



Business Intelligence & Analytics

Implementación de una solución de big data para el monitoreo y simulación de escenarios para la actividad económica tiempo real

Oficina de Presupuestos

Dirección Académica de Economía (DAE)

Julio 2020

Índice

- I. Objetivo
- II. Revisión bibliográfica
- III. Hipótesis de demanda de energía como proxy del PBI
- IV. Arquitectura para el procesamiento de alta frecuencia
- V. Metodología econométrica seleccionada
- VI. Siguietes pasos

Identificación y cuantificación oportuna de riesgos
macroeconómicos para una eficiente previsión
presupuestal

II. Revisión bibliográfica

Revisión de literatura : Nowcasting de la actividad económica

Box - Jenkins
(1976)

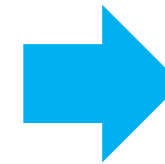
- Metodología ampliamente utilizada para el tratamiento y modelado de series de tiempo no estacionarias.

Scott & Varian
(2014)

- Se estudian diversos indicadores proxy de para la actividad económica encontrándose la generación de electricidad como mejor indicador sobre otros.

Bok & Caratelli
(2018).

- Se utilizan técnicas de Big Data sobre una plataforma automatizada para la extracción continua de datos y proyección inmediata de diversos indicadores de mercado.



Justificación de selección de
metodología
Box - Jenkins

- Metodología base para modelar series de tiempo
- Es parsimoniosa y robusta para muestras suficientemente grandes.
- Muy intuitiva y simple de implementar.

III. Hipótesis de la demanda de energía como proxy del PBI

Intuición de acuerdo a Scott & Varian, 2014
Tasa de crecimiento interanual del PBI real (estimado)



$$y_t^{PBI} = \beta_0(y_t^{DE}) + \beta_1(\varepsilon_{t-1})$$

Donde:

- y_t^{PBI} : Tasa de crecimiento interanual del PBI real (Perú)
- y_t^{DE} : Tasa de crecimiento interanual de la demanda eléctrica
- \hat{y}_t^{PBI} : Tasa de crecimiento interanual del PBI estimada en tiempo real

Prueba de hipótesis
Metodología de Mínimos Cuadrados Ordinarios

Dep. Variable:	y	R-squared (uncentered):	0.929			
Model:	OLS	Adj. R-squared (uncentered):	0.926			
Method:	Least Squares	F-statistic:	322.0			
Date:	Wed, 22 Jul 2020	Prob (F-statistic):	6.47e-29			
Time:	16:35:13	Log-Likelihood:	141.13			
No. Observations:	51	AIC:	-278.3			
Df Residuals:	49	BIC:	-274.4			
Df Model:	2					
Covariance Type:	nonrobust					
=====						
	coef	std err	t	P> t	[0.025	0.975]

x1	0.5695	0.104	5.496	0.000	0.361	0.778
x2	0.7592	0.031	24.571	0.000	0.697	0.821
=====						
Omnibus:	22.351	Durbin-Watson:	2.069			
Prob(Omnibus):	0.000	Jarque-Bera (JB):	48.538	(*)		
Skew:	1.203	Prob(JB):	2.88e-11			
Kurtosis:	7.129	Cond. No.	3.36			
=====						

(*) Se tomó una muestra pequeña por ser de rápido procesamiento, pudiendo extenderse esta hasta el año 2001

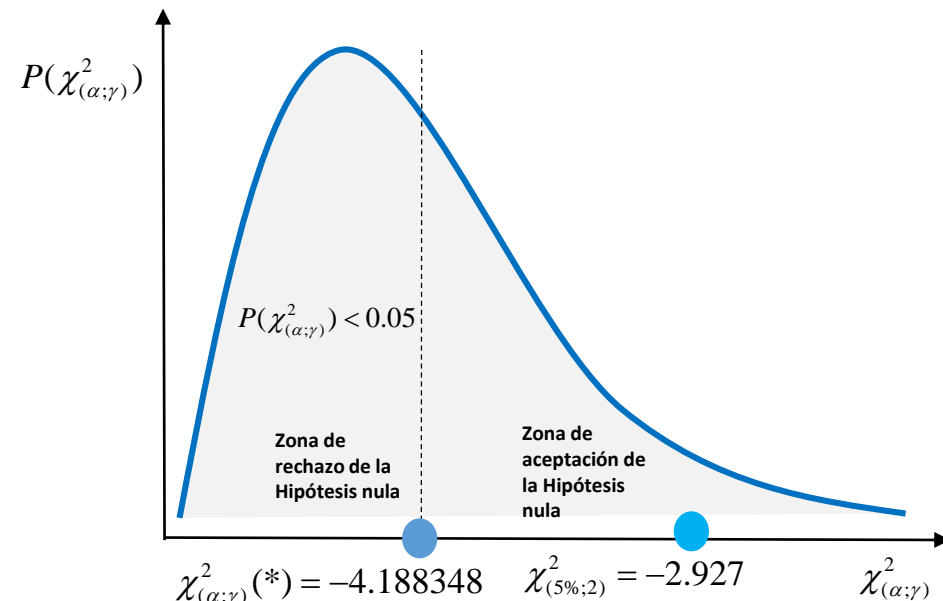
Pruebas de estacionariedad de las series seleccionadas



