Aplicación de Cadenas de Markov en el Mercado de Divisas

Grupo: Los Analíticos

Nicolás Joel Cáceres (273195)

Isabella Endara Chitiva (288852)

Juan David Lasso (292677)

Nestor Andres Tabares (287880)

Universidad de la Sabana

6080: Procesos Estocásticos

Esteban A. López

20 de agosto de 2024

Índice

Introducción

Los mercados son un medio en donde inversores y compradores pueden buscar oportunidades para hacer crecer su capital por medio de acciones, divisas, bonos u otros bienes económicos. Un mercado como el de divisas puede dar una noción acerca del estado económico de una sociedad respecto al mundo (Aronsson & Folkesson, 2023).

Una particularidad que es común entre estos mercados es la constante fluctuación de sus precios y su comportamiento impredecible; en otras palabras, son mercados volátiles con bastantes fuentes de incertidumbre. Por esto mismo, surge el interés de intentar predecir el comportamiento de estos mercados, puesto que, si se logra acertar pronósticos de forma confiable, eso podría representar enormes oportunidades de ganancias a partir de inversiones en estos mercados (Aronsson & Folkesson, 2023). Sin embargo, según teorías económicas, los mercados son eficientes son aquellos cuyos precios son impredecibles, y por tanto, la ganancia de un inversor no puede ser constantemente mayor al promedio de ganancia del mercado, porque la acción misma del comprador influye en el mercado (Svoboda & Říhová, 2021). Históricamente se ha buscado implementar análisis técnicos sobre mercados de todo tipo alrededor del mundo, no obstante, los resultados obtenidos hasta la actualidad se han encargado de demostrar la dificultad y poca eficacia de los mismos, dando mayor consenso a las hipótesis de mercados eficientes (Svoboda & Říhová, 2021).

El interés por comprender a fondo los mercados financieros, se presenta tanto desde una perspectiva teórica como empírica. Este estudio se centra en el mercado de divisas, analizando el comportamiento de las tasas de cambio del Yen Japonés, la Libra Esterlina, el Franco Suizo y el Peso Mexicano frente al Peso Colombiano durante los últimos cuatro años. El objetivo principal es explorar la dinámica de este mercado y evaluar si los movimientos de las tasas de cambio siguen un proceso Markoviano; es decir, si el estado futuro de una variable depende únicamente de su estado presente y no de estados pasados. Al identificar las propiedades estadísticas de estos procesos, se busca desarrollar modelos probabilísticos para pronosticar los estados futuros de las divisas extranjeras.

Marco teórico

Para el análisis del mercado de divisas se ha de hacer uso de las cadenas de Markov y procesos Markovianos. Estos modelos funcionan bajo el supuesto de que la historia de una variable aleatoria no influye en el estado siguiente, sino que simplemente depende del estado actual. Esto se puede describir la siguiente manera (Aronsson & Folkesson, 2023):

$$P(X_{t+1} = j | X_t = i) = P(X_{t+1} = j | X_t = i_n, X_{t-1} = i_{n-1}, ..., X_0 = i_0) \quad (1)$$

Esto permite definir el estado de la variable aleatoria X en el estado $t+1$ a partir únicamente del estado t. Asimismo, existe la propiedad de homogeneidad, la cual sugiere que no importa el periodo en que se comparen los estados, la probabilidades de transición entre estados son iguales. Por ejemplo, es lo mismo la transición del estado t a $t+1$ que del estado $t+100$ al $t+101$. Las probabilidades que se tiene para saltar de un estado a otro van a generar una matriz cuadrada de tamaño (E x E), indicando E la cantidad de estados posibles, que en este ejemplo solo tendrá 3 estados posibles.

$$P_{ExE} = \begin{pmatrix} E1-E1 & E1-E2 & E1-E3 \\ E2-E1 & E2-E2 & E2-E3 \\ E3-E1 & E3-E2 & E3-E3 \end{pmatrix}$$

La matriz de transición P_{ij} muestra los valores de probabilidad de transición entre todos los estados posibles.

Extracción de datos

Los datos que se manejaron para este estudio fueron tomados de la página oficial del Banco de la República de Colombia en la sección de Monedas disponibles (Banrep, 2024). Allí se tiene a disposición una base de datos la cuál se puede filtrar por rango de fecha y por una o varias monedas en simultáneo. Desde allí, se seleccionaron las cuatro divisas, las cuales son: Yen Japonés (JPY), la Libra Esterlina (GBP), el Franco Suizo (CHF) y el Peso

Mexicano (MXN). Estas monedas se escogieron por diversas razones y criterios, entre los cuales se tuvo en cuenta que hubiese al menos una moneda por continente.

Por ejemplo, para elegir al JPY se observó que se considera una moneda de refugio para los inversores. Los factores que influyen en esta percepción general son, la tasa de ahorro, la base industrial sólida y la baja inflación de Japón (FasterCapital, 2024). Para la GBP se investigó que esta divisa es la cuarta moneda más negociada del mundo, luego del Dólar, Euro y Yen; adicionalmente es la tercera más usada para pagos mundiales, al igual que una de las divisas más representativas en la reservas mundiales (MarketScreener, 2024).

En cuanto al peso Mexicano, se tiene en cuenta que es uno de los mercado emergentes con mayor fuerza en el continente, agregando el hecho de que posee una frontera extensa con los Estados Unidos y grandes reservas de petróleo, pues todo lo anterior son elementos que consideran los inversionistas (Imparcial, 2024). Por último, el franco suizo es reconocido como un salvavidas en tiempos de crisis, pues se invierte en esta divisa por su reputación de estabilidad, puesto que las políticas monetarias suelen priorizar la estabilidad monetaria a los intereses de los exportadores, adicionalmente suiza es reconocido como un país neutral ante muchos conflictos y por esto mismo suele tener un entorno menos volátil que otras divisas (Eugster, 2023).

Importación de datos

Al haber seleccionado estas monedas, se exportó de la página del Banco de la República un archivo en formato .csv con toda la información relacionada a esas monedas en el periodo del 6 de Agosto del 2020 al 6 de Agosto del 2024. Este archivo fue subido a un repositorio de GitHub con la finalidad de que la base de datos fuera accesible mediante el siguiente enlace. Este archivo contiene un total de 52.632 registros y 8 columnas con las variables que se muestran en la Cuadro 1.

Limpieza de datos

La limpieza de datos es un componente esencial en el modelado de la matriz de transición para cada divisa. En primer lugar, se revisa el tipo de dato de cada columna en

Cuadro 1

información disponible en la base de datos (Banrep, 2024).

Columna	Valores
Fecha (aaaa/mm/dd)	2020-08-06 a 2024-08-06 (díario)
Moneda	Franco suizo, Yen japonés, Libra Esterlina, Peso Mexicano
Continente	Europa, América, Asia del Pacifico
Tipo de Cambio	USD por moneda, Moneda por USD, COP por moneda
Tipo de Tasa	Venta, Media, Compra
Tasa Cambio	'1,098049', ... , '28,961099'
Id Moneda	CHF, JPY, GBP, MXN

la base de datos y si es necesario, se realiza un cambio de tipo de dato. Paralelamente, se verifica la presencia de datos nulos; en caso de encontrarlos, se aplican técnicas avanzadas de preprocesamiento de datos para completar estos valores faltantes. El Cuadro a continuación presentará la información de la base de datos. Este proceso garantiza la precisión y la eficacia del modelo resultante.

