

Proyecto de tweets y de curvas de tasas

31 de marzo de 2021

- Proyecto de Tweets
 - Extracción de Tweets
 - Filtrado de Tweets
 - Análisis de sentimiento y métricas de eventos
 - Resultados y Conclusiones
- Proyecto de curvas de tasas
 - Modelo DNS: parámetro λ
 - Datos
 - Filtros de Kalman
 - Comparación de métodos
 - Variables exógenas
 - Conclusiones y Extensiones

Proyecto de Tweets

- Proyecto de Tweets
 - Extracción de Tweets
 - Filtrado de Tweets
 - Análisis de sentimiento y métricas de eventos
 - Resultados y Conclusiones

Objetivos

Los principales objetivos de este trabajo son:

- Construir una métrica para la percepción del de los distintos agentes sobre los principales bancos chilenos.
- Índice de Sentimiento Bancario: Crear un Índice de Sentimiento Financiero que permita determinar la percepción del servicio bancario.

Extracción de Tweets: Métodos y librerías: (Python)

Basándonos en varias librerías y métodos existentes (*TwitterAPI*, *BeautifulSoup4*, *Selenium*, entre muchas otras) desarrollamos una aplicación (" β ") que nos permite acceder a tweets históricos sin límite de cantidad ni tiempo y mucho más.



Extracción de Tweets: Métodos y librerías: (Python)

Basándonos en varias librerías y métodos existentes (*TwitterAPI*, *BeautifulSoup4*, *Selenium*, entre muchas otras) desarrollamos una aplicación ("β") que nos permite acceder a tweets históricos sin límite de cantidad ni tiempo.

```
AppSBC ×
C:\ProgramData\Anaconda3\envs\EjemploTweets\Codigo\AppSBC.py
1377289085484683269 2021-03-31 12:58:01 -0300 <bcentralchile> Hoy a las 14.00 horas, el Presidente del Banco Central de Chile, Mario Marcel, ofrecerá una conferencia de prensa sobre la situación económica y monetaria del país. El #IPoM de marzo 2021 se presentará en la conferencia de prensa.
1377288731921575939 2021-03-31 12:56:47 -0300 <bcentralchile> Nota de prensa | Presidente del Banco Central de Chile, Mario Marcel, presentó Informe de Política Monetaria #IPoM de marzo 2021.
1377288536177709058 2021-03-31 12:55:50 -0300 <bcentralchile> La presentación del Presidente del Banco Central de Chile, Mario Marcel, sobre el #IPoM de marzo 2021 efectuada ante los medios de comunicación.
1377283419462336517 2021-03-31 12:35:30 -0300 <bcentralchile> Presidente del Banco Central de Chile, Mario Marcel, responde preguntas de los Senadores en la presentación del #IPoM de marzo 2021.
1377268174111002625 2021-03-31 11:34:55 -0300 <bcentralchile> El Banco Central de Chile presenta, a través de su Presidente Mario Marcel, las principales proyecciones económicas del año.
1377262977838161925 2021-03-31 11:14:16 -0300 <bcentralchile> Presidente del Banco Central de Chile, Mario Marcel, expone escenario para la economía chilena al presentar el #IPoM de marzo 2021.
1377254779815129092 2021-03-31 10:41:42 -0300 <bcentralchile> Comienza la presentación del Presidente del Banco Central de Chile, Mario Marcel, a nombre del Consejo, del Informe de Política Monetaria #IPoM de marzo 2021.
13772366435354157571 2021-03-31 09:29:38 -0300 <bcentralchile> El Presidente del Banco Central de Chile, Mario Marcel, ofrecerá, hoy a las 14.00 horas, una conferencia de prensa sobre la situación económica y monetaria del país.
1377236458808243712 2021-03-31 09:28:54 -0300 <bcentralchile> A las 10.00 horas, el Consejo del Banco Central de Chile expone, a través de su Presidente, Mario Marcel, el #IPoM de marzo 2021.
1377235931921575939 2021-03-31 09:26:48 -0300 <bcentralchile> El #IPoM de marzo 2021 sostiene que el Banco Central y otras instituciones públicas han hecho un esfuerzo importante para mitigar la crisis económica provocada por la pandemia.
1377234909371195399 2021-03-31 09:22:44 -0300 <bcentralchile> Según el #IPoM del Banco Central, persisten muchos riesgos, pues lo inédito de la situación causada por el Covid-19 hace que sea difícil prever su evolución.
1377234425407225858 2021-03-31 09:20:49 -0300 <bcentralchile> El #IPoM de marzo del Banco Central sostiene que varias economías, como China y EE.UU., también se han recuperado antes que Chile.
1377234107369998810 2021-03-31 09:19:33 -0300 <bcentralchile> Según el #IPoM del BCC, la recuperación de la economía ha ido más rápido que lo previsto. La mejor adaptación de las políticas monetarias ha sido clave para este resultado.
137723261039636097 2021-03-31 09:15:36 -0300 <bcentralchile> ¿Qué nos dice este #IPoM? A continuación, un resumen de los principales mensajes del Informe de Política Monetaria de marzo 2021.
1377228021525719042 2021-03-31 08:56:05 -0300 <bcentralchile> Banco Central de Chile publicó los Indicadores de Coyuntura Semanal #ICS al 31 de marzo de 2021. https://t.co/0k0M1Z0fEh
13772251515488647106 2021-03-31 08:44:37 -0300 <bcentralchile> Nota de prensa | Banco Central de Chile presentó Informe de Política Monetaria #IPoM de marzo 2021 https://t.co/enD0cs1
1377222594542370817 2021-03-31 08:33:48 -0300 <bcentralchile> Banco Central de Chile publicó Informe de Política Monetaria #IPoM de marzo 2021. Link al documento: https://t.co/uWMyV
[!] No more data! Scraping will stop now.
found 0 deleted tweets in this search.
```

