

# Preprocesamiento de datos y selección de *features* para HMM/HSMM en trading intradía de Forex

## Contexto y objetivo del modelo

El proyecto consiste en aplicar modelos ocultos de Markov (HMM) – y posteriormente su extensión a modelos ocultos **semi**-Markov (HSMM) con duraciones explícitas – para identificar cambios de **régimen** en el mercado e inferir estados ocultos (por ejemplo, mercados alcistas vs. bajistas, alta vs. baja volatilidad, etc.). En particular se trabajará con datos intradía del par Forex **EUR/USD**, típicamente en intervalos de 15 minutos (M15), aunque con posibilidad de ajustar a temporalidades mayores como 1 hora (H1) para el análisis intradía. El objetivo inicial es **preprocesar los datos** provenientes de MetaTrader y extraer el **máximo número de características observables** relevantes (*features*) para alimentar el modelo, optimizando así su rendimiento en la detección de regímenes. **Más adelante**, una vez definido un buen conjunto de *features* y un HMM base, se incorporarán mejoras en el modelo (como pasar a **HSMM con distribución de duraciones explícita**) y se realizarán pruebas para verificar que las **probabilidades** inferidas por el modelo efectivamente son capaces de **detectar cambios de régimen** y clasificar correctamente los estados ocultos.

## Datos proporcionados por **MetaTrader** (EUR/USD intradía)

MetaTrader (por medio de su *History Center* o exportación a CSV) entrega las series de precios históricas en forma de **barras OHLCV** para cada intervalo de tiempo seleccionado. Cada registro incluye: fecha, hora, precio de apertura (*Open*), precio máximo (*High*), precio mínimo (*Low*), precio de cierre (*Close*) y el volumen (en Forex, volumen *tick* o número de transacciones durante la barra) <sup>1</sup>. En el caso de EUR/USD, los datos abarcan cotizaciones prácticamente las 24 horas del día (mercado continuo de lunes a viernes), por lo que no hay cierres diarios que generen vacíos significativos en la serie (a diferencia de acciones). El **volumen** proporcionado en MetaTrader para Forex es el volumen *tick*, es decir, el número de cambios de precio registrados en ese intervalo <sup>2</sup> <sup>3</sup>. Si bien no es volumen financiero real (el mercado spot Forex no reporta un volumen consolidado), **sirve como aproximación** de la **actividad del mercado** – p. ej., un alto tick-volume indica mayor actividad o volatilidad en esa vela, aunque con ciertas limitaciones de precisión <sup>4</sup> <sup>5</sup>.

En resumen, **nuestro dataset crudo de MetaTrader para EUR/USD en 15 minutos contendrá una secuencia temporal de registros con timestamp (fecha/hora), Open, High, Low, Close y volumen, a partir de los cuales debemos derivar las características observables que alimentarán el HMM/HSMM.**

## Frecuencia temporal intradía (15M vs 1H)

El enfoque del modelo es de **trading intradía**, lo que implica que trabajaremos con frecuencias menores a un día, típicamente **15 minutos (M15)** como base. Esta temporalidad ofrece un balance entre resolución suficiente para capturar dinámicas intradía y un número manejable de datos por día (96 barras por día en 15min). No obstante, es factible ajustar el intervalo a, por ejemplo, **1 hora (H1)** u otros, según las necesidades de experimentación. De hecho, MetaTrader 5 permite generar barras de

mayor temporalidad a partir de series de minuto (M1) o M15 importadas <sup>6</sup> <sup>7</sup> . Si se importan datos M15, la plataforma puede agruparlos para obtener H1, H4, etc., manteniendo consistencia y alineación temporal <sup>8</sup> <sup>9</sup> .

El análisis de **régimen de mercado** puede depender de la escala temporal elegida. Por ejemplo, en datos diarios de bolsa suele observarse largos períodos de baja volatilidad interrumpidos por crisis de alta volatilidad, lo que sugiere quizás 2–3 regímenes a modelar <sup>10</sup> . En intradía Forex, **los regímenes podrían relacionarse con sesiones de mercado** (ej. sesión europea vs. asiática con distinta volatilidad) **o con periodos tranquilos vs. altamente volátiles dentro del día**. La elección de **15min permite captar movimientos intradía relativamente rápidos**, mientras que a 1 hora se suavizan ciertas fluctuaciones menores. En todo caso, **el número y tipo de estados ocultos** apropiados para el HMM dependerá del horizonte temporal y del activo analizado <sup>10</sup> . **Parte de la experimentación será determinar cuántos estados ocultos capturan mejor las dinámicas del EUR/USD intradía** (dos estados podría modelar “mercado tranquilo” vs “mercado volátil”; con tres quizás añadir un estado intermedio, etc.).

