of available data:

Source: Ministry of the Interior of Taiwan
Time period of obtained data:
1981-2013 for newborn babies and crude married rate. (33 periods, yearly data)
1999-2013 for bank deposit rate, female work rate and female education. (15 periods, yearly)
Series to analyze.
2005-2014 for city weight of newborns. (9 periods, yearly)
There's one thing noticeable, we won't get the data of married rate, bank deposit rate, and so on, until the beginning of next year. So if we want to forecast the newborn babies by external data, we need to use lag 1 data rather than the current data.

(2) Time plot of series.

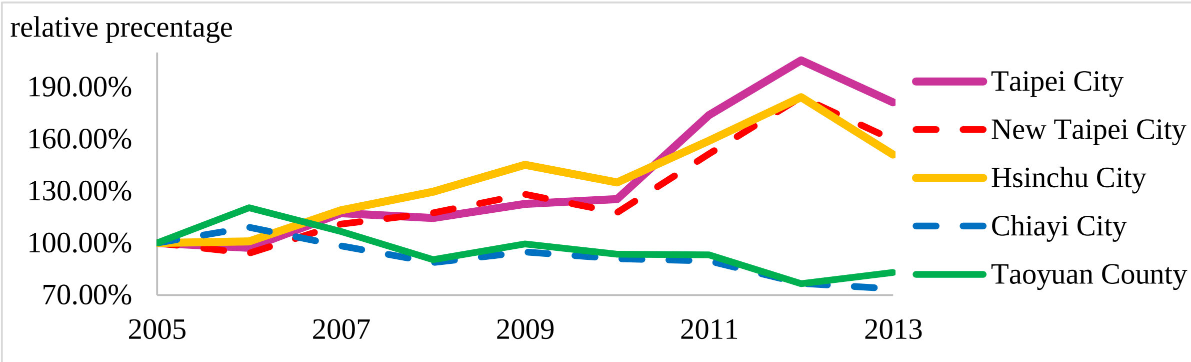
a. Newborn babies per year

There's a seasonality in newborn babies series which is driven by Chinese Zodiac, the amount goes up in dragon years and goes down in tiger years.



b. Potential Market for CSDCC

Since we want to allocate the branches and staff of cram schools and day care center, we need to find out where the market is. Cities weights are the percentage change of the total newborns, relative to 2005 (base year, because this the oldest data we can get). To select strategic cities, we used city weight growth in a 9 year period, (to find the behavior of newborns quantity in each city. From the figure below, we clearly see that Taipei, New Taipei and Hsinchu City are the potential markets for CS/DCC while Chiayi and Taoyuan County are those unfavorable cities for CS/DCC. Notice that the series are forecasted using the actual percentage of the potential markets.

