A Forced Uniqueness Theorem for Inventory Market Making: Axiomatic Foundations of the Avellaneda–Stoikov and Cartea–Jaimungal Frameworks

Frank M. V. Feys

August 10, 2026

Abstract

We consider the dynamic preference functional of an inventory market maker and show that five axioms on it (cash-additivity, normalization, concavity, strong dynamic consistency, and law-invariance) determine it uniquely as the entropic certainty-equivalent on liquidation-adjusted terminal wealth, parametrized by a single positive scalar γ. The Avellaneda–Stoikov framework is the unique representative of this axiom class. The characterization has three structural consequences. First, the Cartea–Jaimungal terminal-inventory coefficient α turns out not to be a preference parameter at all, but a property of the market microstructure; under a mild regularity condition on the liquidation cost function L, it equals $\frac{1}{2}L''(0)$. Second, the Cartea–Jaimungal objective does not arise as a primitive preference functional within the axiom system, but appears instead as the second-order Taylor expansion of the axiomatically-determined entropic functional in inventory magnitude. Third, the parameter identification $\phi = \gamma\sigma^2/2$, already present in Cartea et al. (2015, p. 261) as an ODE-level match between two specific model formulations under constant volatility, arises here as a corollary of the axiomatic characterization. The relation extends to stochastic volatility through clock-invariance of γ against the quadratic-variation clock, and inverts to $\gamma = 2\phi/\sigma^2$, providing a consistency cross-check on independently calibrated desk parameters. Further consequences of the framework include a robust-optimization dual under Knightian ambiguity, incompatibility with dynamic conditional-value-at-risk preferences, and a multi-asset extension that preserves the single-γ structure.

MSC 2020 classification. Primary: 91B16 (utility theory), 91G70 (statistical methods; risk measures). Secondary: 49L20 (dynamic programming in optimal control and differential games), 60H30 (applications of stochastic analysis), 91G10 (portfolio theory), 91G80 (financial applications of other theories), 93E20 (optimal stochastic control).

Keywords. Inventory market making; forced uniqueness; axiomatic preference theory; Avellaneda–Stoikov; Cartea–Jaimungal; dynamic consistency; entropic risk measure.

1 Introduction

A market maker, in the inventory-management tradition, faces a problem with a deceptively simple structure. She posts bid and ask quotes around a prevailing mid-price, earns the spread when her quotes are hit, accumulates inventory (typically unwanted) in the process, and must manage the risk that the mid-price moves against her position before she can unwind it, all under uncertainty about future prices and order arrivals. At the preference-functional level, the mathematical literature is organized around two principal frameworks, each emphasizing a different representation of risk and inventory aversion.

The first is the Avellaneda–Stoikov (AS) framework. Originating with Ho and Stoll (1981) and cast in modern quantitative form by Avellaneda and Stoikov (2008), the AS market maker is endowed with constant absolute risk aversion (CARA) preferences over terminal wealth, comprising any residual inventory marked to the terminal mid-price net of a convex liquidation penalty. The optimal quoting strategy solves a Hamilton–Jacobi–Bellman equation, and the resulting reservation-price and optimal-spread formulas are standard textbook material, with the risk-aversion coefficient γ as the single free preference parameter.

The second is the Cartea–Jaimungal framework, which we shall refer to as the CJ tradition. Beginning with Cartea and Jaimungal (2015) and codified in Cartea et al. (2015), the CJ market maker maximizes expected terminal wealth subject to a running quadratic inventory penalty and a terminal liquidation penalty. The running-penalty coefficient ϕ and the terminal-penalty coefficient α are treated as independent free parameters, typically calibrated to market data or chosen on tractability grounds, with no principled relation between them.

The two traditions are presented as competing alternatives. In Cartea et al. (2015), Guéant (2017), and essentially every applied paper we have surveyed, the choice between AS and CJ is governed by tractability, by the analytic form of the resulting strategy, or by the practitioner's modeling preference, but not by an axiomatic argument on preferences. This is the situation that we wish to revise.

Two prior works unify the AS and CJ frameworks by reducing them to a common object. Cartea et al. (2015, Section 10.3) show that under the parameter identification $\phi = \gamma\sigma^2/2$ (in the notation used here), the value-function ODEs for the CARA and running-penalty market makers coincide in structure, and the two optimal quoting strategies agree up to a constant shift. Guéant (2017) extends this pattern to a broader HJB ansatz. Both reductions operate at the level of the value function and its ODE, for specific model formulations under constant volatility. Our contribution operates one level up, at the preference functional itself, and characterizes the class of preference functionals a coherent market maker must have. The parameter identification $\phi = \gamma\sigma^2/2$ then arises as a corollary of the axiomatic characterization in the constant-volatility case, and extends to stochastic volatility through a quadratic-variation clock. Structural consequences invisible at the ODE level, notably the forcing of α by the liquidation cost curvature, the inadmissibility of the CJ objective as a primitive functional, and incompatibility with dynamic conditional-value-at-risk, follow at the functional level.

Before proceeding, we note that we do not model adverse selection, queue position, toxic flow, or the latency-sensitive components of execution; our results address the preference-functional layer alone, which sits above these microstructural concerns and constrains how their outputs should be aggregated into a quoting policy.

We propose to derive the market maker's objective from a small set of axioms, stated and motivated in Section 3 by asking whether a working market maker would endorse each on reflection. Some of the axioms are rather uncontroversial, such as cash-additivity (a sure dollar adds a dollar of value) and normalization (a zero-terminal-wealth strategy has zero value); others carry economic content, such as concavity, strong dynamic consistency, and law-invariance. We also record three additional natural properties, all derivable from the five core axioms: *monotonicity* (M), the *wealth-summary* property (W), and *right-continuity in time* (R). One structural consequence is microstructural and, to our knowledge, novel: *clock-invariance* of the risk-aversion parameter with respect to the price process's quadratic-variation clock, which we argue is the right time-stationarity property for an inventory market maker. The constant-volatility AS benchmark satisfies clock-invariance vacuously, whereas the standard CJ extension to stochastic-volatility settings violates it, as we show in Section 6 via Proposition 36. Clock-invariance is not an additional axiom but a consequence of the entropic representation, recorded as Corollary 12 below.

Our main result, Theorem 10, is that the five axioms force the market maker's preference functional to be the entropic certainty-equivalent on liquidation-adjusted terminal wealth, parametrized by a single positive scalar γ. Equivalently, the associated risk measure $\rho_t = -J_t$ is the (conditional) entropic risk measure. The Avellaneda–Stoikov framework is the unique axiom-consistent representative; the Cartea–Jaimungal framework is the second-order Taylor expansion of the same functional, with its free hyperparameters ϕ and α no longer free but fixed by the relations $\phi = \gamma\sigma^2/2$ and $\alpha = \frac{1}{2}L''(0)$. The axiomatic characterization of the market maker's preference functional presented here has not, to our knowledge, appeared in the inventory market making literature. The parameter identification $\phi = \gamma\sigma^2/2$ that follows as a corollary in the constant-volatility case is due to Cartea et al. (2015, p. 261) at the ODE level, and the closest structural precursor is Guéant (2017), discussed in detail in Section 7.3. Our contribution lies at the preference-functional level rather than at the value-function level.

The result is conditional on the axiom system. Theorem 10 establishes the forcing as a consequence of the five core axioms. It is not a free-standing claim about all reasonable market makers, and it does not preclude a practitioner from operating outside the axiom class if she has good reasons to do so. The substantive content most likely to be challenged is the wealth-summary property (Proposition 7), which commits the market maker to evaluating strategies solely through their liquidation-adjusted terminal wealth W_T^L, and thereby excludes path-functional preferences such as intraday drawdown aversion (discussed in Section 7.2). A skeptic might object that wealth-summary is too restrictive an assumption. The derived-property framing of Section 3.5 answers this objection precisely: wealth-summary is not an axiom, and the skeptic cannot reject it directly. It is a theorem, derived from cash-additivity (J1), normalization (J2), and strong dynamic consistency (J4) alone. Any objection to wealth-summary must therefore take the form of a rejection of one of those three core axioms, each of which carries independent and transparent economic justification. In particular, a practitioner who endorses cash-additivity, normalization, and strong dynamic consistency (earlier preferences should not knowingly disagree with later ones on the same information) is automatically committed to wealth-summary, whether or not she finds the wealth-summary framing intuitive on its face. The forcing conclusions of the theorem then follow for her as a matter of pure logic.

From our main theorem, four sharp consequences follow. First, the AS framework is uniquely correct among law-invariant dynamically-consistent inventory market making models, and the CARA utility is forced by the axioms rather than chosen for tractability (Corollary 19). Second, under a mild regularity condition on the liquidation cost, the CJ terminal-penalty coefficient α is a property of the market microstructure, equal to $\frac{1}{2}L''(0)$, not a preference parameter (Corollary 23). Third, the CJ objective does not arise as a primitive preference functional within the axiom system, but appears as the second-order Taylor expansion of the entropic functional around zero inventory, with the running coefficient forced to $\phi = \gamma\sigma^2/2$ in the constant-volatility case (Corollary 21). A practitioner calibrating (ϕ, α, γ) as three free parameters is therefore over-parametrizing on both sides of the preference–market boundary. Fourth, dynamic conditional-value-at-risk preferences are time-inconsistent in our framework (Corollary 31), with implications for the gap between desk-internal optimization and CVaR-based regulatory frameworks. In Section 5 we develop eight corollaries, including the four named above.

Executive summary. For the reader who wants the bottom line before the details:

- *What the axioms determine.* The market maker's dynamic preference functional is the entropic certainty-equivalent on liquidation-adjusted terminal wealth, parametrized by a single positive scalar $\gamma > 0$; the Avellaneda–Stoikov framework is the unique

representative of this axiom class, and everything else in both the AS and CJ frameworks follows from γ and the market inputs.

- *What was already known.* The parameter identification $\phi = \gamma\sigma^2/2$ appears in Cartea et al. (2015, p. 261) as an ODE-level match between the AS and CJ frameworks in constant volatility, and in Guéant (2017) as an HJB-level change of variables.
- *What is new from the axioms.* The identification arises here as a corollary of the axiomatic characterization, extends to stochastic volatility through clock-invariance of γ against the quadratic-variation clock (whence the correct inventory penalty is $\int_0^T \sigma_s^2 q_s^2\, ds$, not $\int_0^T q_s^2\, ds$), and inverts to $\gamma = 2\phi/\sigma^2$, providing a preference-parameter audit instrument for desks in the CJ tradition. The CJ objective itself does not arise as a primitive preference functional within the axiom system.
- *What is separately forced.* Under a mild regularity condition on the liquidation cost function, the CJ terminal-penalty coefficient α is a property of the market microstructure, equal to $\frac{1}{2}L''(0)$. A desk calibrating (ϕ, α, γ) as three free parameters is over-parametrizing on both sides of the preference–market boundary.
- *What is not covered.* Adverse selection, queue position, intraday drawdown limits, and regime-switching risk aversion operate at a different layer of the market-making stack and are outside the framework.
- *A regulatory note.* No CVaR-based objective satisfies strong dynamic consistency (J4), so a desk's time-consistent internal optimization is entropic, while the regulator's CVaR limit is a different, non-entropic object.

Outline. Section 2 introduces the market microstructural setup, the strategy space, and the liquidation-adjusted terminal wealth; Section 3 states and motivates the five core axioms and three derived properties. Section 4 proves the main theorem and records clock-invariance (Corollary 12); Section 5 develops eight further corollaries; Section 6 derives the forced coefficient via HJB (Proposition 32) and extends to stochastic volatility (Proposition 36). Section 7 opens with the operational implications for a working desk (Section 7.1), then discusses scope (in particular, the comparison with Cartea et al. (2015) and Guéant (2017), and with drawdown-averse market makers), records further structural connections, and points to four directions for future work; Section 8 concludes.

2 The Market Making Setup

We fix a finite trading horizon $T > 0$ and a filtered probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$, where $\mathbb{F} = (\mathcal{F}_t)_{t\in[0,T]}$ is a fixed filtration satisfying the usual conditions (right-continuity and completeness). We further assume that the initial σ-algebra $\mathcal{F}_0$ is $\mathbb{P}$-trivial, that is, every $A \in \mathcal{F}_0$ satisfies $\mathbb{P}(A) \in \{0, 1\}$. This is a standard assumption in the dynamic risk measure literature (Kupper and Schachermayer, 2009). It expresses the requirement that no information beyond the model primitives is available to the market maker at the start of the trading horizon. We also assume that the terminal σ-algebra $\mathcal{F}_T$ is non-atomic: for every $A \in \mathcal{F}_T$ with $\mathbb{P}(A) > 0$ there exists $B \in \mathcal{F}_T$ with $B \subseteq A$ and $0 < \mathbb{P}(B) < \mathbb{P}(A)$. This is satisfied by any filtered Brownian space or more generally by any space on which a continuously-distributed random variable is $\mathcal{F}_T$-measurable. The filtration $\mathbb{F}$ is given as a primitive of the model and represents the information available to the market maker at each time $t \in [0, T]$; $\mathcal{F}_t$ is interpreted as the σ-algebra of all events whose occurrence is known by time t. All processes in this paper are adapted to $\mathbb{F}$ in the standard sense of stochastic analysis, meaning that the

value at time t is $\mathcal{F}_t$-measurable. The primitives, namely the mid-price S (together with its volatility σ in the stochastic-volatility setting), the order-flow counts N^a, N^b, the strategy $\pi = (\delta^a, \delta^b, C)$, the intensities λ^a, λ^b, and the preference functional J, are made adapted by direct assumption; all derived processes (the quadratic variation $\langle S \rangle$, the canonical clock Λ, the cash X, the inventory q, the wealth W, and the liquidation-adjusted terminal wealth W_T^L) are then adapted automatically. The strategy π, the volatility σ, and the intensities λ^a, λ^b are required to be *predictable*, the strengthening of adaptedness appropriate when integrating against jump or diffusion components.

Throughout, we work in $L^\infty = L^\infty(\mathcal{F}_T)$, the space of bounded $\mathcal{F}_T$-measurable random variables, together with the natural conditional spaces $L^\infty(\mathcal{F}_t)$ for $t \in [0, T]$. This is the same setting as the one used in the dynamic risk measure literature; see, for example, Kupper and Schachermayer (2009) and Bion-Nadal (2009). The boundedness assumption is a technical convenience that simplifies the representation theorems on which our main result rests, and is standard in the axiomatic literature on dynamic risk measures.

Each component is introduced at the minimal level of generality required for the main result, and we note explicitly where stronger or weaker hypotheses would suffice.

2.1 Price Process and Order Flow

The mid-price $(S_t)_{t \in [0,T]}$ is a continuous semimartingale with quadratic variation process $\langle S \rangle$. The constant-volatility benchmark, which we shall use as a running example, is $S_t = S_0 + \sigma B_t$ for a Brownian motion B and a constant $\sigma > 0$, so that $d\langle S \rangle_t = \sigma^2 \, dt$. More generally, we allow S to be of the form $dS_t = \mu_t \, dt + \sigma_t \, dB_t$ with σ a positive predictable process, in which case $d\langle S \rangle_t = \sigma_t^2 \, dt$. The drift μ plays no role in our analysis and we set $\mu = 0$ without loss of generality for the structural results.

Order flow is described by two counting processes, $N^a = (N_t^a)$ and $N^b = (N_t^b)$, representing the cumulative number of fills at the market maker's ask and bid, respectively. We take N^a and N^b to be Cox processes (i.e., doubly stochastic Poisson processes) with intensities λ_t^a and λ_t^b that depend on the market maker's quote distances (δ^a, δ^b) and on the relevant state variables. The classical AS specification is $\lambda^a(\delta^a) = Ae^{-\kappa \delta^a}$ for the ask intensity at distance δ^a from the mid, and similarly for the bid, but we shall not need this functional form for the structural results; we require only that λ^a, λ^b be predictable, nonnegative, and locally integrable.

For the technical conditions of Appendix B to be nonvacuous on the canonical Brownian setup, we work throughout with a stopped version of the price process. Fix a deterministic $M > 0$ and let $\tau_M := \inf\{t \in [0, T] \mid |S_t| > M\}$, with $\tau_M = T$ if the infimum is not attained; the *stopped mid-price* $\bar{S}_t := S_{t \wedge \tau_M}$ is a continuous semimartingale with $|\bar{S}_t| \leq M$ almost surely and $\langle \bar{S} \rangle_t = \langle S \rangle_{t \wedge \tau_M}$. Since M can be chosen so that $\mathbb{P}(\tau_M \leq T)$ is arbitrarily small, the stopping is operationally invisible on intraday horizons, with all structural results holding for S on the high-probability event $\{\tau_M > T\}$; for the rest of the paper, S refers to $\bar{S}$, and we drop the bar notation. A treatment without the stopping, using an Orlicz space in place of L^∞ in the spirit of Cheridito and Li (2008), delivers the same structural conclusions.

2.2 Strategy Space

The market maker's decision profile over the trading horizon has two components: a quoting policy (where to post limit orders relative to the mid-price) and a cash-management policy (how much capital to commit to or withdraw from the trading account). We model the strategy as a triple $\pi = (\delta^a, \delta^b, C)$, where δ^a, δ^b are $\mathbb{R}$-valued predictable processes (the quote distances from the mid-price) and C is an $\mathcal{F}$-adapted càdlàg process of finite variation on $[0, T]$ (the cumulative cash injection into the trading account). Predictability of δ^a, δ^b is required

because the quoting decision at time t must be measurable with respect to information strictly prior to t (to be matched against the jump-arrival processes N^a, N^b, which are integrated against in (1)). The cash injection C is integrated as a Stieltjes integral against time, which requires only adaptedness; predictability would be a stronger condition than necessary and we do not impose it. We allow for more general control structures, so as to accommodate possible extensions to multi-tick or hedge-augmented strategies. The cash-injection component C is more than a mere modeling convenience. Indeed, it is mathematically necessary for the reduction argument underlying Theorem 10, since Lemma 46 (Appendix B) uses terminal cash injections to establish that the range of W_T^L over Π is all of $L^\infty(\mathcal{F}_T)$, a density condition allowing the reduced functional $\tilde{J}_t$ to be defined on $L^\infty(\mathcal{F}_T)$ and the Kupper–Schachermayer representation theorem to apply. Its operational significance in quoting is secondary; for the purposes of the present results, the reader may take C to consist of a single lump-sum injection at time 0. The closure of Π under bounded cash injection is a modeling hypothesis on the strategy space, logically independent of the five preference axioms of Section 3, and is verified for the specific admissibility conditions of Definition 43 in Lemma 46.

We take Π, the set of admissible strategies, to be a convex subset of the space of such triples, satisfying integrability and boundedness conditions sufficient for existence and pathwise uniqueness of the wealth and inventory dynamics introduced below. Precise admissibility requirements are stated in Definition 43 of Appendix B.

We shall make use of the following construction. Given any strategy $\pi = (\delta^a, \delta^b, C) \in \Pi$, any $t \in [0, T]$, and any $\mathcal{F}_t$-measurable bounded random variable c, the *cash-injected strategy* $\pi^{[c]}$ is obtained from π by adding c to the cash-injection process at time t. To be precise, $\pi^{[c]} = (\delta^a, \delta^b, C^{[c]})$ with $C_s^{[c]} := C_s + c\,\mathbf{1}_{\{s \geq t\}}$. The quote-distance and inventory processes are unchanged, the terminal cash satisfies $X_T(\pi^{[c]}) = X_T(\pi) + c$, and the liquidation-adjusted terminal wealth satisfies $W_T^L(\pi^{[c]}) = W_T^L(\pi) + c$ almost surely. Admissibility of $\pi^{[c]}$ is verified in Lemma 46 (Step 1) of Appendix B.

2.3 Wealth and Inventory Dynamics

Given a strategy $\pi \in \Pi$, the market maker's cash holdings $X = (X_t)$ and inventory $q = (q_t)$ evolve according to

$$dX_t = (S_t + \delta_t^a)\, dN_t^a - (S_t - \delta_t^b)\, dN_t^b + dC_t, \tag{1}$$

$$dq_t = dN_t^b - dN_t^a, \tag{2}$$

with initial conditions $X_0 \in \mathbb{R}$ and $q_0 \in \mathbb{R}$. Equation (1) says that each fill at the ask adds $S_t + \delta_t^a$ to the cash account (the market maker sells at her ask), each fill at the bid subtracts $S_t - \delta_t^b$ (she buys at her bid), and dC_t is her cash injection or withdrawal; equation (2) says fills at the bid increase inventory by one unit and fills at the ask decrease it by one unit. We work in units in which the lot size is one share; the extension to arbitrary lot sizes is mechanical. The two processes are coupled: inventory is determined entirely by the fill history, while cash depends on both the fill history and the prevailing mid-price at each fill time. A sequence of round-trip trades (a bid fill followed by an ask fill at the same price) leaves inventory unchanged but contributes positively to cash, reflecting the earned spread; a directional inventory accumulation contributes to cash at each fill price but leaves the market maker exposed to subsequent mid-price moves. The mark-to-market wealth is $W_t := X_t + q_t S_t$, which in a frictionless market with liquidation at the mid-price is the natural measure of instantaneous value; however, W_t does not account for the fact that liquidating q_t at the prevailing mid-price is generally not feasible, since market impact, the bid-ask spread of the market into which one liquidates, and the time required to unwind a large position all reduce the actually-realizable value. We capture this through the liquidation cost function.

2.4 The Liquidation Cost Function

Let $L\colon \mathbb{R} \to \mathbb{R}_+$ denote the *liquidation cost function*, where $L(q)$ is the cash cost of unwinding inventory q at the terminal time T; we require $L(0) = 0$ and L convex on $\mathbb{R}$. A third property, *symmetry* ($L(q) = L(-q)$), holds in settings where long and short positions are interchangeable (e.g., a stock with no short-sale constraint); we mention it for completeness but do not require it.

Convexity captures the standard market-impact intuition that larger inventory positions are disproportionately more expensive to unwind, and is compatible with most explicit market-impact models in the literature (linear, square-root, power-law); we do not need an explicit functional form for L, since the structural results depend only on convexity and on $L(0) = 0$.

Given L, the *liquidation-adjusted terminal wealth* of strategy π is

$$W_T^L(\pi) \coloneqq X_T(\pi) + q_T(\pi)\, S_T - L\big(q_T(\pi)\big),$$

the random variable in terms of which the market maker's preferences are expressed.

Remark 1 (Boundedness of W_T^L)**.** Under the admissibility conditions from Definition 43 (see Appendix B), every strategy $\pi \in \Pi$ satisfies $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$, and the formal statement and proof are recorded as Lemma 45 in the appendix. The boundedness is the regime in which the dynamic-risk-measure machinery of Kupper and Schachermayer (2009) applies directly; it can be relaxed at the cost of additional integrability conditions on the strategy space, and we believe the same uniqueness conclusion holds in a Cheridito–Li-style Orlicz-space treatment (Cheridito and Li, 2008), although we do not verify this here. In practice the bound is innocuous, since every realistic intraday market-making strategy faces hard inventory and loss limits.

Remark 2 (Deterministic Liquidation Cost Function)**.** We have taken L to be a deterministic function of terminal inventory, the standard AS/CJ assumption. A more realistic model takes $L\colon \mathbb{R} \times \Omega \to \mathbb{R}$ with $L(q, \cdot)$ an $\mathcal{F}_T$-measurable random function reflecting the realized depth and microstructure of the book at terminal time; our axiomatic machinery operates on W_T^L as a random variable, not on the internal structure of L, so the structural results survive this generalization unchanged provided $W_T^L \in L^\infty(\mathcal{F}_T)$, and the forced coefficient $\alpha = \frac{1}{2}L''(0)$ would then simply become an $\mathcal{F}_T$-measurable random variable.

2.5 The Preference Functional

The market maker's dynamic preference functional is a family

$$J = (J_t)_{t\in[0,T]}, \quad J_t\colon \Pi \to L^\infty(\mathcal{F}_t),$$

in the sense of the dynamic risk measure literature (Föllmer and Schied, 2016, Cheridito et al., 2006, Acciaio and Penner, 2011). Here $J_t(\pi)$ is the conditional preference value of strategy π as seen at time t, expressed as an $\mathcal{F}_t$-measurable random variable in cash units. Observe that a higher J_t is better: π is preferred to π' when $J_t(\pi) \geq J_t(\pi')$ almost surely. The associated risk measure $\rho_t \coloneqq -J_t$ also takes values in $L^\infty(\mathcal{F}_t)$; we use both J and ρ as needed.

2.6 Inputs, Choice, and Outputs

Before stating the axioms, we fix the framework's inputs and outputs. The *market inputs*, fixed throughout the paper, are the objects defined in Sections 2.1–2.4: the filtered probability space, horizon T, mid-price process S and its quadratic variation, order-flow processes N^a, N^b, admissible strategy space Π, wealth and inventory dynamics, and liquidation cost function L.

These are properties of the market, not choices the agent makes. The *agent's choice* is the dynamic preference functional J, defined in Section 2.5. This is the one object on which she has genuine freedom, and the axioms of Section 3 constrain only J. The conceptual force of this split is that any test of the framework can be attributed cleanly to one side or the other.

The main result, Theorem 10, says that under five axioms on J, the agent's choice collapses to a one-parameter family, in that J is the entropic certainty-equivalent on W_T^L, parametrized by a single positive scalar γ; the scalar γ is the only preference-side degree of freedom that survives the axioms. The two most consequential downstream quantities are the running inventory coefficient $\phi = \gamma\sigma^2/2$ (Section 6) and the terminal penalty coefficient $\alpha = \frac{1}{2}L''(0)$ (Corollary 23); the first is a preference parameter masquerading in the Cartea–Jaimungal tradition as a free running cost but in fact determined by γ and the realized volatility, and the second is a market parameter masquerading as a preference parameter but in fact determined by the curvature of L at zero inventory. A practitioner who calibrates (ϕ, α, γ) as three independent free parameters is therefore over-parametrizing on both sides of the preference–market boundary at once.

3 Axioms for the Market Maker's Objective

In this section we lay out the five axioms which together determine the market maker's dynamic preference functional. We state each axiom precisely, motivate it from the perspective of a working trader, and contrast it with what the existing literature (i.e., the Avellaneda–Stoikov and Cartea–Jaimungal traditions) assumes either implicitly or as a primitive.

The list of axioms is, roughly, the following. The first three (J1, J2, and J3) encode standard monetary properties of the preference functional: cash-additivity, normalization at zero terminal wealth, and the diversification preference (concavity). The fourth (J4) is dynamic consistency, in the strong sense. The fifth (J5) is law-invariance, that is, the requirement that the market maker's ranking depends only on the distribution of liquidation-adjusted terminal wealth. A separate relevance condition, in the sense of Kupper and Schachermayer (2009), is not part of our axiom system; it is a derivable property, recorded in Section 3.4 below and invoked when the representation theorem of Kupper–Schachermayer is applied in the proof of Theorem 10.

In addition, three natural properties are derivable from (J1)–(J5) rather than independent axioms: *monotonicity* (in liquidation-adjusted terminal wealth), the *wealth-summary property* (all dependence of J_t on the strategy is mediated by W_T^L), and *right-continuity in time* (the path $t \mapsto J_t(\pi)$ is right-continuous in probability). The derivations are recorded in Section 3.5 below, with full proofs deferred to Appendix A.

Recall from Section 2.6 that the market data is fixed. The axioms below constrain only the agent's preference functional $J = (J_t)_{t\in[0,T]}$.

3.1 Cash-Additivity, Normalization, and Concavity

The first three axioms are basic monetary and diversification properties shared with most monetary risk measures, namely scale on cash, scale at zero, and aversion to dispersion. For axiomatic treatments of dynamic monetary risk measures in adjacent settings, see also Cerreia-Vioglio et al. (2011) and Drapeau and Kupper (2013); these works characterize broader classes of risk measures, whereas the present paper forces a one-parameter family by combining law-invariance and strong dynamic consistency under the market making setup.

Axiom J1 (Cash-Additivity)**.** For all $t \in [0, T]$, all $\pi \in \Pi$, and all $\mathcal{F}_t$-measurable bounded c, the cash-injected strategy $\pi^{[c]}$ of Section 2.2 satisfies $J_t(\pi^{[c]}) = J_t(\pi) + c$ almost surely.

Axiom J1 expresses the basic monetary character of the preference functional. The value J_t is measured in the same units as wealth, and adding a sure amount of cash to the book translates one-for-one into the preference value. In the risk-measure literature, this property is variously called cash-invariance or also translation-invariance (Artzner et al., 1999, Föllmer and Schied, 2016). The conditional form, with c an $\mathcal{F}_t$-measurable random variable, encodes a zero-discount-rate assumption: cash is the numéraire, with no stochastic discounting between t and T. This is the standard convention in the AS/CJ literature; in a multi-period economy with a stochastic interest rate, Axiom J1 would need modification.

Axiom J2 (Normalization)**.** For every $t \in [0, T]$ and every $\pi_0 \in \Pi$ with $W_T^L(\pi_0) = 0$ almost surely, it holds that $J_t(\pi_0) = 0$ almost surely.

Axiom J2 fixes the scale of the preference functional. A strategy whose liquidation-adjusted terminal wealth is identically zero has zero cash-equivalent value at every time. The axiom is a standard normalization in the dynamic risk measure literature, corresponding to the requirement $\rho_t(0) = 0$ in Kupper and Schachermayer (2009, Definition 1.8). It is independent of cash-additivity. Axiom J1 fixes how J_t responds to cash shifts, but says nothing about the absolute level of J_t on any one strategy.

Before stating the next axiom, we introduce a piece of notation. Given $\pi, \pi' \in \Pi$ and an $\mathcal{F}_t$-measurable $\lambda \in [0, 1]$, we write $\lambda\pi \oplus (1-\lambda)\pi'$ for any strategy in Π whose liquidation-adjusted terminal wealth equals $\lambda W_T^L(\pi) + (1-\lambda)W_T^L(\pi')$ almost surely. Existence of at least one such strategy is guaranteed by Lemma 46 in Appendix B. Multiple such strategies may exist, and Axiom J3 should be read as a constraint on J_t holding for every such representative. Under the wealth-summary property (W) (Proposition 7, derived from Axioms J1, J2, and J4 below), all representatives receive the same J_t-value, and the constraint collapses to a single condition.

