Research proposal

A nowcasting model for Medellín city

María Camila Vásquez Correa[†], Juan Carlos Duque [‡] and Jairo Alejando Gómez[‡]

mvasqu49@eafit.edu.co, jduquec1@eafit.edu.co, jagomeze@eafit.edu.co

[†]Mathematical Engineering, Universidad EAFIT
[‡] Research group in Spatial Economics (RiSE), Universidad EAFIT

21 de marzo de 2019

1. Problem definition

Macroeconomic variables such as unemployment rates and GPD are key for the conduct of monetary policies, macro-prudential policies and fiscal policies. More specifically, the real gross domestic product (GPD) is a summary of the health of an economy (Bragoli & Modugno, 2017) and is one of the most heavily monitored indicators.

Also, unemployment rate, released on a monthly basis (DANE, 2018), helps the government tracking the state of the economy and, in the Colombian case, is the input for new policies and the redesign of existent ones with the aim of reducing this rate and improve the quality of life of Colombians.

The period used to update unemployment rate and GDP does not allow the policy makers to take actions in extreme circumstances, or in anticipation of a decrease in the indicator based on fresh data: they have to wait a fairly long time until the indicator is released to make decisions with it. Our world is changing faster each day and this requires decisions with the same velocity. This implies that the conventional methodologies, such as surveys and the traditional data collection that takes two months to generate an indicator as important as the GDP, begin to be inappropriate to support the decision making.

To face this problem, some authors have proposed nowcasting models to estimate these indicators with a higher frequency, allowing their respective economies to partially reach those gaps. Nowcasting is a contraction of *now* and *forecasting* and it is defined as "the prediction of the present, the very near future and the very recent past"(Elliott & Timmermann, 2013). These models are better described in section 3.

In this work, we aim to investigate a new way of producing a macroeconomic indicator for Medellín city, by exploring the potential of the recent developments in the field of artificial intelligence, especially in convolutional neuronal networks, for the design and implementation of a

nowcasting model that allows the appropriate authorities to measure in real time (or at least with a much higher frequency that is currently available) an economic variable of interest.

2.   Objectives

2.1.   General objective

Design a nowcasting model for Medellín city to generate a macroeconomic indicator with a higher frequency than what is currently available by processing images of cameras located in strategic places of the city.

2.2.   Specific objectives

- Perform a literature review of the state of art for nowcasting models based on digital image processing and artificial intelligence.

- Determine the information to be extracted from the image as well as the track segments to be monitored.

- Determine and characterize the macroeconomic variable to estimate.

- Design an econometric strategy that makes good use of the data obtained and the macroeconomic variable.

3.   State of the art

The nowcasting problem is not new. It has been treated by several authors in several countries in order to generate a real-time indicator for their decision makers, using diverse types of data and models to perform the forecast, as described bellow.

In Canada, Bragoli & Modugno (2017) propose a model for the nowcasting of the GDP considering some variables from the United States. The problem is clear: the value of the gross domestic product of the current quarter is generated with a delay of two months, so the country has a bigger delay compared to other developed countries such as Japan or the United Kingdom. Their target variable is the quarterly GDP and their input series are the purchase management index, employment, manufacturing shipments, retail sales, exports and imports, among others. A lot of these variables are obtained in surveys and trade indicators.

In Urasawa (2014), authors propose another model for the nowcasting of the real GDP of Japan, selecting monthly indicators such as industrial production, employment and retail sales. They follow the lead of some other authors that have models for the short-term forecast of the real GDP in various economies. For instance, they mention Lahiri & Monokroussos (2013), where authors study the effects of survey data in nowcasting the U.S GDP, and Karim *et al.* (2010), where there is a model for nowcasting the French GDP.

For macroeconomic variables, the most common nowcasting models include the GDP. Some other examples include models for the Turkish economy, which is proposed by Modugno *et al.*

(2016), and considers financial data and survey data for the nowcasting; The Czech model, detailed in Rusnák (2016) and the euro-zone model detailed in Maximo & Gabriel (2008).

For the models discussed above, the data used for the nowcasting has economic nature (surveys and financial indicators), and, in most models, the measured variable is GDP. There is a great opportunity for this research project to explore ways of including new types of data (processed images from the city) and to estimate new variables (unemployment, for instance).