En el contexto intradía, es importante notar que existen patrones recurrentes dentro del día (por ejemplo, mayor volatilidad en la apertura de mercados europeos y americanos, y menor en horarios asiáticos o nocturnos). Estas variaciones cíclicas diarias podrían ser tratadas incorporando **features de hora del día** o similares, aunque **inicialmente podríamos dejar que el HMM detecte implícitamente esos regímenes si son relevantes**. En resumen, trabajaremos principalmente con datos M15 de EUR/USD, sabiendo que los métodos desarrollados podrían aplicarse también a H1 u otras frecuencias intradía.

## Preprocesamiento y extracción de *features* observables

Dado el dataset OHLCV de EUR/USD, el **primer paso es derivar el conjunto de variables observables o features que alimentarán el modelo HMM/HSMM**. El objetivo es extraer la mayor cantidad de información relevante para que el modelo pueda distinguir regímenes de mercado y predecir cambios con la mayor precisión posible. A continuación, se describen posibles *features* a considerar y aspectos del preprocesamiento:

- **Retornos y variaciones de precio:** Es común utilizar **retornos logarítmicos** o simples entre barras consecutivas como input principal de modelos financieros. Por ejemplo,  $r_t = \ln(C_t/C_{t-1})$  donde  $C_t$  es el cierre de la barra  $t$ . Los retornos capturan la dirección y magnitud del movimiento de precio eliminando escalas absolutas. Alternativamente, se pueden usar diferencias porcentuales o absolutas entre *Close* y *Open* de la misma barra (para medir el rendimiento intrabarra) o entre cierres de barras sucesivas. Estos retornos/variaciones suelen ser la base para detectar regímenes (un régimen podría tener retornos con media positiva vs otro con media negativa o distinta varianza, etc.). De hecho, muchos enfoques de HMM financieros usan la serie de **retornos como observación para inferir estados de volatilidad y tendencia** <sup>11</sup> <sup>12</sup> . En nuestro caso EUR/USD, los retornos intradía pueden mostrar periodos de alta volatilidad (mayores amplitudes) frente a periodos de consolidación (retornos cerca de cero).
- **Indicadores técnicos de tendencia y momentum:** Incluir **features basadas en indicadores técnicos** puede enriquecer la información que el HMM recibe sobre el estado del mercado. Por ejemplo:
- **Medias móviles (MA):** diferencias o ratios entre la cotización actual y una media móvil (p. ej. posición relativa del precio vs. su MA de 20 periodos) indican si el mercado está por encima/abajo de la tendencia corta.

- **Indicadores de momentum:** el **MACD** (Moving Average Convergence Divergence) o el oscilador **RSI** (Índice de Fuerza Relativa) proporcionan señales de sobrecompra/sobreventa o cambios de momentum que podrían servir como *features*. En estudios previos, la adición de indicadores técnicos como MACD y RSI ayudó a mejorar la adaptación del modelo a distintos comportamientos del mercado <sup>13</sup> <sup>14</sup>, señalando cambios de tendencia de forma más explícita.
- **ADX (Average Directional Index)** u otros indicadores de tendencia vs rango: podrían indicar la fortaleza de una tendencia (alto ADX) frente a un mercado lateral débil (bajo ADX).
- **Osciladores estocásticos** u otros indicadores de corto plazo pueden aportar información complementaria.

Estos indicadores se calculan a partir de las series de precios OHLC, por lo que pueden derivarse del dataset sin fuentes externas. Es importante configurar sus parámetros (ventanas, etc.) adecuados al timeframe (por ejemplo, RSI de 14 periodos sobre 15min abarca 3.5 horas de datos).

