SUERF
THE EUROPEAN MONEY AND FINANCE FORUM

**SUERF Policy Note**
**Issue No 154, April 2020**

# The Prudent Algorithm Principle: A new paradigm for financial supervision

By Jakob Thomä
2° Investing Initiative

*JEL-codes: D53, E61, E60.*

*Keywords: Algorithm, financial supervision, financial policy.*

*The bedrock of financial regulation focuses on the need to regulate the people making the investment and financing decisions on behalf of society. The Prudent Person Principle and notions of fidicuary duty are designed to govern that decision-making in the interest of the common good. Increasingly however, it is not the people but algorithms making decisions across the investment chain, from the investment advice process all the way to actual investment decision-making. In that environment, financial supervision needs to evolve as well. This paper argues for the need fo a "prudent algorithm principle", governing the deployment of algorithms in financial markets. This principle should be predicated on the idea that financial institutions have the responsibility to identify for each algorithm the outcome it informs, the nature by which the algorithm can represent a threat to the public policy goals associated with that outcome, the auditing procedure by which financial institutions plan to prevent this outcome, and the 'fail safe mechanism' designed to prevent contagion or scaling of negative outcomes should they materialize.*

Industrial regulation – that is, the rules governing the "commanding heights" of the economy – can be neatly categorized into two types of rules. First, those governing the people. And second, those governing machines and the goods those machines produce. All regulatory interventions fall under one of these two categories.

On the whole, manufacturers are governed by rules related to the goods and services produced. For instance, Europe does not allow genetically modified food or chlorinated chicken. In medicine, the U.S. Foods and Drug Administration defines the standards that drugs must meet in order to be commercialized. Environmental standards regulate the design and operation of industrial sites (with admittedly mixed effects) and an Alexandrian library of rules is there to ensure the safety of nuclear power plants. As the globe deploys 5G, questions of security inform the discussion of involving Huawei as a technical partner in many jurisdictions.

In the services sector, however, rules have traditionally focused on labour. The guilds of the Middle Ages have made way for a myriad set of rules related to licensing and other qualifications needed to do different jobs. Nurses, hairdressers, and taxi drivers all fall under this category.

Meanwhile, finance – one of those "commanding heights" – has generally been focused on regulating people. Of course, that is not to say that rules on products don't exist (I have a few friends who are strongly nodding their heads as they read this). But rather that the primary set of rules in finance has always been focused on the people conducting the transactions, not the abacuses of the Italian renaissance or the high-powered computers of modern-day Wall Street.

Arguably, the two core regulatory concepts governing transactions in financial markets are "fiduciary duty" and the "prudent person principle". According to the German financial supervisor Bafin, the "prudent person principle stipulates that insurers may only invest in assets and instruments whose risks the undertaking concerned can properly identify, measure, monitor, manage, control and report and appropriately take into account in the assessment of its overall solvency needs. All assets are to be invested in a manner that ensures the security, quality, liquidity and profitability of the portfolio as a whole."

The concept of fiduciary duty (generally found under common law systems) involves a range of responsibilities of the fiduciary to the beneficiary - the most prominent being the "duty to exercise reasonable care, skill, and diligence".

These two rules govern the nature of decision-making by *people* as they invest for the general public, through pensions, insurances, retail markets, or otherwise.

The problem? We are about to – and have arguably already entered – a world where machines, not people, make these decisions. The entire investment chain, from advice to trading, is being taken over by algorithms. In a sense, the financial sector of tomorrow will look more like a digital version of the heavy industries – machines creating things.

As this revolution is under way, financial supervisors and lawmakers will have to think about a framework for controlling these machines. In short, this will entail moving and building from a "prudent person principle" to a "prudent algorithm principle".

Of course, the first question that comes to mind may be, *"Why would we need new rules for machines? Can't we simply ask them to practice the same type of care as people would, to ensure that they only invest in assets they can "properly identify", to act in the best interest of their beneficiary?"*

This article by no means claims that those concepts are outdated. Rather, it argues that they don't meaningfully comment on what we expect algorithms to do or how we expect them to act. Airplane pilots are not allowed to fly intoxicated and must have their wits about them. But it is also critical that the planes they fly are safe. As commercial aviation took off, so the regulatory framework adjusted to develop new rules and principles. The aviation industry soon began to realize that it's no longer enough to govern just the behaviour of the agents: they must also enact rules governing machines. A pilot must be sober. And a machine must base its decisions on inputs from at least two sensors – a principle that was cruelly violated in the Boeing 737 MAX crisis.

The question then is what such a prudent algorithm principle would look like. Finance is not alone in facing the growing concern about the 'interpretability' or 'explainability' challenge associated with algorithms. We know how to build the machines, even if we can't understand why and how they do what they do.

The advantage from a regulatory perspective, however, is that we don't need to necessarily know why and how they do their work; all we need to ensure is that their work doesn't represent a threat to the common good – in this case financial stability, consumer protection, and broader adherence with policy objectives. That is not to say such a task is easy. It does, however, organize the principle itself.