Cuadro 2

Información de la base de datos

columna	Non-null count	Dtype
Fecha	52632 non-null	Object
Año	52632 non-null	Int64
Moneda	52632 non-null	Object
Continente	52632 non-null	Object
Tipo de cambio	52632 non-null	Object
Tipo de tasa	52632 non-null	Object
Id Moneda	52632 non-null	Object

Con lo que se observa en el Cuadro , no existen datos nulos o faltantes en esta base de datos. Además, todos los datos de las columnas son de tipo 'Object', a excepción de la columna 'año', que es de tipo 'int64'. En el Cuadro siguiente, se realizará una verificación para confirmar la ausencia de datos nulos en la base de datos. Esta inspección es crucial para garantizar la integridad y la calidad de los datos antes de proceder con cualquier análisis o modelado.

Cuadro 3

Cantidad de datos nulos en cada columna

Columna	Cantidad Datos nulos
Fecha	0
Año	0
Moneda	0
Continente	0
Tipo de cambio	0
Tipo de tasa	0
Id Moneda	0

Como se puede observar en el Cuadro anterior, no existen datos nulos en toda la base de datos, lo que elimina la necesidad de aplicar técnicas de preprocesamiento para tratar con valores faltantes. No obstante, en esta base de datos de divisas, se observa que existen datos correspondientes a todos los días de la semana, incluyendo los días domingo. Sin embargo, estos datos son repetitivos, ya que el mercado de divisas cierra el día anterior; así que, lo que hace esta base de datos es sostener la tasa de cambio del día sábado. Esto podría generar sesgos debido a estos datos repetidos, por lo tanto, en el siguiente código, se procederá a eliminar todos los datos correspondientes a los domingo. Esta acción mejora la precisión de los datos y evita cualquier distorsión en el análisis posterior.

```
Sundays = df['Fecha'].unique().tolist()[3::7]
```

```
keep = set(df['Fecha'].unique().tolist())
dates = keep.difference(Sundays)
dates =list(dates)
dates.sort()
df = df[df['Fecha'].isin(dates)]
```

La lógica de este código de programación consiste en crear una lista de fechas únicas y seleccionar los domingos (cada séptimo día a partir del cuarto elemento). Luego, crear un conjunto con todas las fechas únicas y eliminar los domingos de este conjunto. Las fechas restantes se convierten en una lista ordenada y se filtra el *DataFrame* original para mantener solo las filas cuyas fechas están en esta lista.

Al hacer esta limpieza, la base de datos queda en el formato como se muestra en la siguiente Cuadro:

Cuadro 4

Base de datos

Fecha	Nombre_Moneda	Tipo_tasa	TRM
2020-08-06	Franco suizo	Venta	1,098049
2020-08-06	Franco suizo	Venta	0,9111
2020-08-06	Franco suizo	Venta	4146,20621
2020-08-06	Libra esterlina	Venta	1,3145
2020-08-06	Libra esterlina	Venta	0,76097

Por razones de espacio y comodidad, se decidió no mostrar las columnas ‘año’, ‘continente’, ‘tipo de cambio’ e ‘id de la moneda’. Sin embargo, en este Cuadro se puede observar que la TRM es de tipo ‘Object’, como se mencionó anteriormente. Además, se utiliza una coma (",") para separar los decimales en lugar de un punto ("."), lo cual podría causar un error ya que se esté trabajando con Python, ya que este lenguaje interpreta los números decimales con un punto. Para el modelado, se trabajará con el tipo de tasa ‘media’, como se evidencia en el código siguiente. Esta decisión garantiza la coherencia y la precisión de los datos utilizados en el análisis.

```
COP_Med_Data = df.query("Cambio == 'Pesos colombianos por cada moneda'
    and Tipo_tasa == 'Media'")
COP_Med_Data['TRM'] = COP_Med_Data['TRM'].str.replace(',','.').astype(
    float)
COP_Med_Data
```

Debido a la selección de la información considerada relevante, se debe crear una nueva base de datos. Esta base de datos contendrá exclusivamente la TRM (Tasa Representativa del Mercado) para cada moneda y cada fecha. Es importante asegurar que esta información esté en formato 'float' para facilitar los cálculos posteriores. Como se puede observar en el siguiente código, se crea una columna para cada moneda, y cada fila corresponderá a una fecha específica de registro. Este enfoque estructurado y organizado facilita el análisis y el modelado para cada una de las divisas.

```
data_original = pd.DataFrame(columns=df['Id_Moneda'].unique().tolist(),
    index=df['Fecha'].unique())
for i in data_original.index:
    row = list(COP_Med_Data.query("Fecha == @i")['TRM'])
    data_original.loc[i] = row
data_original = data_original.astype(float)
```

Adicionalmente, se ha decidido reajustar los índices de esta nueva base de datos, que están en formato de fecha, para facilitar su manejo en Python. Se ha optado por reiniciar los índices para que comiencen desde 0 y continúen hasta el valor de la última fila. A continuación, se presenta la base de datos después de la limpieza respectiva. Este ajuste en los índices mejora la accesibilidad y la manipulación de los datos durante el análisis y el modelado.

Estructura de datos

Analizar la estructura de los datos es crucial para entender su comportamiento, especialmente cuando se trata de modelar cadenas de Markov y verificar la homogeneidad.

Cuadro 5

TRM

CHF	**GBP**	**MXN**	**JPY**
4145.295860	4962.731079	168.702869	35.782510
4130.239940	4919.796309	167.470169	35.601540
4130.239940	4919.796309	167.470169	35.601540
4122.110440	4934.121060	167.425549	35.638570
4115.286790	4933.332870	168.414890	35.412759

Es posible que algunas divisas presenten datos atípicos o sesgos que podrían afectar el modelado. Por lo tanto, es esencial identificar y tratar estos problemas antes de proceder con el modelado.

Una forma efectiva de visualizar la distribución de los datos y detectar posibles anomalías es a través de los histogramas. Los histogramas nos permiten ver la frecuencia de diferentes rangos de valores, lo que puede revelar datos atípicos o sesgos en los datos.

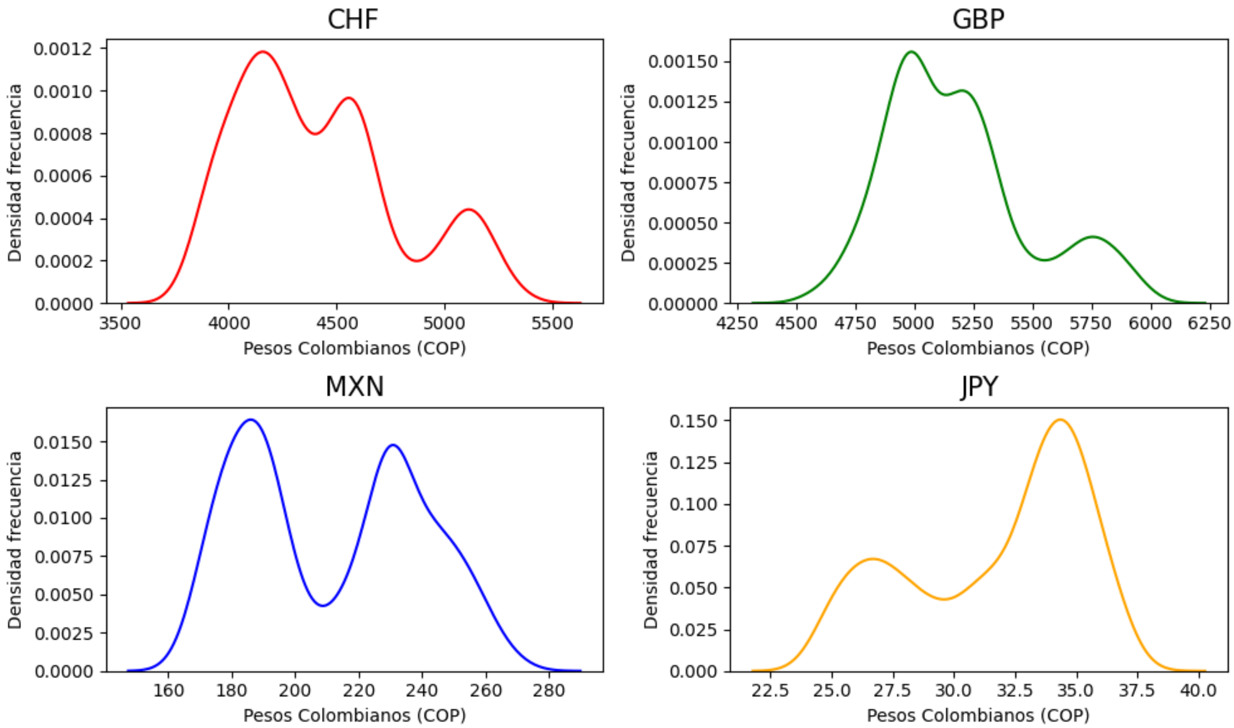
A continuación, se mostrarán los gráficos de densidad de frecuencia de las diferentes divisas. Estos gráficos proporcionarán una visión clara de la distribución de los datos y ayudarán a identificar cualquier irregularidad que necesite ser abordada antes del modelado.