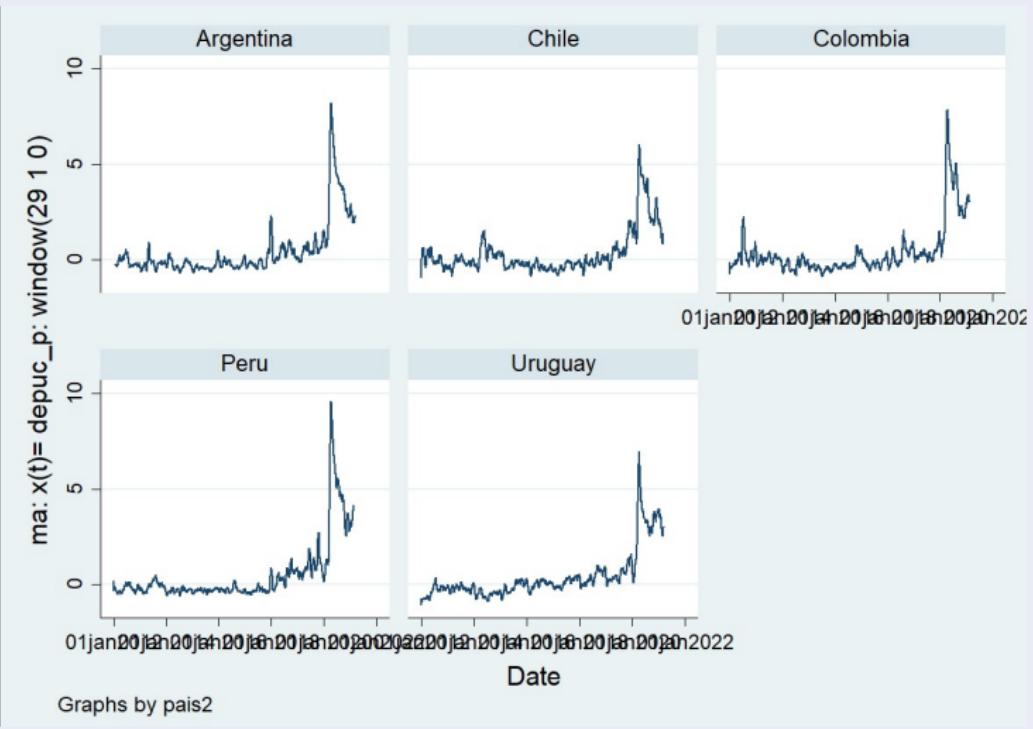
Ejemplo y Colaboración con otros proyectos

Depuc

Impacto en el mercado inmobiliario

Y varios más!!!...

Ejemplo y Colaboración con otros proyectos



Filtrado de Tweets

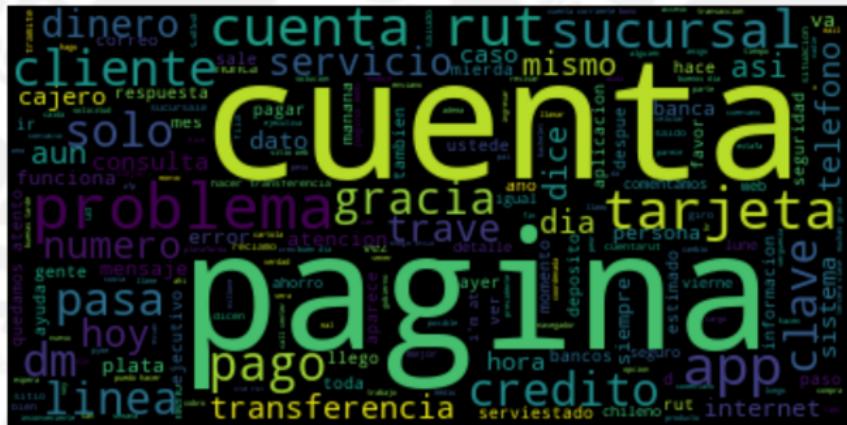
Se descargó:

- Banco Estado: 1.072.421 tweets filtrados, 1Gb
- Banco de Chile: 194.812 tweets filtrados, 450Mb
- Banco BCI: 318.337 tweets filtrados, 300Mb
- Totales filtrados: 1.318.031 tweets

para las fechas: Enero 2010 hasta Marzo 2021

Palabras en Tweets

WordCloud de todos los tweets (1.318.031 tweets):



Análisis de sentimiento y métricas

Análisis de sentimiento

SentimentAnalysisSpanish 0.0.25

Natural Language Toolkit

$S(tweet) = p$:Probabilidad que el texto sea positivo.

e.g.

- '*comienzo 2010 feliz x mi familia, gratitud bancoestado cestas octantis; esperanza x chile y entusiasmado x nuevo equipo y desafio innovacion*' $\Rightarrow p = 0,87$
- '*toda la tarde tratando de realizar un pago y nada solo dice: en estos momentos no podemos atender su solicitud cod01*' $\Rightarrow p = 0$

Análisis de sentimiento

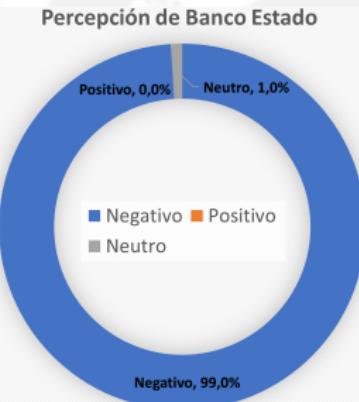
Para la evaluación de los tweets se consideraró la siguiente partición:

- $p = S(tweet) \geq 0,7 \Rightarrow$ tweet es cosiderado positivo
- $p = S(tweet) \leq 0,3 \Rightarrow$ tweet es considerado negativo
- $0,3 < p = S(tweet) < 0,7 \Rightarrow$ tweet es considerado neutro

Análisis de sentimiento



(a) Banco BCI
318.337 tweets
(0.1 % , 2.3 %, 97.6 %)



(b) Banco Estado
1.072.421 tweets
(0.0 % , 1.0 %, 99.0 %)



(c) Banco de Chile
194.812 tweets
(0.1 % , 3.7 %, 96.1 %)

Figura: Percepción de sentimiento medio diario ($\left(\frac{1}{n} \sum S(tweet)\right)$) hacia los distintos bancos.
percepción = (positivo, neutro, negativo)

Métrica

Se consideran 3 factores:

- Volumen de tweets diario : V_d
- Sentimiento diario : S_d
- Subjetividad de sentimiento : $\varphi_d = 1 - (s_d - S_d)$

Donde,

s_d : sentimiento promedio diario, $\frac{1}{n} \sum \mathcal{S}(\text{tweet}_i)$