Axiom J3 (Concavity)**.** *(J3a) Weak inequality.* For all $t \in [0, T]$, all $\pi, \pi' \in \Pi$, and all $\mathcal{F}_t$-measurable λ with $0 \le \lambda \le 1$,

$$J_t(\lambda\pi \oplus (1-\lambda)\pi') \ge \lambda J_t(\pi) + (1-\lambda)J_t(\pi') \quad \text{almost surely.}$$

(J3b) Strict inequality at $t = 0$. For all $\pi, \pi' \in \Pi$ such that $W_T^L(\pi) - W_T^L(\pi')$ is not almost-surely equal to a deterministic constant, and all deterministic $\lambda \in (0, 1)$,

$$J_0(\lambda\pi \oplus (1-\lambda)\pi') > \lambda J_0(\pi) + (1-\lambda)J_0(\pi').$$

Axiom J3 is the formal statement of risk aversion. The weak concavity inequality says that the value of the average payoff is at least the average of the values; equivalently, mixing two strategies is weakly preferred to the convex combination of the two valuations. This is the standard utility-theoretic expression of preference for diversification, applied to the cash-equivalent rather than to an underlying utility function. The strictness clause is stated at $t = 0$ with deterministic λ for the following reason. At $t = 0$ with $\mathcal{F}_0$ $\mathbb{P}$-trivial, every $\mathcal{F}_0$-measurable λ is deterministic anyway, so the distinction between "deterministic λ" and "$\mathcal{F}_0$-measurable λ" collapses. The clause is restricted to $t = 0$ to match the scope of law-invariance (J5), which is itself stated only at $t = 0$. The role of the clause is to rule out the risk-neutral limit ($\gamma = 0$, conditional expectation) of the entropic family that emerges in Section 4.

The asymmetry between J3a (stated for all $\mathcal{F}_t$-measurable $\lambda \in [0, 1]$ and all t) and J3b (deterministic $\lambda \in (0, 1)$, $t = 0$ only) is deliberate. Statement J3a is the strong conditional form, needed to verify the convexity hypothesis of the Kupper–Schachermayer representation theorem in its full conditional generality; statement J3b is the minimal strict-concavity requirement that rules out the boundary case $\gamma = 0$, and suffices at $t = 0$ because the Kupper–Schachermayer theorem identifies γ from a single nondegenerate evaluation (Sub-step 3b of

the proof of Theorem 10), after which strict concavity at general t follows from the explicit entropic form. Stating J3b with $\mathcal{F}_t$-measurable λ for general t would be a stronger axiom, derivable from the rest of the system once the entropic representation is in hand; whether the strengthening follows from a proper subset of the axioms not involving (J5) is an open question we do not pursue here.

Axiom J3 states the concavity of J_t, which is the standard expression of risk aversion; on $\rho_t = -J_t$ this becomes convexity of the risk measure (Föllmer and Schied, 2016), and we move between the two conventions as the literature being discussed requires.

3.2 Strong Dynamic Consistency

We turn now to the dynamic consistency requirement, which ensures that the market maker's preferences at time s are consistent with her preferences at any later time $t \geq s$.

Axiom J4 (Strong Dynamic Consistency)**.** For all $s, t \in [0, T]$ with $s \leq t$ and all $\pi, \pi' \in \Pi$, if $J_t(\pi) \geq J_t(\pi')$ almost surely, then $J_s(\pi) \geq J_s(\pi')$ almost surely.

Of all the axioms, Axiom J4 is the one which most directly drives the uniqueness conclusion of Theorem 10. It says that if at some future time t, in every state of the world, the market maker prefers π to π', then she must also prefer π to π' at every earlier time $s \leq t$. A rational agent cannot, at time s, knowingly anticipate that she will (uniformly) prefer π to π' at time t, and yet rank them in the reverse order at time s. If she did, she would be planning to renegotiate her own preferences against herself.

The literature distinguishes several notions of dynamic consistency; the version above is the *strong* version, sometimes also called rectangularity or recursive consistency, while weaker versions (e.g., acceptance- and rejection-consistency in the sense of Weber (2006)) are also studied. Only the strong version forces the uniqueness conclusion we shall obtain, and many practical risk frameworks (including the Basel-style CVaR framework discussed in Section 5.8) operate with something weaker. Our position is that strong dynamic consistency is the most appropriate rationality requirement for an inventory market maker, since a time-s ranking disagreeing with the time-t ranking on the same information structure is the Strotzian time-inconsistency of Strotz (1955), which creates an exploitable inconsistency relative to time-consistent agents. A reader who prefers a weaker form may read the forcing conclusions as conditional on the choice of (J4), noting that the resulting framework commits to a time-inconsistent agent.

Under the other axioms, Axiom J4 is equivalent to the recursive Bellman-type identity

$$J_s(\pi) = J_s\big(J_t(\pi)\big), \qquad s \leq t,$$

where the right-hand side is interpreted as the time-s certainty-equivalent of the time-t certainty-equivalent of π, viewed as a (possibly random) bounded cash amount. The forward direction (Axiom J4 implies the Bellman identity) is verified at $s = 0$ in Step 2 of the proof of Theorem 10 below, which is all the proof requires; the full $s \leq t$ version then follows from the entropic representation by the tower property. The corresponding discrete-time statement appears as Föllmer and Schied (2016, Theorem 11.18), and in continuous time the equivalence holds under right-continuity and $J_t(\pi) \in L^\infty(\mathcal{F}_t)$, both of which the present framework provides. This is the form in which the axiom is most directly comparable to the time-consistency condition of Kupper and Schachermayer (2009).

3.3 Law-Invariance

The last of the five core axioms is law-invariance, asking that the market maker's ranking of strategies depend only on the unconditional probability law of the liquidation-adjusted terminal wealth W_T^L.

Axiom J5 (Law-Invariance)**.** For all $\pi, \pi' \in \Pi$, if $W_T^L(\pi) \overset{d}{=} W_T^L(\pi')$, then $J_0(\pi) = J_0(\pi')$ almost surely.

Axiom J5 expresses that the market maker's ranking of strategies depends only on the distribution of liquidation-adjusted terminal wealth. Two strategies that produce the same probability law for W_T^L are ranked identically, regardless of any other features that distinguish them, such as inventory sample paths, trade timing, or counterparty identity. The condition is studied in the static setting by Kusuoka (2001) and in the dynamic setting by Jouini et al. (2006) and Kupper and Schachermayer (2009). The axiom is stated at $t = 0$, matching the form of Kupper and Schachermayer (2009, Definition 1.8); since $\mathcal{F}_0$ is $\mathbb{P}$-trivial, the unconditional and time-0 conditional laws of W_T^L coincide, making the statement well-defined. A version at general $t > 0$ would be too strong: the conditional value $J_t(\pi)$ depends on the conditional law of $W_T^L(\pi)$ given $\mathcal{F}_t$, which is generally not determined by the unconditional law. Combined with cash-additivity, concavity, and dynamic consistency, law-invariance is the missing ingredient that determines the entropic form among all dynamic monetary convex risk measures.

Remark 3 (The Knightian-Ambiguity Qualification)**.** Working traders typically endorse Axiom J5 on reflection, with one common qualification: "My preferences are law-invariant given the model, but my ambiguity about the model is not." This is the Knightian-ambiguity objection, and it is well taken. We treat it as strictly outside the scope of the present paper and refer to Gilboa and Schmeidler (1989) or Maccheroni et al. (2006) for the relevant framework. A recent extension of the law-invariance machinery to settings where law-invariance holds only on a sub-σ-algebra of $\mathcal{F}_T$ is developed by Shen et al. (2026), which may provide a natural bridge between the strictly law-invariant axiomatization of this paper and the model-ambiguity reading of the objection.

3.4 Relevance

We close the discussion of the axiom system with the relevance condition, in the sense of Kupper and Schachermayer (2009). This condition is not part of our five core axioms; it is, in fact, a consequence of those axioms, as we prove in Lemma 5 below. We state it explicitly here because it appears as a separate hypothesis in the Kupper–Schachermayer representation theorem, which we invoke in the proof of Theorem 10.

Definition 4 (Relevance)**.** The dynamic preference functional J is *relevant* if, for every $\mathcal{F}_T$-measurable event A with $\mathbb{P}(A) > 0$, every $\varepsilon > 0$, and every pair of strategies $\pi_0, \pi_1 \in \Pi$ with

$$W_T^L(\pi_1) = W_T^L(\pi_0) - \varepsilon \mathbf{1}_A \quad \text{almost surely,}$$

it holds that

$$J_0(\pi_1) < J_0(\pi_0) \quad \text{almost surely.}$$

The condition is stated at $t = 0$, matching the formulation of Kupper and Schachermayer (2009, Definition 1.8). Existence of strategies $\pi_0, \pi_1 \in \Pi$ satisfying the wealth identity is guaranteed by the closure of Π under bounded $\mathcal{F}_T$-measurable shifts (Lemma 46 of Appendix B). Note that under dynamic consistency (Axiom J4), the $t = 0$ condition propagates to every later t, and so the single-time and all-time formulations are equivalent.

Relevance expresses, plainly, that the market maker cannot be indifferent to a positive-probability loss; an agent who fails relevance has a preference functional that registers no welfare cost from a loss that occurs with positive probability, regardless of how large or how likely. The canonical example is the worst-case functional $J_0(\pi) = \operatorname{ess\,inf} W_T^L(\pi)$, which is indifferent to losses on any event not containing the worst-case scenario. From the perspective

of a rational agent operating in a genuinely stochastic environment, indifference to a positive-probability loss is difficult to defend, since it amounts to assigning zero decision weight to an outcome that will materialize with positive frequency over repeated interactions with the market. Relevance is therefore best understood not as a technical regularity condition but as a minimal rationality requirement, ruling out preference functionals that are, in an economically meaningful sense, blind to part of the probability space.

Lemma 5 (Relevance Is Derivable from the Axioms)**.** *Any dynamic preference functional J satisfying Axioms J1, J2, J3, and J4 is relevant in the sense of Definition 4. Specifically, the derivation uses the strict-concavity clause (J3b) of Axiom J3 together with the monotonicity property (M) of Proposition 6, which itself follows from Axioms J1, J2, and J4.*

Proof. Fix an $\mathcal{F}_T$-measurable event A with $\mathbb{P}(A) > 0$, $\varepsilon > 0$, and $\pi_0, \pi_1 \in \Pi$ satisfying $W_T^L(\pi_1) = W_T^L(\pi_0) - \varepsilon \mathbf{1}_A$ almost surely. Set $W := W_T^L(\pi_0)$ and $W' := W_T^L(\pi_1) = W - \varepsilon \mathbf{1}_A$, so $W \geq W'$ almost surely with $\mathbb{P}(W > W') = \mathbb{P}(A) > 0$.

The boundary case $\mathbb{P}(A) = 1$ is immediate. In this case, $W - W' = \varepsilon$ is deterministic, and Axiom J1 (cash-additivity) gives $J_0(\pi_1) = J_0(\pi_0) - \varepsilon < J_0(\pi_0)$ directly. For the remainder of the proof we therefore assume that $\mathbb{P}(A) \in (0, 1)$.

Using Proposition 6 (monotonicity (M), derived from Axioms J1, J2, J4) applied to the pair (π_0, π_1), we obtain that $J_0(\pi_0) \geq J_0(\pi_1)$ almost surely. By contradiction, suppose that $J_0(\pi_0) = J_0(\pi_1)$ on a set of positive $\mathbb{P}$-measure; since $\mathcal{F}_0$ is $\mathbb{P}$-trivial and both quantities are $\mathcal{F}_0$-measurable, this means $J_0(\pi_0) = J_0(\pi_1)$ almost surely. The random variable $W - W' = \varepsilon \mathbf{1}_A$ is not almost-surely equal to a deterministic constant (since $\mathbb{P}(A) \in (0, 1)$).

Choose a midpoint strategy $\pi_{1/2} \in \Pi$ with $W_T^L(\pi_{1/2}) = \frac{1}{2}W + \frac{1}{2}W'$, which exists by Lemma 46. By the strict-concavity clause (J3b) of Axiom J3 with $\lambda = 1/2$,

$$J_0(\pi_{1/2}) > \tfrac{1}{2}J_0(\pi_0) + \tfrac{1}{2}J_0(\pi_1) = J_0(\pi_0).$$

However, $\frac{1}{2}W + \frac{1}{2}W' \leq W$ a.s. (since $W \geq W'$), so by Proposition 6 (M), $J_0(\pi_{1/2}) \leq J_0(\pi_0)$. The two inequalities contradict each other, so $J_0(\pi_1) < J_0(\pi_0)$ almost surely. □

Since Lemma 5 shows that relevance is in fact a consequence of Axioms J1–J4 alone, with law-invariance (J5) playing no role in the derivation, we do not include relevance as a separate hypothesis in the statement of Theorem 10 or in its corollaries. We invoke it explicitly within the proof of Theorem 10, where the representation theorem of Kupper and Schachermayer (2009) lists relevance as one of its hypotheses, and we verify it there via Lemma 5. The analysis here also clarifies how the two boundary cases $\gamma = 0$ and $\gamma = \infty$ are ruled out: both fail the strict concavity clause (J3b), with the linear functional ($\gamma = 0$) giving equality on every pair and the worst-case functional ($\gamma = \infty$) giving equality on any pair with overlapping essential-infimum sets. Relevance corresponds to the $\gamma = \infty$ boundary failure specifically, captured by (J3b) together with monotonicity (M).

3.5 Derived Properties

We close the axiom system by recording three natural properties, namely *monotonicity*, the *wealth-summary property*, and *right-continuity in time*, all derivable from Axioms J1–J5. Each is stated as a numbered proposition, with the same motivational care given to the five core axioms; full proofs are deferred to Appendix A. Every property below is as much a feature of the framework as the five core axioms.

The relevance condition (Definition 4) is likewise derivable from (J1)–(J5), as established in Section 3.4 as Lemma 5. We treat it separately from the three derived properties below because it does not have a stand-alone economic reading of the kind that (M), (W), and (R) admit; it is a technical regularity condition that appears as an explicit hypothesis in the Kupper–Schachermayer representation theorem invoked in the proof of Theorem 10.

Monotonicity (M). The first derived property states that the market maker's preferences are monotone with respect to liquidation-adjusted terminal wealth W_T^L, the cash-equivalent value of her book at time T after deducting the cost of unwinding any remaining position.

Proposition 6 (Monotonicity (M))**.** *Any dynamic preference functional J satisfying Axioms J1 (cash-additivity), J2 (normalization), and J4 (strong dynamic consistency) also satisfies the following monotonicity property: for all $\pi, \pi' \in \Pi$, if $W_T^L(\pi) \geq W_T^L(\pi')$ almost surely, then $J_t(\pi) \geq J_t(\pi')$ almost surely for all $t \in [0, T]$.*

Proof. See Appendix A, Proposition 37. The argument is short: cash-additivity (J1) together with normalization (J2), applied with the cash injection $c = -W_T^L(\pi)$ at time T, gives $J_T(\pi) = W_T^L(\pi)$ for every π; strong dynamic consistency (J4) then propagates the trivial monotonicity at $t = T$ backwards to every $s \leq T$. □

The property is a basic dominance condition requiring that more cash-equivalent at the horizon, in every state of the world, must be weakly preferred. The condition is what makes liquidation-adjusted terminal wealth a meaningful object for the market maker to optimize at all. We take it to be uncontroversial. It is indeed satisfied by every framework in the inventory market making literature that we are aware of, including the original Avellaneda–Stoikov setup (Avellaneda and Stoikov, 2008) and the Cartea–Jaimungal–Penalva textbook (Cartea et al., 2015).

The wealth-summary property (W). The second derived property is the wealth-summary requirement: all dependence of J_t on the strategy is mediated by W_T^L. Concretely, two strategies that produce the same liquidation-adjusted terminal wealth almost surely must receive the same preference value at every time t.

Proposition 7 (Wealth-Summary (W))**.** *Any dynamic preference functional J satisfying Axioms J1, J2, and J4 (equivalently, any J satisfying Proposition 6) also satisfies the following wealth-summary property: for all $\pi, \pi' \in \Pi$, if $W_T^L(\pi) = W_T^L(\pi')$ almost surely, then also $J_t(\pi) = J_t(\pi')$ almost surely for all $t \in [0, T]$.*

Proof. See Appendix A, Proposition 38. The argument is two lines: if $W_T^L(\pi) = W_T^L(\pi')$ a.s., then both $W_T^L(\pi) \geq W_T^L(\pi')$ a.s. and $W_T^L(\pi') \geq W_T^L(\pi)$ a.s., and applying monotonicity (M, Proposition 6) in both directions yields $J_t(\pi) = J_t(\pi')$ a.s. □

Proposition 7 formalizes a methodological commitment. The market maker is a cash-equivalent maximizer whose dislike of inventory derives entirely from its effect on liquidation-adjusted terminal wealth, not from inventory-as-such. Inventory affects W_T^L through two channels, namely the *variance channel*, where holding nonzero inventory exposes the market maker to subsequent mid-price moves, and the *liquidation-cost channel*, where the term $L(q_T)$ charges her for any residual position at the close. A market maker who dislikes inventory through either or both of these channels is fully wealth-summary-compatible; what wealth-summary excludes is a third channel, namely intrinsic discomfort with the bare fact of holding a position, independent of where it leaves her at the close.

Two natural alternatives are therefore excluded as primitives. Path-dependent inventory aversion of the running-quadratic form $\int_0^T q_s^2 \, ds$, which appears as a primitive in the Cartea–Jaimungal tradition, is not excluded by fiat, since our claim (developed in Section 5) is that any such running penalty, if consistent with the five core axioms, must arise as a derived feature of W_T^L-preferences with a forced coefficient. Path-functional preferences in the strict sense (e.g., maximum-drawdown aversion or time-average wealth) are genuinely outside the scope of Proposition 7, as discussed in Section 7.

Remark 8 (On the Circularity Charge)**.** A skeptical reader might object that the wealth-summary property is circular, in that we use it to rule out the Cartea–Jaimungal objective as a primitive functional (Corollary 21), and it is designed precisely to exclude such path-dependent running penalties. The derived-property framing answers this directly, in that wealth-summary is not assumed but forced by cash-additivity, normalization, and strong dynamic consistency (Proposition 7); any objection must reject one of those three axioms, each independently justified earlier in this section. The framework moreover does not exclude running penalties as such, since the running penalty $\phi \int_0^T q_s^2 \, ds$ arises within it as the second-order Taylor expansion of the entropic functional (Corollary 21, Proposition 32); what is excluded is treating such a penalty as a *primitive, free-parameter* object rather than as a derived feature with a forced coefficient $\phi = \gamma\sigma^2/2$.

Right-continuity in time (R). The third derived property is a regularity condition on the path $t \mapsto J_t(\pi)$. Concretely, if $t_k \downarrow t$, then $J_{t_k}(\pi) \to J_t(\pi)$ in probability. It plays a role only in the continuous-time-to-discrete-time passage of the proof of Theorem 10.

Proposition 9 (Right-Continuity in Time (R))**.** *Every dynamic preference functional J satisfying Axioms J1–J5 also satisfies the following right-continuity property: for every $\pi \in \Pi$ and every $t \in [0, T)$, the path $s \mapsto J_s(\pi)$ is right-continuous in probability at $s = t$. That is, if $t_k \downarrow t$ with $t_k \in [t, T]$, then $J_{t_k}(\pi) \to J_t(\pi)$ in probability as $k \to \infty$.*

Proof. See Appendix A, Proposition 39. The derivation proceeds in two distinct stages. First, Axioms J1–J5 (from which relevance is derived; Lemma 5) force J to equal the entropic certainty-equivalent on W_T^L at every dyadic time $t \in \mathcal{D} \subset [0, T]$, via the Kupper–Schachermayer representation theorem applied to the restriction of $\tilde{J}$ to each dyadic sub-filtration, with consistency of the resulting parameter across scales. Second, the Bellman identity (from strong dynamic consistency, Axiom J4) together with cash-additivity (Axiom J1), applied with bounded $\mathcal{F}_{t_0}$-measurable cash injections as test functions, lifts the entropic representation from dyadic times to every $t_0 \in [0, T]$. Right-continuity then follows from the right-continuity of conditional expectations $\mathbb{E}[e^{-\gamma W_T^L} \mid \mathcal{F}_t]$ under the usual conditions on the filtration. □

The information filtration $\mathbb{F}$ is right-continuous and the price process S is a continuous semimartingale, so the primitive processes of the setup are right-continuous in t. Proposition 9 (R) says that the preference functional inherits this regularity: $J_t(\pi)$ does not jump at a non-information time. A jump in $J_t(\pi)$ at some t where the underlying primitives have not jumped would be incoherent with the cash-equivalence interpretation, since the trader could realize neither the pre-jump nor the post-jump value via any actual trade in the limit.

3.6 Summary

The five core axioms.

(J1) Cash-additivity.

(J2) Normalization.

(J3) Concavity at all t; strict at $t = 0$ on pairs with nonconstant W_T^L-difference.

(J4) Strong dynamic consistency.

(J5) Law-invariance (at $t = 0$).

Three derived properties (consequences of the axioms).

(M) *Monotonicity* (Proposition 6): if $W_T^L(\pi) \geq W_T^L(\pi')$ a.s., $J_t(\pi) \geq J_t(\pi')$ a.s. for all t.

(W) *Wealth-summary property* (Proposition 7): $J_t(\pi)$ depends on π only through $W_T^L(\pi)$.

(R) *Right-continuity in time* (Proposition 9): the path $t \mapsto J_t(\pi)$ is right-continuous in probability.

The axioms above are jointly satisfied by the entropic certainty-equivalent on W_T^L, with any strictly positive risk-aversion parameter γ. Theorem 10 below states the converse, namely that they are satisfied *only* by this family.

We prove in Appendix A that the five core axioms (J1)–(J5) form an *independent* system, in the sense that no axiom in the list is derivable from the others. The three derived properties (namely, (M) monotonicity, (W) wealth-summary, and (R) right-continuity) are derivable from (J1)–(J5), with the derivations presented in that appendix. The relevance condition of Definition 4 is also derivable from (J1)–(J5) (Lemma 5); it is treated separately from the three derived properties because it functions as a technical hypothesis of the representation theorem rather than as a stand-alone property of J.

4 The Forced Uniqueness Theorem

The proof proceeds in three steps. First, the wealth-summary property (W) (Proposition 7, derived from Axioms J1, J2, J4) reduces the market maker's preference functional to a functional on $L^\infty(\mathcal{F}_T)$. Second, this reduced functional is shown to satisfy the standard hypotheses for a law-invariant time-consistent representation. Third, the representation result yields a one-parameter family of admissible functionals, with both boundary cases ($\gamma = 0$ and $\gamma = \infty$) ruled out by the strict-concavity clause (J3b) of Axiom J3.

4.1 Statement of the Theorem

Theorem 10 (Forced Uniqueness)**.** *Let the market data $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P}, T, S, N^a, N^b, \Pi, L)$ be as specified in Section 2, and let $J = (J_t)_{t\in[0,T]}$ with $J_t \colon \Pi \to L^\infty(\mathcal{F}_t)$ be a dynamic preference functional. Then J satisfies Axioms J1–J5 of Section 3 if and only if there exists $\gamma \in (0, \infty)$ such that*

$$J_t(\pi) = -\frac{1}{\gamma} \log \mathbb{E}\Big[\exp\big(-\gamma\, W_T^L(\pi)\big) \,\Big|\, \mathcal{F}_t\Big] \tag{3}$$

for all $t \in [0, T]$ and all $\pi \in \Pi$. The scalar γ is unique.

Under the five axioms, the market maker's preference functional is forced to take the entropic certainty-equivalent form on W_T^L, with a unique positive scalar γ that is the coefficient of absolute risk aversion of the underlying utility $u(x) = -\exp(-\gamma x)$ in the sense of Pratt (1964). The scalar γ is, moreover, a property of the agent's preferences alone, independent of the analyst's choice of clock. Corollary 12 states and proves this clock-invariance, and it underwrites the stochastic-volatility forced-coefficient result of Proposition 36. The conclusion depends critically on the strong form of dynamic consistency (Axiom J4). Under acceptance-rejection consistency in the sense of Weber (2006), the uniqueness fails: shortfall risk measures of the form $\rho(X) = \inf\{m \in \mathbb{R} \mid \mathbb{E}[u(X + m)] \geq c\}$ are also consistent with the weakened axiom system, so the one-parameter entropic family is no longer the only admissible class. The dividing line between the two regimes is therefore the strength of the dynamic-consistency axiom, with strong consistency forcing a single risk parameter and weaker consistency admitting a richer family parametrized by the utility-and-threshold pair. This is a caveat discussed in Section 3.2.

Remark 11 (Why CARA Survives the Dynamic Lift)**.** A reader familiar with the static AS setting may note that the entropic functional is the certainty-equivalent of CARA expected utility, $-\frac{1}{\gamma}\log\mathbb{E}[\exp(-\gamma W)] = u^{-1}(\mathbb{E}[u(W)])$ for $u(x) = -\exp(-\gamma x)$. In the static setting this is an identity, CARA-expected-utility maximization coinciding with entropic-functional maximization for any γ. The nontrivial fact is that this coincidence extends to the dynamic setting; the dynamic entropic functional, defined by recursive application of the conditional certainty-equivalent, is the same object as dynamic expected-utility maximization on a CARA agent. The structural reason CARA emerges is that its Bellman value function factorizes as $V = -\exp(-\gamma(X + qS))\,\psi(t, q)$. In the constant-volatility benchmark, after solving the HJB equation with the CARA ansatz (Proposition 32), the associated certainty-equivalent takes the explicit form $J_t(\pi) = (X_t + q_t S_t) + \theta(t, q_t)$, depending on the strategy only through the conditional distribution of $W_T^L(\pi)$ given $\mathcal{F}_t$, and not on (X_t, q_t, S_t) individually. This is precisely the wealth-summary property (W), now visible at the level of the explicit value function. Note that the Bellman value function $V(t, X, q, S)$ itself does depend on the full state (X, q, S); what is special about CARA is this factorization, which makes the associated certainty-equivalent J_t inherit the wealth-summary property. For any other utility curvature, the Bellman value function does not factorize in this way, the dynamic programming state (cash, inventory, price) enters separately into the certainty-equivalent, and the dynamic risk-measure language and dynamic expected-utility language diverge. This is the mathematical reason CARA emerges here, and in the utility-theoretic literature more broadly CARA is essentially the unique utility for which the two languages coincide, a classical fact traceable to Merton (1971) and treated within the market-making context in Cartea et al. (2015, Section 10).

4.2 Proof of Theorem 10

The technical conditions on the strategy space and the boundedness of W_T^L over Π are collected in Appendix B. In particular, Lemma 45 establishes that $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$ for every $\pi \in \Pi$, and Lemma 46 establishes that the range $\{W_T^L(\pi) \mid \pi \in \Pi\}$ is equal to all of $L^\infty(\mathcal{F}_T)$. The proof has two directions. The forward direction (axioms imply the entropic form) proceeds in three steps: first, we reduce the problem to a question about a functional on $L^\infty(\mathcal{F}_T)$; second, we verify that this functional satisfies the axioms of Kupper and Schachermayer (2009); third, we invoke their Theorem 1.10 and translate the conclusion back into the language of our setup. The converse direction (entropic implies the axioms) is a direct verification of each axiom on the entropic functional.

(⇒) Forward direction: the axioms imply the entropic form. Assume J satisfies Axioms J1–J5. By Lemma 5, J is then also relevant in the sense of Definition 4. We shall invoke this fact when applying the Kupper–Schachermayer representation theorem in Step 3 below.

Step 1: reduction to L^∞. By property (W) (Proposition 7), for every $\pi \in \Pi$ the value $J_t(\pi)$ depends only on the random variable $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$. Define, on the range

$$\mathcal{R} := \{W_T^L(\pi) \mid \pi \in \Pi\} \subset L^\infty(\mathcal{F}_T),$$

the *reduced functional*

$$\tilde{J}_t \colon \mathcal{R} \to L^\infty(\mathcal{F}_t), \quad \tilde{J}_t(W) := J_t(\pi) \ \text{ for any } \pi \in \Pi \text{ with } W_T^L(\pi) = W.$$

By property (W) (Proposition 7), $\tilde{J}_t$ is well-defined on $\mathcal{R}$. By Lemma 46 of Appendix B (Step 3, which uses Step 1 of that lemma applied at $t = T$), the range satisfies $\mathcal{R} = L^\infty(\mathcal{F}_T)$

exactly: every bounded $\mathcal{F}_T$-measurable random variable arises as $W_T^L(\pi)$ for some $\pi \in \Pi$, via terminal-cash injection. The reduced functional $\tilde{J}_t$ is therefore defined on all of $L^\infty(\mathcal{F}_T)$ for every $t \in [0,T]$; no separate extension argument is required. Define the associated risk measure $\rho_t := -\tilde{J}_t \colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$.

We use throughout $\tilde{J}_t \colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$ for the reduced functional just defined, and $\widehat{J}_u \colon \Pi \to L^\infty(\mathcal{F}_{\tau(u)})$ for the business-time reparametrization $\widehat{J}_u(\pi) := J_{\tau(u)}(\pi)$ used in Corollary 12 below.

Step 2: verification of the dynamic risk measure axioms. We check that the family $(\rho_t)_{t\in[0,T]}$ on $L^\infty(\mathcal{F}_T)$ satisfies the hypotheses of Kupper and Schachermayer (2009, Definition 1.8): normalization, conditional cash-invariance, monotonicity, convexity, law-invariance (at $t = 0$), time-consistency, and relevance.

