Lost in noise? Some thoughts on the use of machine learning in financial market risk measurement

Received (in revised form): 5th October, 2023

Peter Quell
Head of Portfolio Analytics, DZ Bank AG, Germany

Dr Peter Quell is Head of the Portfolio Analytics Team for Market and Credit Risk in the Risk Controlling Unit of DZ Bank AG in Frankfurt. He is responsible for methodological aspects of internal risk models, economic capital and model risk. He holds a MSc in mathematical finance from Oxford University and a PhD in mathematics. Peter is a member of the editorial board of the *Journal of Risk Model Validation* and a founding member of the Model Risk Managers' International Association (mrmia.org).

DZ Bank AG, Deutsche Zentral-Genossenschaftsbank, Platz der Republik, 60325 Frankfurt am Main, Germany
E-mail: peter.quell@dzbank.de

Abstract Machine learning has permeated almost all areas in which inferences are drawn from financial data. Nevertheless, in financial market risk measurement most machine learning techniques struggle with some inherent difficulties: Financial time series are very noisy, not stationary and mostly considerably short. This paper contains an easy to implement sequential learning algorithm that overcomes some of these disadvantages. It is based on a Kalman filtering mechanism for quite general stochastic processes and provides a first step in the direction of separating parameter dynamics from the ubiquitous noise component. The core idea here is to use some stylised facts inherent to financial markets time series such as time varying measures of volatility. The new approach is tested using real market data in two different settings. First, a hypothetical portfolio containing credit spread and equity risk is analysed over a time frame containing the outbreak of the global pandemic in 2020 and the beginning of the Russian attack on Ukraine in 2022. Another analysis is focused on US$/EUR exchange rate during a time span containing the global financial crisis of 2008 and the subsequent European sovereign crisis. In all test calculations the proposed sequential learning algorithm performs better than the historical simulation approach used by many firms in the banking industry to meet regulatory capital requirements. Due to its simplicity this method has a high degree of explainability and interpretability which will decrease the inherent model risk. The paper concludes with a discussion of model risk for machine learning in financial institutions. Compared to classical model risk frameworks, the emphasis must be put on the more prominent role of data. The simple approach described in this paper shows that machine learning in financial market risk does not have to get lost in noise.

Keywords: *market risk, machine learning, Basel regulation, Kalman filter, adaptive methods, model risk*

INTRODUCTION

Machine learning has permeated almost all areas in which inferences are drawn from data. The range of applications in the financial industry spans from credit rating, loan approval processes in credit risk to automated trading, portfolio optimisation and scenario generation for market risk. Machine learning techniques can also be found in fraud prevention, anti-money laundering, efficiency/cost control and marketing models. Machine learning has demonstrated significant uplift in these business areas and the use of machine learning will continue to be explored in the financial industry.

Nevertheless, there is one area in which machine learning has not (yet) contributed too many innovations: *Time series analysis for financial market risk measurement*. The main reasons for this are rooted in the observation that financial time series are very noisy, not stationary and mostly considerably short. Therefore, traditional machine learning algorithms simply do not find enough data to draw any relevant conclusion.

Here is an example. During the start of the COVID-19 pandemic in the first half of 2020, many banks experienced large numbers of backtest outliers when they compared actual profit and loss numbers to value at risk (VaR) estimates in the trading book. The simple reason was that their regulatory VaR systems were not able to adapt to rapidly changing market conditions as volatility spiked to ever-higher levels. The problem lies with the amount of data that is used to 'train' the VaR system.

In a nutshell, this amount of data is a compromise between having enough observations to compute relevant statistical quantities (here we want lots of data) and the degree to which this data is still relevant for the current environment (here we only want recent data). Whereas in quiet times collecting many data points to reduce measurement uncertainty seems to be a priority in some model risk management approaches, the situation changes once there is a rapid shift as experienced in March 2020.

This paper contains an easy to implement algorithm that overcomes some disadvantages of currently used regulatory systems to capture financial market risk. For that purpose, we first have a look at the basics of regulatory market risk measurement and machine learning. Based on these insights, we propose a method that can easily be added to existing risk measurement frameworks to cope with rapidly changing market conditions. This method is tested using real market data and finally some aspects of model risk management will complement the discussion.