S_d : sentimiento total diario, $\mathcal{S}(\sum(\text{tweet}_i))$

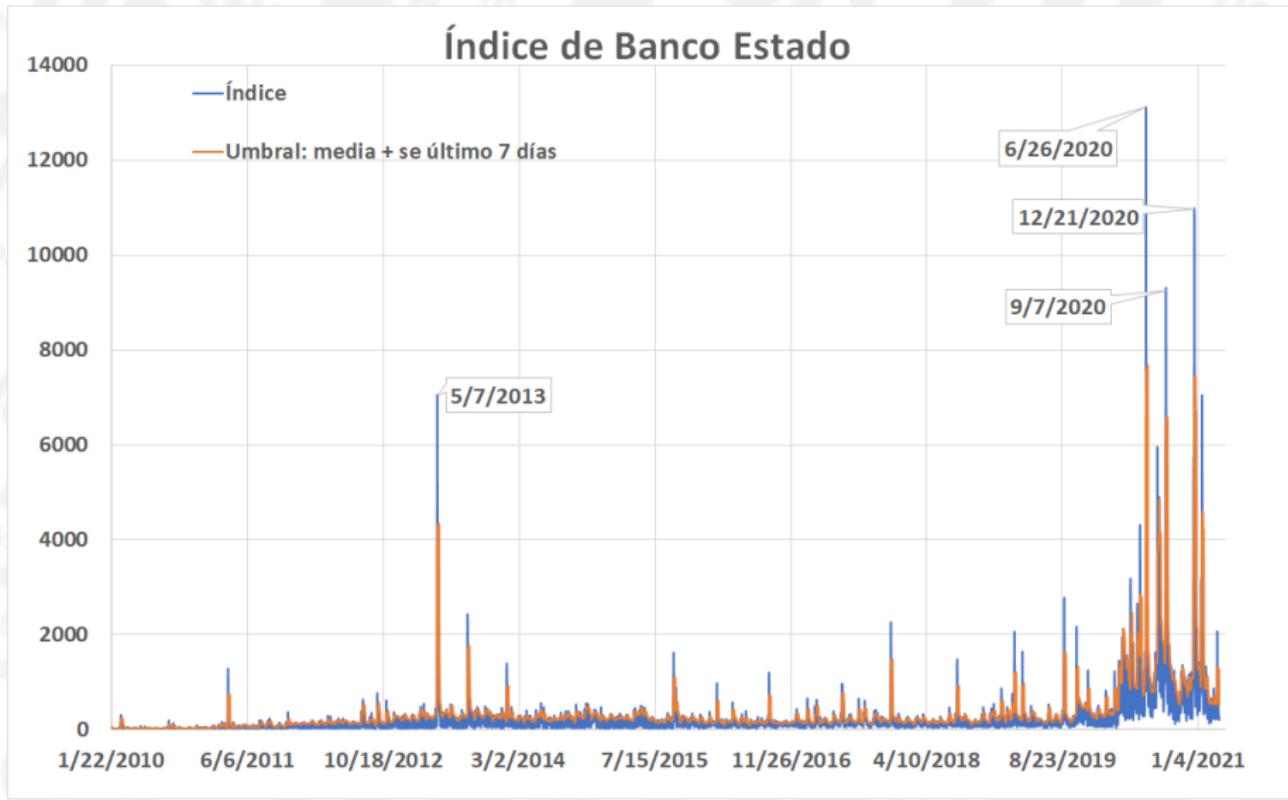
Entónces la métrica se define como:

$$M_d = V_d \cdot \varphi_d \cdot (1 - S_d)$$

└ Proyecto de Tweets

└ Análisis de sentimiento y métricas

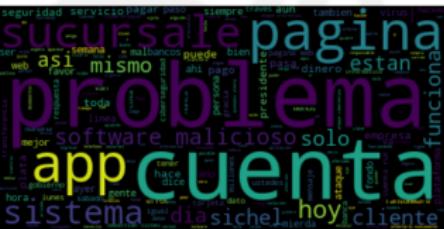
El caso de Banco Estado



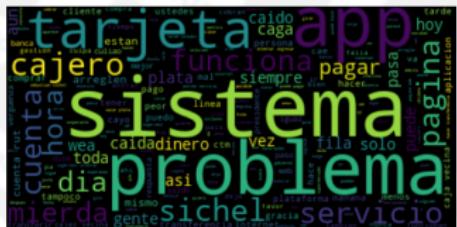
Resultados



(a) 2013-05-07



(b) 2020-09-07



(c) 2020-06-26



(d) 2020-12-21

Figura: WordClouds correspondientes a total de tweets del día para Banco Estado

Proyecto de Tweets

Resultados y Conclusiones

Eventos

Google

2013-05-07 banco estado

[Todos](#) [Imágenes](#) [Maps](#) [Noticias](#) [Videos](#) [Más](#) [Preferencias](#) [Herramientas](#)

Cerca de 64.800 resultados (0.54 segundos)

<https://m.elmostrador.cl> • mercados • 2013/05/07 • el-c... ▾

El caso de abusos del BancoEstado salpica a nombres ...

El caso de abusos del **BancoEstado** salpica a nombres emblemáticos de la Concertación, por Héctor Cárcamo, 7 mayo, 2013. El caso de abusos del ...
Visítate esta página el 30/03/21.

<https://m.elmostrador.cl> • noticias • país • 2013/05/07 •

Director del Sernac: "El dolor es más profundo cuando se trata ...
7 mayo, 2013 ... "El dolor es más profundo cuando se trata de **BancoEstado** y además se trata de afectados que vienen de los sectores más vulnerables, que ...

<https://www.sernac.cl> • portal • w3-article-4629 •

Acuerdo ante Corte Suprema: BancoEstado devolverá casi ...
... el 2004, en [...] 07 de mayo de 2013. BancoEstado devolverá casi \$5.700 millones a consumidores por haber realizado cobros de comisiones ilegales ...

2020-09-07 banco estado - Buscar

<https://google.com/search?q=2020-09-07+banco+estado&sxsrf=ALeKj020yfC>

Google

2020-09-07 banco estado

[Todos](#) [Maps](#) [Noticias](#) [Imágenes](#) [Videos](#) [Más](#) [Preferencias](#) [Herramientas](#)

Cerca de 127.000 resultados (0.72 segundos)

<https://lm.alejandradel.com/nacionales/2020/09/07/> • org ...

Alerta por amenaza informática: Banco Estado cierra ...
por El Mostrador 7 septiembre, 2020. Alerta por amenaza informática. Banco Estado cierra su sistema este lunes para garantizar que clientes no. La situación ...

<https://lm.alejandradel.com/nacionales/2020/09/07/> • org ...

El lunes negro del Banco Estado: Gobierno admite "ataque ...
7 septiembre, 2020 ... En tanto, el presidente del **BancoEstado**, Sebastián Sichel, junto con su equipo de trabajo, se reunió con autoridades nacionales para evaluar las consecuencias de la situación ...

<https://www.enem.com/ciberataque-bancosteado/> •

"Crisis operativa" y 12 mil computadores infectados: Los ...
07 de Septiembre de 2020 | 18:17 | Por A. González | Enel, Sebastián Sichel, presidente de **BancoEstado**. Alguno de los sistemas de la entidad quedaron fuera de servicio durante más de 15 horas.

<https://www.t13.cl/noticia/repercusion-bancosteado.html> •

BancoEstado y software malicioso: Clientes no han sido ... - T13
7 sep, 2020 — BancoEstado avanza que hasta de sus clientes no se han visto afectados por software malicioso. En medio ... Lunes 07, Septiembre 2020 ...

<https://www.t13.cl/noticia/repercusion-bancosteado.html> •

BancoEstado no abrió sus sucursales por software malicioso ...
7 sep, 2020 — BancoEstado avanza que hasta de sus clientes no se han visto afectados por software malicioso. En medio ... Lunes 07, Septiembre 2020 ...

Conclusiones y Extensiones

- Se logra capturar eventos de exposición operacional.
- Existe coherencia en los WordClouds de los eventos y el evento.

Extensiones:

- Extender el análisis a más bancos.
- Explorar uso de distintas métricas y umbrales.
- Desarrollo de librería de *sentiment analysis* propia para bancos en Chile.