III. Hipótesis de la demanda de energía como proxy del PBI



Pruebas de estacionariedad de las series seleccionadas



Pruebas de ADF – Demanda de energía eléctrica (exógena)

Null Hypothesis(H0): Serie has a unit root

ADF Statistic: -4.188348

p-value: 0.000689

Critical Values:

1%: -3.581

5%: -2.927

10%: -2.602

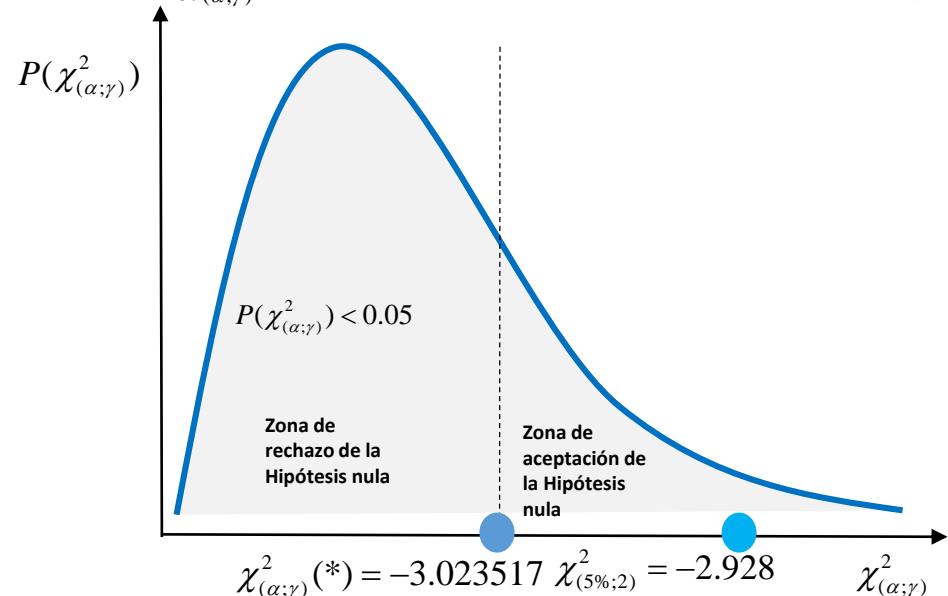
Numbers of lag used : 4.000000

Numbers of observations used: 46.000000

Criterio de Información : -232.031085

Se rechaza H0, de rechazar H0 la probabilidad de equivocarnos sería 0.000689

Por tanto: La serie no tiene raíz unitaria, entonces es estacionaria



Pruebas de ADF – PBI (endógena)

Null Hypothesis(H0): Serie has a unit root

ADF Statistic: -3.023517

p-value: 0.032756

Critical Values:

1%: -3.585

5%: -2.928

10%: -2.602

Numbers of lag used : 5.000000

Numbers of observations used: 45.000000

Criterio de Información : -226.835148

Se rechaza H0, de rechazar H0 la probabilidad de equivocarnos sería 0.032756

Por tanto: La serie no tiene raíz unitaria, entonces es estacionaria

IV. Arquitectura para el procesamiento de alta frecuencia

Big data pipeline (*)

