Ignorance is Power

Judge a man by his questions rather than his answers.

*Pierre-Marc-Gaston, Maximes et réflexions sur différents sujets de morale et de politique (*not *Voltaire)*

In our first, short briefing document, *A Dark Glass Brightens*, we spoke of the value of being able to say when a signal is predictable and when it is not, so that we could focus on forecasting when we had reasonable confidence that forecasting was in fact possible. In this new note we begin to explain how Cognitive Trading exploits this knowledge – and how recognising ignorance is also key.

When considering the predictability of a signal – such as a stock price, index or forex rate – one question usually goes unasked: what do we mean by *the signal*, as opposed *the noise* which one usually works hard to ignore? Simply put, the answer is that a signal is something that signifies, something that has meaning – and behind that naïve description lies a valuable insight: *noise* is that that which does not signify and is therefore a measure of our ignorance. But as soon as we recognise that so-called noise merely signifies something else – ignorance – we also see that measuring, predicting etc. noise is no less important than the analyses of the signal.

This is not really news. Noise is *defined* in statistics as that part of a signal not explained by the independent variables, and if we add more variables we might reduce the noise. There are two problems with doing that: we may not know how to include extra variables, or we do, and the model then becomes too complex to manage. Either way, it is usually taken for granted that there is a theoretically or practically irreducible element of noise in any signal of interest and the best we can do is focus on what we can analyse and ignore the rest.

The first stage in any signal processing pipeline therefore tends to be a signal conditioning stage in which algorithms (e.g. moving averages) and neural-networks (e.g. de-noising auto-encoders) are used to “clean” *the signal* by discarding *the noise* – thereby depriving the modeller of the opportunity to make use of this *signal-of-another-kind*.

A more sophisticated approach may be to *stationarise* the signal, i.e. transform it in such a way as to ensure that certain statistical properties, such as the mean and variance, are constant so that specialised statistical analyses, such as Granger Causality (about which, more later) can be performed. But, either way, noise is generally considered as little more than an impediment to effective forecasting of “the signal”, although as we shall see it can be exploited in far more sophisticated ways.

Forecasters then look for patterns in what is left after the noise has been removed – from simple trends, through cycles to complex curve fitting, Markov processes and so on; and as machine learning (both algorithmic and neural-network based) has advanced, forecasters have been able to make fewer and fewer assumptions and to tease out more useful information from their raw data.

So, given the power of neural networks and Deep Learning, why not just let the system work out for itself what is significant (and when it is significant)?

The answer is again a combination of theoretical and practical difficulties. The amount of financial information available for training price forecasting systems is limited, and there is no reason to expect that even a perfectly capable machine learning system would in fact have enough to learn what does and does not signify, and to what extent – especially given the fact that there is no guarantee that what the machine learning system is trying to learn is not changing faster than back-propagation can keep up with. And given that there is no known way to architect such a perfectly capable system, it should not be at all surprising that, whilst many forecasting systems may be surprisingly good, they are typically not good enough – even before considering the third key consideration in the effective use of Deep Learning for financial trading: the time component.

The question of time manifests simply: what is the optimum trading strategy based on simple price forecasting, given that a profitable trade now might negate the possibility of making a more profitable trade some time later? The answer to that again concerns the way we process raw financial signals into significant features for neural-network forecasting and Deep Learning strategising.

At Cognitive Trading we separate our input signals into a variety of components with different timescales and treat them all equally, knowing that we don't know how to tell "signal" from "noise". We leverage that ignorance into new basic knowledge based on standard analyses and forecasting methods, such as Long-Short Term Memory (LSTM) neural networks, and generate additional signals that encode the results of meta-analysis, such as,

- How predictable is the coming and going of predictability?
- How do signals behave before, during and after each kind of period?
- How do markets behave after genuinely unpredictable fluctuations?

and so on.

We keep the neural network and Deep Learning systems manageable by calculating for them things that they might in principle be able to work out for themselves – but only given more data and time than is actually available.

