Nowcasting economic activity with electronic payments data:
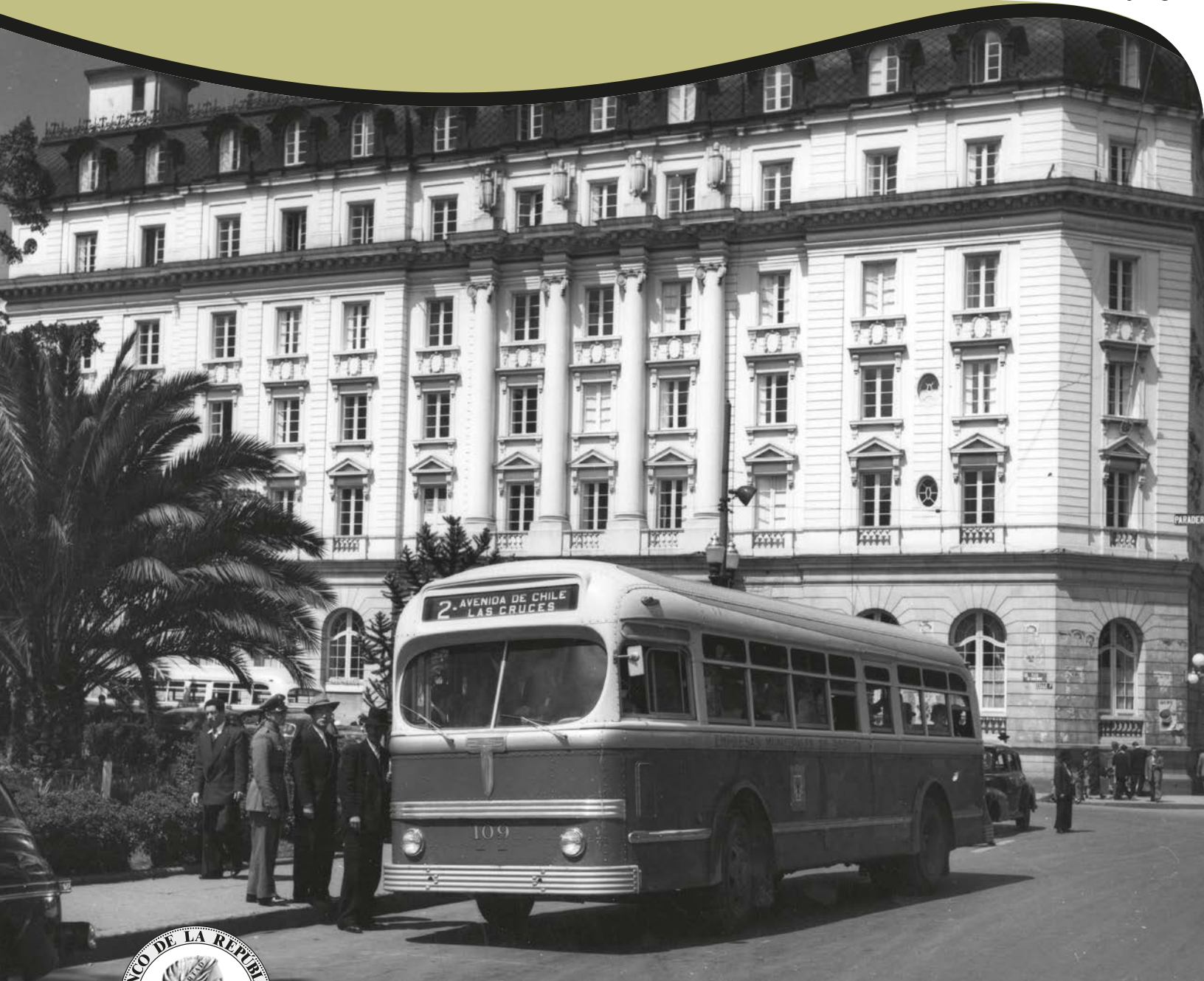
A predictive modeling approach [1]

Carlos León[2]   Fabio Ortega[3]

Abstract

Economic activity nowcasting (i.e. making current-period estimates) is convenient because most traditional measures of economic activity come with substantial lags. We aim at nowcasting ISE, a short-term economic activity indicator in Colombia. Inputs are ISE's lags and a dataset of payments made with electronic transfers and cheques among individuals, firms, and the central government. Under a predictive modeling approach, we employ a non-linear autoregressive exogenous neural network model. Results suggest that our choice of inputs and predictive method enable us to nowcast economic activity with fair accuracy. Also, we validate that electronic payments data significantly reduces the nowcast error of a benchmark non-linear autoregressive neural network model. Nowcasting economic activity from electronic payment instruments data not only contributes to agents' decision making and economic modeling, but also supports new research paths on how to use retail payments data for appending current models.

Keywords: forecasting, machine learning, neural networks, retail payments, NARX.

JEL classification: C45, C53, E27

---

[2] Financial Infrastructure Oversight Department, Banco de la República; CentER, Tilburg University; cleonrin@banrep.gov.co / carlosleonr@hotmail.com; Banco de la República, Carrera 7 #14-78, Bogotá, Colombia; Tel.+57 1 3430731. [corresponding author]

[3] Financial Infrastructure Oversight Department, Banco de la República; fortegca@banrep.gov.co.

1. Introduction

The lack of timely information about the current state of the economy is a well-recognized problem among policy makers (Evans, 2005). That explains why *nowcasting*, defined as *current-period estimates* (Galbraith & Tkacz, 2017) or *the prediction of the present, the very near future and the very recent past* (Bańbura et al., 2017), has become a standard activity for central banks (see Tiffin, 2016, Hinds et al., 2017).

Economic activity is one among many lagged key macroeconomic variables. Gross domestic product (GDP) estimates are released with a four-week lag in the United States and the United Kingdom; six in Japan; six or seven in the Euro area; eight in Canada; and up to one or two years in Lebanon (see Bańbura et al., 2013, Tiffin, 2016, Bragoli, 2017, Galbraith & Tkacz, 2017). Short-term economic activity indicators attempt to reduce this lag. However, as most short-term indicators tend to rely on a subset of GDP inputs or other non-high-frequency and lagged variables, they usually have a non-negligible delay.

In the Colombian case, the national bureau of statistics (*Departamento Administrativo Nacional de Estadísticas*, DANE) releases GDP quarterly figures with an approximate two-month lag (see DANE, 2017). The same bureau releases the Economic Monitoring Index (*Indicador de Seguimiento a la Economía*, ISE), which is a monthly indicator with an approximate two-month lag that provides the market with a dynamic index of economic activity (see DANE, 2016). Consequently, an accurate current-period estimate of ISE could provide the central bank, financial market participants, and other economic agents (e.g. the government, real sector) with better tools for decision-making and modeling.

Our aim in this article is to answer a single question: is it possible to nowcast changes in the ISE with a dataset of electronic payment instruments?[4] Unlike most literature on economic activity nowcasting, we rely on electronic payments among individuals, firms, and the central government as high-frequency indicators (i.e. the inputs) from which reliable signals of economic activity are to be extracted. Using electronic payments as inputs is not

---

[4] A payment instrument enables the holder or user to transfer funds, such as cash, cheques, debit and credit cards, and electronic transfers. Cash is a non-electronic (i.e. physical) payment instrument. Cards and transfers are electronic payment instruments. Cheques are to be considered electronic payments because they are physical at their inception, but they are cleared electronically. As stated below, we focus on electronic transfers and cheques only.

fortuitous: they convey information about economic activity as they are representative of agents' expenditure.

As nowcasting is better understood as a prediction task, we undertake a *predictive modeling* approach (see Shmueli, 2010) that relies on the well-documented performance and flexibility of artificial neural networks (ANNs) for prediction. Hence, unlike a traditional econometric approach, we do not attempt to test any causal theory or to explain which inputs excel as explanatory determinants of economic activity, but to attain an accurate out-of-sample prediction model by discovering complex structures in data that are not specified in advance. In this vein, following Shmueli (2010), Varian (2014) and Mullainathan and Spiess (2017), our approach departs from traditional econometric methods (i.e. explanatory modeling), and pertains to *machine learning* methods.