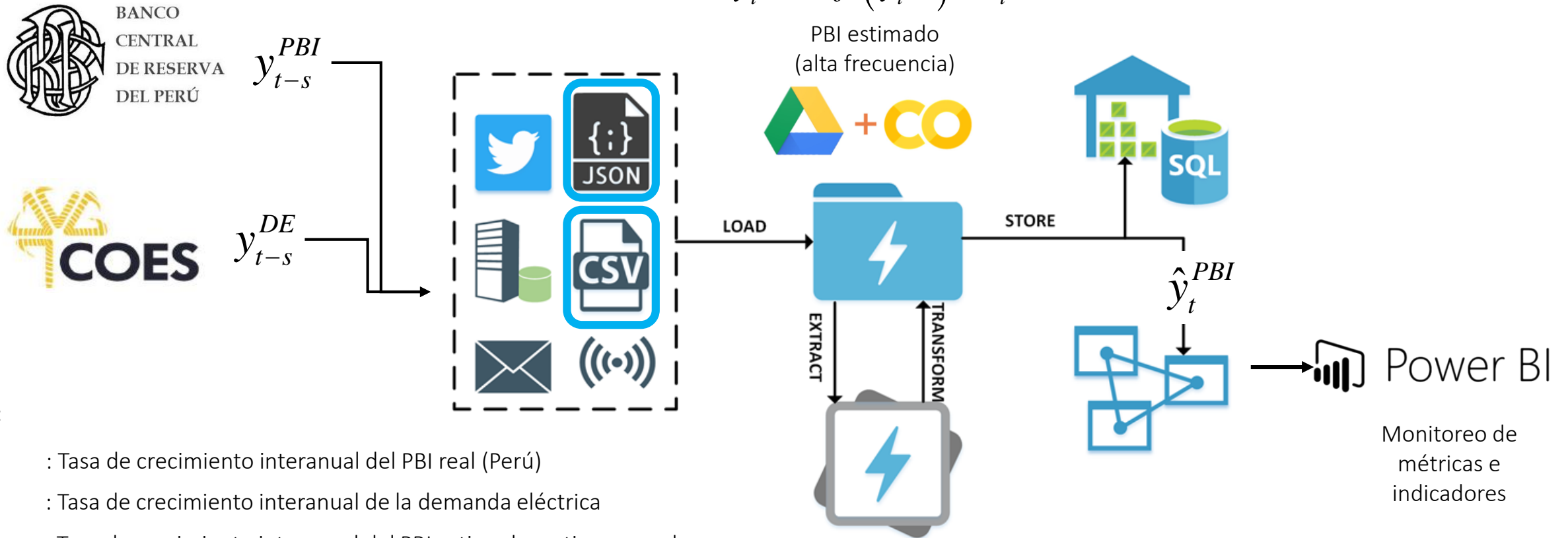
Extracción (Web scraping)

Carga

Transformación

Almacenamiento

Visualización



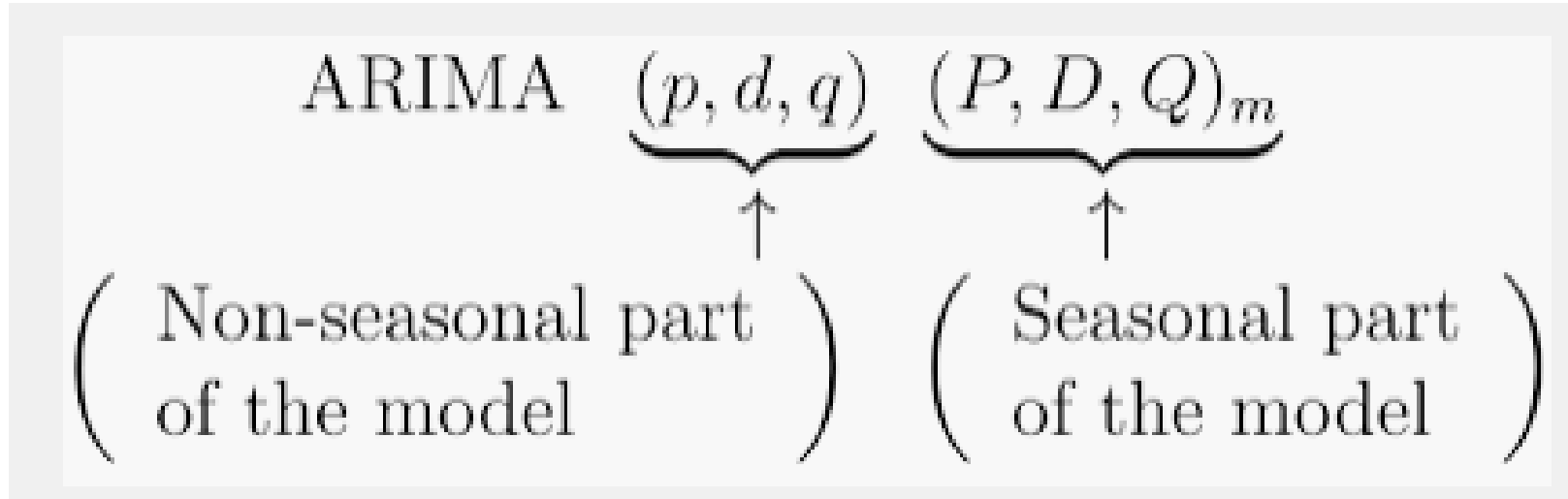
Donde:

