

The Half-Life of a Futures Contract: Technological Obsolescence and Vintage Risk in Compute Derivatives

Fuli Yang

GARCH Institute, Harvard Square, Cambridge, MA 02138, USA

ORCID: 0009-0005-4055-7561 | dr@garch.ac

Working Paper v3.1 | 27 August 2026

Abstract

A futures contract has a legal maturity, but it can lose economic relevance before that date. This paper develops the concept of contract half-life for derivatives written on rapidly evolving technological vintages. Contract half-life is the time required for a hardware-specific contract's economic relevance - measured through commercial exposure, trading liquidity, and price-discovery importance - to fall to one-half of its reference level. The distinction matters for compute derivatives because the announced H100 and B200 futures are tied to named GPU generations whose commercial importance can migrate faster than their rental prices disappear. A simple vintage-migration model separates rental-price depreciation from benchmark-relevance depreciation and shows how innovation intensity, maturity, and liquidity network effects jointly determine economic expiry. The model yields four testable propositions: activity should migrate to a new vintage faster than the old rental index declines; distant maturities of an old-vintage contract should lose viability first; faster innovation should shorten contract half-life; and rolling or performance-adjusted benchmarks should extend durability while increasing conversion-model and governance risk. The paper proposes observable half-life measures based on open interest, volume, information shares, and commercial benchmark adoption, and explains why a still-published settlement index is insufficient evidence that a contract remains economically alive. Because these propositions require post-launch data, the paper also uses three completed benchmark transitions - the shift of a gasoline contract to an RBOB basis, the ultra-low-sulfur respecification of heating oil, and the inclusion of WTI Midland in Dated Brent - plus unsuccessful DRAM-futures proposals as identification templates. These analogues distinguish benchmark replacement, internal respecification, constituent revision, and launch failure; they are not estimates or formal bounds on compute-contract half-life.

Keywords: compute futures; contract half-life; technological obsolescence; vintage risk; benchmark design; liquidity migration

JEL Classification: G13; G14; L86; O33

1. Introduction

The calendar printed in a futures contract is not always the clock that matters economically. A contract can remain legally valid, retain a calculable settlement index, and continue to appear on an exchange's product list after the commercial risk it was designed to transfer has moved elsewhere. This possibility is familiar in a loose sense: exchanges regularly delist unsuccessful products, delivery grades are revised, and trading migrates between contracts. It becomes a first-order design problem when the underlier is a named technological vintage whose relevance can change during the maturity structure of the derivative itself.

Compute futures make that problem unusually visible. CME Group has announced cash-settled H100 and B200 rental-index futures for planned listing on 5 October 2026, subject to regulatory review. Both contracts are defined around specific NVIDIA GPU generations and settle to daily rental benchmarks constructed by Silicon Data (CME Group, 2026a, 2026b). Separately, Intercontinental Exchange (ICE) and Ornn announced on 19 May 2026 a suite of U.S.-dollar-denominated, cash-settled GPU compute futures based on the transaction-based Ornn Compute Price Index, subject to regulatory approval (ICE, 2026). The Commodity Futures Trading Commission (CFTC) has meanwhile requested public comment on the listing of compute derivatives, including the standardization of compute, idiosyncratic contract risks, effects on the cash market, and the possible role of perpetual products (CFTC, 2026). These product plans are public institutional events, but their trading outcomes remain prospective. Yet the economic unit being standardized can age much faster than grain, crude oil, or a broad equity index.

A new GPU generation does not make an older one physically vanish. H100 capacity can remain productive, rentable, and observable after B200 or later systems attract frontier workloads. Older generations may continue serving inference, fine-tuning, and cost-sensitive applications. What changes is the share of commercial risk for which the older vintage is the natural reference. That share can migrate before the old rental price reaches a low level and before an index provider loses the ability to publish a daily observation. A contract may therefore have a price but cease to be the market's focal hedge.

This paper names and formalizes that separation. Contract half-life is defined as the time required for the economic relevance of a hardware-specific derivative to fall to one-half of its reference level. Economic relevance is not the price of the underlying chip or rental service. It is a composite of three roles: the amount of commercial exposure naturally mapped to the contract, the share of trading liquidity concentrated in it, and its contribution to price discovery for the broader market. Legal expiry closes the contract. Economic expiry occurs when these roles have migrated sufficiently that the contract no longer coordinates the main risk-bearing activity it was intended to serve.

The central argument extends the benchmark-before-market framework in Yang (2026), SSRN Working Paper No. 7324139. That earlier paper examines how a settlement benchmark can help constitute a compute market before the underlying service is fully fungible and identifies a trilemma among tradability, exposure representativeness, and technological durability. The present paper takes one mechanism from that framework - technological durability - and develops it as an independent theory of contract life. It does not repeat the broader review, does not claim that announced contracts will attract realized liquidity, and does not use synthetic forward curves or simulated hedges as evidence of actual market success.

The contribution is fivefold. First, the paper distinguishes rental-price depreciation from benchmark-relevance depreciation. Second, it gives contract half-life a measurable definition rather than using obsolescence as a qualitative warning. Third, it introduces a simple two-vintage model in which innovation intensity, maturity, and endogenous liquidity migration determine economic life. Fourth, it uses three completed benchmark transitions and a never-alive control as identification templates that separate relevance decay, administrative redefinition, and launch failure without treating historical episodes as estimates of compute-contract half-life. Fifth, it compares fixed-vintage, rolling-vintage, performance-adjusted, and token-expenditure designs, showing that durability can be purchased only by accepting transition rules, workload assumptions, and benchmark-governance discretion.

This is a prospective theoretical and institutional analysis. The announced compute futures had not begun trading at the paper's cutoff date. The propositions are therefore framed for later falsification using actual volume, open interest, quote depth, information shares, and commercial adoption. A failure to attract trading is not repaired by simulated evidence; it is itself evidence about whether a hardware-specific contract can survive long enough to become a market institution.

2. Institutional Setting and Relation to Existing Work

2.1 From a rental index to a vintage-specific futures contract

The announced CME products convert a heterogeneous service into a monthly cash-settled claim. Silicon Data publishes separate rental indices for several GPU models, including A100, H100, H200, B200, and MI300X. Raw observations are converted to a per-hour, per-GPU rate, filtered, adjusted for observable contract characteristics, aggregated at the provider level, and then combined across providers (Silicon Data, 2025, 2026). The resulting number is more comparable than a raw cloud rate card, but it remains explicitly attached to a hardware model.

