GALYTIX

Banks Must Act on their Early Warning Systems or Risk ROE Downturn

In association with

pwc

Contents

Executive summary

Today's EWS produce

high false positives

Corporate banks' pure lending ROEs averaging

3-6%

Evolving external forces are creating many risks for corporate banks globally. Macro factors like inflationary pressures, pandemics, frequent climate change related events and heightened geopolitical risk have the potential to disrupt supply chains across industries and impact businesses in unexpected ways.

Corporate banks are now witnessing the end of an era. Following the 2008 financial crisis, banks rebuilt capital and invested in technology to strengthen their relationships with clients through improved efficiencies.

The banking industry now faces a new era which will challenge banks' ability to future-proof their risks and exposures without having any historical evidence or data on how to predict or tackle the risk. Banks can no longer depend on their current Early Warning Systems (EWS) using backward looking indicators with high false positives, to protect them from future risks. Previous methods and data sources for assessing credit risk or identifying distress signals are quickly becoming out of date. Banks have a limited window of opportunity to adapt and respond to the emerging scenarios.

Although large systematic shocks are unpredictable, it is clear that the speed at which banks are able to react to these events can have a huge effect on the ability to actively manage the lending book and reduce any negative impact. The same can be said about the ability to have early warnings of more common "everyday" market movements and events too.

Early Warning Systems (EWS) to assess credit risks are critical to maximising corporate bank returns and shareholder value

Banks are struggling to generate returns in their corporate lending books. Galytix estimates that ROEs on a standalone basis without the support of investment and transaction banking business are 3-6%. This is in benign credit markets, where default rates are at near record lows. Taking a through the cycle or recession scenario default rates would reduce these ROEs and, in many cases, turn them negative without management actions. Credit risk costs for banks are likely to rise because of increased loan loss provisioning and a deterioration in credit portfolio quality.

External stakeholders – from investors and regulators to media and activists – increasingly will be seeking reassurances that banks have solid early warning capabilities ensuring robust credit monitoring and risk mitigation across the credit value chain. Credit risk of firms are being impacted by a whole host of new macro factors and stakeholder interests. The sanctions against Russia, the public pressure in the West to divest of any interests in the region, the sanctions against certain Russian individuals and the impact of higher commodity prices have altered the status quo overnight. Businesses that were seen as safe borrowers may look risky. Financial statements alone are no longer enough for risk managers to assess credit worthiness.

Banks that fail to improve their EWS will also face significant regulatory pressures. The European Central Bank (ECB) has highlighted[1] the huge variation in the quality of early warning systems and how credit assessment at a micro as well as macro level is core to risk management and provisioning.

1 Credit risk: Acting now paves the way for sound resilience later | europa.eu ,
Who pays the piper calls the tune: The need for and benefit of strong credit risk management | europa.eu

Effective EWS reduces loan loss provisions by

10-20%

and required regulatory capital by

10%

Today's EWS use

ad hoc manual triggers

producing meaningless alerts

Furthermore, banks are also losing their competitive positioning to bonds and non-bank lenders, in many countries. Galytix estimates that in the US more than three-quarters of corporate financing comes from these sources. This trend is less pronounced in Continental European countries, where bank lending continues to dominate.

Risk aversion alone is not a strategy. Upgrading early warning systems (EWS) are crucial for corporate banks to drive their competitive advantage and improve returns. This needs to be embedded across the credit chain from loan origination to fulfilment to risk monitoring. The strategic case for developing and implementing an EWS is clear: effective risk monitoring will lower both credit losses and capital requirements – directly improving a bank's ROE by over 20%. Experience shows that an effective EWS could help reduce loan loss provisions by 10-20 % and the required regulatory capital by up to 10 %. Moreover, an effective EWS will also maximise shareholder value by materially reducing the volatility of corporate bank earnings. This will underpin higher stock market valuation multiples.

Where traditional Early Warning Systems are faltering today

The experience of Galytix, PwC United Kingdom and PwC Luxembourg (hereafter jointly referred to as "PwC") is that currently most early warning indicators that banks produce could be meaningless. Given the surge in false positives, several manual data checks are implemented by banks to ensure consistency, effectiveness, and accuracy of signals – which ultimately increases the cost and time of managing the early signal detection process. The ECB has also been critical of the use of ad hoc manual triggers being implemented by banks and the need for a more systematic approach to alert monitoring.

For many banks, existing EWS frameworks are based on more easily available and traditionally used data sources including client financials and market data which is readily available. However, such indicators are usually backward looking and fail to predict corporate defaults well enough.

To generate signals, banks still use a traditional univariate modelling approach, but these have rather limited predictive powers. In addition, inflexible legacy data architecture – constrained by static, often hard-coded business rules – prevent banks from rapidly feature engineering their frameworks and adding / testing new relevant data as and when it becomes available.

Many banks lack mechanisms by which to seamlessly integrate non-traditional and unstructured data – data volume has been growing exponentially – into their downstream early warning analytical processes. Often, banks set up structured data warehouses or digital lakes, but these are costly to maintain and cannot handle the rising number of heterogeneous data formats that banks must use for effective early signal detection. As an example, banks often take six or more months to incorporate new and varied sources of data into their core early signal processes.

Designing an effective End-to-End Early Warning System

An effective EWS should identify borrowers at risk of non-performance (High Hit Ratio), distress or default sometime before an actual event (Time before Default). It must enable efficient and reliable assignment of borrowers to different watch-list categories and trigger other actions and escalation scenarios depending on the nature and severity of risk. The system must use indicators that are derived by combining both traditional and non-traditional data sources (internal and external) using a multivariate or decision tree modelling approach. It must be capable of algorithmically streaming and engineering (continuous discovery, ingestion, transformation, and curation of data) both unstructured and structured data through orchestrated pipes – supervised by humans – into a common data ontology – not extracted ad hoc from a static datastore. Finally, the system should have rules based high integrity versioning and traceability of data processes running through the pipeline as part of scheduled batch processing.

LEGO framework can deliver **High Hit Ratio** & improve **Time Before Default**

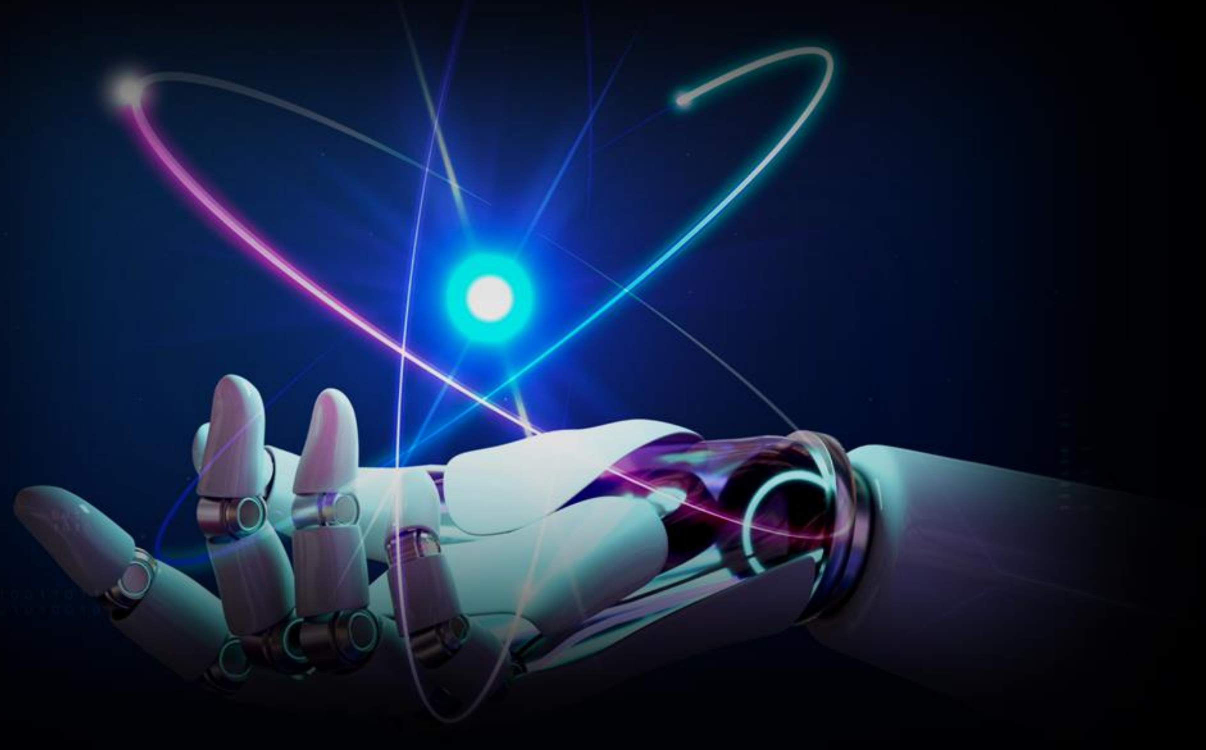
A LEGO (Leverage, External Indicators, Governance and Ontology) framework implemented in an AI-driven pipeline architecture based solution can accelerate EWS quality and effectiveness. This framework involves streaming real-time data and analytics, making adjustments to current indicators, adding a few high impact ones, and providing a systematic and automated capability to manage the EWS without creating unnecessary burdens. These processes allow credit risk managers to quickly assess exposures of counterparties. For instance, it would allow for not only tracking of revenues and investments of Western firms in Russia, but also make sanctions related checks in Western countries through subsidiary filings and court documents. These could tie back to not just Russian companies but also individuals on the sanctions list.

EWS built on **AI-driven pipeline architecture solution**

Both PwC and Galytix believe that expanding the list of early warning indicators must be focused on the highest impact external data sources around equity market signals, governance, fraud, aggressive accounting, or cyber risks. There is a lot of interest in sentiment analysis amongst both financial investors and banks. However, we find that these are often too broad creating a large audit trail of false positives that could take risk teams a long time to parse through. There must also be processes in place to ensure these data points are accurate, specific enough and from high quality sources.

No single list of early warning indicators is a magic bullet. There is a need for a systematic front-to-back approach where data governance is prioritised to ensure accuracy of definitions, lineage, and policies. Banks must also be careful not to rely on mutualised data excessively. Loans tend to be relatively illiquid and following the herd doesn't allow for credible risk management, in enough time. This includes which indicators are used, which data sources are seen as credible, what specific information or financial data percentage changes are seen as a threshold and how these are all combined in a fluid and evolving data ecosystem.

An EWS that generates real early warning indicators compared to late-emerging information such as credit ratings deterioration offers the potential to identify distress signals three to five months before an event. This enables banks to take risk-mitigating actions earlier and more effectively.



A large number of banks expect to revamp their Early Warning Systems with a greater emphasis on real-time data and machine learning models.

1 Why Early Warning Systems are crucial to maximising corporate banking returns

Key messages

a. Centre of gravity moves away from G10 bank lending

---

b. Corporate lending ROEs below cost of capital

---

c. A world-class EWS could improve ROEs materially especially in a rising default environment

Why Early Warning Systems are crucial to maximising corporate banking returns

Banks will have to adjust their **data, methodologies & warning indicators** to reflect the new normal

Risk management is once again the focus of banks, following COVID-19 and the ongoing Ukraine conflict which has resulted in sanctions on trade with Russia. China has recently unveiled a growth target of about 5.5 per cent, its lowest in three decades. European banks and firms are more exposed given the expected supply chain disruptions, with only one percent of US listed S&P500 revenues coming from the region[2]. Assessing counterparty credit risk is even more complex. Risks are particularly great where firms have balance sheet exposure to the region.

The conventional sources of data typically used in credit-risk assessments have become obsolete overnight. Models that the banks depend on to run their business are facing significant issues – as the modelling assumptions and boundaries haves significantly changed in a Post-Covid world. While access to the needed data to recalibrate the models is theoretically possible, integration of this information in an agile manner will be a herculean effort for the banks. This is because the systems and infrastructure on which they are built lack the necessary flexibility and scalability. All of this will test banks' credit risk management processes in coming months.

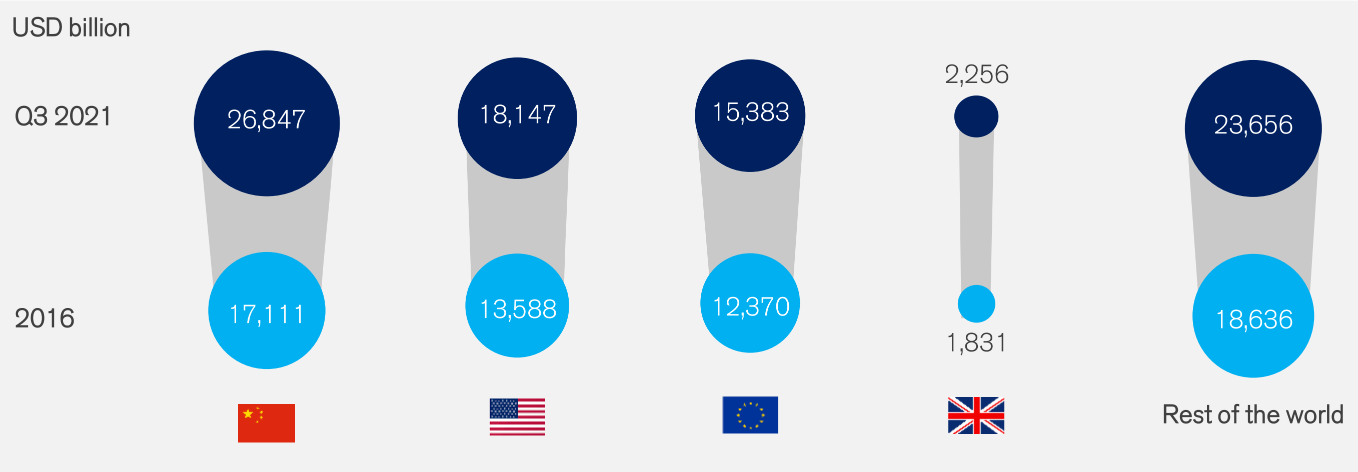
We are also seeing financial services clients on the end of increasing regulatory scrutiny of their ability to monitor early warning indicators to anticipate credit deterioration as quickly as possible. Regulators are focused on the whole credit monitoring and reporting cycle from EWS through to a bank's loan loss provisions and capital requirements. They are demanding more forward-looking indicators that can anticipate credit events and where credit quality could be recently or rapidly deteriorating. They are intended to complement existing risk management tools, without the intent of replacing or substituting any current practices.

Given the significant and widespread effects on the bank-wide operations, we expect regulators in the near-term future to expect the banks to adjust their data and methodologies to reflect the new normal.

Centre of gravity moves away from G10[3] bank lending

Credit to the non-financial corporate sector globally has exploded in recent years. According to the Bank for International Settlements (BIS) it is currently around $90 trillion, a more than 40% increase since 2016. The Exhibit 1 illustrates the large size and growth of corporate debt markets outside of the US and Europe. China at one-thirds of this number is an outsize influence on these statistics and much bigger than the US.

Exhibit 1: Change in size of major corporate debt markets - source BIS



[2] Stocks Fall, Oil Leaps as Ukraine Crisis Deepens – WSJ

[3] G10 countries refer to Belgium, Canada, France, Italy, Japan, the Netherlands, the United Kingdom, the United States, Germany, and Sweden

50% of US corporate loans held by non-banks

Bank corporate lending has faced headwinds both on the demand and supply side. The growth of liquidity in the corporate bond market has been underpinned by electronic trading and exchange-traded funds (ETF's). On the supply side with low rates squeezing margins and increased capital requirements, banks are increasingly selective on who to lend to. This trend has accelerated in recent years.