- **COP frente al CHF**: El histograma muestra tres picos, lo que indica que ciertos valores son más frecuentes, sugiriendo una distribución no uniforme. Esto podría afectar las cadenas de Markov, ya que no todos los estados tienen la misma probabilidad de ocurrencia. Es crucial evaluar si estas frecuencias se mantienen constantes en el tiempo o si hay cambios en la ocurrencia de ciertos tipos de cambio.

- **GBP en COP**: El histograma muestra un pico alrededor de los 5000 COP, lo que sugiere que la tasa de cambio ha sido más frecuente en este rango. Esto podría indicar que las tasas futuras podrían depender solo de la tasa actual, un principio

Figura 1

Gráficos de densidad de frecuencia por divisa



clave de las cadenas de Markov de primer orden. Esto permite predecir futuras tasas basándose en la tasa actual.

- **MXM en COP**: El histograma muestra dos picos prominentes, lo que sugiere que hay dos valores comunes en los que la tasa de cambio se estabiliza o fluctúa. Esto indica que podría haber dos estados más prevalentes en el sistema analizado, lo que sugiere que un modelo de cadena de Markov de primer orden podría tener dos estados con mayores probabilidades. Al construir una matriz de transición basada en estos datos, es importante considerar estos dos picos, ya que probablemente representan estados significativos dentro de la matriz.

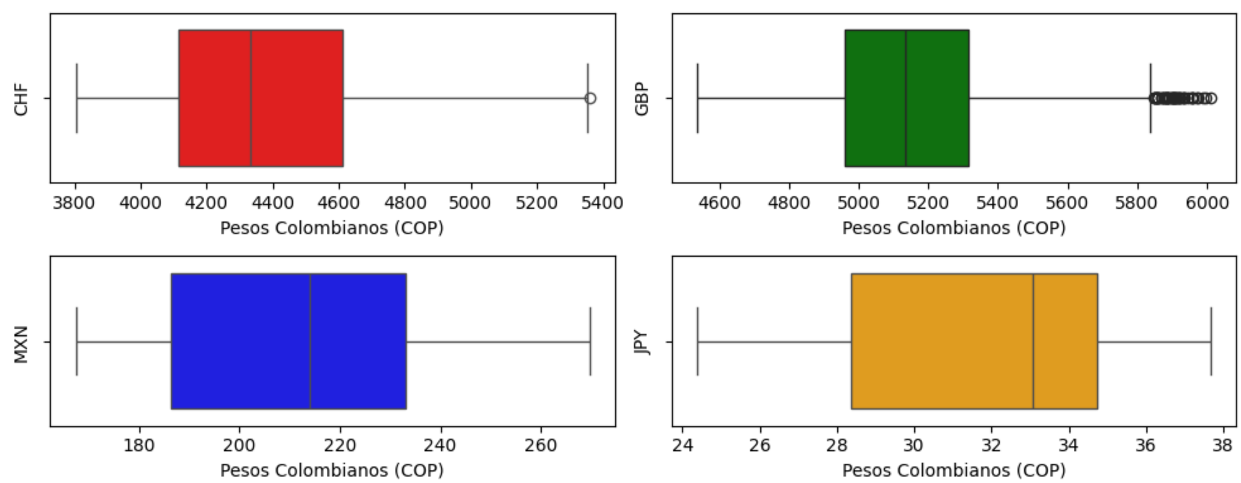
- **JPY en COP**: El histograma muestra una concentración de valores alrededor de un rango específico, lo que indica que la mayoría de las tasas de cambio se encuentran en ese intervalo.

Los histogramas de las tasas de cambio muestran una variedad de distribuciones, lo que puede tener implicaciones significativas para el modelado de las cadenas de Markov. Es importante tener en cuenta estas distribuciones al construir modelos de cadenas de Markov, ya que pueden influir en la probabilidad de transición entre los estados.

Para evidenciar los posibles valores atípicos por medio del criterio intercuartílico, se realizó un *boxplot* para cada divisa, los cuales se presentan en la siguiente figura.

Figura 2

Boxplots de precios para cada divisa



Los gráficos de *boxplot* presentados muestran la variabilidad de la Tasa Representativa del Mercado (TRM) para cuatro monedas diferentes en relación con el peso colombiano (COP). El franco suizo (CHF) muestra un rango de valores que oscila aproximadamente entre 3800 y 5400 COP y contiene un dato atípico. Este amplio rango sugiere una alta variabilidad en las tasas de cambio, lo que podría impactar las pruebas de homogeneidad si se consideran diferentes períodos de tiempo. Además, esta variabilidad podría causar fluctuaciones significativas en las probabilidades de transición en las cadenas de Markov. No obstante, con la finalidad de apegarse a la realidad y evitar omitir más información estos datos se tendrán en cuenta. Vale la pena aclarar que a pesar que sean datos que se alejen un poco de la tendencia normal, no significa que su comportamiento sea errático, pues algún evento puntual potencia el efecto de subida de la moneda pero en vez de sacrificar datos más normalmente distribuidos, estamos ganando información con estas observaciones.

La libra esterlina (GBP) presenta valores que oscilan aproximadamente entre $4.600 y $6.000 COP y por arriba del limite superior se evidencia múltiples datos atípicos que podrían darle sesgos al modelo. Al igual que con el CHF, esta variabilidad podría afectar las suposiciones sobre las transiciones de estado en las cadenas de Markov de primer orden. Además, un gran rango en las tasas de cambio podría sugerir no homogeneidad. No obstante, para evitar remover información del modelo se tendrán en cuenta estos valores para todo el proceso.

Por otro lado, tanto el peso mexicano (MXN) como el yen japonés (JPY) muestran rangos más estrechos y no presentan datos atípicos. Esto sugiere una mayor estabilidad en estas tasas de cambio a lo largo del tiempo, lo que podría resultar en un mejor rendimiento en las cadenas de Markov y las pruebas de homogeneidad.

En conclusión, mientras que el CHF y la GBP muestran una considerable volatilidad que podría desafiar las suposiciones de homogeneidad y afectar significativamente las matrices de probabilidad de transición en las cadenas de Markov, el MXN y el JPY parecen ser más estables, ofreciendo potencialmente un mejor rendimiento tanto para las pruebas de homogeneidad como para la modelización de las cadenas de Markov.

Estructuración del modelo

El modelo de Markov que se va a realizar se lleva a cabo teniendo en cuenta supuestos sobre los datos y el comportamiento de los datos. Lo primero que se supone es que los días domingos no aportan información a la data, y por ende se removieron de la base de datos, asimismo, se considera que los días festivos no afectan los datos y se mantuvieron estos días dentro de la data. Considerando adicionalmente que la frecuencia temporal más útil en contextos reales y en la literatura es la variación diaria, se decidió mantener la frecuencia original de los datos. El alcance de este modelo es generar una matriz de transición que cumpla la propiedad markoviana y de homogeneidad, la cual permite conocer que tan probable es que una divisa, dado su estado actual, se mantenga o cambie a otro estado y que esa probabilidad sea constante a lo largo del tiempo, que en este caso se estudia día a día en las divisas seleccionadas.