That attachment offers an early-market advantage. Hardware names are salient, observable, and already used in procurement. A contract labeled H100 is easier to explain than a latent-performance unit whose conversion depends on workload, precision, interconnect, software stack, and utilization. Salience can concentrate liquidity. At the same time, the label embeds a vintage clock. NVIDIA reports that DGX B200 can deliver substantially higher training and inference performance than DGX H100 for specified workloads, while later Blackwell Ultra and Rubin systems shift the frontier again (NVIDIA, 2024, 2025, 2026). Exact performance ratios are workload-dependent, but the direction of embodied technical change is not.

The settlement index can therefore remain statistically healthy while the contract's economic constituency changes. An old vintage may have many price observations because it is widely deployed, even as new investment, frontier workload demand, and marginal hedging needs move toward a successor. Index observability and contract relevance are separate state variables.

2.2 Three literatures that do not yet meet

The first relevant literature studies futures contract innovation, viability, and survival. Early work emphasizes commercial hedging demand, homogeneous exposure, sufficient speculative participation, and contract terms that avoid manipulation or delivery problems (Silber, 1981; Carlton, 1984; Black, 1986). Later evidence shows that most new derivatives begin with low activity and that network effects can make liquidity persistent once a focal contract emerges (Cavanaugh & Penick, 2018). This literature ordinarily treats the underlying commercial need as evolving more slowly than the product-launch process. Compute reverses that presumption: the product may be trying to establish a network while its named underlier is already being displaced. A related strand studies contracts that fail outright rather than age: Johnston and McConnell (1989) analyze the rise and fall of an established financial futures contract once the match between its terms and hedgers' exposure deteriorated. The boundary of the present contribution should therefore be stated positively. The existing success-and-failure literature studies contract failure - design flaws, insufficient hedging demand, or manipulability that prevent a market from forming or cause it to wither (Johnston & McConnell, 1989; Cavanaugh & Penick, 2018). This paper studies something else: the underlier being redefined by technology, so that the contract design succeeds and the settlement index remains healthy and computable while the economic exposure being hedged migrates to a successor vintage. In the first case the cause of failure is internal to the contract; in the second it is external to it. A further adjacent literature studies competition between trading venues and the migration of liquidity across exchanges rather than across vintages (Cantillon & Yin, 2011), a distinction that matters for identification and is developed in Section 8.2.

The second literature studies vintage capital and investment-specific technical change. New capital embodies superior technology, so an installed asset can lose economic value even when it remains physically productive (Solow, 1960; Greenwood, Hercowitz, & Krusell, 1997). Replacement models distinguish physical wear from obsolescence caused by the arrival of a better vintage (Cooley, Greenwood, & Yorukoglu, 1997). The diffusion of general-purpose technologies likewise emphasizes coexistence and displacement across

vintages (Jovanovic & Rousseau, 2005). Empirical depreciation research similarly separates deterioration from revaluation due to changing relative prices and technology (Hulten & Wykoff, 1981; Fraumeni, 1997). These models concern machines and investment. They do not ask what happens to a derivative whose settlement definition is itself a machine vintage.

The third literature concerns compute pricing and financialization. Bandi and Su (2026) analyze GPU rental indices, non-storability, term rentals, and synthetic forward curves. Fernández-Fuertes and Grégoire (2026) emphasize asymmetric basis risk across flexible buyers and hardware-locked sellers. Xing (2026) proposes token-based contracts, while Yang (2026) focuses on benchmark constitution and the standardization trilemma. These studies recognize technological change, but none makes the economic lifetime of a vintage-specific futures contract the central object.

Contract half-life joins these literatures. It treats market liquidity as an institution built through coordination, commercial exposure as a vintage-dependent state variable, and technology as a force that can move the object of hedging before legal maturity. The concept is not a synonym for delisting probability. Delisting is an exchange decision and often arrives late. Half-life is a threshold in the decline of economic relevance and can be measured while the contract remains listed.

3. What Exactly Has a Half-Life?

3.1 Price life, physical life, and contract life

A GPU has several clocks. Its physical service life ends when failure, maintenance cost, or technical incompatibility makes operation uneconomic. Its rental-price life describes the decline of the hourly rate earned by the hardware. Its frontier life ends when it is no longer the preferred technology for leading workloads. A derivative adds a fourth clock: the time during which the vintage remains a focal unit for transferring price risk.

These clocks need not coincide. Strong demand for cheaper inference can support the rental rate of an older GPU long after frontier training migrates. Conversely, a new architecture announcement can redirect long-horizon hedging and investment immediately even if the old spot rental index barely moves. A physical asset can therefore remain useful while the derivative written on it loses price-discovery leadership.

Let v denote a hardware vintage. Throughout the paper the event-time origin t_0 is the successor vintage's introduction date, and $u = t - t_0 \geq 0$ is event time, so that $u = 0$ at the successor's introduction. If a contract's normalization date differs from t_0 , it is written $u_{ref} = t_{ref} - t_0$. Define $C_v(u)$ as the normalized share of economically relevant commercial exposure naturally mapped to vintage v ; $L_v(u)$ as its normalized liquidity importance; and $I_v(u)$ as its normalized informational importance. Each component equals one at u_{ref} . A parsimonious relevance index is the weighted geometric mean

$$R_v(u) = C_v(u)^\alpha L_v(u)^\beta I_v(u)^\gamma, \quad \alpha + \beta + \gamma = 1 \quad (1)$$

where the weights reflect the research question. Equal weights provide a transparent baseline; hedger-centered work may place more weight on C , while exchange-centered work may emphasize L . In logarithms, the geometric index aggregates component decay rates as a weighted arithmetic average. A component approaching zero ultimately drives R_v toward zero, but the first-passage half-life need not equal the half-life of the fastest-decaying component. This distinction matters for measurement because C - the commercial component - is precisely the one for which the available proxies are weakest. Single-component half-lives and a two-component (L , I) version should therefore be reported as main results, as set out in Section 8.1. An arithmetic aggregation can be reported as an additional robustness measure.

