



# It's time to commit to data

By Abraham Thomas

Capital markets are like Lake Wobegon: everybody is above average. Every active investor thinks their forecast is better than the market consensus. Every algo trader thinks their orders are not being picked off. Every risk manager thinks their model is more robust than anyone else's.

To some extent this is understandable. It's irrational to participate in a zero-sum competition if you don't have some sort of advantage. And it's hard to attract investors or clients if you don't at least claim to have an edge.

But that edge, even if it exists, is always tenuous. An execution algo from 10 years ago would be woefully inadequate today.

It's an evolutionary arms race; stay still and you'll perish. And practitioners understand this, which is why models, algorithms and infrastructure are constantly being reinvented. Firms invest heavily to stay on the cutting edge of trading technology; "constant vigilance!" is the battle cry.

**TAKING DATA FOR GRANTED**  
Amazingly, this degree of diligence does not extend to one of the key inputs to every single capital markets decision: data. Firms tend to take data for granted. Whether in portfolio design, or trade execution, or risk management: after an initial exploration and implementation, the data element is largely ignored. There's rarely an ongoing investment in data; it's assumed to "just work".

This is misguided. Even the best systems fail if they're given incorrect information. And in a zero-sum game, information doesn't even have to be 'incorrect' for the system to fail; merely 'less good' than the competition. A simple model built on good data will beat a sophisticated model built on bad data every single time. And yet, firms invest far more time, effort and resources on building and improving and iterating their quantitative superstructure than they do on reinforcing their data foundations.

FX market participants are especially guilty of this error, because FX data has historically been hard to find and hard to use. FX markets are notoriously fragmented and opaque; as a result, there are few

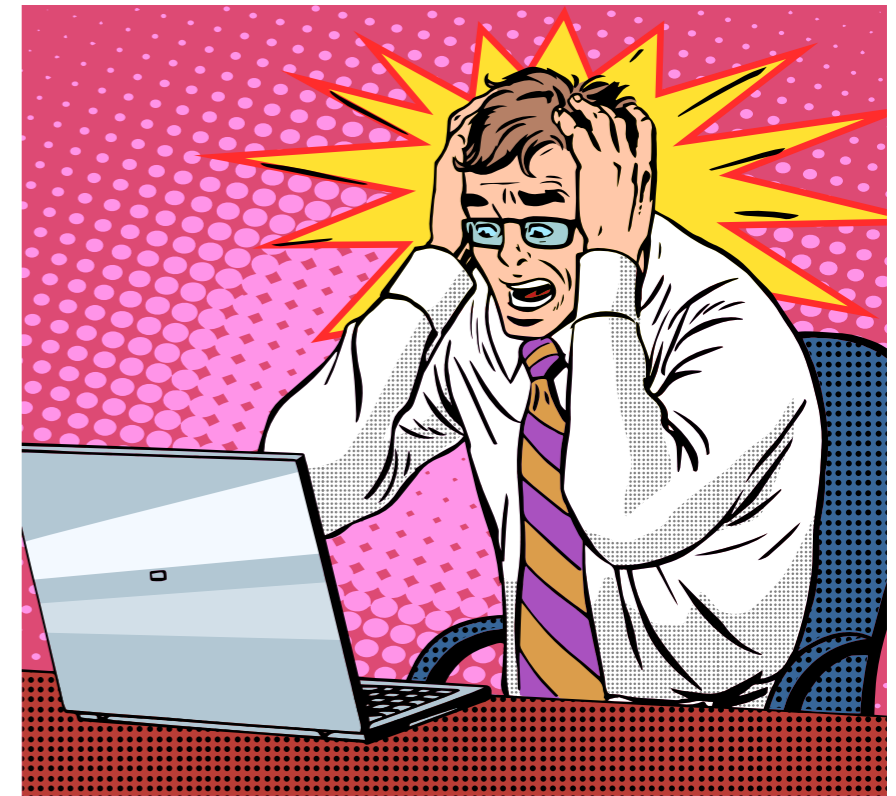
providers of truly comprehensive, accurate and reliable FX data.

Faced with these constraints, most participants decide to double down on the part that is in their control, namely the algorithms and models and technology. Ironically, this means that the potential advantages of a data-first policy are even greater in the FX market than in other asset classes. And FX firms are finally waking up to this fact, and committing themselves to data.

**WHAT DOES IT MEAN: TO BE COMMITTED TO DATA?**  
Most analysts, when asked what "good data" means to them, will invoke some combination of accuracy, consistency, completeness, documentation, timeliness and provenance. This answer is fine as far as it goes, but it doesn't go nearly far enough. It's artificially constrained: it assumes that the data is endogenously given, and that the analyst's job is to accept or reject it. That's simply not true anymore.

A truly data-first policy does not ask, "How good is the data I have?". It asks, "What data can I get that addresses my specific needs?". And it constantly re-asks that question. This requires some thoughtfulness. For example: analysts accustomed to no-arbitrage models assume that "the price is the price". But it's not. As Stuart Farr from Deltix writes, "Forex quotes from a given bank differ between buy-side firms according to individual customer characteristics such as credit quality, assets, trading volume and trading style ... There is little point in [analyzing] market data with quotes unattainable to the firm in question!". A deeper analysis would recognize this, and adjust accordingly.