The BIS estimated that in 2021[4] loans contributed around half of non-financial corporate borrowing in the US but significantly higher percentages across other G10 countries such as France, Germany and Japan. The largest corporate debt market in the world, China, is also largely a bank loan market. These proportions are highest for SMEs that tend to be funded more by bank loans and lower for large multinationals. A growing trend particularly in the US is the emergence of non-bank lenders in the loan market such as private equity and distressed debt hedge funds. Galytix estimates that in Q4 2021 the proportion of US corporate loans held by non-bank lenders was just as large as those held by bank lenders.

The 2019/2020 National Bureau of Economic Research (NEBR) paper[5] "Why do firms borrow directly from non-banks?" provides interesting insights into the success of non-bank lenders. Analysing credit agreement data on a large group of US publicly listed middle-market firms (defined as companies with revenues of between $10m and $1bn) for the 2010-2015 period, this review found that non-banks lend to less profitable and more leveraged firms.

The probability of borrowing from a non-bank jumps by 34% as EBITDA falls below zero, an effect that is largely due to bank regulation. With the rapid growth of tech firms with long scale up investment cycles this proportion of firms is only rising. The Wall Street Journal[6] (WSJ) recently highlighted that around a third of Russell 2000 companies are loss-making, a doubling over the last decade and well above both the dot.com bubble and financial crisis peaks.

The NEBR study found that controlling for firm and loan characteristics, non-banks carry190 basis points higher interest rates suggesting that access to funding, rather than prices, is why firms borrow from non-banks. Corporate banks need to move away from their reliance on trailing profitability metrics to incorporate factors like the recurring revenue base for technology firms and intangible factors.

The NEBR research also found that non-banks appear to engage in greater due diligence in the loan origination phase than banks. Another lesson from non-bank lenders to corporate banks is one of not just focusing on early warning signals a few quarters ahead of a credit default but the need to embed the broad set of financial and non-financial indicators all along the credit chain starting with the origination team.

Corporate lending returns are under stress

Lending ROE much lower than Payments and Investment Banking

Most corporate banking revenues are outside of Europe and the US, reflecting the regional breakdown of bank loans and the higher net interest margins enjoyed by national champions in emerging markets. Corporate banks are facing increased pressure in many mature economies from bonds, non-bank lenders, and low returns on equity.

Although SMEs are typically a small portion of outstanding corporate loans in most countries[7], they are relatively high margin. Hence, they form a crucial part of the success of a corporate bank. With significant disruptions expected across supply chains, profits from SMEs are also likely to come under pressure in the future. In Europe, Galytix estimates that they could be around half of corporate bank lending revenues.

Corporate lending ROEs vary widely depending on global scale, regional market share, competitive dynamics, and the regulatory burden. This range is from low teens at the market leaders to low single digits at weaker players. ROEs are lowest in Europe and relatively high in some emerging markets where there are lower cost structures and oligopolies.

4 BIS Statistics Explorer: Table F4.2, Debt securities statistics | bis.org

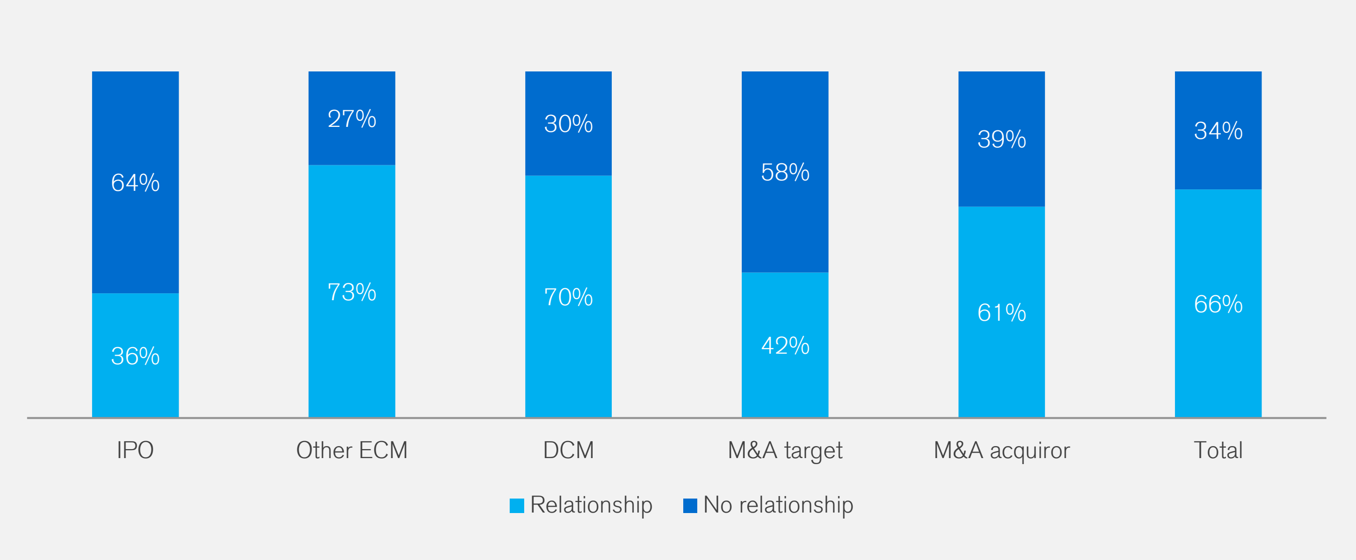
5 Why Do Firms Borrow Directly from non-banks? | nber.org

6 Investors Lose Appetite for Stocks of Unprofitable Companies | WSJ

7 Financial Stability in Focus: The corporate sector and UK financial stability | Bank of England

A major reason that corporate lending ROEs between banks are hard to compare is that lending especially in the large corporate segment is part of a holistic offering that corporate and investment banks (CIB) provide including DCM, M&A[8] and transaction banking. Revenues from these latter sources are significantly greater and higher margin. The chart below (Exhibit 2) from an April 2016 UK FCA study illustrates the importance of lending to support other activities with large corporates. Our channel checks suggest these statistics are just as relevant today.

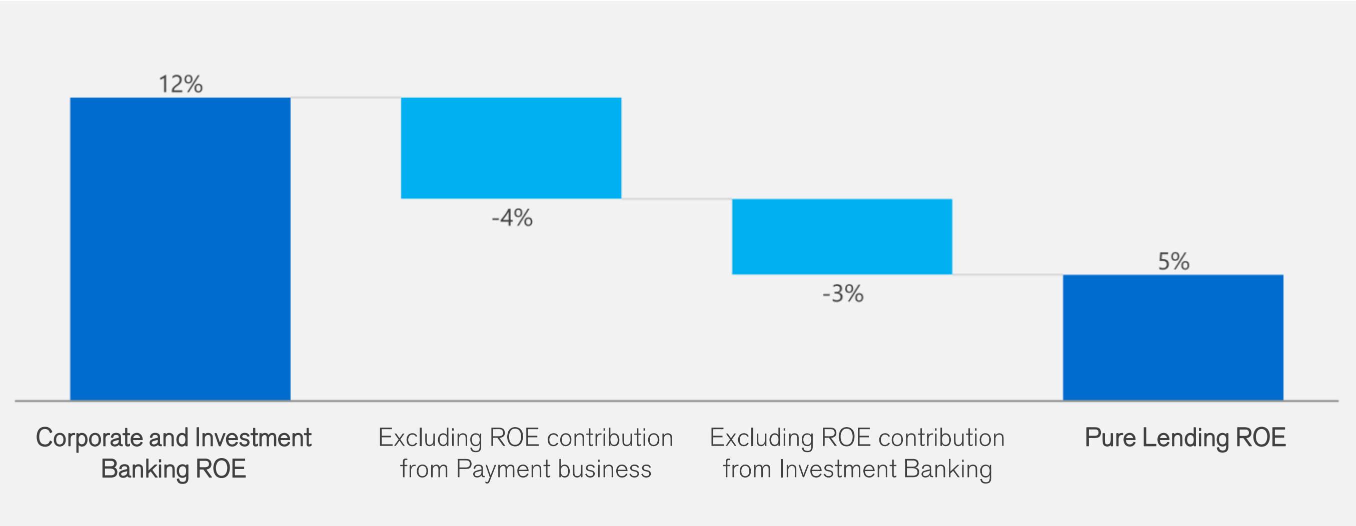
Exhibit 2: Proportion of roles awarded to banks which have and have not provided lending or other corporate banking services in the 2 years prior to the transaction, January 2014 to May 2015 - source FCA study



Pure corporate lending ROEs are
3-6%

In the following waterfall chart (Exhibit 3), Galytix estimates average corporate lending ROEs by taking public data on corporate and investment banking businesses and backing out transaction banking, ECM, DCM, and M&A which generate much higher ROEs in the 20%+ range across the cycle. Of course, depending on the relative size of these businesses the waterfall can vary from bank to bank. Moreover, there are wide geographical variations with Europe being the lowest margin given the high degree of competition and less scale than the US. The implied ROE of 5% for corporate lending is in line with Galytix's channel checks which are in the 3-6% range.

Exhibit 3: Corporate lending ROEs below cost of capital



[8] DCM is Debt Capital Markets and M&A is Mergers and acquisitions

As well as generating low ROEs, corporate lending has a higher proportion of fixed costs than capital markets businesses of banks. This creates scope for significant negative operational leverage. This is exacerbated by the high degree of cyclicality of ancillary transaction fees from DCM and M&A.

All of this makes it crucial to minimise corporate loan losses in an economic downturn. Lending committees need to have the right systems in place for early warning detection. These need to be used all along the credit chain from the time of loan origination all the way through to predicting defaults covering both Line of Defense 1 and 2. Moreover, an effective EWS will also increase shareholder value by materially reducing the volatility of corporate bank earnings. This will underpin higher stock market valuation multiples.

Risks to future corporate lending ROEs from increased defaults

Corporate lenders were already struggling to generate decent returns and worried about the impact of the withdrawal of government stimulus and inflationary pressures. Now they are being hit by an unprecedented wave of sanctions to incorporate into their KYC and credit processes which will be an unavoidable cost heavy endeavour. Although revenues and investments in Russia for most industries is small, the inclusion of sanctions on Russian oligarchs makes this complicated and some banks have also stopped issuing letters of credit for deals involving Russian energy[9].

Corporate lending suffers greater losses in an economic downturn than consumer and property lending, owing to the lower levels of collateral. Add in unprecedented amounts of disruption from technology innovations, supply chains and climate change and there are a wide range of potential scenarios for how bad corporate default rates could get.

3 scenarios on corporate lending ROEs

Pure Lending ROEs to decline from

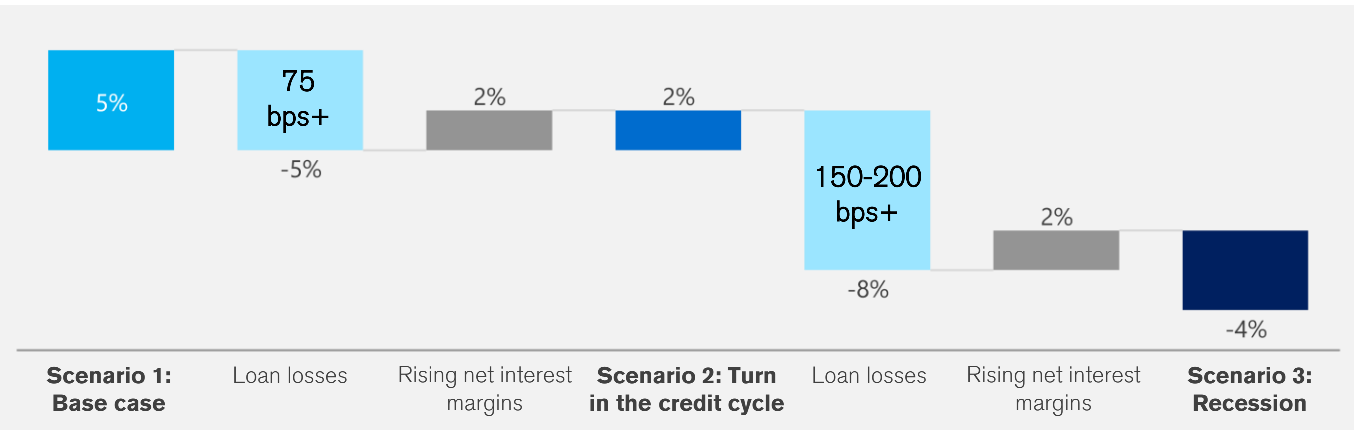
5% to -4%

in recession

Given these uncertainties Galytix has chosen to provide three potential scenarios for the coming years. The first scenario or **Base case** assumes the very low corporate defaults of recent quarters continues. The second scenario labelled "**turn in the credit cycle**" below assumes loan losses in line with what we have seen across the cycle (a circa 75bps+ increase) and rising net interest margins. This generates ROEs of around 2%. Our third **recession scenario** assumes a total 150-200bp+ increase in loan losses and would result in ROEs on corporate lending of around -4%. A continuation of the current Ukraine crisis with its direct and indirect consequences on credit quality has the potential to look like a recession type scenario. These numbers should of course be treated as directionally illustrative rather than precise down to the decimal point.

The figure (Exhibit 4) below illustrates these three scenarios. It shows that there will be difficult choices facing corporate bank senior management when the credit cycle turns.

Exhibit 4: Corporate lending ROEs under different loss assumptions



9 Singapore Banks Halt Russia Commodities Lending to Cut Risks | Bloomberg

A world-class Early Warning System could improve corporate lending ROEs

EWS benefits ROE by **70bps** in benign credit markets but **330bps** in recession

In the new world, the strategic case for developing and implementing an EWS is clear: effective risk monitoring across both Line of Defense 1 and 2 will lower both credit losses and capital requirements – directly improving a bank's ROE by over 20%. Experience shows that an effective EWS could help reduce loan loss provisions by 10-20 % and the required regulatory capital by up to 10 %.

In the below charts (Exhibit 5, 6, 7) we have taken the three scenarios outlined in the prior section and assumed that implementing an EWS reduces loan losses by 20%, in line with our channel checks across the industry. Although there is a likely regulatory capital benefit that we estimate at 10%, this is likely to be lagged and spread over many years hence we have not incorporated it in ROE calculations for the given year. This ability to protect ROEs in tough credit conditions and reduce the volatility in these ROEs makes an effective EWS crucial to maximising shareholder value.

