Newspaper Analytics and the Dynamics of the Exchange Rate in Peru

Erick Lahura and Luciana Figueroa

XXXVI Encuentro de Economistas del BCRP

October, 2018

Contents

Introduction
Motivation

- Under a floating exchange rate regime, forecasting future changes in the exchange rate can improve financial decisions of both policy makers and private agents.
- Especially important in partially dollarized economies such as Peru: balance-sheet effects.
- Since Meese and Rogoff(1983), most of the literature suggests that the exchange rate is a random walk: its growth rate is unpredictable.
- Some recent attempts that suggest that exchange rates can be predictable to some extent (Clark and West, 2006; Wright, 2008; Amat et al., 2018).

Introduction
What do we do?

- **Question**: Can we improve the random walk model?
- **Hypothesis**: macroeconomic news data help to forecasting the dynamics of the exchange rate in Peru.
- **Data**: Online newspaper articles published by El Comercio and Gestión.
- **Empirical methodology**: daily news indexes,
  - Text mining and text analysis of newspaper articles.
  - Linear and non-linear time series models: ARMA and GARCH.
- **Results**: news indexes are helpful to forecast the dynamics of daily exchange rates.

Introduction
Text mining and text analytics

- **Text mining**: to clean up data. Python, R. Categorization, clustering, sentiment analysis.
- **Text analytics**: application of statistical and machine learning techniques to infer/predict. Python, R, Eviews, Stata, etc. Association, forecasting analysis, word clouds.
- We construct two types of indexes:
  - Occurrence and intensity of key macroeconomic words related to the foreign exchange (forex) market: “volatility”, “crisis”, etc.
  - Classification of words into “positive” and “negative”.

Contents

Data and News Indexes

Text mining and text analytics

- Online textual information from “El Comercio” and “Gestión”.
- Web scraping methods implemented in the statistical program *R*.
- Daily covering between January, 2014 and December, 2017.
- A total of 27,942 news articles. Only 864 news articles contained the words “exchange rate”.

Data and News Indexes

A standard word cloud - daily newspaper articles related to exchange rate news

Figure: Word cloud from daily newspaper articles, 2014

Data and News Indexes

A standard word cloud - daily newspaper articles related to exchange rate news

Figure: Word cloud from daily newspaper articles, 2015

Data and News Indexes

A standard word cloud - daily newspaper articles related to exchange rate news

Figure: Word cloud from daily newspaper articles, 2016

Data and News Indexes

A standard word cloud - daily newspaper articles related to exchange rate news

Figure: Word cloud from daily newspaper articles, 2017

Data and News Indexes

Daily exchange rate data

- The "exchange rate growth" at time "t" is:

$$\left[\left(\frac{Exchange\ \ Rate\ \ at\ \ 11:00_{(t)}}{Exchange\ \ Rate\ \ at\ \ 13:31_{(t-1)}}\right)-1\right]*100 \quad (1)$$

- News at time t include those published between 1:31 PM of day $t-1$ and 11:00 AM of the day t.

- News between 11:00 am and 1:31 PM are excluded because at those hours the central bank intervenes in the exchange rate market.