Normalization. By Axiom J2, $J_t(\pi_0) = 0$ a.s. for every $t \in [0,T]$ and every $\pi_0 \in \Pi$ with $W_T^L(\pi_0) = 0$ a.s. Equivalently, $\tilde{J}_t(0) = 0$ a.s. for every t, i.e., $\rho_t(0) = 0$ a.s. for every t. This is the normalization required by Kupper and Schachermayer (2009, Definition 1.8).

Cash-invariance. For any $W \in L^\infty(\mathcal{F}_T)$ and any $\mathcal{F}_t$-measurable bounded m, Axiom J1 gives $\tilde{J}_t(W + m) = \tilde{J}_t(W) + m$, i.e., $\rho_t(W + m) = \rho_t(W) - m$. This is the cash-invariance property of Kupper and Schachermayer (2009).

Monotonicity. For $W, W' \in L^\infty(\mathcal{F}_T)$ with $W \geq W'$ almost surely, Proposition 6 (M) applied to any pair of strategies $\pi, \pi' \in \Pi$ with $W_T^L(\pi) = W$ and $W_T^L(\pi') = W'$ gives $J_t(\pi) \geq J_t(\pi')$ almost surely, so $\tilde{J}_t(W) \geq \tilde{J}_t(W')$ and $\rho_t(W) \leq \rho_t(W')$ almost surely. This is the monotonicity of Kupper and Schachermayer (2009).

Convexity. For $W, W' \in L^\infty(\mathcal{F}_T)$ and any $\mathcal{F}_t$-measurable $\lambda \in [0,1]$, Axiom J3 gives $\tilde{J}_t(\lambda W + (1-\lambda)W') \geq \lambda \tilde{J}_t(W) + (1-\lambda)\tilde{J}_t(W')$, and thus

$$\rho_t(\lambda W + (1-\lambda)W') \leq \lambda \rho_t(W) + (1-\lambda)\rho_t(W').$$

This is the convexity property.

Law-invariance. For $W, W' \in L^\infty(\mathcal{F}_T)$ with $W \overset{d}{=} W'$, take any $\pi, \pi' \in \Pi$ with $W_T^L(\pi) = W$ and $W_T^L(\pi') = W'$. Then $W_T^L(\pi) \overset{d}{=} W_T^L(\pi')$, and Axiom J5 gives $J_0(\pi) = J_0(\pi')$, hence $\tilde{J}_0(W) = \tilde{J}_0(W')$ and $\rho_0(W) = \rho_0(W')$. (Law-invariance is required only at time zero in Kupper and Schachermayer, 2009. The conditional version at $t > 0$ follows automatically from the other K–S hypotheses once the representation is in hand.)

Time-consistency. We need to verify that $\rho_0(W) = \rho_0\big(-\rho_t(W)\big)$ for all $W \in L^\infty(\mathcal{F}_T)$ and all $t \in [0,T]$. Equivalently in $\tilde{J}$-form, $\tilde{J}_0(W) = \tilde{J}_0\big(\tilde{J}_t(W)\big)$, where $\tilde{J}_t(W) \in L^\infty(\mathcal{F}_t) \subseteq L^\infty(\mathcal{F}_T)$ is viewed as a terminal random variable on the right-hand side. The identity follows from Axiom J4 (strong dynamic consistency), combined with Axiom J1 (cash-additivity) and Axiom J2 (normalization), as follows.

Given $W \in L^\infty(\mathcal{F}_T)$, fix any $\pi \in \Pi$ with $W_T^L(\pi) = W$ (this π exists by Lemma 46), so that $\tilde{J}_t(W) = J_t(\pi)$. Pick any $\pi_0 \in \Pi$ with $W_T^L(\pi_0) = 0$ almost surely (existence again by Lemma 46). By Axiom J2 (Normalization), $J_t(\pi_0) = 0$ almost surely for every $t \in [0,T]$. We first verify that $J_t(\pi) \in L^\infty(\mathcal{F}_t)$ with $\|J_t(\pi)\|_\infty \leq \|W_T^L(\pi)\|_\infty$. Since $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$ by Lemma 45, write $M_\pi := \|W_T^L(\pi)\|_\infty < \infty$, so $-M_\pi \leq W_T^L(\pi) \leq M_\pi$ almost surely. Let $\pi_0^{[\pm M_\pi]}$ denote the cash-injected strategies of Section 2.2 applied to π_0, which satisfy $W_T^L(\pi_0^{[\pm M_\pi]}) = \pm M_\pi$ almost surely. By Axioms J1 and J2, $J_t(\pi_0^{[\pm M_\pi]}) = \pm M_\pi$ almost surely. Since $W_T^L(\pi_0^{[-M_\pi]}) \leq W_T^L(\pi) \leq W_T^L(\pi_0^{[M_\pi]})$ almost surely, Proposition 6 (monotonicity (M), already verified above) gives $-M_\pi \leq J_t(\pi) \leq M_\pi$ almost surely, so $J_t(\pi) \in L^\infty(\mathcal{F}_t)$ with the claimed bound. Let π^* denote the strategy obtained from π_0 by injecting the cash amount $J_t(\pi)$ at time t, that is, $\pi^* = \pi_0^{[J_t(\pi)]}$ in the notation of Section 2.2. By Axiom J1 (cash-additivity at time t), $J_t(\pi^*) = J_t(\pi_0) + J_t(\pi) = 0 + J_t(\pi) = J_t(\pi)$ almost surely; and at

the wealth level, $W_T^L(\pi^*) = W_T^L(\pi_0) + J_t(\pi) = J_t(\pi)$ almost surely. Since $J_t(\pi^*) = J_t(\pi)$ almost surely, the equality applied as a pair of inequalities (each direction) under Axiom J4 (strong dynamic consistency) yields $J_0(\pi^*) = J_0(\pi)$ almost surely. Translating to the reduced functional,

$$\tilde{J}_0(W) = J_0(\pi) = J_0(\pi^*) = \tilde{J}_0\big(W_T^L(\pi^*)\big) = \tilde{J}_0\big(J_t(\pi)\big) = \tilde{J}_0\big(\tilde{J}_t(W)\big),$$

almost surely, where the last equality uses $J_t(\pi) = \tilde{J}_t(W)$ by definition of the reduced functional (Step 1) on $W = W_T^L(\pi)$. This is the desired time-consistency identity.

Relevance. By Lemma 5, the five core axioms (J1)–(J5) imply that J is relevant in the sense of Definition 4, which is identical to the relevance condition of Kupper and Schachermayer (2009). We emphasize that relevance is therefore not an additional assumption on J, but a property derived from the five axioms. It appears here as a separate item only because the Kupper–Schachermayer theorem lists it as a separate hypothesis.

Step 3: applying the representation result. By the verification in Step 2, the family (ρ_t) is a law-invariant, time-consistent, relevant, convex dynamic risk measure on $L^\infty(\mathcal{F}_T)$. We invoke Kupper and Schachermayer (2009, Theorem 1.10), which addresses the discrete-time case (with the filtration indexed by the nonnegative integers). The continuous-time extension is the main work of this step. We present below the more direct route, which uses property (R) (Proposition 9) to lift the dyadic representation to all of $[0,T]$ via Lévy's martingale convergence theorem. An alternative, logically independent route is given in Stage 1 of the proof of Proposition 39 in Appendix A, which establishes the entropic representation at every $t \in [0,T]$ directly from (J1)–(J5) by a Bellman + cash-additivity + conditional-MGF argument, with (R) emerging as a consequence rather than a hypothesis. The two routes lead to the same conclusion; we adopt the (R)-based route here because it isolates the role of regularity in the lift. The conclusion delivered by this step is

$$\rho_t(W) = \frac{1}{\gamma} \log \mathbb{E}[\exp(-\gamma W) \mid \mathcal{F}_t], \qquad t \in [0,T],\ W \in L^\infty(\mathcal{F}_T),$$

for a unique $\gamma \in (0,\infty)$. The argument proceeds in three sub-steps.

Sub-step 3a: Dyadic restrictions and inheritance of axioms. Let $\mathbb{F}^{(n)} = (\mathcal{F}_{kT/2^n})_{k=0}^{2^n}$ be the dyadic sub-filtration of $\mathbb{F}$ at level n. The restriction of $\tilde{J}$ to $\mathbb{F}^{(n)}$, denoted $\tilde{J}^{(n)}$, is the family $(\tilde{J}_k^{(n)})_{k=0}^{2^n}$ with $\tilde{J}_k^{(n)}(W) := \tilde{J}_{kT/2^n}(W)$ for $W \in L^\infty(\mathcal{F}_T)$. We claim that $\tilde{J}^{(n)}$ inherits Axioms J1–J5 as well as the derived properties of Proposition 6 (monotonicity (M)) and Proposition 7 (wealth-summary (W)) directly from $\tilde{J}$, since each is stated as a property of $\tilde{J}_t$ for arbitrary $t \in [0,T]$ and continues to hold when t is restricted to $\{kT/2^n\}_{k=0}^{2^n}$. Relevance (Definition 4) is a single condition at $t = 0$, and is inherited by $\tilde{J}^{(n)}$ because $0 \in \{kT/2^n\}_{k=0}^{2^n}$ for every n. Proposition 9 (right-continuity (R) in time) is trivial on a finite discrete time-set and does not need separate verification on $\tilde{J}^{(n)}$. Specifically, dynamic consistency (J4) in the discrete-time form $\tilde{J}_j^{(n)}(W) = \tilde{J}_j^{(n)}(\tilde{J}_k^{(n)}(W))$ for $j \le k$ is an immediate consequence of strong dynamic consistency (Axiom J4) at the continuous-time level. Hence, $\tilde{J}^{(n)}$ is a discrete-time normalized, cash-invariant, monotone, convex, law-invariant (at $t = 0$), time-consistent, and relevant dynamic risk measure on $L^\infty(\mathcal{F}_T)$, satisfying the hypotheses of Kupper and Schachermayer (2009, Theorem 1.10). The theorem yields $\gamma_n \in [0,\infty]$ such that

$$\tilde{J}_k^{(n)}(W) = -\tfrac{1}{\gamma_n} \log \mathbb{E}[e^{-\gamma_n W} \mid \mathcal{F}_{kT/2^n}] \quad \text{for all } k \in \{0,1,\dots,2^n\},$$

with the two boundary conventions given by $\tilde{J}_k^{(n)}(W) = \mathbb{E}[W \mid \mathcal{F}_{kT/2^n}]$ at $\gamma_n = 0$ and $\tilde{J}_k^{(n)}(W) = \operatorname{ess\,inf}(W \mid \mathcal{F}_{kT/2^n})$ at $\gamma_n = \infty$. The strict-concavity clause (J3b) of Axiom J3

rules out $\gamma_n = 0$: at $\gamma_n = 0$ and $k = 0$, the representation reads $\tilde{J}_0^{(n)}(W) = \mathbb{E}[W]$ (since $\mathcal{F}_0$ is $\mathbb{P}$-trivial), which is linear in W; for any W, W' with $W - W'$ not a.s. constant and any deterministic $\lambda \in (0,1)$, $\mathbb{E}[\lambda W + (1-\lambda)W'] = \lambda\mathbb{E}[W] + (1-\lambda)\mathbb{E}[W']$ gives equality, contradicting (J3b). The relevance condition rules out $\gamma_n = \infty$: at $\gamma_n = \infty$ and $k = 0$, the representation reads $\tilde{J}_0^{(n)}(W) = \operatorname{ess\,inf}(W)$. Pick any nondegenerate $W \in L^\infty(\mathcal{F}_T)$ (which exists by the non-atomicity of $\mathcal{F}_T$) and choose an event $A \in \mathcal{F}_T$ of positive probability with $A \subseteq \{W > \operatorname{ess\,inf}(W) + \delta\}$ for some $\delta > 0$; such an A exists by definition of essential infimum. For any $\varepsilon \in (0, \delta]$, the shift $W - \varepsilon \mathbf{1}_A$ then satisfies $\operatorname{ess\,inf}(W - \varepsilon\mathbf{1}_A) = \operatorname{ess\,inf}(W)$, since $W - \varepsilon\mathbf{1}_A \geq W - \delta \geq \operatorname{ess\,inf}(W)$ on A and $W - \varepsilon\mathbf{1}_A = W$ on A^c. Hence, we have $\tilde{J}_0^{(n)}(W - \varepsilon\mathbf{1}_A) = \tilde{J}_0^{(n)}(W)$, contradicting Definition 4 (which requires strict inequality for every $\varepsilon > 0$, in particular for this ε). Therefore, $\gamma_n \in (0, \infty)$.

Sub-step 3b: Consistency of γ_n across scales. The claim is that $\gamma_n = \gamma_{n+1}$ for every n. The argument is by identifiability of the entropic parameter from a single nondegenerate evaluation. Both restrictions of $\tilde{J}$ to $\mathbb{F}^{(n)}$ and $\mathbb{F}^{(n+1)}$ agree at every common time index, in particular at $t = 0$. Since $\tilde{J}_0$ is a single well-defined object (the restriction of the same functional $\tilde{J}$ to time 0, independent of the chosen dyadic scale), both Kupper–Schachermayer representations must reproduce it: for every $W \in L^\infty(\mathcal{F}_T)$,

$$-\tfrac{1}{\gamma_n}\log\mathbb{E}[e^{-\gamma_n W}] = \tilde{J}_0(W) = -\tfrac{1}{\gamma_{n+1}}\log\mathbb{E}[e^{-\gamma_{n+1} W}],$$

equivalently,

$$\tfrac{1}{\gamma_n}\log\mathbb{E}[e^{-\gamma_n W}] = \tfrac{1}{\gamma_{n+1}}\log\mathbb{E}[e^{-\gamma_{n+1} W}]$$

for every $W \in L^\infty(\mathcal{F}_T)$. Pick any nondegenerate $W \in L^\infty(\mathcal{F}_T)$. Such W exists by the non-atomicity of $(\Omega, \mathcal{F}_T, \mathbb{P})$ assumed in Section 2: there exists an event $A \in \mathcal{F}_T$ with $\mathbb{P}(A) \in (0,1)$, and the random variable $W = \mathbf{1}_A - \mathbf{1}_{A^c}$ is nondegenerate. Let $\kappa(\gamma) \coloneqq \log\mathbb{E}[e^{-\gamma W}]$ denote its cumulant-generating function and consider the entropic certainty-equivalent function $f(\gamma) \coloneqq -\frac{1}{\gamma}\kappa(\gamma)$ on $(0,\infty)$. The agreement of the two representations at $t = 0$ on this W reads $f(\gamma_n) = f(\gamma_{n+1})$, so it suffices to show that f is strictly monotone on $(0,\infty)$. A direct calculation gives

$$f'(\gamma) = \frac{1}{\gamma^2}\kappa(\gamma) - \frac{1}{\gamma}\kappa'(\gamma) = -\frac{1}{\gamma^2}\big(\gamma\kappa'(\gamma) - \kappa(\gamma)\big).$$

The quantity in parentheses is exactly the relative entropy of the entropic-tilted measure $\mathbb{Q}$ (defined by $d\mathbb{Q}/d\mathbb{P} = e^{-\gamma W}/\mathbb{E}[e^{-\gamma W}]$) with respect to $\mathbb{P}$. Indeed, with $\kappa'(\gamma) = -\mathbb{E}_{\mathbb{Q}}[W]$, so that $\mathbb{E}_{\mathbb{Q}}[-\gamma W] = \gamma\kappa'(\gamma)$, we have

$$H(\mathbb{Q}\,\|\,\mathbb{P}) = \mathbb{E}_{\mathbb{Q}}\left[\log\frac{d\mathbb{Q}}{d\mathbb{P}}\right] = \mathbb{E}_{\mathbb{Q}}[-\gamma W - \kappa(\gamma)] = \gamma\kappa'(\gamma) - \kappa(\gamma) \geq 0,$$

with equality if and only if W is $\mathbb{P}$-almost-surely constant. Since W is nondegenerate, $H(\mathbb{Q}\,\|\,\mathbb{P}) > 0$. Thus, $f'(\gamma) < 0$ on $(0,\infty)$, so f is strictly monotone decreasing on $(0,\infty)$ for any nondegenerate W. The equation $f(\gamma) = c$ thus has at most one solution in $(0,\infty)$ for any c in its range. Hence, $\gamma_n = \gamma_{n+1}$, and we denote the common value by γ.

Sub-step 3c: Extension to all of $[0,T]$. At this stage we have the entropic representation $\tilde{J}_t(W) = -\frac{1}{\gamma}\log\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t]$ for every dyadic $t \in \mathcal{D} \coloneqq \bigcup_n \{kT/2^n \mid 0 \leq k \leq 2^n\}$. We extend to arbitrary $t \in [0,T]$ as follows. Fix t and a dyadic sequence $t_k \downarrow t$ in $\mathcal{D}$. By right-continuity of $\mathbb{F}$, $\mathcal{F}_t = \bigcap_k \mathcal{F}_{t_k}$ (Kallenberg, 2021, Lemma 7.13), so by the Lévy downward martingale convergence theorem applied to $e^{-\gamma W} \in L^\infty \subset L^1$ (Kallenberg, 2021, Theorem 7.23),

$$\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_k}] \longrightarrow \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t] \quad \text{almost surely.}$$

Using the uniform lower bound $\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_k}] \geq e^{-\gamma\|W\|_\infty} > 0$ and continuity of the logarithm, $-\frac{1}{\gamma}\log\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_k}] \to -\frac{1}{\gamma}\log\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t]$ almost surely. On the left-hand side, take

any $\pi \in \Pi$ with $W_T^L(\pi) = W$; by Proposition 9 (R), $\tilde{J}_{t_k}(W) = J_{t_k}(\pi) \to J_t(\pi) = \tilde{J}_t(W)$ in probability. By uniqueness of limits in probability (with both sides now viewed as convergent in probability, the a.s. convergence of the right-hand side implying its in-probability convergence), we have $\tilde{J}_t(W) = -\frac{1}{\gamma} \log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t]$ for every $t \in [0,T]$ and $W \in L^\infty(\mathcal{F}_T)$.

Step 3 yields $\rho_t(W_T^L) = \frac{1}{\gamma} \log \mathbb{E}[\exp(-\gamma W_T^L) \mid \mathcal{F}_t]$ for some $\gamma \in (0,\infty)$. Translating back, $J_t(\pi) = -\frac{1}{\gamma} \log \mathbb{E}[\exp(-\gamma W_T^L(\pi)) \mid \mathcal{F}_t]$, which is (3). This completes the forward direction.

($\Leftarrow$) Converse direction: the entropic form implies the axioms are satisfied. Suppose now that J has the entropic form (3) for some $\gamma \in (0,\infty)$. We verify each of the five core axioms (J1)–(J5) on J. Throughout, for $\pi \in \Pi$ write $W := W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$, so that $\|W\|_\infty < \infty$ and $e^{-\gamma W}$ is bounded above by $e^{\gamma \|W\|_\infty}$ and below by $e^{-\gamma \|W\|_\infty} > 0$. In particular, $\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t] \in L^\infty(\mathcal{F}_t)$ is bounded away from zero, so $\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t]$ is well-defined and bounded, and we have $J_t(\pi) \in L^\infty(\mathcal{F}_t)$.

(J1) Cash-Additivity. Let $\pi \in \Pi$, $t \in [0,T]$, and $c \in L^\infty(\mathcal{F}_t)$ bounded. By Lemma 46 (Step 1) applied to π at time t, the cash-injected strategy $\pi^{[c]}$ satisfies $W_T^L(\pi^{[c]}) = W_T^L(\pi) + c$ almost surely. Since c is $\mathcal{F}_t$-measurable, $e^{-\gamma c}$ is $\mathcal{F}_t$-measurable and bounded, and hence the conditional expectation pulls it out, so we get

$$\mathbb{E}[e^{-\gamma(W_T^L(\pi)+c)} \mid \mathcal{F}_t] = e^{-\gamma c}\, \mathbb{E}[e^{-\gamma W_T^L(\pi)} \mid \mathcal{F}_t] \quad \text{almost surely.}$$

Taking $-\frac{1}{\gamma} \log$ on both sides and using the identity $\log(e^{-\gamma c} \cdot Y) = -\gamma c + \log Y$ for $Y > 0$ gives $J_t(\pi^{[c]}) = J_t(\pi) + c$ almost surely.

(J2) Normalization. If $W_T^L(\pi_0) = 0$ almost surely, then $e^{-\gamma W_T^L(\pi_0)} = 1$ almost surely, so $\mathbb{E}[1 \mid \mathcal{F}_t] = 1$ almost surely, and $J_t(\pi_0) = -\frac{1}{\gamma} \log 1 = 0$ almost surely.

(J3) Concavity, with strict clause. We verify the two parts of Axiom J3 separately.

Weak inequality. Fix $t \in [0,T]$, $\pi, \pi' \in \Pi$, and an $\mathcal{F}_t$-measurable random variable λ with $0 \le \lambda \le 1$ almost surely. Write $W := W_T^L(\pi)$, $W' := W_T^L(\pi')$, $X := e^{-\gamma W}$, and $Y := e^{-\gamma W'}$, both strictly positive and bounded. Then $e^{-\gamma(\lambda W + (1-\lambda)W')} = X^\lambda Y^{1-\lambda}$ pointwise. Since $Y > 0$ almost surely and $\mathbb{E}[Y \mid \mathcal{F}_t] \ge e^{-\gamma \|W'\|_\infty} > 0$ almost surely, define the probability measure $\mathbb{Q}$ on $\mathcal{F}_T$ by $d\mathbb{Q}/d\mathbb{P} := Y/\mathbb{E}[Y \mid \mathcal{F}_t]$. Then $\mathbb{E}_\mathbb{Q}[Z \mid \mathcal{F}_t] = \mathbb{E}[YZ \mid \mathcal{F}_t]/\mathbb{E}[Y \mid \mathcal{F}_t]$ for all $Z \in L^\infty$, almost surely, and

$$\mathbb{E}[X^\lambda Y^{1-\lambda} \mid \mathcal{F}_t] = \mathbb{E}\Big[Y \cdot (X/Y)^\lambda \mid \mathcal{F}_t\Big] = \mathbb{E}[Y \mid \mathcal{F}_t] \cdot \mathbb{E}_\mathbb{Q}\Big[(X/Y)^\lambda \mid \mathcal{F}_t\Big].$$

On $\{0 < \lambda < 1\}$, for each fixed realization of the $\mathcal{F}_t$-measurable random variable λ the function $z \mapsto z^\lambda$ is concave on $(0,\infty)$; applying conditional Jensen under $\mathbb{Q}$ pointwise on ω, treating $\lambda(\omega)$ as a fixed coefficient (formally, integrate the $\mathcal{F}_t$-measurable kernel $(z,\omega) \mapsto z^{\lambda(\omega)}$ against the regular conditional distribution of X/Y given $\mathcal{F}_t$, which exists by Kallenberg (2021, Theorem 8.5)), gives

$$\mathbb{E}_\mathbb{Q}\Big[(X/Y)^\lambda \mid \mathcal{F}_t\Big] \le (\mathbb{E}_\mathbb{Q}[X/Y \mid \mathcal{F}_t])^\lambda = \left(\frac{\mathbb{E}[X \mid \mathcal{F}_t]}{\mathbb{E}[Y \mid \mathcal{F}_t]}\right)^\lambda \quad \text{almost surely.}$$

On $\{\lambda = 0\}$ and $\{\lambda = 1\}$, the inequality is an equality. Substituting back yields

$$\mathbb{E}[X^\lambda Y^{1-\lambda} \mid \mathcal{F}_t] \le \mathbb{E}[X \mid \mathcal{F}_t]^\lambda\, \mathbb{E}[Y \mid \mathcal{F}_t]^{1-\lambda} \quad \text{almost surely.} \tag{4}$$

Applying the strictly decreasing map $z \mapsto -\frac{1}{\gamma} \log z$ to both sides of (4) reverses the inequality, and using $\log(a^\lambda b^{1-\lambda}) = \lambda \log a + (1-\lambda) \log b$ for $a, b > 0$ yields

$$-\tfrac{1}{\gamma} \log \mathbb{E}[e^{-\gamma(\lambda W + (1-\lambda)W')} \mid \mathcal{F}_t] \ge \lambda J_t(\pi) + (1-\lambda) J_t(\pi') \quad \text{almost surely.}$$

The left-hand side equals $J_t(\lambda\pi \oplus (1-\lambda)\pi')$ by the wealth-summary property (W), which is the weak concavity inequality of Axiom J3.

Strict inequality at $t = 0$. Suppose $W - W'$ is not almost-surely equal to a deterministic constant, and let $\lambda \in (0,1)$ be deterministic. At $t = 0$, the σ-algebra $\mathcal{F}_0$ is $\mathbb{P}$-trivial, so the conditional expectations reduce to unconditional ones. Define the probability measure $\mathbb{Q}$ on $\mathcal{F}_T$ by $d\mathbb{Q}/d\mathbb{P} := Y/\mathbb{E}[Y]$ (well-defined since $\mathbb{E}[Y] \geq e^{-\gamma\|W'\|_\infty} > 0$). It follows that

$$\mathbb{E}[X^\lambda Y^{1-\lambda}] = \mathbb{E}\Big[Y \cdot (X/Y)^\lambda\Big] = \mathbb{E}[Y] \cdot \mathbb{E}_{\mathbb{Q}}\Big[(X/Y)^\lambda\Big].$$

The function $z \mapsto z^\lambda$ is strictly concave on $(0,\infty)$ for deterministic $\lambda \in (0,1)$, and X/Y is a strictly positive bounded random variable. An application of Jensen's inequality gives that $\mathbb{E}_{\mathbb{Q}}[(X/Y)^\lambda] \leq (\mathbb{E}_{\mathbb{Q}}[X/Y])^\lambda$, with equality if and only if X/Y is $\mathbb{Q}$-almost-surely equal to a constant. Since $d\mathbb{Q}/d\mathbb{P} = Y/\mathbb{E}[Y]$ is strictly positive almost surely, we have $\mathbb{Q} \sim \mathbb{P}$ on $\mathcal{F}_T$, so $\mathbb{Q}$-a.s. constant is equivalent to $\mathbb{P}$-a.s. constant. Now $X/Y = e^{-\gamma(W-W')}$, and $z \mapsto e^{-\gamma z}$ is a bijection on $\mathbb{R}$. Therefore, X/Y is $\mathbb{P}$-a.s. constant if and only if $W - W'$ is $\mathbb{P}$-a.s. constant. By hypothesis this fails, hence Jensen's inequality is strict:

$$\mathbb{E}[X^\lambda Y^{1-\lambda}] < \mathbb{E}[X]^\lambda\, \mathbb{E}[Y]^{1-\lambda}. \tag{5}$$

Applying the strictly decreasing map $z \mapsto -\frac{1}{\gamma}\log z$ to (5) yields

$$-\tfrac{1}{\gamma}\log \mathbb{E}[e^{-\gamma(\lambda W + (1-\lambda)W')}] > \lambda J_0(\pi) + (1-\lambda) J_0(\pi'),$$

which is the strict concavity inequality of Axiom J3 at $t = 0$.

(J4) Strong Dynamic Consistency. For $s \leq t$ in $[0,T]$ and any $\pi \in \Pi$, write

$$A_u(\pi) := \mathbb{E}[e^{-\gamma W_T^L(\pi)} \mid \mathcal{F}_u], \qquad u \in [0,T].$$

We have $J_u(\pi) = -\frac{1}{\gamma}\log A_u(\pi)$. Equivalently, $A_u(\pi) = e^{-\gamma J_u(\pi)}$ almost surely. The tower property of conditional expectations gives $A_s(\pi) = \mathbb{E}[A_t(\pi) \mid \mathcal{F}_s]$ almost surely, i.e.,

$$\mathbb{E}[e^{-\gamma W_T^L(\pi)} \mid \mathcal{F}_s] = \mathbb{E}\big[e^{-\gamma J_t(\pi)} \mid \mathcal{F}_s\big] \quad \text{almost surely.} \tag{6}$$

Note that this is the recursivity identity for the entropic functional. Now suppose $J_t(\pi) \geq J_t(\pi')$ almost surely. Then $e^{-\gamma J_t(\pi)} \leq e^{-\gamma J_t(\pi')}$ almost surely, so conditional monotonicity implies that $\mathbb{E}[e^{-\gamma J_t(\pi)} \mid \mathcal{F}_s] \leq \mathbb{E}[e^{-\gamma J_t(\pi')} \mid \mathcal{F}_s]$ almost surely. Using (6), the latter is equivalent to $\mathbb{E}[e^{-\gamma W_T^L(\pi)} \mid \mathcal{F}_s] \leq \mathbb{E}[e^{-\gamma W_T^L(\pi')} \mid \mathcal{F}_s]$ almost surely, and applying $-\frac{1}{\gamma}\log$ yields $J_s(\pi) \geq J_s(\pi')$ almost surely.

(J5) Law-Invariance. Suppose $W_T^L(\pi) \overset{d}{=} W_T^L(\pi')$. The transformation $x \mapsto e^{-\gamma x}$ is Borel-measurable, so $e^{-\gamma W_T^L(\pi)} \overset{d}{=} e^{-\gamma W_T^L(\pi')}$. In particular, $\mathbb{E}[e^{-\gamma W_T^L(\pi)}] = \mathbb{E}[e^{-\gamma W_T^L(\pi')}]$. At $t = 0$, the σ-algebra $\mathcal{F}_0$ is $\mathbb{P}$-trivial, so the conditional expectation reduces to the unconditional one and is deterministic. Thus, $J_0(\pi) = -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W_T^L(\pi)}] = -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W_T^L(\pi')}] = J_0(\pi')$ almost surely.

This completes the verification of the converse direction, and the proof of Theorem 10. □

4.3 The Clock-Invariance Corollary

The forced entropic representation of Theorem 10 has an immediate structural consequence: the scalar γ is invariant under the canonical clock change associated with the price process.