MARKET RISK MEASUREMENT

Standard ways to measure regulatory market risk

What is market risk all about? Market risk is the risk of changes in the value of a given portfolio of financial instruments due to the movement of risk factors like stock prices, interest rates, foreign exchange rates and commodity prices. The focus here is on *short term risk*, ie changes over a risk horizon of up to a few trading days. It is part of the risks that need to be capitalised by financial institutions subject to the suite of regulations contained in Basel Pillar one and Pillar two.[1] In brief, regulatory market risk is measured by evaluating a portfolio under different potential future scenarios. Every scenario will imply a potential future portfolio value which delivers a probabilistic description of the future behaviour of the portfolio. This will then lead to a risk measure that will be used to determine the regulatory or economic capital of the financial institution.

Currently there are different approaches that are used by banks within the internal model framework to measure market risk. Essentially, these methods only differ in the way they identify the potential future scenarios.

Historical simulation approach: The potential future scenarios are described by changes in risk factors that have been observed in the past. There might be differences in the way these shifts are computed (relative shifts for equities and foreign exchange rates, absolute shifts for credit spreads), but, in essence, historical information is directly transferred into the scenarios used to compute market risk. This approach seems to resemble the slogan 'Let the data speak for itself' often encountered in machine learning frameworks.

Monte Carlo simulation approach: The potential future scenarios are sampled from given probability distributions instead of the empirical distribution. The historically observed data is used to calibrate the corresponding parameters of the probability distribution. In principle, a scenario generator can produce an arbitrarily large amount of shift scenarios for the risk factors, which makes this approach more flexible than the historical simulation. Nevertheless, the focus on a certain form of parametric families of distributions for the risk factors will lead to its own model risk aspects.

Parametric approach: The distributions of potential future risk factor shifts are described by parametric families (usually normal distributions, student t-distributions or hyperbolic distributions). The corresponding profit and loss distribution of the portfolio under consideration is then derived under assumptions that guarantee that this distribution remains within a certain parametric family of distributions.

Extreme value approaches: Within the scenario generator there is an explicit distinction between the body of the profit and loss distribution (the region where 'normal' things happen) and the tails of the distribution (the region where 'extreme' things happen). However, extreme value approaches are not so common in practice since it is usually controversial where the transition between the body and tail of the corresponding distributions is located — a feature that will also introduce additional model risks.

Stylised facts of financial time series data

But why should the above-mentioned data driven approach to measure market risk work at all? Some evidence comes from the analysis of so-called stylised facts of financial markets time series. In a nutshell, stylised facts relate to statistical properties of time series data observed across a wide range of financial instruments, markets and time periods. The considerations of this section are mostly based on the work of Cont and the references therein.[2]

Stylised facts are essentially qualitative statements by nature, mostly obtained on an empirical basis. They reveal certain characteristics of large amounts of data — they belong to the realm of 'big data'. The main aspect is therefore to let the data 'speak for itself'. Although this non-parametric approach does not depend on specific models for the data-generating process, there is the shortcoming of not having a theory to guide further conclusions.

As the study of stylised facts will integrate different financial instruments, markets and time periods *per se*, the statements gain in generality but obviously lose in precision. Therefore, it will be observed time and again that stylised facts do not help the modeller to identify a unique data-generating process. On the other hand, these stylised facts are so constraining that in general it is very hard to construct any model of a (parametric) data-generating process that shows all these stylised facts simultaneously.[3] After these general remarks, the focus now moves to a few stylised facts that are particularly important within the context of market risk measurement.

Absence of linear autocorrelations: In liquid markets, the linear autocorrelation function of price changes often decays rapidly as a function of the corresponding time lag. In general, for medium-sized lags in the range of one hour or greater, the autocorrelations could be assumed to be zero. Of course, the corresponding price must be liquid, otherwise it makes no sense to talk about autocorrelation over a time lag of a few minutes.

Volatility clustering: Although there is only little evidence of linear autocorrelation in time series of price changes, this does not mean that these changes are statistically independent. If absolute values are studied, there is very strong evidence of (positive) autocorrelation. Roughly speaking, that means large absolute price changes are likely to be followed by other large absolute price changes, a feature that is often referred to as 'volatility clustering'. Time series models such as these are extensively studied by Box *et al*.[4]

Fat tails, extreme events: Empirical evidence from many different markets and time periods indicates that fat tails in distributions of price changes are the norm and not the exception. This result even holds for conditional distributions — for distributions where the effects of volatility clusters (see above) have been controlled. Nevertheless, fat tails seem to be more pronounced in unconditional distributions. Since fat tails characterise the behaviour of the probability distribution for extreme realisations, they are often associated with extreme events. Although on a qualitative basis this is quite intuitive, due to the rareness of extreme events it is much harder or even impossible to quantify the probability of occurrence of such an extreme event.

