When Is Segmentation Needed in Actuarial Models? Diagnosing Structural Incompatibility

Xana Verte

February 2026

Abstract

In many actuarial applications, portfolio segmentation is commonly considered as a means of improving predictive accuracy. However, under conditions of limited data, strong regulatory scrutiny, and significant social consequences of pricing decisions, segmentation itself becomes a source of model risk. Unwarranted segmentation may lead to overfitting, operational complexity, and difficulties in regulatory justification.

In this paper, segmentation is treated not as a modelling technique, but as a diagnostic hypothesis about the structural incompatibility of a global model. We propose a diagnostic framework that allows the actuary to assess whether observed model instability is supported by the data, or whether it can be more plausibly explained by noise, scale effects, or insufficient information. The approach is based on combining geometric locality in the feature space with measures of local disagreement in the conditional response distribution. A graph-based representation of local neighborhoods is used to detect conflicts between neighboring regimes, and an energy functional provides a principled way to evaluate competing structural hypotheses.

The framework admits both frequentist and Bayesian interpretations. In particular, the Bayesian perspective highlights posterior uncertainty over the number of segments and naturally enforces a preference for simpler structures when data are weak. Flat energy trajectories indicate high uncertainty and support retaining a single global model, while sharp transitions provide evidence of structural heterogeneity. Importantly, the absence of segmentation is treated as a valid and informative diagnostic outcome.

The proposed methodology is designed for practical actuarial settings with limited data and constrained computational resources. The framework emphasizes interpretability, transparency, and conservative decision-making, aligning naturally with regulatory requirements and public-interest considerations. A synthetic example inspired by strategic interactions in concentrated insurance markets illustrates how local structural conflicts may arise without pronounced changes in marginal distributions, and how the diagnostic distinguishes such cases from spurious signals.

Rather than optimizing segmentation for predictive performance, the proposed approach supports responsible control of model complexity by diagnosing when segmentation is structurally justified.

Keywords: Actuarial modelling, Portfolio segmentation, Model diagnostics, Structural heterogeneity, Limited data settings, Bayesian interpretation, Regulatory transparency

1 Introduction: Segmentation as an Actuarial Decision

Portfolio segmentation is widely used in actuarial practice as a way to account for portfolio heterogeneity and improve forecast accuracy. In applied contexts, it is commonly considered as a natural response when the quality or stability of a global model deteriorates: if a model appears not to hold uniformly across the portfolio, segmentation is often considered as a potential remedy.

In this work, segmentation refers specifically to the practice of splitting a portfolio dataset into distinct subsets and building separate models for each subset. This notion is distinct from incorporating segment indicators as features within a single model, from mixture modelling approaches, or from changing the model class to increase expressiveness. The diagnostic framework proposed here addresses the question of whether such dataset splitting is structurally justified.

Segmentation, however, is far from a neutral operation. It increases model complexity, heightens sensitivity to noise in the data, complicates maintenance and interpretation, and may have direct regulatory and social consequences. In many practical actuarial settings, data are limited, fragmented, or only weakly informative at a local level. Under such conditions, unwarranted segmentation can lead to overfitting, unstable premiums, operational complexity, and difficulties in regulatory justification.

From the perspective of actuarial risk management, the key question is therefore less about how to segment a model, and more about when segmentation is truly justified. Observed model instability may arise for various reasons, including noise, scale effects, temporal fluctuations, or insufficient information, and not all such cases reflect the presence of persistent structural heterogeneity.

In this work, we treat segmentation primarily as a diagnostic hypothesis about the structural organization of the data, rather than as a technical optimisation procedure. Introducing segmentation is interpreted as a claim that a single global model is fundamentally incapable of adequately describing the data and that the observed instability has a structural character. Accordingly, the actuary's task is to test this hypothesis, rather than to automatically increase model complexity.

We propose a diagnostic approach aimed at identifying and interpreting evidence for or against the structural incompatibility of a global model. The approach is oriented toward practical conditions in which data, computational capacity, and operational resources are finite. Particular attention is paid to situations in which the absence of convincing diagnostic signals in favour of segmentation should be regarded as a correct and informative result.

The proposed framework is designed to support the actuary's professional judgement in making decisions about model structure, reducing model risk, and ensuring the stability and interpretability of actuarial models, rather than to replace existing segmentation methods.

