Forecasting the Colombian Unemployment Rate Using Labour Force Flows

By: Francisco Lasso-Valderrama
Héctor M. Zárate-Solano

Borradores de ECONOMÍA

No. 1073
2019

Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia - Bogotá - Colombia

Forecasting the Colombian Unemployment Rate Using Labour Force Flows[1]

Francisco Lasso-Valderrama☆ Héctor M. Zárate-Solano☆☆

Abstract

Accurate predictions of future magnitudes of the unemployment rate are crucial for monetary policy. This paper investigates whether the use of disaggregated household survey data improves the forecasts of the Colombian 13 cities unemployment rate. We conduct an out-of-sample forecast exercise to compare the performance of a model that incorporates flows of workers across different states of the labour market to that of various macroeconomic non-structural models. The paper follows the approach proposed by Barnichon & Nekarda (2013). Our results indicate that the two-state-flow model provides substantially better forecasts of the unemployment rate over longer horizons (more than five months ahead). Additionally, when forecasts are combined, significant gains in every forecasting horizon occurs. This combined forecast shows a 23% reduction in overall *RMSE*.

Keywords: Forecasting, unemployment, VAR models, labour market flows

JEL Codes: C53, E24, E27, E3, J64

[1]We thank Luis E. Arango-Thomas for the constructive comments and recommendations which help to improve readability and quality of the paper. We also thank the comments by an anonymous referee and participants at the CEMLA's XXIII Annual Meeting of the Central Bank Researchers Network conference. We also thank Juan Manuel Julio-Roman and the participants at the Seminario del Banco de la República. The views expressed herein are those of the authors and not necessarily those of the Banco de la República, nor its Board of Directors.

☆Junior Researcher at Banco de la República de Colombia (Central Bank of Colombia), email: flassova@banrep.gov.co

☆☆Senior Econometrician at Banco de la República de Colombia (Central Bank of Colombia), email: hzaratso@banrep.gov.co

Pronóstico de la tasa de desempleo en Colombia usando flujos de trabajadores de la fuerza laboral

Francisco Lasso-Valderrama   Héctor M. Zárate-Solano

Resumen

En este documento se evalúan los pronósticos de la tasa de desempleo urbana en Colombia utilizando varias metodologías. La primera se basa en las propiedades estadísticas de la serie de tiempo de la tasa de desempleo. La segunda considera la relación entre el crecimiento del producto y los cambios en el desempleo, conocida como la Ley de Okun. Finalmente, con base en los microdatos de las encuestas de hogares se calculan los flujos de trabajadores del mercado laboral para pronosticar la tasa de desempleo de acuerdo con Barnichon y Nekarda (2013).

La evaluación de los pronósticos fuera de muestra indica que el modelo de dos estados (ocupado-desocupado) es el mejor en horizontes superiores a cinco meses. Por su parte, los modelos ARIMA y la Ley de Okun compiten en precisión en horizontes de corto plazo. Cabe destacar que la combinación de los modelos de pronóstico genera ganancias significativas en todos los horizontes, alcanzando una reducción global de 23% en la raíz del error cuadrático medio.

Palabras clave: Pronósticos, desempleo, modelos VAR, flujos del mercado laboral

Clasificación JEL: C53, E24, E27, E3, J64

1. Introduction

The unemployment rate is a leading macroeconomic indicator of the business cycle that provides information about the condition of the labour market and is a variable of fundamental importance to policy makers. For instance, Kolly (2014) discussed that for monetary policy the unemployment rate serves as an indicator of the stance of the macroeconomy in general and gives information about the inflationary pressure. On the other hand, for fiscal policy makers, this rate is linked to government expenditure and income. Consequently, there is an increasing interest in being able to generate good forecast of the unemployment rate. An empirical fact found in various countries suggests that the unemployment rate reveals an asymmetric pattern, exhibiting a quick upsurge when rising and a gradual decrease when diminishing. This paper develops an empirical exercise of unemployment forecasting to investigate whether the Colombian unemployment rate forecasts can be improved by using the labor force flows which are computed with individual household survey data. Specifically, we compare the out-of sample forecasts with statistical time-series-based models and non structural economic methods.

Traditionally, forecasting methodologies are based on linear and non-linear models, taking solely into account the unemployment rate and their time series properties. Thus, various specifications implemented in different countries have been taken from the *ARIMA*, structural and *LSTAR* models. Examples are Montgomery & Zarnowitz (1998), Skalin & Terasvirta (2002) and Floros (2005). It's not an easy task forecasting the unemployment rate, especially during economic crises due to the complexity of the generating process. Historically, this has been carried out by using two strategies. The first involves statistical models based on identifying, estimating, and verifying the stochastic process selected represents the unemployment rate time series. The second bares in mind the relationship between GDP growth and unemployment variations by means of economic models such as Okun's Law. We follow the discussions by Friedman & Wachter (1974) and Okun (1962) on this topic.

A new focus is based on the flows of workers among different states of the labour market. According to this new methodology proposed by Barnichon & Nekarda (2013), the unemployment rate of a given period may be understood by using the analogy of the quantity (stock) of water in a tank. Therefore, if we have an initial level of water, the level of water at some point in the future is determined by the rate at which water comes in and goes out of the water tank. When the incoming water rate is the same as the outgoing rate, the level of water in the tank remains the same or constant. But if the rate of incoming water increases and the one going out doesn't, then the level of water in the tank will increase in the future. In other words, both incoming and outgoing rates provide information about the future level of water. In the case of the labour market, the incoming and outgoing flow of workers from unemployment is key in predicting the future unemployment rate level.

This aspect of convergence, as used in this paper, means that the current unemployment rate level comes close to a stationary rate arising from the gross flow of workers. Furthermore, this model allows for different types of time series properties, as seen from the flow of workers and their contributions to unemployment rate changes with the economic cycle, thereby allowing us to really capture the asymmetric nature of unemployment rate variations.

