ISI OTTAWA 2023
64TH WORLD STATISTICS CONGRESS

isi

IPS Paper

MODEL-BASED SINGLE-MONTH UNEMPLOYMENT ESTIMATES FOR THE BRAZILIAN LABOUR FORCE SURVEY INCORPORATING GOOGLE TRENDS DATA

Author: Dr Denise Silva

Coauthors: Caio Gonçalves, Luna Hidalgo, Jan van den Brakel

Submission ID: 1093

Reference Number: 1093

Brief Description

The Brazilian Labour Force Survey publishes monthly unemployment estimates based on three month rolling data and the need to produce single-month estimates to monitor the labour market emerged after the COVID-19 outbreak.

Multivariate state space models that integrate survey data and Google Trends series for nowcasting were developed.

High dimensionality problems were solved using a dynamic factor state space model.

(Joint work with Caio Gonçalves/Fundacao Joao Pinheiro, Luna Hidalgo/IBGE and Jan van den Brakel/Statistics Netherlands- Maastricht University).

Abstract

The Brazilian Labour Force Survey publishes monthly unemployment estimates based on three month rolling data at the national level. The need to produce single-month estimates to monitor the labour market emerged markedly after the COVID-19 outbreak at both federal and state levels. Multivariate state space models that integrate survey data and Google Trends series for nowcasting were developed. Their potential to improve estimates for small samples at state level, or for specific population groups such as young people, were evaluated. A different set of searched keyword series were considered for each state as Brazil is a large country and has diverse workforce behaviour. In addition, three approaches for targeting the predictors in the dimensionality reduction process were compared: elastic net, clustering, and bivariate state space models. High dimensionality problems were solved using a dynamic factor state space model. The analysis period spans from January 2012 until December 2021. The models also account for the autocorrelation of sampling errors due to sample overlap and the increased volatility in the labour force series in 2020. The results indicate that the use of bivariate structural models as an intermediate stage for keyword reduction constitutes a better strategy to select the correlated Google Trends series, when compared to elastic net and clustering. The significance of multivariate models incorporating big data is noticeable at the national level and nowcast estimates using Google Trends series outperformed the univariate model. On the other hand, no evidence was found that the trajectories of the selected searches bring useful information to improve state level or youth unemployment estimates.