

# Operating in illiquid markets: How to gather, consolidate and use disparate data sources to enhance returns and more effectively control risk



**Gus Sekhon**  
Head of Product & Marketing  
FINBOURNE Technology

This article is neither an exercise in blame nor a lament at the status quo. Instead, we aim to examine the challenges inherent in operating in these illiquid markets and identify necessary improvements in the process of gathering, consolidating and utilising disparate data sources to enhance returns and more effectively control risk. In effect, we propose the toolset you need at your disposal to navigate these markets. Spoiler – we think you are underpowered without data virtualisation, scripting and simulation as standard.

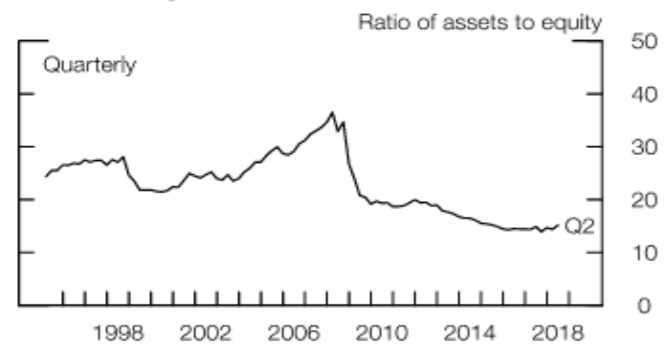
### Illiquid markets: the challenges

In deep and liquid markets such as short-dated FX, a constant stream of reference prices can be obtained from numerous brokers. Such markets, with high information density at all tenors and over all periods, lend themselves well to simulation and learning algorithms, narrowing the knowledge gap between participants. But they are becoming rarer as bank leverage continues to fall (see right). As the list of instruments suffering dramatically reduced liquidity widens (e.g. IRS non-linear products), the demand shifts to alternative products and markets each with their own idiosyncrasies.

Illiquid markets suffer wider bid-offers, thinner volumes and higher volatility. Illiquidity can manifest itself in almost any instrument from equity (e.g. alternative micro-cap listings or privately held stocks) to small cap or non-vanilla corporate debt to high yield issuance. In such illiquid markets, the reference information for pricing through replication is often unavailable or stale (see left). Without regular trades to feed the model, a trader needs to imply prices and risk from instruments that are imperfectly related to the target instruments. This necessitates an improved ability to gather information from related data sources and to seamlessly integrate that information into models

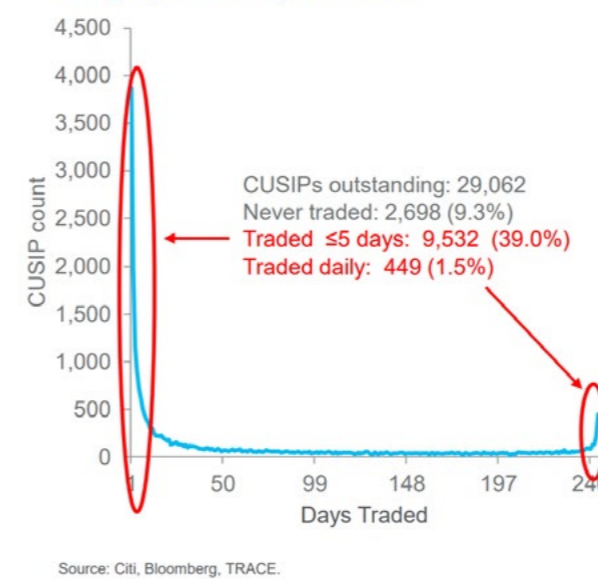
The jury may be out on the precise causes, but there can be little doubt amongst market participants in recent times that there has been a steadily growing bid for illiquid asset classes, venues and ownership structures, whether as a core strategy or to enhance returns on more conventional products. The likely factors in this relentless bid are the inevitable effects on the global yield curve of financial repression, the knock-on impact of this on the hunt for returns, changes in investor demographics - most notably mobility - and of course the shift to passive investing which leads in turn to the imperative for differentiation.

