§ m.01 · methodology note

On the deterministic decomposition of healthcare payment.

A note on the epistemic structure of reimbursement, and on the category error that has shaped the industry's analytical infrastructure for three decades.

H. Chaudhry · Crescent Intelligence
crescentintel.com/research/deterministic-decomposition

13 February 2026
2847 words

abstract

Existing approaches to healthcare reimbursement analysis attempt to model realized payment as a single quantity, applying probabilistic methods inherited from actuarial science to a quantity that is not, in its dominant component, probabilistic. The accuracy ceiling observed across the literature — between seventy and eighty percent against held-out remittance data — is not an artifact of model architecture. It is a structural consequence of modeling a convolution of two epistemically distinct quantities as if they were one. Realized payment decomposes exactly into a contractual component, which is a proof, and a behavioral residual, which is a forecast. Conflating the two is a type error. We argue the error is inherited from an adjacent discipline, that it has been undetected because the deterministic component is stable enough to make the convolutional approach appear functional, and that its downstream consequences include the systematic mispricing of healthcare receivables across diligence, contracting, and capital markets.

1 introduction

Every healthcare claim resolves, eventually, to a realized payment R: an observed cash amount remitted from a payer to a provider for a specified service. The standard analytical approach across the revenue cycle management industry, the major claim-audit firms, and the quality-of-earnings practices used in healthcare M&A is to model R directly. Features are engineered from the claim and its context. Supervised learning is applied against historical remittance data. The output is a point estimate, sometimes accompanied by a confidence interval, of what R will be or should have been.

The accuracy of these methods, measured against held-out remittances, plateaus between seventy and eighty percent depending on payer mix and specialty. The ceiling has been remarked upon in the literature for at least fifteen years. It is generally attributed to insufficient feature engineering, inadequate training data, or the irreducibility of payer-side discretion.[1]

We argue that none of these explanations is correct. The ceiling is structural to the approach, and the approach contains an error that precedes any modeling decision.

The error is this. The realized payment R is the sum of two components that are epistemically distinct. The first component, C, is the contractually correct payment — the unique value that the controlling rule system (statute, fee schedule, coding rules, payer contract) assigns to the claim. The second component, B, is the behavioral residual — the difference between what the payer should pay under the rules and what the payer actually pays. We have $R = C + B$. The two components are not points on a confidence spectrum. They are categorically different objects, produced by different mechanisms, knowable through different epistemic operations, and falsifiable under different conditions. Any methodology that models R directly is modeling the convolution of a proof and a forecast as if it were a forecast. This is a type error.

[1] marginalia

The literature treats the ceiling as a frontier to be pushed against. We argue it is a wall produced by the framing of the problem.

This note develops the implication of that error and names what follows from its correction.

2 the category error

There are two kinds of knowledge available to an analyst of healthcare payment, and the existing literature has conflated them. We name them precisely.

The first is *proof*. A proof is a claim whose truth is established by derivation from primitives, by rule application, in a finite number of steps. A proof is dollar-exact. It is falsifiable in a single check: either the derivation is correct or it is not. Knowledge of the form "given this CPT code, this place of service, this modifier set, this payer contract, this fee schedule, and this set of NCCI edits, the contractually owed amount is $487.23" is a proof. It does not become more or less true with more data. It is either right or wrong, and rightness is verifiable by re-running the derivation.

The second is *forecast*. A forecast is a claim whose truth is established by inference from observation, under uncertainty, with calibration measured across a distribution rather than at a single point. A forecast is probabilistic. It is falsifiable only across many cases, by comparing predicted distributions to realized distributions. Knowledge of the form "this payer, in this region, for claims in this specialty, downcodes from level four to level three approximately eighteen percent of the time in the second quarter of any given year" is a forecast. It can be wrong about any individual case while remaining correctly calibrated, and it can be correct about every individual case while being structurally miscalibrated. The unit of falsification is the distribution, not the case.

These are not two ends of a spectrum. They are different objects. A proof and a forecast cannot be averaged. They cannot be combined into a single accuracy metric. They cannot be produced by the same methodology, because the methodology that produces a proof (rule application over a closed primitive set) is mechanically distinct from the methodology that produces a forecast (statistical inference over an observed sample). To treat them as the same kind of object is the analytical equivalent of treating temperature and time as the same kind of quantity because both can be measured by numbers.