2 Motivation: When Standard Diagnostics Fall Short

2.1 A typical scenario

Consider a situation encountered by many practicing actuaries.

A model is regularly refitted on new data. Monitoring indicates relative stability of marginal distributions of the covariates. Business-level metrics—such as the average portfolio prediction or aggregated loss indicators—remain within acceptable ranges. At the same time, *internal model diagnostics* gradually deteriorate: residual variance increases in certain subgroups, coefficient estimates become less stable across refits, and local predictions lose reliability.

Importantly, such situations often lack the standard signals associated with data degradation. There is no clear evidence of data drift: classical indicators such as KS tests or PSI remain within acceptable bounds, although with limited data their statistical power is low and the interpretation of what constitutes a "normal" range relies heavily on expert judgment. There is also no obvious concept drift, as the global relationship between covariates and the response variable appears to be preserved.

Attempts at local model adjustments—adding features, increasing functional flexibility, or strengthening regularisation—fail to produce a lasting effect. The problem does not manifest as an abrupt failure, but rather as a growing form of procedural instability.

2.2 Common responses and their limitations

In practice, such situations typically prompt one or more of the following responses:

1. **Changing the model class**, moving toward more flexible models (for example, from GLMs to ensemble methods) or, conversely, toward more strongly regularised ones.

2. **Hyperparameter tuning**, which often yields short-term improvements but rarely addresses the underlying source of instability.

3. **Expanding the feature space** through interactions or nonlinear transformations of covariates.

4. **Introducing segmentation**, an organisationally and computationally costly decision that is frequently taken reactively in response to deteriorating internal diagnostics.

The shared limitation of these approaches is that they focus on *adapting the model* rather than on *understanding the source of instability*. Without a diagnostic interpretation, any intervention risks being either excessive or insufficient. In particular, segmentation may be introduced without a clear understanding of whether it reflects genuine structural heterogeneity or merely amplifies noise and marginal effects.

2.3 Structural incompatibility as a diagnostic hypothesis

This work proposes an alternative diagnostic perspective. Observed instability may be unrelated to data quality, model class selection, or insufficient expressive power of the model. Instead, it may arise in situations where *local conditional relationships between covariates and the response variable are internally incompatible*.

In other words, different regions of the feature space may be governed by distinct risk-generating mechanisms. Attempting to represent them within a single global model leads to internal conflicts, which manifest as parameter instability, increased local errors, and reduced reproducibility.

In such cases, segmentation is better understood as a *structural hypothesis* rather than as an ad hoc technical intervention. The central task is to develop diagnostic tools that allow this hypothesis to be assessed, distinguishing genuine structural incompatibility from the effects of noise, marginal data support, or transient fluctuations.

3 Approach: Diagnosing Local Conditional Conflicts

The proposed approach is based on a shift in diagnostic focus from global measures of model fit toward the analysis of local conditional behaviour. Rather than relying on aggregated performance metrics, we examine the coherence of local explanations in the feature space and seek persistent conflicts that cannot be attributed to noise or smooth variation.

3.1 From global fit to local coherence

Standard model assessment typically relies on global metrics such as deviance, AIC, or out-of-sample performance. While these measures are informative, they aggregate information across the entire portfolio and may obscure local structural issues.

We propose to focus instead on *local conditional behaviour*: how does the response variable behave in small neighbourhoods of the feature space? Are local explanations mutually coherent, or do systematic conflicts arise between them?

The key idea is the following. If a model is structurally adequate, local conditional distributions should vary smoothly and remain mutually consistent. If, however, the data contain incompatible local regimes, this incompatibility manifests itself as persistent conflicts between neighbouring regions of the feature space.

3.2 What constitutes a local conflict?

Consider two observations, or two local aggregates, that are close in the feature space X. If the conditional behaviour of the response variable $F \mid X$ differs substantially between them, this provides evidence of a local conflict.

It is important to emphasise that not every difference constitutes a conflict. Observed differences may be:

- **Noise-driven**, reflecting random fluctuations without structural significance;

- **Smooth**, corresponding to gradual changes that can be captured within a single global model;

- **Structural**, exhibiting sharp and persistent discrepancies indicative of incompatible local mechanisms.

The purpose of the diagnostic procedure is to distinguish structural conflicts from noise-driven and smooth effects.

