

# Reconstrucción del Limit Order Book

Juan Delgado, Sebastián Alvarado

September 2025

## 1. Introducción

En los mercados financieros electrónicos, los agentes del mercado operan bajo un sistema de compra y venta de activos financieros donde cada uno puede elegir un precio al que comprar vender alguna cantidad de un activo financiero. Este funciona bajo un sistema de contrapartes, es decir, si alguien desea comprar un activo a un precio dado, la operación de compra se ejecuta si y solo si hay otro agente dispuesto a vender dicha cantidad (o más) a ese mismo precio, bajo este mismo concepto se obtiene el Libro de Órdenes Límite, mejor conocido como: *Limit Order Book (LOB)*.

En el *LOB* completo se pueden observar todos los precios y cantidades disponibles para la compra y venta de un activo financiero, desde el precio mas alto de compra (*best-bid*) y el mejor precio de venta (*best-ask*) y los precios mas altos (bajos) que siguen despues del mejor precio de compra (venta) para una ventana de tiempo en específico, y este se puede ajustar para poder obtener los  $k$  mejores niveles del *LOB*, es decir, si  $k = 3$ , el *LOB* mostrará los 3 precios mas altos de compra y los 3 precios mas bajos de venta disponibles.

Ask		Bid	
Cantidad	Precio	Precio	Cantidad
<b>700</b>	<b>10.2</b>	<b>10</b>	<b>350</b>
150	11	9.8	450
120	12	9.5	400

Cuadro 1: Limit Order Book (LOB) de 3 niveles del activo X

El cuadro 1 muestra un *LOB* hipotético de 3 niveles, donde la primera línea (en negrilla) representa el mejor bid y el mejor ask disponibles en un tiempo  $t$  arbitrario. Dado que este cambia a lo largo del tiempo, es posible visualizar la evolución del precio y cantidad a un nivel como una serie de tiempo común, o, si se toma todo el libro, se puede visualizar como una serie de tiempo de alta dimensión.

### 1.1. Métricas generales del LOB

Entre las métricas calculables de un *LOB* ampliamente reconocidas en la literatura existe el *precio medio*, *el spread*, *el desbalance*, *la profundidad total*, entre otros. El precio medio es calculado como la media aritmética del mejor bid y el mejor ask, el spread es la diferencia entre el mejor bid y el mejor ask, el desbalance es la cantidad disponible en el

bid, dividida por la cantidad disponible en el bid y en el ask y la profundidad del mercado es la suma de las cantidades disponibles en cada lado del libro, es decir la profundidad del bid es la suma de todas las cantidades disponibles en el bid, mientras que la profundidad del ask es la suma de todas las cantidades disponibles en el ask.

## 1.2. Motivación

Consideramos que en Colombia, uno de los mayores factores que hace que su mercado accionario sea ineficiente y ampliamente considerado ilíquidos es la falta de información de fácil acceso y significativa para la toma de decisiones informadas, con lo cual, el objetivo de este proyecto es realizar la reconstrucción/simulación realista de LOB para 2 acciones colombianas, la acción ordinaria de Corficolombiana y la acción preferencial de Davivienda con el fin de poder obtener datasets de entrenamiento para algoritmos de trading y las distintas metodologías de machine learning aplicables en el campo de las finanzas, incluyendo algoritmos de *market making* y de trading bajo información disponible en un LOB, buscando potenciar los métodos cuantitativos en el mercado accionario colombiano. Mediante este paper buscamos brindar una guía de trabajo inicial para poder realizar este tipo de modelamientos de microestructura del mercado en el país y brindar un modelo generalizado que permita hacer esta reconstrucción para cualquier activo financiero y mostrar su funcionamiento para las 2 acciones previamente mencionadas.

El problema que se busca resolver en este documento es poder reconstruir o simular fielmente este libro utilizando solo las puntas de compra y venta, puesto que en el mercado colombiano la información del libro no está disponible, razón por la cual, el *best-bid* y el *best-ask* son los únicos datos que se lograron conseguir para diferentes ventanas de tiempo.

## 2. Estructura de los datos

En el archivo *base\_market\_making.xlsx* se encuentra la información de la negociación de 2 activos, la acción ordinaria de Corficolombiana *CFV.CN* y la acción preferencial de Banco Davivienda *DVI.p.CN*.