Exhibit 5: Below illustrates that in the 'Base case' there is a 70bps benefit to the corporate lending ROE of 5% from implementing an EWS

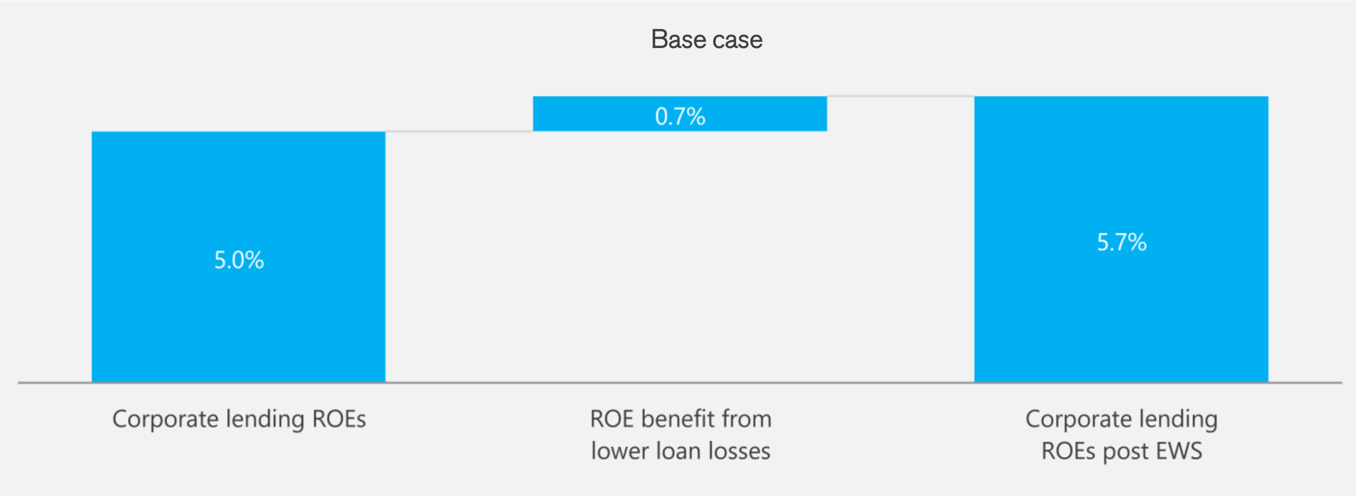


Exhibit 6: Below illustrates that in the 'Turn in the credit cycle' scenario there is a 170bps benefit to the corporate lending ROE of 2% from implementing an EWS

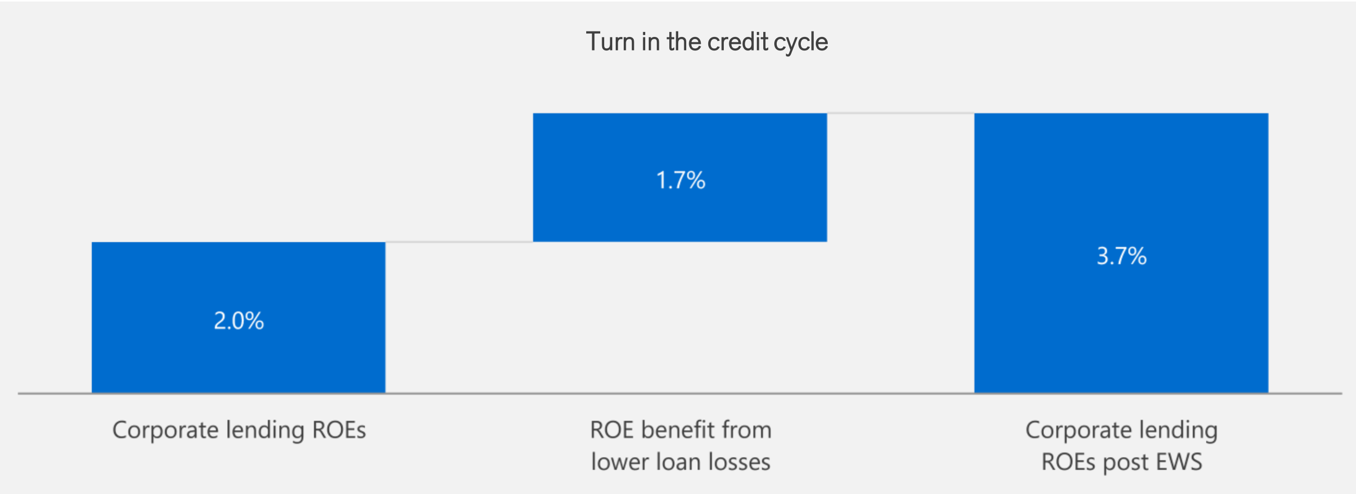
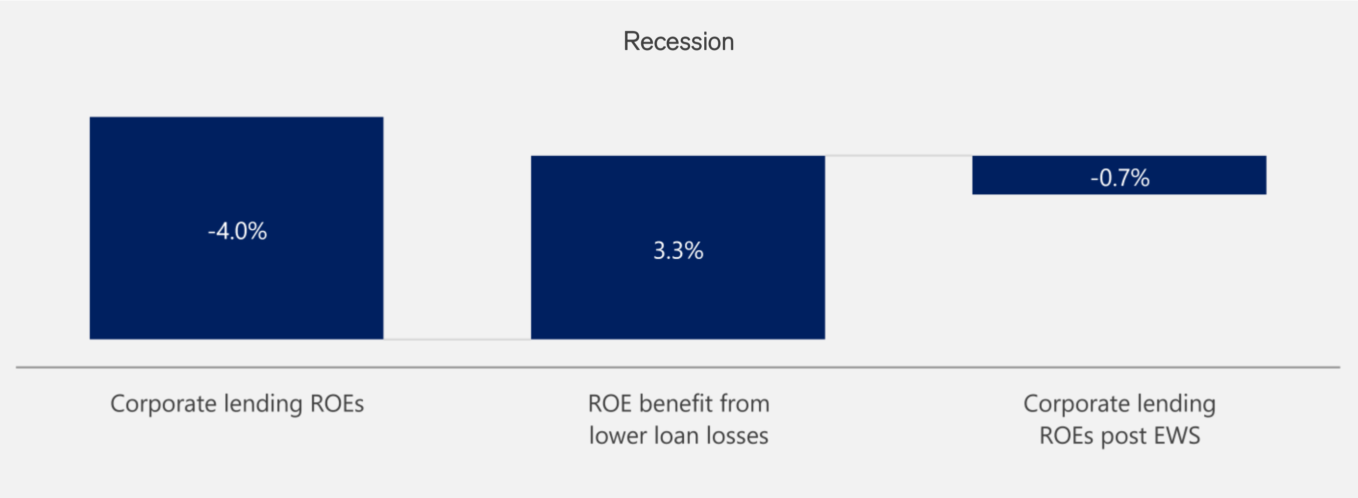


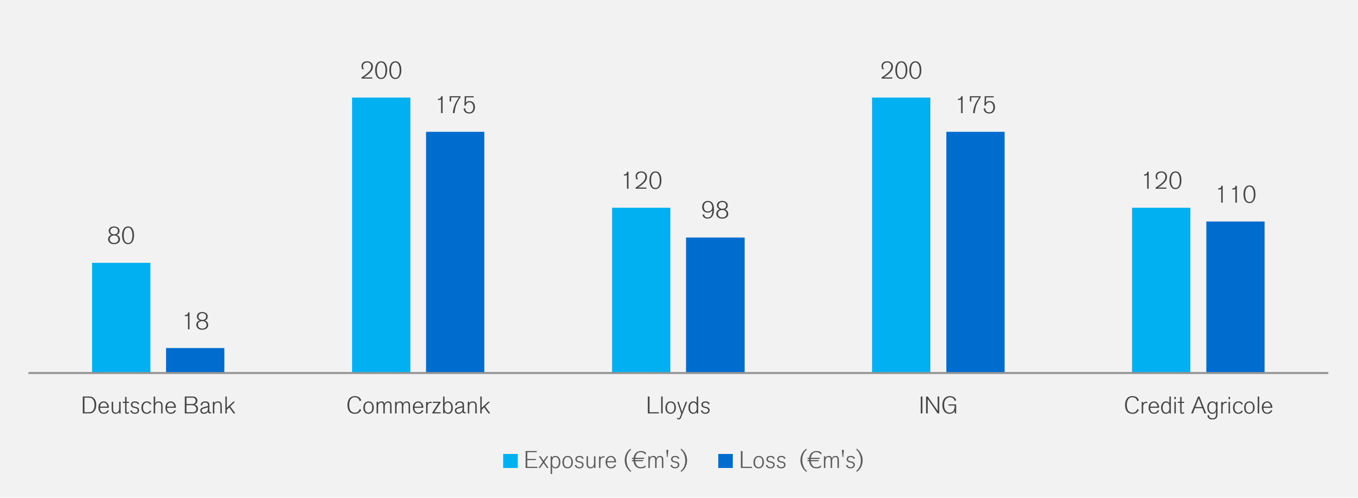
Exhibit 7: Below illustrates that in a recession scenario there is a 330bps benefit to the corporate lending ROE of -4% from implementing an EWS



The core function of every bank is of course to be a risk manager and we have seen many examples of effective risk management in the last few decades as certain banks have managed to either proactively stay away from specific credits and segments or use hedges and active workouts ahead of distress. During the financial crisis this saved the best risk managers tens of billions of dollars and many hundred basis points of ROEs.

Most information on the success of a bank's early warning systems is private and anecdotal with limited public proclamations made quantifying this. However, a high-profile example of disparities in loss rates faced by lenders to the same borrower was the corporate failure of Wirecard in June 2020. Here lenders lost virtually all the €1.75bn of outstanding loans, with almost half of that lent by four banks (Commerzbank, ABN Amro, ING and LBBW). The chart below (Exhibit 8) with data sourced from Bloomberg and the Wall Street Journal[10] illustrates diverging loss rates across major lenders.

Exhibit 8: Wirecard lenders



[10] https://www.bloomberg.com/news/articles/2020-07-03/deutsche-bank-s-wirecard-ties-from-margin-loan-to-merger-talks, https://www.bloomberg.com/news/articles/2020-08-06/wirecard-implosion-tears-through-european-banks-bottom-lines, https://www.wsj.com/articles/wirecard-woe-spreads-as-banks-struggle-to-exit-loans-11594978201 https://www.bloomberg.com/news/articles/2021-01-14/commerzbank-flagged-shady-wirecard-transactions-before-meltdown

If these banks were collectively able to identify distress when the first Financial Times report with allegations of accounting fraud came out in January 2019, they would have been able to save most of the €1.75bn of loans. Moody's initiated coverage with an investment-grade Baa3 rating in August 2019 and the share price only collapsed in mid-2020.

Regulatory pressures on credit risk monitoring continue to build

The European Banking Authority (EBA)'s guidelines[11] specifying internal governance arrangements for the granting and monitoring of credit facilities throughout their lifecycle was a response to the non-performing loan (NPL) crisis. It went into force on new loans on 30 June 2021 but existing loans that require renegotiation come into scope from 30 June 2022. There is also a period for remedying processes until 30 June 2024.

Central banks are increasingly focused on the need for banks to have a robust EWS for credit risks. Most vocal has been the ECB which has spoken regularly on the need for these to provide timely identification, forward-looking measurement, and mitigation of credit risk.

The ECB has been focused on how assessment of individual credits by the banks they supervise has often diverged from overall credit provisioning by these banks. The most notable takeaway of bank earnings releases over the last two years in both Europe and the US has been the huge credit provisions taken and subsequent writebacks.

On 24th November 2021 ECB Supervisory Board member Elizabeth McCaul summarised ECB views well[12] :

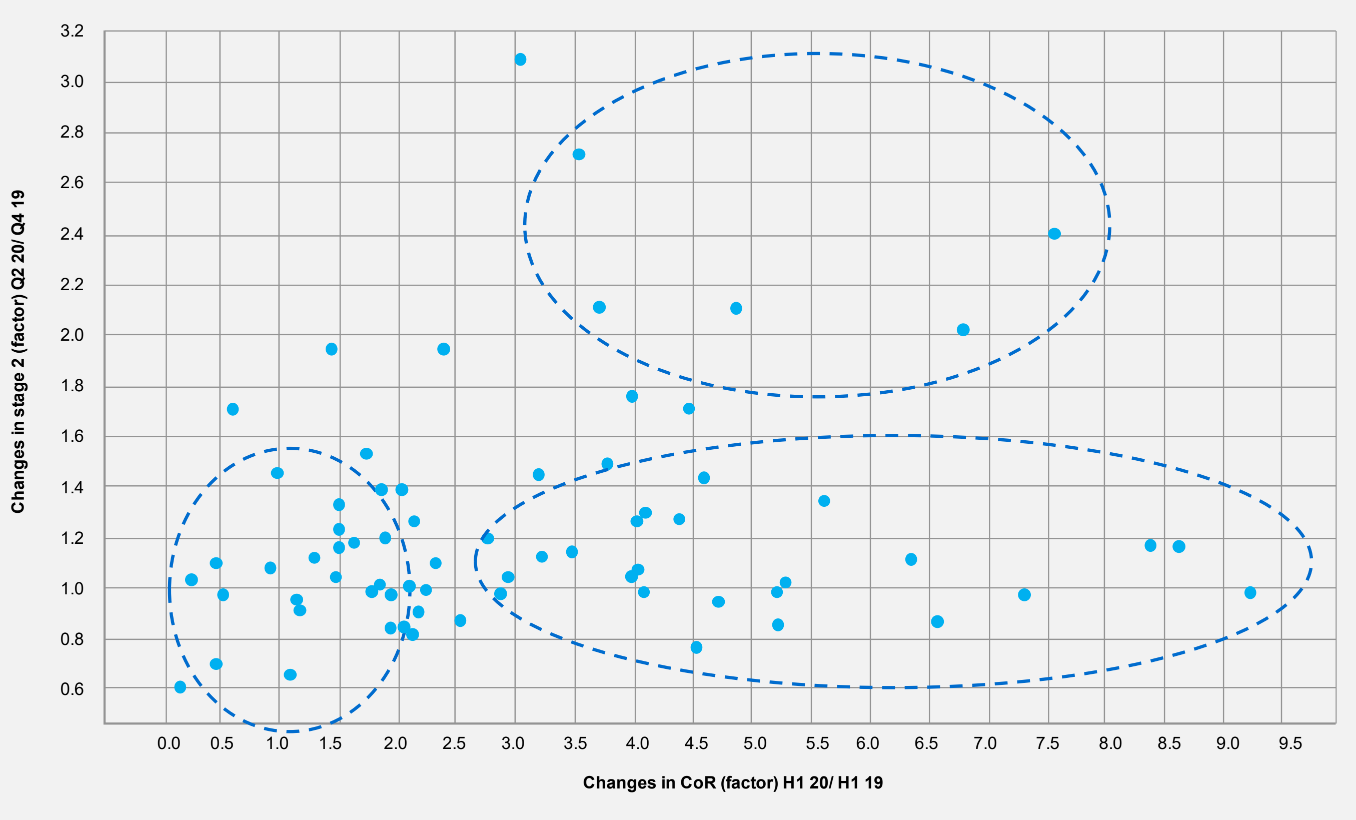
> “Having noticed some undue delays in IFRS 9 stage adjustments in 2020, despite a significant increase in credit risk, we recommended that banks consider a threefold increase in the PD of an obligor as a hard backstop measure to transfer exposures from stage 1 to stage 2. The IFRS 9 benchmarking data for the second half of 2020 show a substantial reduction in the level and dispersion among banks of the share of loans kept in stage 1 despite a tripling of the PD estimates at re-rating. This shows that our recommendation was successfully implemented.
>
> Although this convergence of provisioning practices is a step forward towards implementing IFRS 9, we must not become complacent. For example, we still see a wide variation in stage 3 classifications across banks. Our hypothesis is that this material variability in default classification stems from bank-specific implementations of the definition of default and UTP assessment. And so, we will continue to follow up with banks on these issues in the coming year.
>
> **– Elizabeth McCaul**”

The divergence between IFRS 9 stages and overall loan loss provisioning in the early days of the pandemic were highlighted well by the following ECB chart (Exhibit 9) across euro area banks.