Data and News Indexes

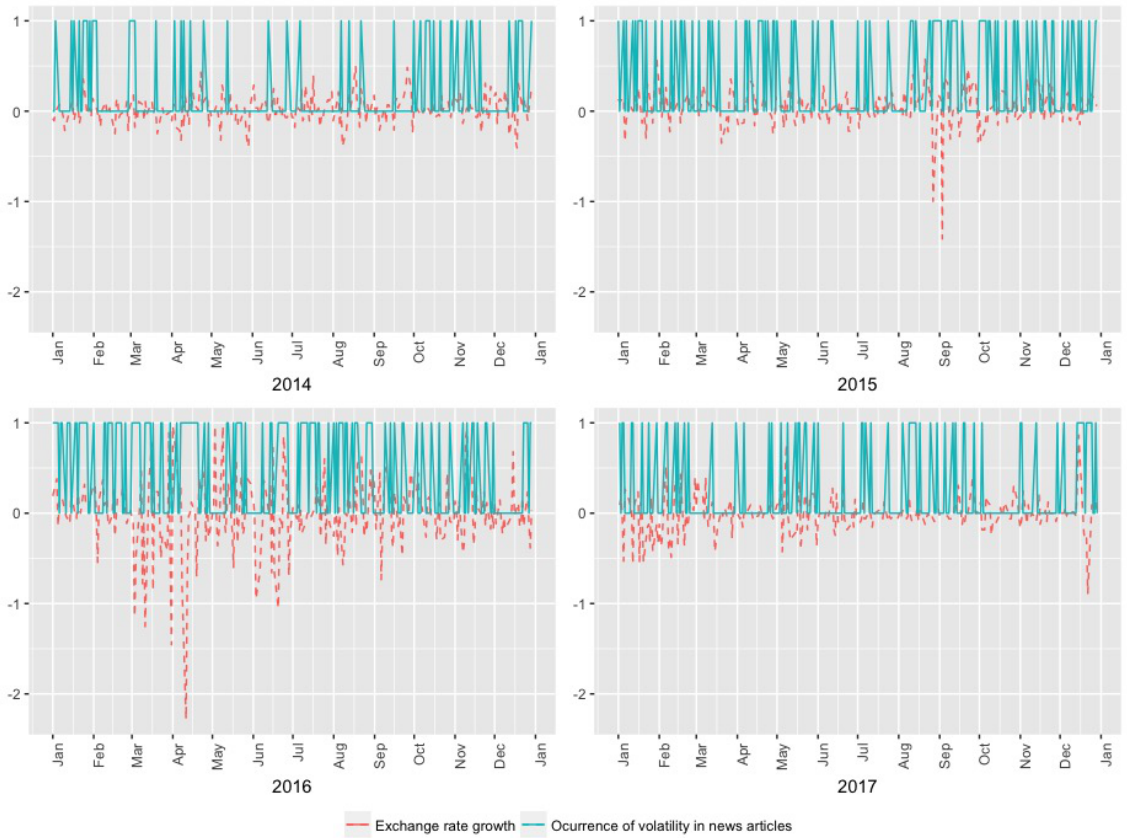
News indexes

- Online textual information from “El Comercio” and “Gestión”.
- Occurrence of different keywords, such as “crisis”, “volatility”, among others. Both dummy and continuous variables.
- Sentiment analysis: classification of the news articles into “positive” and “negative”:

$$\frac{\textit{Positive words used} - \textit{Negative words used}}{\textit{Positive words used} + \textit{Negative words used}}$$

