Red tape: Does economic regulation have an overfitting issue?[1]

Lennart Baumgärtner

Executive summary

"Overfitting" is a phenomenon in statistics where **models are needlessly complex. Overfit models are very elaborate yet inaccurate**. Such models look precise and sophisticated, but the precision is illusory: they reproduce the past while obscuring the future. The same dynamics can emerge in regulatory frameworks and models when detail accumulates faster than insight.

Economic regulation in the UK has increased in complexity over recent decades with the introduction of cost-based regulation, more stringent performance incentives, and the growing capital program across sectors. As a result, **parts of the regulatory model have become overfit**, such as granular adjustments to the WACC, or the Totex modelling in water and wastewater. Other elements are underfit, such as CAPM-based Cost of Equity or Totex modelling in aviation, where there are great gains from making better use of data, even if this implies more complex models. Overfitting is arguably the bigger problem.

To avoid overfitting and develop resilient frameworks, regulators and companies need to **continuously prune the regulatory model**.

- **Adding more data** enables more granular insights. Companies are increasingly reporting real-time performance data that can be used to make regulatory insights more bespoke and timelier.
- **Model validation** is critical to identify and prune existing over- and underfitting. Machine Learning inspired validation can distinguish material insight from noise.
- **Recognizing model imperfections** can help differentiate planning and execution mistakes. Uncertainty mechanisms are an effective tool to mitigate impacts from inevitably imperfect models.

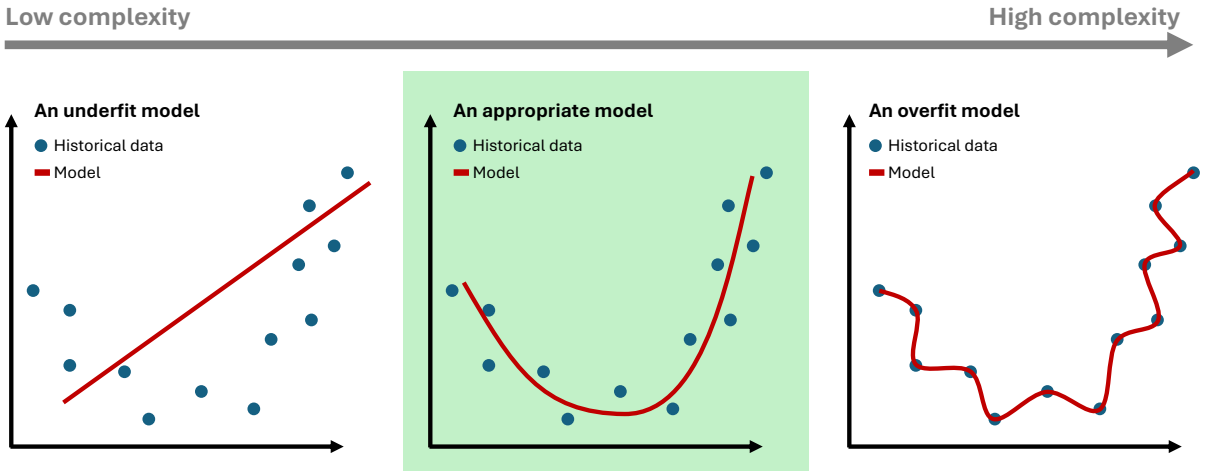
It is in the interest and responsibility of both companies and regulators to support such pruning. In practice, this can be done via **explicit processes** (e.g., post-mortem analysis) and **strict governance** around regulatory complexity. Regulators and companies need to work actively against overfitting. **Powerful structural and behavioural incentives** otherwise lead companies and regulators to add mechanisms that optimize their own position yet collectively push the sector towards fragility. Establishing no-blame culture and robust toolkits is essential to support a sustainable regulatory model.

[1] Many thanks to Robert Ritz, Cameron Hepburn, Jorge Cardenas, Martin Malinowski, Gavin Knott and other regulatory experts for helpful comments. All comments were made in a personal capacity.

What exactly is overfitting and underfitting?

Overfitting and underfitting are statistical terms used to describe how well an empirical model captures reality. A model is "overfit" when it is needlessly complex. An overfit model usually reproduces historical datapoints with great precision. However, when applied to new situations, it falters: the model over-interprets and misses broader patterns. At the other end of the spectrum a model may be "underfit" if it is not complex enough. Its simplicity means that it fails to describe real-world situation and offers limited explanatory value. Models can also suffer from both over- and underfitting at once: capturing some aspects in excessive detail while neglecting others.