3.3 A two-component diagnostic structure

The proposed approach separates the diagnostic task into two conceptually distinct yet interrelated components.

The **geometric component** determines which observations can be meaningfully compared:

- a locality graph is constructed over the feature space X;

- edges connect observations that are close in X;

- this defines a topological structure on which local comparisons are performed.

The **response component** determines what is being compared:

- for each edge in the graph, a measure of disagreement in the conditional behaviour of $F \mid X$ is computed;

- disagreements may be quantified through differences in residuals, local predictions, moments, or more general distributional characteristics;

- large disagreements between neighbouring observations indicate the presence of local conflicts.

Taken together, these components produce a diagnostic picture of where conflicts concentrate in the feature space, how persistent they are, and whether they are likely to be structural in nature.

3.4 Segmentation as a diagnostic hypothesis

Within this perspective, segmentation is understood primarily as a means of assessing a hypothesis about the structural organisation of the data.

Hypothesis H_0 (global model): the data are structurally homogeneous, and a single model provides an adequate description of the entire portfolio.

Hypothesis H_1 (segmented model): the data contain incompatible local regimes that call for structural separation.

The task of diagnostics is to evaluate which of these hypotheses is better supported by the data. Crucially, the absence of sufficient evidence in favour of segmentation should be regarded as a valid and informative diagnostic outcome, rather than as a limitation of the approach.

4 Application Context: Concentrated Markets with Limited Data

The proposed approach is applicable to a broad class of actuarial problems in which the structural adequacy of a global model is in question. In this paper, we focus on the context of concentrated markets with limited data—not because this is the only or primary domain of application, but because under such conditions the limitations of standard diagnostic tools tend to become most apparent, and the cost of incorrect structural decisions is particularly high.

We outline below the key characteristics of markets that call for increased attention to structural diagnostics.

4.1 Market characteristics requiring structural diagnostics

The proposed approach is especially relevant for markets exhibiting the following properties.

Participant concentration. The market is dominated by a small number of large players (typically two to five). Strategic decisions made by one participant directly affect the portfolio composition of others. Behavioural spillover effects may induce structural changes in the data without producing clear global signals of model degradation.

Limited historical data. The volume of available observations is often insufficient for reliable drift detection using standard methods. Low statistical power makes stability assessments conditional and heavily dependent on expert judgement. Under such circumstances, Bayesian or diagnostic approaches are frequently more informative than purely frequentist tests.

Regulatory and resource constraints. Models are subject to requirements of transparency and interpretability. Computational resources for frequent retraining of complex models are limited. In addition, structural decisions—including segmentation—must be justified and explained to regulators and internal stakeholders.

Dynamic market structure. Markets may undergo rapid changes in competitive conditions, product offerings, distribution channels, or regulatory frameworks. These processes can generate local structural effects that remain invisible to aggregated performance metrics.

4.2 Why standard diagnostics may be insufficient

Under the conditions described above, standard diagnostic tools often fail to detect structural problems in a timely and reliable manner.

Methods for detecting marginal drift typically have low statistical power and may miss relevant structural changes. Frequent model retraining is computationally costly and may obscure, rather than resolve, the underlying issue. Increasing model complexity—either by expanding the feature space or by moving to more flexible model classes—raises the risk of overfitting when data are limited. Finally, segmentation introduced in an ad hoc manner increases operational burden without guaranteeing improved model stability.

In such settings, there is a need for a diagnostic tool that:

- remains applicable under limited data;
- is used in a targeted, hypothesis-driven manner rather than as part of continuous monitoring;
- produces interpretable results that can be justified to regulators and business stakeholders;
- allows segmentation to be explicitly rejected when insufficiently supported by the data.

4.3 Illustrative example: strategic interactions in a concentrated market

Consider a simplified but realistic scenario. In an automobile insurance market, three major players—A, B, and C—operate. Company A makes a strategic decision to adjust its underwriting policy for a particular group of policyholders, for instance by tightening conditions for young drivers in urban areas.

As a result, some of A's clients migrate to competitors B and C. The portfolios of these companies receive an inflow of policyholders with a different risk profile. Conditional relationships between features (such as age and location) and loss experience within the portfolios of B and C change accordingly. Structural heterogeneity emerges: "legacy" policyholders exhibit different behaviour compared to those recently acquired from a competitor.