In choosing the most precise unemployment rate forecasting model, one assesses the model by using several statistics related to forecasting errors, based on the flow of workers vis-à-vis other traditional models. In addition, precision is measured when combining different model forecasts. As a reference for the empirical application in Colombia, the work of Barnichon & Nekarda (2013) was the basis for it. In effect, they introduced a forecasting model that has the information on the flow of workers. According to the these authors, this model far surpasses the forecasts of expert surveys, the historical outlooks by the Federal Reserve's Greenbook Admin Council, and the basic time-series-based models, with a forecast error below 30% approximately. In addition, the authors argued that the best forecasting ability of their model was at those periods that surrounded the inflection points of the economic cycle and great recessions. On the other hand, when combining the model's forecasts with those of expert surveys and of basic time-series-based- models, less errors were seen when compared to the expert surveys, i.e., around 35% for the current quarter, 15% for the next quarter forecast, and with slight improvements in longer time horizon outlooks.

Finally, the model has the advantage of also forecasting both the labour participation rate and the occupation rate which are compatible with the unemployment rate. Here the authors found that the model shows a modest forecasting improvement compared with that of the unemployment rate. In this connection, the model is only better than Greenbook's forecast for the current quarter. However, when combining both the model and Greenbook's forecasts, the forecast error diminishes substantially. In summary, the objectives of the model to be applied in Colombia, as proposed by Barnichon & Nekarda (2013), allows for incorporating the information of the labour force through the concept of having a "conditional unemployment rate to that of the stationary state," which in turn provides information on the short-term unemployment rate (see figure 1), and also allows for forecasting the flows of the labour force with time-series-based models, and to incorporate these forecasts in the law of movement that relates flows with the unemployment rate.

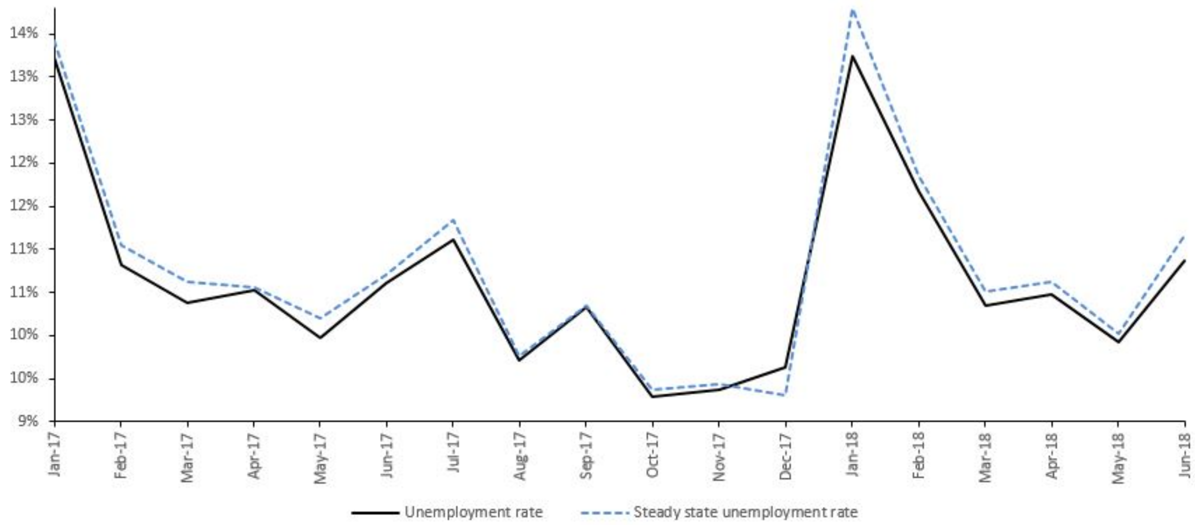
The results found in this article suggest that the two-state-flow model better approximates the long term dynamic structure of the unemployment rate at horizons greater than five months, whereas the *ARIMA* better represent its short term structure (one, two and four months ahead) and the *Okun* law provides superior forecasts at five-months-ahead. Furthermore, when compared with the best model, the forecast combination methodology reduces the *RMSE* from 10% in one-step-ahead to 42.6% in the nine-step-ahead.

This document contains five sections apart from this introduction. The second section describes the unemployment data and the flows among labour states that come from the household surveys. Then the forecasting methods used in the empirical study are described. The fourth contains an out of sample forecast evaluation with a set of standard measures at various time horizons and, finally, the main conclusions are presented, as well as future innovations.

2. The Colombian Unemployment Data

Each month, the official statistics bureau, DANE, collects data on the Colombian labour market with the Great Integrated Household Survey (*GEIH*) which regularly include topics of housing, household demographics, income, and employment, among others. Several indicators that measure the health of this market are summarized and published by DANE on a monthly basis. We used the time period with surveys covering the 13 major cities of Colombia that ranges from January 2008 to June 2018. The sample and period analyzed contain 6,145 thousand people at working age from which 3,358 thousand are employed, 449 thousands are unemployed and 2,338 thousands are not part of the labor force. In the surveys, individuals are categorized as employed, unemployed and not in the labour force. Those classified as unemployed are jobless persons available to work that are searching for a job. The sum of employed and unemployed forms the labour force. DANE has available these micro-data at *http://www.dane.gov.co*. Figure 1 shows that the unemployment rate presents a strong seasonality pattern and from 2017 to 2018 has fluctuated from a local maximum of 13.45% in January 2017 to a minimum of 9.49% in October 2017.

Figure 1: Unemployment rate and steady state from the three states model - 13 main cities



Source: Author's calculations based on Colombian household surveys, 2008.01-2018.06 and Appendix from Barnichon & Nekarda (2013)

We have to notice that these surveys also contain retrospective questions that allow us to compute the gross flows of workers among different states of the labour market: employed, unemployed and inactive. Thus, the labor force flows based on household information from the GEIH surveys are defined in Table 1.

