POSTPONEMENT OF CHILDBEARING DURING THE COVID-PANDEMIC

Using Internet data

Siwen Liu
Faculty of Science
May 17, 2023

0 Outline

1 Outline

1 Introduction

Previous external shocks, such as economic recessions or pandemics, have been shown to have an impact on childbearing patterns. It is also the case for the COVID pandemic.

In this context, immediate and effective forecasts are beneficial to facilitate the formulation of relevant policies. To quickly depict the childbearing pattern under emergency shocks, Internet data is widely used to predict either child-bearing intention or fertility.

In this research, two internet data sources, specifically Twitter and Google Trends, are combined with the birth count to project the change of birth. The predictive power of different platforms is assessed in this research.

2 Outline

2 Research question

- Can internet data help predict changes in childbearing in Flanders?
- Can generated keywords improve prediction performance?
- Which internet platforms have predictive power for changes in childbearing in Flanders?
- Which models result in the best prediction quality for childbearing changes in Flanders?

3 Outline

3 Internet data and official data

- Data source
  1. Official data: birth count in Flanders from 2009 to 2021
  2. Search engine data: Google Trends
  3. Social media platform data: Twitter
- Motivation
  1. birth count function as a fertility indicator in the model.
  2. The primary source of information for women planning to become pregnant is the search engine, and the most popular search engine in Belgium is Google
  3. According to a survey by Kernohan, online social platforms can serve as opportunities for social networking or experience sharing among women who plan to become pregnant.

3 Keywords for Internet data

The four different aspects are considered as the initial keywords to generate relevant keywords in Google Trends and Twitter.

- Initial keywords
  1. Positively correlated with Fertility: eisprong, bevalling, vruchtbaarheid, bevruchting, ovulatie.
  2. Negatively correlated with Fertility: anticonceptie, miskraam, abortus.
  3. Fertility Test: ovulatietest(en), zwangerschapstest(en).
  4. Economic Aspect: werkloosheid.
- Motivation
  1. Fertility-related words(positive, negative, and test) are undoubtedly essential to explore fertility-related information.
  2. As Berrington and Comolli have found: personal economic situations can also lead to changes in childbearing intentions and patterns. Hence, the keyword reflecting economic situation, such as werkloosheid, is considered.

3 Keywords generation

- Keywords generation for Google Trends
  1. Top Searches section in Google Trends recommendation is utilized.
  2. Interrogative words and repeated words are neglected.
- Keywords generation for Twitter
  1. Association strength between Twitter(target corpus) and SoNAR new media corpus (reference corpus) is used to detect potential keywords.
  2. Pointwise mutual information (PMI) & loglikelihood ratio as standard to choose words.
  3. High PMI and loglikelihood indicate the words has a strong association with the preliminary four aspects.

3 Keywords generation

- Generated words for Google Trends

Dimensions	Words
Positive correlated words	menstruatie(period), zwanger(pregnant), innesteling(implantation), baby, bevruchting eicel(fertilization ovum)
Negative correlated words	anticonceptiepil(birth control pill), anticonceptie staafje(contraceptive stick), miskraam symptomen(miscarriage symptoms), abortus pil(pill abortion)
Fertility test	clearblue ovulatietest(ovulation test), clearblue, kruidvat zwangerschapstest(kruidvat pregnancy test)
Economic aspect	werkloosheid(unemployment)

- Generated words for Twitter

Dimensions	Words
Positive correlated words	epidurale, pijnvrije(pain free), partnerverlof(partner's maternity leave), bevallingtrauma(childbirth trauma), oncofreezing(egg freezing)
Negative correlated words	abortuspil, familieplanning, zwangerschapsafbreking(termination of pregnancy), anticonceptiekoste(contraceptive cost), contraceptie(contraception)
Fertility test	clearblue
Economic aspect	werkloosheid

4 Outline

4 Data Exploration

- Graph Exploration

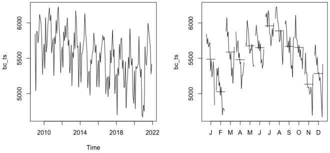


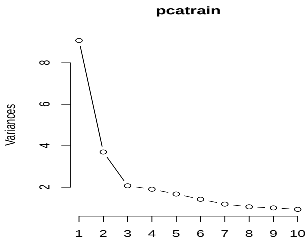
Figure: Time series and Seasonal plot of birth count

- Augmented Dickey-Fuller test for birth count

ADF test	**P-value**	**Stationarity**
$log(bc_t s)$	0.9843	non stationary
$diff(log(bc_t s))$	2.999e-08	stationary
$diff(diff(log(bc_t s), lag = 12))$	5.707e-10	stationary

4 Variable selection

- Data, which starts in January 2009 and ends in December 2019, is split into the training set and the rest data is treated as the test set.
- Principle variables is applied to select variables that account for the majority variance in the training set.
- According to Kaiser's rule, retain 7 components with eigenvalues larger than 1.
- Variables account for most of the variance of each component is further selected.