- **Volatilidad local y rango de precios:** Para detectar regímenes de volatilidad, es útil incorporar medidas de volatilidad intraperíodo. Algunas opciones:
- **Rango Alto-Bajo:** La diferencia *High-Low* de cada barra refleja la amplitud de movimiento en ese intervalo. Una media móvil del rango o un **ATR (Average True Range)** pueden indicar volatilidad en aumento o disminución.
- **Volatilidad estadística:** la desviación estándar de los últimos *N* retornos (por ejemplo 1 hora = 4 barras de 15m, o 1 día) podría usarse para informar al modelo sobre el entorno de volatilidad reciente.
- Estas medidas ayudan a que el HMM identifique estados de alta volatilidad (ej. durante noticias o aperturas) versus estados de baja volatilidad (mercado tranquilo), ya que diferentes regímenes a menudo se caracterizan por varianzas distintas <sup>15</sup> <sup>16</sup>.

- **Volumen/tick volume:** Si bien el volumen *tick* en Forex tiene sus limitaciones, puede añadirse como feature para reflejar la actividad del mercado. Períodos de volumen inusualmente alto pueden indicar participación masiva o eventos (posiblemente marcando un régimen diferente de mercado activo), mientras que volúmenes bajos pueden asociarse a sesiones tranquilas (ej. horario nocturno asiático en EUR/USD). Algunos indicadores derivados como *Money Flow Index (MFI)* combinan precio y volumen y podrían también considerarse.

- **Tiempo (hora del día, día de semana):** En intradía Forex, ciertas horas tienen comportamientos típicos (por ejemplo, la apertura de Londres o Nueva York suelen traer movimientos fuertes en EUR/USD). Incluir variables categóricas o numéricas que representen la hora del día o identificadores de sesión podría ayudar al modelo a distinguir regímenes ligados al reloj. Sin embargo, esto introduce estacionalidad conocida; se puede intentar tanto con como sin estas variables para ver si el HMM capta por sí solo estas diferencias.

- **Otros posibles features:** se puede ser creativo y derivar muchas otras variables: distancia del precio a máximos/mínimos recientes, pendientes de tendencias, spreads entre pares correlacionados (aunque aquí solo se considera EUR/USD), etc. La idea es proveer al modelo de **pistas observables** que puedan correlacionarse con los estados de mercado subyacentes. Por ejemplo, si un estado oculto corresponde a "tendencia alcista sostenida", observaremos probablemente una combinación de: retornos consistentemente positivos, precio por encima de la MA, RSI > 50, volatilidad moderada; versus un estado "mercado en rango" con retornos oscilando  $\pm$  pequeños, RSI neutro, bajo rango High-Low, etc.

VOLAT

Al generar este conjunto amplio de *features*, es crucial aplicar **preprocesamiento adecuado**: esto incluye normalizar o estandarizar las variables para que tengan escalas comparables. Según recomendaciones de la literatura, cuando se usan datos multivariados en HMM es importante normalizar cada serie restando su media y dividiéndola por su desviación estándar <sup>17</sup>, evitando así que una variable de magnitud mayor domine la estimación. También se deben tratar posibles *outliers* o valores atípicos (por ejemplo, spikes debidos a errores de datos o spreads anómalos) que podrían distorsionar la formación del modelo; a veces se aplican winsorizaciones o transformaciones logarítmicas si fuera necesario (aunque en retornos log y osciladores esto es menos problemático).