The cost of the conflation is visible in the variance structure of R. Decomposing the variance of the realized payment:

variance decomposition

$$\mathrm{Var}(R) = \mathrm{Var}(C) + \mathrm{Var}(B) + 2 \cdot \mathrm{Cov}(C, B)$$

The convolutional model is trained to minimize loss against the full $\mathrm{Var}(R)$. Of the three terms on the right-hand side, only $\mathrm{Var}(B)$ is the signal of interest — the part of the variance that reflects actual payer behavior and that the analyst cannot derive from the rule system. $\mathrm{Var}(C)$ is large but structural: it is the variance produced by

the combinatorial space of contract rules, fee schedules, and code modifier interactions, all of which are knowable by computation. The covariance term reflects the systematic relationship between contract terms and payer enforcement patterns and is partially attributable to each component. By training against $\mathrm{Var}(R)$, the convolutional model spends the majority of its learning capacity approximating $\mathrm{Var}(C)$ — a quantity that is, by construction, already known.

The existing approach to healthcare reimbursement analysis treats R as a single quantity to be forecasted. It applies probabilistic methods uniformly. The accuracy ceiling that results is the necessary consequence: any methodology that produces a forecast as its output will, when applied to a quantity that contains a deterministic component, underperform the trivial methodology of computing the deterministic component directly.[2]

The probabilistic model attempts to learn what the rule system already says, and it cannot exceed the rule system on the rule-governed portion. It can only approximate it. The portion of R that is C is being approximated when it should be computed. The portion of R that is B is being approximated badly, because the model is also trying to absorb the variance of C and consequently has less capacity for B.

The convolutional methodology produces two errors at once: it forfeits exactness on the deterministic component, and it degrades the residual signal on the probabilistic component. The decomposition is therefore not a methodological improvement. It is the mathematical consequence of recognizing that R contains two epistemically distinct objects that demand distinct treatments.

$R = C + B$ is the structural form. The decomposition is forced.

[2] marginalia

This is true by construction. A methodology that forecasts a quantity which has a deterministic component is dominated, on that component, by direct computation. The dominance is mathematical, not empirical.

3 the contractual component

The deterministic component C is exactly computable. This claim is stronger than it sounds and we state it precisely.

For any claim, the controlling rule system consists of: the Code of Federal Regulations (Title 42 in particular, for Medicare-governed services), the Medicare Physician Fee Schedule and the Medicare Hospital Outpatient Prospective Payment System schedules, the National Correct Coding Initiative edits (procedure-to-procedure and medically-unlikely-edit components), the relevant Local Coverage Determinations and National Coverage Determinations, the applicable payer-specific medical policy bulletins, and the payer-provider contract including its fee schedule attachments and material modifiers. This rule system is finite, closed over its primitive set, and fully specified at any point in time. It is the joint product of statute, regulation, and bilateral contract. It contains no ambiguity in its terminal form — every ambiguity in the rule system is resolved by a definite rule, either explicitly written or implicitly invoked through precedence ordering.

For any claim, the contractually correct payment C is the unique value that this rule system assigns. Existence is not in question. Uniqueness is not in question. The rule system is a function from claim-state to payment-amount, and functions have unique outputs.

The challenge, historically, has not been the existence of C. It has been the encoding of the rule system at sufficient resolution to compute C at scale and in practice. The various commercial encodings — claim editors, rules engines, payment integrity tools — have been partial. They handle the most common edits and policies and approximate the rest. Their partiality is the reason the industry has come to treat C as if it were approximate. It is not approximate. The encodings have been approximate. The distinction matters because it locates the imprecision in human implementation, not in the mathematics of the underlying problem.[3]

The encoding problem is real but it is a problem of resolution, not of category. With sufficient encoding, C becomes computable to dollar-exactness for any claim, against any payer, under any contract, at any point in time. This is the work of constructing the apparatus, and it is the work that has not been done at the necessary resolution. We do not develop the encoding here. We observe that the existence of an exact C is a mathematical fact, not a methodological aspiration.

[3] marginalia

The existing partial encodings are sometimes described as approximations of C. They are better described as approximations of the encoding task. C itself is exact.