The contract half-life is then

$$\tau_v(1/2; u_{\text{ref}}) = \inf \{ \tau \geq 0 : R_v(u_{\text{ref}} + \tau) \leq 0.5 R_v(u_{\text{ref}}) \} \quad (2)$$

This definition is an observable threshold, not a claim that relevance follows radioactive decay or that one constant governs the full path. The baseline sets $u_{\text{ref}} = 0$ when normalization occurs at the successor's introduction; equation (2) also permits a later joint-active normalization window. Relevance can recover after a supply shock, software improvement, export restriction, or renewed demand for a cost-effective vintage. Researchers should therefore report first-passage half-life, sustained half-life (remaining below one-half for a prespecified period), and any recrossing.

The threshold also differs from a rental-price half-life, τ_p , defined by the time for the vintage's rental index to fall by one-half. If trading and commercial exposure migrate while the old rental rate remains supported by residual users, τ_v can be far shorter than τ_p . This is the distinctive empirical claim.

3.2 Legal expiry versus economic expiry

Legal maturity is contractual and deterministic. Economic maturity is endogenous. A December H100 future can settle exactly as specified even if, by December, most forward-looking procurement risk is quoted in B200 or a newer generation. The old contract still performs settlement; it no longer performs coordination for the main market.

Economic expiry is not identical to zero volume. Thin trading can remain informative for a specialized constituency, and block activity may not appear as continuous order-book depth. Nor is high volume sufficient for economic life: temporary market-making incentives or speculative activity can support turnover without broad commercial adoption. The relevance index therefore separates trading, information, and commercial exposure instead of using one metric as a universal success test.

A further distinction is between a delivery-month contract and a product family. Individual futures always approach expiry. Contract half-life concerns the continuing economic relevance of a vintage-specific family across listed maturities, measured on constant-maturity or aggregated terms. It asks whether the H100 complex remains a focal market after controlling for routine rolling between calendar months.

4. A Simple Vintage-Migration Model

4.1 Commercial relevance after an innovation shock

Consider an incumbent vintage o and a successor n introduced at calendar date t_0 , so that event time is $u = t - t_0$ with $u = 0$ at the successor's introduction, exactly as in Section 3.1. Let λ denote effective innovation intensity: a reduced-form combination of performance improvement, expected availability, software readiness, financing capacity, and the probability that users can substitute toward the new vintage. Two features of λ should be fixed at the outset. It is a reduced-form scale parameter with units of inverse time that summarizes the realized intensity of substitution; it is not itself an arrival rate. The Poisson hazard λH introduced in Section 4.4 is the arrival rate of frontier-displacing events, and in a steady state $\lambda \approx \lambda H$ multiplied by the average substitution magnitude per event. Empirically the two are identified from different data - λ from realized adoption and exposure shares, λH from dated event counts - so they can be estimated and tested separately. The incumbent's natural commercial exposure share follows

$$C_o(u) = \exp(-\delta_c \lambda u) \quad (3)$$

where $\delta_c > 0$ captures workload substitutability. Hardware-locked sellers, regulated workloads, or installed software stacks reduce δ_c ; flexible buyers and frontier applications increase it. The old rental price need not fall at the same rate. Let

$$P_o(u) / P_o(0) = \exp(-\delta_p \lambda u) + q(u), \quad q(0) = 0 \quad (4)$$

where $q(u)$ captures residual scarcity, cost-sensitive demand, supply retirement, and common compute-market shocks. Even with a low δ_p , commercial exposure can migrate rapidly if the successor becomes the marginal unit for investment and frontier risk. Thus price persistence does not imply benchmark persistence.

The distinction is particularly important for service prices. Unlike the resale value of a chip, a rental rate reflects both embodied capability and the equilibrium of heterogeneous workloads. An older GPU can become more intensively used as it becomes cheaper. That response can slow price depreciation while accelerating the migration of high-value risk to the successor.

4.2 Liquidity migration and network effects

Liquidity is endogenous because traders prefer the contract other traders already use. Let $L_o(u)$ denote the incumbent's normalized cross-vintage liquidity importance after removing mechanical calendar-roll effects. A reduced-form transition written in cumulative-hazard form is

$$L_o(u) = \exp\{-\int_0^u [\delta_L \lambda + v A_n(s)] ds\} \quad (5)$$

where δL maps commercial migration into order flow, $A_n(s)$ in $[0,1]$ measures the successor contract's availability and adoption, and $v \geq 0$ captures the network amplification created by depth, tighter spreads, clearing familiarity, and benchmark reference in bilateral contracts. Before the successor is listed or operationally usable, A_n can be near zero and incumbent liquidity can persist despite declining commercial fit. Once the successor crosses a coordination threshold, migration can be discontinuous. The cumulative-hazard form ensures that $\delta L \lambda + v A_n(u)$ is the instantaneous decay rate even while adoption changes over time.

This structure allows both delay and acceleration. A strong incumbent network can initially protect the old contract. The same network logic then works against it after liquidity coordinates on the new vintage. The observed half-life is therefore not a mechanical chip-release interval; it depends on exchange timing, market-maker support, contract fungibility across clearing accounts, and whether the successor benchmark is accepted by commercial users.

The adoption path $A_n(u)$ is not exogenous; it is shaped by the exchange's choice of listing date a for the successor contract. Listing early (small a) facilitates spread trading and position transfer and lowers migration frictions, but it divides the order book while commercial exposure to the successor is still thin. Listing late (large a) protects the incumbent's network but makes the eventual migration more likely to display a liquidity gap. Let $Q_o(u; a)$ and $Q_n(u; a)$ denote usable liquidity levels rather than shares, $C_n(u)$ the commercial relevance of the successor, $\phi > 0$ the cost of simultaneous order-book fragmentation, and $g > 0$ the cost of leaving successor exposure without a usable contract. A minimal design problem is

$$\max_a W(a) = \int_0^T [Q_o(u; a) + Q_n(u; a) - \phi Q_o(u; a) Q_n(u; a) - g C_n(u) (1 - A_n(u; a))] e^{-ru} du \quad (5a)$$

The objective no longer collapses to the sum of two shares. An earlier listing reduces the uncovered-successor term but increases the interval over which both order books coexist and incur fragmentation costs. A later listing does the reverse. Under regularity conditions, the optimal listing date a^* moves earlier when innovation raises C_n more quickly or when the cost g of a liquidity gap rises, and later when the fragmentation cost ϕ rises. The formulation is deliberately reduced-form, but it makes the trade-off in Proposition 4 economically identified rather than merely descriptive.