The stylised facts mentioned so far are associated with financial time series data. But what about alternative data like data form social media platforms, geo data from satellites or data from sensors embedded in different products? Is there a potential to use this kind of data for financial risk forecasts? Usually, alternative data comes with short data history (maybe a few years to a decade) which may make a reliable backtest difficult. Nevertheless, it may complement the classical financial time series when it comes to time series analysis.

Up to now there has not been any conclusive evidence that this kind of alternative data may yield a sustainable improvement of financial market risk forecasts.[5] Nevertheless, the current M6 competition examines the extent to which the so-called efficient market hypothesis[6] can be aligned with the observation that there are participants in the financial markets who seem to deliver very good investment results over long periods of time. The results of this competition are expected in 2024. It remains to be seen how far alternative data could be used to optimise financial time series analysis.

USING MACHINE LEARNING IN FINANCIAL MARKET RISK

What are the main obstacles?

To what extent are machine learning techniques applicable to financial markets data? Even more importantly, since a lot of machine learning algorithms are quite time consuming when it comes to the integration of new data, can we use these techniques when we have to come up with a timely market risk report on each and every trading day?

There are indications that (current) machine learning techniques that have been developed for time series data are not quite reliable when it comes to the application in market risk measurement.[7] The first objection relates to the inherently *low signal-to-noise ratio in financial market data*. To put it another way, time series data used for market risk measurement usually come with a lot of noise that may lead machine learning techniques astray. Some machine learning methods are in danger of overfitting and therefore have large issues with noisy data like time series in financial markets.

The second objection relates to ever-evolving financial markets ie machine learning algorithms must be prepared for a concept drift in the learning environment. This is an area where machine learning techniques may improve in the future.

The third objection relates to the *availability of training data*. Whereas in credit risk, rating assignments rely on data collected over a huge time span (several years or even decades), market risk measurement must cope with the fact that even recent data may no longer be so relevant. In the context of market risk measurement, we should not focus on machine learning techniques designed for 'big data'. Instead, we must think about how to adapt to new situations in a fast and efficient way. This special case of machine learning runs under the name of sequential (online) learning. The focus here is not so much on processing huge amounts of data, but on the efficient usage of new information as it arrives. The next section turns to a prototypical example that provides a promising starting point.

Parameter adaptation in a noisy environment

As mentioned in the introduction, market turmoils during the early phase of the COVID-19 pandemic in 2020 have once again highlighted weaknesses of regulatory VaR estimates when it comes to changes in financial market environments.[8] Although potential remedies have been available for some time, the concrete implementation of solutions in running VaR systems seems to be more challenging. In this section we like to consider a *simple adaptation mechanism* that can be used in existing VaR methods to track changing market environments. Like classical stochastic volatility methods,[9,10] the focus is on the dynamics of the logarithm of portfolio returns.

Consider the following univariate stochastic process for the unknown true parameter

$$\mu_{n+1} = \mu_n + m_{n+1}$$

for $n = 0,1,\ldots$ with initial condition $\mu_0 = 0$ and independent identically distributed increments m_n for $n = 1,2,\ldots$. The increments do not need to be normally distributed; we only require that the corresponding distribution allows for first and second moments. The relevant profit and loss dynamic is described by the stochastic process

$$pl_{n+1} = \mu_n + \varepsilon_{n+1}$$

for $n = 0,1,\ldots$ with initial condition $\mu_0 = 0$ and independent identically distributed ε_n for $n = 1,2,\ldots$. Note that the process ε_n contains the unknown dynamics of the parameter as well as the measurement uncertainty of the profit and loss (therefore the subscript n in μ_n). Both processes m_n and ε_n are supposed to be independent identically distributed, of second order with zero mean and standard deviation $\sigma_m > 0$ and $\sigma_\epsilon > 0$ respectively. The processes m_n and ε_n are correlated with covariance $\sigma_{\epsilon m}$ for $n = 1, 2, \ldots$.

But how does the actual algorithm work once we are given specific measurements? Based on the noisy profit and loss measurements $PL_1, \ldots, PL_{n+1}$ an estimator is given by

$$M_{n+1} = M_n + \lambda\left(PL_{n+1} - M_n\right)$$

for $n = 0, 1, \ldots$ with initial condition $M_0 = \mu_0$ and a parameter $\lambda > 0$. The idea behind this update is quite intuitive. We are starting from the most recent estimate M_n at time t_n and correct for the difference between this quantity and the new measurement PL_{n+1} at time t_{n+1}. So, for example, if the new measured PL_{n+1} is higher than our last best guess M_n, we need to increase our best estimate M_{n+1} for time t_{n+1}.