4 Model Construction

- Baseline Model– SARMA model

$$\left(I-\phi_1 L-\ldots-\phi_p L^P\right)\left(I-\Phi_1 L^s-\ldots-\Phi_P L^{sP}\right) log(BCi^t)$$
$$=c+\left(I-\theta_1 L-\ldots-\theta_q L^q\right)\left(I-\Theta_1 L^s-\ldots-\Theta_Q L^{sQ}\right) u_t \quad (1)$$

- Regression model with time series error

$$log(BC_i^t)=\beta_i+\sum_w^W \beta_{i,w}^t * Google_{i,w}^t+\sum_w^W \omega_{i,w}^t Tweet_{i,w}^t+\epsilon_i^t \quad (2)$$

$$\epsilon_i^t=a+\psi_1 \epsilon_i^{t-1}+\psi_2 \epsilon_i^{t-2}+\ldots+\psi_p \epsilon_i^{t-p}+u_t-\theta_1 u_{t-1}-\ldots-\theta_q u_{t-q} \quad (3)$$

4 Model Results

- Sensitivity analysis on Google Trends' variables

Model	Exogenous variable	AIC	residual correlogram	Box.test	RMSE
SARMA(3,1,0)(1,1,0)		-522.61	white noise	P-value>0.05	0.07321694
Model 1: Regression model with SARMA (3,1,1)(1,1,1) errors	Google Trends: zwangerschapstest, miskraam,	-541.91	white noise	P-value>0.05	0.05421308
Model 2: Regression model with SARMA (3,1,1)(1,1,1) errors	Google Trends: zwangerschapstest, miskraam, miskraam symptomen	-548.61	white noise	P-value>0.05	0.05629741
Model 3: Regression model with SARMA (3,1,1)(1,1,1) errors	Google Trends: zwangerschapstest, miskraam, werkloosheid	-542.28	white noise	P-value>0.05	0.05629754
Model 4: Regression model with SARMA (3,1,1)(1,1,1) errors	Google Trends: zwangerschapstest, miskraam, miskraam symptomen, werkloosheid	-548.83	white noise	P-value>0.05	0.05876636

- Sensitivity analysis on Twitter's variables

Model	Exogenous variable	AIC	residual correlogram	Box.test	RMSE
SARMA(3,1,0)(1,1,0)		-c.61	white noise	P-value>0.05	0.07321694
Model 5: Regression model with SARMA (3,1,1)(1,1,1) errors	Twitter: bevalling, anticonceptie	-540.65	white noise	P-value>0.05	0.05639522
Model 6: Regression model with SARMA (3,1,2)(1,1,1) errors	Twitter: bevalling, anticonceptie, werkloosheid	-538.79	white noise	P-value>0.05	0.0534121

4 Model results

- Sensitivity analysis on werkloosheid from different platforms

Model	Exogenous variable	AIC	residual correlogram	Box.test	RMSE
SARMA(3,1,0)(1,1,0)		-522.61	white noise	P-value>0.05	0.07321694
Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: werkloosheid	-539.66	white noise	P-value>0.05	0.06297479
Regression model with SARMA (3,1,1)(1,1,1)	Twitter: werkloosheid	-538.69	white noise	P-value>0.05	0.05545581

- Sensitivity analysis on all Internet variables

Model	Exogenous variable	AIC	residual correlogram	Box.test	RMSE
SARMA(3,1,0)(1,1,0)		-522.61	white noise	P-value>0.05	0.07321694
Model 7: Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: miskraam, zwangerschapstest Twitter: bevalling, anticonceptie	-540	white noise	P-value>0.05	0.05347966
Model 8: Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: miskraam, zwangerschapstest, miskraam symptomen Twitter: bevalling, anticonceptie	-539.7	white noise	P-value>0.05	0.05333436
Model 9: Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: miskraam, zwangerschapstest, werkloosheid Twitter: bevalling, anticonceptie	-539.93	white noise	P-value>0.05	0.05343789
Model 10: Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: miskraam, zwangerschapstest Twitter: bevalling, anticonceptie, werkloosheid	-538.67	white noise	P-value>0.05	0.05333369
Model 11: Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: miskraam, zwangerschapstest, miskraam symptomen, werkloosheid Twitter: bevalling, anticonceptie	-539.63	white noise	P-value>0.05	0.05344107
Model 12: Regression model with SARMA (3,1,1)(1,1,1)	Google Trends: miskraam, zwangerschapstest, miskraam symptomen Twitter: bevalling, anticonceptie, werkloosheid	-538.61	white noise	P-value>0.05	0.05320282

4 Prediction result

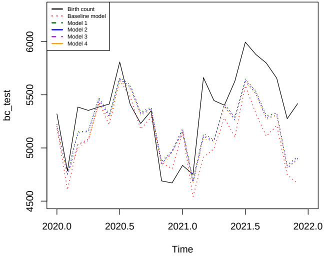


Figure: Prediction result for model with variables from Google Trends

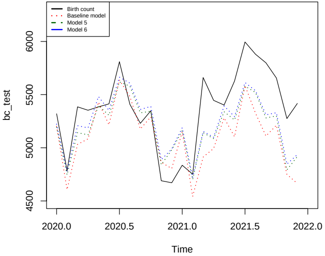


Figure: Prediction result for model with variables from Twitter

4 Prediction result

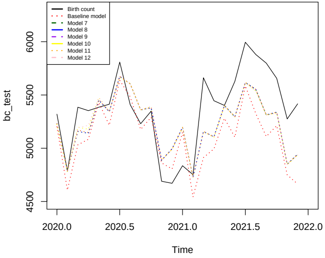


Figure: Prediction result for model with variables from Google Trends

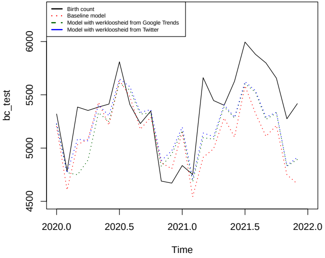


Figure: Prediction result for model with variables from Twitter

4 Conclusion

- Including Internet data can help predict changes in childbearing in Flanders.
- Both Twitter and Google Trends have predictive power for changes in childbearing.
- The most common keywords can already collect the majority of the information from the Internet.
- Fertility-related factors from Google Trends have better predictive performance than those from Twitter.
- Economic factors from Twitter have better predictive performance than that from Twitter.
- The model maximizing the prediction quality is the model combing fertility-related factors from Google Trends and economic factors from Twitter.

Thanks for listening!
Any suggestions or questions?