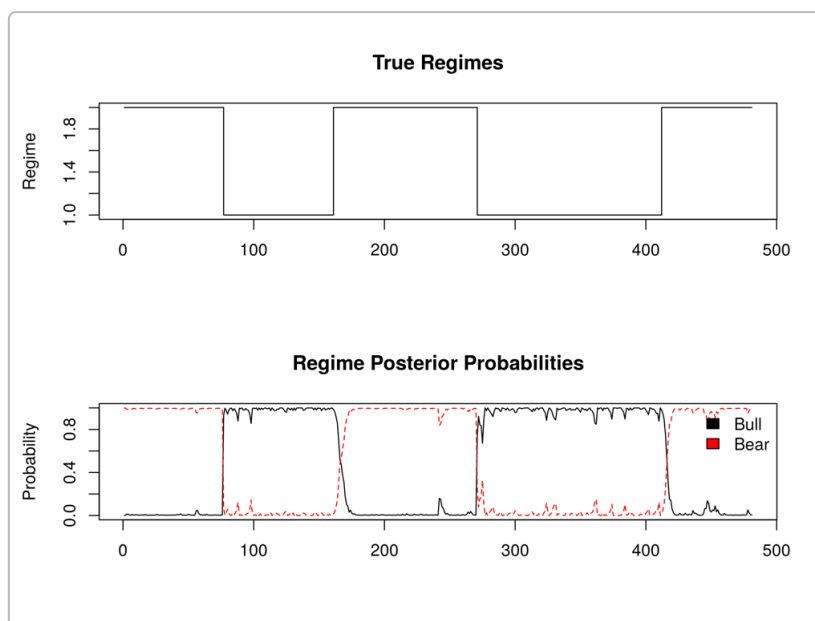
Otra consideración: agregar “el máximo número de *features*” no siempre garantiza mejor desempeño. Existe riesgo de **sobreajuste** si el modelo tiene que estimar demasiados parámetros con dimensiones altas y datos limitados <sup>18</sup> <sup>19</sup>. Para mitigar esto, se pueden: - Limitar correlaciones redundantes (si dos indicadores dicen casi lo mismo, quizás elegir uno). - Comenzar con un subconjunto básico (p. ej. retornos, rango, RSI) y luego ir añadiendo y evaluando impacto. - Asegurar que el modelo HMM multivariable tenga suficientes datos para estimar distribuciones conjuntas (posiblemente suponiendo covarianzas diagonales, es decir, independizar *features* dado el estado, lo cual simplifica la estimación <sup>20</sup>).

En un trabajo de investigación previo, se probaron HMM tanto univariados (solo la serie de retornos) como **multivariados** con hasta 6 *features* adicionales derivadas de precios <sup>21</sup>. Los autores reportan que el HMM puede manejar entradas multivariadas y que la inclusión de variables de apoyo (no linealmente correlacionadas con el retorno principal) tenía potencial para mejorar la predicción <sup>21</sup> <sup>22</sup>. En nuestro caso, se espera que tras un cuidadoso preprocesamiento, el conjunto de *features* seleccionadas (retornos, indicadores técnicos, volatilidad, etc.) provea al HMM/HSMM de la mayor información posible para distinguir los **estados ocultos** del mercado.

## Detección de regímenes con modelo oculto de Markov (HMM)

Una vez preparado el conjunto de observables, se procederá a entrenar un **Hidden Markov Model (HMM)** sobre estos datos. Un HMM asume que hay un proceso subyacente de **estados ocultos** (no observables directamente, por ejemplo “régimen alcista” vs “régimen bajista”) que sigue una cadena de Markov, y que en cada estado el modelo genera observaciones (nuestros *features*) con ciertas distribuciones características. Al *fit* el HMM con los datos de entrenamiento (por ejemplo, usando algoritmo de Baum-Welch/EM para estimar parámetros), el objetivo es que el modelo **aprenda** patrones estadísticos en las observaciones que correspondan a distintos estados ocultos de mercado.

En contexto de trading, los HMM se han usado para **detectar regímenes de mercado** de forma no supervisada <sup>23</sup>. La motivación es que diferentes regímenes (p. ej. mercados tranquilos vs. turbulentos) se manifiestan en cambios de la media, volatilidad, correlación, etc., de las series financieras <sup>15</sup>. El HMM puede *descubrir* estos regímenes porque puede asignar, digamos, un estado oculto 1 a periodos donde los retornos tienen baja varianza y ligera tendencia positiva, y estado 2 a periodos donde los retornos tienen varianza alta o media negativa, por ejemplo. A diferencia de un modelo completamente supervisado, aquí no le damos al algoritmo etiquetas de régimen; el HMM **inferirá** la secuencia más probable de estados escondidos que explica los movimientos observados.