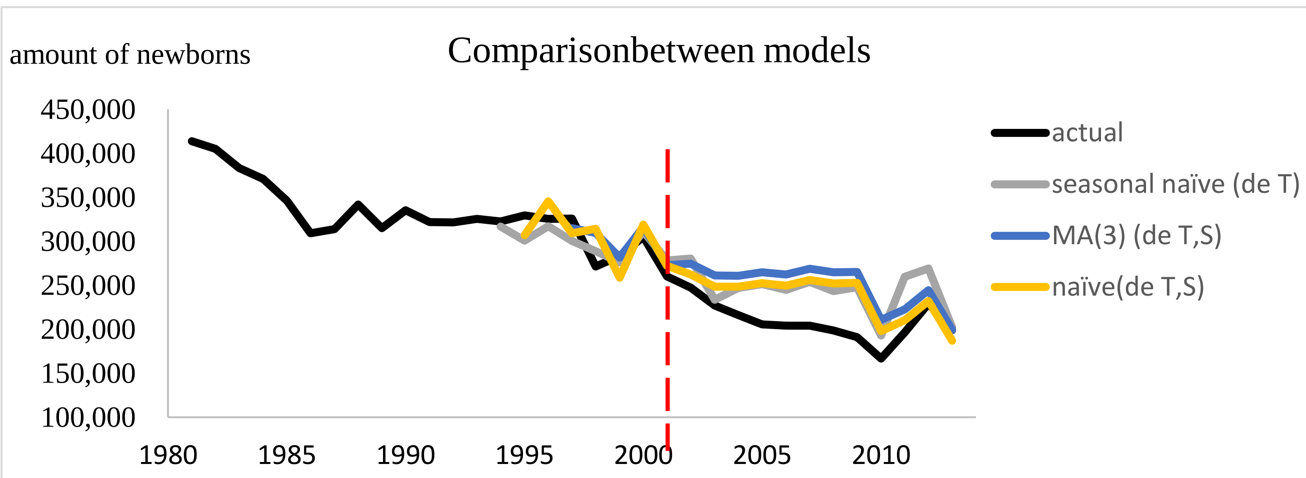


Forecasting solution

(1) Methods and performance evaluation

a. Data driven methods

As a benchmark method to compare more advanced approaches, first we consider seasonal naïve forecasting. Since we identified trend and seasonality in our data, we consider a simple approach using differentiation and then applying moving average method or naïve forecast for the level.



	Seasonal Naïve	Seasonal Naïve de T	MA(3) (de T,S)	naïve(de T,S)
MAPE	50.69%	17.98%	21.15%	16.06%

b. Model based methods

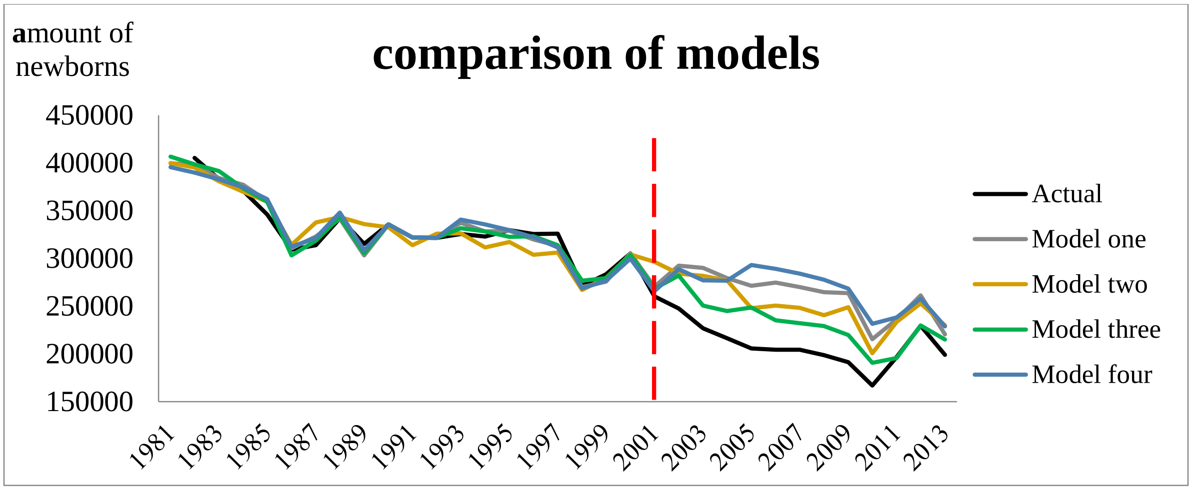
(1) $\ln(\text{Newborn}) = \beta_0 + \beta_1 \text{trend}_t + \beta_2 \text{marriage rate}_{t-1} + \beta_{3\sim 13} \sum(\text{Chinese Zodiac Animal}_t) + \varepsilon_i$

(2) $\ln(\text{Newborn}) = \beta_0 + \beta_1 \text{trend}_t + \beta_2 \text{marriage rate}_{t-1} + \beta_3 \text{dragon}_t + \beta_4 \text{tiger}_t + \varepsilon_i$

(3) $\ln(\text{Newborn}) = \beta_0 + \beta_1 \text{trend}_t + \beta_2 \text{marriage rate}_{t-1} + \beta_3 \text{Ln}(\text{New born})_{t-1} +$