From a diagnostic perspective, marginal feature distributions may change only slightly: the age and geographic composition of the portfolios remain broadly comparable. Global model performance metrics may also stay within acceptable ranges. Locally, however, persistent conflicts arise: a model trained predominantly on legacy policyholders performs systematically worse for newly acquired ones.

Standard diagnostics may fail to flag this issue, as they are not designed to assess local structural coherence. The proposed diagnostic approach makes it possible to detect such local conflicts and to assess the plausibility of segmentation hypotheses—for example, by time of acquisition or by proxy variables reflecting portfolio origin.

5 Mathematical framework: topology-aware energy-based diagnostics

5.1 Problem formulation

Consider a set of observations (X_i, F_i), where $X_i \in \mathbb{R}^d$ denotes a vector of covariates (possibly of mixed type), F_i is the target variable (loss, frequency, severity, etc.), and $i = 1, \ldots, n$.

A global model assumes that the conditional distribution $F \mid X$ is structurally homogeneous over the entire domain. Segmentation introduces an alternative hypothesis: the data are better described by several local regimes, each characterised by its own conditional relationship.

The objective here is not to optimise segmentation for predictive performance, but to *diagnose* whether the data provide support for the hypothesis of structural heterogeneity.

5.2 Locality graph and two-component structure

To perform the diagnostic analysis, we construct a locality graph $G = (V, E)$ over a set of representative points (landmarks). Vertices correspond to representative observations or local aggregates, while edges E encode proximity in the covariate space X.

Geometric component. The graph specifies which observations are locally comparable. For mixed-type data, an appropriate distance metric (e.g. Gower distance) is used, and edges are constructed via k-nearest neighbours or ε-neighbourhoods.

Target component. For each edge $(i, j) \in E$, we define a disagreement weight w_{ij} reflecting differences in the conditional behaviour of $F \mid X$ between vertices i and j.

The key distinction from purely geometric clustering is that edges are weighted not only by proximity in X, but also by inconsistency in $F \mid X$.

5.3 Energy functional

Let a segmentation be defined by an assignment $z = (z_1, \ldots, z_n)$, where $z_i \in \{1, \ldots, K\}$ denotes the zone label for observation i. Each zone k is associated with parameters ϕ_k describing the local behaviour of $F \mid X$ within that zone (e.g. moments, local model parameters, or more general distributional characteristics).

We define the segmentation energy as

$$E(z, \phi) = \sum_{i=1}^{n} \ell_i(\phi_{z_i}) \;+\; \lambda \sum_{(i,j)\in E} w_{ij} \mathbb{1}[z_i \neq z_j] \;+\; \alpha K,$$

where:

- **The first term** (data term) measures the compatibility of observation i with zone z_i, for example via negative log-likelihood or squared deviation;

- **The second term** (boundary penalty) penalises cutting edges with high disagreement weights w_{ij}; stronger local disagreement implies a lower penalty for separating neighbouring vertices;

- **The third term** (complexity penalty) penalises the number of zones and discourages excessive fragmentation.

5.4 Dual interpretation

The proposed functional admits two complementary interpretations.

Frequentist interpretation. The energy corresponds to a regularised risk functional. Introducing segmentation reduces structural disagreement (through the first and second terms) at the cost of increased complexity (the third term). The balance is controlled by the parameter α.

Bayesian interpretation. The same energy can be understood as proportional to the negative log-posterior probability of a structural hypothesis:

$$E(z, \phi) \propto -\log p(z, \phi \mid \text{data}),$$

where the complexity term plays the role of a prior penalising complex structures (implicitly related to priors such as a Dirichlet process or a Chinese Restaurant Process with low concentration). This interpretation is particularly natural under limited data, where posterior mass concentrates on simpler hypotheses until sufficiently strong local evidence in favour of segmentation emerges.

Importantly, both interpretations lead to the same computational procedure, while providing different conceptual frameworks for understanding the results.

5.5 Disagreement weights

Diagnostic performance depends critically on the choice of disagreement weights w_{ij}. Ideally, w_{ij} should reflect rich distributional discrepancies between local conditional distributions $F \mid X$ in the neighbourhoods of i and j. However, computing such measures for all pairs may be computationally expensive.

We therefore adopt a two-phase strategy.