4.3 Information migration

Informational importance $I_o(u)$ measures the old contract's contribution to the common efficient price of compute. It can be estimated from lead-lag relations, forecast-encompassing tests, Hasbrouck information shares, or Gonzalo-Granger component shares once synchronized observations are sufficient (Garbade & Silber, 1983; Gonzalo & Granger, 1995; Hasbrouck, 1995). In the model, let

$$I_o(u) = \exp\{-\int_0^u [\delta_l \lambda + \psi L_n(s)] ds\} \quad (6)$$

where ψ captures the feedback from successor liquidity to information production. Price discovery may migrate faster than volume when informed traders use the new contract while residual hedgers remain in the old. Alternatively, incumbent liquidity can make the old contract informative about common compute scarcity even after commercial demand begins to migrate. The path is empirical, but the model identifies what should be measured.

Combining equations (3), (5), and (6), take logarithms of equation (1):

$$\ln R_v(u) = \alpha \ln C_v(u) + \beta \ln L_v(u) + \gamma \ln I_v(u) \quad (7a)$$

Differentiating with respect to event time gives the instantaneous decay rate of relevance as the weighted arithmetic average of the three component decay rates:

$$\kappa_v(u) = \alpha \delta_c \lambda + \beta [\delta_l \lambda + v A_n(u)] + \gamma [\delta_l \lambda + \psi L_n(u)] \quad (7)$$

Because the component decay rates need not be constant while successor adoption and liquidity are still rising, the half-life is defined by an integrated-decay condition:

$$\int_0^{\tau_v(1/2)} \kappa_v(s) ds = \ln 2 \quad (8)$$

where $\kappa_v(s)$ is the instantaneous decay rate in equation (7). When $A_n(u)$ and $L_n(u)$ have converged to constants - or when their averages over the half-life interval are used - the familiar closed-form special case obtains:

$$\tau_v(1/2) \approx \ln(2) / \bar{\kappa}_v \quad (8')$$

Equation (8') is an approximation rather than an identity. Its error depends on the shape and speed of the successor-adoption and liquidity paths; it can be evaluated only after those paths or a stated numerical parameterization are supplied. The exact integrated condition in equation (8) should be used whenever adoption is still changing materially at the horizon of interest.

The expression makes the comparative statics transparent. Faster embodied innovation, easier workload substitution, a viable successor contract, and stronger liquidity-information feedback all shorten half-life. Contract half-life is shorter than rental-price half-life whenever $\kappa_v > \delta_p \lambda$ after accounting for residual-demand shocks.

4.4 Maturity and prospective obsolescence

A hardware-specific future with remaining maturity m transfers exposure from event time u to $u + m$. If frontier-displacing innovations arrive with Poisson hazard λ_H , distinct from the effective innovation intensity λ in Section 4.1, the probability that the named vintage remains the frontier reference through maturity is $\exp(-\lambda_H m)$. Prospective exposure and installed exposure are affected through different channels. Let ω_F and ω_S be nonnegative weights summing to one, and let $C_{i,o}^{\text{stock}}(u)$ denote the share of participant i 's installed exposure mapped to the incumbent vintage. Define the hedgeability index

$$J_{i,o}(u,m) = \omega_F \exp(-\lambda_H m) + \omega_S C_{i,o}^{\text{stock}}(u) |\text{Corr}[\Delta S_i(u,m), \Delta F_o(u,m)]| \quad (9)$$

The first term is a flow channel: it measures the probability that new frontier-oriented exposure still maps to the named vintage at maturity. The second is a stock channel: it measures the hedgeability of participant i 's installed exposure. The absolute correlation makes the stock term nonnegative. As maturity lengthens, frontier-reference survival and cross-hedge correlation may both fall, producing a maturity inversion in which far-dated old-vintage contracts become economically empty before nearby contracts. The effect is not conventional time-to-expiry decay; it is prospective technological mismatch.

The additive decomposition avoids multiplying two measures that may reflect the same substitution process. If prospective-flow exposure cannot be observed, researchers may report the stock-channel contribution $\omega S_{Ci,0}^{stock(u)} | \text{Corr}[\cdot] |$ alone. Because the omitted flow term is nonnegative, this contribution is a lower bound on the full index defined in equation (9), conditional on the stated weights; the unweighted correlation by itself is a diagnostic rather than a bound.

5. Four Propositions

Proposition 1 - Relevance migrates before price disappears. *Following an identifiable innovation shock and the availability of a successor contract, the incumbent vintage's shares of open interest and price discovery will decline faster than its rental-price index.*

The null is not that the old rental price is constant. Let P denote the incumbent's normalized rental index and let X be either its open-interest share or its Hasbrouck information share. For each co-primary outcome define $DX = (X_{60} - X_0) - (P_{60} - P_0)$ over the 60 trading days following the dated innovation event. Test $H_0: DX \geq 0$ against $H_1: DX < 0$ using a one-sided time-block bootstrap over daily observations within the event window, with Holm correction across the two outcomes and a familywise 5 per cent level. Proposition 1 is supported for an outcome only when the upper endpoint of its adjusted one-sided confidence interval is below zero. An interval containing zero is inconclusive, not a falsification; a statistically precise estimate in the opposite direction supports the competing claim that benchmark relevance is at least as persistent as rental price. Ordinary calendar rolls and common compute-market shocks should be removed before the statistic is constructed. The outcomes, sign convention, window, and inference rule are stated here before trading data become available.

Proposition 2 - Distant old-vintage maturities empty first. *For an incumbent hardware-specific product, liquidity and quoted depth will retreat from longer maturities before they retreat from nearby maturities as the probability of technological displacement rises.*

A clean test compares eligible constant-maturity segments within the same vintage and controls for the exchange's listing schedule. The proposition is weakened if long-dated commercial contracts create natural hedging demand specifically in the old vintage; such evidence would identify hardware lock-in rather than invalidate the framework. A power qualification also applies. In the first six to twelve months after launch, deferred-maturity liquidity is determined largely by the listing schedule and market-maker incentives, so that period has little power. Tests should use only near-versus-deferred maturity pairs that have both been continuously listed and quote-eligible for at least twelve months. Studies should report the minimum detectable deferred-versus-nearby share gap at 80 per cent power under the observed dependence and volatility of the share measures, so that an uninformative null is not reported as support for the alternative.