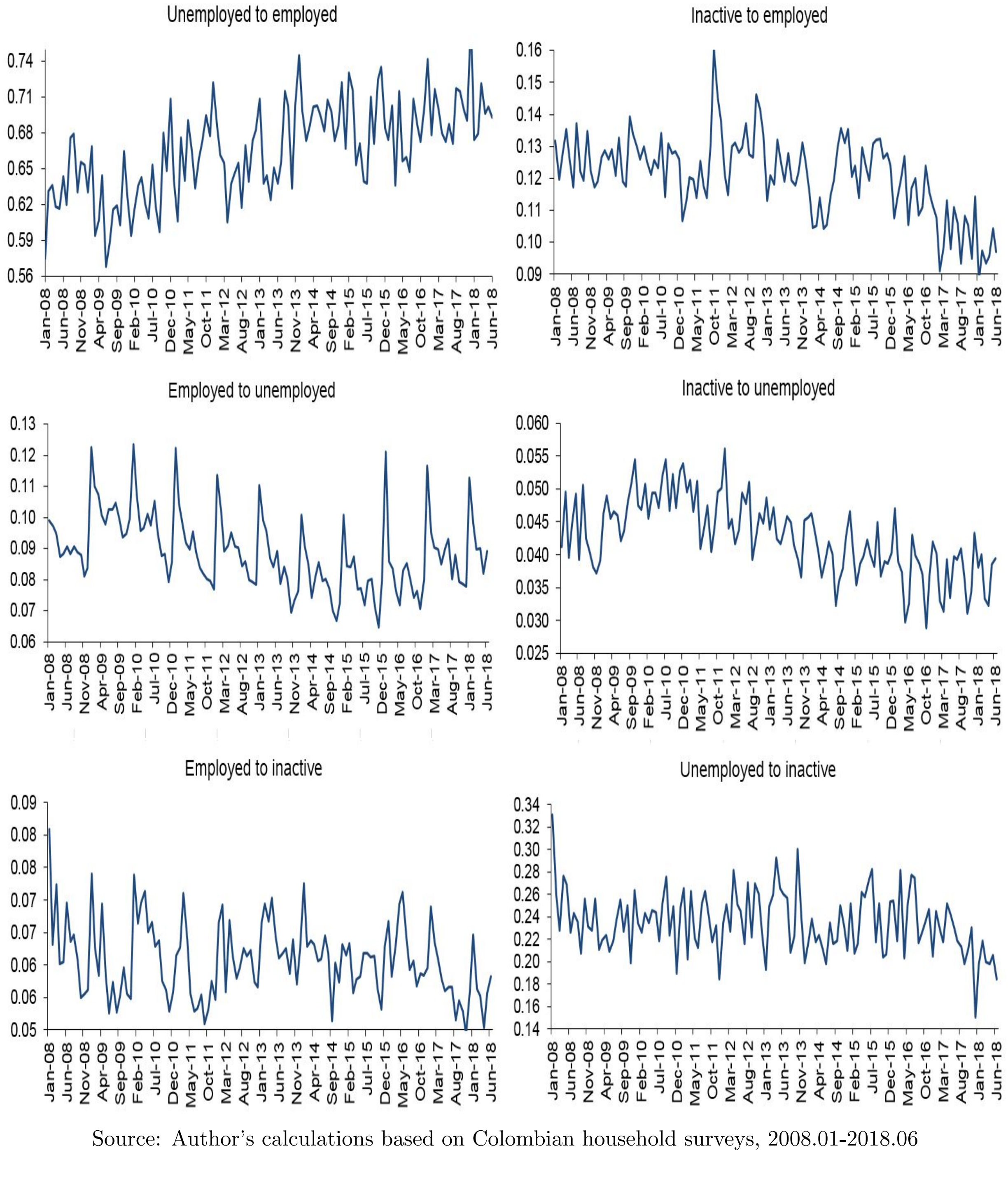
In Figure 2 are displayed the six flows among the states and Table A of the Appendix presents the unit root tests in order to characterize the persistence and possibly nonstationarity behaviour of the time series involved in the empirical comparison. It is value to notice that some forecasting methods evaluated in this paper take into account the flows of the proportion of the working-age population outside the labour force. which are known as inactive persons. Furthermore, this subgroup of persons comprises those recognized as discouraged jobseekers, defined as persons, who are available for work but stop looking for work due to various labour market reasons. For example, during periods of economic recession they have beliefs that there are no jobs available and assesses a low probability of being hired. A percentage of those discouraged jobseekers who temporarily retire to inactivity, as soon as they perceive some improvement in the odds of finding a job, start searching for work, which could cause an increase in the level of unemployment

Table 1: Annual transition matrix based on GEIH surveys

Previous state	Current state		
	Employed	Unemployed	Inactive
Employed	With more than one year in the company or with less than one year in the company and had another job in the year before the Survey.	They stopped working a year or less ago.	They stopped working a year or less ago.
Unemployed	One year or less working in the company and their previous work between one and two years before the Survey application.	They stopped working over a year ago and are looking for work since then. Applicants who have been looking for work for more than a year.	They have been working for one or two years and looked for work less than two years ago. They did not work before and looked for work less than two years ago.
Inactive	With one year or less in the company and they did not have previous work or they had their previous work more than two years ago with respect to the date of application of the Survey.	They stopped working over a year ago and are looking for work. Applicants who are looking for work a year or less ago.	The rest inactive today.
Total	Actual employed	Actual unemployed	Actual inactive

Source: GEIH surveys. Flows computed based on the methodology by Lasso (2012)

Figure 2: Gross flows from different states - 13 main cities



Source: Author's calculations based on Colombian household surveys, 2008.01-2018.06

We also have to notice that the sampling design of the *GEIH* surveys rely on independent repeated cross sectional sampling and do not constitute a longitudinal statistical operation. Therefore, new respondents are continually selected ensuring the reliability for each successive sample. On international applications in developed countries the survey design is based on rotating panel surveys, which allow the entry of new subjects which helps to capture dynamic changes in the population. For instance, the *CPS* for United States and the *LFS* for United Kingdom. Moreover, in other countries there is a matching among individuals by using key variables. It is known from statistical issues that these kinds of surveys face bias from attrition, coverage and design for stationary populations.

3. Empirical and Economic Models.

The Colombian unemployment rate and the flows among the states of the labour market described in the above section can be considered as realizations of unknown stochastic processes, which is named the data generation processes, DGP. In this section we briefly describe the methodologies generating forecasts of the unemployment rate and other indicators of the labour market.