Corollary 12 (Clock-Invariance of the Risk-Aversion Parameter)**.** *Adopt the hypotheses of Theorem 10, and let* $\Lambda := \langle S \rangle$ *denote the quadratic-variation clock of the price process, with right-continuous inverse,* $\tau(u) := \inf\{t \geq 0 \mid \Lambda_t \geq u\}$ *for* $u \in [0, \Lambda_T]$*. The associated business-time filtration is* $\tilde{\mathbb{F}} = (\tilde{\mathcal{F}}_u)_{u \in [0,\Lambda_T]}$ *with* $\tilde{\mathcal{F}}_u := \mathcal{F}_{\tau(u)}$*, and the reparametrized criterion is* $\widehat{J}_u(\pi) := J_{\tau(u)}(\pi)$ *for* $u \in [0, \Lambda_T]$ *and* $\pi \in \Pi$*. Then*

$$\widehat{J}_u(\pi) = -\frac{1}{\gamma} \log \mathbb{E}\Big[\exp\big(-\gamma\, W_T^L(\pi)\big) \,\Big|\, \tilde{\mathcal{F}}_u\Big] \tag{7}$$

for all $u \in [0, \Lambda_T]$ *and* $\pi \in \Pi$*, with the same scalar* γ *as in Theorem 10.*

Proof. By Theorem 10, $J_t(\pi) = -\frac{1}{\gamma} \log \mathbb{E}[e^{-\gamma W_T^L(\pi)} \mid \mathcal{F}_t]$ for all $t \in [0, T]$ with unique $\gamma > 0$. Substituting $t = \tau(u)$ and using $\tilde{\mathcal{F}}_u = \mathcal{F}_{\tau(u)}$ yields (7). □

Remark 13. The scalar γ is a property of the agent's preferences alone, recovered identically from both the wall-clock and business-time formulations. The substantive content arises in the stochastic-volatility setting of Section 6.2: the same γ appearing in (7) determines the instantaneous running-penalty rate as $\phi_t = \gamma\sigma_t^2/2$, so γ is the agent's single time-invariant preference parameter (Proposition 36), and her mean-variance trade-off is stationary in $\langle S \rangle$, the clock against which price risk accumulates, rather than in wall-clock time. A wall-clock constant ϕ, as in direct extensions of the CJ framework to stochastic-volatility settings, is correspondingly inconsistent with the axioms, since it implies a time-varying $\gamma_t = 2\phi/\sigma_t^2$ excluded by Corollary 12.

Remark 14 (Relation to Prior Work on Constancy of γ)**.** The constancy of the risk-aversion parameter γ across time is a known consequence of law-invariance and time-consistency for the dynamic entropic risk measure: Delbaen et al. (2010, Proposition 37) characterize time-consistency for the dynamic entropic functional with a time-varying process (γ_t) and note, following Kupper and Schachermayer (2009), that constant $\gamma_t \equiv \gamma_0$ is the only law-invariant time-consistent case. Corollary 12 sharpens this in a direction not previously recorded, in that the same scalar γ governs the market maker's preferences in both wall-clock and business-time formulations, so that in the stochastic-volatility setting the local inventory-penalty rate $\phi_t = \gamma\sigma_t^2/2$ tracks the instantaneous variance pointwise with a single time-invariant γ. A recent complementary direction is Beiglböck et al. (2026), who strengthen the Kupper–Schachermayer rigidity theorem by replacing terminal-law invariance with adapted law invariance; their axiom weakening goes in a different direction from ours.

Corollary 12 encodes a microstructural commitment satisfied (vacuously) by the constant-volatility AS benchmark, used implicitly by some recent stochastic-volatility extensions (Rosenbaum and Zhang, 2022), and violated by direct wall-clock extensions of CJ. To our knowledge the commitment has not previously been stated at the preference-functional level. The market maker's inventory aversion is a commitment about how she trades off the mean and variance of her terminal wealth, and the natural units in which both accumulate are the units of price risk, namely $\langle S \rangle$; among possible clocks, $\langle S \rangle$ is privileged by the Dambis–Dubins–Schwarz theorem (Kallenberg, 2021, Theorem 18.4), which states that for any continuous semimartingale S with nontrivial quadratic variation there exists a Brownian motion B on a possibly enlarged probability space such that $S_t = B_{\langle S \rangle_t}$, making $\langle S \rangle$ the unique clock under which the price process becomes a standard Brownian motion. When the volatility σ_t is random, wall-clock and business-time diverge, and a wall-clock prescription under-penalizes inventory on high-volatility days and over-penalizes on low-volatility days; Corollary 12 commits to the business-time prescription, with the corrected calibration $\phi_t = \gamma\sigma_t^2/2$ of Proposition 36.

A concrete example makes the nature of the violation explicit. In a Heston model with $d\sigma_t^2 = \kappa(\bar{\sigma}^2 - \sigma_t^2)\,dt + \nu\sigma_t\,dW_t$, the naive CJ running penalty $\phi\int_0^T q_s^2\,ds$ with constant ϕ corresponds, in business-time units $u = \int_0^t \sigma_s^2\,ds$, to the penalty $\phi\int_0^{\Lambda_T} q_{\tau(u)}^2/\sigma_{\tau(u)}^2\,du$, with effective business-time coefficient $\phi/\sigma_{\tau(u)}^2$ random and time-varying, contradicting the forced constant-coefficient entropic structure; the correct running penalty is $\frac{\gamma}{2}\int_0^T \sigma_s^2 q_s^2\,ds$, as given by Proposition 36.

The literature on subordinated stochastic processes, originating with Clark (1973) and developed by Ané and Geman (2000) and Carr and Wu (2004), treats asset prices as time-changed Brownian motions in which the stochastic clock captures fluctuations in market activity; Al Dayri et al. (2011) establish empirically that price variance is approximately linear in trade count, suggesting that trade-count time is a viable proxy for $\langle S\rangle$-time in liquid markets, and our framework absorbs this as a structural feature of the preference functional.

Remark 15 (Constant-Volatility Vacuity)**.** When $\Lambda_t = \sigma^2 t$ is deterministic linear, $\tilde{\mathbb{F}}$ and $\mathbb{F}$ differ only by a deterministic time-rescaling, and Corollary 12 is automatically satisfied as a trivial restatement of Theorem 10; the corollary has real content only when Λ has a nontrivial random structure (stochastic-volatility, rough-volatility, or jump-diffusion).

Remark 16 (Noncanonical Clocks)**.** Corollary 12 structurally privileges the canonical clock $\Lambda = \langle S\rangle$; a market maker insisting on a different clock, say $\Lambda' = N$ (the cumulative trade-count process), is incompatible with the clock-invariance forced by the corollary unless $N_t = f(t)\,\langle S\rangle_t$ almost surely for some deterministic function $f > 0$. The empirical regularity $\mathbb{E}[d\langle S\rangle_t] \approx c\,\mathbb{E}[dN_t]$ documented by Al Dayri et al. (2011) holds in expectation only; under this weaker regularity, the trade-count clock yields at best approximate time-stationarity in the sense of Corollary 12.

4.4 Discussion of the Theorem

Theorem 10 has not, to the best of our knowledge, been stated before in the inventory market making literature. The axiom system is itself a new contribution. In particular, the wealth-summary property (W) (Proposition 7), which we derive from cash-additivity, normalization, and strong dynamic consistency rather than assume, is a new contribution to the theory of dynamic preference functionals. Clock-invariance of γ with respect to the canonical clock $\langle S\rangle$ (Corollary 12) is a further new structural contribution and underwrites the framework's applicability to stochastic-volatility settings. Of the five core axioms, strong dynamic consistency (J4) is the one that most directly drives the uniqueness conclusion, since it forces the recursivity relation and collapses the one-parameter family of Φ-divergence risk measures onto the entropic member (see Remark 27). The non-atomicity of $\mathcal{F}_T$ enters at a single point of the proof, namely the identifiability argument of Sub-step 3b, where it ensures the existence of a nondegenerate test random variable to identify γ_n. Finally, the proof uses, at one step, a standard representation result from the dynamic risk measure literature (Kupper and Schachermayer, 2009, Theorem 1.10); the substantive content of the theorem lies elsewhere, in the consequences developed in Section 5 for the AS-CJ split and for the regulatory CVaR framework.

5 Consequences of the Theorem

Eight consequences of forced uniqueness are developed as formal corollaries of Theorem 10, together with two further structural consequences recorded as remarks (Remarks 20 and 28). Unless otherwise stated, each corollary is read under the hypotheses of the theorem, namely

Axioms J1–J5. The two corollaries on mean-variance and CVaR (Corollaries 29 and 31) are exceptions, in that they are statements about *candidate alternative functionals* and require only the axiom statements, not the theorem's conclusion.

5.1 Single-Parameter Reduction

The first consequence of Theorem 10 is that the entire space of preference functionals satisfying our axioms is one-dimensional, parametrized by a single positive scalar living on an open half-line whose two endpoints admit natural limiting interpretations.

Corollary 17 (Single-Parameter Reduction)**.** *Under the same hypotheses as in Theorem 10, the market maker's dynamic preference functional J is determined, given the canonical clock $\Lambda = \langle S \rangle$ and the liquidation cost function L, by a single free scalar parameter $\gamma \in (0, \infty)$, with the entropic functional (3) at parameter γ as the unique representative. The two endpoints of the half-line have the following limiting interpretations.*

1. *As $\gamma \downarrow 0$, the entropic functional converges to the risk-neutral functional,*

$$J_t(\pi) \longrightarrow \mathbb{E}\big[W_T^L(\pi) \,\big|\, \mathcal{F}_t\big].$$

2. *As $\gamma \uparrow \infty$, the entropic functional converges to the pessimistic worst-case functional,*

$$J_t(\pi) \longrightarrow \operatorname{ess\,inf}\big[W_T^L(\pi) \,\big|\, \mathcal{F}_t\big].$$

Every preference functional satisfying the axiom system lies strictly between these two extremes. The strict positivity and finiteness of γ are both enforced by the strict-concavity clause (J3b) of Axiom J3.

Proof. The single-parameter statement is immediate from the explicit form (3), since the only free quantity on the right-hand side, beyond the data $(\langle S \rangle, L)$ which are properties of the market and not of the agent, is the scalar γ. For the first limit, a Taylor expansion of the exponential gives $-\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t] = \mathbb{E}[W \mid \mathcal{F}_t] - \frac{\gamma}{2}\operatorname{Var}(W \mid \mathcal{F}_t) + O(\gamma^2)$ as $\gamma \to 0$, so the limit is $\mathbb{E}[W \mid \mathcal{F}_t]$. For the second limit, Kupper and Schachermayer (2009, Lemma 2.1(i)) establishes the unconditional fact $\lim_{\gamma\to\infty} \rho_\gamma(X) = \operatorname{ess\,sup}(-X)$ for $X \in L^\infty$, where $\rho_\gamma(X) = \frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma X}]$ is the static entropic risk measure. To obtain the conditional extension $\lim_{\gamma\to\infty} \frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t] = -\operatorname{ess\,inf}(W \mid \mathcal{F}_t)$ for $W \in L^\infty(\mathcal{F}_T)$, the same argument goes through ω-by-ω when applied to the regular conditional distribution of W given $\mathcal{F}_t$ (which exists and is supported in $[-\|W\|_\infty, \|W\|_\infty]$); equivalently, the conditional version is encoded in the dynamic representation of Kupper and Schachermayer (2009, Theorem 1.10, equations (8)–(9)). Both boundary cases are ruled out by the strict-concavity clause (J3b) of Axiom J3. At $\gamma = 0$, the linear functional $\mathbb{E}[\cdot]$ gives equality on every pair π, π' with $W_T^L(\pi) - W_T^L(\pi')$ not a deterministic constant, contradicting (J3b) on such pairs. At $\gamma = \infty$, the worst-case functional $J_0(\pi) = \operatorname{ess\,inf} W_T^L(\pi)$ gives equality on any pair π_0, π_1 with $W_T^L(\pi_1) = W_T^L(\pi_0) - \varepsilon \mathbf{1}_A$ whenever A does not lower the essential infimum. In this case, $W_T^L(\pi_0) - W_T^L(\pi_1)$ is not a deterministic constant yet J_0 gives equality, contradicting (J3b). Equivalently, by Lemma 5, the strict-concavity clause (J3b) together with monotonicity (M) implies relevance, which fails at $\gamma = \infty$. □

Corollary 17 locates the market maker's preferences in a one-dimensional family bounded by two well-understood limits, namely the risk-neutral limit ($\gamma \to 0$), in which she maximizes expected wealth, and the worst-case limit ($\gamma \to \infty$), in which she maximizes the guaranteed outcome. The entire space of admissible preferences interpolates continuously between these

two extremes, with larger γ corresponding to heavier weighting of bad outcomes. The naive intuition that a more risk-averse market maker quotes wider is correct, and the corollary is its rigorous form, in that her risk-aversion sits on a half-line and her position on that half-line determines the entire dynamic preference functional she optimizes, with no second risk-aversion parameter, no separate skewness-aversion parameter, and no path-dependent or state-dependent component. The single scalar γ does all the work.

Cartea–Jaimungal-style models typically parametrize the market maker's preferences by at least three numbers, namely the running inventory penalty coefficient ϕ, the terminal inventory penalty coefficient α, and (when explicit) a risk-aversion parameter. Corollary 17 says, in effect, that two of these are not free, by way of two distinct mechanisms. The running coefficient ϕ is forced to be a specific function of the preference parameter γ and the realized volatility, computed explicitly in Section 6. The terminal coefficient α is forced to be the leading coefficient of the liquidation cost function L, a primitive of the market rather than a preference parameter. The genuine preference degree of freedom is the single scalar γ.

A practitioner used to the CJ parametrization might find independent calibration of ϕ and α empirically convenient. Our framework does not endorse this, in that any combination of (ϕ, α, γ) outside the two-equation constraint surface

$$\begin{cases} \phi = \gamma\sigma^2/2, \\ \alpha = \frac{1}{2}L''(0) \end{cases}$$

violates at least one of Axioms J1–J5. The statement is about agents with a well-defined preference functional satisfying Axioms J1–J5; a CJ-style desk may not be maximizing any preference functional at all, instead using the CJ-formula as a heuristic, and the forcing applies to the class of agents who do have such a preference functional, telling those agents that once γ is chosen, ϕ and α are not independent free parameters.

Remark 18 (Constancy of γ)**.** A direct consequence of the explicit form (3) is that γ is a single positive scalar, and does not depend on time t, on the state of the world ω, on the price level, on realized volatility, on the time-of-day, or on the intraday P&L. What *can* legitimately vary across the trading day is the inventory penalty rate $\frac{\gamma}{2}\sigma_t^2 q^2$, which naturally tracks realized volatility through σ_t^2; the underlying preference parameter γ itself does not move. The constancy result does not preclude γ from being a function of slow-moving structural quantities such as asset class, desk capital, or firm risk policy, only from being a function of intraday state; a market maker whose subjective γ does vary intraday is free to do so, but she is then outside the axiom system. The stronger cross-asset statement is the content of the multi-asset extension (Theorem 48 in Appendix C; see also Section 7.4).

5.2 AS Uniqueness

The second consequence identifies the Avellaneda–Stoikov framework as the unique axiom-consistent model in the constant-volatility benchmark.

Corollary 19 (Avellaneda–Stoikov Uniqueness)**.** *Suppose that the price process S has constant volatility $\sigma > 0$, i.e., $d\langle S\rangle_t = \sigma^2\,dt$. Then, under the same hypotheses as Theorem 10, the market maker's preferences are uniquely given by the entropic certainty-equivalent functional*

$$J_t(\pi) = -\frac{1}{\gamma}\log\mathbb{E}\Big[\exp\big(-\gamma\, W_T^L(\pi)\big)\,\Big|\,\mathcal{F}_t\Big], \tag{8}$$

which by Remark 11 coincides, for the CARA utility $u(x) = -\exp(-\gamma x)$, with the conditional certainty-equivalent of expected utility under u. Under the further specializations of Avellaneda

and Stoikov (2008), with exponential arrival intensities $\lambda^a(\delta) = Ae^{-\kappa\delta}$ and vanishing terminal liquidation cost $L \equiv 0$, the optimal quoting strategy derived from (8) coincides with the strategy in their paper.

Proof. The functional (8) is the constant-volatility specialization of (3). Equivalence with the AS strategy under their further specializations is by standard calculation; see Avellaneda and Stoikov (2008) or Cartea et al. (2015, Section 10.3). The forcing in Corollary 19 is therefore not a new optimization result but a statement about which preference functional an analyst must use to obtain the AS strategy as an optimum. □

The CARA utility in the AS paper is introduced with two motivations, namely tractability and the wealth-independence of the optimal δ^a, δ^b. Corollary 19 changes the picture entirely. Given Axioms J1–J5, CARA is not a modeling choice; it is in fact a theorem. The wealth-independence property, often cited as a reason to use CARA, is a direct consequence of cash-additivity (J1) and does not require the CARA commitment, as recorded in Remark 20 below.

The mathematical content of AS is therefore not expected-utility maximization but dynamic-risk-measure minimization on liquidation-adjusted terminal wealth; the CARA-expected-utility framing is an artifact of the certainty-equivalent identity (Remark 11) that holds for CARA and for no other utility. The standard AS calculation could be presented, without any change in conclusions, as the explicit computation of the optimal strategy for a dynamic entropic risk measure, with no reference to expected utility at all.

A related consequence is that the standard portfolio-theory utility hierarchy is incompatible with our framework: CRRA, HARA, and the rest of the Merton menagerie all fail one or more of Axioms J1–J5, with cash-additivity (J1) the most direct failure (under power utility, the marginal utility of a sure dollar depends on current wealth, contradicting $J_t(\pi + c) = J_t(\pi) + c$ for $\mathcal{F}_t$-measurable bounded c). Market making and portfolio choice, which share much of their mathematical apparatus, require structurally different preference primitives; the portfolio-theory tradition treats CARA as a special case, while ours forces CARA as the only case.

Remark 20 (Wealth-Independence)**.** The optimal quote distances chosen by the market maker do not depend on her current cash holdings. Indeed, writing $\pi + c$ for the strategy whose terminal liquidation-adjusted wealth is shifted by c in the sense of the cash-injection construction of Lemma 46, cash-additivity (Axiom J1) gives $J_t(\pi + c) = J_t(\pi) + c$ for any $\mathcal{F}_t$-measurable bounded c, so taking $c = X_t$ (assumed bounded), the dependence of J_t on X_t is an additive shift and does not affect the ranking of strategies; the optimal quote distances $\delta^{a,*}, \delta^{b,*}$ are therefore functions of (t, S_t, q_t) and the relevant state of the world, but not of X_t. Wealth-independence follows from cash-additivity alone and does not on its own motivate the CARA form, which is forced by the full axiom system, as established in Theorem 10. The AS rationale for CARA thus combines two ingredients, a normative desideratum (wealth-independent quotes) and a heuristic implication (wealth-independence implies CARA), and neither is needed under our axioms; cash-additivity delivers wealth-independence directly, and the CARA form then arises from the deeper uniqueness theorem, for which wealth-independence reasoning plays no role.

5.3 CJ as Forced Second-Order Approximation

The Cartea–Jaimungal framework models the market maker as maximizing

$$\mathbb{E}\big[W_T^L(\pi)\big] - \phi\,\mathbb{E}\left[\int_0^T q_s^2\,ds\right] \tag{9}$$

over admissible strategies, with constant $\phi > 0$ a running inventory penalty. The terminal liquidation term, when present, is absorbed into our L, as discussed in Section 2.

Cartea et al. (2015, p. 261) record, at the value-function level and in the constant-volatility setting, a parameter identification of the form $\phi = \gamma\sigma^2/2$ under which the AS and CJ objectives yield the same optimal quoting strategy up to a constant shift. The corollary below states a different and structurally stronger claim, at the preference-functional level, in that the CJ objective (9) does not arise as a primitive preference functional within the axiom system, and appears only as the second-order Taylor expansion of the axiomatically-determined entropic functional around zero inventory, with the same coefficient $\phi = \gamma\sigma^2/2$.

Corollary 21 (CJ as Forced Second-Order Approximation)**.** *Under the same hypotheses as Theorem 10, the Cartea–Jaimungal objective (9) does not arise as a primitive preference functional within the axiom system. It is, however, the second-order expansion of the entropic functional (3) around zero inventory in the constant-volatility setting, provided that the running coefficient takes the value*

$$\phi = \frac{\gamma\sigma^2}{2}. \tag{10}$$

Any other value of ϕ produces a CJ objective that is not the expansion of any functional satisfying Axioms J1–J5.

Proof. For the first statement, the *primary* failure is path-dependence, in that the running-penalty term $\int_0^T q_s^2(\pi)\,ds$ in (9) is a path functional of the inventory process q, which can take different values for two strategies that produce the same distribution of W_T^L, so (9) violates property (W) (Proposition 7). Even if one ignores the running term, the linear expected-wealth component $\mathbb{E}[W_T^L]$ is the risk-neutral limit excluded by the strict-concavity clause (J3b), an independent *secondary* failure. (To be precise: because the full CJ objective violates (W) via its path-dependent running term, it is not a functional on W_T^L at all, so Axiom J3 does not directly apply to it, and the secondary failure is a statement about the terminal-wealth component $\mathbb{E}[W_T^L]$ considered in isolation.) Either failure alone is sufficient.

For the second statement, the full calculation is given in Section 6 as the proof of Proposition 32; the headline result is that the certainty-equivalent contribution of holding inventory q at time s, accumulated over an infinitesimal interval ds, equals $-\frac{\gamma\sigma^2}{2}q^2\,ds$, with the coefficient $\gamma\sigma^2/2$ arising *exactly* from the HJB diffusion term (no approximation in the coefficient itself; the "second-order" qualifier refers to the inventory expansion of the full functional, not to the coefficient). This is exactly the running-penalty contribution in (9) with coefficient $\phi = \gamma\sigma^2/2$, which establishes both directions of the second statement. See Remark 33 below for the distinction made precise. □

A working market-making desk that uses the CJ framework as a primitive objective is, on our axiomatic terms, using a functional whose path-dependence (through the inventory penalty $\int q^2\,ds$) violates the wealth-summary requirement (W) (Proposition 7). If, however, the desk wishes to interpret CJ as a tractable second-order approximation to entropic preferences, it is welcome to do so, but it must use $\phi = \gamma\sigma^2/2$; any other value corresponds to a CJ objective that cannot be obtained as an expansion of any entropic functional and is therefore inconsistent with the CARA preferences which the second-order interpretation is supposed to encode. At $\phi = \gamma\sigma^2/2$, the CJ-formula strategy and the AS-optimal strategy produce the same reservation price and half-spread formulas, so they coincide exactly in the constant-volatility benchmark.

Remark 22 (Relation to Prior Value-Function Statements)**.** The forced coefficient (10) is visible, though not stated as a parameter constraint, in the AS value function of Avellaneda

and Stoikov (2008), namely

$$V(t, X, q, S) = -\exp\big(-\gamma\big(X + qS - \tfrac{1}{2}\gamma q^2\sigma^2(T-t)\big)\big)\,,$$

in which the certainty-equivalent reduction $-\frac{1}{2}\gamma q^2\sigma^2(T-t)$ is the integrated running cost at rate $(\gamma\sigma^2/2)\,q^2$, exactly equal to ϕq^2 with the value (10). The identification is stated as a parameter setting in Cartea et al. (2015, p. 261), under which the AS and CJ value-function ODEs coincide in structure and the two optimal quoting strategies agree up to a constant shift. Both prior statements operate at the value-function level and in the constant-volatility setting; the present corollary makes a different structural claim, deriving the identification from the axiomatic characterization at the preference-functional level (see Section 7.3 for the full comparison).

5.4 Forced Terminal-Penalty Coefficient

The CJ tradition also features a terminal-inventory penalty αq_T^2, usually treated as an independent hyperparameter alongside ϕ. A parallel forced-coefficient logic, analogous to that of Corollary 21, extends to α.

Corollary 23 (Forced Terminal-Penalty Coefficient)**.** *Under the hypotheses of Theorem 10, suppose in addition that the liquidation cost function admits a second-order Taylor expansion at zero inventory: $L(0) = 0$, $L'(0) = 0$, and $L(q) = \frac{1}{2}L''(0)\,q^2 + o(q^2)$ as $q \to 0$. (The condition $L'(0) = 0$ is the natural requirement that the liquidation cost be minimized at zero inventory.) Then the CJ terminal-penalty coefficient α is forced to take the value*

$$\alpha = \tfrac{1}{2}\,L''(0).$$

Proof. By property (W) (Proposition 7), we have that $J_t(\pi)$ depends on π only through $W_T^L(\pi) = X_T + q_T S_T - L(q_T)$. The terminal-inventory contribution to W_T^L is the term $-L(q_T)$. By the assumed Taylor expansion of L at zero, and using the hypothesis $L'(0) = 0$ to eliminate the linear term, the leading nonzero contribution near $q_T = 0$ is equal to $-L(q_T) = -\frac{1}{2}L''(0)\,q_T^2 + o(q_T^2)$. This is precisely the $-\alpha q_T^2$ contribution that the CJ objective adds, with $\alpha = \frac{1}{2}L''(0)$.

For the rigidity, consider a market maker who specifies $-\alpha q_T^2$ as a terminal penalty with $\alpha \neq \frac{1}{2}L''(0)$, taking α as a free preference parameter. Since Proposition 7 fixes the terminal-wealth variable as $W_T^L = X_T + q_T S_T - L(q_T)$ with L given as market data, any quadratic terminal-inventory coefficient other than $\frac{1}{2}L''(0)$ represents a preference functional depending on something other than W_T^L, and the two functionals disagree on any pair $\pi, \pi' \in \Pi$ with $W_T^L(\pi) = W_T^L(\pi')$ but $q_T^2(\pi) \neq q_T^2(\pi')$. Such pairs exist whenever q_T is not almost surely zero (the generic case), as follows. Fix $\pi_a, \pi_b \in \Pi$ with $q_T(\pi_a) \neq q_T(\pi_b)$ on a positive-probability event, constructed by varying the quote-distance processes asymmetrically between bid and ask, and let $\Delta W \coloneqq W_T^L(\pi_a) - W_T^L(\pi_b) \in L^\infty(\mathcal{F}_T)$. By Lemma 46, the cash-shift construction of Section 2.2 produces $\pi_b^{[\Delta W]} \in \Pi$ obtained from π_b by injecting the bounded $\mathcal{F}_T$-measurable cash amount ΔW at time T, so that $W_T^L(\pi_b^{[\Delta W]}) = W_T^L(\pi_a)$ almost surely while $q_T(\pi_b^{[\Delta W]}) = q_T(\pi_b)$ (the cash injection does not affect inventory). Setting $\pi \coloneqq \pi_a$ and $\pi' \coloneqq \pi_b^{[\Delta W]}$ yields the desired pair, and feasibility within Π is by the admissibility verification of Lemma 46. □

Remark 24 (Relation to Prior Work on Terminal-Penalty Coefficients)**.** The identification of the terminal-inventory coefficient with a market-impact parameter is implicit in the optimal-execution literature. For instance, Barger and Lorig (2018, Section 2.3) use a terminal penalty κq_T^2 alongside a running penalty and treat κ as a temporary price-impact parameter rather than a preference parameter. Corollary 23 sharpens this informal reading into a formal forcing

statement derived from the wealth-summary property (W). Indeed, once the market maker's preferences are axiomatically committed to depend on the strategy only through W_T^L, the quadratic-order dependence on q_T is fixed by $L''(0)$ and cannot be a free preference parameter.

The corollary makes precise an observation that is implicit but not stated in the AS–CJ literature. The coefficient α is not a preference parameter at all but a property of the *market* (specifically, the curvature of the cost-to-liquidate function at zero inventory). A market with deeper liquidity has a smaller $L''(0)$ and hence a smaller α; in the quadratic-impact benchmark $L(q) = \frac{\kappa}{2}q^2$, one finds $\alpha = \frac{\kappa}{2}$.

Figure 1 makes the geometry explicit. The Cartea–Jaimungal terminal penalty αq^2 is the osculating parabola of the true liquidation cost L at zero inventory, agreeing with L in value, slope, and curvature there, which is exactly the content of $\alpha = \frac{1}{2}L''(0)$. For larger inventory the two separate, since a convex L with super-quadratic market impact rises above its second-order approximation, so the quadratic penalty under-penalizes large terminal inventory by the shortfall $L(q) - \frac{1}{2}L''(0)q^2$. The figure also isolates the precise sense in which the Cartea–Jaimungal objective approximates the entropic one. The running coefficient $\phi = \gamma\sigma^2/2$ is exact in the constant-volatility benchmark, with no approximation in the coefficient itself (Remark 33), while the terminal liquidation cost shown here is the sole place at which a genuine second-order approximation enters.

The forcing of α is different in kind from the forcing of ϕ. The forcing $\phi = \gamma\sigma^2/2$ is a consequence of the entropic structure of the preference functional, arising from the second-order Taylor expansion of $-\frac{1}{\gamma}\log\mathbb{E}[e^{-\gamma W_T^L} \mid \mathcal{F}_t]$. The forcing $\alpha = \frac{1}{2}L''(0)$ is a consequence of the market structure: it is the Taylor coefficient of L at zero, and would hold under any preference functional satisfying Proposition 7. Combining Corollaries 21 and 23, the full CJ-style parametrization (ϕ, α, γ) collapses to a single free preference parameter γ, with ϕ a function of γ and market data and α purely market data. This is the sharp form of the single-parameter reduction of Corollary 17.

5.5 Calibration Inversion: Recovering γ from ϕ

Corollary 21 expresses the forced relation $\phi = \gamma\sigma^2/2$ as a constraint surface in (ϕ, γ, σ)-space. The constraint can be read in either direction, and the reverse direction is operationally important.