How to choose the parameter $\lambda > 0$? Since what we really want is the parameter μ_{n+1} at time t_{n+1} it would be a good idea to focus on the estimation error $E_{n+1} = \mu_{n+1} - M_{n+1}$. Using the definitions of the stochastic processes from above yields

$$E_{n+1} = (1-\lambda)(\mu_n - M_n) + m_{n+1} - \lambda\varepsilon_{n+1}$$

such that the mean value $e_{n+1}^2 = E\left(E_{n+1}^2\right)$ is given by

$$e_{n+1}^2 = (1-\lambda)^2\left(e_n^2\right) + \sigma_m^2 + \lambda^2\sigma_\epsilon^2 - 2\lambda\sigma_{\varepsilon m}$$

where $\sigma_m > 0$ and $\sigma_\epsilon > 0$ are the standard deviations of the innovations and $\sigma_{\epsilon m}$ is the corresponding covariance. The condition $\lambda < 2$ is necessary to avoid explosion of the error terms. In this case the asymptotic error is

$$e^2(\lambda) = \frac{\sigma_m^2 + \lambda^2\sigma_\varepsilon^2 - 2\lambda\sigma_{\varepsilon m}}{\lambda(2-\lambda)}$$

The parameter λ can now be specified to obtain a minimal error term $e^2(\lambda)$. Please note that due to the restriction $\lambda > 0$ we only consider one of the roots of $\frac{d}{d\lambda}e^2(\lambda) = 0$. We have

$$\lambda = -\frac{\sigma_m^2}{2\left(\sigma_\varepsilon^2 - \sigma_{\varepsilon m}\right)} + \frac{\sigma_m}{\sqrt{\sigma_\varepsilon^2 - \sigma_{\varepsilon m}}}\sqrt{1 + \frac{\sigma_m^2}{4\left(\sigma_\varepsilon^2 - \sigma_{\varepsilon m}\right)}}$$

which can also be written as

$$\lambda = -\frac{1}{2}\ snr^2 + snr\sqrt{1 + \frac{1}{4}\ snr^2}$$

with the signal-to-noise ratio snr such that

$$snr^2 = \frac{\sigma_m^2}{\left(\sigma_\varepsilon^2 - \sigma_{\varepsilon m}\right)}.$$

Since, in general, the noisy process has increments with a larger standard deviation than the process without noise, we have $1 < \sigma_\varepsilon / \sigma_m$ such that the signal-to-noise ratio is well defined in this case. In Figure 1 we see the asymptotic development of the parameter λ as the signal-to-noise ratio tends to infinity.

As a final remark please note that in the above expression defining the optimal parameter λ only

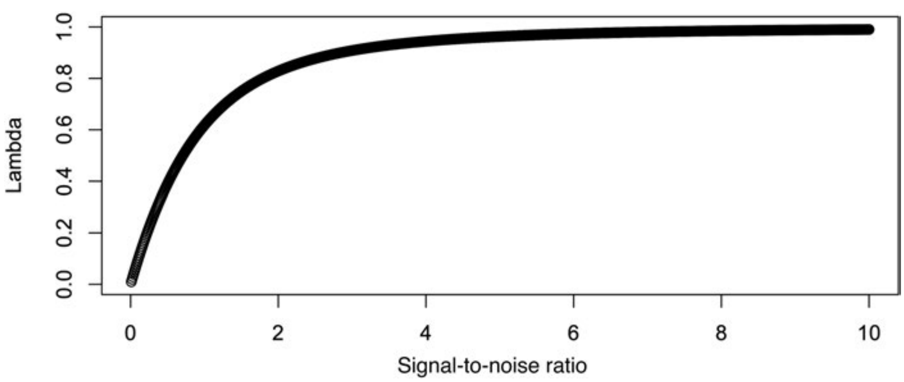

Figure 1: Asymptotic behaviour of the model parameter λ with different signal-to-noise ratios

the signal-to-noise ratio *snr* is relevant; we do not necessarily need the individual values for the standard deviations $\sigma_m > 0$ and $\sigma_\epsilon > 0$ or the correlation coefficient $\varrho_{\varepsilon m}$. One can see the crucial effect of the signal-to-noise ratio; low values of *snr* indicate calm times in which there is no need to pronounce recent observations. As values for signal-to-noise ratio increase, recent observations yield a much bigger influence in estimating the current state.