Definición de Estados de Mercado

El primer paso es definir los posibles estados del sistema. En el contexto del mercado de divisas, estos estados pueden representar diferentes niveles de precios o tendencias del mercado, como alzas, bajas o estabilidad. Es crucial definir los estados con precisión, ya que deben capturar las dinámicas más importantes del mercado para que la cadena de Markov sea efectiva.

Este informe presenta un enfoque para definir y clasificar los estados de las divisas, utilizando la desviación estándar móvil y cambios porcentuales en las tasas de cambio. El objetivo es demostrar cómo estas herramientas pueden segmentar las condiciones del mercado en diferentes estados, proporcionando una visión más clara y detallada de la dinámica del mercado.

La clasificación de estados se basa en dos principales indicadores: el cambio porcentual diario de la divisa y la desviación estándar móvil calculada sobre un período de tiempo determinado. Esta metodología permite capturar no solo la dirección del cambio en la tasa de cambio, sino también, la magnitud de dicho cambio en relación con la volatilidad reciente.

La desviación estándar móvil es una medida de volatilidad que se recalcula constantemente sobre un número fijo de periodos. Esta métrica es crucial para evaluar la magnitud de los movimientos en la tasa de cambio y compararlos con su comportamiento reciente. Como se discute en el documento de Milan y Pavla, "La desviación estándar móvil es fundamental para identificar períodos de alta volatilidad, permitiendo diferenciar entre movimientos significativos y fluctuaciones normales"(Svoboda & Říhová, 2021).

Los estados del mercado se definen combinando la dirección y la magnitud de los cambios porcentuales en la tasa de cambio:

- D2 (Disminución notable): Se asigna cuando la tasa de cambio disminuye significativamente, superando la volatilidad esperada según la desviación estándar móvil.
- D1 (Disminuyo): Se refiere a una disminución en la tasa de cambio que no excede la volatilidad estimada.
- I1 (Incremento): Representa un aumento moderado en la tasa de cambio, que está dentro de los límites de la desviación estándar móvil.
- I2 (Incremento notable): Se asigna a un aumento en la tasa de cambio que excede la volatilidad prevista.

Estos estados permiten no solo identificar si la tasa de cambio está subiendo o bajando, sino también si estos cambios son lo suficientemente grandes para ser considerados como un cambio significativo en el mercado. El enfoque descrito sugiere que "la combinación de dirección y magnitud en la clasificación de estados del mercado permite una segmentación más precisa de las condiciones del mercado, capturando tanto la tendencia como la intensidad de los movimientos"(Svoboda & Říhová, 2021).

El empleo de la desviación estándar móvil, junto con los cambios porcentuales diarios para definir los estados del mercado, son una herramienta valiosa para el análisis de divisas. Este enfoque avanzado no solo identifica la dirección del cambio en la tasa de cambio, sino que también captura la intensidad de estos cambios, permitiendo una segmentación más precisa y una mejor comprensión de la dinámica del mercado.

Para la construcción de la cadena de Markov de primer orden se necesita esta segmentación de datos y también crear tiempos pasados y futuros; a continuación se presenta un ejemplo con la divisas japonesa.

Cuadro 6

Estados de la divisa japonesa

x_t	x_{t+1}	x_{t+1}
I1	I2	I1
12	11	I1
I1	I1	I1
I1	I1	D1
I1	D1	D2
...	...	...

Construcción de la Cadena de Markov

Entre las metodologías más utilizadas se encuentra la cadena de Markov de primer orden, que permite modelar las transiciones entre estados basándose únicamente en el estado actual. Esta simplificación hace que el análisis y la predicción sean más manejables en sistemas complejos, como los mercados financieros (Laipaporn & Tongkumchum, 2021).

Frecuencias Observadas y Esperadas

Para determinar si el proceso sigue una cadena de Markov de primer orden, es necesario comparar las probabilidades observadas con las probabilidades esperadas. Antes de calcular las probabilidades, se deben obtener las frecuencias observadas y esperadas.

- **Frecuencias Observadas (O_{ij})**: Son las frecuencias reales entre estados y el tiempo t, obtenidas directamente de la base de datos de estados de mercado.
- **Frecuencias Esperadas (E_{ij})**: Se calculan bajo la hipótesis nula de que el proceso es markoviano. Estas se obtienen directamente de los estados en la base de datos, considerando únicamente X_{t+1} y X_{t+2}.

Es importante que al menos el 0.25 de las frecuencias esperadas sean mayores a 5. Si este requisito no se cumple, se deben agrupar o concatenar las filas de las frecuencias observadas y esperadas, hasta que se cumpla la condición.

Para la creación de estos tipos de frecuencia se puede utilizar el siguiente código:

```
freq_obs = pd.DataFrame(columns=("Anterior", "Actual"))
for i in Estados:
    for j in Estados:
        freq_obs.loc[len(freq_obs)] = [i,j]
for i in Estados +["Total"]:
    freq_obs[i] = 0
for i in freq_obs.index:
    for j in Estados:
        for k in Data.index:
            if freq_obs.loc[i].tolist()[:2]+[j] == Data.loc[k].tolist():
                freq_obs.loc[i, j] += 1
    freq_obs["Total"] = freq_obs[[Estados[0],Estados[1],Estados[2],
  Estados[3]]].sum(axis=1)
freq_obs
```

Para establecer la frecuencia esperada es exactamente la misma lógica de programación, sin embargo, se elimina la columna 'Anterior' y en la base de datos solo se escoge las dos últimas columnas. El resultado que se observa en las frecuencias son las siguientes:

Para la tabla de frecuencia esperada, es exactamente lo mismo sin la columna 'Anterior'.

Probabilidades Observadas y Esperadas

- **Probabilidades Observadas (PO_{ij})**: Estas son las probabilidades calculadas a partir de las frecuencias observadas. Reflejan las transiciones reales entre los estados y el tiempo X_t, X_{t+1} y X_{t+2}.

- **Probabilidades Esperadas (PE_{ij})**: Se calculan bajo la hipótesis de que el proceso sigue una cadena de Markov de primer orden, utilizando las frecuencias esperadas.

Cuadro 7

Tabla de Frecuencias Observadas

Anterior	Actual	D2	D1	I1	I2	Total
D2	D2	4	17	14	8	43
D2	D1	7	26	27	7	67
D2	I1	10	24	23	11	68
D2	I2	12	21	13	8	54
D1	D2	9	23	24	15	71
D1	D1	20	59	34	29	142
D1	I1	26	43	38	16	123
D1	I2	20	19	18	17	74
I1	D2	16	18	15	19	68
I1	D1	30	38	37	30	135
I1	I1	25	46	38	18	127
I1	I2	14	16	20	6	56
I2	D2	14	9	15	12	50
I2	D1	14	19	25	8	66
I2	I1	7	22	29	10	68
I2	I2	4	10	17	2	33

Para calcular las probabilidades observadas y esperadas, se puede utilizar la siguiente fórmula:

$$P_{ij} = \frac{N_{ij}}{N_i} \tag{2}$$

donde:

- N_{ij} es la frecuencia observada o esperada del estado i al estado j.
- N_i es el número total de frecuencias desde el estado i.