Phase A (screening). Simplified surrogate measures $\tilde{w}_{ij}$ are computed based on differences in moments, residuals, or simple predictive contrasts. These measures are inexpensive and are used for initial candidate identification. False positives are acceptable at this stage and are filtered out later.

Phase B (refinement). For a restricted set of candidates, disagreement weights w_{ij} are recomputed using richer measures such as energy distance, Hellinger distance between local mixtures, or posterior predictive divergence in Bayesian models.

5.6 Algorithmic procedure

The algorithm follows an agglomerative strategy:

1. Select a set of landmark points $L \subset \{1, \ldots, n\}$ with $|L| = m$;
2. Construct a locality graph G over L (e.g. k-NN);
3. Compute surrogate weights $\tilde{w}_{ij}$ for all edges;
4. Initialise with $K = m$ (each landmark forms its own zone);
5. Iteratively merge zones by minimising the surrogate energy $\tilde{E}(z)$;
6. Record the full energy trajectory $E(K)$ for $K = m, m-1, \ldots, 1$;
7. Identify candidates based on the trajectory (elbows, plateaus);
8. Recompute refined energies using w_{ij} for the candidates;
9. Select the final segmentation or return the global model ($K = 1$).

Crucially, if none of the candidates exhibits a stable and interpretable improvement over $K = 1$, the algorithm *explicitly returns the non-segmented solution*. This outcome represents a valid diagnostic conclusion rather than a failure of the method.

5.7 Role of the complexity parameter α

The parameter α plays a role analogous to complexity penalties in classical model selection criteria (AIC, BIC). No universal optimal value exists.

Instead, we recommend:

- exploring a small range of α values;
- analysing energy trajectories and solution stability;
- using α as a diagnostic sensitivity parameter.

Small values of α increase sensitivity to local conflicts (potentially detecting weak signals at the cost of noise), while larger values favour conservative solutions and tend to preserve the global model.

5.8 Relation to existing approaches

Compared to drift detection. Standard drift detection methods (KS tests, PSI, residual monitoring) operate on marginal or low-dimensional projections and do not assess structural coherence of local conditional relationships.

Compared to geometric clustering. Methods such as spectral clustering partition data based on the geometry of X, ignoring the behaviour of $F \mid X$. The proposed approach integrates both components: geometry determines where comparisons are made, while the target variable determines what is compared.

Compared to mixture models. Mixture and regime-switching models assume latent structure and optimise parameters for prediction. Here, no structure is assumed a priori; instead, the presence or absence of structure is diagnosed.

Compared to manifold intersection detection. Approaches based on geometric signals (e.g. local tangent spaces or curvature, such as Deutsch & Medioni, 2015) are effective under clean data and strong geometric structure. The proposed method is designed for low signal-to-noise settings with weak geometric cues, combining topology with target-based disagreement.

6 Synthetic scenario: a diagnostic illustration of strategic interactions

This section is illustrative in nature. Its purpose is not to demonstrate empirical results, but to clarify the types of structural situations the proposed diagnostic approach is designed to detect, and how its conclusions should be interpreted in actuarial practice under limited observability.

6.1 Scenario setup

We consider a stylised scenario reflecting a type of strategic interaction commonly encountered in concentrated insurance markets. Suppose that three large players (A, B, and C) operate in an automobile insurance market, each with a relatively stable portfolio of policyholders.

For all companies, loss experience depends on standard actuarial covariates such as driver age, city type (large versus small), and driving experience. Within each portfolio, a company-specific conditional relationship between covariates and loss outcomes is present, reflecting differences in underwriting policies and client composition.

At time $t = 0$, company A makes a strategic decision to tighten conditions for young drivers in large cities, for example through pricing or underwriting restrictions.

Conceptual consequences of this decision include:

- a portion of company A's clients from the affected region of the feature space migrates to competitors B and C;
- the portfolios of B and C receive an inflow of clients with a different conditional risk profile;
- within the portfolios of B and C, potential structural heterogeneity emerges, as policyholders governed by different conditional relationships $F \mid X$ coexist.

6.2 Remark on the actuary's perspective

It is important to emphasise that, in practical settings, an actuary typically observes **only their own portfolio**, rather than the market as a whole.

Companies A, B, and C are introduced in this scenario not as objects of joint analysis, but as a **conceptual device** for describing a mechanism through which structural heterogeneity may arise.