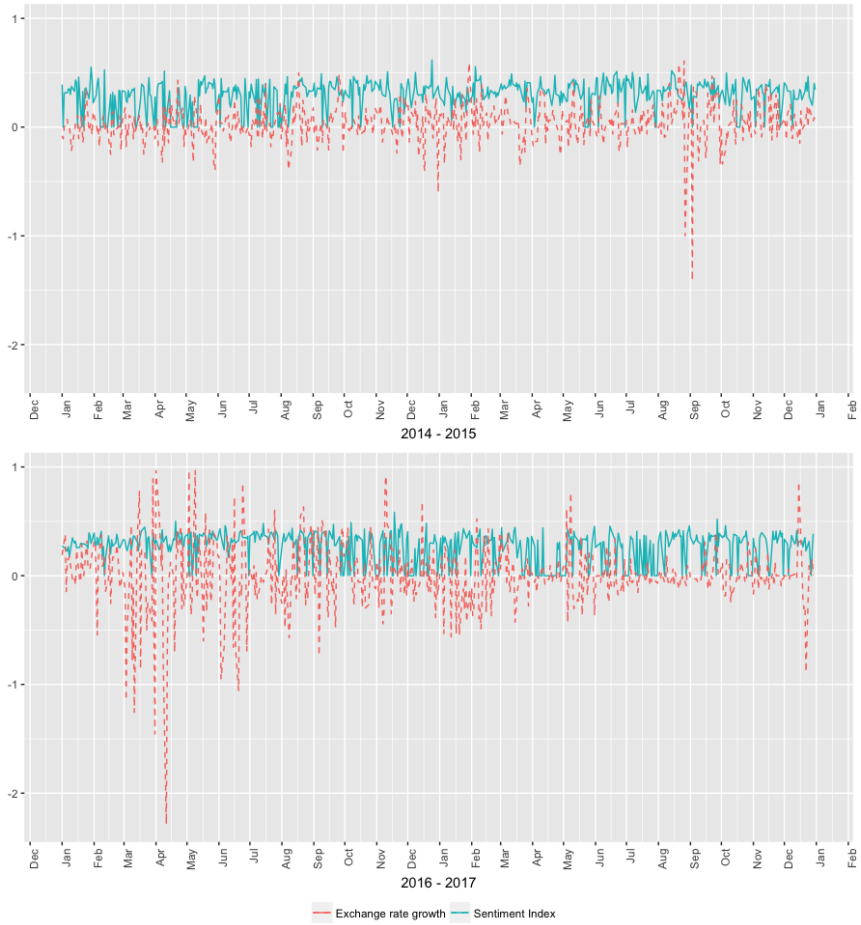
Data and News Indexes

The exchange rate growth and the occurrence of the word “volatility”, 2014-2017



Data and News Indexes

The exchange rate growth and "sentiment" index, 2014-2017



Contents

Methodology
Estimation

- Let y_t be the exchange rate growth at time t and x_t' a row vector containing news indexes observed at time t.
- $ARMA(p, q)$ model:

$$y_t = c + \phi_1 y_{t-1} + \ldots + \phi_2 y_{t-p} + u_t + \theta_1 u_{t-1} + \theta_q u_{t-q} + x_t'\gamma$$

- $GARCH(p, q)$ model with error term given by

$$u_t = \sqrt{h_t} v_t \qquad with \quad v_t \sim i.i.d(0, 1)$$

where $v_t \sim i.i.d(0, 1)$, and:

$$h_t = \kappa + \delta_1 h_{t-1} \ldots \delta_r h_{t-r} + \alpha_1 u_{t-1}^2 + \ldots + \alpha_m u_{t-m}^2 + x_t'\gamma$$

Methodology
Forecasting

- We estimate different models and construct out-of-sample forecasting series.
- We compare alternative models: adjusted R_2, and information criteria (Akaike, Schwarz and Hannan-Quinn).
- We also conduct different specification tests (individual significance, autocorrelation and conditional heteroscedasticity tests).
- Out-of-sample forecasting evaluation: Root Mean Square Error (RMSE), Mean Absolute Error (MAR), Theil inequality coefficient (Theil) and Diebold-Mariano (DM).

Contents

Results

Estimation results and goodness of fit statistics

	M0	M1	M2	M3	M4	M5	M6	M7
Mean equation								
Constant	0.012	0.012	0.004	0.014	0.008	0.013	0.004	0.005
	0.008	0.008	0.007	0.005	0.005	0.005	0.007	0.004
Dev (-1)		0.114	0.062	0.112	0.068	0.116	0.051	0.064
		0.031	0.024	0.032	0.024	0.033	0.028	0.023
Dev (-2)		-0.084	-0.078	-0.041	-0.006	-0.051	-0.035	
		0.031	0.024	0.036	0.026	0.037	0.027	
News*Pos			0.112		0.104		0.096	0.132
			0.019		0.013		0.018	0.006
NNews*Pos			0.003		0.001		0.002	
			0.001		0.000		0.001	
NCrisis*Neg			-0.025		-0.005		-0.006	
			0.005		0.003		0.006	
NVolat*Neg			0.004		-0.002		0.002	
			0.002		0.001		0.002	
Information criteria								
Adj. R2	0.0000	0.016	0.435	0.014	0.396	0.015	0.408	0.377
AIC	0.1068	0.094	-0.457	-0.323	-0.874	-0.349	-0.726	-0.908
SIC	0.1115	0.108	-0.424	-0.294	-0.827	-0.301	-0.660	-0.861
HQ	0.1085	0.099	-0.444	-0.312	-0.856	-0.331	-0.701	-0.890