### 2.1. Datos disponibles

La base de datos dispone de campos que incluyen precios, volúmenes, identificadores de transacciones, marcas de tiempo, entre otros. Los datos disponibles en la base de datos útiles para el desarrollo de este paper son los siguientes:

- **Timestamp:** Indica la marca de tiempo a la que ocurre un *evento* en la base de datos, dada por el formato *fecha:hora:minuto:segundo.milisegundos*.
- **EVENT\_TYPE:** Indica el tipo de *evento* que sucede en el libro en un *timestamp* específico, este puede ser *quote* que representa la llegada de una nueva orden en cualquier nivel del *LOB*, mientras que *trade* representa que una cantidad fue negociada a cierto precio en el libro, eliminando alguna parte del volumen o el volumen total de las puntas y niveles mas alejados en el *LOB*.

- **SOURCE\_DATETIME:** Indica marcas de tiempo con  $10^{-9}$  segundos disponibles en las que sucedió un *evento*, bajo el formato *fechaT:hora:minuto:segundo.10<sup>-9</sup> segundos*.
- **TRDPRC\_1:** Indica el precio al que se ejecutó un evento *trade* en el libro de órdenes.
- **TRDVOL\_1:** Indica la cantidad de acciones que se operaron en un evento *trade* en el libro de órdenes.
- **ACVOL\_UNNS:** Indica la cantidad de acciones que se han operado a lo largo de todos los eventos *trade* realizados en una sesión de negociación.
- **VWAP:** *Volume-Weighted Average Price*, Indica el precio ponderado por volumen operado. matemáticamente definido de la siguiente forma:

$$VWAP = \frac{\sum_{i=1}^n q_i \times p_i}{\sum_{i=1}^n q_i} \quad (1)$$

Donde  $q_i$  representa la cantidad de unidades de un activo (en este caso, acciones) operada en el tiempo  $i$ ,  $p_i$  representa el precio al que se ejecutó la operación de  $q_i$  unidades del mismo activo en el tiempo  $i$ , mientras que  $n$  representa el número de operaciones realizadas en una sesión de negociación (haciendo que exista un *VWAP* para cada sesión de negociación)

- **BID:** Representa *best-bid*, es decir, el mejor precio de compra existente en el libro, disponible para cada ventana de tiempo
- **BIDSIZE:** Representa la cantidad de acciones que se desean comprar al precio del *best-bid* (en la base de datos: *BID*)
- **ASK:** Representa *best-ask*, es decir, el mejor precio de venta existente en el libro, disponible para cada ventana de tiempo
- **ASKSIZE:** Representa la cantidad de acciones que se desean comprar al precio del *best-ask* (en la base de datos: *ASK*)
- **TRNOVR\_UNNS:** Indica el *nocional operado*, es decir, la suma de las multiplicaciones de precio y volumen operado para cada operación (*trade*) disponibles en el

libro, o, lo que es lo mismo, multiplicar el *VWAP* por el volumen acumulado (en la base: *ACVOL\_UN\$*).

- **MIDPRICE:** Representa el *precio medio*, una medida ampliamente utilizada en la literatura como el precio “justo” de un activo, se calcula mediante la media aritmética entre el *best-bid* y el *best-ask*

## 2.2. Datos calculables

Existen 2 datos calculables ampliamente utilizados en la literatura de *trading en un libro de órdenes límite*, estos son el *spread* y el *imbalance*:

- **SPREAD:** Es la diferencia entre el *best-ask* y el *best-bid*. Dado que el mejor precio de venta SIEMPRE estará por encima del mejor precio de compra (de ser menor o igual al precio de compra, se ejecutaría un *trade*), este valor siempre será positivo.
- **SPRDTCK:** Es el *spread* dividido entre el *tick-size*, es decir, la diferencia mínima posible entre precios para un mismo activo (ej. si un activo cotiza a 100 y tiene un *tick-size* de 1, el siguiente precio menos alto posible es de 101, mientras que el siguiente precio menos bajo posible es de 99).
- **IMBALANCE:** Es la proporción de las cantidades disponibles en el *best-bid* (es decir, el *bid size*) con respecto a toda la cantidad disponible en el mejor nivel del libro, formalmente se define de la siguiente forma:

$$I = \frac{Q^B}{Q^A + Q^B} \quad (2)$$

Donde  $Q^K$  representa las cantidades disponibles en  $K = \{A: Ask, B: Bid\}$ .