Proposition 3 - Innovation intensity shortens contract half-life. *Contract half-life will be shorter when performance improvements arrive faster, workloads are more substitutable across generations, and successor capacity becomes commercially available sooner.*

Innovation should be measured by dated events and realized adoption, not by product announcements alone. Shipment delays, software immaturity, power constraints, or export controls can lower effective λ even when advertised performance rises sharply. The proposition concerns the effective innovation intensity λ of Section 4.1, not the frontier-displacement hazard λ_H of Section 4.4. The two imply separable tests in a panel containing multiple vintage transitions: half-life should shorten with adoption-based λ , while, conditional on λ , the viability of distant maturities should shorten with event-count-based λ_H . A single H100-to-B200 transition can illustrate both mechanisms but cannot identify them independently without additional cross-vintage or cross-market variation.

Proposition 4 - Durability transfers risk into governance. *Rolling-vintage or performance-adjusted benchmarks will exhibit longer contract half-lives than fixed-vintage benchmarks, but their hedge errors will become more sensitive to conversion rules, workload weights, rebalancing dates, and benchmark-governance decisions.*

This is not a claim that abstraction is inferior. It identifies a design trade-off. A durable unit avoids repeated product death only by making equivalence across vintages a modeled and revisable judgment.

6. Historical Analogues and Identification Templates

The propositions above require post-launch data that do not yet exist. Historical episodes in other markets can nevertheless clarify what a future measurement exercise must observe. The evidence below consists of three completed benchmark transitions and one never-alive control. They illustrate fixed-label replacement, internal respecification, constituent revision, and launch failure; they do not correspond one-for-one to all four design families in Section 7. Table 1 summarizes the identification role of each episode.

6.1 Specification obsolescence and contract-family replacement (fixed vintage)

When a delivery specification is retired, a fixed-label contract may have to migrate to a successor specification. CME announced that the underlying basis of its miNY unleaded gasoline futures would change from conventional unleaded gasoline to RBOB beginning with the July 2006 contract month (CME Group, 2006). The contemporaneous rule change documents a successor specification and a dated transition point. It does not, by itself, establish how quickly open interest migrated across the broader gasoline complex; that claim requires archived contract-level volume and open-interest records. The episode therefore supplies an identification template for Proposition 1 - track activity by label before and after a successor specification becomes active - rather than a numerical estimate of half-life.

6.2 Respecification within an existing contract (rolling vintage)

The alternative is to redefine the underlier without changing the contract's identity. The NYMEX heating oil contract did precisely that: from the May 2013 contract month the deliverable grade changed from a maximum sulfur content of 2,000 ppm to ultra-low-sulfur diesel at 15 ppm (U.S. Energy Information Administration, 2013). Contract identity, and with it the possibility of preserving the existing liquidity network, survived; the underlying commodity was redefined. This is an institutional analogue of a rolling-vintage design. It illustrates the risk transfer in Proposition 4: avoiding a new contract label concentrates transition basis around the respecification date for hedgers whose physical exposure does not change simultaneously.

6.3 Changing constituents to extend benchmark life (performance-adjusted and governance-based)

A benchmark can also be kept relevant by widening what it includes. S&P Global Commodity Insights announced in 2022 that WTI Midland cargoes would be reflected in Dated Brent and Cash BFOE from the June 2023 loading month, and completed the inclusion in 2023 (S&P Global Commodity Insights, 2022, 2023). The stated purpose was to reinforce the robustness of the benchmark by expanding the deliverable supply reflected in the assessment. Whether correlation with any particular commercial exposure improved is an empirical question, not an assumption made here. The institutional lesson is that comparability across time, and the judgment of when a transition is warranted, become matters for consultation and governance. This is the closest analogue to a performance-adjusted compute unit whose conversion rule is revisable, and hence to Proposition 4.

6.4 The never-alive control: launch failure versus economic expiry

A fourth episode supplies a never-alive control. The Pacific Stock Exchange announced a proposal for DRAM-chip futures in 1989, while a 2003 industry report described renewed DRAM-futures ventures and noted that earlier attempts had not succeeded (Horovitz, 1989; Souza, 2003). These sources document repeated proposals, not a continuously traded contract family. The defensible conclusion is therefore that no durable exchange-traded DRAM market was established in those episodes; it is not that a previously focal contract experienced a measurable half-life. Proposed, approved, listed, quoted, and actively traded states must be coded separately. A half-life should be computed only after a prespecified joint-active threshold is met.

These episodes are analogues, not estimates, calibrations, or formal bounds. Without harmonized dates and activity records, the paper cannot infer that energy-market transition horizons are upper bounds for GPU contracts or that effective GPU innovation intensity λ is numerically larger. Once the relevant series are assembled, historical durations can serve as priors or scale checks. The LIBOR-to-SOFR transition supplies a complementary governance lesson: administrative replacement shifts attention to the fallback provisions of legacy contracts (Alternative Reference Rates Committee, 2022), which connects directly to the discussion of perpetual compute products in Section 9. Taken together, the cases motivate three observable series for a post-launch study: activity share by label, maturity structure around a dated redefinition, and an explicit state classification that prevents never-alive proposals from being treated as rapidly expiring contracts.

Episode	Transition mechanism	Closest design issue	What it identifies
Gasoline contract moved to an RBOB basis (2006)	Underlying specification changed at a dated contract month	Fixed-label replacement analogue	Template for measuring activity migration by label; no half-life estimate is made here
NYMEX heating oil respecified to ULSD (2013)	Underlier redefined inside an existing contract identity	Rolling/respecification analogue	Transition basis around a dated redefinition
WTI Midland included in Dated Brent (2023)	Constituent set widened through benchmark governance	Governance-based constituent revision	Durability-comparability trade-off and governance discretion
DRAM-futures proposals (1989 and 2003)	Proposals did not establish a durable exchange-traded market	Never-alive control; no family correspondence	Separation of launch failure from economic expiry

Table 1. Historical analogues and their identification role.

7. Contract Design Choices

The model suggests four main design families. They differ in where they place technological risk rather than whether they eliminate it.