4 the behavioral residual

Once C is computed, the residual $B = R - C$ is the behavioral component. It captures what the payer actually does relative to what the contract requires.

B is probabilistic, but it is not random. Payer behavior is patterned. It varies systematically by payer, by specialty, by geography, by claim type, by season, by year, and by the prevailing regulatory and economic environment. A given payer, in a given specialty, in a given region, exhibits behavioral patterns that are stable enough to model and unstable enough to require continuous re-estimation. Downcoding rates, denial rates, payment-delay distributions, and selective enforcement of contract terms are all features of B that admit probabilistic characterization.

The important property of B, once C is removed, is that it is a cleaner signal than R. This can be made precise through capacity allocation. A statistical learning model has bounded representational capacity, which we denote K. Training the model against R forces it to allocate some portion K_C of its capacity to approximating the deterministic structure of C — the combinatorial rule space whose values it must learn to reproduce in order to minimize loss on the dominant component of the target. The capacity available for B is therefore bounded:

$$K_B \leq K - K_C$$

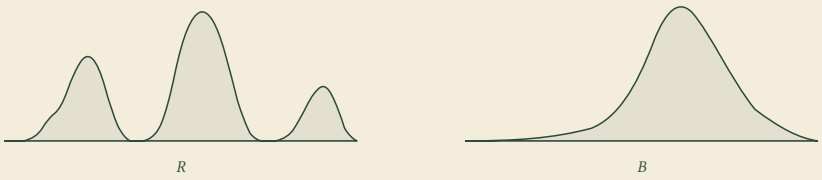
In the decomposed architecture, C is computed exactly outside the model. The model is trained directly on the residual, and the capacity required to represent C drops to zero. The full capacity K is available for B:

$$K_B = K$$

The decomposed model dominates the convolutional model on the forecast side by a margin equal to the capacity reclaimed from C. This is the second source of the dominance argument: decomposed methodology beats convolutional methodology not only on the proof side, where the dominance is by construction, but also on the forecast side, where the dominance is by capacity reallocation. The argument holds across model architectures — neural networks, gradient-boosted trees, regression families — because the capacity constraint is general to bounded statistical learners and the reallocation argument is independent of how the capacity is parameterized.

The cleaner signal is also visible in the geometry of the conditional distributions. The distribution of R given a feature set is structurally multimodal: the discrete pricing tiers of C produce sharp modes separated by gaps, and a model attempting to characterize this distribution must learn to allocate probability mass across discontinuous regions. The distribution of B given a feature set, once C is removed, is unimodal or low-modal, smoother, and centered on the systematic payer-behavior offset from contract. Standard probabilistic modeling techniques are optimized for the latter and structurally degraded on the former.

figure 1



P($R \mid X$) — realized payment, conditional on claim features
multimodal, jagged, skewed by discrete contract pricing tiers.

P($B \mid X$) — behavioral residual, conditional on claim features (C removed)
smooth, well-behaved, amenable to standard probabilistic modeling.

conditional probability densities of the realized payment R (left) and the behavioral residual B after C is removed (right). The decomposition transforms a structurally intractable distribution into a tractable one. Standard probabilistic modeling — gradient descent, maximum likelihood, Bayesian inference — converges faster, scales better, and achieves tighter confidence intervals on the smooth distribution of B than on the fractured distribution of R.

B is itself further decomposable. A portion of B is sub-deterministic — attributable to specific, identifiable payer policies that are not contained in the formal rule system but that are nonetheless consistent enough across cases to be characterized as near-rules. Another portion is genuinely stochastic, reflecting individual adjudicator discretion, timing artifacts, and irreducible idiosyncrasy. The further decomposition of B is the subject of a forthcoming note (§ M.04).

For the purposes of this note, it is sufficient to observe that B, once isolated, is amenable to standard probabilistic modeling and yields better calibrated forecasts than any approach that models R directly.

5 two outputs

The decomposition yields two outputs. The outputs serve different stakeholders, answer different questions, and are produced by different methodologies. The existing literature collapses both into a single number and produces neither well.

The first output is *contractual truth*. This is C, the dollar-exact amount owed under the rule system. It is a proof. Its addressees are the patient (who has the right to know what was owed), the provider (who has the right to bill what is owed), the payer (who has the obligation to remit what is owed), and the auditor (who has the standing to verify what was owed). The unit of analysis is the individual claim. The verification method is re-derivation. The output has zero noise.