We slice each currency pair time series into a set of frequency bands, subtract each from the input series to derive a relative noise signal, create time-series of statistical measures from them, take the derivatives of each input and derived time series, and, most significantly, apply our Ξ (pronounced, "Xi") intra-series causality function to create a time-series of predictability measures. And then we cross-correlate these signals with each other to obtain the Granger Correlations coefficient and parameters that tell us how much knowledge of each contributes to better forecasts of the others.

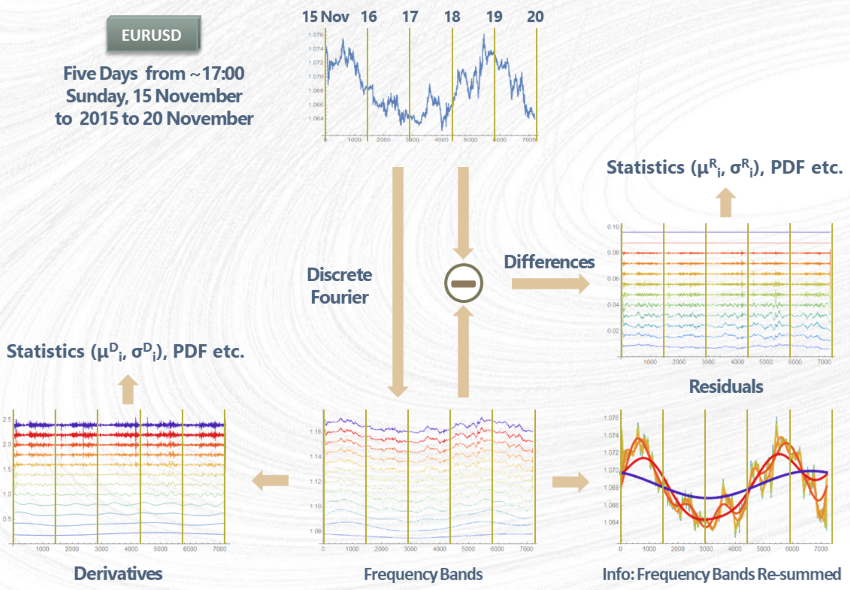

Figure 1 – Signal Processing

Each of these then becomes a feature – part of our "currency tensor" – that we feed into LSTM networks (or similar, such as the recent Clockwork neural network) to obtain forecasts of each feature over a range of timescales that the Deep Learning block can then play a trading game with.

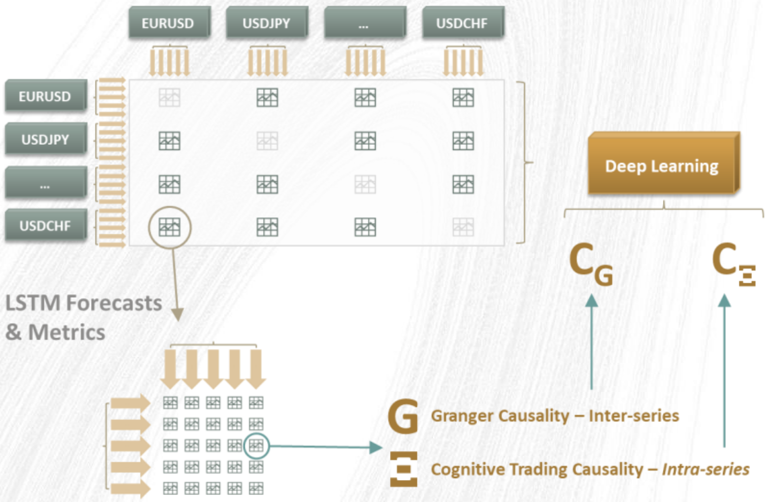

Figure 2 – The Cognitive Trading Currency Tensor

The Deep Learning block continues to explore randomly selected trades so that as circumstances change over time, it can update itself accordingly and, by repeatedly replaying historical data we can bootstrap our limited data into an effective corpus of training data.