Evans (2005) and Giannonne et al. (2008) are among the first and most influential works on nowcasting economic activity. Our work is closely related to that of Galbraith and Tkacz (2017), who claim to be the first to make economic activity nowcasting based on payments data. However, there are several differences with Galbraith and Tkacz (2017) as we (i) use a broader and larger spectrum of payment instruments that incorporates expenditures from individuals, firms, and the central government; (ii) implement a Non-linear AutoRegressive eXogenous Artificial Neural Network model (NARX-ANN) that is documented for its predictive high-performance and flexibility; and (iii) attempt to nowcast economic activity with a model that does not include traditional inputs (e.g. macroeconomic data, financial variables, surveys), but relies on electronic payment instruments data and lags of the corresponding economic activity indicator only.

Results suggest that a dataset containing ISE's lagged data and contemporary electronic payments data allows to make a fair current-period estimate of ISE's change circa two months before its release. That is, nowcasting economic activity with electronic payment instruments data and a NARX-ANN model is feasible. Also, results validate that electronic payments data significantly reduces the nowcast error of a benchmark autoregressive model. Therefore, we contribute to decision-making and modeling by providing a suitable nowcasting alternative for economic activity indicators. Furthermore, our work may be a starting point for other nowcasting attempts based on payments data.

2. Data

The term *nowcasting* is a contraction for *now* and *forecasting*, and has been used for a long time in meteorology and recently also in economics (Bańbura et al., 2013). It consists of exploiting the information that is available early and at higher frequencies to make a current-period or early estimate of a lagged or low-frequency variable.

Nowcasting economic activity entails two main challenges. First, finding a suitable set of high-frequency indicators (i.e. inputs) from which reliable signals of economic activity (i.e. outputs) are to be extracted. Second, selecting an appropriate prediction method for extracting reliable signals of economic activity that serve as fair current-period or early estimates of economic activity. This section presents how we tackle the first challenge.

2.1. Output

Literature on economic activity nowcasting is fairly abundant (see Evans, 2005, Bell et al., 2014, Bec & Mogliani, 2015, Bragoli, 2017). Most literature sets the GDP as the output variable. Our aim is to nowcast the ISE (Economic Monitoring Index, *Indicador de Seguimiento a la Economía*). ISE is a monthly indicator of economic activity that combines information about the production of goods and services pertaining to the most important economic activities in Colombia.[5] It is estimated and released by the Colombian bureau of statistics with an approximate 2-month lag.

By choosing ISE instead of GDP we attain a greater number of observed outputs to work with. ISE is available monthly from January 2000. As payment inputs are available from January 2001 onward (see inputs section below), working on GDP quarterly figures would restrict our sample to about 48 observations, whereas working on ISE provides 192 observations. Moreover, as it is harmonized with quarterly GDP figures, working on ISE provides a fair and coherent estimate of changes in economic activity (see DANE, 2016).

Figure 1 exhibits ISE from January 2001 to December 2016; descriptive statistics and the logarithmic returns of ISE are reported in Table 6 and Figure 6 (in the Appendix). As

---

[5] It corresponds to 111 indicators that track the dynamics of nine economic activities that compose the GDP, and they are weighted according to their contribution to the value added to economic activity. See DANE (2016).

electronic payments data is non-seasonally adjusted (i.e. seasonal patterns are not removed), we use the non-seasonally adjusted ISE, at constant prices. We work with the revised index instead of preliminary estimates; as highlighted by Bell et al. (2014), ultimately, it is the official data that is most important for policy makers.
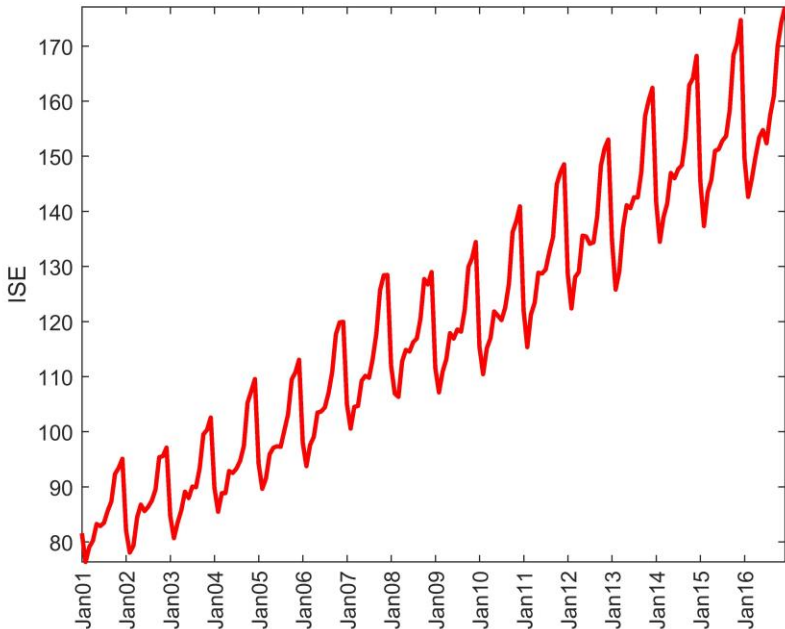

Figure 1. ISE. Non-seasonally adjusted, at constant prices, revised index. Source: DANE.

2.2. Inputs

Most economic activity nowcasting literature relies on macroeconomic data, financial variables, surveys, or a mixture of them (see Evans, 2005, Bańbura et al., 2013, Bell et al., 2014, Bec & Mogliani, 2015, Bragoli, 2017). These inputs tend to convey a non-negligible lag (e.g. days, weeks or even months). Lately, research on how to nowcast economic activity from *online* data has surfaced. In that case, news articles, social media, web search data and other related sources are used to measure economy activity; for instance, McLaren (2011) and Choi and Varian (2012) illustrate how Google query indexes may be used for short-term economic activity prediction.

We choose electronic payment instruments data as our inputs for nowcasting economic activity. Our choice is unusual but has been already put into practice by Galbraith and Tkacz

(2017). Galbraith and Tkacz use a set of electronic payments data comprising the value and number of operations from debit card transactions (i.e. point-of-sale payments) and "small" cheques (i.e. under $50,000) that clear through the Canadian banking system. They argue that both sources of payments information are representative of consumption expenditure, which is a major component of GDP, meaning that their set of inputs provide an incomplete but direct source of information on GDP changes. Also, they argue that electronic payments are available quickly and virtually free of sampling error.

In our case we use a dataset comprising the value and number of operations from two different electronic payment instruments. First, from electronic transfers ordered by individuals, firms, and the central government, which are processed and cleared in both existing automated clearing houses (ACHs), namely ACH Colombia and ACH Cenit. Second, from cheques that are processed and cleared in the local cheque clearing house, Cedec. Unlike Galbraith and Tkacz (2017), we do not consider debit and credit card transactions because their contribution to the total value of electronic payments is rather low (i.e. below 5 percent), and because data is available with a lag that is incompatible with a nowcasting task, i.e. two or three months. Also, we do not limit cheques to being "small". A notable shortcoming common to both works is the absence of information about cash payments –the prevalent payment instrument in Colombia.[6]

ACH Colombia is owned and managed by local banks. It fulfills the nowadays prevalent purpose of ACHs (see McAndrews, 1994), namely to clear transfers between different banks regarding small repetitious payments among economic agents, such as payrolls, supplies, social security, mortgage installments, insurance premiums, dividends, and utility bills. As of 2016, twenty financial institutions clear their clients' transfers in ACH Colombia. On the other hand, ACH Cenit is owned and managed by the central bank. Its main function is processing central government's payments (e.g. payrolls, transfers to municipalities, public social security); thus, its main participant is the central government, which contributes with about 87 percent of its transfers by value (Banco de la República, 2017). Regarding cheques clearing, Cedec is owned and managed by the central bank as well.

---

[6] It has been established that there is a strong preference to use cash (i.e. pay and accept) in the Colombian case (see Banco de la República, 2017).