Proyecto de curvas de tasas

- Proyecto de curvas de tasas
 - Modelo DNS: parámetro λ
 - Datos
 - Filtros de Kalman
 - Comparación de métodos
 - Variables exógenas
 - Conclusiones y Extensiones

Los principales objetivos de este trabajo son:

- Estimar estados latentes del modelo DNS
- Utilizar métodos de Filtros de Kalman.
- Implementar filtros para estructura de datos incompletos.

└ Proyecto de Tasas

└ Objetivos

Modelo DNS

$$R(m)_t = L_t + S_t \frac{1 - e^{-\lambda_t m}}{\lambda_t m} + C_t \left(\frac{1 - e^{-\lambda_t m}}{\lambda_t m} - e^{-\lambda_t m} \right)$$

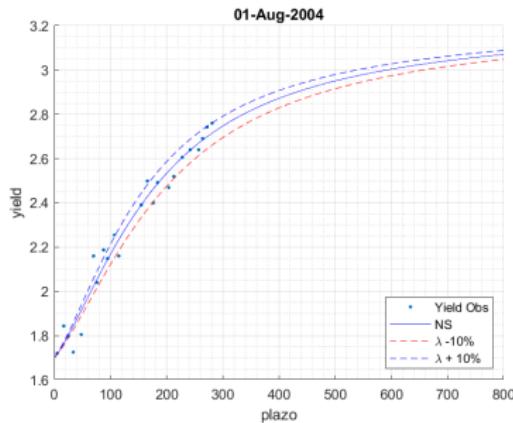
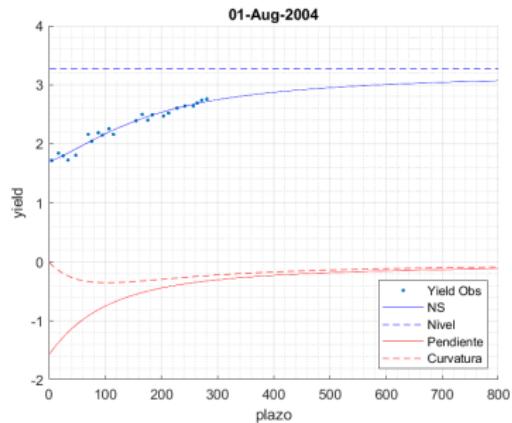


Figura: Descomposición de la curva de NS: $L = 3.270$, $S = -1.5725$, $C = -1.1816$, $\lambda = 0.0172$

Datos

Datos

Los datos corresponden a transacciones en mercado secundario de depósitos a plazo de 3 bancos en estudio.

Banco	Nº Trans.	Plazo medio	Plazo se	P95 %
A	498.052	96	107	344
B	329.581	90	141	236
C	385.474	131	142	366

fechas de 2003-01-02 hasta 2021-03-05 con estructura:

Fecha	Tasa	Monto	Plazo
-------	------	-------	-------

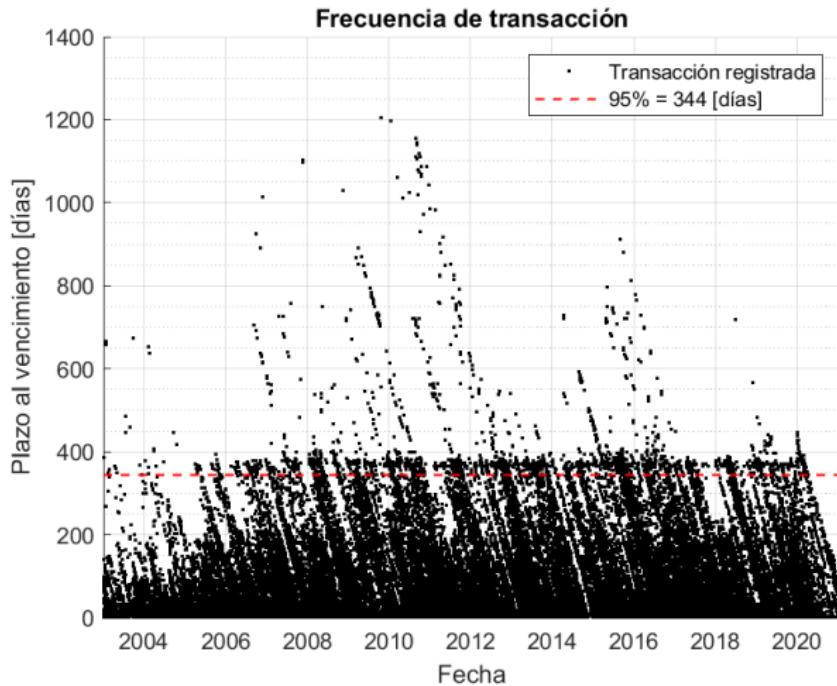
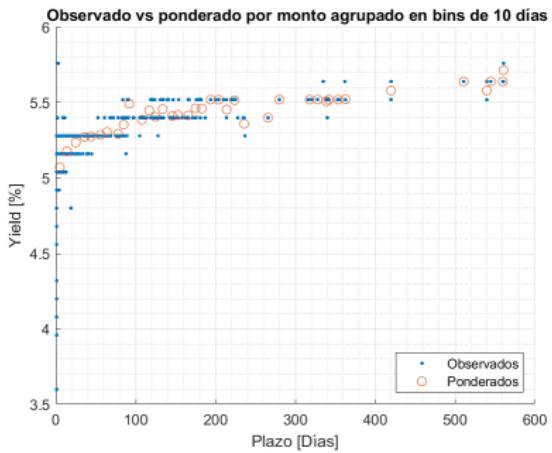


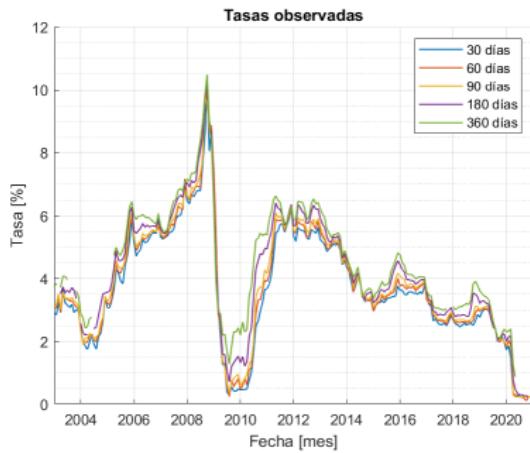
Figura: Transacciones correspondientes a depósitos de banco A.

Tratamiento de datos

- Winsorizing al 90 %
- Plazo ponderado por monto de transacción
- Tasa ponderada por monto de transacción
- Bins en plazo al vencimiento]0,10],]10, 20]...
- Tasas observadas: 30 días = promedio 20-40 días, margen de $\pm 20\%$ para tasas largas.