Extension to more general stochastic processes

The approach described in the last section is easily extended to include noisy auto-regressive time series models instead of noisy random walks; in this case, the optimality criterion can also be derived analytically. If, for example, the parameter would be described by the stochastic process

$$\mu_{n+1} = \alpha \mu_n + m_{n+1}$$

with $0 < \alpha \leq 1$ then the optimality criterion for the parameter λ in

$$M_{n+1} = \alpha M_n + \lambda_\alpha \left(PL_{n+1} - \alpha M_n\right)$$

would be given by

$$\lambda_\alpha = -\frac{1}{2}\ snr^2(\alpha) + \sqrt{snr^2(\alpha) + \frac{1}{4}\ snr^4(\alpha) - \frac{1-\alpha^2}{\alpha^2}}$$

with

$$snr^2(\alpha) = \frac{\alpha^2 \sigma_m^2 + \left(1-\alpha^2\right)\ \sigma_\varepsilon^2}{\alpha^2\left(\sigma_\varepsilon^2 - \sigma_{\varepsilon m}\right)}.$$

This expression reduces to the form derived earlier for the noisy random walk $\alpha = 1$. Quell and Meyer give details concerning the derivation of this optimality criterion.[11]

Even more complex generalisations of the stochastic processes describing the temporal evolution of the parameter μ_n can be handled using state space approaches and the Kalman/Stratonovich Filter.[12–14]

Conclusion 1: It became clear that changing volatilities observed in financial markets time series data require a rapid adaptation of sequential machine learning methods used for market risk measurement. Under the assumption of quite general stochastic processes describing these parameters, efficient filter mechanisms can be found that minimise the tracking error within noisy time series data. In a more informal way, the dynamics of the time varying parameters and their estimates M can be thought of as the signal that has to be identified within noisy data.

IMPROVING REGULATORY MARKET RISK MEASUREMENT

Definition of a benchmark risk model

How can the stylised facts together with the sequential learning from the previous sections be used to improve regulatory market risk measurement? Hull and White suggest standardising risk factor changes using an estimate of the current volatility.[15] Barone-Adesi and Giannopoulos[16] extend the Hull and White approach and examine the behaviour of this method for individual financial instruments.

When it comes to a practical implementation, the limiting factors of these approaches are:

1. They operate on a risk factor level, ie in practice one would have to calibrate several hundreds of parameters to allow for an optimal tracking of risk factor dynamics.
2. If banks like to use these techniques within their regulatory market risk measurement, they usually need to modify large parts of their risk model software.

To overcome these issues, this section contains the description of a benchmark risk model that can be placed 'on top' of already existing regulatory risk models in banks. Since the focus is on the profit and loss of complete portfolios, they require a minimal amount of calibration effort.

To fix terms, let the profits and losses of the portfolio under consideration be given by $PL_1, \ldots, PL_N$. These numbers may have been produced by any of the risk models already mentioned. The crucial fact is that there is a natural order to these scenarios: PL_1 precedes PL_2, and so on until the most recent PL_N.

This feature will allow for the adaptation to changing parameters.

Let M_n be an estimate for the parameter at time n, then the normalised profit and loss QL_n is defined for $n = 0,\ldots,N-1$ by

$$QL_{n+1} = PL_{n+1} - M_n$$

At this point, it is important to note that there is no typo in the above formula. The term QL_{n+1} should contain the 'surprise' that QL_{n+1} contains relative to the knowledge of M_n from the previous time step. While the classical historical simulation approach would use the profits and losses $PL_1,\ldots,PL_N$ to estimate the risk measure, the *adaptive benchmark model* will use

$$RL_n = M_N + QL_n$$

for $n = 1,\ldots,$ N as the profit and loss distribution based on the current parameter estimate M_N.

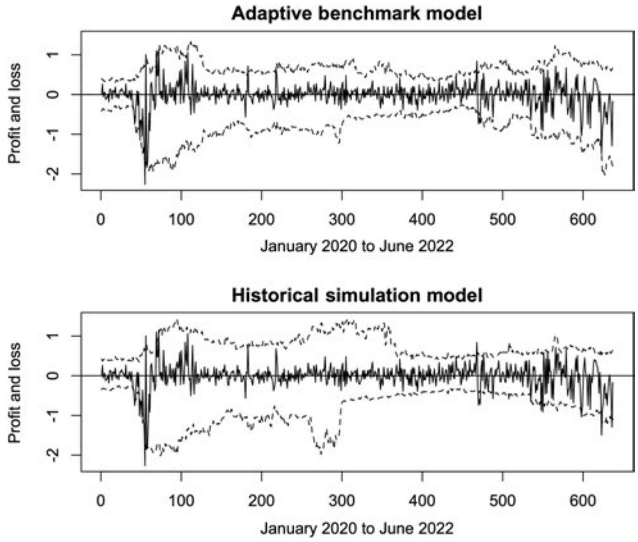

Figure 2: Adaptive benchmark model and historical simulation for a hypothetical portfolio