[11] EBA GL 2020 06 Final Report on GL on loan origination and monitoring.pdf | europa.eu

[12] Banks' credit risk management and IFRS 9 provisioning during the COVID-19 crisis | europa.eu

Exhibit 9: Differences in loan loss provisioning practices across Euro area banks



Source: ECB. Horizontal axis: Cost of risk in 2020H1 divided by Cost of risk in 2019H1. Example: Value of 2 means a 100% increase in cost of risk in 2020H1 vs. 2019H1. Vertical axis: IFRS stage 2 ratio in 2020Q2 divided by same ratio in 2019Q4. IFRS stage 2 ratio = stage 2 loans / total loans subject to staging. Example: Value of 1 means the IFRS9 stage 2 ratio is unchanged between 2020Q2 and 2019Q4

Elizabeth McCaul described this trend on 4th December 2020[13]

> Stage 2 classifications, which should anticipate credit risk deterioration, have also remained low for many banks. While some banks – those in the bottom-left hand side of the chart – seem to be in a wait-and-see mode in that low Stage 2 classifications are matched with low provisioning, others – those in the bottom-right hand side of the chart – are making provisions by means of overlays, albeit without clearly identifying the exposures subject to increased credit risk.
>
> **– Elizabeth McCaul**

Pillar 2 requirement could capture EWS frameworks

Non-performing loans (NPLs) in Europe declined from 8% of outstanding loans or €1 trillion at the end of 2014 to 2.3% or €422 billion in June 2021. In 2018 the ECB announced a bank-specific approach to their Supervisory Review and Evaluation Process (SREP). This included ensuring banks conduct an adequate classification and measurement of their balance sheet risks and are prepared for dealing with distressed debtors in a timely fashion.

[13] Who pays the piper calls the tune: The need for and benefit of strong credit risk management | europa.eu

Average Euro bank pillar 2 charge up to

2.3%

When SREP decisions are reached near the end of each year, the ECB imposes a Pillar 2 requirement add-on for many banks. This de facto increase in capital requirements of some banks can be used to compensate for any shortfall against ECB expectations regarding provisions for NPLs.

The Pillar 2 requirement is a powerful tool to encourage banks to ensure their EWS of credit risk assessment are rigorous. In 2021, the Pillar 2 requirement across euro banks was an average of around 2.3% up from 2.1%. That is mostly due to the introduction of a specific requirement (a provisioning shortfall add-on) imposed on banks which have not booked enough provisions to cover the credit risk on non-performing loans (NPLs) granted before 26 April 2019.

The ECB has also warned European banks to be ready for the impact of Russia related sanctions both in terms of adhering to the new requirements and the risks on credit quality deterioration. [14]

The significance of crisis and the speed with which banks have had to react has triggered new risks

Covid-19 has also affected model reliability across all bank functions and operations. For example:

- Rating models have become inaccurate because they are unable to update scores rapidly
- Model based market-risk approaches are over-reacting to stressed price and credit
- Regulatory models are mechanically increasing capital and liquidity requirements
- EWS indicators are producing high false positives

Models need to be updated to adapt to the **new normal**

Banks have taken rapid mitigating actions that are tactical in nature by replacing models with expert views, recalibrating models and building expert judgement into models. These tactical actions have helped the banks circumvent short-term needs but generated a host of new risks in the long term:

- Model underperformance and failure
- Model versioning and traceability
- Model re-development
- Predictability of models to produce an effective modelling outcome

To mitigate the long term risks, banks need to take a step back and ultimately work towards adjusting their data and methodologies to reflect the new normal.

[14] Western Banks That Stuck With Russia Face Their Biggest Test of Nerve | Bloomberg

In the new normal, banks believe that building an effective EWS for credit assessment is more critical than ever.

2 Where traditional Early Warning Systems are faltering today

Key messages

a. Current EWS are producing a high degree of late indicators, false positives and fail to incorporate external data

b. Multivariate models are rarely adopted with most EWS relying on ad hoc manual triggers

c. EWS is not embedded across the credit chain. Data and technology architecture is siloed and not flexible

Where traditional Early Warning Systems are faltering today

8 out of 10 EWIs are **false positives**

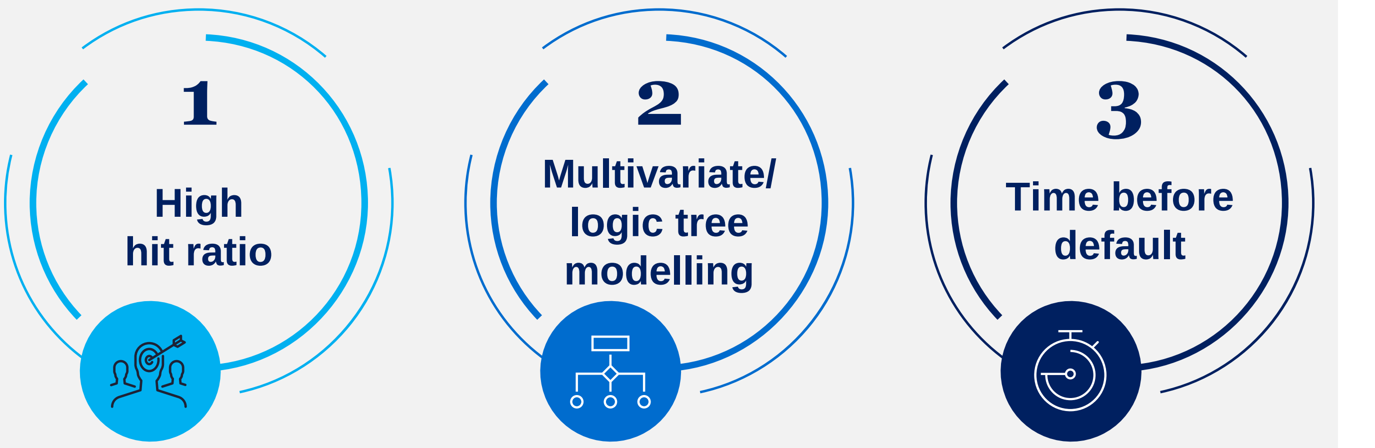
Over recent months we conducted a gap analysis of what factors traditional EWS have missed. We spoke to a wide range of market participants and analysed several dozen of the largest bankruptcies in the world in recent years. A risk professional stated to Galytix "Today, eight out of ten early warning signals produced create false positives". Key areas that need improvement are the hit ratio of early warning indicators, the timeliness of these indicators and the methodologies being applied to embed these indicators into broader credit risk management.

As may be expected, the maturity level of an early warning framework at different financial institutions varies significantly. However, a common trend we are seeing is an increasing use of more advanced data analysis of trends and AI/ML techniques on a more sophisticated data set to be better able to generate early alerts of potential credit deterioration. Newer data sources such as scanning of news articles to understand market sentiment may also be included.

Most EWS struggle to achieve high hit ratios, use traditional data with limited external data

For most banks, existing frameworks have been based on more easily available and traditionally used data sources including client financials and market data which is readily available. However, such indicators are usually backward looking and fail to predict corporate defaults well enough. Added to this, corporates are being impacted by an unprecedented level of disruption and industry change increasing the speed of change and potential for a credit default event.

Exhibit 10: EWS systems fail to address three core issues



- How many customers are flagged by the system?
- What proportion of customers are transferred to the watch list?

- Which combination of indicators are statistically significant in a multivariate or logic tree context?

- What is the average time before default when the system or trigger identifies a non-performing customer for the first time?

Each bank typically has its own slightly different approach to EWS but there is a great deal of overlap amongst a core set of Early Warning Indicators (EWI) used.

Exhibit 11: Banks using traditional metrics to assess obligors

Segment	Metric category	Metrics (examples)
Corporate	Profitability	• EBITDA performance over the last twelve months versus budget
	Indebtedness	• Gross debt divided by last twelve month's EBITDA (or change of this indicator over time)
	Obligor risk profile	• Credit rating downgrades of certain amount of notches • Bond/CDS prices, late payments • Covenants and internal ratings
SME	Transactional data	• Company's borrowing history with the bank
	Indebtedness	• Ratio of loans to capital • Company's level of bank deposits relative to loans
	Obligor risk profile	• Company age

The experience of both Galytix and PwC is that currently most early warning indicators that banks produce could be ineffective. Financial information extracted from public annual reports can be up to a year out of date for public companies and even longer for private companies. EWIs need to change to remain relevant and leading. They also need constant fine-tuning to stay up to date.

Internal data received by banks from the borrower are typically payments of interest on loans, management accounts and other information included in loan covenants. These are more up to date than last year's annual reports. But information quality can vary materially limiting their usefulness as effective early warning signals of credit default.

However, those lenders that have large transaction banking business with the borrower can benefit from near real-time payments data. Late payments from customers or to suppliers can give early hints of corporate stress. Where a borrower may use another bank for payments, a leading transaction banking business can still provide signals of the financial health of peers, customers, and suppliers.

Existing EWS keep missing many high-profile bankruptcies. This is a big risk when the credit cycle turns, and the absolute number of corporate failures increases. The high number of false negatives is also a function of the reliance by most banks on internal data and historical financial statements.

That is why financial institutions are starting to see clear benefits for early alert generation with increasing use of external data sources to understand market sentiment. Use of such external data is facilitated through rapid development of newer technologies. Traditional data such as analysts' commentary and corporate or industry news flow in the press is more easily discoverable online than in the past. There has also been the emergence of new alternative data sources such as social media, payments data and satellite imagery.

There is a risk that if banks don't embrace all of this, they will be left behind by their competitors. New fintech firms are data-first, algorithm-driven and use new alternative sources. As do hedge funds and private equity firms. Financial investors have made huge investments in their data capabilities both in terms of the amount of data they consume and the skills of their teams of quants and software engineers to be able to extract signals from this noise.

There is pressure from regulators such as the ECB for more forward-looking and granular early warning indicators. The following risks have emerged or accelerated in recent years to such a degree that for certain firms and industries they could increase the probability of default on a loan:

- Technology disruption results in legacy industries disappearing at speed
- De-globalization is creating supply chain bottlenecks
- Climate risks
- Cyber and technical debt

Time to default – getting ahead of the game

Many of the traditional indicators that are currently being used are more "late" than early warning indicators. Late payments on loans are one such internal data source. External data that often fit into this category are credit ratings downgrades and bond price selloffs. Unlike equities or even bonds, loans can be relatively illiquid to sell or hedge so for enough time to act signals need to appear well ahead of a default event.

In the US many bankruptcies do allow returns for creditors as was illustrated by 2020's largest filing Hertz[15]. However, many corporate defaults in recent years have been companies that were highly indebted, struggling for profitability and had poor credit ratings for many years. The largest bankruptcies in 2020 were concentrated in the aviation, energy and mining and retail industry[16].

Methodologies used in EWS

Using more indicators may mean an increased number of false positive signals from EWS which could dramatically undermine confidence in the system. For example, many attempts to create online news aggregator driven EWS indicators in the credit space have seen mixed results owing to the inability to extract the relevant signals from a mountain of data.

Univariate models lead to high false positives

The ECB has praised banks for being rigorous at screening vulnerable sectors and geographies during the pandemic. At the same time, they have outlined the need for improvements to the consistency[17] of EWS and been critical of the excessive use of manual ad hoc triggers at many banks. A 2019 ECB working paper[18] outlined ECB research on the potential to use machine learning leveraging decision trees, in building an EWS to monitor smaller European banks. The following diagram (Exhibit 12) is an example of such an approach from this paper.

---

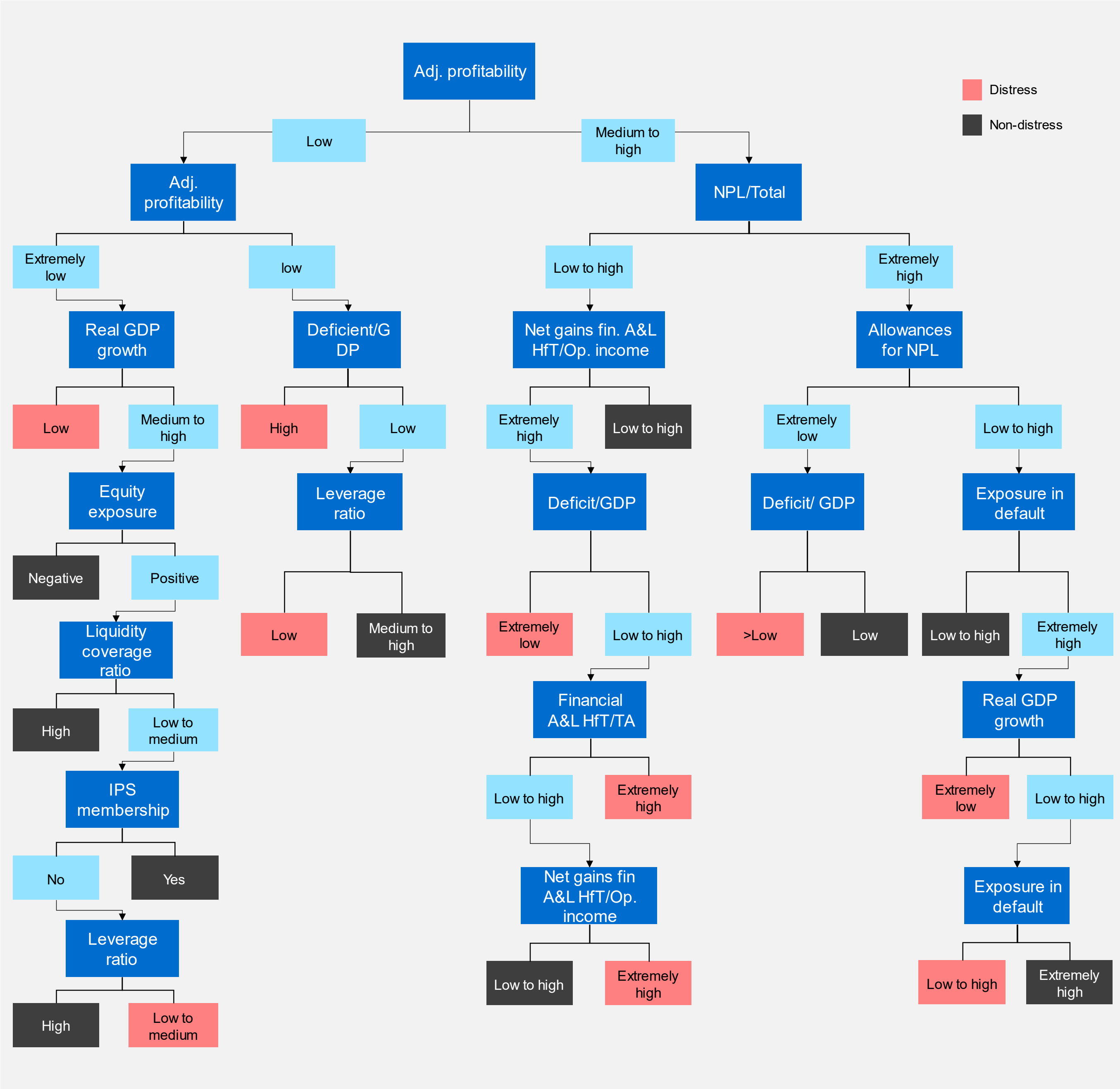
[15] Hertz Bondholders Seek More Cash Amid Stock's Comeback | wsj.com

[16] The World's Biggest Bankruptcies 2020 | Global Finance Magazine (gfmag.com)

[17] Credit risk: Acting now paves the way for sound resilience later | europa.eu

[18] A new approach to Early Warning Systems for small European banks | europa.eu