Results

Estimation results and goodness of fit statistics

	M0	M1	M2	M3	M4	M5	M6	M7
Variance Equation								
Constant				0.002	0.001	0.001	0.011	0.000
				0.000	0.000	0.000	0.001	0.000
ARCH(1)				0.156	0.129	0.143	0.272	0.122
				0.015	0.012	0.016	0.029	0.010
GARCH(1)				0.824	0.863	0.822	0.512	0.860
				0.012	0.009	0.015	0.020	0.008
News*Pos						-0.003	-0.006	-0.001
						0.001	0.001	0.000
NNews*Pos						0.000	0.000	0.000
						0.000	0.000	0.000
NCrisis*Neg						0.002	0.003	0.001
						0.001	0.001	0.000
NVolat*Neg						0.001	-0.001	0.000
						0.000	0.000	0.000
Information criteria								
Adj. R2	0.0000	0.016	0.435	0.014	0.396	0.015	0.408	0.377
AIC	0.1068	0.094	-0.457	-0.323	-0.874	-0.349	-0.726	-0.908
SIC	0.1115	0.108	-0.424	-0.294	-0.827	-0.301	-0.660	-0.861
HQ	0.1085	0.099	-0.444	-0.312	-0.856	-0.331	-0.701	-0.890

Results

Out-of-sample forecasts evaluation

	M0	M1	M2	M3	M4	M5	M6	M7
Mean equation								
RMSE	0.2022	0.2033	0.1674	0.2033	**0.1511**	0.2055	0.1576	**0.1512**
MAE	0.1379	0.1380	0.1287	0.1390	**0.1086**	0.1414	0.1186	**0.1051**
Theil	0.9402	0.8727	**0.4149**	0.8859	0.4309	0.8650	**0.4151**	0.4520
Diebold-Mariano test								
M0		0.50	**-3.06**	0.65	**-5.86**	1.47	**-4.40**	**-6.45**
		0.62	0.00	0.52	0.00	0.14	0.00	0.00
M1	-0.50		-3.07	-0.04	-5.66	1.66	-4.35	-6.29
	0.62		0.00	0.97	0.00	0.10	0.00	0.00
M2	3.06	3.07		3.12	-3.68	3.25	-3.81	-2.89
	0.00	0.00		0.00	0.00	0.00	0.00	0.00
M3	-0.65	0.04	-3.12		-5.84	2.36	-4.44	-6.50
	0.52	0.97	0.00		0.00	0.02	0.00	0.00
M4	5.86	5.66	3.68	5.84		5.91	2.44	0.03
	0.00	0.00	0.00	0.00		0.00	0.02	0.97
M5	-1.47	-1.66	-3.25	-2.36	-5.91		-4.55	-6.54
	0.14	0.10	0.00	0.02	0.00		0.00	0.00
M6	4.40	4.35	3.81	4.44	-2.44	4.55		-1.66
	0.00	0.00	0.00	0.00	0.02	0.00		0.00
M7	6.45	6.29	2.89	6.50	-0.03	6.54	1.66	
	0.00	0.00	0.00	0.00	0.97	0.00	0.10	

Contents

Conclusions

- We conducted text mining and analyses of online newspaper articles published at irregular frequencies within a day and construct daily “news indexes”.
- We estimate and forecast different ARMA and GARCH models which include news indexes in the mean and variance equation.
- Overall, the results show evidence that different news indexes are helpful to forecast the dynamics of daily exchange rates.

Thanks

Newspaper Analytics and the Dynamics of the Exchange Rate in Peru

Erick Lahura and Luciana Figueroa

XXXVI Encuentro de Economistas del BCRP

October, 2018