As exhibited in Table 7 (in the Appendix), as of 2016, payments processed in ACH Colombia, ACH Cenit and Cedec account for 96 and 38 percent of the value and number of operations, respectively, of all electronic payments cleared and settled among banks.[7]

Hence, our argument for using electronic payment instruments data is similar to that of Galbraith and Tkacz (2017). By using a broad set of electronic payments ordered by individuals, firms and the central government we aim at providing a direct and comprehensive source of information about expenditure dynamics of individuals, and the corporate and public sectors. That is, from the national income identity, our choice of inputs attempts to incorporate consumption, investment and government expenditure to economic activity nowcasting. Nonetheless, our source of information is partial: We lack data corresponding to some payment instruments (e.g. cash, credit and debit cards), and to transfers and cheques cleared and settled in the books of banks (i.e. not processed in a clearing house) –among others.

Figure 2 exhibits the six inputs from January 2001 to December 2016. Inputs are non-seasonally adjusted; those corresponding to values are at constant prices –deflated by the consumer price index. Descriptive statistics and visual inspection (see Table 6 and Figure 7, in Appendix) reveal that all series are to be considered non-normally distributed (except the number of operations of Cedec) and potentially heteroscedastic. As in Galbraith and Tkacz (2017), inputs are monthly aggregates of daily transactions, and they are available simultaneously; issues related to mixed-frequency or asynchronous data are absent.

---

[7] As of 2016 (Banco de la República, 2017), the value and number of cheques cleared in banks' books (i.e. not cleared and settled among banks in Cedec) account for about 25 and 44 percent of all cleared cheques, respectively. The value and number of electronic transfers cleared in banks' books account for about 72 and 53 percent, respectively.
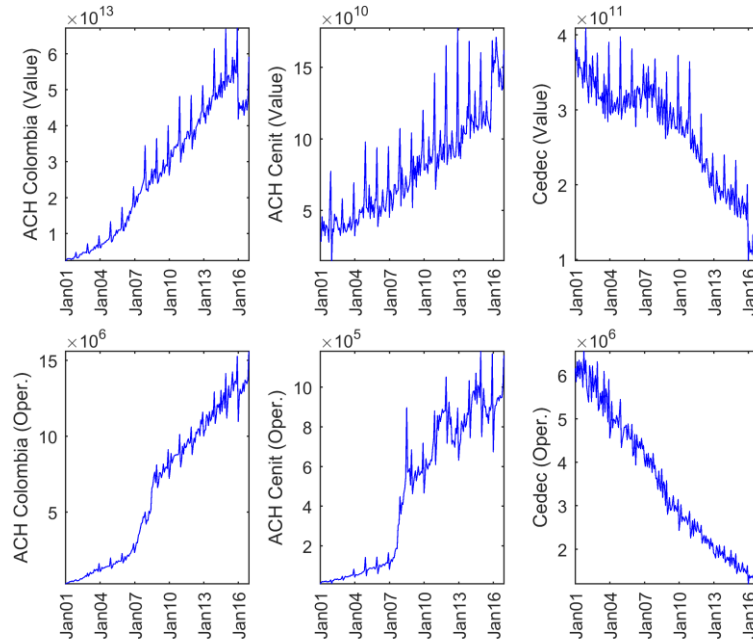
Figure 2. Electronic payment instruments data. Non-seasonally adjusted. First row presents the value (in constant Colombian pesos, deflated by the consumer price index) and the second the number of operations. Source: ACH Colombia and Banco de la República.

Figure 2 and Figure 7 (in Appendix) suggest that inputs display a seasonal behavior that resembles that of ISE (in Figure 1 and Figure 6, in Appendix). The correlation matrix between ISE and the six inputs reveals that there is some degree of linear dependence among series' logarithmic returns.

Variable	1	2	3	4	5	6
1. ISE						
2. ACH Colombia (Value)	**.668**					
3. ACH Cenit (Value)	**.506**	.681				
4. Cedec (Value)	**.569**	.849	.512			
5. ACH Colombia (Oper.)	**.611**	.894	.589	.814		
6. ACH Cenit (Oper.)	**.524**	.780	.733	.658	.803	
7. Cedec (Oper.)	**.546**	.789	.414	.861	.828	.652

Table 1. Correlation matrix of output and inputs. Estimated on logarithmic returns of monthly series (191 observations); inputs corresponding to values are at constant prices, deflated by the consumer price index. Source: DANE, ACH Colombia and Banco de la República; authors' calculations.

7

Also, it is rather clear that the series disclose a well-known dynamic in electronic payment instruments in the Colombian case: the decay of cheques as payment instruments and the simultaneous rise of electronic transfers (see Table 7, in Appendix). Therefore, it is worth highlighting that inputs not only reflect economic activity dynamics but also the evolution of payment instruments. As suggested by Galbraith and Tkacz (2017), using all inputs in a convenient manner –instead of using each one of them in isolation- should allow avoiding false signals about economic activity that may be related to the evolution of payment technology and habits.

3. Prediction method

Forecasting and nowcasting are, by definition, prediction problems. Accordingly, machine learning approaches, which are intended for predictive modeling (see Shmueli, 2010, Varian, 2014), have been used for forecasting and nowcasting economic activity. For instance, Tkacz and Hu (1999) and Aiken (2000) use ANNs to forecast the GDP of Canada and the United States; Tiffin (2016) uses elastic net regression and random forests to nowcast GDP in Lebanon; and Hinds et al. (2017) use elastic net regression in the United Kingdom's GDP case. However, most nowcasting literature is about econometric approaches.[8]

Three main reasons support our choice of machine learning in the form of an ANN for this prediction problem. First, given enough hidden layers and enough training samples, ANNs can closely approximate any function, thus they are able to deal with non-linear relationships between factors in the data (see Bishop, 1995, Han & Kamber, 2006, Demyanyk & Hasan, 2009, Sarlin, 2014, Hagan et al., 2014). This has been reported by Tkacz and Hu (1999) and Zukime and Junoh (2004) as the most plausible reason for ANNs delivering improved forecast accuracy over linear methods when forecasting GDP in the Canadian and the Malaysian cases, respectively. Second, ANNs make no assumptions about the statistical properties of the data (see Zhang et al., 1999, Aiken, 2000, Zukime & Junoh, 2004, McNelis, 2005, Demyanyk & Hasan, 2009, Sarlin, 2014). This is particularly convenient in our case

---

[8] A review of econometric approaches to economic activity nowcasting is beyond the scope of this paper. A comprehensive review is presented in Bańbura et al. (2013).

as the output and most inputs exhibit non-normal distributions and appear to display some degree of heteroscedasticity (see Table 6, figures 6 and 7, in Appendix). Finally, ANNs have proven to be very effective in time series prediction problems, even better than standard econometric approaches (see Kohzadi et al., 1995, Zhang et al., 1999, Misas et al., 2003, McNelis, 2005, Jalil & Misas, 2006, Han & Kamber, 2006, Chaudhuri & Ghosh, 2016, Di Piazza et al., 2016). This is key in our case because, again, nowcasting is a prediction task.

The main disadvantage of ANNs, commonly known as the *black box criticism*, is related to results' opacity and limited interpretability (see Han & Kamber, 2006, Angelini et al., 2008, Witten et al., 2011, Chakraborty & Joseph, 2017). However, as the black box criticism comes from a desire to tie down empirical estimation with an underlying economic theory (McNelis, 2005), this disadvantage is not an issue in our case: our only goal is to test whether electronic payment instruments data and an ANN are able to nowcast economic activity with fair precision –no underlying economic theory is to be tested. We do not aim at producing good estimates of parameters that underlie the relationships between inputs and outputs: we aim at good output estimation. Our implementation of ANNs does not intend to yield a policy model, but only a nowcasting model.

