

# LSEG Tick History – Query をバックテストに活用 – トレーディング戦略の信頼性を さらに向上

システマティック・トレーディング戦略を構築するうえで重要な要素となるのが、バックテストです。このバックテストに高頻度のティック・データを利用できると、日中取引戦略だけでなく、さまざまな執行アルゴリズムを用いた日次取引戦略を正確にシミュレーションするうえでも便利です。しかし、ティック・データベースを自社管理するとなると多大な時間とコストがかかるため、これまでティック・データの利用は容易ではありませんでした。LSEG Tick History – Queryは、膨大な数の銘柄と市場の厚み(デプス)に関する完全管理型データベースです。このソリューションを使うことで、ティック・データベースを自社管理する必要がなくなります。

## 取引戦略にバックテストが必要な理由

取引戦略を策定する際、それが実際のマーケットにおいて、将来的にどのようなパフォーマンスを出すかは、当然ながらわかりません。しかし、過去のデータを使って、過去においてその戦略を実行していたら、どのようなパフォーマンスとなっていたかを評価することは可能です。将来のパフォーマンスが過去のパフォーマンス通りになるとは限りませんが、バックテストは取引戦略の策定における重要な要素となっています。

もし、ある取引戦略が過去のデータ上で一貫して損失となるのなら、その戦略を実際に適用するかどうかを判断するうえで重要なデータポイントとなります(もちろん、市場環境の変化が見込まれるという確実な根拠があれば別ですが)。バックテストは、取引コストの損益分岐点や、戦略の推定規模など、さまざまな要素を知るうえでも役に立つのです。



## 正確なバックテストに求められる要素とは

バックテストには、適切にタイムスタンプが付与された、クリーニング済みのデータを使う必要があります。タイムスタンプが不正確なデータだと、先読みバイアスが入ってしまう(つまり将来のデータを使ってしまう)ことになり、過去のリターンが実際よりも大きくなる可能性があります。また、バックテストにおいては取引コストを正確に想定することも重要です。これが不正確だと、やはり過去のリターンが誇張されてしまいます。バックテストは、過去のリターンが正確に反映されて初めて、取引戦略の策定において有益となるのです。

## 日次取引戦略のバックテストに役立つ高頻度データ

バックテストは、取引する金融商品の種類や取引ルールの高頻度により、その複雑性が異なります。高頻度の取引戦略のバックテストには、当たり前ですが高頻度のマーケット・データが不可欠です。日次取引ルールのバックテストなら、それほど多くのマーケット・データはいらないと思われる方もいるでしょう。

しかし、この場合も高頻度データが役に立ちます。もちろん、適切にタイムスタンプが付与された、クリーニング済みのデータを使う必要があります。高頻度のマーケット・データが特に活かせるのは取引シグナルのバックテストです。通常、取引シグナルは日次データのみを使って取引終了時に生成されますが、高頻度データがあれば、1日に何回も生成することができます。

## 意外に多い、高頻度のきめ細かいティック・データが必要なケース

高頻度のティック・データがあれば、従来の取引シグナルの生成頻度に関わらず、さまざまな執行アルゴリズムを用いた取引のシミュレーションを行うことができます。したがって、終値注文のシグナルに基づくバックテストではなく、執行アルゴリズムを用いた取引執行(1日に複数回に分けた執行)を行った場合の戦略パフォーマンスのバックテストもできるようになります。この選択肢は、規模の大きなファンドにとっては特に重要です。こうしたファンドの想定執行額は規模が大きく、マーケットに大きな影響を与えます。大量の取引を一度に執行すると、取引戦略のリターンに悪影響を及ぼす可能性があります。

板情報の上位注文(最良売買気配値)だけでなく、それ以下の売り買い注文の個別の取引量についてのデータが必要となることは多々あります。レベル2、特にレベル3のデータがあれば、板情報の最良気配だけでなく、さまざまなレベルの情報にアクセスでき、市場の厚みを評価することができます。板情報の最良気配を見るだけでは、市場の流動性の一部しか知ることができません。しかし、その厚みがわかれば、より大きな全体像が把握できるようになります。

こうしたきめ細かいデータセットを活用することで、出来高加重平均価格(VWAP)などの執行アルゴリズムのコンピュータ処理や、独自のリスク移転価格の推計が可能となり、それぞれの時点におけるマーケット注文のコストをより正確に評価することができます。また、きめ細かなマーケット・データを使用すれば、板情報の歪みをモニタリングし、市場のインバランスなどの評価指標の計算も可能となるため、取引アルゴリズムの構築に役立てることができます。

## ティック・データベース自社管理に時間がかかる理由

高頻度のティック・データを用いて取引戦略のバックテストを行うには、膨大な量のティック・データベースにアクセスする必要があります。ティック・データベースを最新の状態に保つには、ストリーミング・データの定期的な取り込みや、ダウンロードによる定期的更新など、作業にとっても時間がかかります。また、ティック・データベースのハードウェア管理も必要であり、データ量の増加に伴い、拡張しなければなりません。また、コンピュータ能力に対するユーザーの要求も高まるでしょう。レベル 2、または特にレベル 3 のデータを使用していれば、求められるコンピュータ能力やストレージの規模も大幅に増加します。

データベースを自社管理する時間やコストをかけずに、ティック・データを用いたバックテストをする方法はあるのでしょうか。

## LSEG Tick History – Queryにより、ティック・データを用いたバックテストが容易に

LSEG Tick History – Queryを使えば、ティック・データベースの自社管理は不要となります。LSEG Tick Historyは、膨大な数の銘柄の25年以上にわたるティック・データを収録したデータベースで、毎週2テラバイトのデータが継続的に更新されています。レベル 2、レベル 3 のよりきめ細かいデータも含まれています。

LSEG Tick HistoryはBigQuery経由でアクセスできるため、使い慣れたSQL言語を用いてバックテストでのクエリ実行が可能です。また、BigQueryはウェブ・インターフェースまたはPythonなどのAPIを通じてアクセスでき、社内のバックテストのフレームワークとも統合できます。

社内のハードウェアやスケーリングの自社管理も不要となり、それに代わってBigQueryがコンピュータ能力の必要性に応じて、自動的にスケーリングします。データを自社コンピューターに取り込む(つまり、データをローカルシステムにダウンロードして管理する)のではなく、コンピュータ能力をデータに移すことにより、時間とコストを削減できるという仕組みです。



**Saeed Amen**

Cuemacro 創業者。過去15年にわたり、リーマン・ブラザーズや野村證券などの大手投資銀行でシステムティック・トレーディング戦略の開発に従事。著書に『タレス派の投資術: 古代に学ぶ現代の投資 (Trading Thalesians: What the ancient world can teach us about trading today)』(Palgrave Macmillan 刊)、共著書に『オルタナティブ・データブック (The Book of Alternative Data)』(Wiley 刊)。現在は、Cuemacroでシステムティック・トレーディング分野でのコンサルティングやリサーチ事業を手掛ける。取引コスト分析用のPythonライブラリ(finmarketpy、tcopyなど)を多数開発。大手クオンツ・ファンドやデータ企業を顧客に持つ。主要会議や、ECB、IMF、イングランド銀行、FRBなどの機関での研究発表多数。ロンドン大学クイーン・メアリー校客員講師、Thalesians共同創設者でもある。