From the perspective of an actuary at company B or C:

- company A is not directly observable;
- client inflows are perceived as endogenous changes in the portfolio;
- information about client origin is either unavailable or only indirectly accessible.

Accordingly, the diagnostic question can be formulated as follows:

> Can signs of structural inadequacy of a global model be detected using only internal portfolio data?

The proposed diagnostic approach is designed precisely with this question in mind.

6.3 Diagnostic formulation

Suppose that the actuary at company B or C observes the following:

- marginal distributions of covariates change only slightly;
- global model quality metrics remain within acceptable ranges;
- locally, however, the model exhibits instability or systematic errors within a restricted region of the feature space.

In such a situation, standard diagnostic tools may fail to provide a clear signal. This creates a need for a method that analyses not marginal stability, but **local conditional coherence**.

6.4 Expected diagnostic behaviour

Within the described scenario, the following diagnostic expectations can be formulated.

Control case (structurally homogeneous portfolio). If the portfolio is structurally homogeneous, the diagnostic procedure should support the global hypothesis:

- the energy trajectory decreases smoothly;
- no stable localisation of graph cuts is observed;
- the preferred solution corresponds to $K = 1$.

Scenario with local structural heterogeneity. If the portfolio contains a limited region of the feature space in which the conditional relationship $F \mid X$ differs from the rest of the portfolio, the diagnostic procedure should:

- reveal a stable improvement for $K > 1$;
- localise graph cuts in an interpretable region of X;
- demonstrate that segmentation explains local conflicts rather than noise.

A key point is that the diagnostic signal emerges **locally**, without requiring a global deterioration of model-level metrics.

6.5 Interpretation of diagnostic conclusions

The proposed scenario illustrates the distinction between two notions of stability:

- **marginal stability** (covariate distributions, aggregated metrics);
- **structural stability** (coherence of local conditional relationships).

In actuarial practice, this distinction is essential. The method allows diagnostic conclusions to be formulated in terms of structural hypotheses:

- either the global model remains adequate and segmentation is not warranted;
- or local regimes exist for which the global model is internally incompatible.

Both outcomes represent valid diagnostic results.

Thus, the synthetic scenario should be understood not as empirical evidence, but as a **conceptual map** illustrating the types of structural problems the proposed diagnostic approach is capable of revealing, and how its conclusions should be interpreted under conditions of limited data and partial market observability.

7 Diagnostic principles and interpretation of results

The proposed method is not an automatic procedure that mechanically produces a segmentation decision. It is intended as a **diagnostic tool** that supports, rather than replaces, actuarial professional judgment.

This section formulates principles of interpretation that help distinguish justified structural segmentation from artefacts driven by noise, scale effects, or data limitations.

7.1 Multi-level diagnostics: avoiding reliance on a single criterion

Decisions regarding segmentation should be supported by **multiple consistent signals**. No single diagnostic indicator is sufficient on its own.

Key diagnostic signals include:

1. **Energy trajectory** — whether increasing the number of zones yields a structurally meaningful improvement or merely a monotonic decrease without characteristic transitions.

2. **Spatial localisation** — whether zone boundaries concentrate in interpretable regions of the feature space.

3. **Boundary load** — whether identified boundaries carry a substantial share of local conflicts, rather than resulting from random fragmentation.

4. **Improvement of local models** — whether segmentation leads to stable improvements in local data description.

5. **Between- and within-zone variation** — whether differences between zones are substantively meaningful in terms of $F \mid X$.

Segmentation is considered diagnostically justified only when the majority of these signals point in the same direction. When signals are weak or conflicting, preference should be given to the global model.

7.2 Comparison with geometric clustering

A useful reference point is comparison with purely geometric clustering based solely on the structure of the feature space X.

Typical interpretative scenarios include:

- **Geometry = 1, energy-based diagnosis > 1:** The signal originates from the behaviour of the target variable rather than from feature geometry. This indicates structural conflicts without clear separability in X.

- **Geometry > 1, energy-based diagnosis consistent with it:** Segmentation reflects a natural geometric separation and can be interpreted as structurally meaningful.

- **Geometry > 1, energy-based diagnosis = 1:** Geometric separation is not supported by the target variable and should not automatically lead to segmentation.

Such comparisons help identify the source of the diagnostic signal and avoid unjustified structural decisions.