3.1. The NARX artificial neural network model

In the case of time-series econometric models, a commonly used model is the linear autoregressive exogenous model (ARX). In this model [1], the output variable (y_t) is modeled as a linear function of k lagged y_t variables and h exogenous variables (x_t). Solving this linear closed-form problem consists of minimizing the sum of square differences between observed and estimated output variables, y_t and $\hat{y}_t$, respectively.

$$y_t = \sum_{i=1}^{k} \beta_i y_{t-i} + \sum_{j=1}^{h} \rho_j x_{j,t} + \epsilon_t \qquad [1]$$

The simplicity of the ARX model entails a great shortcoming. The linear approximation may not be flexible enough to model a system in a complete manner, as there may be non-

linear processes in the data. A non-linear autoregressive exogenous model may provide a more general and convenient approach to this type of problem.

Based on Lin et al. (1996), let Ω represent a non-linear function, and k and m the output and inputs lag order, respectively, a Non-linear AutoRegressive eXogenous (NARX) model may be represented as in [2]. In this case, the output variable (y_t) is modeled as a non-linear function Ω of k lagged y_t variables and m lagged exogenous variables (x_t).

$$y_t = \Omega[x_t, x_{t-1}, \dots, x_{t-m}, y_{t-1}, \dots, y_{t-k}] \qquad [2]$$

Lin et al. (1996) suggest that the non-linear function Ω may be approximated by an ANN structure, i.e. a NARX-ANN model. Unlike standard econometric approaches to time-series problems, the NARX-ANN model can efficiently be used for modeling non-stationary and non-linear dynamic systems (see Mahmoud et al., 2013, Chaudhuri & Ghosh, 2016, Di Piazza et al., 2016).

In the sample structure exhibited in Figure 3, the output is modeled as a two-layer ANN. The first layer, also known as the *hidden layer*, consists of two *tapped delayed lines* (TDL) that contain the vectors of lagged inputs; three bias scalar terms (b_1^1, b_2^1, b_3^1) with a constant input of 1; and all w^1 weights that connect each input with the three neurons (i.e. the sum operators and their corresponding activation functions, $f^{\,1}$). The second layer, also known as the *output layer*, consists of a single neuron with an operator that sums the product of the weights ($w_{1,1}^2$, $w_{2,1}^2$, $w_{3,1}^2$) and the result from the neurons in the first layer ($f^1(n_1^1)$, $f^1(n_2^1)$, $f^1(n_3^1)$); one bias scalar term, b_1^2, with a constant input of 1; and the second activation function ($f^2(n_1^2)$), which yields the output.
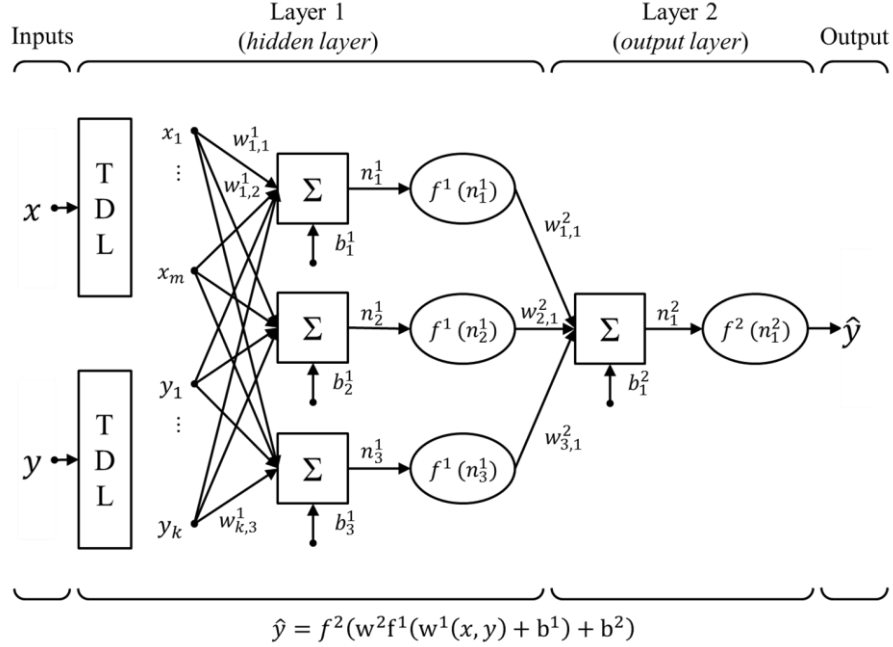
Figure 3. Sample NARX-ANN with two layers. The first (hidden) layer consists of two *tapped delayed lines* (TDL) that contain the vectors of lagged inputs; three bias scalar terms (b_1^1, b_2^1, b_3^1) with a constant input of 1; all w^1 weights that connect each input with the three neurons (i.e. the sum operators and their corresponding activation functions, f^1). The second layer consists of a single neuron with an operator that sums the product of the weights ($w_{1,1}^2$, $w_{2,1}^2$, $w_{3,1}^2$) and the result from the neurons in the first layer ($f^1(n_1^1)$, $f^1(n_2^1)$, $f^1(n_3^1)$); one bias scalar term, b_1^2, with a constant input of 1; and the second activation function ($f^2(n_1^2)$), which yields the output. Based on Hagan et al. (2014) and Di Piazza et al. (2016).

The structure of the ANN in the NARX model depends on the number of layers and neurons, and the choice of transfer functions. About the number of layers, a two-layer network (i.e. one hidden and one output layer) is all that is necessary in most cases (see Zhang et al., 1999, Witten et al., 2011), and it is the most commonly used ANN in economic and financial applications (McNelis, 2005). The number of neurons in the hidden layers corresponds to the complexity of the function that is being approximated, whereas the number of neurons in the output layer depends on the number of outputs –usually, one per output. Regarding the transfer functions in NARX-ANN models, a customary choice for hidden layers is the sigmoid transfer function ($f^1 = 1/(1 + e^{-n})$), whereas a linear transfer function (e.g. $f^2 = n$) is the standard for the output layer (see Hagan et al., 2014).

3.2. Training the NARX artificial neural network model

The learning process of an ANN consists of adjusting all w and b parameters in the structure in order to minimize the error between the prediction and the target (i.e. observed) value for a set of observations. This process, also called training, is somewhat similar to fitting the parameters of a regression model in econometrics.

The most popular ANN learning algorithm is *backpropagation*.[9] Backpropagation learns by iteratively processing a dataset of training observations, comparing network's prediction for each observation with the target value. All parameters w and b are modified in backwards direction, from the output layer, through each layer down to the first layer. In our case, for predicting continuous variables the fit of the ANN is typically measured as a function of square errors, say the root mean square error ($RMSE = \sqrt{\sum_{i=1}^{n}(\hat{y} - y)^2/n}$).

As before, our goal under the nowcasting objective is to provide useful out-of-sample predictions. In this sense, machine learning methods attempt to avoid the *overfitting problem*. The overfitting problem may be described as the model's ability to succeed at fitting in-sample but to fail at fitting out-of-sample (see Shmueli, 2010, Varian, 2014). Overfitting occurs when a model depends too strongly on the details of the particular examples used to produce those (Witten et al., 2011). In this sense, the role of ANNs is to provide general non-linear mappings between a set of input variables and a set of output variables. As in Bishop (1995), the goal is not to memorize the training data, but to model the underlying generator of the data.

In the case of ANNs the overfitting problem may be avoided by means of *early stopping*.[10] In early stopping, instead of allowing the algorithm to attain the minimal in-sample error (at the expense of out-of-sample error), the minimization process is halted before the complexity of the solution inhibits its generalization capability. In this sense, if

---

[9] A complete explanation on the functioning of the backpropagation algorithm is beyond the scope of our paper. The interested reader may refer to Bishop (1995), Mitchell (1997), and Han and Kamber (2006).

[10] The other most common method to avoid overfitting is *regularization*. In regularization, the performance of the ANN includes a term that penalizes network complexity and –thus- forces the function to be smooth and to attain a general mapping of the inputs and outputs. Hagan et al. (2014) demonstrate that there is an approximate equivalence between early stopping and regularization.

training is stopped before the minimum in-sample is reached, then the ANN will effectively be using fewer parameters and will be less likely to overfit.