Illustrative comparison between an underfit, appropriate, and overfit model. The underfit model is too simple. The overfit model is too complex. Appropriate models capture key effects and discard noise.



Overfitting is widespread, even in the academic literature. When academics over-interpret their data, their outcomes are misleading[2]. Underfitting is less common but similarly problematic. When academics are missing key pieces of data, their results can be biased and oversimplified[3]. Over- and underfitting dynamics can also appear in organisational processes. When procedures are designed around the specifics of past situations, they become less effective when circumstances change. We end up learning the quirks of history rather than the principles that matter.

Overfit models and processes are problematic for two reasons. Firstly, their insights cannot be applied outside the narrow context in which they were developed. Most commonly, an overfit model closely matches historical data but tells you very little about the future. They work well ex post, but poorly a priori. Secondly, overfit models are expensive. They require large amounts of data, computational effort, and lengthy discussions to develop and interpret whilst having limited real insights.

[2] A prominent example: Kotz, Levermann and Wenz retracted their Nature paper on climate change damage function after Schötz (2025) demonstrated that their model was overfit.
[3] See also Baumgärtner et al. (2025), *The need for better statistical testing in data-driven energy technology modelling*, Joule

Is UK regulation overfit?

Lengthy discussions with limited insights have become common in regulation. The CAAs headcount grew from 1,195 in 20/21 to 1,602 in 24/25 (34% growth)[4]. Ofwat's headcount grew from 260 in December 2021 to 458 in December 2025 (76% growth)[5]. Ofgem's permanent staff grew from 1,012 in 20/21 to 2,110 in 24/25[6]. Much of this is from additional regulatory mandates (e.g., Ofgem's supply regulation). However, the fact that Ofgem's final determination for RIIO-ET3, GD3, GT3 counts 1,420 pages speaks for itself, not accounting for any of the associated consultations and company submissions.

Yet consumer outcomes are falling behind international comparators. Heathrow is amongst the most expensive international airports[7]; UK electricity prices sit among the highest in Europe[8]; and the water and wastewater sector faces unprecedented operational and financial problems. This discrepancy begs the question if regulation has become needlessly complex. In other words, has UK regulation become overfit?

We focus on economic regulation in water & wastewater, electricity transmission networks, gas distribution networks, and aviation, regulated by Ofwat, Ofgem, and the CAA. The three central pieces of economic regulation are cost modelling, operational incentives, and cost of capital (CoC) modelling. Cost and CoC modelling are used to set ex ante revenue allowances. Operational incentives are ex post penalties or rewards based on company performance. All of these rely on substantial consultation processes and econometric models during price reviews.

As **Table 1** illustrates, economic regulation shows signs of both over- and underfitting. We observe a mixture across sectors and regulatory tools that is worth untangling.

Regulators follow different cost modelling approaches. Ofwat relies heavily on econometric modelling. Historical data is fitted to company-specific variables to arrive at a company-specific Totex benchmark. This approach was criticized by the Cunliffe report as it "has not taken sufficient account of company-specific conditions and challenges.[9]" This is not a critique on the econometric approach per se. Instead, this is due to the econometric model being overfit—its accuracy is overstated due to the limits in historical data and future insight. Getting this accuracy right is important since it can help navigate potential ex-ante and ex-post adjustments to the econometric approach[10].

---

[4] CAA annual reports across Safety and Airspace, Consumers and Market, Aviation Security, CAA International, Space, and Miscellaneous regulation
[5] Ofwat monthly workforce management information
[6] Ofgem annual reports
[7] CAA (2025), *Working paper on regulatory models*
[8] IEA (2024), *Industrial electricity price comparison* & DESNZ (2025), *Quarterly Energy Prices April–June 25*
[9] Independent Water Commission, Final Report, July 2025
[10] The current discussions around the CMA's use of shrinkage estimators to set baseline Totex is another good example of this challenge. Shrinkage estimators, such as LASSO can reduce overfitting but need to be carefully calibrated and reviewed.