3.1. Models based on labour force flows

This novel strategy of forecasting the market labour described by Barnichon & Nekarda (2013) is broadly based on two steps. First, the time series properties for the labor force flows are analyzed and through vector autoregressive models, VAR, their forecasts are computed for various time horizons. Second, economic theory incorporate these flows forecasts in order to form the unemployment rate's forecasts along with other indicators as in Shimer (2012). Conditional on the number of states considered, the models could be divided into a two-state-flow model or a three-state-flow model. For example, in a two-state-model, there are two situations in which an individual might find himself in, i.e., employed or unemployed, this model can be described in the following phases:

First, a law of motion for unemployment is obtained through a differential equation. According to Barnichon & Nekarda (2013): Assume $D_{t+\tau}$ the unemployment rate in $t+\tau$, where $t\epsilon\{0,1,2,\ldots\}$ periods and $\tau\epsilon[0,1]$ is a continuous measure of time within a month. In addition, finding a job f_{t+1} and separation from the employment rates s_{t+1} follows Poisson processes, the change of the unemployment rate is given by the difference between the inflows and outflows from the unemployment:

$$\frac{dD_{t+\tau}}{d\tau} = \dot{D}_{t+\tau} = s_{t+1}(1-D_{t+\tau}) - f_{t+1}D_{t+\tau} \tag{1}$$

Solving equation (1) yields:

$$D_{t+\tau} = \frac{(1 - e^{-f_{t+1} - s_{t+1}}) s_{t+1}}{s_{t+1} + f_{t+1}} + e^{-f_{t+1} - s_{t+1}} D_t \tag{2}$$

And the stationary unemployment rate D^*_{t+1} is:

$$D^*_{t+1} \equiv \frac{s_{t+1}}{s_{t+1} + f_{t+1}} \tag{3}$$

Replacing (3) in (2) and defining the convergence rate to the stationary state as

$$\beta_{t+1}(\tau) = 1 - e^{-f_{t+1} - s_{t+1}}$$

Thus:

$$D_{t+1} = \beta_{t+1}(\tau) D^*_{t+1} + [1 - \beta_{t+1}(\tau)] D_t \tag{4}$$

Equation (4) determines the variation of the unemployment stock D_{t+1} in a given period of time as function of the variation of the entry and exit flows to unemployment, f_{t+1} and s_{t+1}. Second, a forecast of the labor force flows that determines the unemployment rate in the stationary state and the velocity of convergence of the observed unemployment toward the stationary state.The forecast for the first time period of the unemployment rate is given by:

$$\hat{D}_{t+1|t} = \hat{\beta}_{t+1} D^*_{t+1} + [1 - \hat{\beta}_{t+1}] D_t \tag{5}$$

Where $\hat{\beta}_{t+1}$ is the forecast for the convergence velocity rate β_{t+1} between period t and t+1. Thus, according to equation (5), the unemployment rate could be predicted one period ahead given the actual entry and exit rates values. We follow Shimer (2012) and use the *GEIH* data of the number of unemployed (stocks of unemployed persons) and of those unemployed for less than one month (short-term unemployment) to compute both the job finding and job separation hazard rates. On the other hand, the unemployment rate forecasts for longer horizons require building forecast for both the entry and exit rates. Taking into account the lack of persistence of those rates, we used a VAR model to predict the entry and exit rates. We also include the actual unemployment rate (TD) and the vacancy rate (TV) obtained by Arango (2013). Thus, the specification is given by:

$$y_t = (lns_t, lnf_t, \Delta lnTD_t, \Delta lnTV_t) \tag{6}$$

Having the forecasts for the entry and exit rates, the unemployment rate forecasts for the j-period is formed by iterating the differential equation from the equation (5).

In a similar way, a three state market labour is formed by occupied (O), unemployed (D) and inactive (I), the law of movement are governed by the following system of equations:

$$\left\{ \begin{array}{c} \Delta O_t = D_{t-1}\lambda_t^{DO} + I_{t-1}\lambda_t^{IO} - O_{t-1}(\lambda_t^{OD} + \lambda_t^{OI}) \\ \Delta D_t = O_{t-1}\lambda_t^{OD} + I_{t-1}\lambda_t^{ID} - D_{t-1}(\lambda_t^{DO} + \lambda_t^{DI}) \\ \Delta I_t = O_{t-1}\lambda_t^{OI} + D_{t-1}\lambda_t^{DI} - I_{t-1}(\lambda_t^{IO} + \lambda_t^{ID}) \end{array} \right\} \tag{7}$$

Where λ_t^{XY} is the transition probability from the state $X \epsilon \{O, D, I\}$ to the state $Y \epsilon \{O, D, I\}$. This is calculated by the ratio between the number of individuals that transit between these two states during the year and the number of individuals in the beginning of the year. We have to notice that these probabilities follows the Poisson processes and are assumed constants during the year. For solving the system of equations (7), from the three state model, we follow the appendix of Barnichon & Nekarda (2013).

We rely on the VAR methodology in order to forecast the six transition probabilities according to:

$$y_t = c + \vartheta_1 y_{t-1} + \vartheta_2 y_{t-2} + \vartheta_3 y_{t-3} + \epsilon_t \tag{8}$$

Where:

$$y_t = (\lambda_t^{OD}, \lambda_t^{OI}, \lambda_t^{DO}, \lambda_t^{DI}, \lambda_t^{IO}, \lambda_t^{ID}, \Delta lnTD_t, \Delta lnTV_t) \tag{9}$$

Research on *VAR* forecasting is based on expanding the sample and iterating one step ahead according to the principle of multistep point forecasts, *IMS* approach. See Todd & McCracken (2013).

3.2. Okun Law

One of the most important conceptual and empirical relations in labour economics is that between production of real output and employment of labour proposed by Okun (1962). The empirical association states a fixed relation between movements in the unemployment rate and movements in the size of the output gap, which is the difference between actual output and the potential level. In other words, this rule of thumb is used to translate unemployment greater than the natural rate into an output gap and thus to measure the cost to society of excessive unemployment. The natural rate of unemployment when operating at full employment means that the economy is producing its potential GDP.

$$UR_t = f(IME_t, UR_{t-k}, Z_t)$$

Where: UR_t is the unemployment rate at the current period, IME_t is the index that monitor the economy, which is released on monthly basis by DANE. Z_t include other control variables such as dummy variables for some specific dates. Finally, f is assumed to be linear.