Application to financial market data

To see how the adaptive benchmark model could improve current regulatory methods for financial market risk estimation, we focus on the rescaled profit and loss characteristics of a hypothetical portfolio between January 2020 and June 2022. This time frame covers a period of increased economic uncertainty due to the pandemic that started in the first quarter of 2020 and the Russian attack on Ukraine in February 2022. The portfolio consists of credit spread risk (government and corporates) as well as equity risk.

Figure 2 shows the results of daily market risk estimations during this period based on the adaptive benchmark model and the historical simulation approach. In both diagrams, the solid line is the backtest profit and loss time series for the hypothetical portfolio under consideration. The dashed line indicates the daily update of the 1 per cent and the 99 per cent quantile of the profit and loss distribution. In terms of the discussion from the last section, the profit and loss distribution should be used to track the backtest profit and loss form of the hypothetical portfolio. At first glance, both methods seem to work quite well tracking the characteristics of the time series. Nevertheless, there are some important differences.

In both diagrams, the lower dashed line represents the 1 per cent quantile of the profit and loss distribution, a quantity which is called value at risk (VaR) in the Basel regulatory framework. Since it is calibrated on 1 per cent, there should be around 700*1 per cent or around 7 trading days, where this level is breached by the solid line. Such an event is then called a backtest outlier. In this case, the adaptive benchmark model produced 8 backtest outliers whereas the historical simulation approach produced 17 backtest outliers. One immediate cause for this difference is the worse adaptation speed of the historical simulation to increasing market volatilities. In a certain sense, the adaptive benchmark model uses the sequential learning to better keep track of the market dynamics which then results in fewer backtest outliers. But are 8 backtest outliers still tolerable when one expects only 7 backtest outliers in this example? Due to the standard Basel traffic light approach,[17] the 8 outliers of the adaptive benchmark model are still in the green zone of this test (ie will not be rejected on a 95 per cent level) whereas the 17 backtest outliers of the historical simulation approach fall into the yellow zone (rejection at 95 per cent but not rejected at 99 per cent).

To get a better feeling for the behaviour of the two risk models over longer historical time spans, let us look at foreign exchange markets. The log returns of

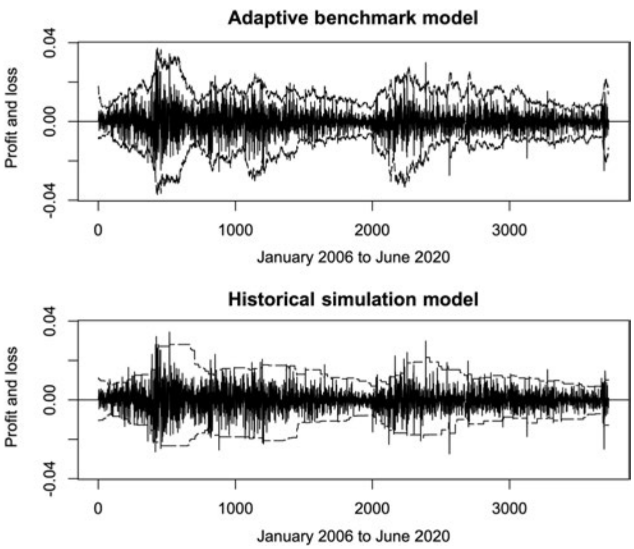

Figure 3: Adaptive benchmark model and historical simulation for US$/EUR exchange rate

the US$/EUR exchange rate from early 2006 until mid-2020 together with the performance of the adaptive benchmark model as well as the historical simulation approach are shown in Figure 3. The impact of the global financial crisis of 2008, the subsequent European sovereign crisis as well as the start of the global pandemic in 2020 are clearly visible.

It is obvious that the adaptive benchmark model follows the risk factor dynamics much better than the historical simulation; this is also reflected in the number of backtest outliers on the 99 per cent confidence level. The historical simulation produces 53 backtest outliers (amber zone of the Basel traffic lights approach[18] with 3724 samples) whereas the adaptive benchmark model stays in the corresponding green zone with 40 backtest outliers.