Corollary 25 (Calibration Inversion)**.** *Under the same hypotheses as Theorem 10 and in the constant-volatility setting, a desk that has calibrated the CJ running-penalty coefficient ϕ from inventory-path and P&L data implicitly determines the risk-aversion parameter γ via*

$$\gamma = \frac{2\phi}{\sigma^2}. \tag{11}$$

With stochastic volatility, the same holds pointwise in time: if a desk has calibrated a time-varying local penalty rate ϕ_t from the data, then by Corollary 12 (clock-invariance) the inversion $\gamma = 2\phi_t/\sigma_t^2$ must yield the same scalar γ at every time t, with $\phi_t = \gamma\sigma_t^2/2$ the clock-coherent local rate of Proposition 36.

Proof. Immediate from the forced relation (10). The stochastic-volatility version follows from Proposition 36, which establishes that the running penalty rate must be $\gamma\sigma_t^2/2$ with a single γ across t; the inversion is the same algebraic relation read pointwise. □

The corollary has two practical readings.

The first is methodological. A working market-making desk typically has two independent sources of preference-parameter information, namely the quoting side (where γ can be estimated

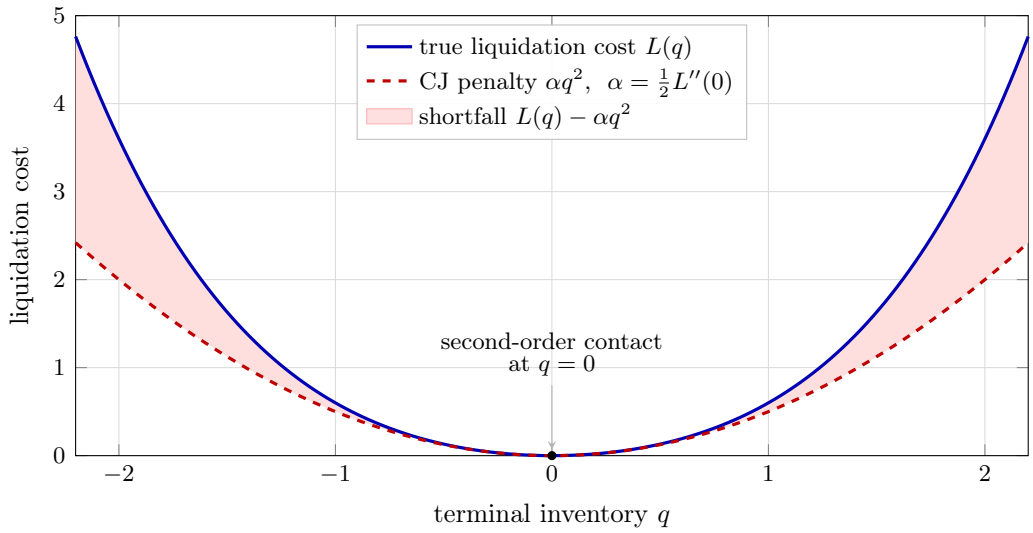


Figure 1: The Cartea–Jaimungal terminal penalty as the second-order image of the liquidation cost. The solid curve is the true liquidation cost $L(q)$, the cash lost to slippage when unwinding a position of size q, drawn here as the illustrative convex form $L(q) = \frac{1}{2}\kappa q^2 + \frac{1}{4}c\,q^4$ with $\kappa = L''(0)$ and $c > 0$. The dashed curve is the Cartea–Jaimungal terminal penalty, given by αq^2. By Corollary 23, the axioms let terminal inventory affect the objective only through this real cost $-L(q_T)$, so a freely chosen α is inadmissible, and the only admissible value is the second-order coefficient of L at zero inventory, $\alpha = \frac{1}{2}L''(0)$. Geometrically the forced parabola makes second-order contact with L at $q = 0$, matching it in value, slope, and curvature there. For larger inventory L rises above the parabola, since the marginal cost of unwinding grows with size ($c > 0$, super-quadratic), so the quadratic penalty under-charges the unwinding of large positions by the shaded shortfall $L(q) - \alpha q^2 \geq 0$. The coefficient α is therefore a property of the market, the curvature of its liquidation cost at zero inventory, and not a preference of the trader. This is the terminal counterpart of the running relation $\phi = \gamma\sigma^2/2$, which is exact in the constant-volatility benchmark (Remark 33), with the terminal cost the sole place at which a genuine second-order approximation enters.

from the shape and time-variation of optimal quote distances, since the AS reservation-price and optimal-spread formulas depend on γ explicitly; see, e.g., Cartea et al. (2015, Section 10.3)) and the inventory side (where ϕ can be calibrated from the historical relationship between inventory paths and realized P&L, via standard CJ-style HJB calibrations). Corollary 25 says these two estimators target the *same scalar*, up to the deterministic rescaling $2/\sigma^2$, and computing both is a sanity check on whether the desk is operating on the constraint surface. If the quoting-side γ and the inventory-side $2\phi/\sigma^2$ disagree by more than estimation error, the desk's quoting behavior and inventory management are inconsistent with any single dynamically-consistent preference functional, and a desk that has been calibrating ϕ for years in the CJ tradition has been operating with an implicit γ all along, whether or not its risk officers were aware of it.

The second reading concerns cross-regime stability. Equation (11) implies that the implied γ should be stable across volatility regimes, since it is a property of the agent's preferences and not of the realized volatility path (Corollary 12). A desk's calibrated ϕ should scale with σ^2 across regimes; if it is approximately constant (the wall-clock implementation), the implied γ varies as σ^{-2}, a structural misspecification and the failure mode identified by Corollary 12.

5.6 Robust-Optimization Interpretation

The entropic functional admits a classical dual representation as a robust optimization over alternative probability measures with a relative-entropy penalty (Föllmer and Schied, 2016, Delbaen et al., 2010). In our market-making context, the dual has substantive content beyond the static one.

Corollary 26 (Robust-Optimization Interpretation)**.** *Under the hypotheses of Theorem 10, the forced preference functional admits the dual representation*

$$J_t(\pi) = \operatorname*{ess\,inf}_{\mathbb{Q}\ll\mathbb{P}} \left\{ \mathbb{E}_{\mathbb{Q}}\big[W_T^L(\pi) \,\big|\, \mathcal{F}_t\big] + \frac{1}{\gamma} H_t(\mathbb{Q} \,\|\, \mathbb{P}) \right\}, \tag{12}$$

where the essential infimum runs over probability measures $\mathbb{Q}$ absolutely continuous with respect to $\mathbb{P}$, and

$$H_t(\mathbb{Q} \,\|\, \mathbb{P}) := \mathbb{E}_{\mathbb{Q}}\left[\log \tfrac{d\mathbb{Q}}{d\mathbb{P}} \,\middle|\, \mathcal{F}_t\right]$$

is the conditional relative entropy of $\mathbb{Q}$ with respect to $\mathbb{P}$ given $\mathcal{F}_t$. The unique minimizing measure $\mathbb{Q}^*$ in (12) has $\mathcal{F}_T$-measurable density (with $\mathcal{F}_t$-measurable normalizing constant)

$$\frac{d\mathbb{Q}^*}{d\mathbb{P}} = \frac{e^{-\gamma\, W_T^L(\pi)}}{\mathbb{E}[e^{-\gamma\, W_T^L(\pi)} \mid \mathcal{F}_t]}.$$

At $t = 0$, the σ-algebra $\mathcal{F}_0$ is trivial and the density reduces to $d\mathbb{Q}^*/d\mathbb{P} = e^{-\gamma W_T^L(\pi)}/\mathbb{E}[e^{-\gamma W_T^L(\pi)}]$, the static entropic-tilted measure.

Proof. The representation is an application of the conditional Donsker–Varadhan variational identity. Indeed, we have

$$-\tfrac{1}{\gamma}\log \mathbb{E}\big[\exp(-\gamma W) \,\big|\, \mathcal{F}_t\big] = \operatorname*{ess\,inf}_{\mathbb{Q}\ll\mathbb{P}} \left\{ \mathbb{E}_{\mathbb{Q}}[W \mid \mathcal{F}_t] + \tfrac{1}{\gamma} H_t(\mathbb{Q} \,\|\, \mathbb{P}) \right\},$$

applied with $W = W_T^L(\pi)$. The conditional minimizer is identified by the conditional first-order condition, namely that $d\mathbb{Q}^*/d\mathbb{P}$ is proportional to $e^{-\gamma W}$ with the proportionality factor $1/\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t]$ chosen to make $\mathbb{Q}^*$ a probability measure conditionally on $\mathcal{F}_t$. See Föllmer and Schied (2016, Proposition 4.33) for the conditional version. □

Corollary 26 has a sharp and economically meaningful interpretation: the forced preference functional is mathematically identical to a robust optimization in which the market maker entertains all $\mathbb{Q} \ll \mathbb{P}$, evaluates each strategy with a conditional-relative-entropy penalty at rate $1/\gamma$, and takes the essential infimum, with $\mathbb{Q}^* \neq \mathbb{P}$ corresponding to her implicit pessimistic measure for the strategy under evaluation. Note that $\mathbb{Q}^*$ depends on π through $W = W_T^L(\pi)$, so the dual is a strategy-by-strategy max-min restatement of the forced functional rather than an adversarial game against a fixed Knightian prior. This places the paper in conversation with the variational-preferences literature of Maccheroni et al. (2006) and resolves the apparent tension with the single-reference-measure commitment of law-invariance (Remark 3).

Remark 27 (On the Relation to Φ-Divergence Risk Measures)**.** The relative-entropy penalty in Corollary 26 is the Kullback-Leibler divergence, one member of the broader family of Φ-divergences (Rényi, Tsallis, χ^2, and others), each of which would generate a different robust-optimization dual and a different family of risk measures (the so-called Φ- or f-divergence risk measures; see Föllmer and Schied (2016, Chapter 4)). Our forced-uniqueness theorem identifies the conditional entropic functional, which by Corollary 26 corresponds to a robust optimization with KL as ambiguity penalty, as the unique preference functional consistent

with the five axioms; a market maker who satisfies our axioms therefore selects KL, rather than some other Φ-divergence, as the implicit ambiguity penalty on her beliefs. We do not prove the converse axiomatic characterization at the level of Φ-divergences themselves (that KL is the unique Φ-divergence whose conjugate dual produces a functional satisfying our axiom system), which would require a parallel analysis at the level of divergence functionals and is an interesting open question.

Remark 28 (Position-Size Nonlinearity)**.** A coherent risk measure in the sense of Artzner et al. (1999) is positively homogeneous, in that doubling all positions doubles the risk; the entropic functional is convex and monetary but not positively homogeneous. A direct calculation gives the rescaling identity $\rho_t(\lambda W) = \lambda\, \rho_t^{(\lambda\gamma)}(W)$ for any $\lambda > 0$, where $\rho_t^{(\lambda\gamma)}$ denotes the entropic risk measure at parameter $\lambda\gamma$; combined with the strict monotonicity in γ' of $f_t(\gamma') := -\frac{1}{\gamma'} \log \mathbb{E}[e^{-\gamma' W} \mid \mathcal{F}_t]$ on any set where W is conditionally nondegenerate given $\mathcal{F}_t$ (Sub-step 3b of the proof of Theorem 10), this yields $\rho_t(\lambda W) \neq \lambda \rho_t(W)$ with positive probability whenever $\lambda \neq 1$. Operationally, a market maker who scales every position in her book by a factor λ does not face λ times the risk but a risk evaluated under a different (and larger, if $\lambda > 1$) effective risk-aversion parameter $\lambda\gamma$, so the forced functional penalizes large positions super-linearly in size. The failure has a clean reading through the robust dual of Corollary 26, in that doubling her positions does not double her relative-entropy ball, but the worst-case loss inside that fixed ball grows super-linearly because $W \mapsto \mathbb{E}_{\mathbb{Q}}[W]$ becomes more sensitive to tail-tilts of $\mathbb{Q}$ when W is scaled up.

5.7 Mean-Variance Incompatibility

The mean-variance objective, in the tradition of Markowitz (1952), is widely used in practitioner finance as a tractable proxy for risk-averse expected-utility preferences. For a market maker, the natural mean-variance objective would be of the form

$$\mathbb{E}[W_T^L(\pi)] - \tfrac{\lambda}{2} \operatorname{Var}\big(W_T^L(\pi)\big) \tag{13}$$

for some $\lambda > 0$, using the variance rather than the cumulant-generating function as the risk-penalty. The failure of monotonicity and dynamic consistency for the mean-variance functional is well-known; see Basak and Chabakauri (2010) and Björk and Murgoci (2014). What is specific to the present setting is the corollary status of these failures within the axiom system (Axioms J1–J5), and their implication that no mean-variance objective, however calibrated, arises within the axiom-consistent class; the proof below records the two failures for the market-making formulation.

Corollary 29 (Mean-Variance Incompatibility)**.** *Let the mean-variance preference functional be the dynamic functional defined by*

$$J_t^{\mathrm{MV}}(W) := \mathbb{E}[W \mid \mathcal{F}_t] - \tfrac{\lambda}{2} \operatorname{Var}(W \mid \mathcal{F}_t),$$

for $W \in L^\infty(\mathcal{F}_T)$ and $\lambda > 0$, which is the naive-dynamic form of (13), obtained by conditioning the static functional at each time. The pre-committed formulation of Basak and Chabakauri (2010) is a distinct construct that sidesteps the time-inconsistency by fixing the objective at time zero, and its failure modes are not the ones addressed here. The functional in fact violates both monotonicity (M, Proposition 6) and strong dynamic consistency (Axiom J4). Consequently, mean-variance preferences are incompatible with our axiom system.

Proof. Failure 1: Monotonicity (Proposition 6). Consider $W, W' \in L^\infty(\mathcal{F}_T)$ with $W \geq W'$ almost surely. We exhibit an explicit counterexample to $J_t^{\mathrm{MV}}(W) \geq J_t^{\mathrm{MV}}(W')$. Fix $t = 0$ and let $W' = 0$ (the zero random variable), so $J_0^{\mathrm{MV}}(W') = 0$. Since λ is a fixed preference

parameter, choose any nonconstant random variable ξ with $\xi \geq 0$ almost surely, $\mathbb{E}[\xi] = 1$, and $\mathrm{Var}(\xi) = v > 2/\lambda$. Such a ξ exists because the variance of nonnegative random variables with mean 1 is unbounded above. Take $W = \xi$. Then $W \geq W'$ almost surely, but

$$J_0^{\mathrm{MV}}(W) = 1 - \tfrac{\lambda v}{2} < 0 = J_0^{\mathrm{MV}}(W'),$$

in violation of monotonicity.

Failure 2: Strong dynamic consistency (Axiom J4). By the law of total variance,

$$\mathrm{Var}(W \mid \mathcal{F}_s) = \mathbb{E}[\mathrm{Var}(W \mid \mathcal{F}_t) \mid \mathcal{F}_s] + \mathrm{Var}(\mathbb{E}[W \mid \mathcal{F}_t] \mid \mathcal{F}_s),$$

the cross-term $\mathrm{Var}(\mathbb{E}[W \mid \mathcal{F}_t] \mid \mathcal{F}_s) - \mathrm{Var}(\mathbb{E}[W' \mid \mathcal{F}_t] \mid \mathcal{F}_s)$ has indefinite sign and is not controlled by the time-t ranking of W, W'. We exhibit a concrete two-period counterexample. Let $s = 0$, $t \in (0, T)$, $\lambda = 1$, with $\mathcal{F}_s$ trivial and $\mathcal{F}_t$ generated by a fair coin flip yielding atoms ω_+, ω_- each with probability $1/2$. Refine $\mathcal{F}_T$ so that on ω_+ there are two equiprobable sub-atoms where W takes the values $+1$ and -1, while on ω_- the random variable W is constant equal to 5. Set $W' \equiv -1/2$ deterministically. (Existence of such a refinement on $(\Omega, \mathcal{F}_T, \mathbb{P})$ is by the non-atomicity of the probability space, as in Sub-step 3b of the proof of Theorem 10.)

At time t, the conditional moments are: on ω_+, $\mathbb{E}[W \mid \mathcal{F}_t] = 0$ and $\mathrm{Var}(W \mid \mathcal{F}_t) = 1$, so $J_t^{\mathrm{MV}}(W)(\omega_+) = 0 - 1/2 = -1/2$; on ω_-, $\mathbb{E}[W \mid \mathcal{F}_t] = 5$ and $\mathrm{Var}(W \mid \mathcal{F}_t) = 0$, so $J_t^{\mathrm{MV}}(W)(\omega_-) = 5$. Meanwhile $J_t^{\mathrm{MV}}(W') \equiv -1/2$ identically. Hence, $J_t^{\mathrm{MV}}(W) \geq J_t^{\mathrm{MV}}(W')$ almost surely, with equality on ω_+ and strict inequality on ω_-.

At time $s = 0$, we have $\mathbb{E}[W] = \frac{1}{2} \cdot 0 + \frac{1}{2} \cdot 5 = 5/2$ and $\mathbb{E}[W^2] = \frac{1}{2} \cdot 1 + \frac{1}{2} \cdot 25 = 13$, so $\mathrm{Var}(W) = 13 - 25/4 = 27/4$. Therefore,

$$J_s^{\mathrm{MV}}(W) = \tfrac{5}{2} - \tfrac{1}{2} \cdot \tfrac{27}{4} = \tfrac{20}{8} - \tfrac{27}{8} = -\tfrac{7}{8},$$

while $J_s^{\mathrm{MV}}(W') = -1/2 = -4/8$. Thus, $J_s^{\mathrm{MV}}(W) = -7/8 < -4/8 = J_s^{\mathrm{MV}}(W')$, contradicting the conclusion of (J4). Fully developed counterexamples of this form are given by Basak and Chabakauri (2010) and Björk and Murgoci (2014). □

Remark 30 (Cash-Additivity and Wealth-Summary Are Not the Issue)**.** The mean-variance functional satisfies both cash-additivity (Axiom J1) and the wealth-summary property (W) (Proposition 7), which is worth stating since both are nonobvious. For cash-additivity, an $\mathcal{F}_t$-measurable bounded c acts as a constant in the conditional distribution of W given $\mathcal{F}_t$, so that $\mathrm{Var}(W + c \mid \mathcal{F}_t) = \mathrm{Var}(W \mid \mathcal{F}_t)$ and therefore $J_t^{\mathrm{MV}}(W + c) = J_t^{\mathrm{MV}}(W) + c$. For (W), since $J_t^{\mathrm{MV}}(W) = \mathbb{E}[W \mid \mathcal{F}_t] - \frac{\lambda}{2}\mathrm{Var}(W \mid \mathcal{F}_t)$ is a function of $W = W_T^L$ alone, the condition "$W_T^L(\pi) = W_T^L(\pi')$ almost surely implies $J_t(\pi) = J_t(\pi')$ almost surely" is immediate. The incompatibility with our axiom system does not come from monetary structure or path-dependence, but from the variance term's failure to respect both monotonicity and time-consistency.

5.8 CVaR Critique

The Basel Committee (Basel Committee on Banking Supervision, 2019) adopted expected shortfall as the official market-risk measure for dealer desks, so the market maker on a regulated desk faces a CVaR-based external risk constraint. The time-inconsistency of dynamic conditional-value-at-risk is a known result in the dynamic risk measure literature; see Artzner et al. (2007) and Cheridito and Stadje (2009) for the underlying counterexamples, and Backhoff Veraguas and Tangpi (2018) for a treatment of related time-inconsistent risk measures via PDE methods in a Brownian filtration. Our contribution here is to place this

negative result within the axiom system and to extract the specific implication for market making, namely that no CVaR-based objective is consistent with the axiomatically-forced preference functional, so the desk's internal time-consistent optimization and the regulator's CVaR limit are structurally different objects.

Corollary 31 (CVaR Time-Inconsistency)**.** *Any dynamic preference functional that takes the form of a dynamic conditional-value-at-risk on W_T^L, at any fixed confidence level $\beta \in (0,1)$, violates Axiom J4 (strong dynamic consistency). As a result, no CVaR-based dynamic objective is consistent with our axiomatic framework.*

Proof. We argue this claim directly from the structure of the candidate preference functional $J_t^{\mathrm{CVaR}} := -\mathrm{CVaR}_\beta(\cdot \mid \mathcal{F}_t)$, with the sign chosen so that higher J^{CVaR} corresponds to the strategy being preferred, matching the convention of Section 2.5. By Cheridito et al. (2006) (see in particular Remark 6.8 therein) and Cheridito and Stadje (2009), the dynamic conditional CVaR functional fails strong dynamic consistency in the sense of Axiom J4. The notion of time-consistency used in those works is the recursive Bellman form discussed after Axiom J4, and is equivalent to Axiom J4 under Axioms J1 and J2 in the present bounded-random-variable setup. A standard counterexample exhibits $W, W' \in L^\infty(\mathcal{F}_2)$ on a two-period filtration $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2$ with $J_1^{\mathrm{CVaR}}(W) \geq J_1^{\mathrm{CVaR}}(W')$ almost surely yet $J_0^{\mathrm{CVaR}}(W) < J_0^{\mathrm{CVaR}}(W')$, in direct violation of Axiom J4. Consequently, no dynamic CVaR functional can satisfy our axiom system, and in particular cannot be the unique functional forced by Theorem 10. □

A regulator who sets a CVaR-based market-risk limit and a dealer desk that runs a time-consistent dynamic optimization in our axiomatic sense are using incompatible objectives, in that the static end-of-day CVaR computed by the regulator is generally different from any dynamic CVaR objective the desk could have optimized against intraday. The desk's internal risk management, under strong time-consistency, is entropic; its external regulatory constraint is CVaR. We make no normative claim about which is right. The result identifies a structural gap which has not previously been pointed out in the inventory market making literature. As we noted in Section 3.2, the negative result is specific to strong dynamic consistency. A desk willing to weaken (J4) to acceptance- and rejection-consistency in the sense of Weber (2006) could in principle accommodate CVaR-like (shortfall) objectives.

6 Supporting Results

In this section we develop two supporting results which were stated, but not proved, in earlier sections. They appear here rather than alongside the axiomatic results for two reasons: both require explicit HJB machinery not needed for the axiomatic results of Sections 3–5, and both specialize the general axiomatic conclusion to concrete model parameters. The first is the explicit Hamilton-Jacobi-Bellman derivation of the forced coefficient $\phi = \gamma\sigma^2/2$ in the constant-volatility benchmark, invoked in the proof of Corollary 21. The second is the stochastic-volatility generalization, which uses the Dambis–Dubins–Schwarz time-change theorem and gives precise operational content to Corollary 12 (clock-invariance) in the nontrivial volatility regime.

6.1 The Forced Coefficient via HJB

The next proposition gives a direct HJB-based derivation of the forced coefficient $\phi = \gamma\sigma^2/2$ from Corollary 21, independent of the axiomatic derivation and suitable for readers more familiar with the stochastic-control tradition.

Proposition 32 (HJB Derivation of the Forced Coefficient)**.** *Let the market making setup be as in Section 2, with constant volatility $\sigma > 0$ and quadratic liquidation cost function $L(q) = \frac{\kappa}{2}q^2$. Define the AS-style value function*

$$V(t, X, q, S) \coloneqq \sup_{\pi \in \Pi} \mathbb{E}\Big[u\big(W_T^L(\pi)\big) \,\Big|\, X_t = X,\, q_t = q,\, S_t = S\Big], \quad u(x) = -e^{-\gamma x}.$$

Under the same hypotheses as Theorem 10, the value function admits the factorization

$$V(t, X, q, S) = -\exp\big(-\gamma(X + qS)\big) \cdot \psi(t, q), \tag{14}$$

for a function $\psi\colon [0, T] \times \mathbb{R} \to \mathbb{R}_+$, and the associated certainty-equivalent function given by $\theta(t, q) \coloneqq -\frac{1}{\gamma} \log \psi(t, q)$ satisfies the Bellman equation (18) below, whose running inventory term has coefficient exactly *$\gamma\sigma^2 q^2/2$. This is the HJB-level realization of the forced relation $\phi = \gamma\sigma^2/2$ asserted in Corollary 21.*

Proof. We work in the constant-volatility benchmark: $dS_t = \sigma\, dB_t$ for a Brownian motion B and constant $\sigma > 0$, with strategies $\pi = (\delta^a, \delta^b, C) \in \Pi$, wealth and inventory dynamics, and $L(q) = \frac{\kappa}{2}q^2$ with $\kappa > 0$, all as in Section 2. By Axiom J1, the value of any cash injection dC_t is exactly the injected amount, so the optimal control problem is translation-invariant in C and we may set $C \equiv 0$; the reduced control variable is the pair $(\delta_t^a, \delta_t^b) \in \mathbb{R}^2$.

By Theorem 10 and the certainty-equivalent identity (Remark 11), maximizing $J_t(\pi)$ over $\pi \in \Pi$ is equivalent to maximizing the CARA expected utility $\mathbb{E}[u(W_T^L(\pi)) \mid \mathcal{F}_t]$ with $u(x) = -e^{-\gamma x}$, the value function in the proposition. The terminal condition is equal to $V(T, X, q, S) = u(X + qS - \frac{\kappa}{2}q^2)$.

The structural feature of the CARA-Brownian setup is the factorization (14), with ψ a function to be determined, and terminal condition $\psi(T, q) = \exp(\gamma\kappa q^2/2)$. The HJB equation for the AS control problem is

$$\partial_t V + \tfrac{1}{2}\sigma^2 \partial_{SS} V + \sup_{\delta^a} \lambda^a(\delta^a)\, \Delta^a V + \sup_{\delta^b} \lambda^b(\delta^b)\, \Delta^b V = 0, \tag{15}$$

where $\Delta^a V$ and $\Delta^b V$ are the impulse operators recording the change in V upon an ask fill and a bid fill, respectively,

$$\begin{aligned} \Delta^a V(t, X, q, S) &\coloneqq V(t, X + S + \delta^a, q - 1, S) - V(t, X, q, S), \\ \Delta^b V(t, X, q, S) &\coloneqq V(t, X - S + \delta^b, q + 1, S) - V(t, X, q, S). \end{aligned}$$

Substituting the ansatz (14) into (15), two structural features emerge. First, the diffusion term $\frac{1}{2}\sigma^2\partial_{SS}V$ gives $-\frac{1}{2}\sigma^2\gamma^2 q^2 \exp(-\gamma(X + qS))\,\psi$, since $\partial_{SS}V = -\gamma^2 q^2 \exp(-\gamma(X + qS))\psi$; this is the term that delivers the $\phi = \gamma\sigma^2/2$ coefficient. Second, using $S + (q-1)S = qS$ for the ask impulse and the symmetric identity for the bid, the wealth-state factor $\exp(-\gamma(X + qS))$ pulls out of both impulse expressions, leaving ψ-only quantities. After dividing through by $-\exp(-\gamma(X + qS))$, the HJB equation reduces to a two-dimensional PDE for $\psi(t, q)$,

$$\partial_t \psi + \tfrac{1}{2}\sigma^2\gamma^2 q^2 \psi + \mathcal{I}^a[\psi] + \mathcal{I}^b[\psi] = 0, \tag{16}$$

where the impulse operators acting on ψ are

$$\begin{aligned} \mathcal{I}^a[\psi](t, q) &\coloneqq \sup_{\delta^a} \lambda^a(\delta^a)\big[e^{-\gamma\delta^a}\psi(t, q - 1) - \psi(t, q)\big], \\ \mathcal{I}^b[\psi](t, q) &\coloneqq \sup_{\delta^b} \lambda^b(\delta^b)\big[e^{-\gamma\delta^b}\psi(t, q + 1) - \psi(t, q)\big]. \end{aligned} \tag{17}$$

This is the form obtained by substituting the CARA ansatz into the AS HJB equation, as in standard derivations following Avellaneda and Stoikov (2008, Section 3).

The forced coefficient becomes visible after passing to the certainty-equivalent function $\theta(t,q) = -\frac{1}{\gamma}\log\psi(t,q)$, which records the time-t certainty-equivalent reduction, relative to mark-to-market wealth, of holding inventory q and following the optimal strategy thereafter, with terminal condition $\theta(T,q) = -\frac{\kappa}{2}q^2$. From $\psi = \exp(-\gamma\theta)$, equation (16) translates into

$$\partial_t\theta - \tfrac{\sigma^2\gamma}{2}q^2 + \mathcal{J}^a[\theta] + \mathcal{J}^b[\theta] = 0, \tag{18}$$

where the θ-impulse operators are

$$\mathcal{J}^a[\theta](t,q) = -\tfrac{1}{\gamma}\sup_{\delta^a}\lambda^a(\delta^a)\big[e^{-\gamma(\delta^a+\theta(t,q-1)-\theta(t,q))}-1\big],$$

and $\mathcal{J}^b[\theta]$ is the symmetric formula with $\theta(t,q-1)\to\theta(t,q+1)$. Equation (18) is structurally a Bellman equation for θ with running inventory cost accruing at rate $\gamma\sigma^2q^2/2$ per unit time, partially compensated by impulse activity through spread capture; the coefficient $\gamma\sigma^2/2$ is exact in the PDE, having entered via the diffusion term $(\sigma^2\partial_{SS}V)/2$ in (15) without any approximation. □

Remark 33 (Two Senses of "Forced" in the Coefficient $\phi = \gamma\sigma^2/2$)**.** Proposition 32 and Corollary 21 together describe the relation $\phi = \gamma\sigma^2/2$ in two complementary senses.

- *Exact in the coefficient.* In the θ-PDE (18) the running inventory term has coefficient $\gamma\sigma^2q^2/2$ *exactly*, arising from the diffusion term $(\sigma^2\partial_{SS}V)/2$ in the HJB equation after substitution of the CARA ansatz, with no Taylor expansion involved. This is also visible in the closed-form AS value function (Remark 22) as the exact term $-(\gamma q^2\sigma^2(T-t))/2$.
- *Approximate in the functional form.* The full CJ objective functional (9), with the running integral $\phi\int_0^T q_s^2\,ds$, is a leading-order (second-order in inventory magnitude) expansion of the full entropic functional (8), with higher-order terms in q neglected; the forced coefficient that this approximation must take, however, is $\gamma\sigma^2/2$ exactly.

A practitioner who reads Corollary 21 as saying "ϕ is the leading-order coefficient" should understand it in the second sense; the coefficient itself is determined to be $\gamma\sigma^2/2$ exactly by the HJB analysis above.