$\beta_{4\sim 14} \sum(\text{Chinese Zodiac Animal}_t) + \varepsilon_i$

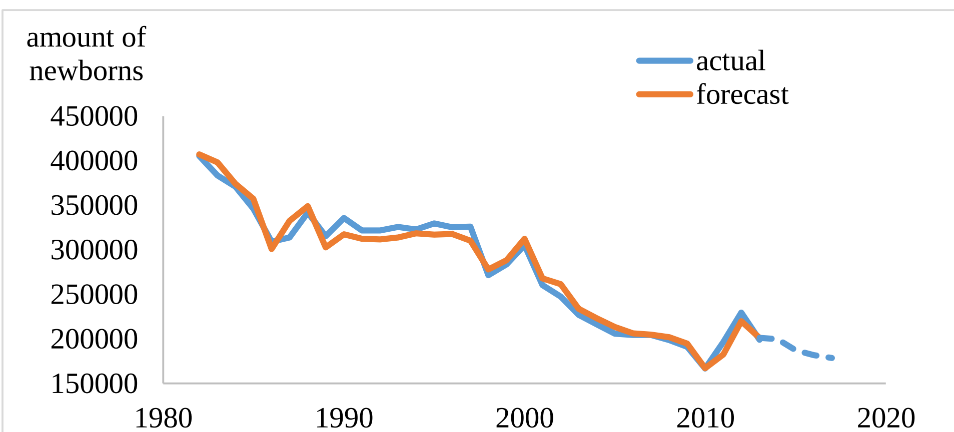
(4) $\ln(\text{Newborn}) = \beta_0 + \beta_1 \text{trend}_t + +\beta_{2\sim 12} \sum(\text{Chinese Zodiac Animal}_t) + \varepsilon_i$



	Regression 1	Regression 2	Regression 3	Regression 4
MAPE	26.57%	20.65%	11.70%	29.76%

c. Our best model

By comparing MAPEs of all models, we found that regression 3 with lag1, external data, 11 dummies and t in the regression would be our best model. Since this method need external data and can only forecast next year, so we roll-forward our forecast, and use our forecast for 2014 (2015, 2016) to forecast 2015 (2016,2017). For the external data (married rate), we use double exponential to forecast the following 3 years and apply the forecast to our model. The forecasts of newborns in the next 3 years are as follows.



	Newborn (f)
2015	187,122
2016	181,904
2017	178,580

d. Applying city weight into our forecast.

We also use double exponential to forecast city weight since there's trend but no seasonality in the series. Then we multiply the city weight by our forecast amount, the amount of newborns are as follows.

You can see that although the amount of newborns in the whole country is going down, our potential markets are experiencing slightly growth.

	New Taipei City	Taipei City	Taichung City	Tainan City	Kaohsiung City
City weight					
2015	15.81%	20.80%	3.36%	0.67%	5.87%
2017	17.44%	22.86%	3.73%	0.58%	5.15%
Amount of newborns in each city					
2015	29,584	38,921	6,287	1,254	10,984
2017	31,144	40,823	6,661	1,036	9,197

Conclusions

1. Newborns per city data is too short to capture a full cycle of seasonality, we decided to create city weights and use the country global series of newborns. After a few years of data collection a city dedicated forecast could be deployed to increase performance and precision.