3-4. Leverage of Broker-Dealers



Note: Leverage is calculated by dividing financial assets by equity.  
Source: Federal Reserve Board, Statistical Release Z.1, "Financial Accounts of the United States."

The liquid core is very, very small....  
Trading days for US corporate bonds



Source: Citi, Bloomberg, TRACE.

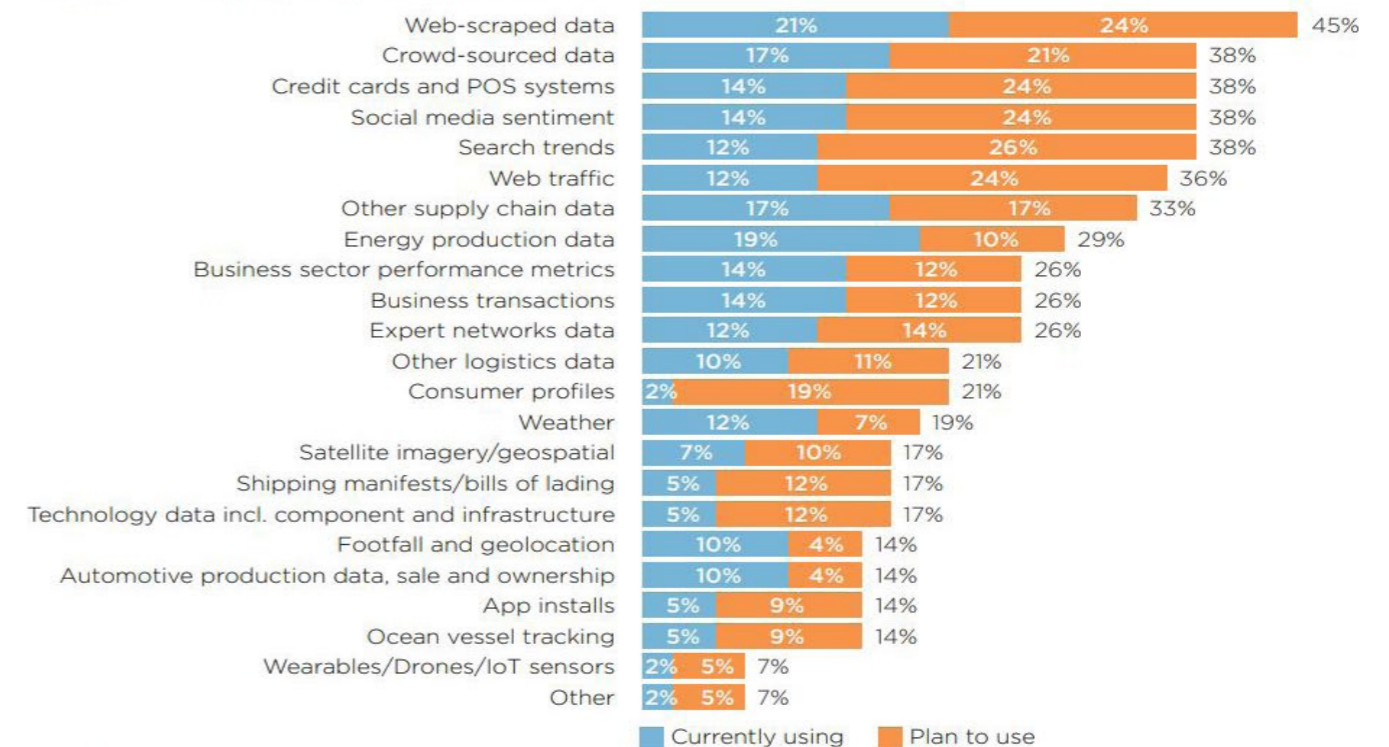
### How can you best gather intelligence from additional data sources?

The challenge with alternative data is well known. Whilst it is now [the norm](#) (for hedge funds at least) to incorporate alternative data into allocation decisions (see right), fewer than a quarter of funds are doing so for risk management and more than three-quarters had trouble back-testing. Clearly quality and consistency is still problematic. Data is typically a mix of structured and unstructured, inconsistent in format (e.g. something as simple as the meaning of timestamps for non-transacted data), lacking in common identifiers, often without clear ontology or lineage, prone to input errors, non-standardised units and stitching together for meaningful time series is challenging.