4.   Justification

In the literature reviewed in section 3, the data used to build and test the nowcasting models came from conventional sources: surveys of financial data, historical data, and other economic indicators with higher frequency than the GDP to build and test a nowcasting model. These models work primary by constructing a different approach to the interest value using another economic variable that explains a portion of the interest value. However, in this project we aim to provide a way to build an indicator using a totally different type of data: images from cameras. The model we propose is rooted in real-time image processing of the city life, what is happening at each moment and how can it help us to determine the value of an important economic variable.

Also, the method proposed for this project combines convolutional neural networks (CNN) with time series econometrics, which is an important innovation in the field. Providing a network with the appropriate information, we are going to be able to generate forecasting of a time series with a totally different input data and produce results with a higher frequency than it is currently available.

Finally, a nowcasting of a macroeconomic variable in Medellín city will allow the competent authorities to take action in a lot of circumstances, to design policies around the real state of the economy of the city, for example, to decrease the level of unemployment by knowing the effectiveness of the existing policies in real time.

5.   Scope

This is an ambitious project and we expect to develop it in a year and a half. To this end, we are going to go through three investigations to develop each one of the stages of the project and write a final paper with the results of the whole investigation. In section 6 we describe the steps that we will take in this first research stage. The reasons that lead us to divide the project are:

- The data processing and collection processes imply a lot of sub-processes, such as learning new technologies and contacting city agents.

- This is a multidisciplinary problem, and we expect to involve some people inside and outside the Research Group in Spatial Economics (RiSE). Then, we need to take into account the required time to contact them, as their inputs are essential to produce the model we want to deliver.

As said above, in this first stage, we attempt to design a first trained model, with simulated data, using a convolutional neural network, that allows us to extract data from the principal track

segments of Medellín in order to feed the final nowcasting model proposed in the final practice. In sections 6 and 7 we describe better the steps to be taken this stage of the research.

6.   Proposed methodology

For the development of the first phase of this research, the methodology will have the next stages:

- The use of CNN to propose a model for image analysis.

- Working with RISE group in order to get a good understanding of the image processing technology.

- Understanding of the variables taken into account, and models of nowcasting proposed for that variables, in order to select the most appropriate one.

- Designing a model containing information extracted from the principal track segments of Medellín and the information they bring. A CNN is implemented to extract information from images coming from this segments in order to measure our interest variable.

- Weekly meetings with the tutors and the Research Group in Spatial Economics in order to check the project progress.

7.   Schedule

The next paragraphs describe the activities to develop in this project. Table 1 shows the schedule.

1. Writing the present research proposal.

2. Perform a complete literature review, part of which was included in the state of art in section 3.

3. Designing the final model:

    - Determining the information that will be extracted from the image.
    - Determining the macroeconomic variable that will be measured and its frequency.
    - Determining the econometric strategy to apply to the data once they are obtained.

4. Learning and exploring digital image processing technology, and associated libraries.

5. First implementation of the image procesing methodology with example data

6. Determination of the track segments to be monitored in Medellín.

7. Search and contact key actors to get historical videos of selected places in Medellín.

8. Start writing a paper:

    a) Paper's introduction: first draft.
    b) Paper's state of art: first draft.
    c) Paper's methodology section: first draft.

Cuadro 1: Schedule overview.

	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19
1	■	■																	
2		■	■	■	■														
3						■	■	■	■										
4									■	■		■	■						
5												■	■	■	■	■	■		
6																		■	■
7			■	■	■	■	■	■	■	■	■	■	■	■	■	■			
8 (a)			■	■	■	■													
8 (b)							■	■	■	■	■								
8 (c)												■	■	■	■	■	■		

8. Budget

Universidad EAFIT provides databases for the literature review, software licenses to implement the computer models and the required time of the tutor to supervise this project. Also, RiSE provides software and the time of its members to guide the development of this project.

9. Intellectual property

According to the internal regulation on intellectual property within Universidad EAFIT, the results of this research practice are product of *María Camila Vásquez Correa*, *Juan Carlos Duque Cardona* and *Jairo Alejando Gómez Escobar*.

In case further products, beside academic articles, that could be generated from this work, the intellectual property distribution related to them will be directed under the current regulation of this matter determined by Universidad EAFIT (2017).