Design	Durability	Main benefit	Risk transferred elsewhere
Fixed hardware vintage	Low to medium	Salient label; simple settlement; concentrated early liquidity	Rapid economic expiry; repeated successor launches; vintage basis
Rolling front-vintage index	Medium to high	Maintains exposure to commercially leading hardware	Roll discontinuity; rule gaming; forced migration; transition basis

Design	Durability	Main benefit	Risk transferred elsewhere
Performance-adjusted basket	Potentially high	Comparability across hardware generations	Workload-model risk; software dependence; conversion discretion
Token-expenditure index	Potentially high	Closer to application-layer cost	Provider pricing, model efficiency, quality, and margin enter the underlier

Table 2. Technological durability and the location of residual risk.

Fixed-vintage contracts minimize methodology discretion. Their weakness is the need to relaunch liquidity around each generation: short maturity limits, early successor listings, and spread products ease migration, yet every added vintage fragments the order book.

A rolling-vintage index preserves a recognizable rule but creates a discrete transition. The roll date becomes economically important: hedgers can face a gap between installed hardware and the new constituent, and participants may try to influence eligibility.

The comparison becomes sharper when the two designs are written as minimum-variance hedge-error variances over a fixed evaluation horizon H . Let $\rho_0(u; \lambda) = \rho_0 \exp(-\chi\lambda u)$, with $\chi > 0$, so that the correlation between the hedger's exposure and the incumbent future decays with effective innovation intensity. Using the optimal hedge ratio, the fixed-vintage error variance is

$$\text{Var}[e_{\text{fixed}}(H)] = \sigma_s^2 [1 - \rho_0^2 \exp(-2\chi\lambda H)] \quad (10)$$

Under a rolling design, suppose the exposure-futures correlation is maintained near an average level $\bar{\rho}$ over H . If N_h rolls occur within the horizon, each contributes a zero-mean transition error with variance σ_t^2 , and transition errors are mutually independent and uncorrelated with the continuous hedge error, then

$$\text{Var}[e_{\text{roll}}(H)] = \sigma_s^2(1 - \bar{\rho}^2) + N_h\sigma_t^2 \quad (11)$$

Equating equations (10) and (11) defines the critical intensity implicitly. Under the stated functional form, $\lambda^* = -(2\chi H)^{-1} \ln\{[\bar{\rho}^2 - N_h\sigma_t^2/\sigma_s^2]/\rho_0^2\}$, provided $0 < \bar{\rho}^2 - N_h\sigma_t^2/\sigma_s^2 < \rho_0^2$. The rolling design has lower error variance for $\lambda > \lambda^*$, while the fixed-vintage design dominates for $\lambda < \lambda^*$. The expression also makes the governance trade-off explicit: λ^* increases with transition-basis variance and with the number of rolls. If the independence assumptions fail, covariance terms must be added and λ^* should be estimated from the full error covariance matrix rather than from equation (11).

Performance adjustment converts hardware into a common effective-compute unit. Bandi and Su (2026) show the limit with H100-equivalent throughput: prices per normalized unit still differ across generations, because peak operations per second omit memory, interconnect, software, and utilization. The conversion rule is a hedonic model whose revision changes the exposure being hedged.

Token expenditure moves further downstream. It may be durable across chips, but it bundles hardware cost with model architecture, inference efficiency, and provider margins, so it is not a neutral substitute for GPU-hour risk.

8. Measuring Contract Half-Life After Launch

8.1 Observable components

The framework is designed for post-launch measurement with actual trading records. The commercial component C can be proxied by the share of disclosed rental contracts, provider quotations, capacity listings, procurement references, or financing documents mapped to each vintage. No single source is sufficient; the

measure should preserve provider, region, commitment term, and service-tier provenance. Coverage is a real constraint rather than a detail: procurement disclosures, lease agreements, and financing documents are difficult to obtain systematically and are rarely comparable across providers. A tiered strategy is therefore appropriate. Tier 1 uses the full three-component index $R = C^\alpha L^\beta I^\gamma$. Tier 2 drops the commercial component and reports $\tilde{R} = L^{\beta'} I^{\gamma'}$ with $\beta' + \gamma' = 1$, to be used whenever C is unavailable; because commercial exposure usually migrates before order flow does, omitting C biases the estimated half-life upward (too long), whereas under hardware lock-in with early liquidity migration the bias runs the other way, so the expected direction must be stated for the sample at hand. Tier 3 reports single-component half-lives and the disagreement among them. Main results should be reported at Tier 2, with Tier 1 as an extension where the data permit.

The liquidity component L should combine open-interest share, volume share, quoted spread, depth, zero-trade frequency, and maturity concentration. Open interest is more informative about risk transfer than one-day launch volume, while spread and depth distinguish a few large blocks from a continuously usable market. Routine expiration rolls must be removed by aggregating a vintage family or using constant-maturity series.

The information component I should begin with event-aligned lead-lag and forecast comparisons. Formal information-share methods require sufficiently frequent synchronized nonzero changes; they should not be forced onto sparse early data. The unit of analysis is relative contribution across vintages, not whether one contract alone predicts its own settlement index.

A transparent baseline can normalize each component to one over the first stable 20-trading-day window after both incumbent and successor products are active. Equal-weight geometric relevance then produces a first-passage and sustained half-life. Alternative weights, arithmetic aggregation, and single-component half-lives should be reported to show whether the result is driven by liquidity, commercial adoption, or information.

8.2 Event time and identification

Calendar time is an imprecise innovation clock. Relevant events include architecture announcements, general availability, major cloud-provider deployment, benchmark inclusion, software support, export-control changes, and the listing of a successor future. Each event changes a different element of λ or A_n . A credible study should preclassify events and avoid selecting only those followed by visible price movements.

The strongest early design is a within-market comparison of old and new vintage shares around dated events. The outcome is relative migration, not a causal estimate from two treated contracts. The announced CME/Silicon Data and ICE/Ornn families already create a prospective competing-benchmark setting (ICE, 2026). If both products begin trading, their relative liquidity, benchmark composition, and participant clienteles can be compared descriptively, but exchange choice and benchmark composition are jointly determined. Until the ICE/Ornn products actually begin trading, the comparison remains institutional rather than empirical.