Execution algos depend critically on volume. But "true" volume is one of the hardest things to determine in the FX market. It's easy to see quotes on a screen; it's harder to know the depth of liquidity that underlies these quotes, or the volume of transactions that are happening in real-time across all venues, or the types of participants behind those transactions. A new generation of data sources such as CLS actually have this information, comprehensive and in real-time, but



Performing TCA on flawed data will inevitably yield flawed results

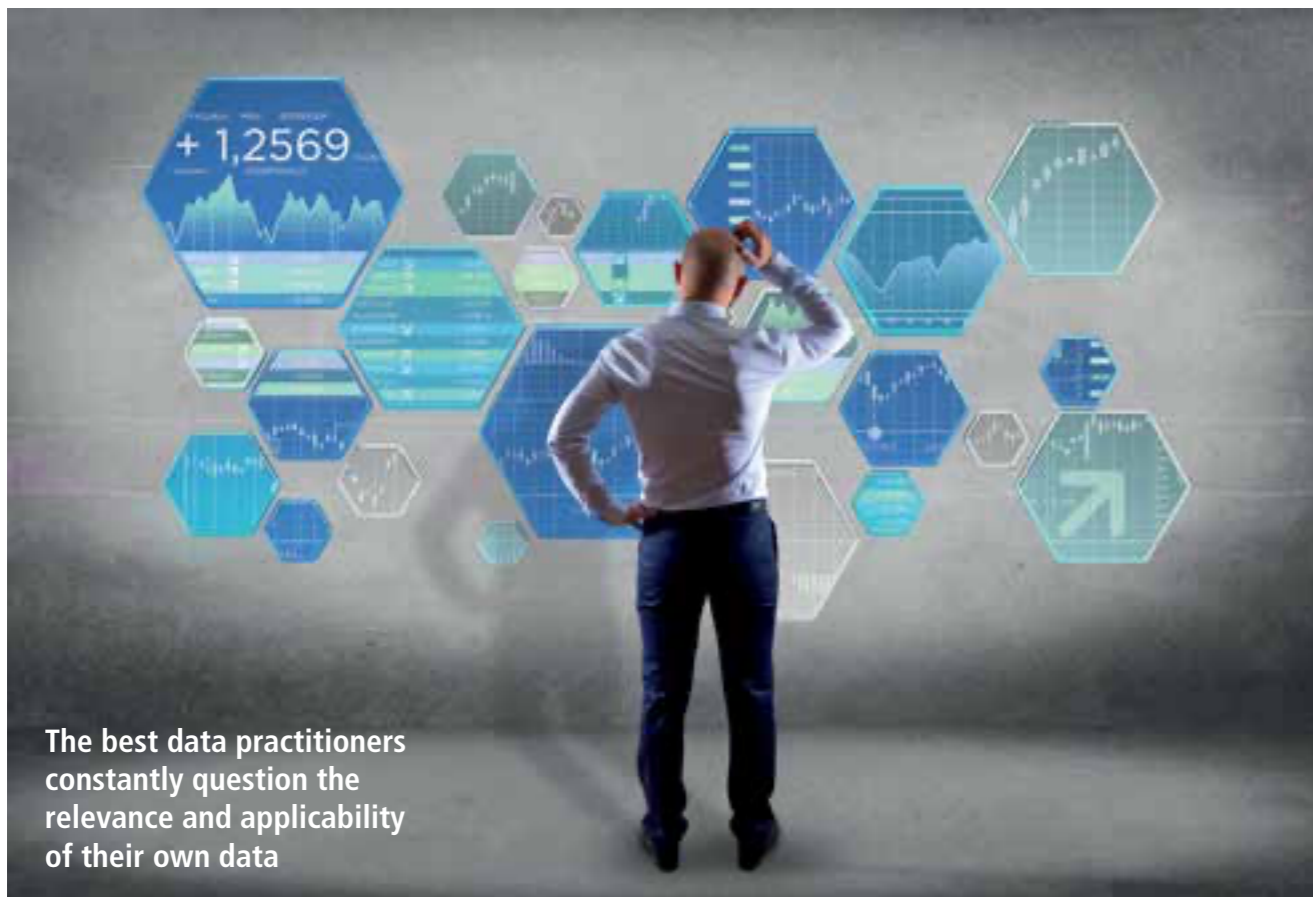
only the most data-progressive firms are using it.

Another volume-based example comes from TCA. It's common to use VWAP to estimate execution costs post-trade; but most commercially-available VWAP datasets are based on less than 1% of the volume in the market; furthermore, they multiply prices and volumes in aggregate, not on a per-trade basis. Performing TCA on flawed data will inevitably yield flawed results; yet many vendors continue to do so.

**PRECISION WITHOUT ACCURACY**  
These examples serve to illustrate an important point when it comes to data: the danger of precision without accuracy. It's all too easy to assume that your data is best-in-class, but doing so creates a confidence that is illusory. Just like the best quants constantly second-guess and double-check and stress-test their models, and the best algo traders constantly try to game their own execution systems, the best data practitioners constantly question the relevance and applicability of their own data, and ask if there are other, better sources out there for the insights they seek.

And this brings us to another key attribute of modern data practice: flexibility. Data sets evolve, newer ones outpace older ones; yet most practitioners are not set up to upgrade. It's easier for most firms to swap in a new model or a new algorithm than it is to swap in a new data source. Fortunately, modern data providers like Quandl are solving this problem for them by providing single-API access to source-agnostic feeds and light web-based delivery tools.

**CONCLUSION**  
All in all, we're seeing a sea change in the way FX participants of every stripe are treating data. Good firms are committing to data like never before: they've realized that resources spent on data underpin all their investments in other areas. And the very best firms have moved beyond this: they've begun to treat data, not as a cost to be minimized, but as a genuine profit center. Data advantages are real: by investing consistently and substantially in data, the best firms ensure that they are always above average – whether in trade selection, execution, or risk management. Now that's something that would not be out of place in Lake Wobegon!



The best data practitioners constantly question the relevance and applicability of their own data