Catastrophe modelling and metaphors in financial markets: A reply to Etzion, Kypraios, and Forgues

Andreas Tsanakas[1] and Laure Cabantous[2]

15 January 2020

Abstract: We welcome Etzion et al.'s (2019) effort to critically assess the role of cat models in insurance markets, by combining a sociology of finance lens with statistical analysis. Nonetheless, we believe there are two flaws in their analysis. First, their interpretation of the model-as-engine metaphor, as well as the way they test for this metaphor, is questionable, in part because it is not clear what form of performativity (generic, effective, Barnesian?) is to be expected in the case of cat models and what constitute rigorous tests for these various forms of performativity. Second, we disagree with the broad conclusion drawn by the authors from the statistical analysis, in particular that the predictive performance of cat models is "not demonstrably better than guesswork" and hence, that such models are not (like) cameras. Overall, we find it hard to see how the two main points discussed by the authors – namely that "catastrophe bonds do not lend themselves to analysis through conventional sociological theories of financial markets" (p. 1) and that cat bonds are not appropriate risk transfer instruments to tackle sustainable development goals – are proven through their discussion of cat models via the model-as-engine and model-as-camera metaphors.

Introduction

Etzion, Kypraios, and Forgues (2019) have recently taken on the challenge of assessing the effectiveness of catastrophe bonds ('cat bonds') – a financial instrument used by insurers and reinsurers to transfer the risks associated to the underwriting of catastrophes and natural disasters to market investors (e.g., institutional investors, hedge funds). Using

[1] Cass Business School, City, University of London; a.tsanakas.1@city.ac.uk
[2] Cass Business School, City, University of London; Laure.cabantous.1@city.ac.uk

insights from the sociology of financial markets and building on statistical analysis, the authors conclude that the "*modeling which underlies catastrophe bonds is not demonstrably better than guesswork at predicting the financial consequences of extreme events*." (p. 1). They also warn against the use of cat bonds to meet sustainable development goals.

We welcome the authors' attempt to use a sociology of finance lens to discuss the value of cat bonds as financial instruments and their critical assessment of catastrophe models' ('cat models') performance. At the same time, we believe that there are two flaws in their analysis. First, we disagree with Etzion et al.' (2019) interpretation of the "model as engine" metaphor and challenge the way they use this metaphor to test for the role cat models play in the insurance market. Second, we disagree with many of the conclusions that the authors draw from their statistical analysis of cat models' effectiveness. Before addressing each point in turn, we briefly remind readers of the 'model-as-engine' and 'model-as-camera' metaphors.

Metaphors about economic models: Models as engines and as cameras.

In "The Methodology of Positive Economics", Friedman (1953) defended the view that economics is an *engine* to make predictions about market outcomes. This view, however, does not imply that models have to accurately describe economic agents' individual behaviours. The realism of behavioural assumptions is not important inasmuch as the theory is able to predict market outcomes (e.g., price, quantity at the equilibrium). Instead, Friedman suggested that economics could be judged along two criteria: simplicity and fruitfulness.

In "An Engine not a Camera. How Financial Models Shape Markets", the sociologist of financial markets, Donald MacKenzie (2006), highlights the *performative* relationship between economics and the economy. He argues that economics "is an engine, [but] in a sense not intended by Friedman: an active force transforming its environment, not a camera recording it" (MacKenzie 2006, p. 12). The 'model-as-camera' metaphor hence, refers to economic models' ability to represent market behaviours. It expresses the idea that models mirror economic phenomena that pre-exist their (abstract) representations. The 'model as engine' metaphor instead refers to a performative view of economic models: economists, through their theorization and modelling practices, 'construct' the economic phenomena they seek to understand (e.g., they name, measure, and give shape to them).

Testing for the engine metaphor: are cat models ‘engines’?

Etzion et al. (2019) interpret the model-as-engine metaphor as follows: “*in the engine metaphor, (…), the theorization that underlies models also shapes what happens in markets*.” (p. 19).

Generic and effective performativity in the case of cat models

At the most generic level, MacKenzie’s argument that economics is a *performative* engine means that economic models, when used by market actors (*generic performativity*), change the way markets work (*effective performativity*). Economics actively participates in the functioning of markets in two ways: i) by providing market actors with concepts and models that can be used like tools, for instance to compute prices, and ii) by changing economic processes (Gond et al., 2016). With these definitions in mind, one can see that even without a formal test or systematic qualitative study of market participants’ practices, the fact that insurers use cat models – even if they were to use them only for symbolic purposes – means that these models have a generic form of performativity.

As for ‘testing’ for the effective performativity of cat model, such as test would require, in our view, a counter-factual case: for instance, an insurance market without cat models. We could then study whether the use of cat models by market actors makes a difference: for instance, do models change actuaries and underwriters’ daily practices? Does their use lead to significantly different market outcomes?