An intuitive and customary early stopping criterion is *cross-validation*. Data is divided into three independent datasets: *training*, *validation*, and *test*. The three datasets are usually selected with an approximate 70%, 15%, 15% random allocation, respectively (see Hagan et al., 2014). The training dataset is used to train the network (i.e. to minimize the in-sample error between the prediction and the target value). The validation dataset is used simultaneously (i.e. during training) to check how the estimated parameters fit out-of-sample data. When the error on the validation dataset increases for several iterations of the training algorithm, the training is halted, and the parameters that produced the minimum error on the validation dataset are used as the final trained network parameters. After training the network, the error obtained on the test dataset is used to check the future performance of the ANN on out-of-sample data. With cross-validation, a *firewall principle* (Mullainathan & Spiess, 2017) is implemented to guarantee that none of the data involved in the training is used to evaluate the out-of-sample predictive performance of the network.

3.3. A *committee* of artificial neural networks

We train a NARX-ANN model with two layers, as the one presented in Figure 3.[11] Our choice of transfer functions is the customary sigmoid and linear transfer functions for the hidden and output layers, respectively. The model's output and inputs are the logarithmic returns of the corresponding series. Consistent with the year-long seasonality of economic activity, the lag order of the output variable (y_t) is twelve months, whereas exogenous variables are not lagged (i.e. only concurrent values of payment instruments are used).[12] Regarding the number of neurons, several scenarios are considered to implement an *empirical tuning* procedure to pick the choice with the best estimated performance (see Mullainathan & Spiess, 2017). To

---

[11] The NARX-ANN model was implemented in Matlab's Neural Network Toolbox.
[12] As ISE is released with an approximate 2-month lag, the lag order of the output variable starts at the second lag. Including lagged exogenous variables did not improve the prediction performance; for parsimony, the output variable is the only one with lags.

avoid overfitting, we employ early stopping by cross-validation with a 70%, 15%, and 15% random allocation for the training, validation, and test datasets.[13]

A single training process of any ANN may not produce optimal performance, because of the possibility of reaching a local minimum of the performance surface (Hagan et al., 2014). This is caused by ANNs' dependence on the initialization parameters. Therefore, we implement the *committee* method (see Bishop, 1995, Mitchell, 1997, Hastie et al., 2013, Hagan et al., 2014). Under this method we run 1,000 independent training processes for each of the scenarios considered, and estimate their joint output, i.e. the simple unweighted average. This is useful because the joint output of the committee of networks will usually achieve higher performance than any single network used in isolation (Bishop, 1995, Hagan et al., 2014), and because it enables to obtain a *density forecast* for the target variable (Chakraborty & Joseph, 2017).

4. Main results and validations

We employ two standard measures of out-of-sample (i.e. on the test dataset) performance. First, the root mean square error ($RMSE$) between the observed and the predicted logarithmic returns of ISE. Second, the correlation (r) between the observed and the predicted logarithmic returns of ISE.

Table 2 exhibits the average of these two measures, estimated on the 1,000 independent training processes of each scenario, denoted as $\overline{RMSE}$ and $\bar{r}$. Standard deviation for both measures are reported in brackets. In the four scenarios considered, both measures of average out-of-sample performance appear to be fair. The measures do not change noticeably with variations in the number of neurons. The scenario that results in the lowest coefficient of variation (i.e. dispersion relative to the mean) is the one with 30 neurons.

---

[13] Performance is robust to changes in the allocation method. Non-random allocations (e.g. interleaved non-overlapping blocks, successive blocks) of training, validation, and test sets delivered similar performance.

	Number of neurons			
	10	20	**30**	40
$\overline{RMSE}$	.0084 [.0022]	.0082 [.0018]	**.0081 [.0016]**	.0082 [.0017]
$\bar{r}$	.9869 [.0072]	.9873 [.0061]	**.9875 [.0048]**	.9871 [.0055]

Table 2. Predictive out-of-sample performance for selected scenarios of number of neurons in the hidden layer. $\overline{RMSE}$ corresponds to the average *RMSE* estimated on 1,000 independent training processes for each scenario. $\bar{r}$ corresponds to the average correlation coefficient between the observed and predicted log-return series for each scenario. Standard deviation is reported in brackets. The highest performance scenario is in bold. Source: authors' calculations.

Figure 4 displays three series, corresponding to the 30-neuron scenario: first, the observed ISE logarithmic return series, displayed as an empty circle for each observation (circles' centers correspond to the observed series); second, 1,000 predicted logarithmic return series, displayed as a green dot for each prediction, for each observation; third, the corresponding average predicted logarithmic return, portrayed as a black dot for each observation. It is apparent that the 1,000 predictions scatter around the target variable; in some cases, the dispersion around the target variable is high. As expected under a committee of networks method, the average predicted logarithmic returns (black dots) are particularly close to the observed logarithmic returns (empty circles).
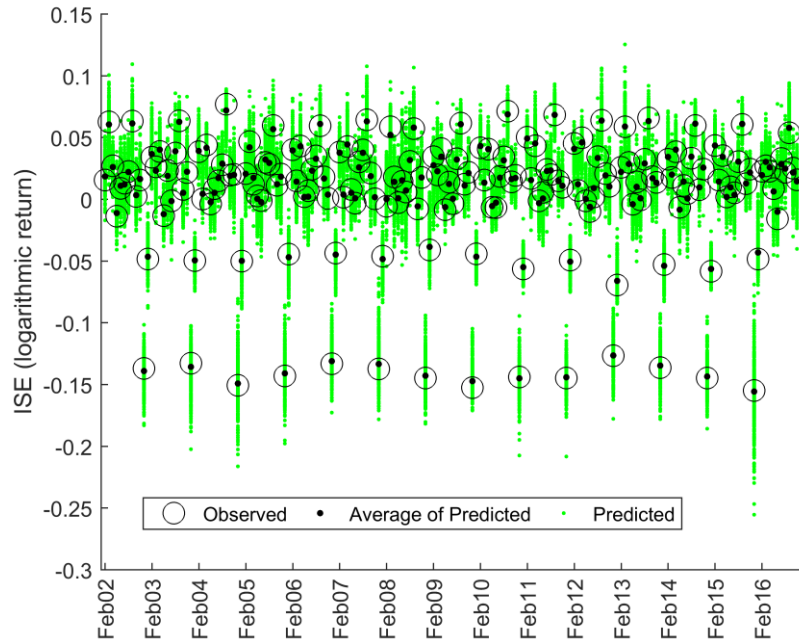
Figure 4. Observed (empty circles), predicted (green dots), and average of predicted (black dots) ISE logarithmic returns. Predicted ISE series correspond to 1,000 independent training processes of the scenario corresponding to 30 neurons. Source: DANE and authors' calculations.

Based on the committee of networks' prediction (i.e. the average prediction of the 1,000 independent training processes), we calculate the committees' corresponding out-of-sample performance measures, $\overline{\overline{RMSE}}$ and $\bar{\bar{r}}$, in Table 3. As expected, the performance increases manifestly: For the 30-neuron scenario, $\overline{RMSE} = .0081$ and $\overline{\overline{RMSE}} = .0026$, which signifies a 67.90 percent decrease in out-of-sample error. Taking into account that Hinds et al. (2017) points out that being able to nowcast United Kingdom's economic quarterly growth with a *RMSE* around .0030 is *relatively impressive*, it is fair to say that nowcasting economic activity in the Colombian case with our choice of dataset and predictive method is feasible.[14]

---

[14] Based on data extracted from the United Kingdom's Office for National Statistics (www.ons.gov.uk), the average quarterly absolute change of GDP is .0067 for the period studied by Hinds et al. (2017). Thus, the ratio of *RMSE* to the average quarterly absolute change is about $0.0030/0.0067 = .4476$; that is, the relative size of the error is about 45 percent. In the case of ISE, the average monthly absolute change is .0361. For the 30-neuron scenario, the $\overline{RMSE}$ and $\overline{\overline{RMSE}}$ are .0081 and .0026, therefore their ratio to the average monthly absolute change is $0.0081/0.0361 = .2244$ and $0.0026/0.0361 = .0720$. That is, the relative size of the error is manifestly lower.