Exhibit 12: Example of decision tree for EWS - source ECB



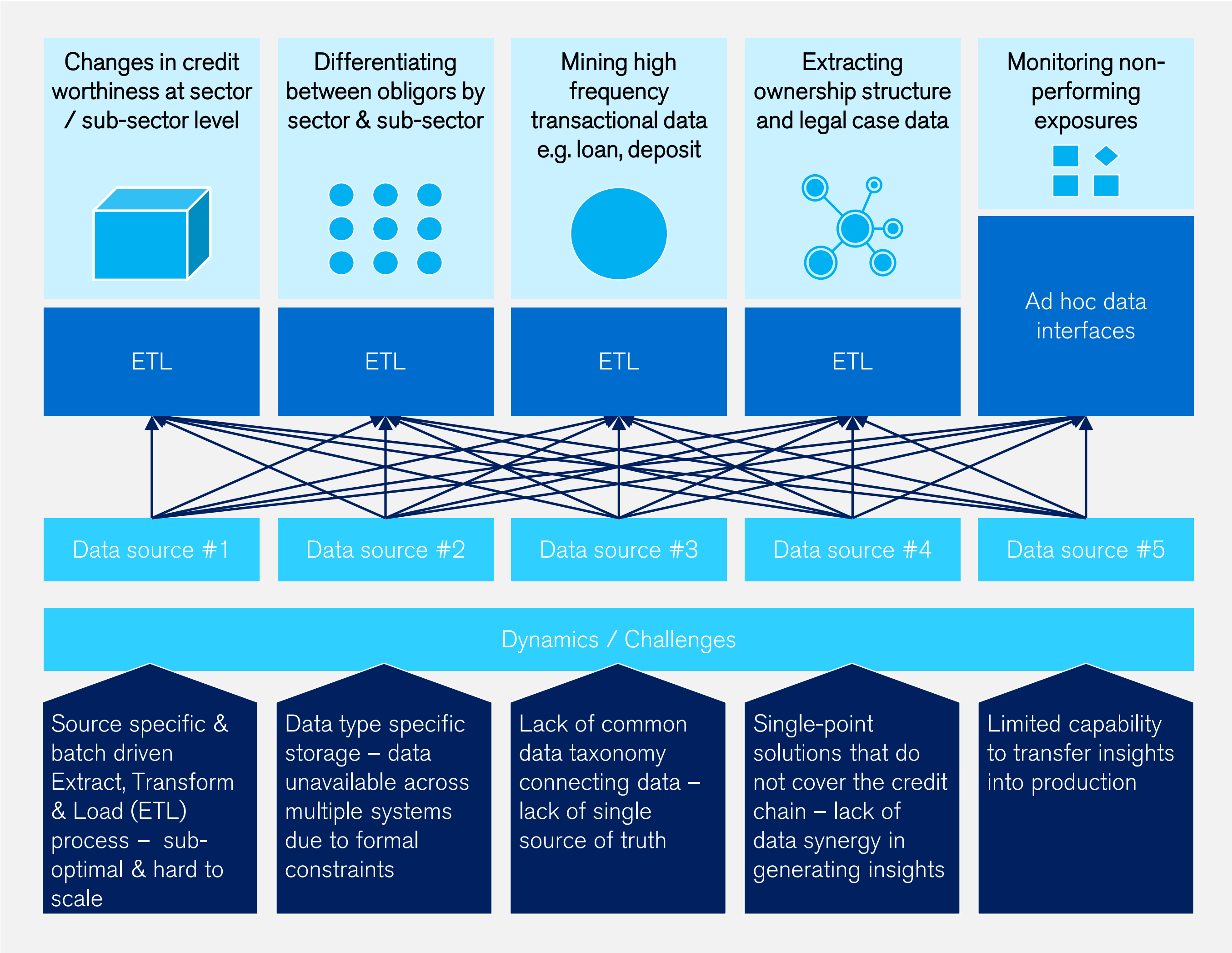
Simplistic univariate models with one discrete financial metric as a trigger point can lead to a high level of false positives. Any single metric could fail to distinguish between what is short term volatility in revenues and what is more a long-term issuing impacting solvency.

Multivariate models bring their own challenges of determining the right mix of inputs with associated flexibility to reflect not just back-testing but also fine-tuning of weightings to reflect the factors that are more relevant today in the past.

The data and infrastructure challenges of embedding EWS

Traditional data approaches in the form of Extract, Transform and Load (ETL) used by banks today lack the ability to engineer and digitise data at scale and speed. This approach is not able to overcome the data challenges necessary to produce effective EWIs e.g. analysing obligors at sector / sub-sector level, assessing borrower resilience at granular level and developing high-frequency transaction data related analytics

Exhibit 13: Single point solutions and lack of a scalable data engineering capability wastes resources & increases complexity



Embedment

EWS should be properly embedded into the wider credit risk management cycle with a robust system of actions and escalation scenarios. A simplistic approach where alerts are triggered to a human on a daily, weekly or monthly basis does not work and could be counterproductive.

Many banks lack a clear understanding on how and where the EWS will be applied in the credit chain and on the interaction between credit and risk teams. For EWS to be effective they need to be applied throughout the credit chain from origination through to monitoring.

Data

A process that is so data reliant is never easy due to the data quality issues that can be expected when considering such large disparate data sets. In the early days of modern computing a phrase was commonly used to warn of the effects of poor data quality "garbage in, garbage out" and this is no less true today.

Data issues are further compounded by having to navigate confidentiality and data protection laws, particularly when using external providers to assist with some of the most complex data processing techniques as is often the case due to the expertise required.

Automation

Risk professionals spent **60%+** of time on finding and cleaning relevant data

Despite living in the digital era EWS in many cases still rely on manual data processing. Risk professionals today spend over 60% of their time on finding relevant data and making it ready for analysis i.e., cleaning and curating data. This is because:

- Data engineering efforts are currently duplicated across multiple functions
- 95% of available data lies in unstructured forms such as annual reports and covenants where information must be extracted manually
- Anecdotally data exploration for machine intelligence enabled system deployment averages 12 to 24 months

Corporate banking has lagged other parts of financial services in the digitisation journey. The growth of relevant available data has materially outpaced the insights being monetised over recent years in corporate credit analysis. Banks that have been built over decades of acquisitions and legacy technology typically have siloed IT systems with scattered data and a piecemeal approach. There is a lack of a dedicated data engineering capability in a fully scalable modern data architecture. In recent years the priority of many banks' technology teams has been on building algorithms and running model pilots without creating a robust data ontology of definitions, lineage and relationships to underpin this.

Credit assessment whether it is in corporate banking credit teams, or the risk monitoring team has typically been a highly manual and ad hoc approach rather than being a systematic and automated approach. There has been a lack of use of algorithms traditionally and not enough precision when they have been introduced to the process.

Banks are moving fast to build advanced analytics capabilities and add new data sources into their EWS engines.

3 How to build an effective Early Warning System

Key messages

1. Need to incorporate financial data and soft indicators in same EWS in systematic fashion

2. LEGO framework focuses on specific indicators related to Leverage, External factors and Governance

3. No single list of EWI is a magic bullet. Need for end-to-end data ecosystem, mixing tech and expert judgement

How to build an effective Early Warning System

Lessons from Industry

Based on Galytix and PwC experience, four factors are critical to a world class EWS capability :

- The right set of early warning indicators at sector and sub-sector level are selected
- High quality data and a robust methodology to model outcomes is paramount
- Cover the entire credit chain and ensure flexibility exists to fine-tune risk processes to incorporate early warning signals
- Have the right technology and infrastructure in place to ensure EWS can be embedded across the entire credit chain – not focused on monitoring alone

Exhibit 14: A world class EWS exhibits four key capabilities

Right set of early warning indicators

- Financial data (Balance sheet) signals
- Transactional data signals
- Non-traditional data signals

Robust modelling methodology and high quality data

- Multivariate modelling approach
- Internal and external data integrated in an enterprise -wide data model ontology

EWS 2.0

Right underlying technology and infrastructure

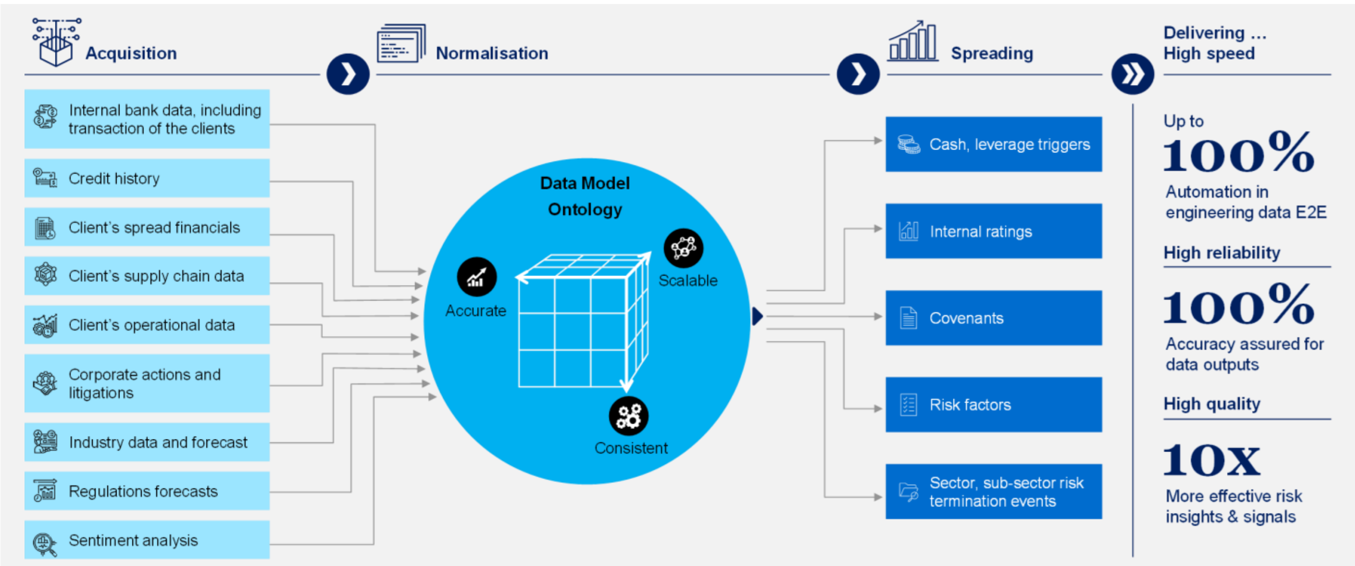
- E2E EWS covering the entire data lifecycle
- Pipeline-driven Data-As-An-Infinity Loop architecture

Covers the entire credit chain – from origination to monitoring

- EWIs applied during time of loan origination, Line of Defense 1 & 2
- Robust governance framework for any early warning alerts generated

A fully connected data ecosystem underpinning core EWS capabilities is critical to success. For example, banks do not typically use transaction data very much, due to the unstructured nature and high volumes of data. In order to build an effective an EWS that produces high hit ratio, banks have to pay serious attention to developing a dynamic and fully connected data ecosystem built on financial data, transaction data and alternative (external) data.

Exhibit 15: A combination of traditional and non-traditional data sources for an effective EWS



Understand set of indicators relevant for your credit risk exposure

The first step in establishing an early warning framework is to consider what metrics to monitor.

In doing this, it is important to draw the distinction between internal and external data sources that can be used to calculate those metrics. Alternatives to the traditional data sources may include the following:

- Internal bank data (e.g. exposure, high frequency transaction covering loan, deposits)
- Internal accounting data of a company
- Corporate actions and litigations
- Sentiment analysis (Internet, press, analytical reports, social networks, and platform)
- Fact checks

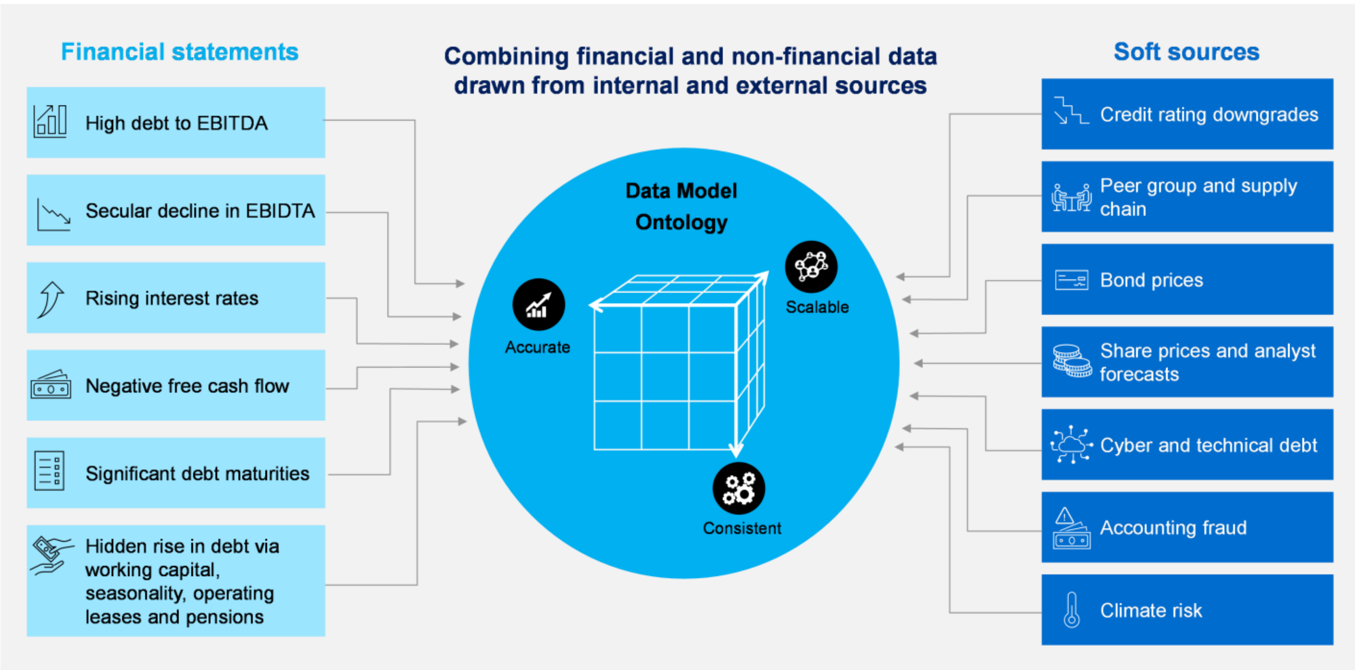
The range of quantitative and qualitative indicators available is vast with some giving better results than others in terms of effective risk management. It is important to consider the specific intricacies of the portfolio under management to determine the effective set of indicators. In this case less is more as a smaller set of truly effective indicators may provide far more useful information than a wide range of mildly effective ones. Some banks have specific sets of indicators on large exposures or client segments. Metric prioritisation is also extremely important.

Surge in interest for sentiment analysis

ESG data is a major new area of focus for banks. However, it is important to distinguish between what ESG data is relevant for a bank's credit risk engine and what is relevant for broader strategic direction and shareholder disclosures. The former must be material enough to impact the ability of a corporate to repay outstanding loans when they mature. An example may be a high concentration of production or customers in an area of immediate climate change impact such as wildfires or flooding where a company may not be able to purchase liability cover.

Sentiment analysis has seen a surge in interest in recent years as investors look for clues beyond financial statements and natural language processing (NLP) models have improved in terms of accuracy. There are many technology firms that read earnings call transcripts looking for key words and relevant semantics around them. This is also being tried out by large banks for their early warning systems that look at credit and market risks across a variety of business lines from trading to corporate and retail banking. However, to date these have struggled with excessive false positives. Online news scrapping for a credit EWS should focus on specific areas of interest such as profit warnings, earnings downgrades by equity analysts, material cyber security breaches or rumours of material fraud.