Remark 34 (Consistency with the AS Value Function)**.** In the AS benchmark in the absence of fills, the running cost is the dominant contribution, and integrating $\gamma\sigma^2q^2/2$ over $[t,T]$ at fixed inventory q gives $(\gamma\sigma^2q^2\,(T-t))/2$, which matches the term $-(\gamma q^2\sigma^2(T-t))/2$ appearing in the AS value function (see Remark 22).

Remark 35 (The Certainty-Equivalent Function as a Marked-to-Market Liquidity Cost)**.** The function $\theta(t,q)$ has the following operational interpretation. By construction, $\theta(t,q)$ is the conditional certainty-equivalent reduction, relative to the mark-to-mid wealth $X_t + qS_t$, attributable to holding inventory q at time t and following the optimal quoting strategy thereafter. Equivalently, $\theta(t,q)$ is the cash-equivalent value that the market maker would pay, at time t, to be flat (i.e., to have $q=0$ instead of q) for the remainder of the horizon, conditional on continuing to quote optimally. Under the forced functional, $\theta(t,q)$ thus plays the role of an inventory-specific liquidity discount: the cash amount by which the desk's true valuation of its book falls short of mark-to-mid.

A desk that wishes to mark its inventory book consistently with the dynamically-consistent preference functional we have derived should value the book at $X_t + qS_t + \theta(t,q)$, not the naive

mark-to-mid $X_t + qS_t$. The quantity $\theta(t,q)$ is the certainty-equivalent haircut that the desk's risk-aversion forces, and it is strictly negative whenever $q \neq 0$ and $t < T$, since the closed-form AS value function (Remark 22) gives $\theta(t,q) = -(\gamma q^2 \sigma^2 (T-t))/2$ plus impulse-induced corrections of the same sign. A desk that uses mark-to-mid for inventory while running an AS- or CJ-style quoting strategy is using two inconsistent valuations of its book, in that one drives quoting decisions (which implicitly applies the haircut) and another reports P&L (which does not). This is a structural-gap observation parallel in shape to Corollary 31 and Remark 28, a piece of standard practice inconsistent, in a precise sense, with the time-consistent dynamic functional our axioms force.

6.2 The Stochastic-Volatility Extension

We now allow the price volatility to be stochastic, letting S be a continuous semimartingale with $d\langle S\rangle_t = \sigma_t^2\, dt$, where σ is a positive predictable process. No assumption is made on the dynamics of σ beyond predictability and local boundedness; in particular, the result below covers Heston-type and rough-volatility models. The natural question is whether the forced coefficient $\phi = \gamma\sigma^2/2$ of Proposition 32 extends to this setting. By Corollary 12 (clock-invariance) and the Dambis–Dubins–Schwarz theorem, it does, with σ^2 replaced by the instantaneous σ_t^2.

Proposition 36 (Clock-Corrected Penalty)**.** *Let the price process satisfy $d\langle S\rangle_t = \sigma_t^2\, dt$ with σ a predictable process such that $\sigma_t > 0$ almost surely for almost every $t \in [0,T]$. Under the same hypotheses as Theorem 10, the market maker's certainty-equivalent function $\theta(t,q,\omega) = -\frac{1}{\gamma}\log\psi(t,q,\omega)$ satisfies a Bellman equation with running inventory cost*

$$\frac{\gamma}{2}\, q^2\, d\langle S\rangle_t = \frac{\gamma\sigma_t^2}{2}\, q^2\, dt. \tag{19}$$

Equivalently, in business-time units $u = \langle S\rangle_t$ (which carries the units of price-variance, not time), the running cost is the constant $\gamma q^2/2$ per unit of business-time.

Proof. We use a time-change argument in three steps, with hat-decorated symbols ($\widehat{J}_u$) denoting business-time reparametrizations on the wall-clock strategy space Π, tilde-decorated symbols ($\tilde\theta, \tilde\psi, \tilde V, \tilde{\mathbb{F}}, \tilde S, \tilde M, \tilde\pi, \tilde\lambda^{a,b}, \tilde N^{a,b}, \tilde{\mathcal{J}}^{a,b}$) denoting the analogous objects on the business-time setup, and undecorated symbols denoting wall-clock quantities. The two systems are related by the time-change $u = \Lambda_t$ with inverse $t = \tau(u)$.

Step 1: Time-change. Define $\Lambda_t := \langle S\rangle_t = \int_0^t \sigma_s^2\, ds$. Under the proposition's assumption that $\sigma_t > 0$ for almost every $t \in [0,T]$ almost surely, the process Λ is continuous (by absolute continuity of the integral with respect to its upper limit) and strictly increasing on $[0,T]$ (since $\int_s^t \sigma_u^2\, du > 0$ for every $0 \le s < t \le T$ on the full-probability event $\{\sigma_u > 0$ for a.e. $u\}$), with $\Lambda_0 = 0$. The right-continuous inverse $\tau(u) := \inf\{t \ge 0 \mid \Lambda_t > u\}$ is a continuous time-change, and the time-changed filtration is $\tilde{\mathcal{F}}_u := \mathcal{F}_{\tau(u)}$. Write $S = M + A$ for the canonical decomposition of S as a continuous semimartingale, with M the continuous local-martingale part (so $\langle M\rangle = \langle S\rangle$) and A the continuous finite-variation part; by the Dambis–Dubins–Schwarz theorem applied to M, the time-changed martingale part $\tilde M_u := M_{\tau(u)}$ is a standard Brownian motion with respect to $\tilde{\mathcal{F}}$. In the general stochastic-volatility case the HJB equation acquires a drift term $\mu_t \partial_S V$ relative to (15), which under the CARA ansatz produces $-\gamma q\, \mu_t \cdot V$: a linear-in-$q$ term (the standard alpha-skewing of optimal quotes) that does not affect the quadratic-in-q running inventory cost coming from $\frac{1}{2}\sigma_t^2 \partial_{SS} V$. Since the proposition's conclusion is about the quadratic-in-q coefficient only, the drift is irrelevant, and we proceed as if $A \equiv 0$.

The proposition's hypothesis $\sigma_t > 0$ a.s. for almost every t ensures that $\tau'(u) = 1/\sigma^2_{\tau(u)}$ is finite for almost every $u \in [0, \Lambda_T]$, which is the regularity required for Brémaud's time-change theorem below. The time-change theorem for point processes (Brémaud, 1981, Theorem T16, p. 41) implies that $\tilde{N}^a_u := N^a_{\tau(u)}$ and $\tilde{N}^b_u := N^b_{\tau(u)}$ are Cox processes under $\tilde{\mathbb{F}}$; since $\Lambda'_t = \sigma^2_t$ and $\tau'(u) = 1/\Lambda'_{\tau(u)}$, a Cox process with wall-clock intensity $\lambda^{a,b}_t$ has business-time intensity

$$\tilde{\lambda}^{a,b}_u = \lambda^{a,b}_{\tau(u)} \cdot \tau'(u) = \frac{\lambda^{a,b}_{\tau(u)}}{\sigma^2_{\tau(u)}}.$$

A strategy $\pi = (\delta^a, \delta^b, C) \in \Pi$ maps under the time-change to $\tilde{\pi} := (\tilde{\delta}^a, \tilde{\delta}^b, \tilde{C})$ with $\tilde{\delta}^{a,b}_u := \delta^{a,b}_{\tau(u)}$ and $\tilde{C}_u := C_{\tau(u)}$, giving a bijection between Π on $[0, T]$ and the analogous admissible strategy space on $[0, \Lambda_T]$ with $\widetilde{W^L}_{\Lambda_T}(\tilde{\pi}) = W^L_T(\pi)$.

Step 2: HJB in business-time. By Corollary 12, the business-time reparametrization $\widehat{J}_u(\pi) := J_{\tau(u)}(\pi)$ has the entropic form on $\tilde{\mathbb{F}}$ with the same parameter γ as J, hence is stationary in Λ in the sense that the five axioms of Section 3, together with the three derived properties, hold for $\widehat{J}$ on $\tilde{\mathbb{F}}$ with t replaced by u throughout. The HJB analysis of Section 6.1 therefore applies directly in business-time, with σ replaced by 1 (since $\tilde{M}$ is standard Brownian) and arrival intensities $\tilde{\lambda}^{a,b}$ as in Step 1. Substituting the ansatz $\tilde{V}(u, X, q, \tilde{S}) = -\exp(-\gamma(X + q\tilde{S}))\,\tilde{\psi}(u, q)$ and passing to $\tilde{\theta}(u, q) = -\frac{1}{\gamma}\log\tilde{\psi}(u, q)$, the business-time Bellman equation reads

$$\partial_u\tilde{\theta} - \frac{\gamma}{2}\,q^2 + \tilde{\mathcal{J}}^a[\tilde{\theta}] + \tilde{\mathcal{J}}^b[\tilde{\theta}] = 0, \tag{20}$$

with running cost rate $\frac{\gamma}{2}\,q^2$ per unit of business-time.

Step 3: Pull-back to wall-clock. Set $\theta(t, q, \omega) := \tilde{\theta}(\Lambda_t(\omega), q, \omega)$; the chain rule then gives $\partial_t\theta = \sigma^2_t\,\partial_u\tilde{\theta}$, and by Step 1 the business-time intensities satisfy $\sigma^2_t\,\tilde{\lambda}^{a,b}_{\Lambda_t} = \lambda^{a,b}_t$, so multiplying (20) by σ^2_t and using $\tilde{\mathcal{J}}^{a,b}[\tilde{\theta}] = \mathcal{J}^{a,b}[\theta]$ at corresponding (u, t) pairs yields

$$\partial_t\theta - \frac{\gamma\sigma^2_t}{2}\,q^2 + \lambda^a_t\,\mathcal{J}^a[\theta] + \lambda^b_t\,\mathcal{J}^b[\theta] = 0,$$

which corresponds to the wall-clock running cost rate (19). □

In words, the running inventory cost is constant in business-time and proportional to σ^2_t in wall-clock time: a volatility spike proportionally raises the maker's inventory penalty rate while a quiet regime keeps it low, and the axiomatic framework forces this empirically natural behavior.

CJ in stochastic volatility, and the connection to rough-volatility market making. The Cartea–Jaimungal running penalty $\int_0^T q^2_s\,ds$, applied uncritically in a stochastic-volatility setting, is inconsistent with the clock-invariance of Corollary 12; the clock-coherent version is $\int_0^T \sigma^2_s q^2_s\,ds$, which scales the inventory penalty with the instantaneous variance rate, and this is the form Proposition 36 forces.

This prescription connects directly to the multi-asset framework of Rosenbaum and Zhang (2022), where the inventory risk is expressed in the quadratic-variation clock of an SPX underlying under the quadratic rough Heston model; their objective specializes Proposition 36 to the rough Heston price process, and what our framework adds is an axiomatic foundation for that scaling. The Dambis–Dubins–Schwarz theorem applies as soon as S has a continuous local-martingale part, so the forced relation $\phi_t = \gamma\sigma^2_t/2$ survives in rough-volatility models (Bergomi, Stein–Stein, log-modulated, and the like) without model-specific re-derivation, as a

consequence of the axiom system rather than of any particular volatility model. Any alternative scaling (wall-clock CJ or an ad-hoc volatility-dependent heuristic) would be inconsistent with the structural clock-invariance of Corollary 12, hence outside the axiom class.

The risk-aversion parameter γ is a single positive scalar by Theorem 10 and is constant across the realized variance path; a market maker who wishes to allow γ to vary across regimes would need to enrich the framework by indexing the preference functional over a state variable, which we do not develop here.

7 Discussion

7.1 Operational Implications

In operational terms, the framework delivers three concrete things to a practitioner.

First, the running-penalty coefficient ϕ in any CJ-style implementation is not a free parameter. It should be set as $\phi_t = \gamma\sigma_t^2/2$, with σ_t the instantaneous volatility, so that the running penalty against the quadratic-variation clock has the constant coefficient $\gamma/2$. Calibrating ϕ independently of γ is over-parametrizing the preference side of the model.

Second, the implied risk aversion γ is recoverable from observed quoted spreads via the calibration-inversion of Corollary 25: $\gamma = 2\phi/\sigma^2$. For a CJ-tradition desk, this inversion provides a sanity check, since the γ implied by the desk's running-penalty calibration should match, up to noise, the γ implied by an AS-style fit to the same data. Persistent disagreement between the two is a signal of miscalibration or of a preference-functional inconsistency.

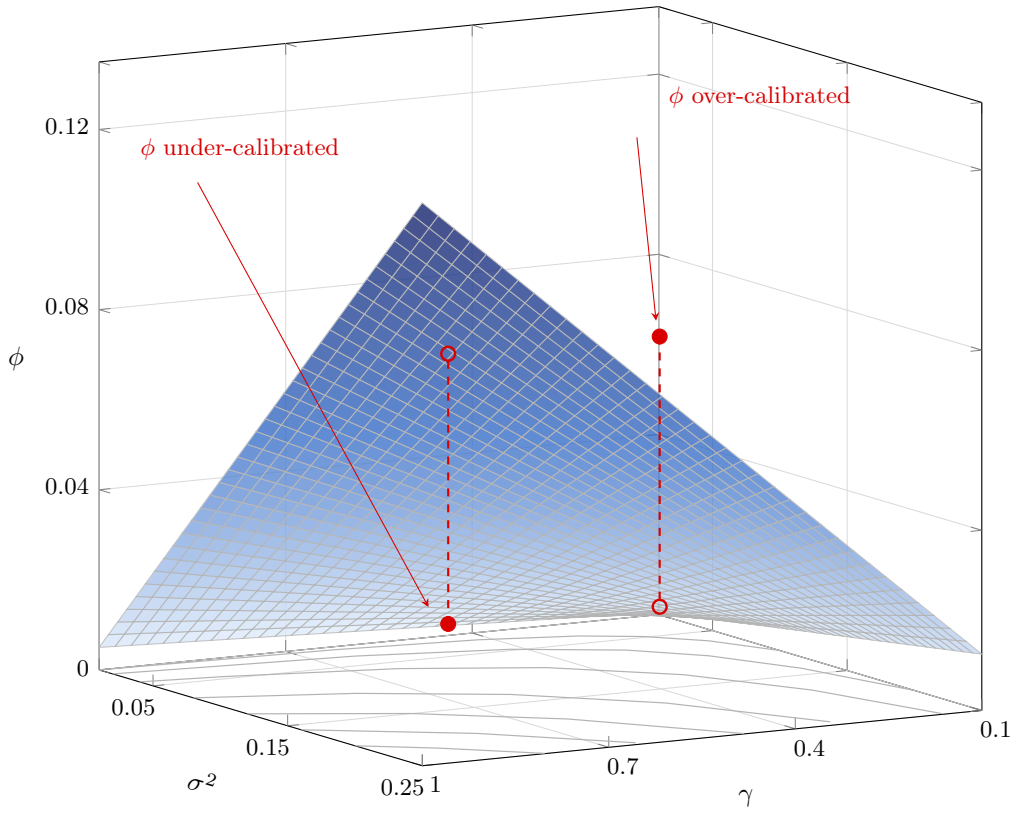
Third, the same forced relation has an implementation consequence in stochastic-volatility settings, in that the inventory penalty must be integrated against the quadratic-variation clock, not wall-clock time. The clock-coherent CJ running cost is $\frac{\gamma}{2}\int_0^T q_s^2\, d\langle S\rangle_s = \frac{\gamma}{2}\int_0^T \sigma_s^2 q_s^2\, ds$, not $\phi\int_0^T q_s^2\, ds$ with constant ϕ; a wall-clock penalty with constant coefficient systematically under-charges inventory on high-volatility days and over-charges it on low-volatility days.

These three prescriptions follow from the axioms alone, and a desk that accepts the five axioms is committed to all three. Figure 2 summarizes the constraint geometry and the stochastic-volatility correction.

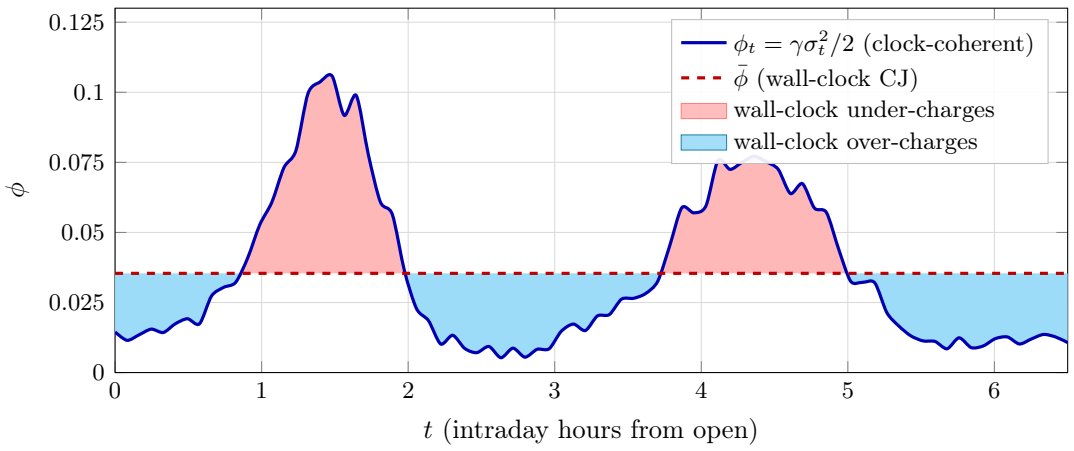
7.2 Drawdown-Averse Market Makers

Proposition 7 commits us to preferences that depend on the strategy only through liquidation-adjusted terminal wealth W_T^L, so path-functional preferences (aversion to maximum drawdown $\sup_{s\le T}(W_0^L - W_s^L)$, to time-average wealth, or to drawdown over a rolling window) are genuinely outside this framework and cannot be absorbed into W_T^L by any reparametrization. This is a real limitation.

Proprietary trading firms typically impose intraday drawdown limits on their market makers, whereby if cumulative losses since the start of the trading day exceed a threshold, the desk is stopped out and required to flatten its book. Such limits are externally imposed constraints, not preferences in the usual sense, but the behavior they induce (more conservative quoting after a losing streak) is genuinely path-functional and is not captured by any W_T^L-based functional. An externally imposed intraday drawdown constraint, viewed as a hard rule on admissible strategies, changes the effective strategy space Π and hence the optimization problem but need not change the preference functional J itself; the limitation of our framework is therefore about *preference modeling*, not about whether hard constraints can be imposed on the strategy space, since such constraints can be added orthogonally to Π while a market maker whose preferences are themselves drawdown-averse remains outside the wealth-summary framework.



(a) The forced relation $\phi = \gamma\sigma^2/2$ as a surface in (γ, σ^2, ϕ) space, with the hyperbolic level curves $\phi = c$ projected onto the (γ, σ^2) floor. Any axiom-consistent calibration lies on this surface. The two filled red dots show off-surface desk calibrations, one with ϕ over-calibrated and one with ϕ under-calibrated; each is connected by a dashed segment to the corresponding open red circle on the surface, which marks the forced value.



(b) Clock-coherent inventory penalty rate $\phi_t = \gamma\sigma_t^2/2$ on a stylized intraday volatility path with two bursts, against the wall-clock CJ constant $\bar{\phi}$ calibrated to the daily average. The wall-clock constant under-charges inventory during high-volatility regimes and over-charges it during quiet periods, inconsistent with the clock-invariance of Corollary 12 and corrected by Proposition 36.

Figure 2: The forced relation $\phi = \gamma\sigma^2/2$ in two views. (a) The constraint surface in parameter space; off-surface calibrations are over-parametrized. (b) The stochastic-volatility version, $\phi_t = \gamma\sigma_t^2/2$, tracks realized variance pointwise in time, while a wall-clock constant systematically misprices inventory.

The result is best read as a uniqueness theorem within the class of W_T^L-based preferences, a substantive but not universal modeling commitment. Note also the scoping distinction from optimal execution (Almgren and Chriss, 2000), which is a cost-minimization problem with fixed terminal target $q_T = 0$, and to which our forced-uniqueness theorem does not apply. The right axiomatic treatment of a genuinely drawdown-averse market maker is an open question; relevant entry points are Cherny and Madan (2009) on path-dependent acceptability indices and Cheridito et al. (2006) on dynamic monetary risk measures for bounded discrete-time processes.

7.3 Relation to Cartea–Jaimungal–Penalva (2015) and Guéant (2017)

Two prior works reduce the AS and CJ frameworks to a common object at the value-function level.

Cartea et al. (2015, Section 10.3, p. 261) show that under the parameter identification $\phi = \gamma\sigma^2/2$ in the notation used here, together with matching intensity and depth parameters, the value-function ODEs for the CARA and running-penalty market makers coincide in structure, and the two optimal quoting strategies agree up to a constant shift under this identification. The result operates in the constant-volatility setting and for the specific intensity form $\lambda^{\pm}e^{-\kappa^{\pm}\delta^{\pm}}$.

Guéant (2017) extends this pattern to a broader HJB ansatz. He shows that in both his Model A (AS with CARA expected-utility maximization) and Model B (CJ with a running inventory penalty), the value function $v(t, q)$ admits an ansatz reducing the HJB equation to a linear tridiagonal ODE system, and the two ODE systems are related by an explicit change of variables, so that a CJ-style problem with a given ϕ can be mapped to an AS-style problem with a corresponding γ.

Our contribution sits at a different structural level. Both prior works unify the two frameworks at the level of the value function and its ODE, by direct calculation in specific HJB problems; the present paper unifies them at the level of the preference functional itself, by axiomatic characterization. The difference has three consequences. First, parameter-forcing rather than parameter-mapping: the prior results map between two free-parameter model families, while ours identifies a constraint surface $\{\phi = \gamma\sigma^2/2,\ \alpha = \frac{1}{2}L''(0)\}$ on which any axiom-consistent preference must lie. Second, functional-level rather than HJB-level: the prior reconciliations require specific HJB problems with specific intensity forms, whereas the axiomatic result holds whenever the axioms apply, and Proposition 32 of Section 6 is a specialization rather than the primary content. Third, further structural consequences, notably the forcing of α by the liquidation cost curvature, the robust-optimization dual, position-size nonlinearity, and incompatibility with dynamic conditional-value-at-risk, are visible at the functional level and invisible at the ODE level.

A third ODE-level precedent for the identification $\phi = \gamma\sigma^2/2$ appears in Cartea et al. (2017), who derive an equivalence between a relative-entropy ambiguity penalty and a running quadratic inventory penalty in the AS market-making setting, with a coefficient relation of the same shape in their notation. Their work operates at the value-function level and takes the entropic penalty as a modeling primitive, motivated by ambiguity aversion, rather than deriving it axiomatically; the dual reading of our Corollary 26 (robust-optimization interpretation) is closely related to their framework and can be viewed as the axiomatic justification for the ambiguity-averse formulation they adopt on modeling grounds.

7.4 Further Structural Connections

Theorem 10 connects to several adjacent literatures. Further connections, to the prediction-market tradition and to minimal-entropy pricing measures, are noted in Section 7.5.

Multi-asset market making with a single risk-aversion parameter. The forced uniqueness theorem is stated for a single-asset market maker, but its natural multi-asset extension preserves the single-scalar-γ structure. Once one reduces, via the wealth-summary property (W) (Proposition 7), to a functional of the scalar random variable $W_T^L \in L^\infty(\mathcal{F}_T)$, the entire multi-asset structure is absorbed into the joint distribution of W_T^L, and the Kupper–Schachermayer representation continues to apply unchanged. Asset-specific structure (correlations, asset-specific liquidation costs, asset-specific volatilities) enters through the joint distribution of W_T^L, not through any preference-side asymmetry. The formal statement and proof are in Appendix C (Theorem 48). The standard multi-asset AS-CJ extensions in the literature, notably Bergault et al. (2021) and Rosenbaum and Zhang (2022), likewise treat γ as a single scalar across assets on modeling grounds; our multi-asset extension provides the axiomatic justification for this convention.

Operationally, this gives a sharp cross-asset falsifiability criterion. A desk that fits, for each of K assets, its implicit γ via the inversion $\gamma = 2\phi/\sigma^2$ of Corollary 25 should obtain K approximately equal values. Persistent cross-asset disagreement is evidence either of a violation of the multi-asset axiom system or of inconsistent calibration.

7.5 Directions for Future Work

We close with four directions which our framework suggests but does not pursue.

Empirical testability of the forced coefficient. The forced relation $\phi = \gamma\sigma^2/2$ is an empirically falsifiable prediction. A market-making desk that calibrates γ from AS-style quoted-spread data and ϕ from CJ-style P&L or inventory-penalty data can compare the two implied values against $\gamma\sigma^2/2$ at the prevailing realized volatility, using the calibration-inversion $\gamma_\phi = 2\phi/\sigma^2$ of Corollary 25. A desk for which γ_ϕ and the AS-implied γ_{AS} disagree persistently is either over-parametrized or violating at least one of the five axioms; this is a productive direction for applied work with desk-internal or high-frequency data.

Strategy space: the limit order book. We have taken the strategy space to be the AS specification of two real-valued quote-distance processes (augmented with a cash-injection process), whereas real limit order books operate under price-time priority, where modifying a quote loses queue priority; see Law and Viens (2019), Lu and Abergel (2018), and Moallemi and Yuan (2016). Our preference-side axioms are independent of the strategy space, and a natural extension would combine them with explicit queue-aware strategy-space axioms; an interesting question is whether the AS continuous-quote control arises as the small-tick limit of a queue-aware control.

Endogenous flow. The assumption made in this paper that the counterparty intensities λ^a, λ^b are exogenous functions of quote distances is in fact a simplification, since sophisticated counterparties ("skew sniffers") read the market maker's position from her quote skew and adjust their arrivals (Barzykin et al., 2025). A natural extension of this framework would investigate whether equilibrium-rational counterparty behavior is consistent with Axioms J1–J5; preliminary investigation suggests the framework absorbs endogenous flow as state-dependence of the intensities rather than as a violation of any axiom.

Convergence with the prediction-market tradition. The prediction-market tradition, originating with the market-scoring-rule mechanism of Hanson (2003, 2007) and developed by Chen and Pennock (2007), Abernethy et al. (2013), and Abernethy et al. (2014), identifies the entropic family as canonical under different axioms (path-independence, bounded loss,

translation-invariance, monotonicity), with the cost function forced to be the conjugate of a generalized entropy (Abernethy et al., 2014). Although the two axiom systems are structurally different, the shared emergence of the entropic family suggests a deeper connection through cash-additivity and a consistency requirement, whose unified treatment we leave to future work.

Connection to minimal-entropy pricing measures. The entropic functional is intimately connected to the minimal-entropy martingale measure of Frittelli (2000), which arises in utility-indifference pricing in incomplete markets (Föllmer and Schweizer, 1991), and identified via the dual representation of the entropic risk measure in the CARA case; the connection is suggestive rather than exact, since our strategy space (quoting plus cash injection) differs from the self-financing space of indifference pricing.

8 Conclusion

We have proposed five core axioms for a dynamic preference functional of an inventory market maker (cash-additivity, normalization, concavity with strict concavity at $t = 0$, strong dynamic consistency, and law-invariance), and three additional properties ((M) monotonicity, (W) wealth-summary, and (R) right-continuity) that are derivable from them. The five axioms force the market maker's preferences to be the entropic certainty-equivalent on liquidation-adjusted terminal wealth, parametrized by a single positive scalar γ which is clock-invariant in the sense of Corollary 12.

The Avellaneda–Stoikov framework is the unique representative of the axiom class, and the Cartea–Jaimungal objective does not arise as a primitive preference functional within the axiom system, but appears instead as the second-order Taylor expansion of the axiomatically-determined entropic functional around zero inventory. The running coefficient is $\phi = \gamma\sigma^2/2$, recorded at the value-function level by Cartea et al. (2015, p. 261) in the constant-volatility case; the identification arises here as a corollary of the axiomatic characterization, extends to stochastic volatility through clock-invariance of γ against the quadratic-variation clock, and inverts to $\gamma = 2\phi/\sigma^2$, providing a consistency cross-check for desks operating in the CJ tradition. Under a mild regularity condition on the liquidation cost function, the terminal coefficient α is a property of the market microstructure, equal to $\frac{1}{2}L''(0)$. Beyond the AS–CJ split, the forced functional admits a robust-optimization dual under a Knightian-ambiguity reading, is not positively homogeneous (so linear-in-size risk-capital allocation misprices large positions), and is structurally incompatible with dynamic CVaR. The preference structure has one free scalar parameter on the open half-line, interpolating between risk-neutral and worst-case limits, and the constant-absolute-risk-aversion property of the AS utility is a corollary of the deeper uniqueness rather than a primitive choice.

The framework has a real limitation, namely that drawdown-averse market makers, for whom intraday loss limits induce path-functional preferences, are not covered. Four directions for future work emerged from the analysis, namely empirical testability of the forced coefficient, the strategy-space side of limit-order-book microstructure, endogenous counterparty flow, and the connection to the prediction-market tradition.

The message is this. The inventory market making literature has carried two parallel traditions since Avellaneda and Stoikov (2008), with the choice between them governed by tractability rather than by axiomatic argument. The present paper places both traditions under a common axiomatic foundation, at the preference-functional level rather than the value-function level of the prior ODE-based reductions of Cartea et al. (2015, p. 261) and Guéant (2017). A practitioner who calibrates ϕ and γ as independent parameters is leaving a falsifiable prediction on the table.

A Derived Properties and Independence of Axioms

This appendix establishes two complementary results about the axiom system of Section 3. First, we prove the three derivability propositions mentioned in Section 3.5: monotonicity (M) (Proposition 37), the wealth-summary property (W) (Proposition 38), and right-continuity in time (R) (Proposition 39). The relevance condition is likewise derivable from the five core axioms, as already established by Lemma 5 in Section 3.4. Second, we prove that the five core axioms (J1)–(J5) are *logically independent*: for each axiom, there exists a preference functional satisfying the other four but not that one (Theorem 40).