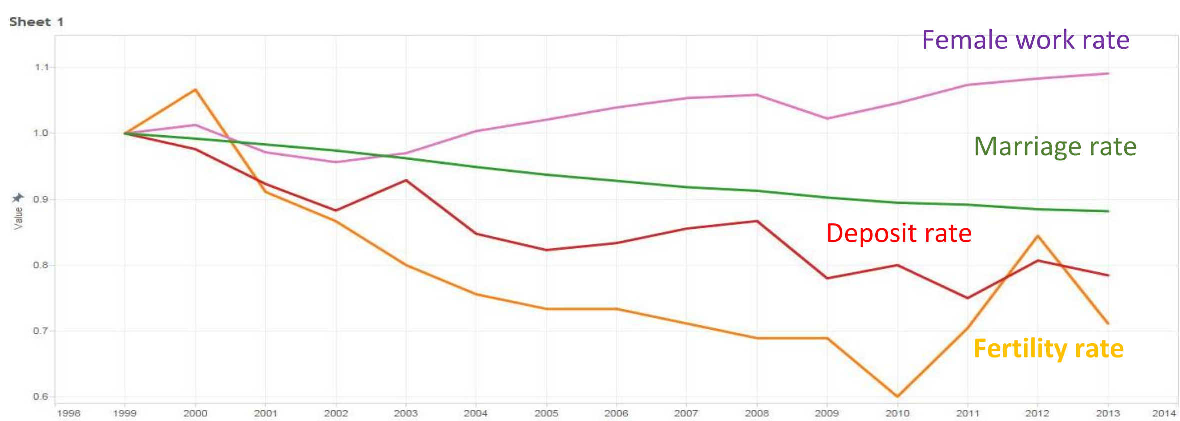
2. We conclude that cities of Taipei, New Taipei and Hsinchu should be considered for CS/DCC expansion while Chiayi and Taoyuan County should downsize their resources.

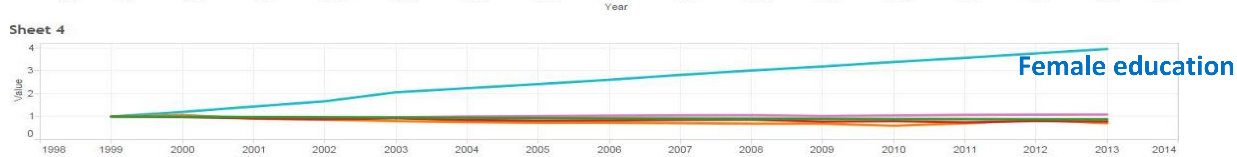
3. Decisions are divided into short term and long term actions, each term associated with a specific method and horizon. Data collection of this forecast depends on government disclosure, and we need to update the forecast every year.

Appendix

a. Standardized related series

We can tell that married rate, deposit rate and fertility rate (newborn babies) have a downward trend while female education and female work rate have an upward trend from the plot below.





b. Sample of a 10 rows per series.

data/year	81	82	83	84	85	86	87	88	89	90
Newborn babies (amount per year (000))	414	405	383	371	346	309	314	342	315	336
Married rate (new married couples per year (0/00))	9.29	8.83	8.5	8.21	8.02	7.51	7.46	7.84	7.89	7.05
Bank deposit rate (percentage per household (%))	26	26	24	23	24	22	22	22	22	23
Female work rate (female/male ratio, age 15< (%))	77	79	80	81	81	82	82	83	83	84
Female education (percentage of women attended university, age 15< (%))	7	8	10	12	15	16	17	18	20	21

c. Statistic table of different regression

	dependent variable: ln(new born)			
	(1)	(2)	(3)	(4)
Intercept	12.5084***	12.4036***	4.8787	12.8308***
trend	-0.0106***	-0.012***	-0.0047	-0.0125***
lag-1--ln(new born)			0.6238	
lag-1--married rate	0.0378	0.0547**	-0.0066	
Rat	0.0537		0.0067	0.0525
Ox	0.029		-0.0082	0.0311
Tiger	-0.1008	-0.1127**	-0.1318**	-0.1046**
Rabbit	-0.0321		-0.01	-0.0624
Dragon	0.0367	0.0306	0.0594	0.0296
Snake	-0.0862		-0.1036**	-0.0782
Horse	0.0235		0.0457	0.0176
Sheep	0.0243		-0.0357	-0.0116
Cock	0.0497		0.0372	0.0701
Dog	0.0573		0.0199	0.0677
Pig	0.0487		0.0166	0.0626
R square	0.9573	0.8401	0.9694	0.947
adjusted R square	0.8779	0.8001	0.8979	0.8676
MAPE	0.2657	0.2065	0.1170	0.2976