*Ejemplo ilustrativo:* La gráfica superior muestra los **regímenes reales** (simulados) alternando entre dos estados (bull vs. bear), mientras que la inferior presenta las **probabilidades posteriores** estimadas por un HMM de dos estados entrenado sobre la serie de retornos correspondiente <sup>24</sup>. Se observa que el modelo oculto de Markov logra identificar correctamente los cambios de régimen: cuando el mercado pasa a modo “bear”, la probabilidad roja (estado bajista) sube casi a 1 y la negra (alcista) cae a 0, y viceversa. Existe un ligero **desfase temporal** en la detección – el HMM necesita ver algunas observaciones diferentes (p. ej. retornos negativos persistentes) antes de cambiar alta probabilidad al nuevo estado <sup>24</sup>. Esto es esperable, ya que el HMM detecta un cambio **una vez que** la evidencia en los datos sugiere que se ingresó a otro régimen. Aún así, las **probabilidades filtradas/suavizadas** que produce el HMM en cada instante permiten **identificar con bastante precisión los intervalos de cada régimen** y, por tanto, detectar cambios de uno a otro. En la práctica, tras entrenar el HMM obtendremos: - Una matriz de probabilidades de transición  $A$  entre estados (que nos dice, por ejemplo, la probabilidad de permanecer en el mismo régimen vs cambiar en la siguiente barra). - Los parámetros de emisión de cada estado (por ejemplo, media y varianza de retornos u otros *features* en cada estado, distribuciones multi-variadas, etc., que caracterizan estadísticamente a cada régimen oculto). - Dado un nuevo dato o sobre los datos históricos, podemos calcular la **probabilidad posterior** de cada estado en cada momento (usando el algoritmo de filtrado o *forward-backward*). Estas probabilidades por ejemplo podrían graficarse como en la figura, mostrando cómo en ciertos tramos el modelo asigna ~90% prob al estado 1, luego cambia gradualmente y asigna alta prob al estado 2, etc.

Con esta salida, podremos **clasificar cada punto temporal** como perteneciente al régimen X o Y según cuál tenga mayor probabilidad posterior, lo que en esencia segmenta la serie en diferentes **fases de mercado**. Estudios han mostrado que los HMM pueden identificar correctamente regímenes de mercado escondidos en datos históricos – por ejemplo, distinguieron períodos de calma vs crisis en índices bursátiles <sup>16</sup> <sup>25</sup>, o detectaron cambios de tendencia de alcista a bajista en simulaciones <sup>24</sup>. Incluso aplicados a series reales, vemos que un HMM de 2 estados en S&P500 dio alta probabilidad a un estado durante 2004-2006 (mercado tranquilo) y cambió al otro estado durante 2007-2009 (crisis subprime, alta volatilidad) <sup>16</sup>.

En nuestro caso con EUR/USD intradía, un HMM podría por ejemplo descubrir un estado oculto que corresponde a “mercado con baja volatilidad y sin tendencia clara” (quizás predominante durante ciertas horas muertas) y otro estado de “mercado en tendencia/volátil” (ej. durante aperturas con movimientos direccionales). La **clasificación de estados** se validará viendo si las secuencias estimadas

tienen sentido financiero (e.g., comparar con eventos conocidos, horarios, o simplemente verificar si en un estado las distribuciones de retornos tienen mayor varianza que en el otro, etc.). También podemos calcular **verosimilitudes** o métricas para comparar modelos con distinto número de estados y *features*, a fin de elegir la configuración que mejor se ajuste sin sobreajustar (criterios AIC, BIC, etc., podrían ayudar dado que no tenemos *labels* de régimen).

En resumen, el HMM nos proporcionará una **probabilidad en cada instante de estar en cada régimen**; cuando esas probabilidades cambian drásticamente, podemos interpretar que ocurrió un **cambio de régimen**. El análisis de si esas probabilidades efectivamente *anticipan* o al menos coinciden con cambios importantes en el comportamiento del precio será parte de las pruebas. Según la experiencia común, el HMM identificará los regímenes con un leve retraso (necesita evidencias), pero aun así es muy útil para *contextualizar* el estado actual del mercado de forma probabilística <sup>24</sup>. Esto puede integrarse luego en una estrategia (por ejemplo, si la probabilidad de régimen “turbulento” supera cierto umbral, reducir el tamaño de posición, etc., tal como se explora en la literatura de risk management con HMM <sup>26</sup> <sup>27</sup> ).