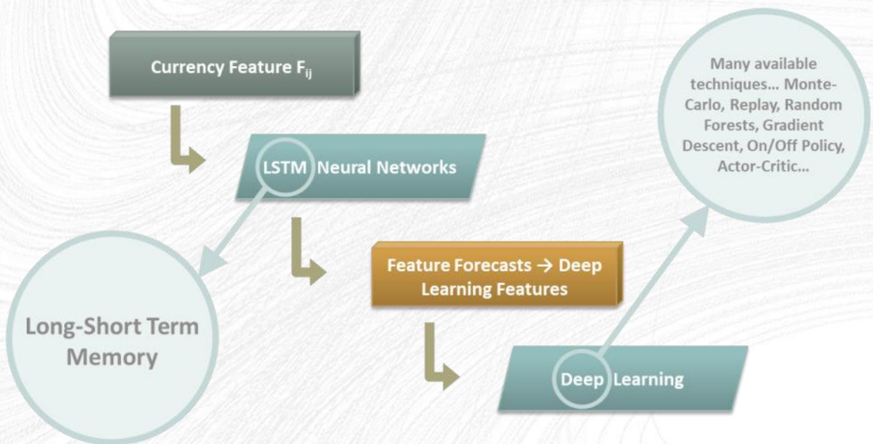

Figure 3 – Forecasting and Trading Architecture (simplified)

In the final sections of this note, we provide some additional details on particular aspects of our approach and methods

Causality

When forecasting time-series, such as forex rates, with neural-network architectures such as LSTM, one can improve the accuracy of forecasts for a target signal by including other signals as model *features.* For example, in a multivariate LSTM model of air quality, one could include as an extra feature the weather conditions rather than just track the air quality measure alone – on the reasonable assumptions that e.g. hot, dry weather causes dust to be raised and wetter weather causes it to be washed out.

However, the question for modellers is always *what other features should be included*? One could include any number of extra features in a model and assume – reasonably – that, given sufficient training data, processing power and time, the model will learn what is relevant and when, and give the best possible forecast.

But it may be difficult – if not actually impossible – to determine a priori how much training data etc. is needed to obtain an accurate model, and it is always best to include as additional features only those things likely to be effective in improving forecasting; but how might one determine *that*?

Fortunately, there is a standard statistical test called *Granger Causality* that allows one to calculate the extent to which knowledge of the history of one time series improves the predictability of another[1]. In the context of forex time series, this might mean that if the USDGBP rate significantly leads the USDCAD rate (in a statistically way), including the amount by which USDCAD *lags* USDGBP (and how strong the Granger Causality is at that time) as features of an LSTM model provides it with a strong indicator it can use to condition the USDCAD forecast – and this is more efficient and more accurate that trying to train the LSTM model to work this out for itself.

Cognitive Trading's new Ξ ("Xi") function complements Granger Causality by providing a measure of *intra*-series causality, in contrast to the Granger Causality which measures *inter*-series causality.

Ξ Causality is based on mathematical research in to the detection of chaotic behaviour without calculating Lyapunov exponents, and Cognitive Trading has developed the published approaches to deal with non-uniformly sampled time series (such as pip data) and to normalise the data so that it is not misled by known in-sample trends and cycles.

Multi-scale Trading with Deep Learning

There is no such thing as absolute unpredictability: even chaotic systems with strange or complex attractors are predictable to the extent that the system is emergently constrained to remain on the attractor, and even if one cannot accurately predict the state of system arbitrarily far into the future, in principle one can always predict – within quantifiable margins – the state of the system over short timescales. The problem is the medium term: a rate may be volatile, but there may still be an underlying pattern such that we cannot say, with any confidence what the rate might be in five minutes time, but – and this is key – with our approach we may be able to say with confidence that in an hour's time the rate will be above or below a certain threshold, and this is information that the Deep Learning system will exploit to maximise the expected return over all timescales under consideration

For further information contact: Volker Dischler, CEO +49 1523 423 76 33

Cognitive Trading: *We think, before we trade*

[1] Strong *Granger Causality* A → B does not mean that A causes B, because both A & B might be caused by some third factor, C and respond over different timescales; this is why statisticians say that A *Granger causes* B, rather than A (intrinsically) *causes* B, when the Granger Causality coefficient his high.