	Number of neurons			
	10	20	**30**	40
$\overline{\overline{RMSE}}$	.0042	.0029	**.0026**	.0026
$\overline{\overline{r}}$	.9971	.9984	**.9987**	.9987

Table 3. Committee of networks' predictive (out-of-sample) performance for selected scenarios of number of neurons in the hidden layer. $\overline{\overline{RMSE}}$ corresponds to the *RMSE* calculated with the average predicted log-return on the 1,000 independent training processes of each scenario. $\overline{\overline{r}}$ corresponds to the correlation coefficient between the observed and the average predicted log-return on the 1,000 independent training processes of each scenario. The highest performance scenario is in bold. Source: authors' calculations.

Figure 5 exhibits the density forecast (i.e. distribution of errors) corresponding to the 30-neuron scenario. It compares the out-of-sample errors attained by each one of the 1,000 predictions to those attained by the committee (i.e. the average prediction). As expected, committee's prediction exhibits a manifestly lower dispersion of errors: Errors' range decreases from [-.1006, .0930] to [-.0069, .0070], with negligible variation in the mean error.
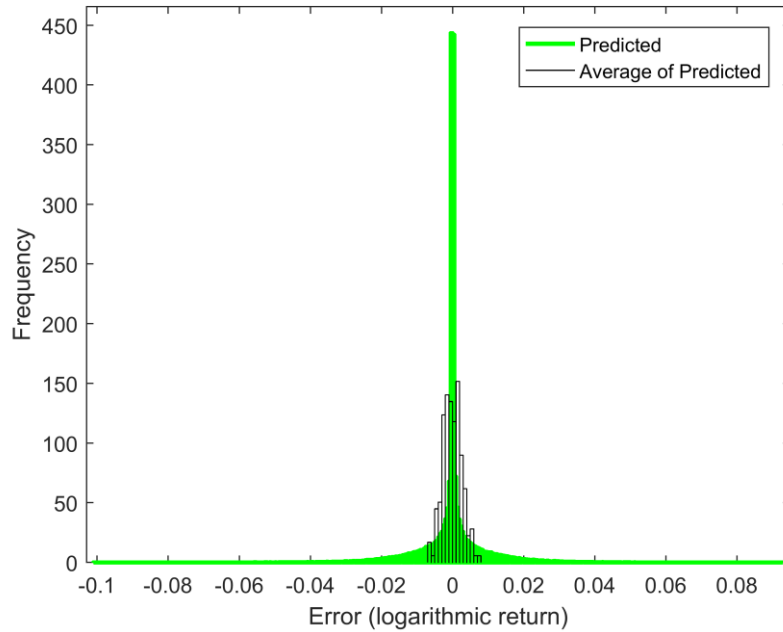
Figure 5. Distribution of errors (logarithmic return) of predicted and the average of predicted. Predicted correspond to the errors of 1,000 independent training processes of the scenario corresponding to 30 neurons. Source: authors' calculations.

There are other validation tools besides attaining a high-performance model in terms of low out-of-sample prediction error and fit. As in Hagan et al. (2014), there are two basic properties of accurate prediction models that should be regarded. First, prediction errors should be uncorrelated with each other. Second, prediction errors should be uncorrelated with the exogenous inputs sequence. In both cases, if errors were correlated the prediction could be further improved (Hagan et al., 2014).

Table 4 presents the incidence of significant correlation of prediction errors. It is calculated as the fraction of cases in which the 95 percent confidence intervals are exceeded in any of the twenty lags considered, with respect to 1,000 independent training processes of each scenario. The incidence of both types of correlation of prediction errors is rather high in all scenarios.

18

	Number of neurons			
	10	20	30	40
Autocorrelation	.8450	.7860	.7450	.7190
Correlation with exogenous inputs	.9680	.9190	.9230	.9320

Table 4. Incidence of significant correlation of prediction errors. Calculated as the fraction of cases in which the 95 percent confidence intervals are exceeded in any of the twenty lags considered, with respect to 1,000 independent training processes of each scenario. Source: authors' calculations.

The standard solution is to change the lag order of both inputs. This has no substantial effects on the incidence of exceedances with respect to the 95 percent confidence intervals, and adds computation time –when lags are increased. However, taking into account the performance (see tables 2 and 3), correcting for these two correlations in search of higher performance may be secondary.

An additional validation comes in the form of verifying whether the NARX-ANN model outperforms a non-linear autoregressive artificial neural network model (NAR-ANN).[15] This validation enables to assess the contribution of electronic payments data to reducing economic activity nowcast error. Table 5 compares the best performing NARX-ANN and the NAR-ANN models by their $RMSE$ performance.[16] The NARX-ANN attains a decrease in $\overline{RMSE}$ and $\overline{\overline{RMSE}}$ of 24.29 and 51.85 percent, respectively.[17] Thus, we validate that electronic payments data significantly reduces the nowcast error of a benchmark model.

---

[15] The implementation of the NAR-ANN model replicates the NARX-ANN in all aspects but excludes the exogenous inputs (i.e. electronic payments data). That is, in [2] and in Figure 3 all x variables are omitted. Not only a benchmark based on ANN allows a fair comparison for the NARX-ANN model, but it is a judicious choice as well because it encompasses linear and non-linear models.

[16] Table 8 (in Appendix) exhibits the performance of the NAR-ANN model for the four number-of-neuron scenarios. The best performing NAR-ANN by $\overline{RMSE}$ and $\overline{\overline{RMSE}}$ are the 10-neuron and 40-neuron models, respectively. The best performing NARX-ANN is the 30-neuron model (see tables 2 and 3).

[17] Correspondingly, the mean and range of out-of-sample errors is manifestly lower in the NARX-ANN model: The mean and range of errors in the NAR-ANN's committee are -1.3247×10^{-4} and [-.0298, .0244], respectively, whereas in the NARX-ANN's committee they are -5.4450×10^{-5} and [-.0069, .0070]. Furthermore, comparing tables 2, 3 and 8, the worst performing NARX-ANN (the 10-neuron model) outperforms the best NAR-ANN model.

	NAR [a]	NARX [b]	Change (%) [c]
$\overline{RMSE}$	.0107	.0081	-24.29
$\overline{\overline{RMSE}}$	.0054	.0026	-51.85

Table 5. Comparison of predictive out-of-sample performance for the best performing NAR-ANN and NARX-ANN models. $\overline{RMSE}$ corresponds to the average $RMSE$ estimated on 1,000 independent training processes for each scenario. $\overline{\overline{RMSE}}$ corresponds to the $RMSE$ calculated with the average predicted log-return on the 1,000 independent training processes of each scenario. [a] The best performing NAR-ANN by $\overline{RMSE}$ and $\overline{\overline{RMSE}}$ are the 10-neuron and 40-neuron models, respectively (see Table 8, in Appendix). [b] The best performing NARX-ANN is the 30-neuron model (see tables 2 and 3). Source: authors' calculations.

5. Final remarks

In this article, we address the economic activity nowcasting issue. We attempt to nowcast changes in the Economic Monitoring Index (*Indicador de Seguimiento a la Economía*, ISE), which is a two-month-lagged monthly indicator of Colombian economic activity. Unlike most literature on the subject, yet concurrent with the work of Galbraith and Tkacz (2017), we rely on electronic payment instruments data as high-frequency indicators from which reliable signals of economic activity are to be extracted. Regarding the methodological approach, as nowcasting is better understood as *predictive modeling* (instead of *explanatory modeling*), we rely on the well-documented performance and flexibility of artificial neural networks (ANNs) for prediction.

Results suggest that electronic payment instruments data and a NARX (Non-linear AutoRegressive eXogenous) ANN enable us to nowcast ISE with fair accuracy, about two months before its release. Besides, a committee of ANNs attained a further enhanced predictive performance. Some issues related to the correlation of prediction errors (e.g.

autocorrelation) are present and could not be circumvented by standard means (e.g. varying the lag order of inputs); nonetheless, it is debatable whether correcting for correlation of prediction errors to enhance performance is a primary concern. Also, we validate that the NARX-ANN model outperforms a benchmark non-linear autoregressive artificial neural network model (NAR-ANN); this confirms the contribution of electronic payments data to reducing economic activity nowcast error. Therefore, we conclude that it possible and convenient to nowcast changes in the Colombian short-term economic activity indicator (i.e. ISE) with a set of electronic payment instruments data.