Table 1: *Risks of over- and underfitting in economic regulation.*

	Water & Wastewater	Electricity transmission	Gas distribution	Airports
Cost modelling (Totex)	**Overfitting risk** – Ofwat relies on econometric benchmark modelling for its Totex allowance. The models closely match historical data but are not tested for their forecasting ability, putting them at risk of over overfitting. Moreover, poor historical performance of water companies does not drive confidence in using this data to forecast future costs.	**Granular engineering reviews** – Ofgem emphasizes granular cost reviews based on market benchmarks and engineering reviews, both for Opex and the large Net Zero Capex. This approach aims to mitigate the risk of over- or underfitting but is resource intensive.	**Overfitting risk for Net Zero pathways** – Ofgem relies on multiple cost models for gas distribution costs. Modelling costs on historical data is appropriate during stable regimes. However, the changing role of the gas network during Net Zero puts this approach at risk of overfitting on historical precedents that no longer hold.	**Underfitting risks, addressed for H8** – For the current price review (H7), the CAAs approach of baseline + adjustments is simple, risking underfitting of costs. The CAA aims to address this for H8 using additional benchmarking. Whether or not that will lead to overfitting remains to be seen.
Operational incentives (e.g., ODI, PCD)	**Over & Underfitting risks, partially mitigated through engineering reviews** – All regulators apply a wide range of incentives (Ofwat: 24, Ofgem ET: 21, Ofgem GD: 30, CAA: 37), covering all or some of the regulated companies and with different degrees of financial implications. Each comes with its own target and exclusion principles. The large number of metrics covers diverse consumer impacts but risks *overfitting* where incentives are correlated. Individual assets can impact multiple incentives. Regulators rely on engineering reviews and committees to mitigate some of these challenges. Company performance against outcome metrics depends on different internal and external factors. Discarding external effects risks *underfitting* operational incentives. In many cases, it is unclear if worse performance is due to mismanagement of assets or previous errors in system planning and Totex allowance.			
Cost of Capital modelling (WACC)	**Over & Underfitting risks** – UK regulators apply different variants of CAPM, as per UKRN guidance. Within the CAPM framework, regulators are at risk of *overfitting*. Averaging and adjustment mechanisms on the Cost of Equity (CoE) and Cost of Debt (CoD) increase complexity while providing limited insights of statistical significance, such as inflation wedges, beta averaging, or debt index calibration. At the same time, it is well known in the economic literature that CAPM itself is *underfit*. There are multiple additional factors that impact rational investor behaviour not captured by CAPM, leading to biased CoE estimates. Amongst other factors, investor returns depend on cost allowances and operational incentives, but neither are considered for the WACC.			

Compared to Ofwat, Ofgem relies more heavily on bottom-up engineering reviews to set cost allowances in electricity networks. This reduces overfitting and improves accuracy but comes at high administrative costs. The CAA reflects a less costly modelling approach. They reduce overfitting by simplifying the econometric model. This in turn was criticised for underfitting, leading to additional cross-checks in the future.

The gas sector points to another central concern of overfitting. A model may be appropriate in one context but overfit in another. Gas distribution cost modelling was appropriate in the past but is at risk of being overfit when the sector's environment changes through Net Zero. As a result, there is growing concern in the sector that gas networks will be insufficiently funded in scenarios of prevailing gas boilers.

The operational incentive and CoC modelling show less differentiation between sectors but need to be viewed with some nuance. In both cases, we observe simultaneous risks of over- and under-fitting. All regulators use at least 20 operational incentives for their sectors and specific companies. This covers wide-ranging consumer preferences but risks overfitting the operational outcomes. Since some outcomes are correlated (e.g., asset health impacting multiple outcomes) the operational incentives may be needlessly complex. At the same time, individual incentives are at risk of being underfit if they do not consider the specific context of the assets they are applied to. For example, operational outcomes today depend on historical planning decisions. Companies might be rewarded for historical gold-plating or penalised for planning mistakes.

A similar mixture of over- and underfitting risks can also be observed in the CoC modelling. The use of CAPM is universal across regulators. It is well known that CAPM badly underfits the CoC. Investors care about additional risks beyond market volatility yet these are discarded under CAPM[11]. Furthermore, regulation shows clear signs of overfitting *within* the CAPM framework. In a heroic effort to correct the failures of CAPM, regulators apply complex averaging and adjustment mechanisms, many of them ad hoc and not statistically significant. CAPM still does not explain why utility betas fluctuate widely or why the return on equity in regulated utilities (5-6% real) is so different from other infrastructure projects, such as Tideway (4-5% real) or Sizewell C (10.8% real). Economic theory tells us that this leads to suboptimal investment decisions in the sector.

How to avoid over- and underfitting

There is no silver bullet to avoid overfitting. Neither is there an optimal model complexity since models need to be tailored to the data available. There are, however, two general ways to reduce model misspecification. One is to add more data. The other is to identify the appropriate level of model complexity given the data available. Some model

[11] Most prominently, Arbitrage Pricing Theory and the Fama French models demonstrate how investors price additional macro and micro risks. The inclusion of additional risks does not necessarily lead to a higher CoC but lowers the CoC for around half of the equity market.

uncertainty always remains and regulators need to identify the right model-uncertainty mechanisms to mitigate its impacts.