Cat models and Barnesian performativity

The model-as-engine metaphor can also refer to the idea of *Barnesian performativity* (MacKenzie 2006): when the use of models increases the correspondence between the theory and the phenomenon that the theory is meant to describe. While the notions of generic and effective performativity do not directly point to any increasing correspondence between the economic phenomenon and its theoretical representation, the notion of Barnesian performativity focuses precisely on the feedback loop between a phenomenon and its representation.

Barnesian performativity seems to be the interpretation of the model-as-engine metaphor that the authors favour in this paper, since they explain that saying that cat models are (like) engines implies “*that changes in the theorisation that underlies the models* *generate corresponding changes in how market actors price the financial instruments predicated on these models*" (…) (p. 26). Furthermore, when testing for this metaphor, they refer to the "*linkage between the modelled losses (potential costs) and cat bond spread (potential benefits) would tighten. In other words, increasingly accurate modelling should translate into some measure of financial certitude…"* (p. 27). We interpret this sentence as meaning that if cat models drive cat bond market prices then, as time passes, with models becoming more accurate, the spread between cat bonds and similarly rated corporate bonds should shrink. While such a view is not unreasonable, we would expect such a possible trend to have a minor effect, not easily observable from a short time series, given the persistent uncertainty around e.g. climate trends.

The authors then write that, since "*spreads are not at all related to changes in expected losses or in the underlying catastrophe risk models,*" (…) *"the factors that have been driving cat bonds over the past 20 years are not an outcome of financial theorization, despite the sophistication of catastrophe modelling. Cat models, unlike derivatives (MacKenzie & Millo, 2003) are not engines*." We struggle to follow the logic behind this conclusion. It is expected that the prices of financial products are influenced by, say, macroeconomic conditions. How does this observation invalidate the view of cat models as engines (in the Barnesian performativity sense)? Conversely, what kinds of econometric data would one realistically need to see, in order to conclude that models are indeed engines?)

More importantly, in our view, applying the notion of Barnesian performativity to cat models means that the use of cat models contributes to 'create' a phenomenon that is closer to its theoretical depiction. Thus, in order to assess the claim that cat models are (like) engines, it is central to understand what kind of theory is embedded in cat models.

Unlike many economic models, cat models are not meant to "record market behaviours" nor are they supposed to "encourage market actors to behave in accordance with [economic] theory" (p. 18). Cat models are much less close to economic theories than, say the Black-Scholes model. Cat models are constructed as representations of (a) the random processes by which the natural phenomena are generated and (b) the mechanisms through which these phenomena are mapped to insurance claims. Together, these two elements allow the calculations of probabilities of events of interest, such as the event that at particular cat bond is triggered and the corresponding expected loss. These probabilities consequently 'inform' the prices of cat bonds, but cat models themselves do not compute, or predict, cat bond prices. If economic theory plays a role, it is located in how exactly this 'informing' works and is thus external to cat models' structure.

There is a broad sense in the market in that cat model outputs would drive market behaviour – for instance, if cat models show that there is a larger probability of the bond being triggered, then investors would seek higher returns on investments in that cat bond. Yet, the link between cat model outputs and market behaviour is not rigidly dictated by economic theory; there does not seem to be a clear agreement between insurance practitioners as to how loss estimates generated by cat models are mapped to cat bond prices, even in the primary market (e.g. Bodoff and Gan, 2009). It is hence not clear that the notion of Barnesian performativity is applicable to cat models, unless we are ready to consider the possibility that the modelling of say, hurricanes, impacts on the probability of hurricanes.

Testing for the 'camera metaphor'

Eztion et al. (2019) explain that "*In the **camera** metaphor, models are merely devices that help understand the reality of what transpires in markets. They are mathematical representations of market processes and produce knowledge." (p. 19).* A direct implication of this metaphor is that, for the authors, cat models – like cameras – must precisely and accurately record the behaviour of the phenomena they are meant to depict. In other words, cat models' forecasts must have small confidence intervals and must "*correspond with historical record*" (p. 19). While we agree with Etzion et al.'s (2019) view that the "model as camera" metaphor conveys a representationalist view of models (and theories), and hence points to the criteria of precision and accuracy, we challenge the way they interpret their statistical analyses.

Testing for 'precision'

Etzion et al. (2019) first consider whether catastrophe models are 'precise'. They demonstrate that variations of key model parameters have a very high impact on the outputs of catastrophe models. For example, Figure 4 shows that, if the distribution chosen for a key random variable related to windspeeds ('V_{max}') is set to its 5% and 95% confidence bounds, there is a very large variance in modelled expected loss costs. This is the vexing problem of 'parameter uncertainty', which is well understood in the loss modelling and actuarial communities (e.g. Cairns, 2000).

We welcome the clear illustration of such uncertainty by the authors. Nonetheless, we do not see how this justifies the conclusion that models are "indeterminate" or that they "don't predict better than guesswork" (p. 6). First, we note that wide confidence intervals do not imply that all values within them (e.g. between 0.5 and 1.5 in Figure 4) are equally plausible. While a frequentist construction of confidence intervals does not explicitly allow for assigning probability weights to the values they contain, central estimates are used because of their higher credibility, not as random guesses within confidence intervals. Second, there is a certain arbitrariness in how such uncertainty is represented. The width of confidence intervals depends on ad hoc values of the confidence levels: on could plausibly represent parameter uncertainty via the 80% or 99% confidence levels, with rather different results.