- y_t^{PBI} : Tasa de crecimiento interanual del PBI real (Perú)
- y_t^{DE} : Tasa de crecimiento interanual de la demanda eléctrica
- \hat{y}_t^{PBI} : Tasa de crecimiento interanual del PBI estimada en tiempo real

(*) Se esta evaluando migrar a la arquitectura cloud AWS

IV. Metodología econométrica seleccionada

Metodología de Box – Jenkins



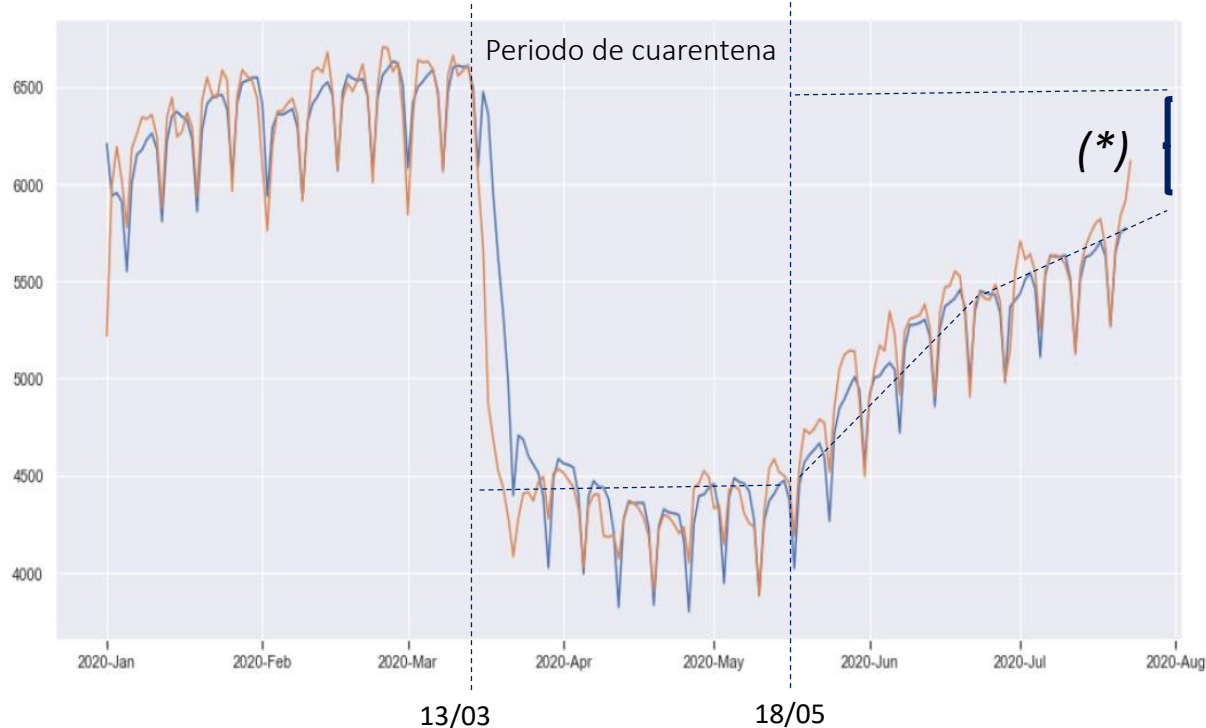
Donde

- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.
- **P**: Seasonal autoregressive order.
- **D**: Seasonal difference order.
- **Q**: Seasonal moving average order.
- **m**: The number of time steps for a single seasonal period.

IV. Metodología econométrica seleccionada

Intuición de acuerdo a Scott & Varian, 2014

Evolución de la Demanda de energía eléctrica (Expresado en GW)



$$y_t = SARIMA(1, 0, 1)(1, 0, 1)_7$$

- **p**: Trend autoregression order.
- **d**: Trend difference order.
- **q**: Trend moving average order.
- **P**: Seasonal autoregressive order.
- **D**: Seasonal difference order.
- **Q**: Seasonal moving average order.
- **m**: The number of time steps for a single seasonal period.