Conclusion 2: This example shows that sequential learning methods like the adaptive benchmark model have the potential to track the behaviour of a profit and loss quantity based on real financial market data. The profit and loss decomposition

$$RL_n = M_N + QL_n$$

helps us not to get lost in noise, but to use the noise to improve financial market risk measurement. During the construction of the method there is no need to make any specific assumption concerning the distribution of the random variables, the existence of first and second moments will suffice.

MACHINE LEARNING AND MODEL RISK MANAGEMENT

The banking industry is becoming increasingly aware of model risks related to the use of machine learning techniques for risk management purposes. What are the main challenges when it comes to the application of machine learning in a regulatory context? Figure 4 provides a first overview.

Explainability/Interpretability: One should be able to explain how the algorithm makes a prediction or decision for one specific case at a time.

Overfitting: One should recognise that there is some amount of randomness in the training data. If not taken care of, algorithms show good performance on training data — but fail on data not seen before.

Robustness and transient environments: One should account for the fact that markets or environments can change, which calls for a good balance of adaptability and robustness.

Bias and adversarial attacks: Compared to classical statistics there is a much more prominent role for (training) data in machine learning applications.

How should the model risk governance react to these challenges? Here are some questions to guide the validation process.

Model review: If machine learning algorithms frequently change their 'inner workings', how should model validation react? What should be the contents of the validation activity? How should aspects of conceptual soundness be treated?

Model development, implementation and use: How to account for the more prominent role of data? What level of complexity can users handle? What kind of explanations would be accepted by users or by senior management?

Model identification and registration: How to account for model complexity, the role of data, model recalibration within the model inventory?

Excellent quality standards: Existing frameworks need to be enhanced by additional checks for overfitting and sensitivity analysis to test for robustness. Tests for possible bias and discrimination may be reviewed with respect to reputational risk.

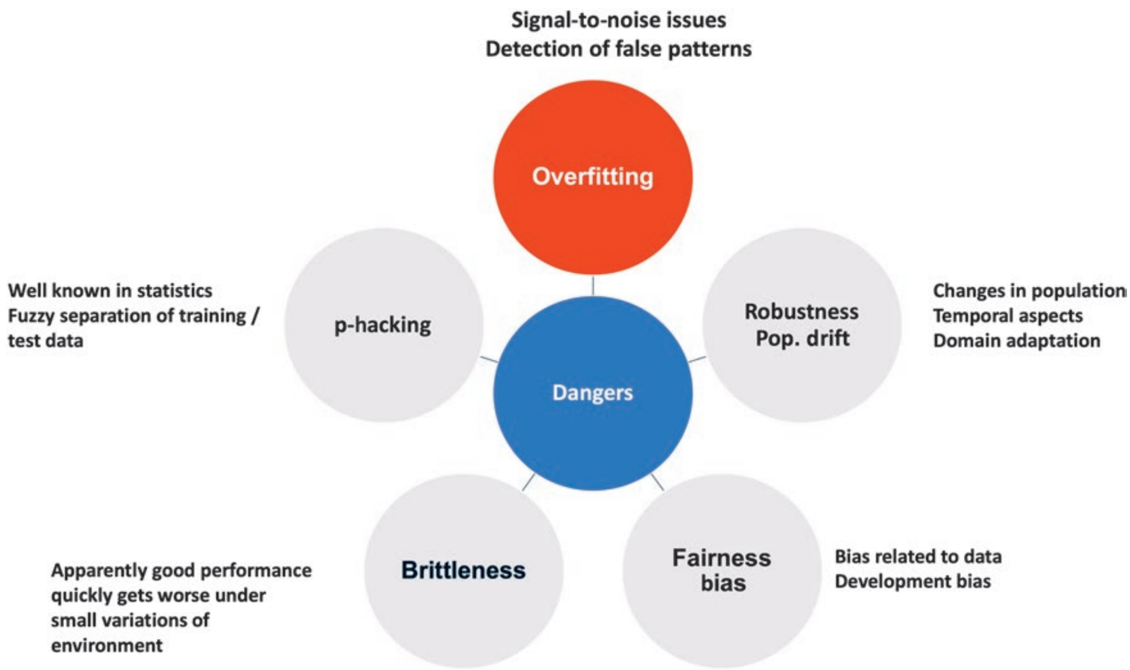

Figure 4: Potential dangers of AI applications

Some banks have already developed frameworks to deal with the model risks of machine learning applications, while other banks are searching for viable starting points. There definitely is a need to share emerging industry best practices and to develop a comprehensive framework to assess model risks in machine learning applications. A good platform for all risk professionals to share their views on model risk and machine learning is the Model Risk Managers' International Association.[19]

CONCLUSION

The preceding considerations have shown that machine learning methods could use the stylised facts within financial markets time series data to come up with real improvements for regulatory market risk measurement. Sequential learning algorithms may take a first step into the direction of separating parameter dynamics (aka signals) from the ubiquitous noise component. Due to their simplicity, these methods have a high degree of explainability and interpretability which will decrease the inherent model risk. Of course, the separation of signal and noise in financial markets time series data will never be perfect. Nevertheless, already the simple adaptive benchmark method described in this paper shows that machine learning does not have to get lost in noise.