Exhibit 16: Illustration of combination of financial statements related data and soft data sources



Proving causation rather than accidental correlation

Have a robust methodology and pay attention to data quality

Once a set of indicators to monitor has been agreed, the data must be sourced and a workflow established to monitor the indicators and compare them to agreed thresholds, generating an alert in case of a breach that might indicate action needs to be taken.

Should trigger levels be set at binary breakpoints such as “If Interest cover is less than 1x and credit ratings are less than BB then an EWS alert takes place” or should they have a more flexible weighting system where a multitude of data points must be incorporated and weighted for the magnitude of their change or breach rather than binary alerts. In both cases proving causation rather than accidental correlation can be difficult as the real world is of course messy in comparison to theoretical models. Hence the importance of industry knowledge being used in creating and constantly making sure these systems are recalibrated appropriately.

While the calibration, and frequency of recalibration, of alert trigger thresholds is important to be able to effectively generate early warnings, it is also important to consider the generation of excess false positives when setting the levels. Insufficient alerts and you miss important warnings. However, too many alerts that do not require action may result in credit officers starting to ignore alerts and thus missing important warnings. Setting of the appropriate threshold level requires a fine balance.

When using modern data analytics techniques, it is important to test the alert generation on different data sets. Over fitting to the data set that has been used may give poor results. It should be back tested to observe the ability to generate alerts for events that have happened in the past.

Change your risk processes to incorporate EWS

It is also important to ensure there is a robust governance framework around any early warning alert generation capability. Effective action is as important as the alert generation itself otherwise the whole process is ineffective in its ultimate purpose of risk management. It is vital that trigger thresholds are effectively calibrated with appropriate frequency to assist with monitoring and detection. And when an alert is generated, a framework should clearly stipulate what action needs to be taken to effectively manage the portfolio. As with any other risk management framework regulators are concerned that a robust and effective governance structure is put around this to avoid costly mistakes or misuse. Reducing false positives is crucial to reducing costs, improving efficiency and getting senior management buy-in for investment.

Have the right data and technology infrastructure in place

All the above may be only enabled by the robust centralised data and IT infrastructure. A new industry leading approach of such infrastructure is outlined in Exhibit 17 and Exhibit 18 below:

Exhibit 17: New approach to EWS – a pipeline driven Data-As-An-Infinity Loop architecture

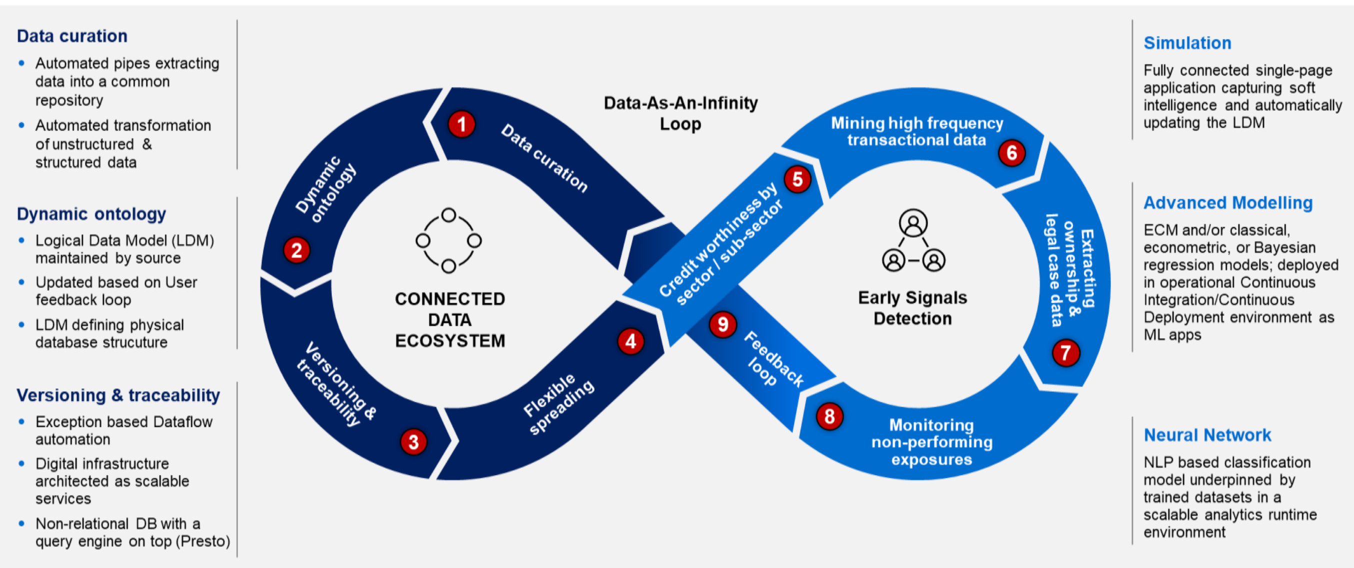


Exhibit 18: Comparison of pipeline driven Data-As-An-Infinity Loop architecture vs traditional 'ETL' approach to data

Category	Traditional solution: 'ETL' data domain based	New approach: End-to-end smart data platform
Data architecture	Siloed architecture with pocket solutions	Pipeline driven Data-as-an-Infinity loop architecture
Data lineage and attribution	Data treated as a static stock asset: locked in a closed environment, limited versioning & traceability	Data treated as a fluid asset: connected to data pipes with full versioning and traceability
Manual vs. AI and automation	Addresses a single specific business need via manual transformation and translation processes	Uses neural network frameworks to solve the problem at scale and speed e.g., NER (Named Entity Recognition), NLP (Natural Language Processing) vocab, Semantic similarity
Data ingestion/ transformation process	Implements a data model with limited flexibility, unable to cater for unstructured data, prone to errors & hard to maintain	Employs a flexible Ontologic model capable of modelling any data type & operates as a lossless data abstraction layer
Reusability and scalability	Minimum efficiency due to lack of reusability	Maximises efficiency through reusable components

Use Machine Intelligence capabilities effectively in building the EWS

A Machine Intelligence driven approach covering the entire data lifecycle can accelerate effectiveness of EWS significantly. The approach can enable banks to ingest and source structured and unstructured data from operational and third party systems algorithmically using tailored extractors, bots and crawlers. Data can be transformed into structure ready for analysis using document structuring and clustering techniques. Automated data quality detection and correction of transformed data for anomalies, gaps and errors can be tackled with AI (human-in-the-loop) - using reinforcement learning and natural language processing capabilities. Relationships between datasets in the form of entities and objects can be developed using automated tagging, parameterisation of data and entity relationship detection. Rule-mining and clustering algorithms can be used to recommend corrections to data quality errors and quantify confidence. Data should be stored in its rawest form (accessible to query and processing systems) and ready for model training including annotation for supervised learning. Data model should operate as a lossless layer capable of modelling almost any type of user defined data model – irrespective of the format. Machine learning models should be trained against processed data – often building on top of a model pre-trained on a corpus of external data. Model performance should be analysed, validated and audited as part of an iterative loop often including re-training through additional data. Trained models should be executed against input data in real time and outputs with CI/CD capability and integrated into user-facing applications in a structured and repeatable way as an infinity loop.

Exhibit 19: Machine Intelligence capabilities in an E2E EWS solution

Data lifecycle

	1 Data sourcing and ingestion	2 Data transformation	3 Data ontology	4 Data analysis	5 Data spread
Objectives	Automatic ingestion of financial, transactional and alternative data from multiple sources	An algorithmic template approach applied to transforming and extracting data from documents	Builds a connected data ecosystem that is accessible to query and processing systems	Builds, trains and deploys ML driven models with robust methodology against data either real time or offline	An application generating outputs for users. An interface for analysts & DS to derive insights and collaborate
How machine intelligence helps	Automated downloading of data from files into pipes using Python based scripts or APIs (where possible)	Data are tagged automatically & collection of analytical transformations on curated data – not just transforming to a single format	A highly flexible ontologic model which is capable of modelling almost any type of user defined data model	Development of advanced analytics & ML models. Evaluation of advanced analytics / ML model outcomes	Embeds data models into user-facing applications. Ready-for-analysis data is served as a REST API
Illustration					
AI tools integrated	Data source discovery Custom web crawlers Data extraction via pipes Data upload auditing	Document structure analysis Data clustering Data sanitisation Data formatting	Metadata tags applied & parameters calculated Data versioning & time stamping Data model – conceptual, logical and physical data model Data lineage and data dictionary	Advanced analytics / ML models for analysis Evaluation of advanced analytics / ML model outcomes	Serving of analytics / ML model outcomes UX design developed for model outcomes Models connected to the Front-End and data exposed via API DB & API maintenance, DevOps and MLOps

LEGO framework approach to effective predictive early warning triggers

The ingredients outlined in this report on what is necessary for creating a world-class EWS led Galytix to our LEGO (Leverage, External Indicators, Governance and Ontology) framework. This makes necessary adjustments to existing indicators, adds new indicators and improves the overall hit ratio. This is all underpinned by a systematic and automated data and technology infrastructure.

Implementing a LEGO framework gives banks a far reaching, detailed and timely understanding of their customers. From a credit risk perspective, banks will be able to make smarter, faster and more-informed credit underwriting decisions. These capabilities are not just useful for credit and risk monitoring but also for the business as a whole.

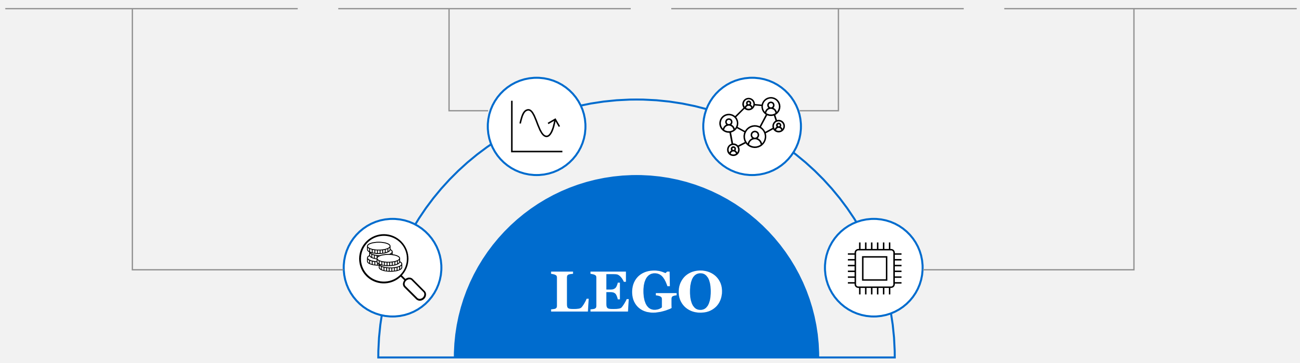
Exhibit 20: LEGO framework generating insights at the sector and sub-sector level can make all the difference

Leverage
Financial statements need to be adjusted for relevant balance sheet items to reveal the cash position and leverage

External indicators
Market prices, analyst earnings forecasts/ ratings, bankruptcies, climate risks and projections by sector

Governance
Indicators of aggressive accounting, fraud and legal risks

Ontology
Fully connected & dynamic data ecosystem with accurate data definitions, labelling, lineage and relationships



Leverage

Need to **adjust financial ratios** for other balance sheet items

Leverage is a focus on making sure the current financial ratios used are relevant to predicting corporate distress. For too many recent bankruptcies there has been a dependence on gross debt to EBITDA ratios without making the necessary adjustments. High profile examples in the UK in which material changes in accounts receivable and payable and end of period fluctuations covered up much greater actual indebtedness were Thomas Cook and Carillion[19].

In an age of increased investment, free cash flow ratios can't be ignored as a supplement to EBITDA. Rising interest rates across most major nations also makes percentage change in interest cover (EBIT divided by interest expense) just as relevant as Debt to EBITDA.

For the SME segment, banks must utilise various additional inputs in determining early warnings of distress. This includes the number of recent credit checks on the owner, whether owner borrows him/herself, degree to which owner utilises his own credit facilities, company's borrowing history with any bank, owner's credit card repayment record, owner's age and owner's number of banking relationships.

An EWS platform must integrate all this data automatically from financial statements into banks financial spreading models and through to early warning systems. Where there are firm and group specific adjustments that need to be made, we provide the flexibility for corporate credit analysts to make these changes themselves.

[19] Why-we-chose-not-to-jump-on-board-thomas-cook | www.schroders.com ,
Loophole that Brought Down Carillion May Be Widely Used in US, Europe | businessinsider.com ,
Carillion hid debt implications of reverse factoring | theconstructionindex.co.uk

External indicators

40% decline in share price or

25% decline in earnings estimates can be an early warning indicator

Equity market signals are often earlier indicators of impending corporate distress than the bond market. Most credit analysts still focus on the latter though. In the examples of Thomas Cook[20] and Carillion[21] , the share prices started to collapse one year and six months before bankruptcy owing to profit warnings. In other examples where there has been accounting fraud involved such as Wirecard and NMC Health share prices reacted swiftly and sharply while credit rating agencies acted only days before bankruptcy.

Although thresholds will vary for each lender, at Galytix we believe material enough changes to be recognised as early warnings are when share prices of a borrower (or its peer group) has fallen by more than 40% over the last twelve months or equity research analysts' earnings estimates have fallen by more than 25% for coming years. Share prices are of course volatile and often driven by equity market sentiment as much as fundamentals so must be used carefully. They can be a major source of false positives if used incorrectly. At Galytix, we believe they must be combined with the other indicators in our LEGO framework in a multivariate approach.

Where there are major suppliers or a concentrated client base, information on financial distress at these firms can be relevant. This is particularly relevant today with banks looking at the exposure of their corporate counterparties to Russia in terms of clients or suppliers. Those borrowers that currently rely on commodities imported from Russia are likely to see large changes to their cost structures.

Another external indicator that the Galytix platform looks at are hard bankruptcies that have occurred in the past either across industry peers or at subsidiaries of a borrower. Many large firms can have dozens of subsidiaries and we make automated checks across bankruptcy courts and company house type institutions in numerous countries. In a recent client case Galytix identified 1,186 company defaults including 713 subsidiary defaults and 473 group company defaults over multiple decades. The neural network that underpins this data discovery, ingestion and curation can also be leveraged to look for connections to sanctioned firms and individuals from the same or different data sources.

A good example of a new external risk is climate change. With insurers pulling away from certain markets, companies like utilities can be exposed. The bankruptcy of California power utility PG&E Energy in 2019 was owing to liability claims of $25bn related to the use of faulty equipment that caused wildfires. But climate changes that causes major wildfires, hurricanes and floods are likely to disproportionately impact such industries and their credit quality.

Governance

LEGO catches **fraud**

This is a huge growth area for early warning detection. Galytix's approach is to focus on a short list of high impact items such as Board/management turnover and insider selling (more than 25% change would be suitable threshold to be considered material), accounting and legal (allegations of fraud) and cyber. This is based on empirical evidence of which factors are material enough to drive an increased probability of default on a firm's credit. We reduce false positives by following a strict multistage data validation process starting with entity matching that the right firm is being picked up in any news flow scrapping and then a "hard" match which needs specific and accurate information that could trigger a bankruptcy from a credible information source. For instance, with Wirecard, the Galytix platform would have reached this "hard" match in January 2019 (eighteen months before bankruptcy) when there were allegations of falsifying accounts & money laundering in the Financial Times. Once a "hard" match has been reached the search can be broadened to a wider list of related topics.