Prueba de hipótesis

Metodología SARIMA

Dep. Variable:	Ejecutado	No. Observations:	5318			
Model:	SARIMAX(1, 0, 1)x(1, 0, 1, 7)	Log Likelihood	-35872.769			
Date:	Wed, 22 Jul 2020	AIC	71757.537			
Time:	23:19:35	BIC	71797.011			
Sample:	01-01-2006	HQIC	71771.329			
	- 07-23-2020					
Covariance Type:	opg					
=====						
	coef	std err	z	P> z	[0.025	0.975]

intercept	73.2865	3.531	20.752	0.000	66.365	80.208
ar.L1	0.6809	0.008	89.924	0.000	0.666	0.696
ma.L1	-0.2470	0.008	-31.517	0.000	-0.262	-0.232
ar.S.L7	0.9512	0.003	367.809	0.000	0.946	0.956
ma.S.L7	-0.5665	0.005	-124.528	0.000	-0.575	-0.558
sigma2	4.116e+04	151.256	272.120	0.000	4.09e+04	4.15e+04
=====						
Ljung-Box (Q):	1672.66	Jarque-Bera (JB):	1759532.26			
Prob(Q):	0.00	Prob(JB):	0.00			
Heteroskedasticity (H):	0.27	Skew:	-4.98			
Prob(H) (two-sided):	0.00	Kurtosis:	91.55			
=====						

(*) “¿Cuánto Tardará la económica peruana en recuperarse?”

Por calcular las diferentes trayectorias y plazos para volver a los niveles de demanda de energía eléctrica previos a la pandemia.

<https://www.facebook.com/watch/?v=318813109273668>

VI. Siguietes pasos

Cronograma de actividades

Etapa	Item	Actividades	Estatus
Marco Teórico	01	Revisión bibliográfica	Hecho
	02	Selección de bibliografía base	Hecho
Web Scraping	03	Desarrollo de codigos para la extraccion de datos intradiarios de demanda de energia (COES)	Hecho
	04	Pruebas de hipotesis iniciales (PBI vs electrica)	Hecho
Estrategía metodológica	05	Selección metodologica	Hecho
	06	Especificacion econometrica	Hecho
	07	Pruebas de hipotesis, proyecciones y simulacion de escenario para el PBI	Haciendo
Estrategía ETL	08	Desarrollo de la infraestructura para la carga y procesamiento de datos masivos	Por hacer
	09	Diseño e implementacion de reportesBI cloud	Por hacer
Integracion al sistema de monitoreo	10	Integracion al modelo para la proyeccion de ingresos academicos y inmobiliarios PUCP	Por hacer

Predicting daily electricity loads for a building on the UC Berkeley campus

<https://towardsdatascience.com/time-series-forecasting-with-a-sarima-model-db051b7ae459>

- Bok, B., Caratelli, D., Giannone, D., Sbordone, A., & Tambalotti, A. (2018). Macroeconomic Nowcasting and Forecasting with Big Data. *Annual Review Of Economics*, 10(1), 615-643. doi: 10.1146/annurev-economics-080217-053214
- Bragoli, D. (2017). Now-casting the Japanese economy. *International Journal Of Forecasting*, 33(2), 390-402. doi: 10.1016/j.ijforecast.2016.11.004
- Bragoli, D., & Modugno, M. (2017). A now-casting model for Canada: Do U.S. variables matter?. *International Journal Of Forecasting*, 33(4), 786-800. doi: 10.1016/j.ijforecast.2017.03.002
- Bragoli, D., Metelli, L., & Modugno, M. (2014). The Importance of Updating: Evidence from a Brazilian Nowcasting Model. *SSRN Electronic Journal*. doi: 10.2139/ssrn.2529168
- Evans, M. (2005). Where Are We Now? Real-Time Estimates of the Macro Economy. *SSRN Electronic Journal*. doi: 10.2139/ssrn.646103
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal Of Monetary Economics*, 55(4), 665-676. doi: 10.1016/j.jmoneco.2008.05.010
- Raftery, A., Madigan, D., & Hoeting, J. (1997). Bayesian Model Averaging for Linear Regression Models. *Journal Of The American Statistical Association*, 92(437), 179-191. doi: 10.1080/01621459.1997.10473615
- Scott, S., & Varian, H. (2014). Predicting the present with Bayesian structural time series. *International Journal Of Mathematical Modelling And Numerical Optimisation*, 5(1/2), 4. doi: 10.1504/ijmmno.2014.059942
- Box, G. E. P., & Jenkins, G. M. (1976). *Time series analysis forecasting and control* - rev. ed. Oakland, California, Holden-Day, 1976, 37 (2), 238 - 242

Business Intelligence & Analytics

Sergio Valera
Data Analytics
s.valera@pucp.pe

Carlos Loayza
Business Intelligence
Carlos.loayza@pucp.edu.pe



De parvis grandis acervus erit