(a) **Datos procesados :**
wíndorizados y ponderados por monto
según bin de plazo.



(b) **Tasas observadas :**
promedio de datos procesados en rango
según plazo al vencimiento.

Filtros de Kalman

Filtros

Para el sistema general:

$$\begin{array}{ll} \text{State ec.} & x_t = F(x_{t-1}, u_t, \varepsilon_t, w_x) \quad , \varepsilon_t \sim N(0, Q) \\ \text{Measurement ec.} & y_t = H_t(x_t, v_t, w_y) \quad , v_t \sim N_t(0, R_t) \end{array}$$

se tiene el esquema de predicción óptimo:

$$\hat{x}_k = (\text{predictor óptimo de } x_{k|k-1}) + G_k[y_k - (\text{predictor óptimo de } y_{k|x_{k-1}})]$$

donde,

$$\begin{aligned} x_{k|k-1} &= E[x_{t-1}, u_{t-1}, \varepsilon_t, w_x] | x_{t-1} \\ y_{k|k-1} &= E[H(x_k, v_t, w_y)] | x_{t-1} \\ G_k &= \text{Cov}[x, y] \text{Cov}[y, y]^{-1} \end{aligned}$$

Modelo

State ec. $x_t = \mu + Ax_{t-1} + B\varepsilon_t$, $\varepsilon_t \sim N(0, Id)$
Measurement ec. $y_t = \Gamma_t(x_t) + \sigma\nu_t$, $\nu_t \sim N_t(0, Id)$

donde $Q = BB'$, con B triangular inferior y

$$\Gamma_t(x_t) = \begin{pmatrix} L_t + S_t \frac{1 - \exp(-\lambda_t m_1^t)}{\lambda_t m_1^t} + C_t \left(\frac{1 - \exp(-\lambda_t m_1^t)}{\lambda_t m_1^t} - \exp(-\lambda_t m_1^t) \right) \\ L_t + S_t \frac{1 - \exp(-\lambda_t m_2^t)}{\lambda_t m_2^t} + C_t \left(\frac{1 - \exp(-\lambda_t m_2^t)}{\lambda_t m_2^t} - \exp(-\lambda_t m_2^t) \right) \\ \vdots \\ L_t + S_t \frac{1 - \exp(-\lambda_t m_n^t)}{\lambda_t m_n^t} + C_t \left(\frac{1 - \exp(-\lambda_t m_n^t)}{\lambda_t m_n^t} - \exp(-\lambda_t m_n^t) \right) \end{pmatrix}$$

Kalman Filter Stochastic Processes

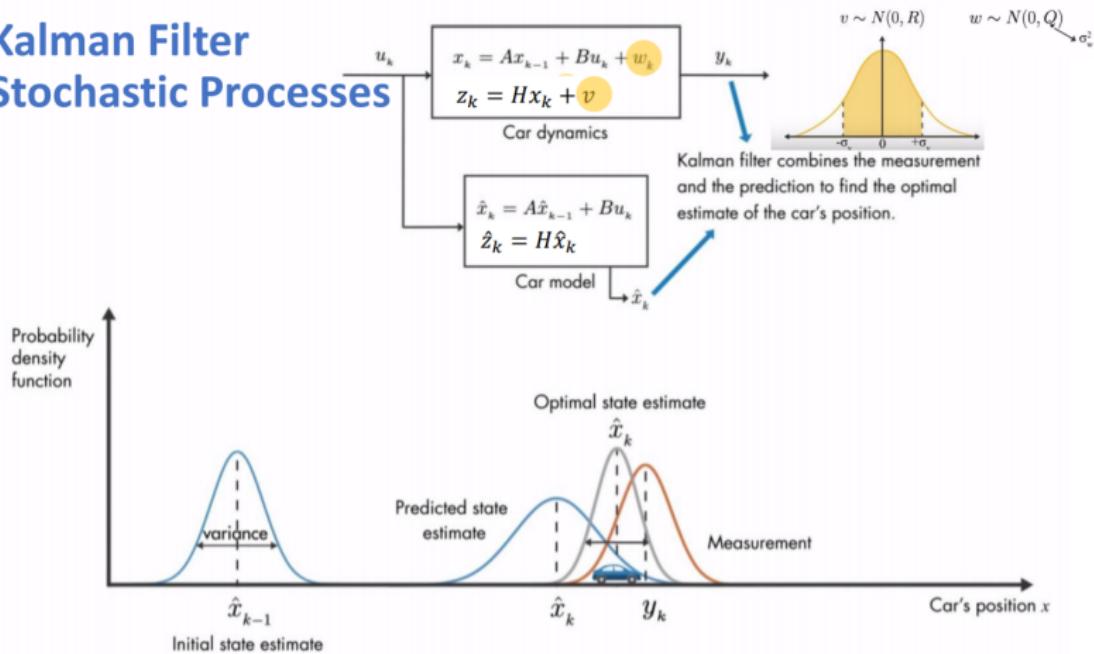
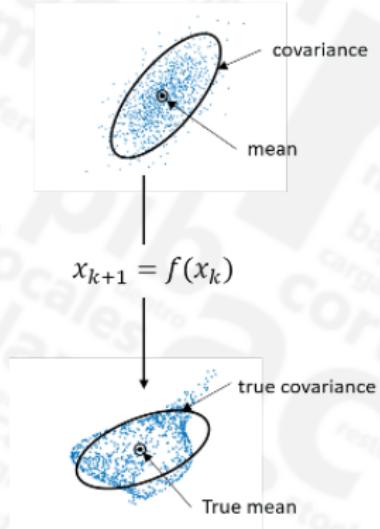


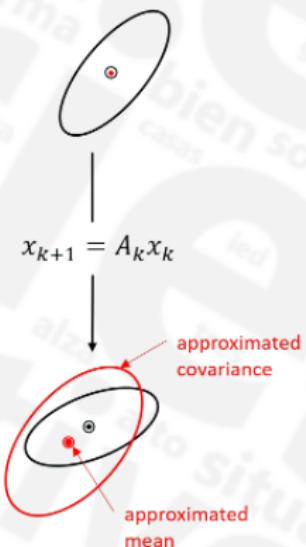
Figura: <https://www.mathworks.com/videos/understanding-kalman-filters-part-3-optimal-state-estimator-1490710645421.html>

Unscented Kalman Filter

True Transformation



Linear Transformation (EKF)



UT Transformation (UKF)