7.3 Temporal stability (when time slices are available)

When data are available across multiple time slices, the diagnostic procedure can be applied independently to each slice.

Zones that are consistently reproduced over time and exhibit substantial overlap (for example, as measured by the adjusted Rand index) can be interpreted as **structural**, rather than accidental or transient.

Temporal stability is not a required condition, but serves as an important supplementary confirmation when relevant data are available.

7.4 Bayesian perspective and the role of uncertainty

Under the Bayesian interpretation, the energy function corresponds to the negative logarithm of the posterior probability of a structural hypothesis. This allows diagnostic outcomes to be viewed as distributions of uncertainty rather than point decisions.

A **flat energy trajectory** indicates that the data do not provide sufficient evidence in favour of segmentation. In such cases, preference should be given to the simpler hypothesis ($K = 1$).

A **sharp transition in energy** suggests concentration of posterior mass and points to the presence of localised structural heterogeneity.

This perspective is particularly valuable under limited data, where explicit treatment of uncertainty reduces the risk of overfitting.

7.5 When diagnostics return the global model

The absence of segmentation constitutes a **valid and informative diagnostic result**.

The method returns $K = 1$ when:

- the data are structurally homogeneous;
- local conflicts are insufficiently stable or interpretable;
- noise dominates the available signal;
- increased structural complexity is not justified by diagnostic benefit.

In all such cases, retaining the global model represents a correct and professionally defensible decision.

7.6 Practical recommendations for use

Recommended use cases include:

- suspected structural instability of a model;
- significant changes in the business or regulatory environment;
- prior to decisions on portfolio segmentation.

Use is not recommended:

- as a routine monitoring tool;
- in cases of clear marginal drift without structural analysis;
- under extremely limited data, where diagnostic power is known to be low.

Interpretation and communication:

- treat results as diagnostic evidence rather than automatic decisions;
- visualise localisation of conflicts and zones;
- relate diagnostic conclusions to business context;
- use the method as a transparent and reproducible way to justify structural decisions to internal and external stakeholders.

8 Limitations and discussion

8.1 Methodological limitations

The proposed diagnostic approach has several limitations that should be taken into account in practical applications.

Data requirements.

- The method requires a sufficient number of observations to construct a locality graph and to assess local conditional disagreements. For $n < 500$–1000, diagnostic power may be limited.
- Rare regimes with very small numbers of observations may remain undetected.

- The approach assumes that structural conflicts are localised in the feature space. If incompatibility is distributed uniformly, detection becomes more difficult.

Choice of metrics and parameters.

- The quality of the locality graph is sensitive to the choice of metric in the feature space X, particularly for mixed-type data.
- The complexity parameter α does not admit a universal optimal value and requires diagnostic calibration.
- The choice of the number of landmarks and neighbourhood size affects the trade-off between sensitivity and stability.

Computational aspects.

- The method is intended for offline diagnostics rather than continuous monitoring.
- Computation of rich disagreement measures may be resource-intensive when many candidates are considered.
- The two-phase strategy (screening followed by refinement) reduces computational burden but does not eliminate it entirely.

Interpretability.

- Diagnostic zones are not required to coincide with natural business segments.
- Zones may have complex shapes in the feature space, requiring additional analytical effort for explanation.
- Transition from diagnostics to operational rules necessarily involves actuarial judgment.

8.2 Relation to modelling and operational practice

The proposed diagnostics are intended to complement, rather than replace, the modelling process.

Before diagnostics: the standard model development cycle applies, including data preparation, feature selection, model class choice, and validation.

Diagnostics: applied when structural instability is suspected, addressing the question of whether observed instability reflects incompatibility between local regimes.

After diagnostics:

- when $K = 1$, sources of instability should be sought in data quality, feature design, modelling assumptions, or business processes;
- when $K > 1$, there is a justified basis to consider structural interventions such as segmentation, feature interactions, or underwriting adjustments.

It is important to emphasise that diagnostically identified zones do not constitute ready-to-use operational solutions. They are defined through the topology of the locality graph and do not have a direct representation in terms of original features.

Practical implementation requires a managerial compromise: approximating complex structural patterns with simple and implementable rules. Examples include:

- introducing segmentation based on simple logical criteria;
- adding feature interactions within a global model;
- adjusting underwriting policy for a narrowly defined subgroup.