Some extensions for our work are readily available. First, we focus on nowcasting, but electronic payment instruments data and the ANN model may be useful for forecasting as well; that is, instead of making a reliable current-period estimate of economic activity, we could attempt to predict one or several periods ahead. Second, as suggested in Evans (2005), it may be possible to attain a high-frequency (e.g. daily, weekly) estimate of economic activity as payments data becomes available with a lower lag (e.g. one-day lag). Third, the nowcasting outcome could feed other models that may profit from a current-period estimate of economic activity. Fourth, evaluating whether adding other traditional macroeconomic and financial variables compatible with the nowcasting objective (e.g. interest rates, stock indexes) may further enhance the predictive performance. Fifth, debit and credit card data should be added as inputs. In our case they are not considered because the corresponding data is reported with a lag that is incompatible with a nowcasting exercise. It is arguable that omitting credit and debit card data is nowadays unimportant because credit and debit card payments contribute with about 4 percent of all electronic payments by value of operations in Colombia (as of 2016), and because the attained predictive performance is fair. Yet, the dynamics of payment habits and technology may require including them in the near future. Finally, a reliable high-frequency input of cash usage is worth adding.

6. References

Aiken, M. (2000). Forecasting the United States gross domestic product with a neural network. *Journal of International Information Management, 9 (1)*, 67-75.

Angelini, E., di Tollo, G., & Roli, A. (2008). A neural network approach for credit risk evaluation. *The Quarterly Review of Economics and Finance, 48*, 733-755. [doi:10.1016/j.qref.2007.04.001]

Bańbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Nowcasting and the real-time data flow. In G. Elliott & A. Timmermann (Eds.), *Handbook of Economic Forecasting* (pp.195-237). Amsterdam: North Holland-Elsevier.

Banco de la República (2017). *Reporte de Sistemas de Pago*, Banco de la República.

Bec, F. & Mogliani, M. (2015). Nowcasting French GDP in real-time with surveys and "blocked" regressions: combining forecasts or pooling information? *International Journal of Forecasting, 31*, 1021-1042. [doi:10.1016/j.ijforecast.2014.11.006]

Bell, V., Wah Co, L., Stone, S., & Wallis, G. (2014). Nowcasting UK GDP growth. *Quarter Bulletin, Q1*, Bank of England.

Bishop, C.M. (1995). *Neural Networks for Pattern Recognition*. Clarendon Press: Oxford.

Bragoli, D. (2017). Now-casting the Japanese economy. *International Journal of Forecasting, 33*, 390-402. [doi:10.1016/j.ijforecast.2016.11.004]

Chakraborty, C. & Joseph, A. (2017). Machine learning at central banks. *Staff Working Paper, 674*, Bank of England.

Chaudhuri, T.D. & Ghosh, I. (2016). Artificial neural network and time series modeling based approach to forecasting the exchange rate in a multivariate framework. *Journal of Insurance and Financial Management, 1 (5)*, 92-123.

Choi, H. & Varian, H. (2012). Predicting the present with Google Trends. *Economic Record, 88*, 2-9. [doi:10.1111/j.1475-4932.2012.00809.x]

Departamento Administrativo Nacional de Estadística – DANE (2016). *Metodología General Indicador de Seguimiento a la Economía ISE*. Retrieved from: https://www.dane.gov.co/files/investigaciones/fichas/cuentas-nacionales/DSO-CT-MET-Metodologia-General-Indicador-de-Seguimiento-a-la-Economia-ISE-Agosto-2016.pdf.

Departamento Administrativo Nacional de Estadística – DANE (2017). *Metodología General Cuentas Trimestrales – CT – Base 2005*. Retrieved from: http://www.dane.gov.co/files/investigaciones/fichas/Cuentas_trimestrales_b2000_ma rzo_17.pdf.

Demyanyk, Y. & Hasan, I. (2009). Financial crises and bank failures: A review of prediction methods. *Federal Reserve Bank of Cleveland Working Papers Series*, *09-04R*, September.

Di Piazza, A., Di Piazza, M.C., & Vitale, G. (2016). Solar and wind forecasting by NARX neural networks. *Renewable Energy and Environmental Sustainability, 1 (39)*. [doi: 10.1051/rees/2016047]

Evans, M.D.D. (2005). Where are we now? Real-time estimates of the macroeconomy. *International Journal of Central Banking, 1(2),* 127-175.

Galbraith, J.W. & Tkacz, G. (2017). Nowcasting with payments system data. *International Journal of Forecasting*. [doi:10.1016/j.ijforecast.2016.10.002]

Giannonne, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics, 55*, 665-676. [doi:10.1016/j.jmoneco.2008.05.010]

Hagan, M.T., Demuth, H.B., Beale, M.H, & De Jesús, O. (2014). *Neural Network Design.* Martin Hagan: Oklahoma.

Han, J. & Kamber, M. (2006). *Data Mining: Concepts and Techniques*. Morgan Kaufman Publishers: San Francisco.

Hastie, T., Tibshirani, R., & Friedman, J. (2013). *The Elements of Statistical Learning*. Springer: New York.

Hinds, S., Rimmington, L., Dance, H., Gillham, J., Sentence, A. & Hawksworth, J. (2017). A machine learning approach to estimating current GDP growth. *UK Economic Outlook*, July, PWC.

Jalil, M. & Misas, M. (2006). Evaluación de pronósticos del tipo de cambio utilizando redes neuronales y funciones de pérdida asimétricas. *Borradores de Economía, 376*, Banco de la República.

Kohzadi, N., Boyd, M.S., Kaastra, I., Kermanshahi, B.S., & Scuse, D. (1995). Neural networks for forecasting: An introduction. *Canadian Journal of Agricultural Economics, 3 (43)*, 463-474. [doi:10.1111/j.1744-7976.1995.tb00135.x]

Lin, T., Horne, B.G., Tino, P., & Giles, C.L. (1996). Learning long-term dependencies in NARX recurrent neural networks. *IEEE Transactions on Neural Networks, 7 (6)*, 1329-1338.

Mahmoud, S., Lofti, A., & Langensiepen, C. (2013). Behavioral pattern identification and prediction in intelligent environments. *Applied Soft Computing, 13*, 1813-1822. [doi:10.1016/j.asoc.2012.12.012]

McAndrews, J. (1994). The automated clearinghouse system: Moving toward electronic payment. *Business Review, July/August*, Federal Reserve Bank of Philadelphia, 15-23.

McLaren, N. (2011). Using internet search data as economic indicators. *Quarterly Bulletin, Q2*, Bank of England, 134-140.

McNelis, P.D. (2005). *Neural Networks in Finance*. Elsevier: Burlington.

Misas, M., López, E., Arango, C., & Hernández, J.N. (2003). La demanda de efectivo en Colombia: Una caja negra a la luz de las redes neuronales. *Borradores de Economía, 268*, Banco de la República.

Mullainathan, S. & Spiess, J. (2017). Machine learning: An applied econometric approach. *Journal of Economic Perspectives, 31 (2)*, 87-106. [doi: 10.1257/jep.31.2.87]

Mitchell, T.M. (1997). *Machine Learning*. McGraw-Hill: Boston.

Sarlin, P. (2014). On biologically inspired predictions of the global financial crisis. *Neural Computing and Applications*, *24 (3)*, 663-673. [doi:10.1007/s00521-012-1281-y]

Shmueli, G. (2010). To explain or to predict?. *Statistical Science, 25(3)*, 289-310. [doi:10.1214/10-STS330]

Tiffin, A. (2016). Seeing in the dark: A machine-learning approach to nowcasting in Lebanon. *IMF Working Paper, WP/16/56*, International Monetary Fund.