The second output is *cash truth*. This is the calibrated expectation of R given C and the modeled payer behavior B. It is a forecast. Its addressees are the lender (who must price advances against expected cash), the investor (who must price securities against expected cash), the acquirer (who must value receivables against expected cash), and the negotiator (who must set contract terms against expected cash). The unit of analysis is the portfolio or the distribution. The verification method is calibration over a large sample. The output has irreducible noise, but the noise is bounded and characterized.

These outputs are not competing. They are complementary. Any sophisticated user of healthcare receivables data requires both: the proof, to establish entitlement and ground negotiation; the forecast, to discount entitlement to cash. The two outputs cannot be substitutes for each other. A proof cannot be a forecast (it does not characterize uncertainty); a forecast cannot be a proof (it does not establish entitlement). The industry's collapse of both into a single approximate quantity has obscured the distinction and consequently the value of either.

6 the inherited error

The category error has an origin and a consequence, and both deserve naming.

The origin is historical. The analytical infrastructure of modern healthcare RCM was built in the 1990s and 2000s by practitioners who had been trained in adjacent disciplines — primarily actuarial science and insurance underwriting. The methodological

inheritance was natural. Actuarial science is the discipline of modeling claims under uncertainty, and insurance claims are, in their underlying physics, genuinely probabilistic: an accident either occurs or it does not, an illness either manifests or it does not, a person either survives or does not. The actuarial methodology — large datasets, probabilistic forecasting, claims modeling as distributional inference — is correct for its native domain. The error was the import of the methodology into reimbursement analysis without recognizing that the underlying object had changed. A healthcare claim is not an insurance loss. It is a contractual computation. The same word "claim" was used for both, and the methodological inheritance followed the word rather than the structure.

The consequence is that the entire downstream apparatus of healthcare receivables analysis has been built on a quantitative foundation that contains an unrecognized type error. Every quality-of-earnings adjustment for revenue cycle performance, conducted under sampling methodology and convolutional modeling, contains the type error. Every healthcare-receivable-backed security priced against historical realized-payment distributions contains the type error. Every payer-provider contract negotiation that uses historical realized payment as the baseline for prospective contracting contains the type error. The error has not been catastrophic, because C is the dominant component of R for most claims, and a methodology that approximates C reasonably will recover most of the truth. But the residual error — the gap between approximate C and exact C, compounded with the degraded estimate of B — has been priced into every transaction in the industry as if it were irreducible uncertainty, when it is in fact industry-wide methodological failure.[4]

The cost is not visible at the level of any individual transaction. It is visible at the level of the asset class. Healthcare receivables, in aggregate, are mispriced. The direction of the mispricing depends on which side of the negotiation the party sits. Providers concede contractual value as behavioral uncertainty. Payers absorb contractual obligation as discretionary cost. Diligence firms produce findings that are directionally correct but not exact. Capital markets price the resulting paper with risk premia that reflect methodological noise as much as underlying uncertainty. The system functions, in the sense that transactions clear, but it functions on a foundation that contains a category error that no participant has had reason to correct, because correction would require rebuilding the analytical infrastructure from the primitive set rather than from the inherited methodology.

[4] marginalia

The error survives because it does not produce catastrophic failure. It produces consistent, distributed, low-grade mispricing that all parties have learned to accept as the cost of doing business.

7 closing

The decomposition is not a tool. It is a recognition.

The contractual quantity has always existed. The rule system has always been closed. The mathematics of the deterministic component has always been exact. The industry has not been failing to compute it for lack of capability. It has been declining to compute it because the inherited methodology, imported from a parent discipline where it was correct, did not call for the computation. The methodological convolution rendered the categorical distinction invisible, and the invisible category has remained uncomputed for thirty years.

The work of correction is the work of separating the proof from the forecast: building the apparatus to compute C exactly, and modeling B against the cleaner signal that emerges once C is removed. This is exactly what we are investigating at Crescent Research, and why we believe a commercial product with these characteristics must exist.

---

see also

§ M.02 On the obsolescence of sampling methodology in healthcare diligence.
§ M.03 On the architecture of decomposable targets.