En la lógica de programación es exactamente igual a la formula dada anteriormente, que se puede aplicar en la probabilidad esperada y en la probabilidad observada como se evidencia a continuación:

```
prob_esp = freq_esp[Estados].div(freq_esp["Total"], axis=0)
print(prob_esp)
```

Y el resultado de esa operación debería parecerse a el siguiente Cuadro:

Cuadro 8

Tabla de Probabilidades Esperadas

	D2	D1	I1	I2
0	0.185345	0.288793	0.293103	0.232759
1	0.173171	0.346341	0.300000	0.180488
2	0.176166	0.349741	0.331606	0.142487
3	0.230415	0.304147	0.313364	0.152074
4	0.185345	0.288793	0.293103	0.232759
5	0.173171	0.346341	0.300000	0.180488
6	0.176166	0.349741	0.331606	0.142487
7	0.230415	0.304147	0.313364	0.152074
8	0.185345	0.288793	0.293103	0.232759
9	0.173171	0.346341	0.300000	0.180488
10	0.176166	0.349741	0.331606	0.142487
11	0.230415	0.304147	0.313364	0.152074
12	0.185345	0.288793	0.293103	0.232759
13	0.173171	0.346341	0.300000	0.180488
14	0.176166	0.349741	0.331606	0.142487
15	0.230415	0.304147	0.313364	0.152074

Prueba de Hipótesis de la Propiedad Markoviana

Las cadenas de Markov de primer orden (MC1) tienen la característica de que las probabilidades de transición de cada observación en la serie temporal dependen sólo del valor del miembro anterior de la serie temporal. Esta característica es conocida como la Propiedad de Markov, que se expresa de la siguiente manera(García, 2021):

$$P_{ij} = Pr(X_{t+1} = j|X_t = I) \tag{3}$$

Para validar si los datos pueden modelarse como una cadena de Markov de primer orden, se realiza una prueba de hipótesis. La hipótesis nula (H_0) sugiere que el proceso puede ser modelado como una cadena de Markov de primer orden, mientras que la hipótesis alternativa (H_1) indica que no es así.

La prueba chi-cuadrado se utiliza para evaluar si las diferencias entre las probabilidades observadas y las esperadas son estadísticamente significativas. Si las frecuencias observadas no difieren significativamente de las esperadas, se concluye que el proceso puede ser modelado como una cadena de Markov de primer orden.

El estadístico chi-cuadrado se calcula para medir la discrepancia entre las frecuencias observadas y las esperadas:

$$\chi^2 = \sum_{i,j} \frac{(PO_{ij} - PE_{ij})^2 \times N_i}{PE_{ij}} \tag{4}$$

Donde:

- PO_{ij} es la probabilidad observada de transiciones del estado i al estado j.
- PE_{ij} es la probabilidad esperada bajo la hipótesis nula.
- N_i es el número total de frecuencias observadas en el estado i.

Finalmente, se compara el estadístico chi-cuadrado calculado con un valor crítico, usando un nivel de significancia α de 0.05 y un grado de libertad adecuado. Si el estadístico chi-cuadrado calculado es mayor que el valor crítico, se rechaza la hipótesis nula, lo que indica que el proceso no sigue una cadena de Markov de primer orden. Si el estadístico chi-cuadrado es menor, no hay evidencia suficiente para rechazar la hipótesis nula, lo que sugiere que el proceso puede ser modelado como una cadena de Markov de primer orden.

Estimación de las Probabilidades de Transición

El siguiente paso es calcular las probabilidades de transición entre los estados. Este cálculo se basa en las frecuencias observadas de transiciones de un estado a otro en la base de datos. Se construye una matriz de transición, donde cada entrada P_{ij} representa la probabilidad de que el sistema pase del estado i al estado j.

La fórmula para calcular P_{ij} es:

$$P_{ij} = \frac{N_{ij}}{N_i} \tag{5}$$

Prueba de la Homogeneidad

La prueba de homogeneidad en el modelo estocástico se realizó mediante la comparación de distribuciones a lo largo de varios periodos para determinar si estas son iguales en todos los periodos evaluados. La prueba implica lo siguiente:

1. **Hipótesis Nula (Ho):** Se establece que las distribuciones de las poblaciones son iguales en todos los periodos.

2. **Estadístico Chi-cuadrado:** Se calcula un estadístico Chi-cuadrado para comparar las distribuciones observadas con las esperadas.

3. **Nivel de Significancia:** Se determina un valor crítico de Chi-cuadrado para un nivel de significancia del 5 %.

4. **Resultado:** Si el estadístico calculado es menor que el valor crítico, no se rechaza la hipótesis nula, indicando que las distribuciones son homogéneas; de lo contrario, se rechaza.

La verificación de esta prueba se llevó a cabo mediante un código en Python, que se detalla a continuación:

El código comienza extrayendo los estados finales de un proceso estocástico y preparando los datos para el análisis.

```
dh = final_states(currency= "JPY")
dh.insert(0, 'Periodo', None)
dh["Siguiente"] = dh["Estado"].shift(-1)
dh.columns = ["Periodo", "Actual", "Siguiente"]
dh.index = dates[6:]
```

Aquí, 'dh' es un *DataFrame*, que contiene los estados actuales y los estados siguientes (es decir, el estado en el próximo periodo).

Posteriormente, se asignan los periodos de tiempo a cada observación:

```
    fechas = dh.index.to_list()
    k= 0
    periodos = []
    for i in ["2020","2021","2022","2023","2024"]:
        f = [ano[:4] for ano in fechas if i in ano]
        f = f[:len(f)-1] +["-"]
        periodos.extend(f)
    dh['Periodo'] = periodos
    dh.reset_index(drop= True, inplace= True)
```

Luego, se calculan las frecuencias observadas (`freq_obs`) y las esperadas (`freq_esp`) para cada estado en cada periodo:

```
for i in freq_obs.index:
    for j in Estados:
        for k in dh.index:
            if dh.loc[k].tolist() == freq_obs.loc[i].tolist()[:2]+[j]:
                freq_obs.loc[i, j] += 1
```

Este bloque cuenta el número de veces que cada estado 'Actual' transita al estado

'Siguiente' dentro de cada periodo.

Posteriormente, se calculan las probabilidades observadas (`prob_obs`) y esperadas (`prob_esp`) dividiendo las frecuencias observadas y esperadas por los totales correspondientes.

El siguiente paso es calcular el estadístico Chi-cuadrado utilizando la fórmula:

$$\chi^2 = \sum \frac{(O_i - E_i)^2}{E_i} \times \text{Total Observado} \tag{6}$$

El valor de χ^2 se compara con el valor crítico de la distribución Chi-cuadrado con los grados de libertad apropiados:

```
resultado = ((prob_obs - prob_esp) ** 2 / prob_esp).sum(axis=1) *
    freq_obs["Total"]
chi_2 = resultado.sum()
deg_free = (prob_esp.shape[1]-1)*(prob_esp.shape[0]-1)
alpha = 0.05
chi2_inv = stats.chi2.ppf(1-alpha, deg_free)
```

Aquí, '`deg_free`' representa los grados de libertad, y '`chi2_inv`' es el valor crítico de la distribución Chi-cuadrado al nivel de significancia $\alpha = 0{,}05$.

Finalmente, se compara el valor calculado de χ^2 con el valor crítico:

```
if chi_2 < chi2_inv:
    Homogenea = "No se rechaza Ho"
else:
    Homogenea = "Se rechaza Ho"
```

Si χ^2 es menor que el valor crítico, no se rechaza la hipótesis nula (`Ho`), indicando que el modelo es homogéneo.

Además de la decisión sobre la homogeneidad, el código también calcula y presenta la matriz de transición basada en las frecuencias esperadas.

Cadenas de orden superior

En este apartado, se abordarán las cadenas de Markov de orden dos, que siguen la siguiente fórmula:

$$P(X_{t+2} = i \mid X_{t+1} = j, X_t = k) \tag{7}$$

Para la creación de la prueba de hipótesis de orden dos, es necesario introducir un estado adicional, X_{t+3}, para comprobar si la cadena de orden dos explica adecuadamente la matriz de transición o si es necesario considerar un orden superior.