3.3. Time-series-based models for the unemployment rate

The leading approach to obtain forecasts is using either models based on stochastic seasonality or deterministic seasonality. The former class strategy relies on the traditional time series building methodology of Box & Jenkins (1970) which fit autoregressive integrated moving average models to the observed unemployment rate. These models are assumed to have seasonal unit roots and to induce stationarity, it is necessary to seasonally differentiate the series (Chatfield & Prothero, 1973). Under this strategy of forecasting a current value of the unemployment rate depends linearly on its own previous values plus a combination of current and previous values of a white noise error term. The *ARIMA* models are built by iterating the identification, estimation and diagnosis checking steps. The compact formulation for seasonal models are given in both polynomial and *ARIMA* forms by:

$$\vartheta_p(B)\phi_p(B^s\Delta^d\Delta_s^D y_t) = \Theta_Q(B^s)\theta_q(B)\varepsilon_t$$

or

$$ARIMA(p, d, q)(P, D, Q)_s$$

Where the lowercase indicate the non- seasonal order of the model and uppercases represent the seasonal order. The goal is to estimate the parameters and produce the forecasts guided by the parsimony principle.

3.4. Out-of-sample evaluation of forecasts accuracy

In this section we present the evaluation of forecasts produced by the competing models. The criteria for assessing the forecast accuracy is based on the smaller mean forecast error root, *RMSFE*. However we also compare with other statistics and hypothesis testing at different time horizons. Finally, we combine forecasts from the rival models following the Granger & Newbold (2014) claim: "is better to combine forecasts since the dominant forecast still contains information that can be utilized to obtain a superior forecast".

In order to assess the predictive ability of the rivals models, we used a rolling window scheme in which we set aside the observations for the last four years to conform the testing set to evaluate the forecasts. And fit the competing models in the training period. In other words, the statistics are calculated at different time horizons $h = 1, \ldots, 12$ in the testing set. The end of the estimation period T changes one period and the process is repeated recursively.

Table 2 presents the main results of the evaluation by using different standard statistics based on the forecasting error. The rows of Table 2 describe the models used for the sample covering the period from 2008-01 to 2018-06: dynamic combination proposed by Coulson & Robins (1993), combination by Granger & Ramahathan (1984), two-state-flow models, three state-flow models, *ARIMA* models and the *Okun* law specification. The columns include the horizon at which is evaluated each model, the number of observations for the evaluation period, the mean absolute error, the root mean squared error, the U-Theil statistic, the Giacomini and White p-value, the Diebold Mariano statistics and an independence test.

Looking at the forecast performance for each model through error related statistics, from Table 2, we conclude that for the sample period covering from 2008-01 to 2018-06, on the *RMSE* average (Horizon 0), the models that take into account the two-state-flow model and the differential equation to form the unemployment rate forecast tend to outperform the other models. For instance, this model is better for forecasting horizons bigger than five-months-ahead, according with the ordering based on the the *RMSE*. However, combining the forecasts from different models could reduce the *RMSE* in approximately 23% for twelve-step-ahead. Finally, Figure 3 presents the forecasts from the alternative models. Thus, according to the two-state-flow model, the predicted annual growth of the unemployment rate will increase steeply during the remaining months of 2018 and with the exception of February it will decrease throughout 2019, year in which the unemployment rate will be 9.6 percent in December .

Table 2: Forecast performance for alternative models

Forescating Model	HORIZON	NOBS	RMSE*	UTHEIL	GW	GW_PVAL	DM	DM_PVAL	S3A0_PVAL	S3A2_PVAL	FI
Coulson and Robins	0	61	0.62	0.42	8.21	0.01	-2.32	0.01	0.01	0.01	0.54
Granger and Ramanathan	0	61	0.64	0.43	8.14	0.01	-2.31	0.01	0.01	0.01	0.55
Two States	0	61	0.80	0.54	0.00	1.00	0.00	0.50	0.00	0.00	0.49
Okun Law	0	61	0.83	0.57	1.60	0.39	0.34	0.63	0.81	0.82	0.53
Three States	0	61	0.86	0.59	0.44	0.59	0.64	0.74	0.68	0.69	0.50
ARIMA Model	0	61	0.96	0.67	7.74	0.26	1.23	0.89	0.95	0.97	0.50
Coulson and Robins	1	66	0.63	0.46	5.83	0.02	-2.09	0.02	0.03	0.04	0.67
Granger and Ramanathan	1	66	0.63	0.46	5.77	0.02	-2.26	0.01	0.08	0.07	0.64
ARIMA Model	1	66	0.70	0.52	1.26	0.26	-0.49	0.31	0.15	0.17	0.50
Okun Law	1	66	0.74	0.55	0.53	0.47	-0.40	0.35	0.71	0.67	0.49
Two States	1	66	0.79	0.58	0.00	1.00	0.00	0.50	0.00	0.00	0.46
Three States	1	66	0.83	0.61	0.46	0.50	0.50	0.69	0.89	0.92	0.36
Coulson and Robins	2	66	0.66	0.40	5.33	0.02	-1.40	0.08	0.01	0.03	0.58
Granger and Ramanathan	2	66	0.67	0.40	5.20	0.02	-1.44	0.07	0.01	0.02	0.61
Two States	2	66	0.77	0.46	0.00	1.00	0.00	0.50	0.00	0.00	0.42
ARIMA Model	2	66	0.78	0.47	0.04	0.84	0.10	0.54	0.24	0.30	0.61
Okun Law	2	66	0.79	0.47	0.15	0.70	0.19	0.57	0.91	0.92	0.64
Three States	2	66	0.80	0.48	0.09	0.77	0.35	0.64	0.44	0.47	0.55
Coulson and Robins	3	65	0.57	0.32	10.34	0.00	-3.43	0.00	0.00	0.00	0.43
Granger and Ramanathan	3	65	0.64	0.36	9.92	0.00	-2.97	0.00	0.00	0.00	0.50
ARIMA Model	3	65	0.75	0.42	1.73	0.19	-0.53	0.30	0.18	0.28	0.52
Okun Law	3	65	0.75	0.42	2.30	0.13	-0.80	0.21	0.20	0.28	0.55
Two States	3	65	0.83	0.47	0.00	1.00	0.00	0.50	0.00	0.00	0.40
Three States	3	65	0.87	0.49	0.34	0.56	0.43	0.67	0.74	0.70	0.53
Coulson and Robins	4	64	0.59	0.34	8.32	0.00	-2.66	0.00	0.00	0.00	0.61
Granger and Ramanathan	4	64	0.60	0.35	7.70	0.01	-2.83	0.00	0.00	0.00	0.61
ARIMA Model	4	64	0.75	0.44	0.23	0.63	-0.27	0.39	0.43	0.55	0.59
Two States	4	64	0.80	0.46	0.00	1.00	0.00	0.50	0.00	0.00	0.52
Okun Law	4	64	0.81	0.47	0.02	0.90	0.10	0.54	0.28	0.31	0.55
Three States	4	64	0.83	0.48	0.28	0.60	0.31	0.62	0.43	0.34	0.64
Coulson and Robins	5	63	0.55	0.35	16.20	0.00	-2.76	0.00	0.00	0.00	0.64
Granger and Ramanathan	5	63	0.56	0.36	16.00	0.00	-2.95	0.00	0.00	0.00	0.67
Okun Law	5	63	0.73	0.47	0.18	0.67	-0.42	0.34	0.38	0.49	0.61
ARIMA Model	5	63	0.75	0.48	0.14	0.71	-0.15	0.44	0.47	0.58	0.66
Two States	5	63	0.78	0.50	0.00	1.00	0.00	0.50	0.00	0.00	0.52
Three States	5	63	0.83	0.53	0.30	0.58	0.52	0.70	0.21	0.34	0.58
Coulson and Robins	6	62	0.64	0.41	6.11	0.01	-2.49	0.01	0.00	0.00	0.52
Granger and Ramanathan	6	62	0.64	0.41	5.77	0.02	-2.71	0.00	0.00	0.00	0.55
Two States	6	62	0.82	0.53	0.00	1.00	0.00	0.50	0.00	0.00	0.49
Three States	6	62	0.86	0.56	0.06	0.81	0.40	0.66	0.69	0.73	0.55
Okun Law	6	62	0.88	0.57	0.33	0.56	0.56	0.71	0.82	0.83	0.52
ARIMA Model	6	62	0.88	0.57	0.50	0.48	0.37	0.65	0.64	0.74	0.62
Coulson and Robins	7	61	0.65	0.41	8.62	0.00	-1.99	0.02	0.02	0.02	0.43
Granger and Ramanathan	7	61	0.66	0.41	8.09	0.00	-2.05	0.02	0.01	0.01	0.46
Two States	7	61	0.80	0.50	0.00	1.00	0.00	0.50	0.00	0.00	0.61
Okun Law	7	61	0.89	0.55	1.01	0.32	0.83	0.80	0.87	0.89	0.42
Three States	7	61	0.90	0.56	0.46	0.50	1.02	0.85	0.86	0.89	0.52
ARIMA model	7	61	0.98	0.61	4.28	0.04	1.21	0.89	0.99	0.99	0.35