## HSMM con duraciones explícitas para modelar persistencia de estados

Una limitación de los HMM estándar es que implican una distribución geométrica (memoria sin memoria) para la duración de los estados ocultos. Es decir, en un HMM clásico la probabilidad de que un régimen continúe decae exponencialmente con el tiempo y **no se puede imponer directamente una duración media específica** – el modelo puede, por ejemplo, a veces alternar estados muy frecuentemente si los datos lo permiten, porque matemáticamente no tiene “memoria” de cuánto lleva en un estado. En los mercados, sin embargo, sabemos que los regímenes suelen tener **duraciones típicas** o al menos *persistencia*: por ejemplo, una tendencia fuerte tiende a durar varios períodos, o una fase lateral puede prolongarse cierta cantidad de tiempo antes de romper. Para incorporar este conocimiento, se recurrirá a los **Hidden Semi-Markov Models (HSMM)**, que extienden el HMM permitiendo modelar explícitamente la distribución de duración de cada estado.

En un **modelo oculto semi-Markov**, además de las probabilidades de transición entre estados, se especifica para cada estado una distribución de **sojourn time** (tiempo de permanencia). Esto significa que podemos definir, por ejemplo, que el “Estado 1” (régimen alcista) suele durar entre 5 y 20 velas con una distribución determinada (p. ej. Poisson o Gamma centrada en 10), el “Estado 2” quizá dure menos, etc., según lo observado. El HSMM *recordará* cuánto tiempo lleva el sistema en un estado y ajustará la probabilidad de transición no solo según el estado actual sino también según el **tiempo transcurrido en él**, rompiendo la memoria puramente geométrica del HMM. En términos prácticos, el HSMM es más **flexible** que el HMM: podemos fijar cualquier distribución (paramétrica o no paramétrica) para las duraciones de estado <sup>28</sup>. Esto permite amoldar mejor el modelo a ciertas **características estilizadas** de las series financieras – por ejemplo, en mercados eléctricos se ha usado HSMM para modelar explícitamente que los spikes de precio son de corta duración mientras que otros estados tienen medias móviles más persistentes <sup>28</sup>. En nuestro caso de Forex intradía, podríamos pensar que un régimen de alta volatilidad tiene a veces ráfagas breves (por noticias, quizá solo dura 3–4 barras) mientras que un régimen de tendencia podría durar varias horas; el HSMM nos dejaría incorporar esas expectativas.

La decisión de “usar HSMM para duraciones explícitas más adelante” implica que primero entrenaremos y entenderemos el HMM básico con nuestras *features*, y luego compararemos con la versión HSMM para ver si mejora la detección de regímenes. ¿Qué ventajas se esperan? Principalmente: - **Menos cambios espúreos de estado**: Al imponer una distribución de duración, el modelo no alternará estados

demasiado rápidamente más allá de lo especificado. Esto puede evitar que interprete ruido de corto plazo como múltiples cambios de régimen. El HMM básico a veces tiene el problema de *sobre-segmentar* si dos estados tienen probabilidades relativamente cercanas y los datos fluctúan – un HSMM podría estabilizar esto imponiendo que un estado típico dure, digamos, al menos  $X$  barras en promedio. - **Mejor adecuación a la realidad de mercado:** Si los análisis empíricos muestran que, por ejemplo, los mercados alcistas tienden a durar más que los bajistas (o viceversa), el HSMM puede capturar esa asimetría. De hecho, hay evidencia de que en mercados bursátiles las duraciones de mercados alcistas y bajistas no siguen una memoria sin duración, sino que tienen ciertas dependencias de tiempo (hazard function no constante) <sup>29</sup>. Un estudio encontró que la probabilidad de terminar un mercado bajista aumenta conforme el bajista envejece (duración dependiente) y que una distribución Gamma modelaba mejor esas duraciones que la geométrica implícita de HMM <sup>30</sup>. Esto sugiere que un HSMM podría modelar más fielmente ese comportamiento de duración. - **Identificación de regímenes más granulares:** El HSMM también permite modelar más estados si se desea, aprovechando la dimensión de duración para diferenciarlos. Por ejemplo, un trabajo reciente empleó un HSMM de **cinco estados** en el S&P500, distinguiendo 3 tipos de bull markets (cortos vs largos, e incluso una fase de burbuja) y 2 tipos de bear markets (uno regular y otro tipo crash) <sup>31</sup>. Gracias a esta mayor diferenciación y a la incorporación de duraciones, pudieron obtener un modelo de regímenes más rico que reflejaba bien las fases históricas del mercado <sup>32</sup>. Además, al utilizar esos 5 estados en estrategias de asignación de activos, lograron mejorar el desempeño con menor riesgo respecto a usar solo 2 estados o buy-and-hold <sup>33</sup>. Esto muestra el potencial de los HSMM para *refinar* la detección de regímenes.