Identification must also separate cross-venue from cross-vintage migration. An observed decline in an old vintage's share can reflect competition between venues rather than succession between vintages; the network-effect mechanism is the same in both cases, but the policy implication is not (Cantillon & Yin, 2011). The design implication is to compare vintage shares within a single exchange, holding the venue fixed, and to treat cross-venue share changes as a separate control rather than as evidence of obsolescence. The migration of Bund futures from LIFFE to DTB illustrates the identification logic for a cross-venue liquidity shift. It does not, without a separately estimated mapping, numerically calibrate the network-amplification parameter v .

No obsolescence conclusion should be drawn from a secular decline in H100 rent alone. Proposition 1 requires relative movement toward a successor in open interest or information importance. Missing data must

remain distinct from true zero volume. Provider quotes can support proxy cross-hedging analysis but cannot establish participant-level hedge effectiveness. These constraints preserve the distinction between a testable theory and a retrospective narrative.

Object	Primary measure	Timing	Failure criterion
Commercial relevance	Vintage share of mapped procurement, quotes, capacity, and contract references	Monthly or event time	No relative migration after successor availability
Liquidity relevance	Open-interest and volume share; spreads; depth; zero-trade frequency	Daily/intraday	Old-vintage activity is as persistent as its rental price
Information relevance	Lead-lag, forecast contribution, information/component share	Intraday/daily after data threshold	No shift in price discovery toward successor
Contract half-life	First passage and sustained passage of R_v below 0.5	From joint active window	Estimate is weight-sensitive or reflects missing observations

Table 3. Empirical implementation of contract half-life.

9. Implications for Exchanges, Hedgers, and Regulators

For exchanges, launch is not the end of design: a vintage ladder requires a migration protocol. Successor contracts should be listed early enough to support spreading and position transfer, but not so early that liquidity fragments across thin labels - the trade-off in equation (5a). Exchanges should also publish open interest by vintage and maturity.

For hedgers, the core risk is benchmark lock-in. A seller financing an H100 fleet may need a fixed-vintage hedge; a flexible buyer may prefer rolling exposure. Hedge quality cannot be inferred from a shared GPU label: hedge ratios must be estimated on actual cash flows (Ederington, 1979; Lien & Tse, 2002).

For benchmark administrators, methodology continuity is not economic continuity. An unchanged index preserves auditability while relevance decays; revised constituents preserve relevance while breaking comparability, as Section 6.3 shows. Governance should specify transition triggers, parallel publication, and back-history treatment (IOSCO, 2013).

For regulators, contract half-life is a customer-protection and market-integrity issue. A perpetual maturity does not solve vintage obsolescence: it removes calendar expiry while making benchmark substitution more consequential, as the LIBOR fallback experience in Section 6 shows. A perpetual rolling contract survives only through discretionary transitions.

The broader lesson is sharpest for semiconductor process nodes. A contract written on a named node inherits a testable criterion: when the volume-production ramp of the next node is shorter than half the contract's listed maturity horizon, the fixed-node contract's half-life will be shorter than its nominal term. Battery chemistries and telecommunications capacity raise the same issue.

10. Conclusion

Compute derivatives expose a clock that conventional futures analysis often leaves implicit. A contract may reach its legal settlement date and still have expired economically much earlier. The reason is not that the old underlier becomes worthless. It is that commercial exposure, liquidity, and price discovery migrate to a successor vintage while the old price remains observable.

Contract half-life provides a precise way to study that migration. It is the time for a vintage-specific derivative's economic relevance to fall by half, measured across commercial, trading, and informational functions. A simple model shows why faster innovation, easier substitution, longer maturity, and a viable successor contract shorten half-life. Fixed-vintage designs concentrate liquidity but age quickly; rolling and performance-adjusted designs last longer by transferring risk into transition rules and benchmark governance.

The propositions await real trading data. That is a feature rather than a defect of the research design. The concept is stated before the first announced CME compute contracts begin trading, so later evidence can show whether H100 and B200 activity actually migrates as predicted. Even a failure to launch will be informative: a benchmark can remain publishable without a futures contract becoming economically alive. The central question is therefore not only whether compute futures trade, but whether their definition of compute survives long enough to matter.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Declaration of competing interests: The author declares no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability: No empirical dataset is analyzed in this conceptual paper. The post-launch measures proposed here are intended for future implementation using actual exchange, benchmark, and commercial-exposure data.

Ethics statement: This research did not involve human participants, personal data, animals, or interventions requiring ethical approval.

AI-use disclosure: Generative AI tools were used for language and document-production assistance. The author is responsible for the argument, source selection, verification, and final text.

Version and date: Working Paper v3.1, 27 August 2026.

References

- Alternative Reference Rates Committee. (2022). LIBOR legacy playbook. Federal Reserve Bank of New York. <https://www.newyorkfed.org/arrc/publications/overall-transition-materials>
- Bandi, F. M., & Su, Y. (2026). (Early) AI compute asset pricing. arXiv:2607.12156. <https://arxiv.org/abs/2607.12156>
- Black, D. G. (1986). Success and failure of futures contracts: Theory and empirical evidence. Monograph Series in Finance and Economics, Monograph No. 1986-1. Salomon Brothers Center for the Study of Financial Institutions, New York University.
- Cantillon, E., & Yin, P.-L. (2011). Competition between exchanges: A research agenda. *International Journal of Industrial Organization*, 29(3), 329-336.
- Carlton, D. W. (1984). Futures markets: Their purpose, their history, their growth, their successes and failures. *Journal of Futures Markets*, 4(3), 237-271. <https://doi.org/10.1002/fut.3990040302>
- Cavanaugh, G., & Penick, M. (2018). The lifecycle of exchange-traded derivatives. *Journal of Commodity Markets*, 10, 47-68. <https://doi.org/10.1016/j.jcomm.2018.05.007>
- CME Group. (2006). NYMEX to change underlying contract for NYMEX miNY unleaded gasoline futures. Press release, 19 May 2006.