Furthermore, we note that the choice of what is simulated as a random variable within a model and what is considered as a parameter (that is, a deterministic variable, whose chosen value may be 'wrong') is contingent on practical constraints and conventions. For example, an alternative technically valid way of representing uncertainty, would be to select a different distribution for V_{max} within each simulated scenario, e.g. from the uncertainty set in Figure 35 of Risk Management Solutions (2013). As a result, the uncertainty depicted in Figure 4 would disappear – or, rather, be repackaged as higher loss volatility (possibly leading to higher triggering probabilities). This is not to claim that uncertainty in predicting extreme events does not exist (it *does* exist and it *i*s high), but that care should be taken when interpreting constructs such as confidence intervals or intermediate variables.

Testing for 'accuracy'

Etzion at al. (2019) make the crucial point that the attachment probabilities generated by catastrophe models are not observable, but can be still validated via observed frequencies of cat bonds being triggered – in other words by cat models' predictive performance. The authors demonstrate that the number of cat bonds triggered is much smaller than what one would expect, if the probabilities produced by cat models were accurate, concluding that the modelled attachment probabilities are overstated.

While the conclusion is drawn correctly, we find the strength of evidence to support it exaggerated. First, we note that the data set involved is too small to provide reliable estimates of predictive performance. The attachment probabilities of 612 tranches are analysed. But these are not independent observations: natural catastrophes do cluster; triggering events in the same year are highly correlated as they are driven by the same events; tranches in the same deal will be sensitive to the same underlying event. Many of those tranches will either be triggered, or not, at the same time. Hence the effective sample size, in terms of information content, is much more modest. This is implicitly acknowledged in footnote 7, where the authors explain that, while their data collection stops at 2016, a large number of cat bonds may be triggered in 2017.

Furthermore, the reliability plot in Figure 6 is misleading. It appears that cat models perform very badly for higher forecast probabilities of attachment, for which no actual triggering events are observed. But, from Figure 5 we can see that there are very few deals that correspond to that range of attachment probabilities – hence the number of data points stuck at the horizontal axis are not particularly credible. The first few points on the bottom left of the graph, still showing an over-prediction of triggering events by cat models, give a more meaningful picture.

Finally, Etzion et al. (2019) conduct a simulation study to support their conclusion on cat models over-predicting triggering events. Unfortunately, this study is flawed. The authors justify their choice of a Poisson distribution by referring to Poisson processes, which are used in actuarial modelling to represent the numbers of loss events arising from a portfolio. However, the Poisson distribution is not a meaningful choice for the modelling of variables that only take values in the set {0,1}, which corresponds to the simulation taking place (a simple Bernoulli model could be used instead). This inconsistency drives strange modelling choices; in particular, the implicit condition $p(0) = 1 - p(1)$, leading to equation (5), has no meaning in the context of the Poisson distribution.

A separate issue with the simulation is the assumed independence between triggering events, referred to above. The distribution of numbers of events is extremely sensitive to dependence assumptions, with even a small correlation between events changing radically the shape of the distribution (e.g. McNeil et al., 2015, Figure 11.1). Accounting for such correlation would lead to much greater spread in the distributions of Figure 7, making the observed event frequencies appear less implausible, given the model.

Conclusion

In their conclusion, Etzion et al. (2019) argue that financialization, exemplified by cat bonds, is not necessarily an appropriate path towards tackling the Sustainable Development Goals. This may well be an accurate assessment. Nonetheless, we find it hard to see how the point is proven through their discussion of cat models via the model-as-engine and model-as-camera metaphors.

References

Cairns, A. J. 2000. A discussion of parameter and model uncertainty in insurance. ***Insurance: Mathematics and Economics***, 27(3): 313-330.

Bodoff, N. M., & Gan, Y. 2009. An analysis of the market price of cat bonds. ***CAS E-Forum***, May 2009.

Etzion, D., Kypraios, E., Forgues, B. 2019. Employing finance in pursuit of the Sustainable Development Goals: The promise and perils of catastrophe bonds. ***Academy of Management Discoveries***.

Friedman, M. 1953. The methodology of positive economics. ***Essays in Positive Economics***, pp. 3–43. University of Chicago Press.

Gond, J.-P., Cabantous, L., Learmonth, M., & Harding, N. 2016. What do we mean by performativity in organization studies? A review of the literature. ***International Journal of Management Reviews***, 18(4), pp. 440-463.

MacKenzie, D. 2006. ***An Engine, Not a Camera. How Financial Models Shape Markets***. MIT Press.

McNeil, A. J., Frey, R., & Embrechts, P. 2015***. Quantitative Risk Management: Concepts, Techniques and Tools.*** Revised edition. Princeton University Press.

RMS Management Solutions. 2013. North Atlantic Hurricane Model RMS® RiskLink 13.0 (Build 1509).