Tkacz, G. & Hu, S. (1999). Forecasting GDP growth using artificial neural networks. *Bank of Canada Working Paper Series*, *99-3*, Bank of Canada.

Varian, H.R. (2014). Big data: New tricks for econometrics. *Journal of Economic Perspectives*, *28 (2)*, 3-28. [doi: 10.1257/jep.28.2.3]

Witten, I.H., Frank, E., & Hall, M.A. (2011). *Data Mining: Practical Machine Learning Tools and Techniques*. Morgan Kaufman Publishers: Burlington.

Zhang, G., Hu, M.Y, Patuwo, B.E., & Indro, D.C. (1999). Artificial neural networks in bankruptcy prediction: General framework and cross-validation analysis. *European Journal of Operational Research*, *116*, 16-32.

Zukime, M. & Junoh, M. (2004). Predicting GDP growth in Malaysia using knowledge-based economy indicators: A comparison between neural network and econometric approaches. *Sunway College Journal*, *1*, 39-50.

7. Appendix

	Series	Mean	Standard deviation	Skewness	Kurtosis	Jarque Bera (p-value)
	ISE	0.004	0.051	-1.747	5.727	.000
Value of Operations	ACH Colombia	0.017	0.160	-0.854	5.749	.000
	ACH Cenit	0.009	0.290	-1.412	9.098	.000
	Cedec	-0.005	0.140	-0.569	5.862	.000
Number of Operations	ACH Colombia	0.021	0.111	-0.872	5.824	.000
	ACH Cenit	0.026	0.236	-1.241	8.336	.000
	Cedec	-0.008	0.100	-0.383	3.019	.075

Table 6. Descriptive statistics of output and inputs. Estimated on logarithmic returns of monthly series from January 2001 to December 2016 (191 observations). Inputs corresponding to values are at constant prices, deflated by the consumer price index. Source: authors' calculations.

	Number of operations			Value of operations		
Year	Debit and credit cards[a]	Transfers via ACHs	Cheques	Debit and credit cards[a]	Transfers via ACHs	Cheques
2001	0,33	0,04	0,63	0,01	0,15	0,83
2002	0,37	0,06	0,57	0,01	0,18	0,80
2003	0,41	0,09	0,49	0,02	0,22	0,76
2004	0,47	0,11	0,42	0,02	0,27	0,71
2005	0,52	0,13	0,36	0,02	0,32	0,66
2006	0,53	0,15	0,32	0,02	0,34	0,63
2007	0,51	0,22	0,26	0,02	0,41	0,57
2008	0,49	0,32	0,20	0,02	0,46	0,52
2009	0,50	0,35	0,15	0,02	0,50	0,48
2010	0,52	0,36	0,12	0,02	0,55	0,43
2011	0,54	0,36	0,10	0,03	0,58	0,39
2012	0,56	0,35	0,08	0,03	0,63	0,34
2013	0,58	0,35	0,07	0,03	0,68	0,29
2014	0,59	0,35	0,06	0,03	0,70	0,26
2015	0,60	0,35	0,05	0,04	0,73	0,24
2016	0,62	0,34	0,04	0,04	0,74	0,22

Table 7. Electronic payment instruments (2001-2016). Contribution to the number and value of operations. Payments cleared and settled in the books of banks are not considered. [a] Purchases with debit and credit cards. Source: Banco de la República (2017).
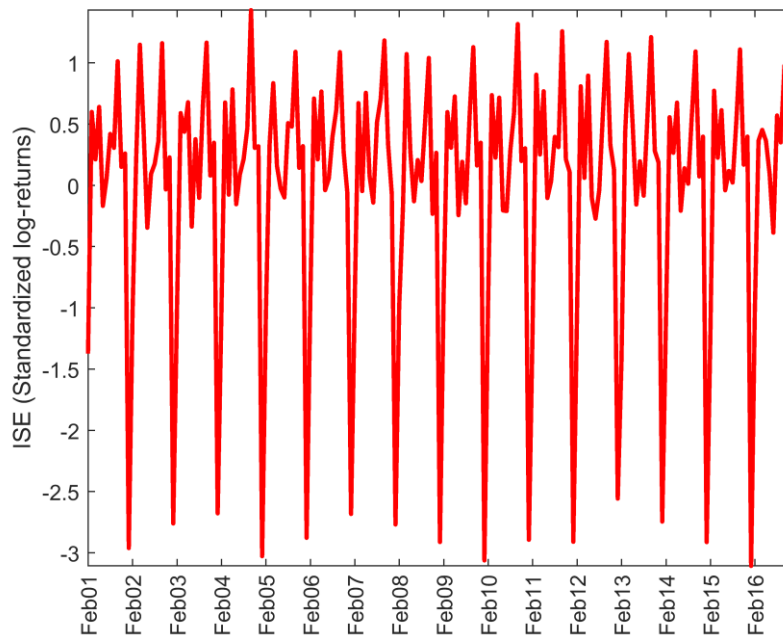
Figure 6. Standardized logarithmic returns of ISE. Non-seasonally adjusted, at constant prices, revised index. Source: DANE.
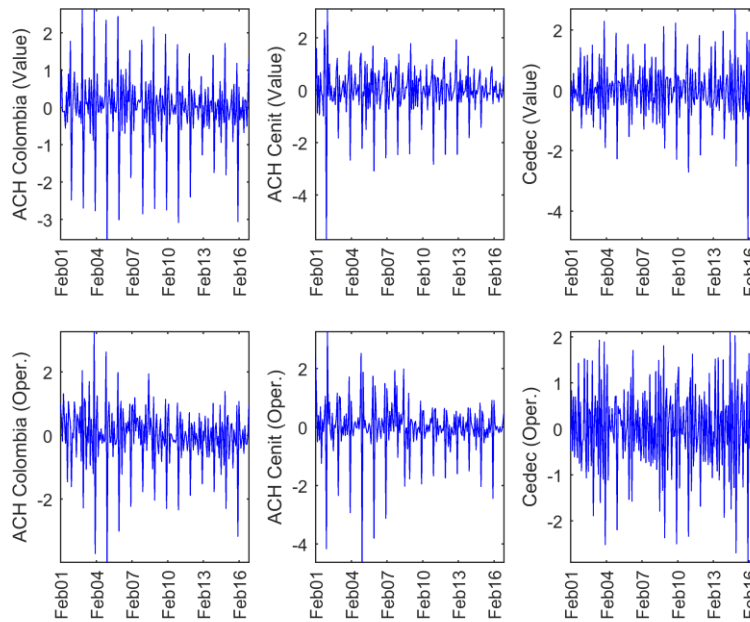

Figure 7. Standardized logarithmic returns of electronic payment instruments data. First row corresponds to logarithmic returns on the value (in constant Colombian pesos, deflated by the consumer price index) and the second to logarithmic returns on the number of operations. Source: ACH Colombia and Banco de la República.

	Number of neurons			
	10	20	30	40
$\overline{RMSE}$	**.0107** **[.0017]**	.0114 [.0025]	.0121 [.0030]	.0126 [.0032]
$\bar{r}$	**.9784** **[.0068]**	.9760 [.0106]	.9732 [.0135]	.9706 [.0193]
$\overline{\overline{RMSE}}$	.0075	.0065	.0059	**.0054**
$\bar{\bar{r}}$	.9895	.9921	.9935	**.9946**

Table 8. NAR-ANN predictive out-of-sample performance for selected scenarios of number of neurons in the hidden layer. $\overline{RMSE}$ corresponds to the average *RMSE* estimated on 1,000 independent training processes for each scenario. $\bar{r}$ corresponds to the average correlation coefficient between the observed and predicted log-return series for each scenario. Standard deviation is reported in brackets. $\overline{\overline{RMSE}}$ corresponds to the *RMSE* calculated with the average predicted log-return on the 1,000 independent training processes of each scenario. $\bar{\bar{r}}$ corresponds to the correlation coefficient between the observed and the average predicted log-return on the 1,000 independent training processes of each scenario. The highest performance scenario is in bold. Source: authors' calculations.