[20] Thomas Cook issues profits warning after UK heatwave | Thomas Cook | The Guardian
Thomas Cook falls 20% as short-sellers pile in after profit warning | Business | The Times
Why did Thomas Cook collapse after 178 years in business? | Thomas Cook | The Guardian
[21] Timeline - Carillion collapses under debt pile after profit warnings | Reuters
Carillion Collapse: What To Know About Its Liquidation | Fortune

Filtering of news flow whether by humans or machines is only as good as the relevance and quality of data sources used. These can vary by firm and industry. For instance, in logistics there may be well known trade press or blogs in shipping and transportation that may be relevant. Although incorporating social media can create excessive amounts of data noise pollution, where the individuals are well respected this can provide insights before mainstream publications. A good example is a famous short seller like Muddy Waters, which in numerous bankruptcies has made accurate and substantiated claims several months and quarters ahead of the press.

Ontologic Data model connecting data in a single ecosystem

Ontology – All Data in One Model. One Model for All Data.

No single list of early warning indicators is a magic bullet, without the right data and technology infrastructure to support it. There is a need for a platform that provides front-to-back automated systems from data ingestion through to curation, analytics and alerts. It must be able to incorporate structured and unstructured data from internal and external sources into the same data model, demonstrate clear derivation of outcomes and combine modelling with expert judgement.

The combination of data engineering, data science and technology teams with industry domain expertise is crucial to building a scalable framework with data governance around lineage, accuracy and definitions ensuring user journeys can be automated.

The use of weighted multivariate and decision-tree models is a key component of ensuring the data from L, E & G discussed already comes together with limited false positives

Case example: Implementing LEGO in a machine intelligence driven smart data platform to transform credit underwriting and monitoring

A smart data platform with automated pipelines engineering Data-As-An-Infinity loop generates EWIs that support better, faster and more-informed credit underwriting and monitoring. A Pipeline-Driven architecture implemented in a unified data infrastructure that engineers all data types (structured, semi-structured and unstructured data) covering both internal and external data sources with high-integrity versioning and traceability on data. The engineering process includes:

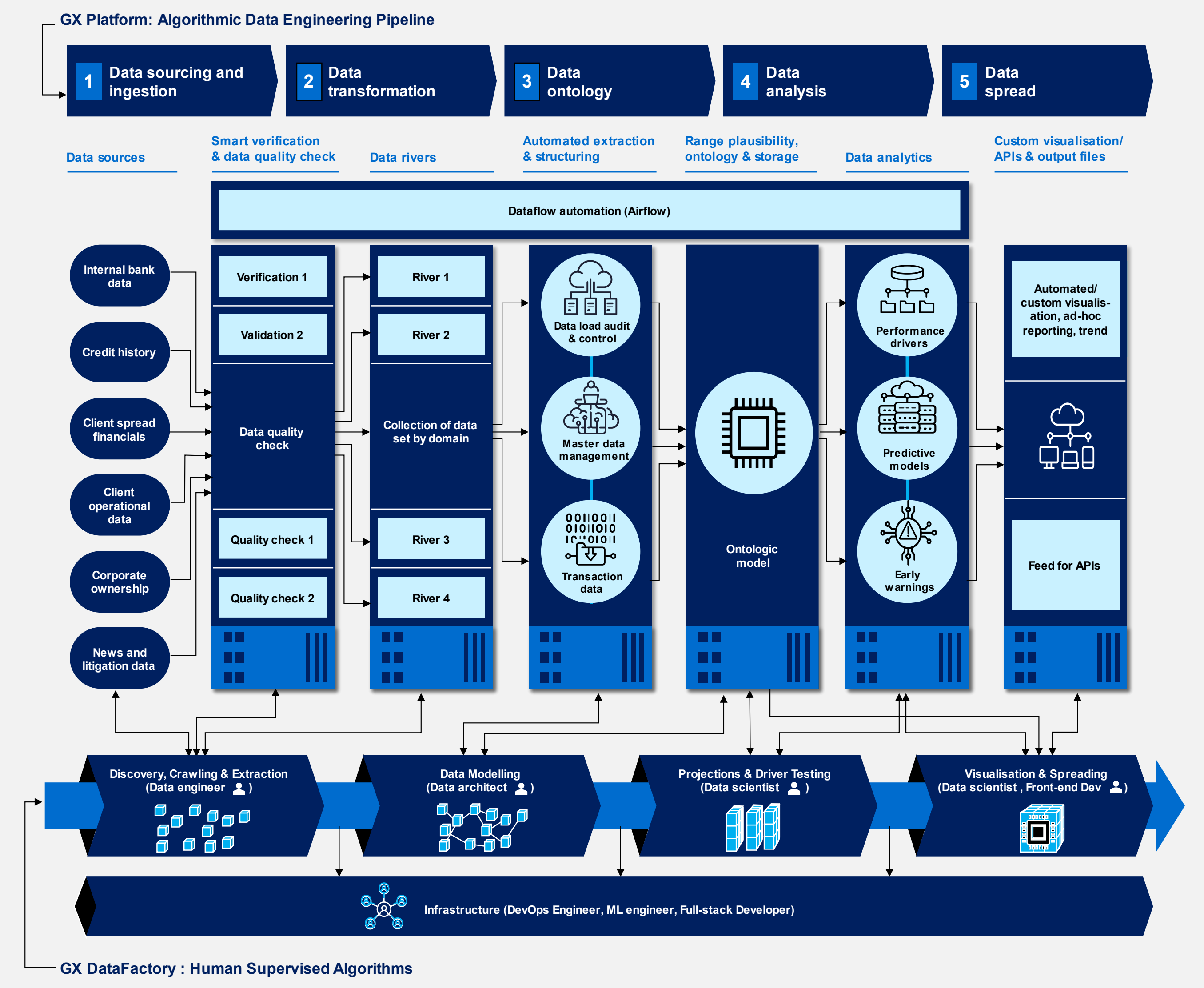
100% data accuracy ensured

- Data sourcing and ingestion: Algorithmically extracts relevant external data from multiple sources and from in-house operational and third party systems. System caters for heterogeneous formats covering structured, semi-structured and unstructured data sources. Pipes are attached to rivers of data to ensure sourced data is continuous – not extracted as a batch.
- Data transformation: Depending on the data format, an algorithmic template approach is applied to transforming and extracting data from unstructured / structured documents e.g. supply chain, sub-sector performance, annual report. Data are tagged automatically based on prescribed rules as per business requirements. The data architecture allows for a collection of analytical transformations on curated data – not just transforming to a single format. Furthermore, it delivers data to storage, building and aligning the schemas between source and destination.

10x more predictive power and data granularity at segment and sub-segment level

- Data ontology: Builds a connected data ecosystem that is accessible to query and processing systems – optimised for consistency, performance and scale. Uses non-relational database with ability to handle heterogeneous data formats. The data ontology and data dictionary is built dynamically and kept up-to-date based on periodic functions as the new data becomes available.
- Data analytics: Builds, trains and deploys ML driven models with robust methodology against data either real time or offline with Continuous Integration / Continuous Deployment capability.
- Data spreading: Visualises data in a single page application generating outputs for internal and external users. Provides an interface for analysts and data scientists to derive insights and collaborate. Embeds data models into user-facing applications. Ready-for-analysis data is served as a REST API.

Exhibit 21: Machine Intelligence capabilities embedded in a pipeline-driven architecture



Human supervised
Algorithmic DataFactory

Algorithmic DataFactory can significantly accelerate data quality capabilities – high quality data delivered as an infinity loop

Data quality in data sourced from unstructured and structured data sources is a significant hurdle for banks in developing effective EWIs. To overcome this issue, banks are deploying human supervised algorithms in a DataFactory set-up. The machine intelligence driven DataFactory engineers data into fuel – feeding continuous discovery, ingestion, curation and transformation of data (irrespective of format) into EWS engine.

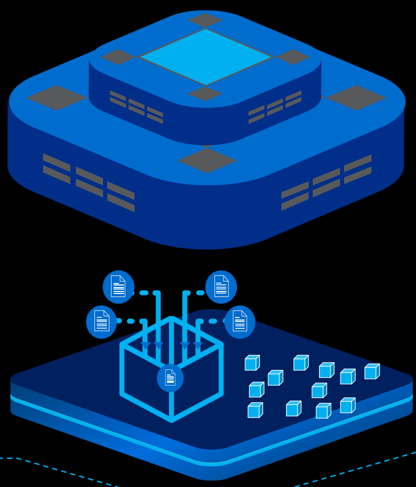
The DataFactory is set up as a production line, similar to what one may find in a manufacturing factory. The factory includes dedicated production cells which operate in a single data pipeline environment covering the entire data engineering lifecycle. Most processes in the factory are algorithmic, supervised by humans who apply reusable tools and repeatable processes.

Data-as-an-Infinity loop
with zero errors and anomalies

The automated pipelines connect and stream relevant data from sources as and when it becomes available. Depending on the data format, an algorithmic template approach is applied to transforming and extracting data from unstructured/structured documents e.g. supply chain, sub-sector performance, annual report. Data are tagged automatically based on prescribed rules as per business requirements. The data ontology and data dictionary is built dynamically and kept up-to-date based on periodic functions as the new data becomes available. A single, flexible document-based data ecosystem houses all data types. An automated high-integrity versioning and traceability to data process runs throughout the pipeline as part of scheduled batch processing. Finally ready-for-analysis data is served as a REST API.

GX

Algorithmic DataFactory (human-in-the-loop)



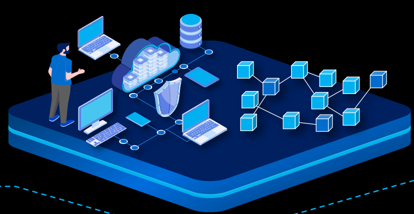
1 Data sourcing and ingestion cell

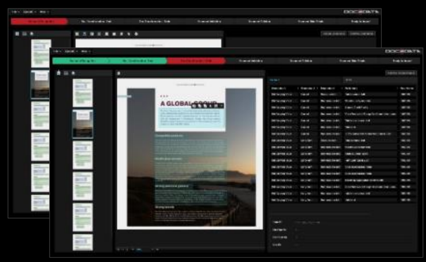
- Data source discovery
- Custom web crawlers
- Data extraction via pipes
- Data upload auditing

2 Data translation & transformation cell

- Document structure analysis
- Data clustering
- Data sanitisation
- Data formatting

Data Engineer supervised algorithms





3 Data credentialing, tagging & Parametrisation cell

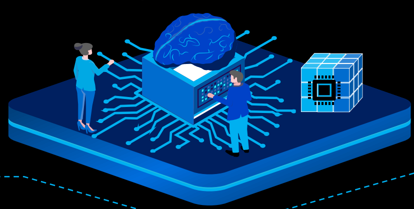
- Metadata tags applied
- Parameters calculated
- Data versioning
- Time stamping

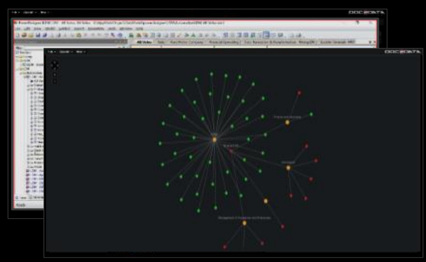
Data Engineer supervised algorithms

4 Data ontology and analysis cell

- Data ontology – conceptual, logical and physical data model
- Data lineage and data dictionary
- Advanced analytics / ML models for analysis
- Evaluation of advanced analytics / ML model outcomes

Data modellers and Quant Analyst supervised algorithms

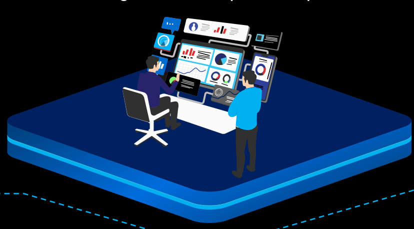


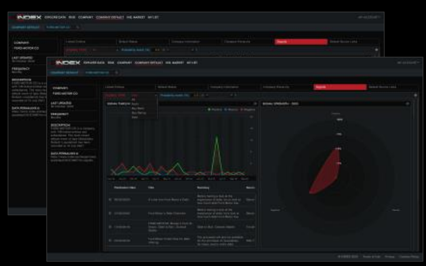


5 Data spreading & infrastructure cell

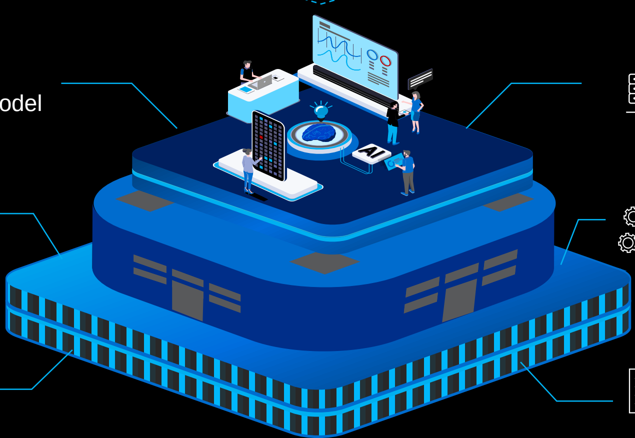
- Serving of analytics / ML model outcomes
- UX design developed for model outcomes
- Models connected to the Front-End and data exposed via microservices
- DB & API maintenance, DevOps and MLOps on Cloud (AWS and Azure)

Front-End, Back-End, Software Engineers, DevOps & MLOps





Key benefits



More scalability: All data connected in a single data model

More speed: 100% automation in engineering data E2E

More predictive power: 10x more effective risk insights & signals

More granularity: Up to **60%** additional data at subsegment level

More consistency: Standardisation across regions, accounting norms

More reliability: 100% accurate data ready for end users

4 stage Neural Network
productionised in the GX Platform

Neural network framework to track early warning distress signals

A disciplined approach to data discovery and curation through alerts is crucial to minimising false positives. A neural network that seeks to recognise underlying relationships between datasets can achieve this at scale and speed. The four stage GX neural network outlined in the exhibit below extracts features from more than 240 different unstructured and structured data sources through a continuous streaming data pipeline.