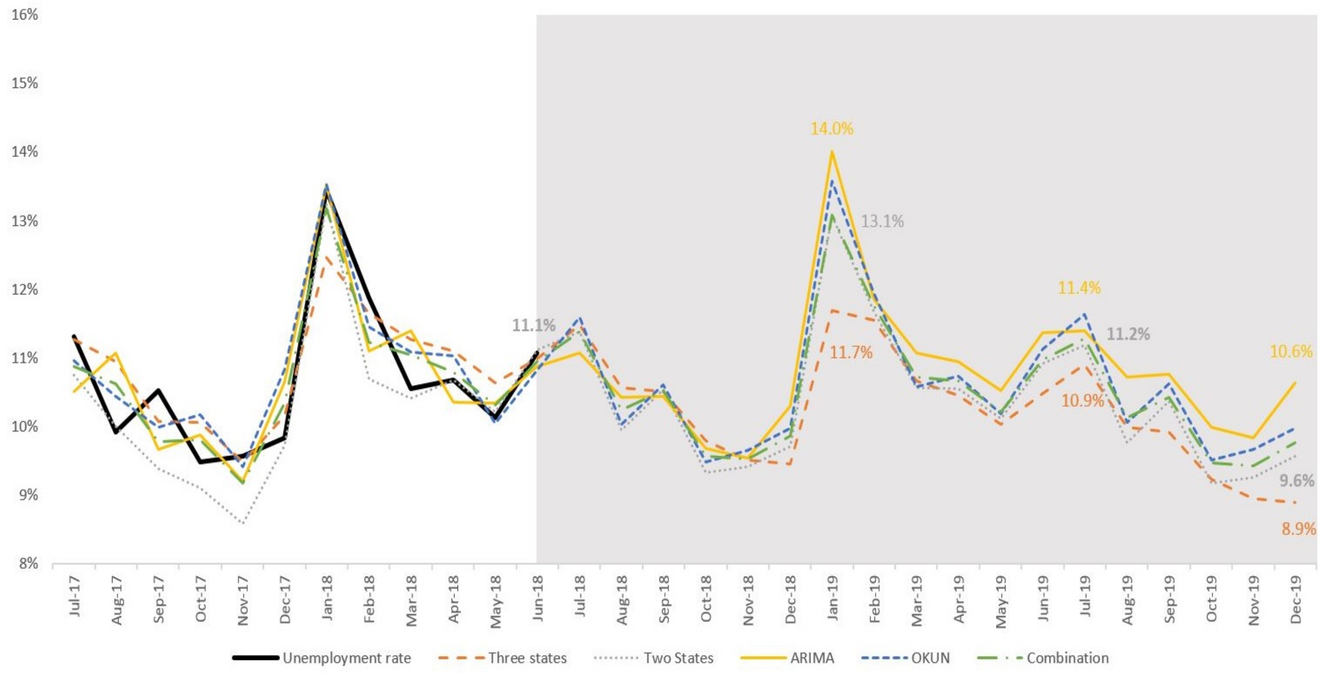
Continuation Table 2: Forecast performance for alternative models