Adding more data can enable higher model complexity. This is akin to Ofgem's cost model in electricity or some project finance where practitioners aim to create "complete contracts" that consider all potential up and downside scenarios. Companies have made significant progress in making data available to them and the public in recent years, such as the widely available APIs on water company operational performance[12] and electricity network performance data published by network companies and NESO. Financial data has also evolved since the beginning of CAPM in the 1960s. On top of traditional stock prices, we now have data on order-books, ownership, and company fundamentals available to us. Most of these novel datasets remain untapped by regulators and companies alike although AI is starting to shift this boundary. Unfortunately, adding more data does not always help. In gas, for example, there are no precedents to decommissioning a legacy asset base of this scale and relevance, so adding more performance data is unlikely to help navigate the energy transition.

Model validation helps identify the right model for a given dataset. Model validation is a complex topic where Machine Learning research has made significant progress in the last decades. Modern model validation techniques can substantially reduce the risk of overfitting by quantifying data correlations and forecasting ability. Model validation can move us towards the Goldilocks zone, where a model is robust (and handles new data well) but also parsimonious (as simple as possible). So far, regulators and companies have done little model validation, leaving substantial room for improved model performance[13]. Commonly used **model cross-checks**, where the results from different models are benchmarked against each other, can be an effective means of model validation. They, however, need to be treated with care. When cross-checks are correlated to the original model outcomes, they can provide a false sense of accuracy. One of the most effective model validation methods are **materiality thresholds**. Model complexity is only increased if it materially improves results. Regulators have made some use of this but do not apply it universally or retroactively on past model changes.

A key point of model validation is to recognise that no **model will be perfect**. This is important in the regulatory context since it opens the conversation to alternative regulatory measures that deal with model imperfection. For example, Totex over- or underspend in water could be caused by imperfections in the Totex model rather than company over- or underperformance. Companies are not in control of model imperfections and should neither benefit nor be penalised for them.

[12] The "Surfers for sewage" website is a good example of how public data is used, where national sewage spills are reported live.

[13] For example, Ofwat's Totex model validation relies entirely on in-sample validation methods, leaving out critical assessment of overfitting.

Processes and governance to reduce overfitting

Overfitting persists because powerful structural, cognitive and social incentives reinforce it. This applies to companies and their consultants where overfit models may benefit individuals in the short term[14]. It also applies to regulators that that may optimise specific mandates (e.g., low bills over environmental outcomes) temporarily imposed by themselves and government. Nevertheless, it is in the consumers and investors long-term interest to reduce overfitting in regulation. Reducing overfitting thus requires an active impulse in the form of modelling processes, governance and culture, and the toolkit used in regulatory decisions.

Rigorous processes can counteract cognitive and social incentives. This means reviewing current regulatory models ex-post and new regulatory changes ex-ante. In Machine Learning, practitioners typically follow a gated process. Model features need to pass pre-defined quality gates before they can be applied in practice. The number of model iterations matters also matters. Models will become overfit if they are revised too often since they will "learn" the model selection criteria. It is therefore important to record the number of model iterations that were made. Better to stop development early before the models become overfit, even if the final model appears imperfect.

Governance and culture are critical to limit overfitting: model weaknesses only become visible long after systems are established, by which time short-run success and apparent precision have reinforced confidence in the existing design. Individuals' careers have been built on overfit models, making a move away from them ever more difficult. Regulators can counteract these drivers by establishing a culture that supports simplification, thresholds against small adjustments to the existing framework, and deploying advanced tools that focus on transparency rather than accuracy. Decision makers need to be held to account for regulatory complexity on top of regulatory outcomes.

Regulatory tools can further reduce the reliance on granular, narrative-driven reports with anecdotal evidence. Building on the current momentum in data transparency, regulators can embrace real-time transparency over granular ex ante assessments. This applies particularly to operational measures where real-time data is increasingly available. The use of AI can create additional narratives without leading into overfit models. Stress tests, scorecards, and proportional capital mechanisms can replace bespoke studies with instruments that are transparent, repeatable, and easier to supervise.

[14] Particularly on the Cost of Capital, companies continue to propose complex adjustments to the CAPM framework that have collectively driven the model into its overfit state. These adjustments have benefited individuals in the past, making it difficult for practitioners to let them go today.