Para construir la prueba de hipótesis, es necesario calcular las frecuencias observadas y esperadas. Las frecuencias observadas contienen todas las combinaciones posibles entre los tiempos X_t, X_{t+1}, X_{t+2} y X_{t+3}. Al tener todas las combinaciones de todos los estados en los diferentes tiempos, se revisa en la base de datos para determinar si estos datos coinciden exactamente con las frecuencias observadas.

Simultáneamente, se debe calcular la frecuencia esperada, que es similar a la tabla de frecuencias observadas, pero eliminando la columna X_t, y luego se verifica en la base de datos cuáles combinaciones coinciden.

En este proceso, el 25 % de los datos debe tener frecuencias mayores a 5. Si no se cumple este requisito, es necesario combinar filas de las tablas de frecuencias observadas y esperadas hasta que se cumpla el criterio.

Una vez calculadas las frecuencias observadas y esperadas con sus respectivos totales por filas, se calculan las probabilidades observadas y esperadas. Esta fórmula ya se mencionó anteriormente en el documento.

Una vez obtenidas las probabilidades, se comparan utilizando una prueba de chi-cuadrado para determinar si son significativamente diferentes o no. Si se rechaza la hipótesis nula, significa que la cadena de Markov no es de orden dos; si no se rechaza, significa que no hay suficiente evidencia para asegurar que la cadena de Markov es de orden dos.

Para la creación de la matriz de transición, se utilizan duplas, en las cuales se deben tener en cuenta los dos estados anteriores para calcular el siguiente. En este caso, para las divisas, la matriz de transición resultante fue de 16×16, lo cual es un tamaño considerablemente grande para incluir en este documento. Sin embargo, se pudo concluir que todas las divisas no rechazaron la prueba de hipótesis, lo que sugiere que todas ellas siguen una cadena de Markov de orden 2.

Resultados

Al aplicar la metodología mencionada anteriormente, sobre la construcción de la cadena de Markov de primer orden y las pruebas de hipótesis, se obtuvo para las diferentes divisas los siguientes resultados:

JPY

En el caso de la divisa japonesa, se concluyó que tanto en la prueba de hipótesis Markoviana de primer orden como en la de homogeneidad, no se rechazan las hipótesis nulas. Esto implica, que no existe suficiente evidencia para afirmar, que el modelo de primer orden no es adecuado para describir el proceso en estudio, ni tampoco, que las distribuciones de las diferentes muestras son distintas.

Por lo tanto, se podría interpretar que para el Yen, en la matriz de transición, solo se necesita el estado actual para predecir la probabilidad del día siguiente o incluso del año siguiente, lo cual cumple con la regla Markoviana de primer orden. Además, el hecho de que no se rechace la hipótesis de homogeneidad podría inferir que, independientemente del año en que se encuentre la matriz de transición, sus distribuciones de probabilidades son similares.

A continuación se mostrará la matriz de transición.

En la matriz presentada anteriormente, se observa que los estados con mayor probabilidad de permanecer en su propio estado son Disminución (D1) e Incremento (I1), con probabilidades de 0.34 y 0.33 respectivamente. Dado que estos estados se encuentran en el centro de la matriz de transición, se podría inferir que la divisa del yen tiende a

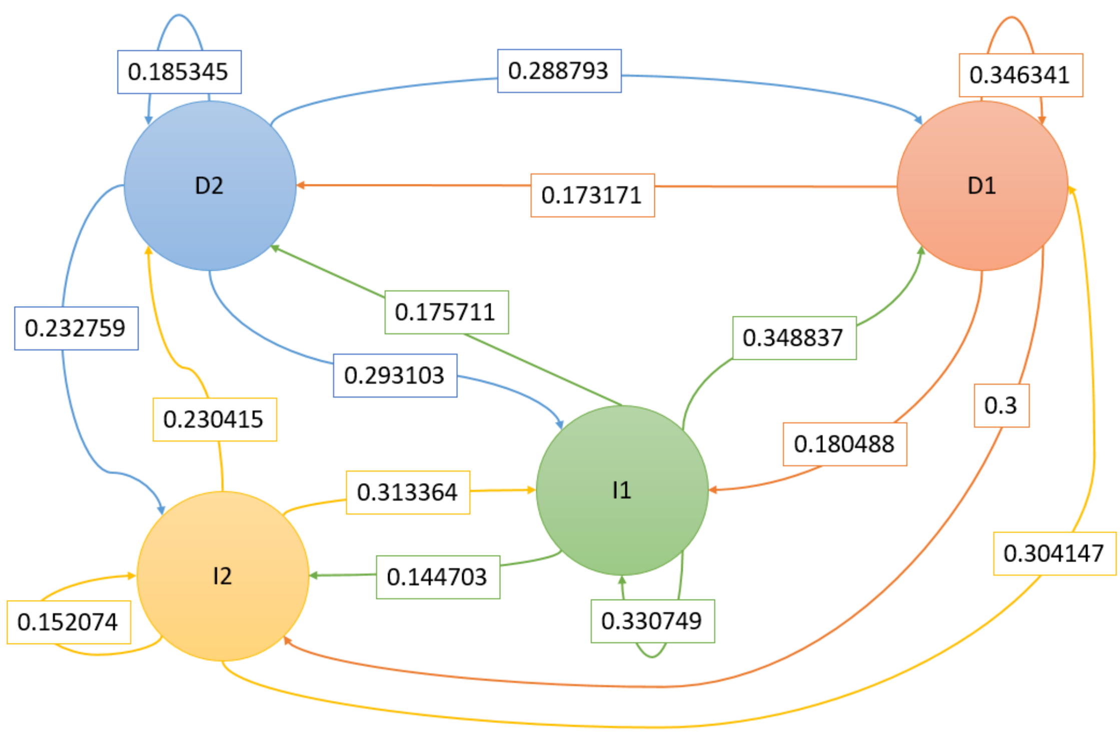
Cuadro 9

Matriz de transición para el YEN

	D2	D1	I1	I2	Total
D2	0.185345	0.288793	0.293103	0.232759	1.0
D1	0.173171	0.346341	0.300000	0.180488	1.0
I1	0.175711	0.348837	0.330749	0.144703	1.0
I2	0.230415	0.304147	0.313364	0.152074	1.0

Figura 3

Diagrama de Markov para JPY



mantenerse estable. Esta afirmación se ve respaldada por el *boxplot* inicial, que no presentaba valores atípicos y mostraba que los datos estaban concentrados en un rango más estrecho en comparación con otras monedas.

Además, se puede notar que si el estado actual es D2, la mayor probabilidad es de transitar a I1. Si se está en D1, la mayor probabilidad es de permanecer en el mismo estado. Si se está en I1, la mayor probabilidad es de transitar al estado D1. Finalmente, si se está en I2, la mayor probabilidad es de transitar al estado I1.

Por lo tanto, esta matriz de transición sugiere que los estados que oscilan con mayor frecuencia son Disminución e Incremento. Este análisis proporciona una visión valiosa de la estabilidad y las tendencias de la divisa del yen.

MXN

Al igual que la divisa de japón, la divisa mexicana no rechaza la prueba de hipótesis Markoviana ni la de homogeneidad, por lo cual se interpreta que para el peso mexicano, en la matriz de transición, solo se necesita el estado actual para predecir la probabilidad del día siguiente o incluso del año siguiente, lo cual cumple con la regla Markoviana de primer orden. Además, el hecho de que no se rechace la hipótesis de homogeneidad podría significar que, independientemente del año en que se encuentre la matriz de transición, sus distribuciones de probabilidades son similares.

A continuación se mostrará la matriz de transición para la divisa mexicana.