Forescating Model	HORIZON	NOBS	RMSE*	UTHEIL	GW	GW_PVAL	DM	DM_PVAL	S3A0_PVAL	S3A2_PVAL	FI
Coulson and Robins	8	60	0.67	0.37	3.41	0.06	-1.45	0.07	0.05	0.05	0.52
Granger and Ramanathan	8	60	0.67	0.38	3.63	0.06	-1.55	0.06	0.07	0.07	0.52
Two States	8	60	0.78	0.44	0.00	1.00	0.00	0.50	0.00	0.00	0.45
Three States	8	60	0.82	0.46	0.09	0.76	0.44	0.67	0.73	0.78	0.49
Okun Law	8	60	0.89	0.50	3.60	0.06	1.06	0.86	0.97	0.97	0.39
ARIMA Model	8	60	1.07	0.60	11.85	0.00	1.99	0.98	1.00	1.00	0.42
Coulson and Robins	9	59	0.54	0.30	10.22	0.00	-3.55	0.00	0.00	0.01	0.58
Granger and Ramanathan	9	59	0.59	0.33	8.73	0.00	-3.33	0.00	0.01	0.01	0.58
Okun Law	9	59	0.77	0.43	0.55	0.46	-0.60	0.27	0.55	0.61	0.56
Two States	9	59	0.83	0.46	0.00	1.00	0.00	0.50	0.00	0.00	0.45
Three States	9	59	0.84	0.47	0.00	0.98	0.10	0.54	0.80	0.72	0.43
ARIMA Model	9	59	1.04	0.58	7.68	0.01	1.42	0.92	1.00	1.00	0.45
Coulson and Robins	10	58	0.64	0.37	12.39	0.00	-2.14	0.02	0.00	0.00	0.55
Granger and Ramanathan	10	58	0.65	0.38	16.29	0.00	-2.20	0.01	0.01	0.01	0.48
Two States	10	58	0.81	0.47	0.00	1.00	0.00	0.50	0.00	0.00	0.52
Okun Law	10	58	0.84	0.50	0.88	0.35	0.36	0.64	0.88	0.84	0.63
Three States	10	58	0.86	0.50	0.38	0.54	0.51	0.69	0.44	0.43	0.45
ARIMA Model	10	58	1.17	0.69	14.25	0.00	2.63	1.00	1.00	1.00	0.36
Coulson and Robins	11	57	0.67	0.46	5.71	0.02	-1.45	0.07	0.04	0.08	0.52
Granger and Ramanathan	11	57	0.69	0.48	6.00	0.01	-1.30	0.10	0.05	0.13	0.55
Two States	11	57	0.78	0.54	0.00	1.00	0.00	0.50	0.00	0.00	0.58
Three States	11	57	0.90	0.62	1.07	0.30	1.14	0.87	0.73	0.81	0.36
Okun Law	11	57	0.92	0.64	6.90	0.01	1.34	0.91	1.00	1.00	0.46
ARIMA Model	11	57	1.29	0.89	22.30	0.00	3.81	1.00	1.00	1.00	0.49
Coulson and Robins	12	56	0.64	0.82	6.08	0.01	-2.40	0.01	0.00	0.00	0.49
Granger and Ramanathan	12	56	0.68	0.88	4.58	0.03	-2.10	0.02	0.00	0.01	0.43
Two States	12	56	0.83	1.07	0.00	1.00	0.00	0.50	0.00	0.00	0.52
Okun Law	12	56	0.99	1.28	2.82	0.09	1.64	0.95	0.99	0.98	0.52
Three States	12	56	1.00	1.29	1.73	0.19	1.81	0.97	0.91	0.86	0.48
ARIMA Model	12	56	1.35	1.74	28.56	0.00	4.14	1.00	1.00	1.00	0.44

Model : specification used to forecast
Horizon: The forecasting horizon. the zero horizon is an average of the evaluation statistics of the different forecast horizons
Nobs : sample size used to assess the performance of the forecasts
RMSE : root mean squared forecast error to measure the forecasting accuracy
UTHEIL: Theil's U statistics for forecast accuracy compared with the random walk alternative
GW: Giacomini and White statistic for equal mean squared error, conditioned on the forecasts.
DM : Diebold Mariano statistic
*DM*_PVAL : P-value for the Diebold Mariano statistics
*S3AO*_PVAL: p-value for the rank statistic for DM: d=abs(observed-forecast)-abs(observed-reference forecast)
*S3A2*_PVAL : P-value for the Diebold Mariano for the statistic d=abs(forecast/obs-1)-abs(reference forecast/obs-1)
FI : independence test to assess the degree accuracy in forecasting monthly ups and downs.

Source: Author's calculations based on Colombian household surveys, 2008.01-2018.06

Figure 3: Forecast performance for alternative models



Source: Author's calculations based on Colombian household surveys, 2008.01-2018.06. The forecasts obtained from this document are the sole responsibility of the authors and do not commit Banco de la República nor its Board of Directors

4. Conclusions

Forecasting the unemployment rate is one of the most important applications for economists and policy makers. In this study we have focused on constructing a set of forecasts from time-series-based models and economic principles that consider the flows among the states of the individuals forming the labour market. That is, we considered *ARIMA*, *Okun* and models based on flows. The forecast comparison is based mainly on the smaller mean forecast error and other hypothesis tests usually found in the time series application involving linearity. The results indicate that the models that use the labor force flows in a two state representation improve other simple models. However, combining the forecasts is a better strategy for achieving superior unemployment forecasts. We have to notice that in this contest, the models that include labor force flows take an additional advantage of forecasting other labour indicators compatible with the unemployment rate such as the occupation and the participation rate. Future research should seek to expand more complex models by considering a class of nonlinear autoregressive models. For example, the smooth transition autoregressions (*STARs*) which are able to smoothly switch between regimes. The accuracy of different forecast methods is a topic of continuous research and other measures outside of the linearity assumption for evaluating the models could be implemented.

References

Aaronson, S., Fallick, B., Figura, A., Pingle, J., & Wascher, W. (2006). The recent decline in the labor force participation rate and its implications for potential labor supply. *Brookings Papers on Economic Activity*, *37*, 69–154.

Arango, L. E. (2013). Puestos de trabajo vacantes según anuncios de la prensa escrita de las siete principales ciudades de Colombia. *Borradores de Economía*, .

Baghestani, H. (2008). Federal reserve versus private information: Who is the best unemployment rate predictor. *Journal of Policy Modeling*, *30*, 101–110.

Barnichon, R. (2012). Vacancy posting, job separation and unemployment fluctuations. *Journal of Economic Dynamics and Control*, *36*, 315–330.

Barnichon, R., & Figura, A. (2010). *What drives movements in the unemployment rate? a decomposition of the Beveridge curve*. Finance and Economics Discussion Series 2010-48 Board of Governors of the Federal Reserve System (U.S.).

Barnichon, R., & Nekarda, C. (2013). *The ins and outs of forecasting unemployment: Using labor force flows to forecast the labor market*. Finance and Economics Discussion Series 2013-19 Board of Governors of the Federal Reserve System (U.S.).