The next are the general obligations withing the RiSE group for practice students:

- The results of this practice are framed within the PEAK Urban project. PEAK Urban is a 4-year international and multidisciplinary research program led by the University of Oxford and funded by the Research Councils UK Global Challenges Research Fund (RCUK GCRF), in which the RiSE group participates as a partner along with research groups from the University of Oxford, University of Peking, University of Cape Town, and the Indian Institute for Human Settlements.

- By accepting the research practice, students are aware that during the development of the same they can join other researchers from the PEAK Urban network as co-authors, according to the training needs and specific experience that arise during the first phases of the practice.

- The practice students commit themselves to continue participating actively in the development of the research until its completion, as co-authors, if the objectives are not achieved within the deadlines stipulated in the research practice proposal.

- The writing of texts related to the development of the practice will be done in English and in a research article format for academic or scientific publication. Students should request a detailed revision of the texts written to avoid plagiarism and ensure the highest quality standards.

- The datasets acquired or generated during the development of the research practice are for exclusive use in the investigations carried out by the RiSE group.

- The data licensed from third parties, such as the case of the Flickr database, satellite images and licensed use cartography, can NOT be copied or downloaded from the RiSE group's computer equipment. Any use that is made of this information must be done in the RiSE laboratory equipment and this information can not leave the same laboratory. Failure to comply with this clause will result in the immediate cancellation of the practice and the opening of a disciplinary process at the University.

- In the academic texts produced in collaboration with any of the RiSE researchers, the textual copy, or translation, of any academic text (even a phrase) or the use of ideas without due citation is strictly prohibited.

- **Failure to comply with any of these obligations will result in the automatic suspension of the practice and RiSE reserves the right to continue the investigation when and with whom it deems appropriate.**
  As a record, the student signs:

Referencias

Bragoli, Daniela, & Modugno, Michele. 2017. A now-casting model for Canada: Do U.S. variables matter? *International Journal of Forecasting*, **33**(4), 786 – 800. "http://www.sciencedirect.com/science/article/pii/S0169207017300341".

DANE, Departamento Administrativo Nacional de Estadística. 2018 (Jun). *Gran encuesta integrada de hogares (GEIH) Mercado laboral.* https://www.dane.gov.co/index.php/estadisticas-por-tema/mercado-laboral/empleo-y-desempleo.

Elliott, G., & Timmermann, A. 2013. *Handbook of Economic Forecasting.* Handbook of Economic Forecasting. Elsevier Science. https://books.google.com.co/books?id=-_dSuakNHWYC.

Karim, Barhoumi, Olivier, Darné, & Laurent, Ferrara. 2010. Are disaggregate data useful for factor analysis in forecasting French GDP? *Journal of Forecasting*, **29**(1-2), 132–144. https://onlinelibrary.wiley.com/doi/abs/10.1002/for.1162.

Lahiri, Kajal, & Monokroussos, George. 2013. Nowcasting US GDP: The role of ISM business surveys. *International Journal of Forecasting*, **29**(4), 644 – 658. http://www.sciencedirect.com/science/article/pii/S0169207012000374.

Maximo, Camacho, & Gabriel, Perez-Quiros. 2008. Introducing the euro-sting: Short-term indicator of euro area growth. *Journal of Applied Econometrics*, **25**(4), 663–694. `https://onlinelibrary.wiley.com/doi/abs/10.1002/jae.1174`.

Modugno, Michele, Soybilgen, Barış, & Yazgan, Ege. 2016. Nowcasting Turkish GDP and news decomposition. *International Journal of Forecasting*, **32**(4), 1369 – 1384. `http://www.sciencedirect.com/science/article/pii/S0169207016300693`.

Rusnák, Marek. 2016. Nowcasting Czech GDP in real time. *Economic Modelling*, **54**, 26 – 39. `http://www.sciencedirect.com/science/article/pii/S0264999315004034`.

Universidad EAFIT. 2017. *Reglamento de propiedad intelectual.* `http://www.eafit.edu.co/institucional/reglamentos/Documents/Reglamento_Propiedad_Intelectual.pdf`.

Urasawa, Satoshi. 2014. Real-time GDP forecasting for Japan: A dynamic factor model approach. *Journal of the Japanese and International Economies*, **34**, 116 – 134. `http://www.sciencedirect.com/science/article/pii/S0889158314000422`.