A.1 Derived Properties

Throughout this subsection, $J = (J_t)_{t\in[0,T]}$ is a dynamic preference functional on the strategy space Π satisfying a subset of Axioms J1–J5 specified separately for each result. We make repeated use of the standing assumption (Section 2) that Π is closed under bounded $\mathcal{F}_T$-measurable cash injection. That is, for any $\pi \in \Pi$ and any bounded $c \in L^\infty(\mathcal{F}_T)$, the cash-injected strategy $\pi^{[c]}$ (defined to inject c at time T) lies in Π, with $W_T^L(\pi^{[c]}) = W_T^L(\pi) + c$ almost surely (Lemma 46 of Appendix B). We also use the existence of a strategy $\pi_0 \in \Pi$ with $W_T^L(\pi_0) = 0$ almost surely (the constant-zero strategy, also Lemma 46).

Proposition 37 (Monotonicity (M), Appendix Version)**.** *Any map J satisfying Axioms J1 (cash-additivity), J2 (normalization), and J4 (strong dynamic consistency) also satisfies the monotonicity property (M) of Proposition 6: for all $\pi, \pi' \in \Pi$, if $W_T^L(\pi) \geq W_T^L(\pi')$ almost surely, then $J_t(\pi) \geq J_t(\pi')$ almost surely for every $t \in [0,T]$.*

Proof. Step 1: At terminal time T, the functional J_T equals liquidation-adjusted terminal wealth. Fix $\pi \in \Pi$ and let $W := W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$; the strategy $\pi^{[-W]} \in \Pi$ obtained by injecting the bounded $\mathcal{F}_T$-measurable cash amount $-W$ at time T has $W_T^L(\pi^{[-W]}) = 0$ almost surely, so Axiom J2 gives $J_T(\pi^{[-W]}) = 0$ a.s., while Axiom J1 at time T gives $J_T(\pi^{[-W]}) = J_T(\pi) - W$ a.s. Combining yields $J_T(\pi) = W_T^L(\pi)$ almost surely, and since π was arbitrary,

$$J_T(\pi) = W_T^L(\pi) \quad \text{almost surely, for every } \pi \in \Pi. \tag{21}$$

Step 2: Propagate monotonicity backwards via dynamic consistency. Let $\pi, \pi' \in \Pi$ satisfy $W_T^L(\pi) \geq W_T^L(\pi')$ almost surely; by (21), $J_T(\pi) \geq J_T(\pi')$ almost surely, and applying Axiom J4 with $(s,t) = (t,T)$ for any $t \in [0,T]$ yields $J_t(\pi) \geq J_t(\pi')$ almost surely. □

Proposition 38 (Wealth-Summary (W), Appendix Version)**.** *Any J satisfying monotonicity (M) of Proposition 37 satisfies the wealth-summary property of Proposition 7: for all $\pi, \pi' \in \Pi$, $W_T^L(\pi) = W_T^L(\pi')$ a.s. implies $J_t(\pi) = J_t(\pi')$ a.s. for every $t \in [0,T]$. In particular, any J satisfying Axioms J1, J2, and J4 satisfies the wealth-summary property.*

Proof. If $W_T^L(\pi) = W_T^L(\pi')$ almost surely, both $W_T^L(\pi) \geq W_T^L(\pi')$ and $W_T^L(\pi') \geq W_T^L(\pi)$ hold almost surely, so applying Proposition 37 in both directions yields $J_t(\pi) = J_t(\pi')$ almost surely. □

Proposition 39 (Right-Continuity in Time (R), Appendix Version)**.** *Any J satisfying Axioms J1–J5 also satisfies the right-continuity property (R) of Proposition 9: for every $\pi \in \Pi$ and every $t \in [0,T)$, the path $s \mapsto J_s(\pi)$ is right-continuous in probability at $s = t$.*

Proof. The proof proceeds in two stages: Stage 1 establishes the entropic representation $J_t(\pi) = -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W_T^L(\pi)} \mid \mathcal{F}_t]$ at every $t \in [0,T]$ via a Bellman + cash-additivity argument that does not invoke right-continuity, and Stage 2 derives right-continuity in probability from the right-continuity of conditional expectations under the usual conditions on the filtration.

By Proposition 38, $J_t(\pi)$ depends on π only through $W := W_T^L(\pi)$, so we work with the reduced functional $\tilde{J}_t \colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$, defined on all of $L^\infty(\mathcal{F}_T)$ by Lemma 46.

Stage 1, Step 1.1: Bellman identity at every time. We claim that for every $t_0 \in [0,T]$ and every $W \in L^\infty(\mathcal{F}_T)$,

$$\tilde{J}_s\big(\tilde{J}_{t_0}(W)\big) = \tilde{J}_s(W) \quad \text{almost surely, for all } s \le t_0. \tag{22}$$

Fix t_0 and W. Let $\pi, \pi_0 \in \Pi$ with $W_T^L(\pi) = W$ and $W_T^L(\pi_0) = 0$. Writing $M_W := \|W\|_\infty$, the cash-injected strategies $\pi_0^{[\pm M_W]}$ satisfy $W_T^L(\pi_0^{[\pm M_W]}) = \pm M_W$ a.s. and, by Axioms J1 and J2, $J_{t_0}(\pi_0^{[\pm M_W]}) = \pm M_W$ a.s. Since $W_T^L(\pi_0^{[-M_W]}) \le W_T^L(\pi) \le W_T^L(\pi_0^{[M_W]})$ a.s., Proposition 37 gives $-M_W \le J_{t_0}(\pi) \le M_W$ a.s., so $\tilde{J}_{t_0}(W) \in L^\infty(\mathcal{F}_{t_0})$ with $\|\tilde{J}_{t_0}(W)\|_\infty \le M_W$, and the strategy $\pi^* := \pi_0^{[\tilde{J}_{t_0}(W)]}$ is admissible with $W_T^L(\pi^*) = \tilde{J}_{t_0}(W)$ a.s. By Axioms J2 and J1 at t_0, $J_{t_0}(\pi^*) = \tilde{J}_{t_0}(W) = J_{t_0}(\pi)$ a.s., and applying Axiom J4 in both directions to the pair (π, π^*) yields $J_s(\pi) = J_s(\pi^*)$ a.s. for all $s \le t_0$, which translates to $\tilde{J}_s(W) = \tilde{J}_s(\tilde{J}_{t_0}(W))$, giving (22).

Stage 1, Step 1.2: Entropic form at dyadic times. Let $\mathcal{D} := \bigcup_{n\ge 1}\{kT/2^n \mid 0 \le k \le 2^n\}$ be the set of dyadic rationals in $[0,T]$. By the Kupper–Schachermayer representation argument (the discrete-time content of Sub-steps 3a–3b in the proof of Theorem 10, which uses Axioms J1–J5 and relevance but does not use right-continuity), there exists a unique $\gamma \in (0,\infty)$ such that for every dyadic $s \in \mathcal{D}$ and every $W \in L^\infty(\mathcal{F}_T)$,

$$\tilde{J}_s(W) = -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_s] \quad \text{almost surely.} \tag{23}$$

At dyadic times we therefore have the desired entropic form.

Stage 1, Step 1.3: A conditional MGF identity below t_0. Fix an arbitrary $t_0 \in [0,T]$ (not necessarily dyadic) and $W \in L^\infty(\mathcal{F}_T)$. For any dyadic $s \in \mathcal{D}$ with $s < t_0$, combining (22) and (23) gives

$$-\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_s] = \tilde{J}_s(W) = \tilde{J}_s\big(\tilde{J}_{t_0}(W)\big) = -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma \tilde{J}_{t_0}(W)} \mid \mathcal{F}_s].$$

Since both sides are bounded random variables and the logarithm is injective on the strictly positive bounded range, we obtain

$$\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_s] = \mathbb{E}[e^{-\gamma \tilde{J}_{t_0}(W)} \mid \mathcal{F}_s] \quad \text{almost surely, for every dyadic } s < t_0. \tag{24}$$

Stage 1, Step 1.4: Pass to the left-limit $s \uparrow t_0$. Let $(s_n)_{n\ge 1} \subset \mathcal{D}$ be a sequence of dyadic times with $s_n < t_0$ and $s_n \uparrow t_0$, which exists since $\mathcal{D}$ is dense in $[0,T]$. The σ-algebras $\mathcal{F}_{s_n}$ increase to $\mathcal{F}_{t_0^-} := \sigma\big(\bigcup_n \mathcal{F}_{s_n}\big) = \sigma\big(\bigcup_{s<t_0} \mathcal{F}_s\big)$. By the Lévy upward martingale convergence theorem applied to the bounded random variables $e^{-\gamma W}$ and $e^{-\gamma \tilde{J}_{t_0}(W)}$, both sides of (24) converge almost surely as $n \to \infty$, yielding

$$\mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_0^-}] = \mathbb{E}[e^{-\gamma \tilde{J}_{t_0}(W)} \mid \mathcal{F}_{t_0^-}] \quad \text{almost surely.} \tag{25}$$

Stage 1, Step 1.5: Use cash-additivity to interrogate with $\mathcal{F}_{t_0}$-measurable test functions. We have established that (25) holds for every $W \in L^\infty(\mathcal{F}_T)$. Fix any bounded $\mathcal{F}_{t_0}$-measurable random variable $c \in L^\infty(\mathcal{F}_{t_0})$ and apply (25) to $W + c$ in place of W (which is valid since $W + c \in L^\infty(\mathcal{F}_T)$). By Axiom J1 (cash-additivity at time t_0), $\tilde{J}_{t_0}(W + c) = \tilde{J}_{t_0}(W) + c$ almost surely, hence $e^{-\gamma \tilde{J}_{t_0}(W+c)} = e^{-\gamma c} \cdot e^{-\gamma \tilde{J}_{t_0}(W)}$ almost surely. Substituting, we obtain that

$$\mathbb{E}\big[e^{-\gamma c}\, e^{-\gamma \tilde{J}_{t_0}(W)} \,\big|\, \mathcal{F}_{t_0^-}\big] = \mathbb{E}\big[e^{-\gamma c}\, e^{-\gamma W} \,\big|\, \mathcal{F}_{t_0^-}\big] \quad \text{almost surely.}$$

We then apply the tower property to the right-hand side, using $\mathcal{F}_{t_0^-} \subseteq \mathcal{F}_{t_0}$ and the $\mathcal{F}_{t_0}$-measurability of $e^{-\gamma c}$, to get

$$\mathbb{E}\big[e^{-\gamma c}\, e^{-\gamma W} \,\big|\, \mathcal{F}_{t_0^-}\big] = \mathbb{E}\big[e^{-\gamma c}\, \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_0}] \,\big|\, \mathcal{F}_{t_0^-}\big].$$

Define

$$\Delta := e^{-\gamma \tilde{J}_{t_0}(W)} - \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_0}].$$

Both terms are $\mathcal{F}_{t_0}$-measurable and bounded (the first because $\tilde{J}_{t_0}(W) \in L^\infty(\mathcal{F}_{t_0})$, the second by definition of conditional expectation), so $\Delta \in L^\infty(\mathcal{F}_{t_0})$. Subtracting, we obtain

$$\mathbb{E}\big[e^{-\gamma c}\, \Delta \,\big|\, \mathcal{F}_{t_0^-}\big] = 0 \quad \text{almost surely, for every } c \in L^\infty(\mathcal{F}_{t_0}). \tag{26}$$

Stage 1, Step 1.6: Conditional MGF argument forces $\Delta = 0$. Since $\Delta \in L^\infty(\mathcal{F}_{t_0})$, for any $k \in \mathbb{R}$ the random variable $c_k := k\Delta/\gamma$ belongs to $L^\infty(\mathcal{F}_{t_0})$. Upon substituting $c = c_k$ in (26), we obtain that

$$\mathbb{E}\big[e^{-k\Delta}\, \Delta \,\big|\, \mathcal{F}_{t_0^-}\big] = 0 \quad \text{almost surely, for every } k \in \mathbb{R}. \tag{27}$$

Note that, for each $k \in \mathbb{R}$, identity (27) holds outside a $\mathbb{P}$-null set that may a priori depend on k. To extract a statement uniform in k, we first restrict to a countable subset and then extend by continuity. There exists a single $\mathbb{P}$-null set $N \subset \Omega$ such that, for every $\omega \notin N$ and every rational k, $\mathbb{E}[e^{-k\Delta}\Delta \mid \mathcal{F}_{t_0^-}](\omega) = 0$. For each $k \in \mathbb{R}$ and $\omega \in \Omega$, define $\phi(k,\omega) := \mathbb{E}[e^{-k\Delta} \mid \mathcal{F}_{t_0^-}](\omega)$. For each fixed ω, the regular conditional distribution of Δ given $\mathcal{F}_{t_0^-}$ is a Borel probability measure on the bounded interval $[-\|\Delta\|_\infty, \|\Delta\|_\infty]$, and $\phi(\cdot,\omega)$ is the moment-generating function of that measure. The moment-generating function of a compactly supported probability measure is entire on $\mathbb{C}$ (and in particular real-analytic on $\mathbb{R}$); this is a standard consequence of differentiation under the integral sign, justified by the dominating bound $|\Delta^n e^{-k\Delta}| \le \|\Delta\|_\infty^n\, e^{|k|\,\|\Delta\|_\infty}$ valid for every $n \ge 0$, which gives an absolutely convergent Taylor series for $\phi(\cdot,\omega)$ on every bounded set in $\mathbb{R}$ (Kallenberg, 2021, Theorem 5.30). Its derivative

$$\partial_k \phi(k,\omega) = -\,\mathbb{E}\big[\Delta\, e^{-k\Delta} \,\big|\, \mathcal{F}_{t_0^-}\big](\omega)$$

is therefore continuous in k for each fixed $\omega \notin N$, and vanishes at every rational k; by continuity on the dense set of rationals, $\partial_k \phi(k,\omega) = 0$ for every $k \in \mathbb{R}$ and every $\omega \notin N$. We therefore have that $k \mapsto \phi(k,\omega)$ is constant on $\mathbb{R}$ for every $\omega \notin N$. Evaluating at $k = 0$ gives $\phi(0,\omega) = 1$, so $\phi(k,\omega) = 1$ for every $k \in \mathbb{R}$ and every $\omega \notin N$. Therefore, almost surely on Ω, the conditional moment-generating function of Δ given $\mathcal{F}_{t_0^-}$ equals 1 for every $k \in \mathbb{R}$. Since Δ is bounded, its conditional law given $\mathcal{F}_{t_0^-}$ (a regular conditional distribution exists since $\Delta \in L^\infty$ is real-valued, by Kallenberg (2021, Theorem 8.5)) is uniquely determined by this conditional MGF on any open interval containing 0, in particular on $(-\|\Delta\|_\infty^{-1}, \|\Delta\|_\infty^{-1})$. The unique Borel probability measure on $\mathbb{R}$ with MGF identically equal to 1 on such an interval is the Dirac mass at 0. Hence, the conditional law of Δ given $\mathcal{F}_{t_0^-}$ is almost surely the Dirac mass at 0, which implies $\mathbb{E}[\Delta^2 \mid \mathcal{F}_{t_0^-}] = 0$ almost surely, and consequently $\Delta = 0$ almost surely. We have therefore established that

$$e^{-\gamma \tilde{J}_{t_0}(W)} = \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_0}] \quad \text{almost surely.}$$

Taking $-\gamma^{-1}\log$ of both sides (valid since both sides are strictly positive and bounded away from zero by $e^{-\gamma\|W\|_\infty} > 0$), we obtain that

$$\tilde{J}_{t_0}(W) = -\frac{1}{\gamma} \log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_0}] \quad \text{almost surely.} \tag{28}$$

This holds for every $t_0 \in [0,T]$ and every $W \in L^\infty(\mathcal{F}_T)$.

Stage 2: Right-continuity in probability follows. Fix $\pi \in \Pi$, write $W := W_T^L(\pi)$, and let $t \in [0,T)$ with $t_k \downarrow t$, where $t_k \in [t,T]$. By (28),

$$J_{t_k}(\pi) = \tilde{J}_{t_k}(W) = -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_{t_k}].$$

The process $M_u := \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_u]$ is a uniformly integrable martingale; under the usual conditions on $\mathbb{F}$ (right-continuity and completeness, as in Section 2), M is almost surely right-continuous, and in particular $M_{t_k} \to M_t$ almost surely as $t_k \downarrow t$. Since $M_u \geq e^{-\gamma\|W\|_\infty} > 0$ uniformly in u, the function $x \mapsto -\gamma^{-1}\log x$ is continuous and bounded on the range of M, hence

$$J_{t_k}(\pi) = -\frac{1}{\gamma}\log M_{t_k} \longrightarrow -\frac{1}{\gamma}\log M_t = J_t(\pi) \quad \text{almost surely as } t_k \downarrow t.$$

Almost-sure convergence implies convergence in probability. This is the right-continuity property. □

A.2 Independence of the Five Core Axioms

Theorem 40 (Independence of (J1)–(J5))**.** *The five axioms (J1) (cash-additivity), (J2) (normalization), (J3) (concavity), (J4) (strong dynamic consistency), and (J5) (law-invariance) form a logically independent system: for each $k \in \{1,2,3,4,5\}$, there exists a dynamic preference functional $J^{(k)}$ on $L^\infty(\mathcal{F}_T)$ satisfying every axiom in the list except (Jk).*

Proof. In this proof we work directly with the reduced functional $\tilde{J}_t \colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$. Each separating model below is defined as a functional of $W = W_T^L(\pi)$ directly, so the wealth-summary property (W) holds by construction in each case, without requiring the general derivation of Proposition 38 (which assumes J1, J2, and J4 simultaneously). Fix any reference parameter $\gamma > 0$ and denote by J^γ the entropic certainty-equivalent on $L^\infty(\mathcal{F}_T)$, namely

$$J_t^\gamma(W) := -\frac{1}{\gamma}\log \mathbb{E}[e^{-\gamma W} \mid \mathcal{F}_t], \qquad W \in L^\infty(\mathcal{F}_T),\ t \in [0,T].$$

Note that this functional satisfies each of (J1)–(J5) (by the converse direction of Theorem 10); each separating model below is obtained as a small perturbation of J^γ designed to violate exactly one axiom. We treat the five cases in turn.

Model M_1: violates (J1), satisfies (J2)–(J5). Define $J_t^{(1)}(W) := 2J_t^\gamma(W)$.

(J1) fails. For any nonzero $\mathcal{F}_t$-measurable bounded c,

$$J_t^{(1)}(W+c) = 2J_t^\gamma(W+c) = 2J_t^\gamma(W) + 2c \neq 2J_t^\gamma(W) + c = J_t^{(1)}(W) + c.$$

(J2) holds. $J_t^{(1)}(0) = 2 \cdot 0 = 0$.

(J3) holds. We verify J3a and J3b directly at the functional level, without invoking the strategy-level reduction. By the (weak) concavity of J^γ (verified in the converse direction of Theorem 10, which uses only the explicit entropic form on L^∞ and does not depend on the strategy-level J1), for any $W, W' \in L^\infty(\mathcal{F}_T)$ and $\mathcal{F}_t$-measurable $\lambda \in [0,1]$,

$$\begin{aligned} J_t^{(1)}\big(\lambda W + (1-\lambda)W'\big) &= 2J_t^\gamma\big(\lambda W + (1-\lambda)W'\big) \\ &\geq 2\Big(\lambda J_t^\gamma(W) + (1-\lambda)J_t^\gamma(W')\Big) \\ &= \lambda J_t^{(1)}(W) + (1-\lambda)J_t^{(1)}(W'), \end{aligned}$$

giving J3a; multiplying the strict inequality J3b for J^γ (which holds since $\gamma > 0$) by $2 > 0$ preserves it and gives J3b for $J^{(1)}$.

(J4) holds. Since $J_t^{(1)}(\pi) = 2J_t^{\gamma}(\pi)$ at the strategy level (by definition of $J^{(1)}$ as twice the entropic functional), the equivalence $J_t^{(1)}(\pi) \geq J_t^{(1)}(\pi')$ a.s. iff $J_t^{\gamma}(\pi) \geq J_t^{\gamma}(\pi')$ a.s. holds at the strategy level as well as at the reduced-functional level. Strong dynamic consistency of J^{γ} in the strategy-level form (Axiom J4) therefore transfers directly to $J^{(1)}$.

(J5) holds. Law-invariance is immediate: $W \overset{d}{=} W'$ implies $\mathbb{E}[e^{-\gamma W}] = \mathbb{E}[e^{-\gamma W'}]$, so $J_0^{(1)}(W) = 2J_0^{\gamma}(W) = 2J_0^{\gamma}(W') = J_0^{(1)}(W')$.

Model M_2: violates (J2), satisfies (J1) and (J3)–(J5). Define $J_t^{(2)}(W) \coloneqq J_t^{\gamma}(W) + 1$.

(J2) fails. $J_t^{(2)}(0) = 0 + 1 = 1 \neq 0$.

(J1) holds. $J_t^{(2)}(W + c) = J_t^{\gamma}(W + c) + 1 = J_t^{\gamma}(W) + c + 1 = J_t^{(2)}(W) + c$ for any $\mathcal{F}_t$-measurable bounded c.

(J3) holds. Adding a constant preserves (strict) concavity, in both clauses.

(J4) holds. Adding the same constant to two functionals does not change orderings: $J_t^{(2)}(W) \geq J_t^{(2)}(W')$ iff $J_t^{\gamma}(W) \geq J_t^{\gamma}(W')$, and dynamic consistency of J^{γ} transfers.

(J5) holds. If $W \overset{d}{=} W'$ then $J_0^{(2)}(W) = J_0^{\gamma}(W) + 1 = J_0^{\gamma}(W') + 1 = J_0^{(2)}(W')$.

Model M_3: violates (J3), satisfies (J1), (J2), (J4), (J5). Define $J_t^{(3)}(W) \coloneqq \mathbb{E}[W \mid \mathcal{F}_t]$, the conditional expectation.

(J3) fails. For any $W, W' \in L^{\infty}(\mathcal{F}_T)$ and any $\mathcal{F}_t$-measurable $\lambda \in [0, 1]$, it holds that $J_t^{(3)}(\lambda W + (1-\lambda)W') = \lambda\, J_t^{(3)}(W) + (1-\lambda)\, J_t^{(3)}(W')$ identically, so J3a is satisfied with equality and the strict-concavity clause J3b fails for every pair (W, W') with $W - W'$ nonconstant.

(J1) holds. For $c \in L^{\infty}(\mathcal{F}_t)$, we have $\mathbb{E}[W + c \mid \mathcal{F}_t] = \mathbb{E}[W \mid \mathcal{F}_t] + c$.

(J2) holds. $\mathbb{E}[0 \mid \mathcal{F}_t] = 0$.

(J4) holds. If $\mathbb{E}[W \mid \mathcal{F}_t] \geq \mathbb{E}[W' \mid \mathcal{F}_t]$ almost surely and $s \leq t$, then by the tower property,

$$\mathbb{E}[W \mid \mathcal{F}_s] = \mathbb{E}[\mathbb{E}[W \mid \mathcal{F}_t] \mid \mathcal{F}_s] \geq \mathbb{E}[\mathbb{E}[W' \mid \mathcal{F}_t] \mid \mathcal{F}_s] = \mathbb{E}[W' \mid \mathcal{F}_s] \quad \text{almost surely.}$$

(J5) holds. If $W \overset{d}{=} W'$ then $\mathbb{E}[W] = \mathbb{E}[W']$; at $t = 0$ with $\mathcal{F}_0$ trivial, it holds that $J_0^{(3)}(W) = \mathbb{E}[W] = \mathbb{E}[W'] = J_0^{(3)}(W')$.

Model M_4: violates (J4), satisfies (J1), (J2), (J3), (J5). We construct a time-dependent entropic functional with distinct risk-aversion parameters on a partition of $[0, T]$. Choose any two distinct parameters $0 < \gamma_1 \neq \gamma_2 < \infty$ and define

$$J_t^{(4)}(W) \coloneqq \begin{cases} J_t^{\gamma_1}(W) & \text{if } t \in [0, T/2), \\ J_t^{\gamma_2}(W) & \text{if } t \in [T/2, T]. \end{cases}$$

That is, the functional uses risk-aversion γ_1 on the first half of the horizon and risk-aversion γ_2 on the second half.

(J4) fails. We exhibit $W, W' \in L^{\infty}(\mathcal{F}_T)$ and times $s = 0$, $t = T/2$ for which it holds that $J_t^{(4)}(W) \geq J_t^{(4)}(W')$ almost surely yet $J_s^{(4)}(W) < J_s^{(4)}(W')$. Without loss of generality $\gamma_1 < \gamma_2$; take $W' \in L^{\infty}(\mathcal{F}_T)$ a nondegenerate symmetric bounded random variable, independent of $\mathcal{F}_{T/2}$ (e.g., $W' = \pm a$ each with probability $1/2$ for some $a > 0$; existence follows from the non-atomicity of $(\Omega, \mathcal{F}_T, \mathbb{P})$). By the strict monotonicity of $\gamma \mapsto J_0^{\gamma}(W')$ on $(0, \infty)$ for nondegenerate W' (Sub-step 3b of the proof of Theorem 10), $J_0^{\gamma_2}(W') < J_0^{\gamma_1}(W')$, so the open interval $(J_0^{\gamma_2}(W'), J_0^{\gamma_1}(W'))$ is nonempty; let $W \coloneqq w$ be a deterministic constant in this interval. Since W' is independent of $\mathcal{F}_{T/2}$, $J_{T/2}^{(4)}(W') = J_0^{\gamma_2}(W')$ almost surely and $J_{T/2}^{(4)}(W) = w > J_0^{\gamma_2}(W') = J_{T/2}^{(4)}(W')$, so the hypothesis of (J4) holds at $t = T/2$; at $s = 0$, however, $J_0^{(4)}(W) = w < J_0^{\gamma_1}(W') = J_0^{(4)}(W')$, contradicting the conclusion.

(J1), (J2), (J3) hold. At each t, $J_t^{(4)}$ is an entropic functional at parameter γ_1 or γ_2, and entropic functionals satisfy cash-additivity, normalization, and concavity; the strict-concavity clause (J3b) is at $t = 0$ where $J^{(4)}$ uses $\gamma_1 > 0$, so J3b holds.

(J5) holds. $J_0^{(4)} = J_0^{\gamma_1}$, which is law-invariant.

Model M_5: violates (J5), satisfies (J1)–(J4). Let $\mathbb{Q}$ be a probability measure on $(\Omega, \mathcal{F})$ equivalent to $\mathbb{P}$ with Radon–Nikodým density $\varphi := d\mathbb{Q}/d\mathbb{P} \in L^\infty(\mathcal{F}_T)$, $\varphi > 0$ almost surely, $\mathbb{E}[\varphi] = 1$, and φ not $\mathbb{P}$-almost-surely constant. Define

$$J_t^{(5)}(W) := -\frac{1}{\gamma} \log \mathbb{E}^{\mathbb{Q}}[e^{-\gamma W} \mid \mathcal{F}_t],$$

the entropic certainty-equivalent computed under $\mathbb{Q}$.

(J5) fails. We construct $W, W' \in L^\infty(\mathcal{F}_T)$ with $W \overset{d}{=} W'$ under $\mathbb{P}$ but $J_0^{(5)}(W) \neq J_0^{(5)}(W')$. Take $Z \in L^\infty(\mathcal{F}_T)$ with $\mathbb{E}[Z] = 0$ and $\mathbb{E}[Z^2] > 0$, and write $\varphi = c_0 + \varepsilon Z$ for $c_0 := \mathbb{E}[\varphi] = 1$ and $\varepsilon > 0$ small enough that $\varphi > 0$. By the non-atomicity of $(\Omega, \mathcal{F}_T, \mathbb{P})$, there exists $W' \in L^\infty(\mathcal{F}_T)$ with $W' \overset{d}{=} Z$ under $\mathbb{P}$ but independent of φ; set $W := Z$. Then $W \overset{d}{=} W'$ under $\mathbb{P}$, but

$$\begin{aligned}
\mathbb{E}^{\mathbb{Q}}[e^{-\gamma W}] &= \mathbb{E}[\varphi\, e^{-\gamma Z}] = c_0\, \mathbb{E}[e^{-\gamma Z}] + \varepsilon\, \mathbb{E}[Z\, e^{-\gamma Z}], \\
\mathbb{E}^{\mathbb{Q}}[e^{-\gamma W'}] &= \mathbb{E}[\varphi]\, \mathbb{E}[e^{-\gamma W'}] = c_0\, \mathbb{E}[e^{-\gamma Z}],
\end{aligned}$$

using the independence of W' and φ in the second line. The difference is $\varepsilon\, \mathbb{E}[Z\, e^{-\gamma Z}]$, which is nonzero whenever Z is nondegenerate: setting $h(k) := \mathbb{E}[e^{-\gamma k Z}]$, we have that $h'(k) = -\gamma\, \mathbb{E}[Z\, e^{-\gamma k Z}]$, and strict convexity of h (from strict convexity of $k \mapsto e^{-\gamma k z}$ and the nondegeneracy of Z) forces h' strictly increasing with unique zero at $k = 0$ (since also $h'(0) = -\gamma\, \mathbb{E}[Z] = 0$), so $h'(1) \neq 0$. Hence, $J_0^{(5)}(W) \neq J_0^{(5)}(W')$.

(J1), (J2), (J3) hold. For $c \in L^\infty(\mathcal{F}_t)$, $\mathbb{E}^{\mathbb{Q}}[e^{-\gamma(W+c)} \mid \mathcal{F}_t] = e^{-\gamma c}\, \mathbb{E}^{\mathbb{Q}}[e^{-\gamma W} \mid \mathcal{F}_t]$ (since c factors out under $\mathbb{Q}$-conditional expectation), giving $J_t^{(5)}(W + c) = J_t^{(5)}(W) + c$, as well as $J_t^{(5)}(0) = -\gamma^{-1} \log 1 = 0$; strict concavity in W follows from strict convexity of $x \mapsto e^{-\gamma x}$ and Jensen's inequality applied under $\mathbb{Q}$.