Cuadro 10

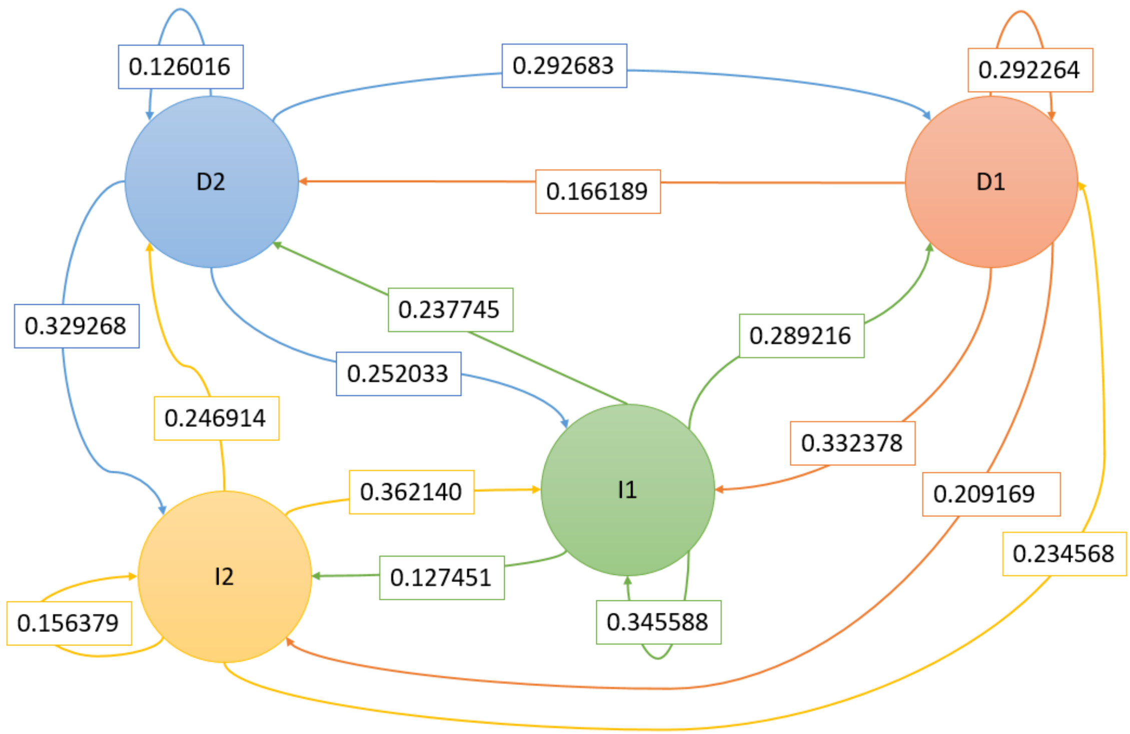
Tabla de Probabilidades de Transición para la divisa mexicana

	D2	**D1**	**I1**	**I2**	**Total**
D2	0.126016	0.292683	0.252033	0.329268	1.0
D1	0.166189	0.292264	0.332378	0.209169	1.0
I1	0.237745	0.289216	0.345588	0.127451	1.0
I2	0.246914	0.234568	0.362140	0.156379	1.0

En la matriz de transición correspondiente a la moneda mexicana, se destaca que el estado con la mayor probabilidad de permanecer constante es el Incremento (I1), con un valor de 0.34. Esto podría indicar que el peso colombiano se está devaluando en menor medida en comparación con el peso mexicano.

Figura 4

Diagrama de Markov para MXN



Si el estado actual es D2, es más probable que el siguiente estado sea I2. Esta dinámica se refleja en el histograma de la divisa mexicana, que es bimodal y muestra caídas y subidas drásticas.

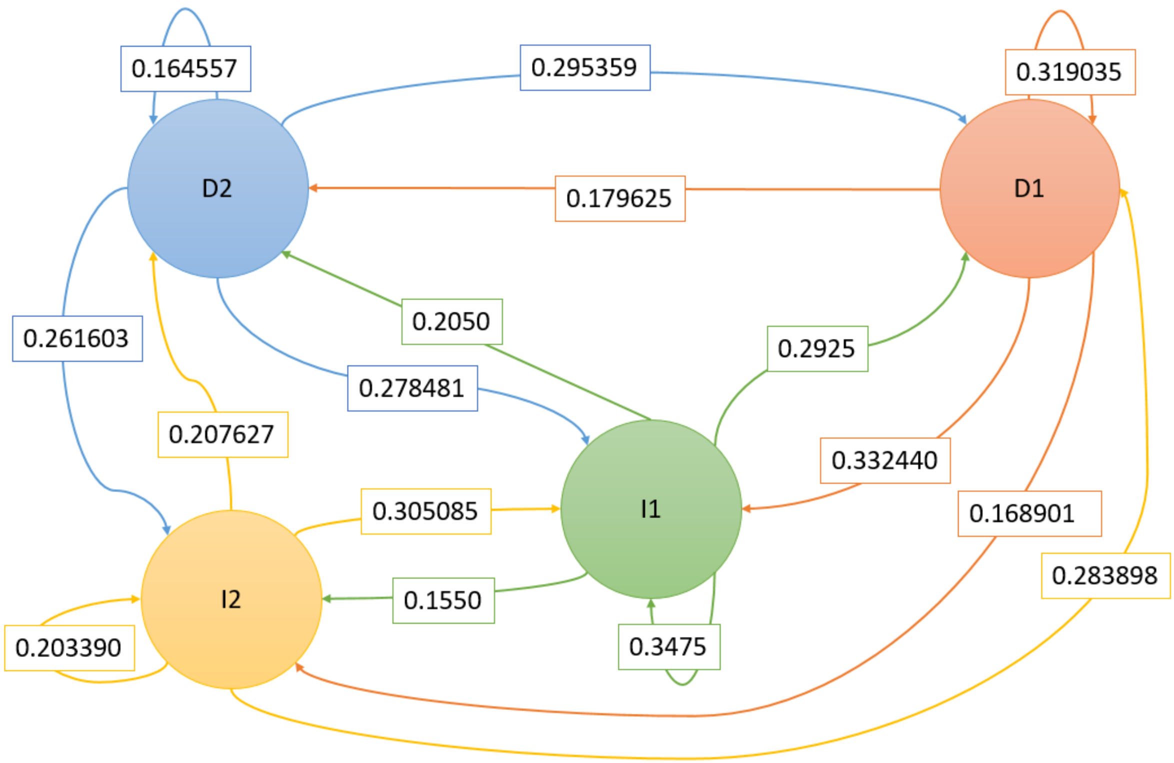
Si el estado actual es D1, es más probable que el siguiente estado sea I1. Si el estado actual es I1, es más probable que permanezca en el mismo estado. Finalmente, si el estado actual es I2, es más probable que el siguiente estado sea I1.

CHF

Al igual que las anteriores divisas, el franco suizo no rechaza la prueba de hipótesis Markoviana ni la de homogeneidad, por lo cual se podría interpretar que para el franco suizo, en la matriz de transición, solo se necesita el estado actual para predecir la probabilidad del día siguiente o incluso del año siguiente, lo cual cumple con la regla

Figura 5

Diagrama de Markov para CHF



Markoviana de primer orden. Además, el hecho de que no se rechace la hipótesis de homogeneidad podría señalar que, independientemente del año en que se encuentre la matriz de transición, sus distribuciones de probabilidades son similares. A continuación se mostrará la matriz de transición para el franco suizo.

En la matriz de transición correspondiente al franco suizo, se destaca que el estado con la mayor probabilidad de permanecer constante es el Incremento (I1), con un valor de 0.34. Esto sugiere una cierta estabilidad en este estado, lo que podría interpretarse como un fortalecimiento del franco suizo con respecto al peso colombiano.

Este incremento constante podría estar reflejado en el dato atípico observado en el *boxplot.* Además, este crecimiento constante podría ser el responsable de la formación de los tres picos en el histograma.