The choice among these options depends on interpretability, operational feasibility, stability, and regulatory constraints.

Crucially, simplification occurs **at the operationalisation stage**, not at the diagnostic stage. Decisions continue to be grounded in the analysis of local conditional conflicts rather than marginal feature comparisons.

8.3 Context of limited-data markets

In markets characterised by limited data and concentrated structure, the proposed approach becomes particularly relevant.

- **High cost of erroneous decisions:** unjustified segmentation may lead to overfitting, operational complexity, and regulatory risk.

- **Strategic interactions:** actions by one participant may induce structural shifts in other portfolios without clear signals in marginal distributions.
- **Regulatory requirements:** the need for transparent justification of structural decisions increases the value of diagnostic evidence.
- **Bayesian perspective:** explicit treatment of uncertainty and preference for simplicity improve robustness under limited data.

8.4 Directions for further research

The present work suggests several directions for future development:

- more rigorous Bayesian formalisation of posterior distributions over segmentations;
- stability analysis of zones using bootstrap and cross-validation;
- integration of tools from topological data analysis;
- empirical application to real actuarial portfolios;
- development of infrastructure for integrating diagnostics into model governance processes.

8.5 Discussion summary

This section has discussed the limitations of the proposed approach and its relationship to actuarial modelling practice. The central conclusion is that the diagnostics are intended to identify and localise structural conflicts, rather than to automate segmentation decisions.

The method narrows the space of plausible structural choices and makes competing hypotheses explicit, while leaving the final decision to actuarial judgment informed by business and regulatory context.

9 Conclusion: Implications for actuarial practice

Under conditions of limited data and high requirements for decision stability, actuarial practice requires not only more sophisticated models, but also more rigorous diagnostic criteria for managing model complexity.

The proposed approach shifts the focus from optimising segmentation to justifying its necessity. By treating segmentation as a diagnostic hypothesis rather than as a modelling objective, the actuary gains a framework for making structural decisions in a disciplined and accountable manner.

It is essential that the absence of segmentation is recognised as a valid and professionally justified outcome. This perspective improves transparency, reduces model risk, and facilitates communication of modelling decisions with stakeholders and regulators.

The method is designed for practical use and can be integrated into existing actuarial governance processes without reliance on large-scale computational experiments. This makes it particularly relevant for markets with limited data and resources, and supports the development of stable and inclusive insurance systems.

References

[1] Board of Governors of the Federal Reserve System (2011). *Supervisory Guidance on Model Risk Management (SR 11-7)*. Federal Reserve Bulletin.

[2] European Insurance and Occupational Pensions Authority (2015). *Guidelines on the Use of Internal Models*. EIOPA.

[3] Sculley, D., Holt, G., Golovin, D., et al. (2015). Hidden technical debt in machine learning systems. In *Advances in Neural Information Processing Systems*.

[4] Molnar, C. (2022). *Interpretable Machine Learning* (3rd ed.). Available at `https://christophm.github.io/interpretable-ml-book/`

[5] Gama, J., Žliobaitė, I., Bifet, A., Pechenizkiy, M., & Bouchachia, A. (2014). A survey on concept drift adaptation. *ACM Computing Surveys*, 46(4), 1–37.

[6] Coifman, R. R., & Lafon, S. (2006). Diffusion maps. *Applied and Computational Harmonic Analysis*, 21(1), 5–30.

[7] Deutsch, S., & Medioni, G. (2015). Intersecting manifolds: Detection, segmentation, and labeling. In *Proceedings of the International Joint Conference on Artificial Intelligence (IJCAI)*.

[8] Mumford, D., & Shah, J. (1989). Optimal approximations by piecewise smooth functions and associated variational problems. *Communications on Pure and Applied Mathematics*, 42(5), 577–685.

[9] Hamilton, J. D. (1989). A new approach to the economic analysis of nonstationary time series and the business cycle. *Econometrica*, 57(2), 357–384.

[10] von Luxburg, U. (2007). A tutorial on spectral clustering. *Statistics and Computing*, 17(4), 395–416.

[11] Verte, X. (2026). *When to Segment: Topology-Aware Energy-Based Diagnostics for Structural Model Uncertainty*. SSRN Working Paper. Available at `https://doi.org/10.2139/ssrn.6081486`