En la práctica, incorporar HSMM significa modificar el algoritmo de estimación (el Baum-Welch se adapta para duraciones, o se usa un enfoque diferente como *forward-backward* con duración). Se definirá una distribución de duración  $D_i$  para cada estado  $i$ . Podríamos inferir estas distribuciones de los datos (p. ej., del HMM inicial estimar duración media de estados a posteriori y ajustarle una distribución) o basarnos en conocimiento de dominio para inicializarlas. Luego el HSMM recalculará las secuencias de estados ocultos imponiendo esas duraciones.

**Verificación de desempeño:** Una vez implementado el HSMM, compararemos si sus **probabilidades de estado** y cambios detectados hacen más sentido o mejoran alguna métrica. Posibles evaluaciones: - Ver si el HSMM reduce el número de cambios de estado demasiado rápidos que el HMM pudo haber marcado. - Si tuviéramos datos rotulados aproximadamente de regímenes (quizá identificados por algún criterio externo o inspección visual), ver si el HSMM acierta más en longitud de episodios. - Evaluar la log-verosimilitud o BIC del HSMM vs HMM en datos de entrenamiento y prueba para ver si el ajuste mejora sin sobreajustar (aunque, como nota, una investigación encontró que la ventaja del HSMM sobre HMM puede manifestarse más en ajuste dentro de muestra, mientras que fuera de muestra las diferencias podrían no ser tan significativas dependiendo del caso <sup>34</sup> <sup>35</sup>; en todo caso, vale la pena explorar la diferencia).

En resumen, el uso del HSMM nos permitirá **incorporar conocimiento temporal** sobre cuánto suelen durar los regímenes, haciendo el modelo más realista para detectar **cambios de régimen** en el momento correcto. La hipótesis es que las probabilidades filtradas por el HSMM podrían tener transiciones más suaves y alineadas con duraciones típicas, potencialmente dando señales más claras de cambio de estado. Por ejemplo, si sabemos que un régimen alcista normalmente dura ~2 horas, el HSMM tras hora y media en ese estado comenzaría a aumentar gradualmente la probabilidad de cambio, reflejando la creciente posibilidad de que termine – algo que un HMM estándar no haría, ya que siempre “cree” que la chance de cambio es igual en cada paso independientemente del tiempo en estado actual.

## Conclusiones y siguientes pasos

Inicialmente nos enfocaremos en **preparar un dataset robusto** de EUR/USD intradía, extrayendo todas las *features* relevantes (retornos, indicadores técnicos, volatilidades, etc.) y normalizándolas adecuadamente para entrenar un HMM. Este modelo oculto de Markov servirá para identificar patrones en las observaciones que correspondan a **diferentes regímenes ocultos** de mercado. Evaluaremos si el HMM, a través de sus probabilidades de estado, **detecta cambios de régimen** de forma coherente (p.ej., comparando con intuición de mercado, viendo si las fechas de alta probabilidad de cierto estado corresponden a eventos volátiles conocidos, etc.).

Una vez optimizado el HMM con las *features* que maximicen su rendimiento (ya sea en términos de verosimilitud, estabilidad de estados, o incluso desempeño de alguna estrategia si se integra), pasaremos a implementar el **HSMM** para incorporar duraciones explícitas. Con el HSMM, se espera lograr una **clasificación de estados más estable y realista**, reduciendo falsos cambios muy breves y capturando la persistencia típica de cada régimen. Comprobaremos empíricamente si las **probabilidades posteriores** del HSMM efectivamente reflejan mejor los **cambios de régimen** (idealmente marcando los inicios y fines de fases de mercado con mayor precisión temporal).