Cuadro 11

Probabilidades de Transición del franco suizo

	D2	**D1**	**I1**	**I2**	**Total**
D2	0.164557	0.295359	0.278481	0.261603	1.0
D1	0.179625	0.319035	0.332440	0.168901	1.0
I1	0.205000	0.292500	0.347500	0.155000	1.0
I2	0.207627	0.283898	0.305085	0.203390	1.0

Adicional, se puede notar que si el estado actual es D2, es más probable que el siguiente estado sea D1. Si el estado actual es D1, es más probable que el siguiente estado sea I1. Si el estado actual es I1, es más probable que permanezca en el mismo estado. Finalmente, si el estado actual es I2, es más probable que el siguiente estado sea I1.

GBP

Por último, la libra esterlina al igual que las anteriores divisas, no rechaza las pruebas de hipótesis de Markov y homogeneidad, por lo cual se podría interpretar que para la libra esterlina, en la matriz de transición, solo se necesita el estado actual para predecir la probabilidad del día siguiente o incluso del año siguiente, lo cual cumple con la regla Markoviana de primer orden. Además, el hecho de que no se rechace la hipótesis de homogeneidad podría apuntar que, independientemente del año en que se encuentre la matriz de transición, sus distribuciones de probabilidades son similares. En la matriz de transición correspondiente a la libra esterlina, se observa que el estado de Disminución (D1) tiene la mayor probabilidad de permanecer en el mismo estado, con un valor de 0.34. Esto sugiere una cierta estabilidad en este estado. Sin embargo, los datos atípicos observados en el *boxplot* podrían indicar un sesgo hacia este estado. Aunque, se sabe que la libra esterlina es una moneda más fuerte que el peso colombiano, este patrón sugiere que la libra esterlina podría estar debilitándose frente al peso colombiano.

Además, se puede notar que si el estado actual es D2, es más probable que el

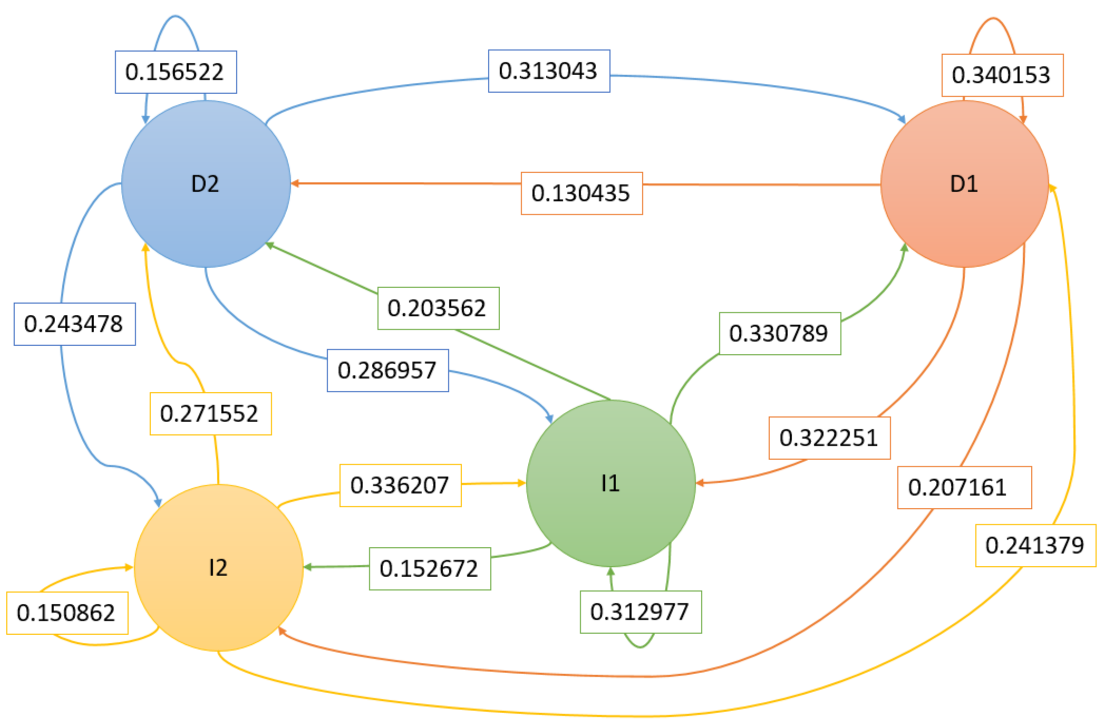
Cuadro 12

Probabilidades de Transición

	D2	D1	I1	I2	Total
D2	0.156522	0.313043	0.286957	0.243478	1.0
D1	0.130435	0.340153	0.322251	0.207161	1.0
I1	0.203562	0.330789	0.312977	0.152672	1.0
I2	0.271552	0.241379	0.336207	0.150862	1.0

Figura 6

Diagrama de Markov para GBP



siguiente estado sea D1. Si el estado actual es D1, es más probable que permanezca en el mismo estado. Si el estado actual es I1, es más probable que el siguiente estado sea D1. Finalmente, si el estado actual es I2, es más probable que el siguiente estado sea I1.

Referencias

Aronsson, M., & Folkesson, A. (2023). Stock market analysis with a Markovian approach: Properties and prediction of OMXS30. https://kth.diva-portal.org/smash/get/diva2:1823899/FULLTEXT01.pdf

Banrep. (2024). Billetes y Monedas - Monedas Disponibles. https://www.banrep.gov.co/es/estadisticas/monedas-disponibles

Eugster, D. (2023). Esplendor y azote: la firmeza del franco suizo. https://www.swissinfo.ch/spa/cultura/esplendor-y-azote-la-firmeza-del-franco-suizo/48517310

FasterCapital. (2024). Yen Explorando el papel del yen en la unidad monetaria asiatica. https://fastercapital.com/es/contenido/Yen--Explorando-el-papel-del-yen-en-la-unidad-monetaria-asiatica.html

García, K. G. (2021). *Análisis probabilístico de sequía meteorológica mediante cadenas de markov y redes bayesianas; cuenca del río tempisque, guanacaste* [Tesis de maestría, Universidad de costa rica sistema de estudios de posgrado].

Imparcial. (2024). ¿Por qué el peso mexicano es tan importante en el mercado de divisas. https://imparcialoaxaca.mx/economia/519963/por-que-el-peso-mexicano-es-tan-importante-en-el-mercado-de-divisas/

Laipaporn, J., & Tongkumchum, P. (2021). Estimating the Natural Cubic Spline Volatilities of the ASEAN-5 Exchange Rates. *Journal of Asian Finance, Economics and Business*, *8*(3), 0001-0010. https://doi.org/https://doi.org/10.13106/jafeb.2021.vol8.no3.0001

MarketScreener. (2024). La importancia de la libra esterlina en los mercados mundiales de divisas. https://es.marketscreener.com/noticias/ultimas/La-importancia-de-la-libra-esterlina-en-los-mercados-mundiales-de-divisas-47219639/#:~:text=No%20s%C3%B3lo%20los%20mercados%20financieros%20utilizan%20la%20libra,del%20d%C3%B3lar%20y%20casi%20el%2023%25%20del%20euro

Rodríguez-González, A., García-Crespo, Á., Colomo-Palacios, R., Guldrís Iglesias, F., & Gómez-Berbís, J. M. (2011). CAST: Using neural networks to improve trading systems based on technical analysis by means of the RSI financial indicator. *Expert Systems with Applications*, *38*(9), 11489-11500. https://doi.org/https://doi.org/10.1016/j.eswa.2011.03.023

Svoboda, M., & Říhová, P. (2021). Stock price prediction using markov chains analysis with varying state space on data from the czech republic. *E+M Ekonomie a Management*, *24*, 142-155. https://doi.org/10.15240/tul/001/2021-4-009