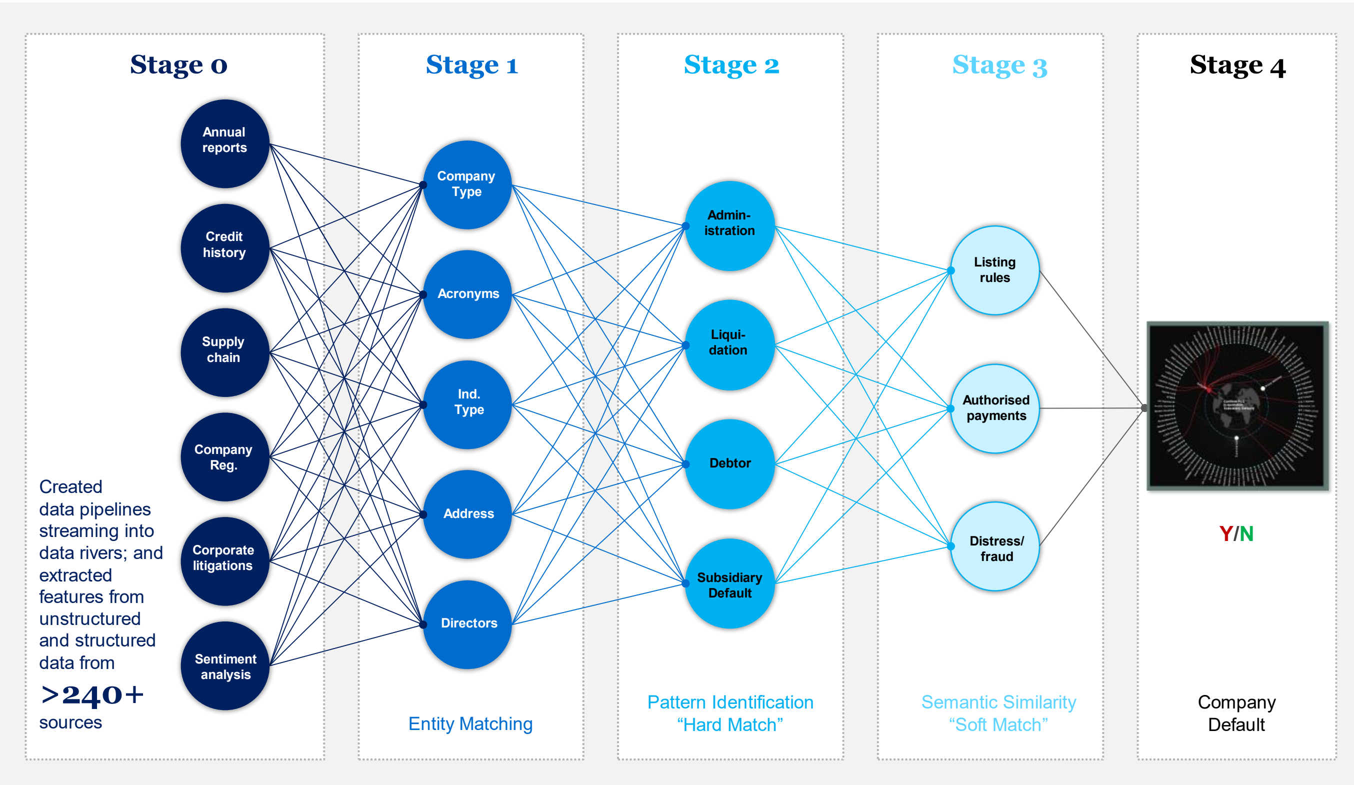
The first stage is taking all the data discovered and screening it to ensure only the data on the right parties, in this case corporate borrower, are selected. Although this may sound simple, data sources can pick up many companies or subsidiaries with similar names or abbreviations. Hence entity recognition including checking company types, addresses and other characteristics is crucial.

The second stage or "hard match" involves the GX algorithms automatically finding detailed information on debts including on all subsidiaries of parent companies. Twenty years of relationship histories and data on previous bankruptcies are analysed from court cases, companies' registries, and other sources.

The third stage involves analysing new data and news sources that show signals of distress or fraud leading to the final automated signal in Stage 4.

Like other parts of the GX data ecosystem, this data engineering process and neural network algorithms are flexible and reusable and so the data pipeline below can be altered, as necessary. The broad approach is for a reverse funnel structure where specific and accurate data from high quality sources is needed early in the process and once some signs of distress or matches create strong signals then the neural network looks further out at a wider range of weak signals. This reverse funnel approach guarantees that a substantial number of alerts are not created and that risk managers in both the first and second line of defence are not inundated for no reason.

Exhibit 22: Bank case example: Neural framework identifying early warning distress signals for a counterparty



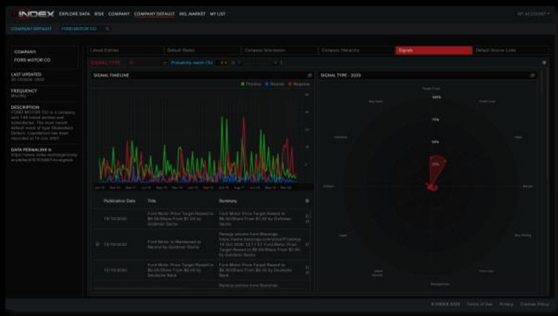
Dynamic visualisation layer connected to the pipeline with smart features and convenient to use – not a dashboarding tool

An effective EWS needs a visualisation capability that enables end users to consume the derived output in order that it leads to actionable insight. The visualisation layer should provide connectivity into the pipeline and exhibit 5 capabilities:

- **High impact visuals:** Well designed data visuals that regularly leads to actionable insights driving multiple operational early warning related operational workflow
- **EWI feature forecasting:** Offers users the ability to forecast each feature using pre-defined forecasts or adjust modelled forecasts. The feature should also allow users the ability to access forecasting accuracy results for prior periods for both features and target.
- **Feature selection and processing:** Support the users in selecting the most important features for modelling, e.g., correlation with target indicator. The UI should also make potential trade-offs of features apparent. In case the features need to be transformed, the user is presented with a set of standard processing operations which can be applied.
- **Simulate models real-time:** Offers users the functionality to simulate modelling outcomes. The functionality also allows users to define coefficients and the users have the ability to access a base case for feature weights and adjust each weight individually. The models can also be staged allowing users to work at local level and publish models when completed
- **Data Exploration:** Allows users to explore the data corpus by its metadata and time series view. The users can select/tag data for further analysis processing

Exhibit 23: Bank case example: Implementation of LEGO Framework to track 12 unique distress early warning signals

Predictive Early Signals of Distress

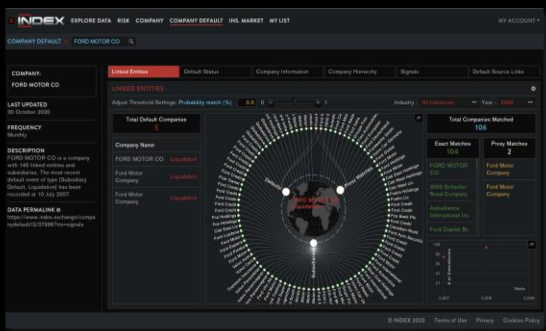


12
Unique signals tracked

240+
sources

1 Financial Strength

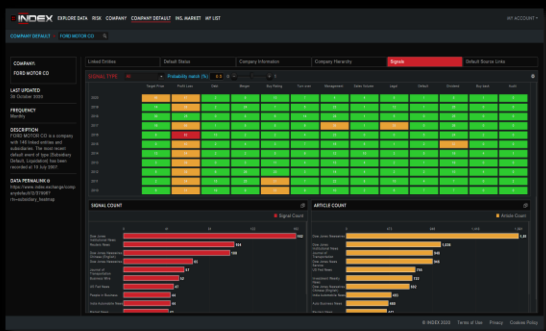
- Capital pressure
- Profitability pressure
- Business performance
- Credit ratings



~40%
Additional predictive insights based on sectors & obligors

2 Regulation Impact

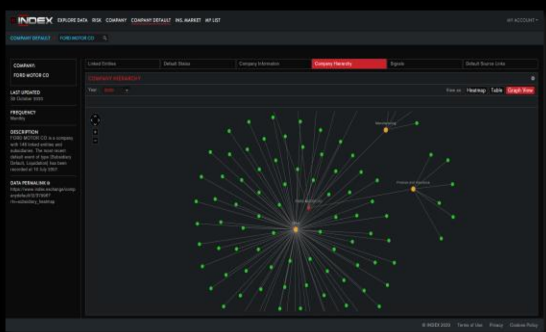
- Regulatory changes
- Risk factors
- Macro-economics



~95%
Risk factor related changes tracked

3 Company Networks

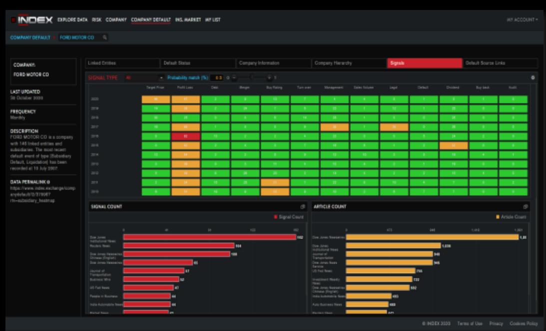
- Company mentions
- Company interactions
- Peer analytics



>300
Subsidiaries tracked on average per each counterparty

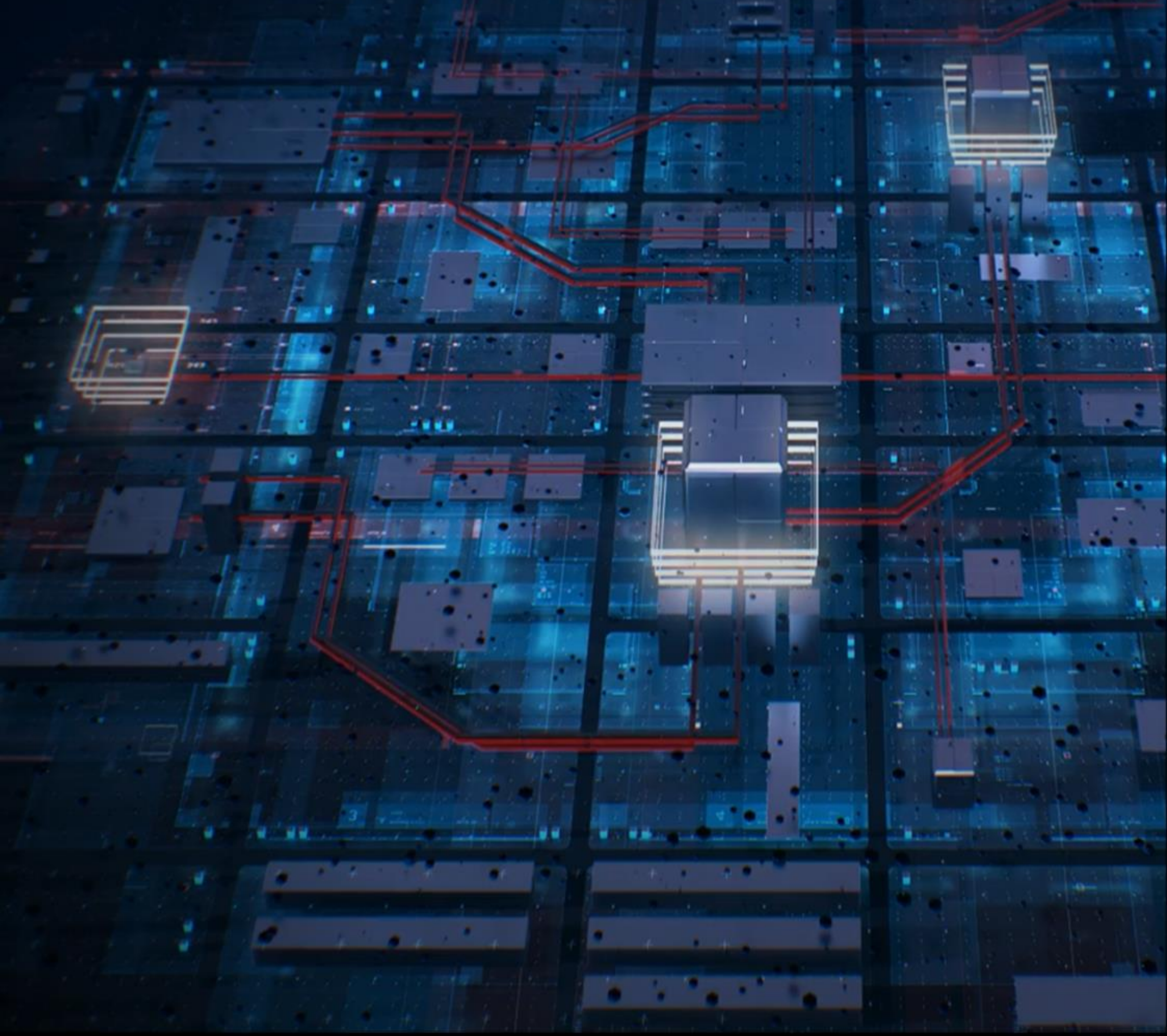
4 Management Changes

- Change of governance
- Board / CEO
- Key personnel



100%
Management changes tracked

It is impossible for any bank to identify all its risky customers before their default. However, it is possible for banks to establish a prudent system and processes to identify and monitor a significantly higher proportion of accounts with potential for default. Establishing state of the art EWS is a journey that requires decision making at both strategic and tactical level. The next few years are crucial for any bank with aspirations to land on the right side of the credit cycle described in this report. The important thing, in our view, is to start the EWS journey today by picking the right challenges and confronting them without reservation. In an era of a changing credit cycle, doing nothing is probably the riskiest approach.



Appendix

Basis for ROE projections

We have used multivariate methodology to calculate ROE across the three scenarios outlined earlier in this report. The assumptions in this ROE tree are based on our client experiences and engagement with a wide range of banking industry subject matter experts. For ease of use we have used $100m as our hypothetical equity base.

Looking at our three scenarios of pure corporate lending ROEs we have assumed a loan book which is around 8x the size of equity. This is equivalent to a double digit Equity/risk-weighted assets ratio and a circa one for one risk weighting of the loan book. In a tougher credit environment there is a chance that banks may shrink the size of their loan book but potentially lower loan losses owing to this would be offset by lower revenues and higher cost income ratios:

- **In Scenario 1 "Base case"** loan losses of 50bps are a 330bps drag on the ROE of 5%. These loan losses are broadly in line with the recent run-rate across major global markets. Using a hypothetical equity base of $100m pretax profits would be $6.25m, pretax loan losses would be $4.2m and pretax profits before loan losses would be $10.4m.
- **In Scenario 2 "Turn in the cycle"** we assume an increase in loan losses of 75bps to 125bps. This is more in line with the average levels seen in recent decades including both good and bad times. Using the hypothetical equity base of $100m, the pre-tax profits before loan losses of $10.4m would be wiped out completely by loan losses. In reality this will typically be offset by a re-pricing especially in a rising interest rate environment. We have assumed 200bps benefit to ROE from wider net interest margins based on our work with leading industry participants. These assumptions generate our Scenario 2 ROE of 2%.
- **In Scenario 3 "Recession"** we have assumed a further increase of loan losses by 120bps to a total loan loss ratio of 2.45%. This is more in line with the dot.com bubble and financial crisis. Using the easy to follow equity base of $100m for illustrative purposes and the 8x leverage pretax loan losses would increase to $20.4m. The increase in loan losses versus Scenario 2 is equivalent to an 800bps drag on ROE. Offsetting this we assume that net interest margins expand further and benefits ROE by another 200bps, given such a scenario is likely to be accompanied by higher interest rate rises in the current inflationary environment. Hence we assume a Scenario 3 ROE of -4%.

For the analysis of the benefit of EWS to ROE we assume a 20% reduction in the loan losses outlined above. This is derived from our client experience. The loan losses may of course vary from bank to bank depending on the sophistication of their EWS implementation. Applying a 20% loan loss reduction assumption to Scenario 1, the pre-tax loan losses of $4.2m would be reduced by $0.8m and add 70bps to the ROE of 5%. In Scenario 3 this benefit increases to $4m pretax or circa 330bps uplift to ROE. There is also likely to be a regulatory capital benefit but this is unlikely to be immediate. Our client experiences suggest this would be in the 10% range.

□ □ □

Contacts

Galytix

Raj Abrol
CEO and Chief Product Architect
Email: Rajiv.abrol@galytix.com

Rupak Ghose
COO and Head of EWS product team
Email: Rupak.ghose@galytix.com

PwC

Symon Dawson
Partner, PwC United Kingdom
Email: Symon.k.dawson@pwc.com

Matt Moran
Partner, PwC Luxembourg
Email: Matt.moran@pwc.com

Natasha Rakova
Director, PwC United Kingdom
Email: Nataliya.x.rakova@pwc.com

Roxane Haas
Partner, PwC Luxembourg
Email: Roxane.haas@pwc.com

Think Data Differently

Galytix Ltd
www.galytix.com
Designed by GX Design Team

https://twitter.com/galytix
https://www.linkedin.com/company/galytix