- **BIDASK\_DISTANCE:** Indica la distancia *nominal* del *best-bid* y el *best-ask* respecto al *mid price* en el mismo *source datetime*. Se calcula dividiendo el *spread* entre 2, puesto que la distancia al *mid price* es igual hacia el *bid* y *ask*
- **INCREASE\_BID:** Es el número acumulado de órdenes en el *best-bid* que han aumentado el *bid size* en la misma sesión de negociación. Se obtiene al sumar la cantidad de *veces* que sucede un aumento en el *bid size* respecto al *source datetime* anterior.
- **INCREASE\_BIDSIZE:** Es el total de acciones en el *best-bid* que han aumentado el *bid size* durante un evento *quote*. Se obtiene al observar la cantidad de acciones que aumentan el *bid size* respecto al *source datetime* anterior.
- **INCREASE\_AC\_BIDSIZE:** Es la suma (es decir, el acumulado) de los *increase bidsizes* en cada *source datetime*, en caso de no aumentar las acciones en esa ventana de tiempo se mantiene el acumulado previo.

- **INCREASE\_ASK:** Es el número acumulado de órdenes en el *best-ask* que han aumentado el *ask size* en la misma sesión de negociación. Se obtiene al observar un aumento en el *ask size* respecto al *source datetime* anterior.
- **INCREASE\_ASKSIZE:** Es el total de acciones en el *best-ask* que han aumentado el *ask size* durante un evento *quote*. Se obtiene al observar la cantidad de acciones que aumentan el *ask size* respecto al *source datetime* anterior.
- **INCREASE\_AC\_ASKSIZE:** Es la suma (es decir, el acumulado) de los *increase asksize* en cada *source datetime*, en caso de no aumentar las acciones en esa ventana de tiempo se mantiene el acumulado previo.
- **REDUCTION\_BID:** Es el número acumulado de órdenes en el *best-bid* que han sido reducidas en la misma sesión de negociación. Se obtiene al observar una reducción en el *bid size* que no es producto de un evento *trade*.
- **REDUCTION\_BIDSIZE:** Es el total de acciones reducidas en el *best-bid* en un mismo *source datetime*. Se obtiene al observar una reducción en el *bid size* que no es producto de un evento *trade* respecto al *source datetime* anterior.
- **REDUCTION\_AC\_BIDSIZE:** Es la suma (es decir, el acumulado) de los *reduction bidsize* en cada *source datetime*, en caso de no reducir las acciones en esa ventana de tiempo se mantiene el acumulado previo.
- **REDUCTION\_ASK:** Es el número acumulado de órdenes en el *best-ask* que han sido reducidas en la misma sesión de negociación. Se obtiene al observar una reducción en el *ask size* que no es producto de un evento *trade*.
- **REDUCTION\_ASKSIZE:** Es el total de acciones reducidas en el *best-ask* en un mismo *source datetime*. Se obtiene al observar una reducción en el *ask size* que no es producto de un evento *trade* respecto al *source datetime* anterior.
- **REDUCTION\_AC\_ASKSIZE:** Es la suma (es decir, el acumulado) de los *reduction asksize* en cada *source datetime*, en caso de no reducir las acciones en esa ventana de tiempo se mantiene el acumulado previo.
- **OFI:** Es el *Order Flow Imbalance*, planteado por Cont et al. (2013). Se define como la diferencia de las variaciones del volumen en el *best-bid*, y tiene en cuenta el volumen cuando ingresa una nueva *punta*, si hay una reducción en la misma, o si el volumen varía en el mismo precio para ambos lados en una ventana de tiempo  $i \in [t_0, t]$ . Matemáticamente definen la variación mediante las siguientes formulas:  
Para el *bid*:

$$\Delta B_n = \begin{cases} +B_n, & b_n > b_{n-1} \text{ (aparece nueva cola más alta)} \\ -B_{n-1}, & b_n < b_{n-1} \text{ (se retira la cola previa)} \\ B_n - B_{n-1}, & b_n = b_{n-1} \text{ (cambia el tamaño al mismo precio)} \end{cases}$$

Para el *ask*

$$\Delta A_n = \begin{cases} A_n, & a_n < a_{n-1} \text{ (ask mejora: más oferta cercana)} \\ -A_{n-1}, & a_n > a_{n-1} \text{ (ask se aleja: menos oferta)} \\ A_n - A_{n-1}, & a_n = a_{n-1} \text{ (cambia el tamaño al mismo precio)} \end{cases}$$

Un evento en la ventana de tiempo  $n$  se define como:

$$e_n = \Delta B_n - \Delta A_n$$

Y finalmente, el *Order Flow Imbalance* está definido por:

$$\text{OFI}(t) = \sum_{n: t_0 < \tau_n \leq t} e_n$$