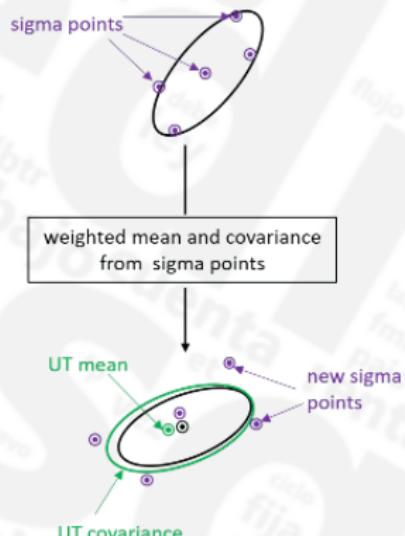
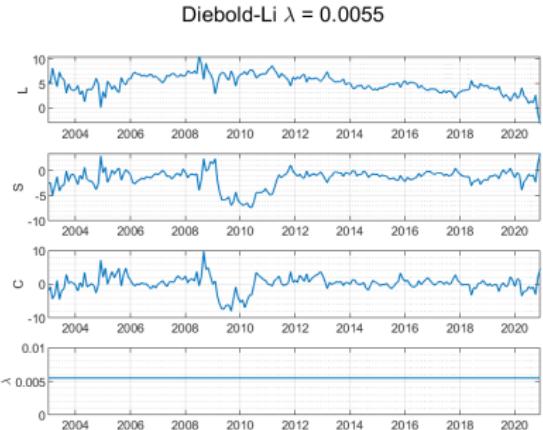


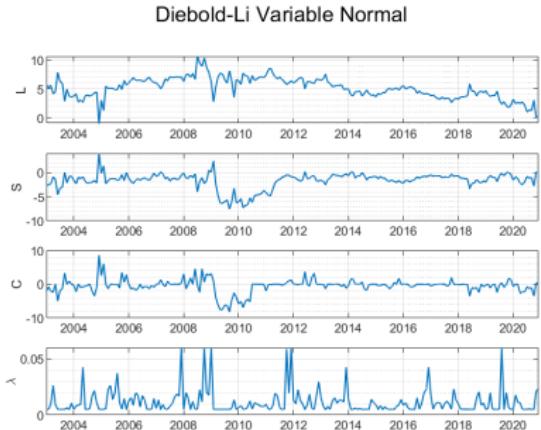
Figura: <https://www.mathworks.com/help/fusion/ug/introduction-to-estimation-filters.html>

Comparación de métodos

Métodos de Diebold-Li, λ fijo y variable



(a) Método de Diebold-Li Banco A

(b) Método de Diebold-Li con λ variable para banco A.
Mallado $\lambda = 0.005:1e-4:0.06$

Diebold-Li y Filtro de Kalman para λ fijo

KF vs DL

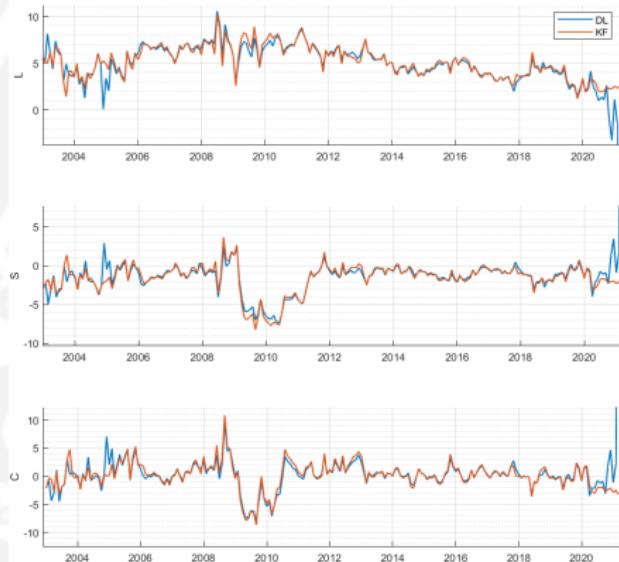


Figura: Estados latentes estimados utilizando Filtro de Kalman lineal junto con método de Diebold-Li. Parámetros $\lambda = 0.0049$ y 0.0055 respectivamente.

Unscented Kalman Filter

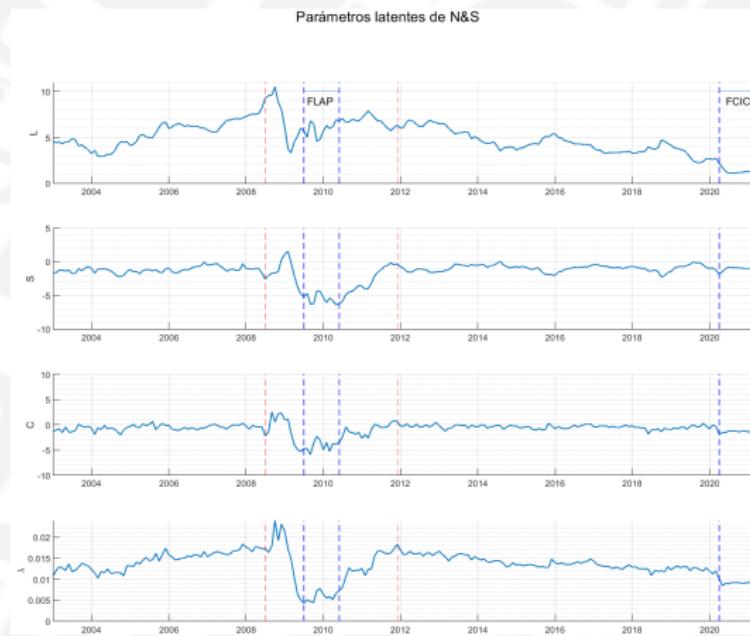
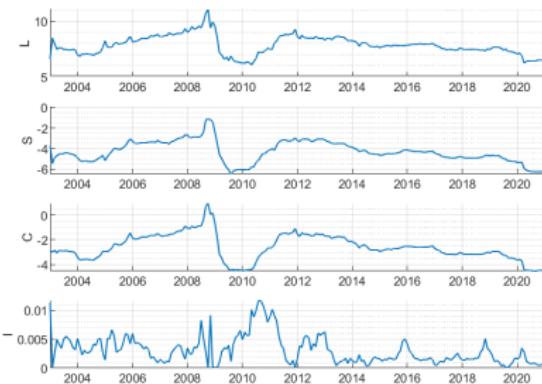


Figura: Estados latentes estimados utilizando Unscented Kalman Filter, Banco A.

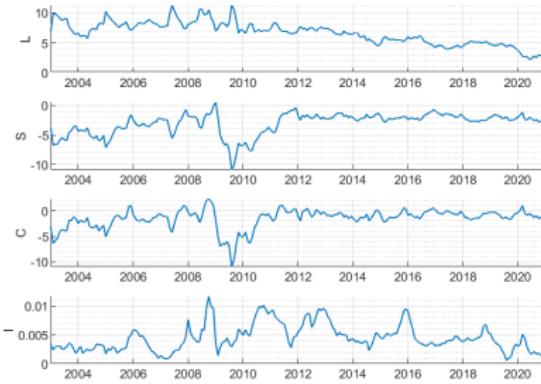
Para los demás bancos...