Traditionally the chosen solution was the export, transform, load (ETL) process - assume a clear and consistent ruleset and contort the data to fit your target format. Clearly, this process fails in an inscrutable fashion with even minor exceptions as every fix requires code debugging and upgrades. Similarly, a data lake-based process (the solution taking the SaaS market by storm in the US) may not require ETL but does have concurrency issues. Static series of loads with rapid querying layered over the top are not fit for valuation or risk models because of the multi-stage nature of the process.

By contrast, a virtualisation mechanism, where the 'gather' stage is run as a distributed, concurrent process has the advantage of being real-time and also of surfacing the transformation or model logic, which makes it discoverable to the user and thus pushes the remediation onus back onto the provider.

### USAGE OF ALTERNATIVE DATA SETS



Note: Based on 42 respondents.  
Source: Greenwich Associates 2019 Alternative Data Study

**Once you have that data, how do you organise it into something useful?**

Once the data is in place, the next step is to run a set of data cleansing algorithms, each targeting one set of flaws and where the intermediate steps themselves can be easily checked. If your virtualisation process is re-entrant you can run the gather and cleanse stages in one. With effective data partitioning and entitlement (ensuring you specify the correct access levels for each participant) you can also stage the cleansing for different use cases. The output is typically a resampled stream of data where the flaws have, according to some target metric, been removed and where additional tags have been added. These tend to include standard instrument identifiers and datetime units to facilitate discovery and joining algorithms further in the process.

The data is now suitable for the application of learning or optimisation algorithms. If not already done, remaining gaps in the data are filled, for instance using a pricing model to imply a missing set of bond prices from rates and credit spreads or generating additional data such as risk measures.

Combining in liquidity data sources is vital given the inherently diminished liquidity we are already dealing with. The stakes tend to be much higher in these markets, and there are countless examples of experienced traders missing their off-ramp. More specifically, information on monetary activity, market trading regime, whether stressed or calm, and simple seasonal volume data can facilitate modelling of decisions such as callable bond exercise probability. Alternatively, projected earnings might be combined with balance sheet information to enhance calibration of credit curves for valuing debt.

A set of such data streams can then be selected for feeding through an optimisation framework to produce a classifier that can be used to generate trading signals. Alternatively, one might attempt to combine pricing models using maximum likelihood/risk-neutrality with priors obtained from related market state or economic data.

**Model: How can you then push that data into a customised/scripted pricing and risk process?**

We know markets, particularly illiquid ones, are capricious at the best of times. A novel insight from yesterday may be obsolete today. This necessitates that the creation of models be flexible. The time to organise and model data should also allow the model to evolve or be easily reconfigured. Likewise, if a new data source becomes available, joining it and incorporating it into the risk process should not invalidate previous work. A scripting and configuration layer that allows one to quickly switch the underlying models or integrate a new data source without weeks or months of coding by specialists is desirable. Scripting allows these layers to be separated, and as with the virtualisation process, surfaces the logic to the user, where entitled. In this way model improvements can be identified and implemented rapidly and rolled out to trading strategies by the portfolio manager with immediacy.

To operate successfully in illiquid markets, it is clear that risk management tools and specifically the data input to that process must not be an afterthought to the asset allocation stage. This necessitates holding alternative data ingestion to much higher standards both in quality and immediacy in the gather, join and cleansing stages and in its logical partitioning and entitlement by participant roles and access. This is entirely within the remit of an integrated virtualisation process. Additionally, there needs to be an externalised scripting process that is owned by the client and allows for agile integration to the optimisation and simulation stages. Suddenly your risk process can evolve with your trading strategy, rather than being an afterthought. The industry seems to indicate that combining these tools within the risk process is not a challenge that has been universally met; while this remains true, there is a clear competitive edge in illiquid markets to those funds that take the step.

# DRIVE GROWTH, INCREASE CONTROL AND ENHANCE DATA ACCESS

# LUSID

BY FINBOURNE

THE FUTURE OF HEDGE FUND TECHNOLOGY