Finally, here are some thoughts for further exploration.

Use domain knowledge: Stylised facts provide a first step towards explainability and may indicate which kinds of algorithms will work.

Use sequential learning: A 'sliding time frame' approach may help to some extent, to counter the effects of changing market environments.

Avoid overfitting: The low signal-to-noise ratio within financial markets time series data calls for explicit treatment of the noise component. This may also support the estimation of uncertainty for the point forecasts as well as providing information for the construction of confidence intervals around such forecasts.

Avoid development bias: Try to automate preprocessing and avoid the extra decisions required on the part of the user.

AUTHOR'S NOTE

The views expressed in this paper are those of the author and do not represent the views of any organisations to which he is affiliated.

References and notes

1 Quell, P. and Meyer, C. (2020) 'Risk Model Validation', 3rd edition, Infopro Digital, London.
2 Cont, R. (2001) 'Empirical Properties of Asset Returns: Stylized Facts and Statistical Issues', *Quantitative Finance*, Vol. 1, No. 2, pp. 223–36.
3 *Ibid*.
4 Box, G. E. P., Jenkins, G. M., Reinsel, G. C. and Ljung, G. M. (2015) 'Time Series Analysis: Forecasting and Control', 5th edition, Wiley, Hoboken.
5 Makridakis, S., Spiliotis, E. and Assimakopoulos, V. (2018) 'Statistical and Machine Learning Forecasting Methods: Concerns and Ways Forward', *PLoS ONE*, Vol. 13, No. 3, available at https://doi.org/10.1371/journal.pone.0194889 (accessed 20th September, 2023).
6 Fama, E. F. (1970) 'Efficient Capital Markets: A Review of Theory and Empirical Work', *The Journal of Finance*, Vol. 25, No. 2, pp. 383–417.
7 Israel, R., Kelly, B. and Moskowitz, T. (2020) 'Can Machines "Learn" Finance?', *Journal of Investment Management*, Vol. 18, No. 2, pp. 23–36.
8 Abboud, A., Anderson, C., Game, A., Iercosan, D., Inanoglu, H. and Lynch, D. (July, 2021) 'Banks' Backtesting Exceptions during the COVID-19 Crash: Causes and Consequences', *Federal Reserve Notes*.
9 Shephard, N. (1996) 'Statistical Aspects of ARCH and Stochastic Volatility', in Cox, D. R., Hinkley, D. V. and Barndorff-Nielsen, O. E. (eds), 'Time Series Models in Econometrics, Finance and Other Fields', Chapman & Hall, London, pp. 1–67.
10 Taylor, S. J. (1982) 'Financial Returns Modelled by the Product of Two Stochastic Processes: A Study of Daily Sugar Prices 1961–79', in Anderson, O. D. (ed.), 'Time Series Analysis: Theory and Practice 1', Elsevier, Amsterdam, pp. 203–26.
11 Quell and Meyer, ref 1 above.
12 Harvey, A. C. (1989) 'Forecasting, Structural Time Series Models and the Kalman Filter', Cambridge University Press, Cambridge.
13 Kalman, R. E. (1960) 'A New Approach to Linear Filtering and Prediction Problems', *Journal of Basic Engineering*, Vol. 82, No. 1, pp. 35–45.
14 Stratonovich, R. L. (1959) 'On the Theory of Optimal Non-linear Filtering of Random Functions', *Theory of Probability and its Applications*', Vol. 4, pp. 223–5.
15 Hull, J. and White, A. (1998) 'Incorporating Volatility Updating into the Historical Simulation Method for Value-at-Risk', *Journal of Risk*, Vol. 1, No. 1, pp. 5–19.
16 Barone-Adesi, G. and Giannopoulos, K. (2001) 'Non-parametric VaR Techniques: Myths and Realities', *Economic Notes*, Vol. 30, No. 2, pp. 167–81.
17 Quell and Meyer, ref 1 above.
18 Quell and Meyer, ref 1 above.
19 See www.mrmia.org.