- https://www.cmegroup.com/media-room/press-releases/2006/5/19/nymex_to_change_underlyingcontractfornymexminytunleadedgasoline.html
- CME Group. (2026a). CME Group and Silicon Data to launch compute futures on October 5 to unlock new way to hedge AI risks. Press release, 11 August 2026. https://www.cmegroup.com/media-room/press-releases/2026/8/11/cme_group_and_silicondatatolaunchcomputeuturesonoctober5tounloc.html
- CME Group. (2026b). Initial listing of two (2) compute futures contracts. Special Executive Report SER-9785, 11 August 2026. <https://www.cmegroup.com/notices/ser/2026/08/ser-9785.html>
- Commodity Futures Trading Commission. (2026). Request for comment on the listing of compute derivatives contracts. 91 Federal Register 54259, 21 August 2026. <https://www.federalregister.gov/documents/2026/08/21/2026-17163/request-for-comment-on-the-listing-of-compute-derivatives-contracts>
- Cooley, T. F., Greenwood, J., & Yorukoglu, M. (1997). The replacement problem. *Journal of Monetary Economics*, 40(3), 457-499.
- Ederington, L. H. (1979). The hedging performance of the new futures markets. *Journal of Finance*, 34(1), 157-170. <https://doi.org/10.1111/j.1540-6261.1979.tb02077.x>
- Fernández-Fuertes, R., & Grégoire, V. (2026). Compute futures: Who is left exposed? SSRN Working Paper No. 7224223. <https://doi.org/10.2139/ssrn.7224223>
- Fraumeni, B. M. (1997). The measurement of depreciation in the U.S. national income and product accounts. *Survey of Current Business*, 77(7), 7-23.
- Garbade, K. D., & Silber, W. L. (1983). Price movements and price discovery in futures and cash markets. *Review of Economics and Statistics*, 65(2), 289-297.
- Gonzalo, J., & Granger, C. W. J. (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics*, 13(1), 27-35. <https://doi.org/10.1080/07350015.1995.10524576>
- Greenwood, J., Hercowitz, Z., & Krusell, P. (1997). Long-run implications of investment-specific technological change. *American Economic Review*, 87(3), 342-362.
- Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *Journal of Finance*, 50(4), 1175-1199. <https://doi.org/10.1111/j.1540-6261.1995.tb04054.x>
- Horovitz, B. (1989). Pacific Exchange may list futures in computer chips. *Los Angeles Times*, 30 May 1989.
- Hulten, C. R., & Wyckoff, F. C. (1981). The measurement of economic depreciation. In C. R. Hulten (Ed.), *Depreciation, Inflation, and the Taxation of Income from Capital* (pp. 81-125). Urban Institute Press.
- Intercontinental Exchange (ICE). (2026). ICE and Ornn to launch GPU compute futures contracts. Press release, 19 May 2026. <https://ir.theice.com/press/news-details/2026/ICE-and-Ornn-to-Launch-GPU-Compute-Futures-Contracts/default.aspx>
- International Organization of Securities Commissions (IOSCO). (2013). Principles for financial benchmarks. Final Report FR07/13. <https://www.iosco.org/library/pubdocs/pdf/IOSCOPD415.pdf>
- Johnston, E. T., & McConnell, J. J. (1989). Requiem for a market: An analysis of the rise and fall of a financial futures contract. *Review of Financial Studies*, 2(1), 1-23. <https://doi.org/10.1093/rfs/2.1.1>
- Jovanovic, B., & Rousseau, P. L. (2005). General purpose technologies. In P. Aghion & S. N. Durlauf (Eds.), *Handbook of Economic Growth*, Vol. 1B (pp. 1181-1224). Elsevier.
- Lien, D., & Tse, Y. K. (2002). Some recent developments in futures hedging. *Journal of Economic Surveys*, 16(3), 357-396. <https://doi.org/10.1111/1467-6419.00172>

- NVIDIA. (2024). NVIDIA Blackwell platform arrives to power a new era of computing. News release, 18 March 2024. <https://nvidianews.nvidia.com/news/nvidia-blackwell-platform-arrives-to-power-a-new-era-of-computing>
- NVIDIA. (2025). Inside NVIDIA Blackwell Ultra: The chip powering the AI factory era. NVIDIA Technical Blog, 22 August 2025. <https://developer.nvidia.com/blog/inside-nvidia-blackwell-ultra-the-chip-powering-the-ai-factory-era/>
- NVIDIA. (2026). NVIDIA HGX platform. Product information, accessed 27 August 2026. <https://www.nvidia.com/en-us/data-center/hgx/>
- Silber, W. L. (1981). Innovation, competition, and new contract design in futures markets. *Journal of Futures Markets*, 1(2), 123-155. <https://doi.org/10.1002/fut.3990010205>
- Silicon Data. (2025). Silicon Data announces major revision to A100 & H100 rental indices, launches first-ever B200 index, and introduces new hyperscaler rental benchmarks. 5 December 2025. <https://www.silicondata.com/news-room/silicon-data-announces-major-revision-to-a100-h100-rental-indices-launches-first-ever-b200-index-and-introduces-new-hyperscaler-rental-benchmarks>
- Silicon Data. (2026). GPU rental price indices: Daily benchmarks across the AI compute market. Accessed 27 August 2026. <https://www.silicondata.com/products/silicon-index>
- Solow, R. M. (1960). Investment and technical progress. In K. J. Arrow, S. Karlin, & P. Suppes (Eds.), *Mathematical Methods in the Social Sciences, 1959* (pp. 89-104). Stanford University Press.
- Souza, C. (2003). DRAM futures trading stages a comeback. *EBN/EE Times*, 28 July 2003.
- S&P Global Commodity Insights. (2022). Platts to include WTI Midland in Dated Brent-related assessments from June 2023. Subscriber note, 8 June 2022. <https://www.spglobal.com/energy/en/news-research/latest-news/crude-oil/060822-platts-to-include-wti-midland-in-dated-brent-related-assessments-from-june-2023>
- S&P Global Commodity Insights. (2023). Platts Dated Brent now reflects WTI Midland crude oil, completes inclusion in Brent complex. Press release, 2 May 2023. <https://press.spglobal.com/2023-05-02-Platts-Dated-Brent-Now-Reflects-WTI-Midland-Crude-Oil%2C-Completes-Inclusion-in-Brent-Complex>
- U.S. Energy Information Administration. (2013). Heating oil futures contract now uses ultra-low sulfur diesel fuel. *Today in Energy*, 10 May 2013. <https://www.eia.gov/todayinenergy/detail.php?id=11211>
- Xing, Y. (2026). AI token futures market: Commoditization of compute and derivatives contract design. arXiv:2603.21690. <https://arxiv.org/abs/2603.21690>
- Yang, F. (2026). Futures before fungibility: Compute derivatives and the financialization of an unstandardized market. SSRN Working Paper No. 7324139. <https://doi.org/10.2139/ssrn.7324139>