UKF estimado



(a) Banco B

UKF estimado



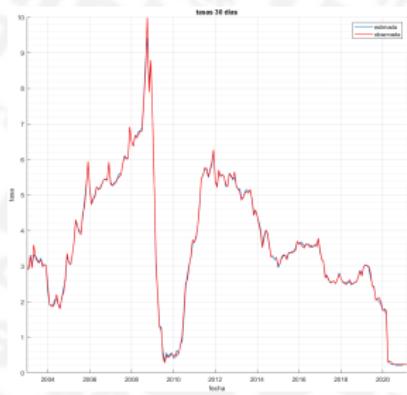
(b) Banco C

Comparación de métrica de performance

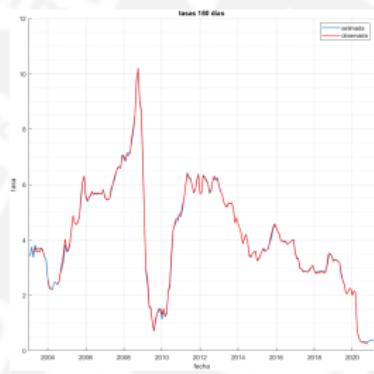
Utilizando el MSE para las tasas observadas a plazos 30, 60, 90, 180 y 360 como comparación:

Banco	DL	DLV	KF	UKF
A	0,0067	0,0057	0,0060	0,0052
B	0,0127	0,0114	0,0119	0,0179
C	0,0069	0,0064	0,0052	0,0061

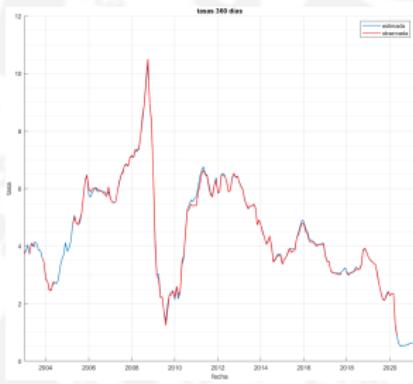
Traducido en tasas



(c)



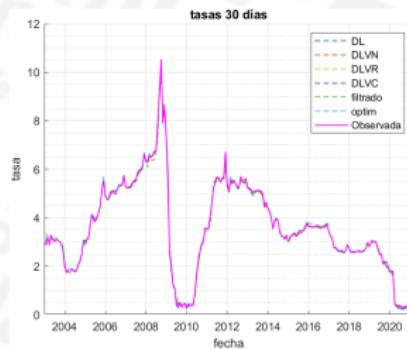
(d)



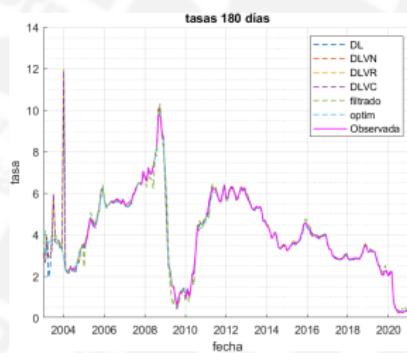
(e)

Figura: Tasas estimadas vs observadas a 30, 180 y 360 días

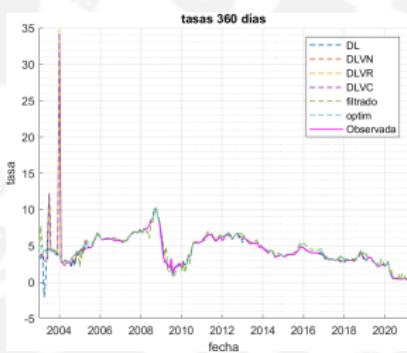
Tasas observadas y estimadas



(a)



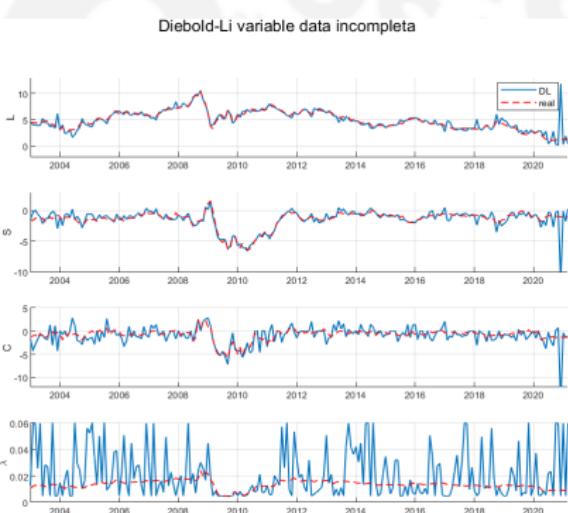
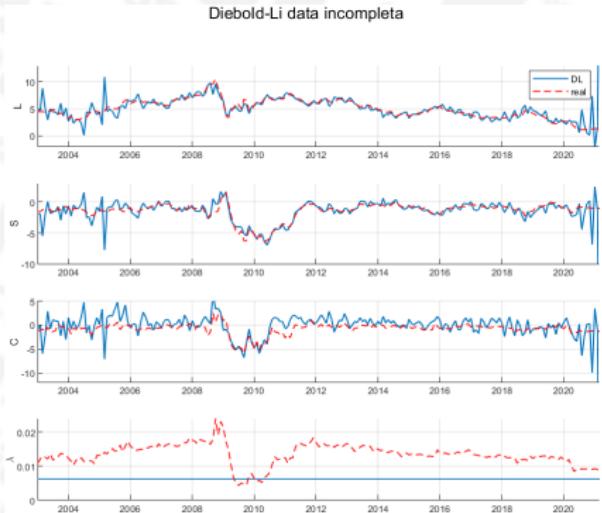
(b)



(c)

Figura: Tasas estimadas por distintos métodos vs observadas a 30, 180 y 360 días para banco B

Simulación: datos incompletos métodos de Diebold-Li



(a) Método de Diebold-Li para estructura de datos incompleta

(b) Método de Diebold-Li con λ variable para estructura de datos incompleta

Simulación: datos incompletos método de UKF

UKF optimizado data incompleta

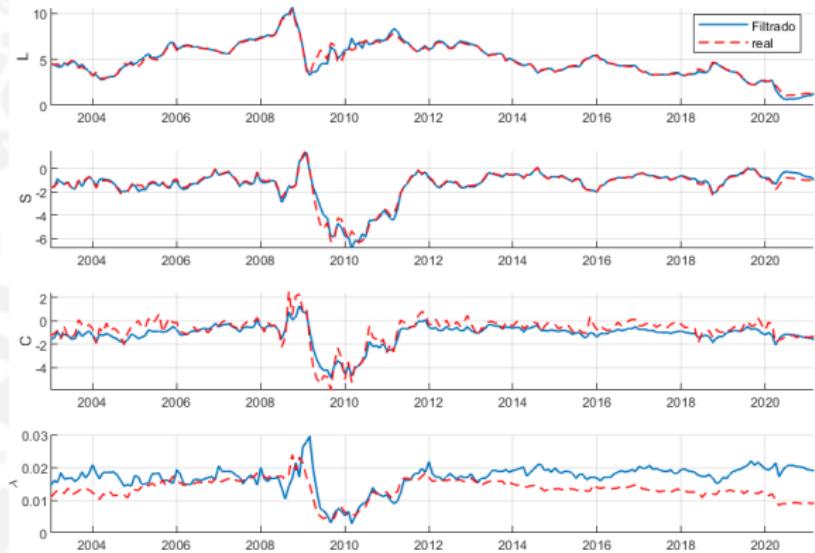


Figura: Estimación UKF para estructura de datos incompleta.