En resumen, el plan es: **(1)** extraer el máximo de información de los datos de MetaTrader (OHLCV) mediante un sólido preprocesamiento y selección de *features*; **(2)** utilizar un HMM para aprender regímenes ocultos y validar que con esas observables el modelo puede distinguir estados de mercado y señalar cambios; **(3)** introducir el HSMM para refinar la dinámica temporal de estados, esperando con ello una mejora en la detección de regímenes prolongados y en la interpretabilidad de los resultados. Con estas pruebas, se podrá determinar en qué medida las probabilidades emitidas por el modelo logran **anticipar o confirmar cambios de régimen** en el mercado Forex intradía, contribuyendo en última instancia a estrategias de trading más adaptativas a dichas fases.

**Fuentes:** Los conceptos y enfoques discutidos se apoyan en la literatura y experiencias previas en modelado de regímenes financieros con HMM/HSMM. Por ejemplo, QuantStart demuestra el uso de HMM para identificar cambios de volatilidad en mercados reales <sup>16</sup> <sup>25</sup>. Estudios académicos exploran mejoras de HSMM en identificación de múltiples estados de mercado <sup>31</sup> <sup>33</sup> y discuten la importancia de modelar la duración de tendencias <sup>29</sup>. Además, trabajos de tesis han investigado HMM en Forex con *features* técnicas <sup>13</sup> <sup>14</sup> e indican la necesidad de normalización y cuidado con sobreajuste en modelos multivariados <sup>17</sup> <sup>18</sup>. Estas referencias respaldan la metodología aquí propuesta y sirven de guía para la implementación detallada.

---

<sup>1</sup> <sup>6</sup> <sup>7</sup> <sup>8</sup> <sup>9</sup> Importing Historical Data into MetaTrader 5 for Robust Backtesting - Greaterwaves  
<https://greaterwaves.com/importing-historical-data-to-metatrader-5-for-robust-backtesting/>

<sup>2</sup> <sup>3</sup> <sup>4</sup> <sup>5</sup> Tick Volume in Forex  
<https://www.earnforex.com/guides/tick-volume-in-forex/>

<sup>10</sup> <sup>11</sup> <sup>12</sup> <sup>15</sup> <sup>16</sup> <sup>23</sup> <sup>24</sup> <sup>25</sup> <sup>26</sup> <sup>27</sup> Hidden Markov Models for Regime Detection using R | QuantStart  
<https://www.quantstart.com/articles/hidden-markov-models-for-regime-detection-using-r/>

<sup>13</sup> <sup>14</sup> scholar.tecnico.ulisboa.pt  
<https://scholar.tecnico.ulisboa.pt/api/records/WuxQcfRb824vpZzuCEkB0TjOBMcIxDn-VdYQ/file/7dce27523655b649f28d694c9305bc6c79935e25234aba9f92be53a87d32dc31.pdf>

<sup>17</sup> <sup>18</sup> <sup>19</sup> <sup>20</sup> <sup>21</sup> <sup>22</sup> Algorithmic Trading : Hidden Markov Models on Foreign Exchange Data  
<https://scispace.com/pdf/algorithmic-trading-hidden-markov-models-on-foreign-exchange-7red5597ws.pdf>



28 Incorporating improved directional change and regime change detection to formulate trading strategies in foreign exchange markets

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/370513515_Incorporating_improved_directional_change_and_regime_change_detection_to_formulate_trading_strategies_in_foreign_exchange_mar)

370513515\_Incorporating\_improved\_directional\_change\_and\_regime\_change\_detection\_to\_formulate\_trading\_strategies\_in\_foreign\_exchange\_mar

29 30 31 32 33 Simplicity versus Complexity: A Comparative Analysis of HMM and HSMM for Regime-Based Asset Allocation | Request PDF

[https://www.researchgate.net/publication/](https://www.researchgate.net/publication/379884579_Simplicity_versus_Complexity_A_Comparative_Analysis_of_HMM_and_HSMM_for_Regime-Based_Asset_Allocation)

379884579\_Simplicity\_versus\_Complexity\_A\_Comparative\_Analysis\_of\_HMM\_and\_HSMM\_for\_Regime-Based\_Asset\_Allocation

34 35 A Comparative Analysis of HMM and HSMM for Regime-Based ...

[https://plu.mx/ssrn/a/?ssrn\\_id=4796238](https://plu.mx/ssrn/a/?ssrn_id=4796238)