Blanchard, O., & Diamond, P. (1990). The cyclical behovior of the gross flows of u.s. workers. *Brookings Papers on Economic Activity*, *21*, 85–156.

Box, G., & Jenkins, G. (1970). Time series analysis: Forecasting and control, .

Brown, L., & Moshiri, S. (2004). Unemployment variation over the business cycles: a comparison of forecasting models. *Journal of Forecasting*, *23*, 497–511.

Chatfield, C., & Prothero, D. L. (1973). Box-jenkins seasonal forecasting: Problems in a case-study. *Journal of the Royal Statistical Society. Series A (General)*, *136*, 295–336.

Coulson, E., & Robins, R. (1993). Forescast combination in a dynamic setting. *journal of forecasting*, .

Diebold, F., & Mariano, R. (1995). Comparing predictive accuracy. *Journal of Business Economic Statistics*, *13*, 253–63.

Elsby, W. L., Michael, Bart, H., & Ayşegül, S. (2011). Unemployment dynamics in the oecd. *Federal Reserve Bank of San Francisco.*, .

Fleischman, C., & Roberts, J. (2011). *From many series, one cycle: improved estimates of the business cycle from a multivariate unobserved components model*. Finance and Economics Discussion Series 2011-46 Board of Governors of the Federal Reserve System (U.S.).

Floros, C. (2005). Forecasting the u.k. unemployment rate: model comparison. *International Journal of Applied Econometrics and Quantitative Studies*, *2*.

Friedman, B., & Wachter, M. (1974). Unemployment: Okun's law, labor force, and productivity. *The Review of Economics and Statistics*, *56*, 167–176.

Fujita, S. (2011). *Dynamics of worker flows and vacancies: evidence from the sign restriction approach*. Working Papers 1.

Giacomini, R., & Rossi, B. (2010). Forecast comparisons in unstable environments. *Journal of Applied Econometrics*, *25*, 595–620.

Giacomini, R., & White, H. (2006). Tests of conditional predictive ability. *Econometrica*, *74*, 1545–1578.

Golan, A., & Perloff, J. (2004). Superior forecasts of the u.s. unemployment rate using a nonparametric method. *The Review of Economics and Statistics*, *86*, 433–438.

Granger, C., & Newbold, P. (2014). *Forecasting Economic Time Series*. (2nd ed.). Elsevier.

Granger, C., & Ramahathan (1984). Improved method of combining forecast. *Journal of Forecasting*, *3*, 197–204.

Kolly, J. (2014). Formal and heuristic model averaging methods for predicting the unemployment rate. chapter 2. (pp. 133–142). Springer Proceedings in Mathematics and Statistics. Bayesian Statistics from Methods to Models and Applications.

Lasso, F. J. (2012). La dinámica del desempleo urbano en colombia. chapter 3. (pp. 131–166). Banco de la República. En: Arango, L. and Hamann, F. ed., (2017). La dinámica del desempleo urbano en Colombia. In: El mercado laboral en Colombia: hechos, tendencias e instituciones.

Montgomery, A., & Zarnowitz, V. (1998). Forecasting the u.s. unemployment rate. *Journal of the American Statistical Association*, *93*, 478–493.

Okun, A. M. (1962). Potential gnp: Its measurement and significance, . (pp. 98–104). Proceedings of the Business and Economic Statistics Section of the American Statistical Association.

Rothman, P. (2008). Reconsideration of the markov chain evidence on unemployment rate asymmetry. *Studies in Nonlinear Dynamics and Econometrics*, *12*.

Shimer, R. (2012). Reassessing the ins and outs of unemployment. *Review of Economic Dynamics*, *14*, 127–148.

Skalin, J., & Terasvirta, T. (2002). Modeling asymmetries and moving equilibria in unemployment rates. *Macroeconomic Dynamics*, *6*, 202–241.

Solon, G., Michaels, R., & Elsby, M. (2009). The ins and outs of cyclical unemployment. *American Economic Journal: Macroeconomics*, *1*, 84–110.

Todd, & McCracken (2013). Evaluating the accuracy of forecasts from vector autoregressions, . (pp. 117–168). Var models in macroeconomics - new developments and applications : Essays in honor of Christopher Sims.

Appendices

A. Unit root tests

UNIT ROOT TESTS

		DO	Δ DO	IO	Δ IO	OD	ΔOD	ID	ΔID
ADF	Trend + Constant	-4.5285*	-3.1498	-3.6981	-9.8871*	-3.321	-7.658*	-2.0773	-6.4463*
	Constant	-1.3086	-3.2548	-3.0769	-9.9332*	-1.5659	-7.5767*	-1.2874	-6.3751*
	None	-1.4237	-2.9853	0.237	-9.9786*	0.3714	-7.6034*	0.5396	-6.379*
KPSS	Constant	1.7237	0.0682*	0.2651*	0.0366*	1.7785	0.179*	1.1618	0.0907*

		OI	ΔOI	DI	ΔDI	TD	ΔTD	GOOGLE	ΔGOOGLE
ADF	Trend + Constant	-4.3126*	-4.0222	-9.4236*	-8.2318*	-2.976	-2.0708	-1.0936	-3.7268
	Constant	-4.2732*	-4.0339*	-2.8174	-8.2411*	-1.6184	-2.8113	-1.6002	-3.5336*
	None	0.1102	-4.0772*	0.1948	-8.2851*	1.9439	-1.9904	-0.4227	-3.5345*
KPSS	Constant	0.1012*	0.0407*	0.0928*	0.1076*	1.9873	0.2106*	0.8759	0.0955*

(*) Denotates rejection of the null hypothesis at the 1% significance level. Under the null hypothesis Y(t) has unit root.

DO*: Logarithm of the flow from unemployed to employed
IO*: Logarithm of the flow from out of the labor force to employed
OD*: Logarithm of the flow from employed to unemployed
ID*: Logarithm of the flow from out of the labor force to unemployed
OI*: Logarithm of the flow from employed to out of the labor force
DI*: Logarithm of the flow from unemployed to out of the labor force
TD*: Logarithm of the unemployment rate
GOOGLE*: Logarithm of the job search rate

*Seasonally adjusted variable through seasonal filter X-12 available in E-views. The multiplicative X-11 method was selected.

Source: Author's calculations based on Colombian household surveys, 2008.01-2018.06


BANCO DE LA REPÚBLICA
LIBERTAD
COLOMBIA