The **Prudent Algorithm Principle** can thus look like this:

*"Financial institutions using algorithms to inform decision-making must ensure that both the decision-making process within that algorithm and the outcomes it suggests do not threaten or otherwise work against public policy goals in financial markets (e.g. stability). Financial institutions have the responsibility to identify for each algorithm the outcome it informs, the nature by which the algorithm can represent a threat to the public policy goals associated with that outcome, the auditing procedure by which financial institutions plan to prevent this outcome, and the 'fail safe mechanism' designed to prevent contagion or scaling of negative outcomes should they materialize. The Prudent Algorithm Principle also confirms that financial institutions are legally liable for the outcomes created by the algorithms."*

Let's unpack this. Ultimately, we are worried about two things when we think about algorithms and the need to supervise them. First, algorithms can create 'sensitive outcomes' (e.g. flash crashes, herding behaviour, inefficient capital allocation due to some systemic biases that are undesirable on the whole). Second, algorithms may deploy sensitive or illegal processes as it arrives at that outcome (e.g. sourcing sensitive or illegally obtained data, insider trading, violating consumer protection principles in its decision-making process).

So, the first step in deploying any algorithm is clearly defining i) the objective of the algorithm and ii) the mechanism by which the algorithm could hypothetically fall into one of the two categories defined above (sensitive process or outcome).

Second, the financial institution needs to identify mechanisms to identify whether or under which circumstances this undesirable outcome or process might come into play. Here, financial institutions can build on the burgeoning literature addressing the question of 'interpretability' of algorithms. This includes random simulations at scale to identify conditions of undesirable outcomes, and targeted simulations related to specific issues (e.g. testing outputs based on input data designed to identify racism or other types of discrimination). This could also include simulating a pre-defined set of events based on specific macro trends (e.g. Brexit, climate change).

Third, it needs to identify 'fail safe' or response mechanisms should these undesirable outcomes or processes materialize. Crucially, the Prudent Algorithm Principle is general in the sense of not outlawing algorithms that may under certain circumstances deliver such undesirable outcomes within the Principle itself. For example, if financial supervisors want to prevent lending in sectors that involve illegal practices (e.g. child labour), then the rule needs to be built into the system itself, and one can't rely on the algorithm to identify that independently. In addition, macro analysis across algorithms deployed to identify systemic correlations is also under the purview of the supervisor, although part of the prudent algorithm principle also involves recognizing and considering such realities.

Each of these items above relate more broadly to the idea that algorithms may not be fully understood, but understanding them is necessary so as not to render ignorant its user and supervisor as to its action and potential reactions.

Finally, the Principle does not prescribe that algorithms under no circumstances ever create undesirable outcomes. Indeed, if we wanted such a world, we should probably start by getting rid of humans first before we start attacking algorithms. It does, however, reiterate the accountability of financial institutions for these algorithms where they break the law.

This final point raises questions similar to other areas where algorithms create potential legal liabilities – such as autonomous driving. Navigating the brave new legal world of crime and punishment in a world where algorithms commit such crimes will be a challenge, the scale of damage being potentially much higher than that a person can inflict. Supervisors and governments may be worried raising those liabilities and the effect they may have on a smooth functioning of markets. However, they also act as a deterrent for creating algorithms that could be at risk of creating such damages – and not monitoring their behaviour. Some might argue that it is exactly this dynamic that leads to moral hazard as banks "bet on the bailout". However, it is exactly that moral hazard that a Prudent Algorithm Principle has the chance of addressing.

As with every new concept, it must be steeled in the furnaces of public opinion and ground in the mill of public policymaking. It must be turned and weighed so as to ensure it is not found wanting. Whatever the dynamic, as algorithms rise, so must the policy response.

## About the author

**Jakob Thomä** is Managing Director at the 2° Investing Initiative, Senior Research Fellow at the Japanese Financial Services Agency, and Teaching Fellow at the School of Oriental and African Studies, University of London. Jakob previously served as Academic Advisor to the Bank of England and currently supports the development of the climate scenario analysis work of EIOPA. He led the development of the first tool to measure the alignment of financial portfolios with climate goals – PACTA – now applied by over 1,000 organizations around the world. Jakob has also been recognized by a range of organizations, including Forbes "30 under 30 Germany" list in 2019. He holds a Ph.D in Finance & Management from the Conservatoire National des Arts et Métiers.

## SUERF Policy Notes (SPNs)

| | | |
|---|---|---|
| No 149 | A diverse monetary union creates invisible transfers that justify conditional solidarity | by Enrico Perotti and Oscar Soons |
| No 150 | How deep will it fall? Comparing the euro area recessions of 2020 and 2009 | by Andreas Breitenfellner and Paul Ramskogler |
| No 151 | Mother Nature: The gender-climate nexus | by Jessica Murray |
| No 152 | European banks in the corona crisis | by Jan Schildbach |
| No 153 | Passive Funds Actively Affect Prices: Evidence from the Largest ETF Makets | by Karamfil Todorov |

SUERF
THE EUROPEAN MONEY AND FINANCE FORUM

SUERF is a network association of central bankers and regulators, academics, and practitioners in the financial sector. The focus of the association is on the analysis, discussion and understanding of financial markets and institutions, the monetary economy, the conduct of regulation, supervision and monetary policy. SUERF's events and publications provide a unique European network for the analysis and discussion of these and related issues.

SUERF Policy Notes focus on current financial, monetary or economic issues, designed for policy makers and financial practitioners, authored by renowned experts.

The views expressed are those of the author(s) and not necessarily those of the institution(s) the author(s) is/are affiliated with.





SUERF Secretariat
c/o OeNB
Otto-Wagner-Platz 3
A-1090 Vienna, Austria
Phone: +43-1-40420-7206
www.suerf.org • suerf@oenb.at