Incluyendo variables exógenas al modelo

Variables exógenas

Cambiando el modelo incluyendo variables exógenas:

$$\begin{array}{ll} \text{State ec.} & x_t = \mu + Ax_{t-1} + Uu_{t-1} + B\varepsilon_t \quad , \varepsilon_t \sim N(0, Id) \\ \text{Measurement ec.} & y_t = \Gamma_t(x_t) + \sigma\nu_t \quad , \nu_t \sim N_t(0, Id) \end{array}$$

Se incluyeron las siguientes variables:

- *Tpm*
- *Inflación*: Expectativa 1 mes
- *Imacec*: Log dif. des.
- *Roa*
- *Holgura*
- *VIX*
- *FLAP*
- *FCIC*

Modelo VARX

$$x_t = \mu + Ax_{t-1} + Bu_{t-1} + \varepsilon_t$$

Constantes	Coef	se	Pvalue
L	0,6523	0,3173	0,0398
S	-0,0671	0,3061	0,8266
C	-0,8170	0,5227	0,1180
λ	0,0113	0,0003	0,0000

(a) coeficiente de la media

Coef	L	S	C	λ
L	0,8101	-0,1144	-0,0893	0,0000
S	-0,2841	0,4084	0,2188	0,0000
C	0,2114	0,1635	0,6492	0,0000
λ	0,0005	0,0001	0,0015	-0,0297
SE	L	S	C	λ
L	0,1298	0,1314	0,0428	0,0000
S	0,1252	0,1268	0,0413	0,0000
C	0,2137	0,2164	0,0705	0,0000
λ	0,0001	0,0001	0,0000	0,0157
Pvalue	L	S	C	λ
L	0,0000	0,3839	0,0370	NaN
S	0,0232	0,0013	0,0000	NaN
C	0,3227	0,4501	0,0000	NaN
λ	0,0001	0,4989	0,0000	0,0578

(b) coeficientes de matriz A

Figura: Coeficientes modelo VARX estimados para banco A.

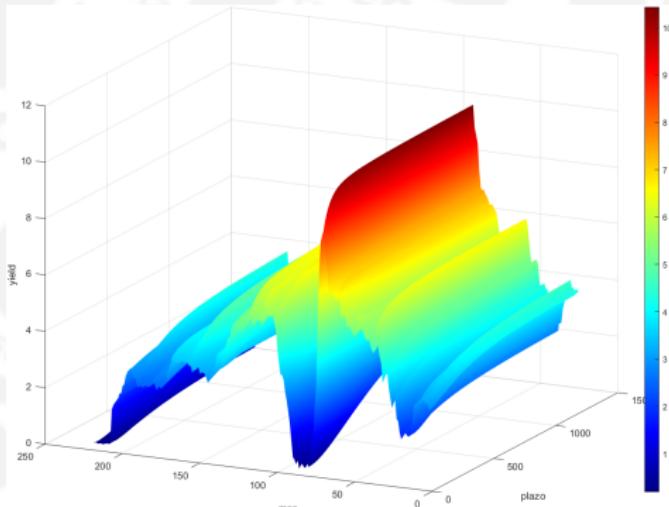
Modelo VARX

Coef	TPM	lnf_e1	VIX	Dln_imc_des	flap	fcic	ROA	Holgura
L	0,1101	0,4252	-0,0098	0,0196	-0,0523	0,1041	-0,0193	-0,0959
S	0,3692	-0,1792	-0,0102	0,0017	-0,3676	0,3954	-0,2065	0,0391
C	-0,0609	0,4336	-0,0153	0,0043	-0,6153	0,4377	-0,0488	0,0661
λ	0,0007	0,0004	0,0000	0,0000	-0,0005	0,0008	-0,0003	-0,0001
SE	TPM	lnf_e1	VIX	Dln_imc_des	flap	fcic	ROA	Holgura
L	0,1274	0,1360	0,0041	0,0114	0,1932	0,1907	0,1453	0,0381
S	0,1229	0,1312	0,0039	0,0110	0,1863	0,1840	0,1402	0,0367
C	0,2098	0,2241	0,0067	0,0187	0,3182	0,3142	0,2394	0,0627
λ	0,0001	0,0001	0,0000	0,0000	0,0002	0,0002	0,0001	0,0000
Pvalue	TPM	lnf_e1	VIX	Dln_imc_des	flap	fcic	ROA	Holgura
L	0,3874	0,0018	0,0158	0,0841	0,7866	0,5851	0,8944	0,0117
S	0,0027	0,1722	0,0093	0,8752	0,0485	0,0317	0,1408	0,2869
C	0,7715	0,0529	0,0225	0,8196	0,0531	0,1636	0,8386	0,2919
λ	0,0000	0,0014	0,0000	0,0084	0,0057	0,0000	0,0370	0,1380

Figura: Coeficientes estimado para matriz B en modelo VARX para banco A.

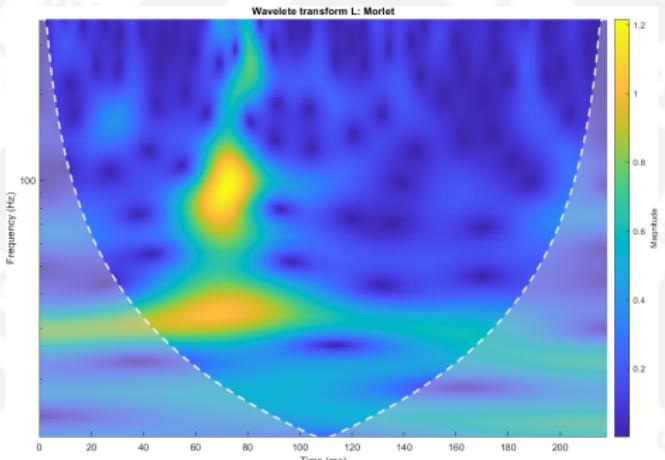
Conclusiones

- Estructura dinámica de parámetros latentes
- Método robusto ante falta de observaciones
- Inclusión de variables exógenas en estructura dinámica
- Capacidad de simulación y pronóstico



Extensiones

- Estimación máximo verosímil de parámetros
- Filtros de partículas
- Análisis espectral de estados latentes:
Wavelets, NeuralNetwork entre otros.



Proyecto de tweets y de curvas de tasas

31 de marzo de 2021