(J4) holds. By the tower property under $\mathbb{Q}$, $\mathbb{E}^{\mathbb{Q}}\big[e^{-\gamma J_t^{(5)}(W)} \,\big|\, \mathcal{F}_s\big] = \mathbb{E}^{\mathbb{Q}}[e^{-\gamma W} \mid \mathcal{F}_s]$ for $s \leq t$, which combined with (J1) and (J2) implies strong dynamic consistency: $J_t^{(5)}(W) \geq J_t^{(5)}(W')$ a.s. gives $e^{-\gamma J_t^{(5)}(W)} \leq e^{-\gamma J_t^{(5)}(W')}$ a.s., and taking $\mathbb{Q}$-conditional expectation given $\mathcal{F}_s$ and applying $-\gamma^{-1} \log$ yields $J_s^{(5)}(W) \geq J_s^{(5)}(W')$ a.s.

This exhibits, for each $k \in \{1, 2, 3, 4, 5\}$, a functional satisfying every axiom except (Jk), establishing logical independence. □

Remark 41 (On the Derived Properties)**.** The three derived properties ((M) monotonicity, (W) wealth-summary, (R) right-continuity) are consequences of (J1)–(J5), but individual separating models can satisfy some of them despite violating a core axiom. The conditional-expectation model M_3 (violating (J3)) satisfies (M) and (W) by direct calculation, and satisfies (R) as well, since $t \mapsto \mathbb{E}[W \mid \mathcal{F}_t]$ is a uniformly integrable martingale with an almost surely right-continuous modification under the usual conditions on $\mathbb{F}$. The proof of (R) in Proposition 39 uses all of (J1)–(J5), since it first establishes the entropic representation and then derives right-continuity, but this does not mean every nonentropic functional fails (R); a precise analysis of which derived properties survive each axiom violation is straightforward and we omit it.

B Technical Conditions on the Strategy Space

In this appendix we record the technical conditions on the strategy space Π which justify the reduction step in the proof of Theorem 10. The conditions are essentially standard for the

Avellaneda–Stoikov framework and its extensions; a reader content with informal statements may safely skip this appendix. Its key conclusion, used in Step 1 of the proof of Theorem 10, is the following.

Lemma 42 (Well-Definedness of the Reduced Functional)**.** *Under the admissibility conditions of Definition 43 below, for every $\pi \in \Pi$ the liquidation-adjusted terminal wealth $W_T^L(\pi)$ is in $L^\infty(\mathcal{F}_T)$, and there is a functional $\tilde{J}_t \colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$ such that $J_t(\pi) = \tilde{J}_t(W_T^L(\pi))$ for every $\pi \in \Pi$. The functional $\tilde{J}_t$ satisfies the dynamic risk measure axioms checked in Section 4.2, Step 2, if and only if J satisfies Axioms J1–J5.*

The proof occupies the rest of this appendix and proceeds in three short steps.

B.1 Admissibility

We work with the setup of Section 2, which we briefly recall. The mid-price is $dS_t = \sigma_t \, dB_t$ with σ predictable and locally bounded; we set $d\langle S \rangle_t = \sigma_t^2 \, dt$. Order flow is described by the Cox processes N^a, N^b with intensities λ_t^a, λ_t^b which depend on the strategy and on relevant state variables, and which we take to be predictable and locally integrable. The quote distances δ^a, δ^b are predictable processes taking values in $\mathbb{R}$.

Definition 43 (Admissible Strategy)**.** A strategy $\pi = (\delta^a, \delta^b, C)$ is *admissible* if all of the following hold.

1. *Predictability and boundedness of quotes.* The processes δ^a, δ^b are predictable, and there exists a deterministic constant $\bar{\delta} = \bar{\delta}(\pi) > 0$ such that $|\delta_t^a|, |\delta_t^b| \le \bar{\delta}$ for all $t \in [0, T]$ almost surely.

2. *Inventory bound.* For each $\pi \in \Pi$ there exists a deterministic constant $Q_\pi > 0$ such that $|q_t(\pi)| \le Q_\pi$ for all $t \in [0, T]$ almost surely.

3. *Mid-price bound.* The mid-price satisfies $\sup_{t \in [0,T]} |S_t| \le M$ almost surely, where M is the deterministic stopping constant fixed in Section 2.1.

4. *Arrival-count bound.* There exists a deterministic constant $\bar{N} = \bar{N}(\pi) > 0$ such that $N_T^a + N_T^b \le \bar{N}$ almost surely.

5. *Cash-injection bound.* The process C is $\mathcal{F}$-adapted, càdlàg, of finite variation, and there exists a deterministic constant $\bar{C} = \bar{C}(\pi) > 0$ such that the total variation satisfies $|C|_T \le \bar{C}$ almost surely. (By convention, C_0 contributes to the initial cash position together with X_0, so the value C_0 is allowed to be any deterministic real number. Without loss of generality we may absorb C_0 into X_0 and assume $C_0 = 0$ when convenient.)

Only the price bound M is fixed globally (by the stopping convention of Section 2.1); the other constants are per-strategy as indicated. The space Π is the set of all admissible strategies, viewed as a convex subset of the space of $(\mathbb{R}^2 \times \mathrm{BV}([0, T]))$-valued predictable processes.

Each of the first four conditions of Definition 43 admits a practical reading. The quote-distance bound corresponds to the fact that no real market maker quotes arbitrarily far from the mid-price; quote distances are bounded by the width of the order book at any reasonable depth. The inventory bound corresponds to the position limit imposed by the desk or the risk officer; every realistic intraday market-making operation has a hard cap on $|q|$. The mid-price bound is automatic by the stopping convention of Section 2.1, with $|S_t| \le M$ holding by construction. The arrival-count bound is the one technical condition that is not standard in the AS literature. We discuss it in Remark 44 below.

Remark 44 (On the Arrival-Count Bound)**.** Condition (4) of Definition 43 is the only condition that is not directly inherited from the standard AS/CJ framework. The reading is entirely practical: every actual market maker faces hardware and exchange throughput constraints which deterministically bound the number of fills per unit time, so a horizon-T deterministic cap $\bar{N}$ is innocuous in practice. The bound $\bar{N}$ is strategy-dependent (as are $\bar{\delta}, Q, \bar{C}$): a strategy with narrower quote distances induces higher fill rates and therefore a larger $\bar{N}$. A global cap on $\bar{N}$ independent of the strategy would correspond to a hardware or exchange throughput constraint; the weaker per-strategy bound suffices for all our results. Theoretically, the bound can be relaxed at the cost of working with a weighted L^p space rather than L^∞, in which case the Kupper–Schachermayer theorem requires the corresponding extension to general Orlicz spaces (Cheridito and Li, 2008). The structural conclusion of Theorem 10 is unchanged under that relaxation, with γ continuing to parametrize the forced family. The technical bookkeeping is, however, considerably heavier, and we have chosen the L^∞ route in this paper for transparency. We expect that a careful Orlicz-space treatment can be carried out by methods that are parallel to those of Cheridito and Li (2008).

The inventory bound and the arrival-count bound (Conditions 2 and 4) together imply that q is a bounded adapted process whose increments are governed by a Cox arrival process with deterministically bounded total count over $[0, T]$. We emphasize that the intensities λ^a, λ^b themselves need not be bounded, but the resulting cumulative counts N_T^a, N_T^b are.

B.2 Boundedness of Liquidation-Adjusted Wealth

The next step is to verify that, under admissibility, $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$ for every $\pi \in \Pi$.

Lemma 45. *For $\pi \in \Pi$, the liquidation-adjusted terminal wealth $W_T^L(\pi)$ is essentially bounded, with a deterministic essential bound depending only on $|X_0|$, M, $\bar{\delta}$, $\bar{N}$, $\bar{C}$, Q, and L.*

Proof. Decompose $W_T^L(\pi) = X_T + q_T S_T - L(q_T)$ and bound each term. The cash term satisfies

$$|X_T| \le |X_0| + \int_0^T |S_t + \delta_t^a| \, dN_t^a + \int_0^T |S_t - \delta_t^b| \, dN_t^b + |C|_T \le |X_0| + (M + \bar{\delta}) \, \bar{N} + \bar{C},$$

using Conditions (1), (3), (4), and (5); the inventory term $q_T S_T$ is bounded by QM in absolute value by (2) and (3); and $|L(q_T)| \le \max_{|q| \le Q} |L(q)|$, a deterministic constant by convexity of L and (2). Combining,

$$|W_T^L(\pi)| \le |X_0| + (M + \bar{\delta}) \, \bar{N} + \bar{C} + QM + \max_{|q| \le Q} |L(q)|,$$

which is deterministic, so $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$. □

B.3 Extension to $L^\infty(\mathcal{F}_T)$

We now construct the functional $\tilde{J}_t$ on $L^\infty(\mathcal{F}_T)$ and verify the dynamic risk measure axioms.

By property (W) (Proposition 7) (inventory-via-liquidation), the value $J_t(\pi)$ depends on π only through $W_T^L(\pi)$. Define

$$\mathcal{R} := \{W_T^L(\pi) \mid \pi \in \Pi\} \subset L^\infty(\mathcal{F}_T),$$

the range of the liquidation-adjusted-wealth map. Using property (W) (Proposition 7), the assignment

$$\tilde{J}_t \colon \mathcal{R} \to L^\infty(\mathcal{F}_t), \quad \tilde{J}_t\big(W_T^L(\pi)\big) := J_t(\pi),$$

is well-defined. The remaining question is whether $\tilde{J}_t$ extends to all of $L^\infty(\mathcal{F}_T)$.

Lemma 46 (Density and Extension)**.** *Define* $\mathcal{R} := \{W_T^L(\pi) \mid \pi \in \Pi\}$. *Then,* $\mathcal{R} = L^\infty(\mathcal{F}_T)$. *In particular,* $\mathcal{R}$ *is closed under addition by* $\mathcal{F}_t$*-measurable bounded random variables for every* $t \in [0,T]$*, and contains all real constants. The functional* $\tilde{J}_t \colon \mathcal{R} \to L^\infty(\mathcal{F}_t)$ *is therefore defined on all of* $L^\infty(\mathcal{F}_T)$ *for every* $t \in [0,T]$.

Proof. The proof proceeds in three steps.

Step 1: $\mathcal{R}$ *is closed under bounded* $\mathcal{F}_t$*-measurable shifts.* Let $W = W_T^L(\pi) \in \mathcal{R}$ for some $\pi = (\delta^a, \delta^b, C) \in \Pi$, fix $t \in [0,T]$, and let $c \in L^\infty(\mathcal{F}_t)$ with $\|c\|_\infty \le C_*$. Then define $\pi^{[c]} = (\delta^a, \delta^b, C^{[c]})$ with $C_s^{[c]} := C_s + c \cdot \mathbf{1}_{\{s \ge t\}}$; since c is $\mathcal{F}_t$-measurable, $C^{[c]}$ is $\mathcal{F}$-adapted and càdlàg with $|C^{[c]}|_T \le \bar{C} + C_*$ deterministic. Conditions (1)–(4) of Definition 43 depend only on quote-distance, inventory, price, and order-flow processes, none of which is changed by the cash injection, so they hold for $\pi^{[c]}$ with the same constants as π; Condition (5) holds with $\bar{C} + C_*$, giving $\pi^{[c]} \in \Pi$. Inserting $C^{[c]}$ into (1) yields $X_T(\pi^{[c]}) = X_T(\pi) + c$, while $q_T(\pi^{[c]}) = q_T(\pi)$ and the liquidation cost are unchanged, so $W_T^L(\pi^{[c]}) = W + c \in \mathcal{R}$.

Step 2: $\mathcal{R}$ *contains all real constants.* The strategy space Π is nonempty, since for any constant choice $(\delta^a, \delta^b) \equiv (\bar{\delta}, \bar{\delta})$ and $C \equiv 0$, all five admissibility conditions hold. Pick any $\pi_0 \in \Pi$ and let $W_0 := W_T^L(\pi_0) \in \mathcal{R}$. For any $c \in \mathbb{R}$, applying Step 1 at $t = T$ with shift $c - W_0 \in L^\infty(\mathcal{F}_T)$ (this is bounded since W_0 is essentially bounded by Lemma 45) yields the strategy $\pi_0^{[c-W_0]} \in \Pi$ with $W_T^L(\pi_0^{[c-W_0]}) = W_0 + (c - W_0) = c$ almost surely. Hence, every real constant c lies in $\mathcal{R}$.

Step 3: Full equality $\mathcal{R} = L^\infty(\mathcal{F}_T)$. Fix $W \in L^\infty(\mathcal{F}_T)$ and pick $\pi_0^{(0)} \in \Pi$ from Step 2 with $W_T^L(\pi_0^{(0)}) = 0$; applying Step 1 at $t = T$ with shift $c = W$ yields $\pi_0^{(0)[W]} \in \Pi$ with $W_T^L(\pi_0^{(0)[W]}) = W$. The Condition (5) constant of the resulting strategy is $\bar{C} + \|W\|_\infty$, consistent with Definition 43 allowing $\bar{C}$ to depend on the strategy, so $\mathcal{R} = L^\infty(\mathcal{F}_T)$.

By property (W), $J_t(\pi)$ depends on π only through $W_T^L(\pi)$, so $\tilde{J}_t \colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$ is well-defined; by Axiom J1, $\tilde{J}_t(W + c) = \tilde{J}_t(W) + c$ for $W \in L^\infty(\mathcal{F}_T)$ and bounded $c \in L^\infty(\mathcal{F}_t)$, and $\rho_t := -\tilde{J}_t$ is the object on which Kupper and Schachermayer (2009, Theorem 1.10) applies directly. □

Remark 47 (On the Role of the Cash-Injection Component)**.** The cash-injection component of the strategy space is a *technical device*, included to ensure that the reduced functional $\tilde{J}_t$ is defined on all of $L^\infty(\mathcal{F}_T)$, the domain on which the Kupper–Schachermayer representation theorem applies; it is not a modeling claim about how desks operate. Concretely, the strategies $\pi_0^{(0)[W]}$ constructed in Step 3 realize any bounded $\mathcal{F}_T$-measurable W via a single cash injection of size W at time T: the market maker quotes constantly at $(\bar{\delta}, \bar{\delta})$ and injects W into her cash account at time T. The forced entropic form and the unique scalar γ produced by Theorem 10 are unaffected by whether such strategies are economically meaningful.

C The Multi-Asset Extension

In this appendix we prove the multi-asset claim of Section 7.4.

Multi-asset setup. The probability space $(\Omega, \mathcal{F}, \mathbb{F}, \mathbb{P})$ is as in Section 2. The market consists of $K \ge 1$ assets with price processes $S^{(1)}, \dots, S^{(K)}$, each a continuous semimartingale adapted to $\mathbb{F}$, and per-asset cumulative ask and bid fills $N^{a,(k)}, N^{b,(k)}$. The strategy space $\Pi^{(K)}$ satisfies the natural multi-asset analogue of the admissibility conditions of Appendix B, with a single shared cash account. Cash and per-asset inventory evolve as

$$dX_t = \sum_{k=1}^{K} \big[(S_t^{(k)} + \delta_t^{a,(k)})\, dN_t^{a,(k)} - (S_t^{(k)} - \delta_t^{b,(k)})\, dN_t^{b,(k)} \big], \qquad dq_t^{(k)} = dN_t^{b,(k)} - dN_t^{a,(k)}.$$

The joint liquidation cost function $L\colon \mathbb{R}^K \to \mathbb{R}_+$ is convex with $L(0) = 0$, and the joint liquidation-adjusted terminal wealth is

$$W_T^L(\pi) = X_T(\pi) + \sum_{k=1}^{K} q_T^{(k)}(\pi) S_T^{(k)} - L\big(q_T^{(1)}(\pi), \dots, q_T^{(K)}(\pi)\big). \tag{29}$$

The preference functional $J^{(K)} = (J_t^{(K)})_{t\in[0,T]}$ with $J_t^{(K)}\colon \Pi^{(K)} \to L^\infty(\mathcal{F}_t)$ is assumed to satisfy the multi-asset analogues of Axioms J1–J5, with Π, J_t, W_T^L replaced by their multi-asset counterparts throughout.

Theorem 48 (Multi-Asset Forced Uniqueness)**.** *Under the setup above, $J^{(K)}$ satisfies the multi-asset analogues of Axioms J1–J5 if and only if there exists a unique $\gamma > 0$ such that*

$$J_t^{(K)}(\pi) = -\frac{1}{\gamma} \log \mathbb{E}\Big[\exp\big(-\gamma\, W_T^L(\pi)\big) \,\Big|\, \mathcal{F}_t\Big] \tag{30}$$

for all $t \in [0,T]$ and all $\pi \in \Pi^{(K)}$, with the same γ governing all K assets.

Proof. By the multi-asset analogue of Proposition 7 (derived from the multi-asset analogues of Axioms J1, J2, J4 exactly as in the single-asset case), $J_t^{(K)}(\pi)$ depends on π only through the scalar random variable $W_T^L(\pi) \in L^\infty(\mathcal{F}_T)$ defined by (29). There exists therefore a functional $\tilde{J}_t^{(K)}$ such that

$$J_t^{(K)}(\pi) = \tilde{J}_t^{(K)}\big(W_T^L(\pi)\big) \quad \text{for all } \pi \in \Pi^{(K)}.$$

The multi-asset analogue of Lemma 46, which exploits the shared cash account and is otherwise identical in structure to the single-asset version, gives $\{W_T^L(\pi) \mid \pi \in \Pi^{(K)}\} = L^\infty(\mathcal{F}_T)$, so $\tilde{J}_t^{(K)}\colon L^\infty(\mathcal{F}_T) \to L^\infty(\mathcal{F}_t)$ is defined on the full space; the role of the shared cash account is essential here, since a single bounded $\mathcal{F}_T$-measurable cash injection c at time T shifts the joint wealth W_T^L by c regardless of the per-asset inventories, so the range of W_T^L covers all of $L^\infty(\mathcal{F}_T)$. With K separate cash accounts, the corresponding density statement would not hold. Each multi-asset axiom on $J^{(K)}$ transfers to $\tilde{J}^{(K)}$ as a property of a functional on a single scalar random variable, and by the single-asset argument of Theorem 10 applied to $\tilde{J}^{(K)}$, there exists a unique $\gamma \in (0,\infty)$ such that

$$\tilde{J}_t^{(K)}(W) = -\tfrac{1}{\gamma} \log \mathbb{E}\big[e^{-\gamma W} \,\big|\, \mathcal{F}_t\big] \quad \text{for all } W \in L^\infty(\mathcal{F}_T),$$

and setting $W = W_T^L(\pi)$ yields (30). The same γ governs all K assets because the multi-asset structure is entirely absorbed into the scalar W_T^L before the Kupper–Schachermayer representation is applied; the converse is identical to that of Theorem 10. □

A market-making desk that trades K correlated assets has, under the multi-asset axioms, a single risk-aversion parameter governing all of them, and cross-asset correlations, liquidation costs, and volatilities enter through the joint distribution of W_T^L, not through any per-asset preference parameter. A desk implementing different risk aversions on different sub-books is operating two distinct preference functionals which cannot both arise from a single dynamically-consistent agent; the falsifiability criterion of Remark 18 sharpens accordingly, in that the implicit γ recovered from each sub-book via Corollary 25 should agree across sub-books.

Proposition 49 (Multi-Asset Forced Running Coefficients)**.** *Under the same hypotheses as Theorem 48, with constant joint covariance Σ of the price processes, the per-asset CJ-style running penalty matrix $\Phi = (\Phi_{ij})_{i,j=1}^K$ for the second-order expansion of the multi-asset entropic functional around zero inventory is forced to take the form*

$$\Phi_{ij} = \frac{\gamma \Sigma_{ij}}{2},$$

i.e., the running inventory cost is $\frac{\gamma}{2} q^\top \Sigma\, q\, dt$ where $q = (q^{(1)}, \dots, q^{(K)})$.

Proof. The argument is a multi-asset Taylor expansion of (30) around zero inventory, parallel to Proposition 32. Fix t and $q = q_t \in \mathbb{R}^K$, and consider $[t, t+dt]$ over which inventory is held constant; the mark-to-market wealth increment is $dW = q^\top dS_t$, where S has joint quadratic covariation $d\langle S\rangle_t = \Sigma\, dt$. By the Bellman identity of the entropic functional, the conditional certainty-equivalent contribution equals

$$-\tfrac{1}{\gamma}\log\mathbb{E}\Big[e^{-\gamma\, q^\top dS_t}\,\Big|\,\mathcal{F}_t\Big] = -\tfrac{1}{\gamma}\log\Big(1+\tfrac{\gamma^2}{2}\,q^\top\Sigma\, q\, dt + o(dt)\Big)\,,$$

using the zero-drift convention of Section 2.1 (a nonzero drift $\mu_t\, dt$ contributes only to the linear-in-q alpha term, not to the quadratic-in-q running cost). Expanding $-\frac{1}{\gamma}\log(1+x) = -\frac{1}{\gamma}x + o(x)$, the leading-order quadratic-in-q contribution as $dt \to 0$ is $-\frac{\gamma}{2}\, q^\top\Sigma\, q\, dt$, the running-penalty contribution with matrix coefficient $\Phi = \frac{\gamma}{2}\Sigma$; the argument is uniform in q and applies at every t. □

Remark 50 (Stochastic-Covariance Generalization)**.** The stochastic-covariance generalization, with Σ_t predictable and a.s. positive semidefinite, follows by combining the multi-asset Taylor expansion above with the time-change argument of Proposition 36 applied to the canonical multi-asset clock; the forced matrix coefficient is then $\Phi_{ij,t} = \frac{\gamma\,\Sigma_{ij,t}}{2}$, with a single time-invariant scalar γ. The argument is structurally identical to that of Proposition 36, applied componentwise to $d\langle S^{(i)}, S^{(j)}\rangle_t = \Sigma_{ij,t}\, dt$.

Declarations

Conflict of interest. The author declares no conflicts of interest.

Funding. No funding was received for this work.

Data availability. No data were used in this study.

References

Abernethy, J., Y. Chen, and J. W. Vaughan (2013). Efficient Market Making via Convex Optimization, and a Connection to Online Learning. *ACM Transactions on Economics and Computation* 1(2), 12:1–12:39.

Abernethy, J., R. Frongillo, and S. Kutty (2014). On Risk Measures, Market Making, and Exponential Families. *ACM SIGecom Exchanges* 13(2), 21–25.

Acciaio, B. and I. Penner (2011). Dynamic Risk Measures. In *Advanced Mathematical Methods for Finance*, pp. 1–34. Springer.

Al Dayri, K., E. Bacry, and J.-F. Muzy (2011). The Nature of Price Returns During Periods of High Market Activity. In F. Abergel, B. K. Chakrabarti, A. Chakraborti, and M. Mitra (Eds.), *Econophysics of Order-driven Markets*, New Economic Windows, pp. 155–172. Springer, Milano.

Almgren, R. and N. Chriss (2000). Optimal Execution of Portfolio Transactions. *Journal of Risk* 3(2), 5–39.

Ané, T. and H. Geman (2000). Order Flow, Transaction Clock, and Normality of Asset Returns. *Journal of Finance* 55(5), 2259–2284.

Artzner, P., F. Delbaen, J.-M. Eber, and D. Heath (1999). Coherent Measures of Risk. *Mathematical Finance* 9(3), 203–228.

Artzner, P., F. Delbaen, J.-M. Eber, D. Heath, and H. Ku (2007). Coherent multiperiod risk adjusted values and Bellman's principle. *Annals of Operations Research* 152(1), 5–22.

Avellaneda, M. and S. Stoikov (2008). High-frequency trading in a limit order book. *Quantitative Finance* 8(3), 217–224.

Backhoff Veraguas, J. and L. Tangpi (2018). On the dynamic representation of some time-inconsistent risk measures in a Brownian filtration. *arXiv preprint* arXiv:1608.07498.

Barger, W. and M. Lorig (2018). Optimal liquidation under stochastic price impact. *arXiv preprint* arXiv:1804.04170.

Barzykin, A., P. Bergault, O. Guéant, and M. Lemmel (2025). Optimal Quoting under Adverse Selection and Price Reading. *arXiv preprint* arXiv:2508.20225.

Basak, S. and G. Chabakauri (2010). Dynamic Mean-Variance Asset Allocation. *Review of Financial Studies* 23(8), 2970–3016.

Basel Committee on Banking Supervision (2019). *Minimum capital requirements for market risk.* Bank for International Settlements, Basel, January 2019.

Beiglböck, M., S. M. Pesenti, and M. Sylvestre (2026). Adapted Law Invariance and Time-Consistent Dynamic Risk Measures. *arXiv preprint* arXiv:2607.04392.

Bergault, P., D. Evangelista, O. Guéant, and D. Vieira (2021). Closed-form Approximations in Multi-asset Market Making. *Applied Mathematical Finance* 28(2), 101–142.

Bion-Nadal, J. (2009). Time consistent dynamic risk processes. *Stochastic Processes and their Applications* 119(2), 633–654.

Björk, T. and A. Murgoci (2014). A theory of Markovian time-inconsistent stochastic control in discrete time. *Finance and Stochastics* 18(3), 545–592.

Brémaud, P. (1981). *Point Processes and Queues: Martingale Dynamics.* Springer Series in Statistics. Springer-Verlag, New York.

Carr, P. and L. Wu (2004). Time-changed Lévy processes and option pricing. *Journal of Financial Economics* 71(1), 113–141.

Cartea, Á., R. Donnelly, and S. Jaimungal (2017). Algorithmic Trading with Model Uncertainty. *SIAM Journal on Financial Mathematics* 8(1), 635–671.

Cartea, Á. and S. Jaimungal (2015). Risk Metrics and Fine Tuning of High-Frequency Trading Strategies. *Mathematical Finance* 25(3), 576–611. Originally circulated as a 2010 working paper.

Cartea, Á., S. Jaimungal, and J. Penalva (2015). *Algorithmic and High-Frequency Trading.* Cambridge University Press, Cambridge.

Cerreia-Vioglio, S., F. Maccheroni, M. Marinacci, and L. Montrucchio (2011). Risk Measures: Rationality and Diversification. *Mathematical Finance* 21(4), 743–774.

Chen, Y. and D. M. Pennock (2007). A Utility Framework for Bounded-Loss Market Makers. In *Proceedings of the Twenty-Third Conference on Uncertainty in Artificial Intelligence (UAI)*, pp. 49–56.

Cheridito, P. and T. Li (2008). Dual characterization of properties of risk measures on Orlicz hearts. *Mathematics and Financial Economics* 2(1), 29–55.

Cheridito, P., F. Delbaen, and M. Kupper (2006). Dynamic Monetary Risk Measures for Bounded Discrete-Time Processes. *Electronic Journal of Probability* 11, 57–106.

Cheridito, P. and M. Stadje (2009). Time-inconsistency of VaR and time-consistent alternatives. *Finance Research Letters* 6(1), 40–46.

Cherny, A. and D. Madan (2009). New Measures for Performance Evaluation. *Review of Financial Studies* 22(7), 2571–2606.

Clark, P. K. (1973). A Subordinated Stochastic Process Model with Finite Variance for Speculative Prices. *Econometrica* 41(1), 135–155.

Delbaen, F., S. Peng, and E. Rosazza Gianin (2010). Representation of the penalty term of dynamic concave utilities. *Finance and Stochastics* 14(3), 449–472.

Drapeau, S. and M. Kupper (2013). Risk Preferences and Their Robust Representation. *Mathematics of Operations Research* 38(1), 28–62.

Föllmer, H. and A. Schied (2016). *Stochastic Finance: An Introduction in Discrete Time.* Fourth edition. De Gruyter Studies in Mathematics 27. De Gruyter, Berlin.

Föllmer, H. and M. Schweizer (1991). Hedging of Contingent Claims under Incomplete Information. In M. H. A. Davis and R. J. Elliott (Eds.), *Applied Stochastic Analysis*, Stochastics Monographs, Vol. 5, pp. 389–414. Gordon and Breach, New York.

Frittelli, M. (2000). The Minimal Entropy Martingale Measure and the Valuation Problem in Incomplete Markets. *Mathematical Finance* 10(1), 39–52.

Gilboa, I. and D. Schmeidler (1989). Maxmin expected utility with non-unique prior. *Journal of Mathematical Economics* 18(2), 141–153.

Guéant, O. (2017). Optimal market making. *Applied Mathematical Finance* 24(2), 112–154.

Hanson, R. (2003). Combinatorial Information Market Design. *Information Systems Frontiers* 5(1), 107–119.

Hanson, R. (2007). Logarithmic Market Scoring Rules for Modular Combinatorial Information Aggregation. *Journal of Prediction Markets* 1(1), 3–15.

Ho, T. and H. R. Stoll (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics* 9(1), 47–73.

Jouini, E., W. Schachermayer, and N. Touzi (2006). Law invariant risk measures have the Fatou property. *Advances in Mathematical Economics* 9, 49–71.

Kallenberg, O. (2021). *Foundations of Modern Probability.* Third edition. Probability Theory and Stochastic Modelling 99. Springer, Cham.

Kupper, M. and W. Schachermayer (2009). Representation results for law invariant time consistent functions. *Mathematics and Financial Economics* 2(3), 189–210.

Kusuoka, S. (2001). On law invariant coherent risk measures. *Advances in Mathematical Economics* 3, 83–95.

Law, B. and F. Viens (2019). Market making under a weakly consistent limit order book model. *High Frequency* 2(3–4), 215–238.

Lu, X. and F. Abergel (2018). Order-Book Modeling and Market Making Strategies. *Market Microstructure and Liquidity* 4(1n2), 1950003.

Maccheroni, F., M. Marinacci, and A. Rustichini (2006). Ambiguity Aversion, Robustness, and the Variational Representation of Preferences. *Econometrica* 74(6), 1447–1498.

Markowitz, H. (1952). Portfolio Selection. *Journal of Finance* 7(1), 77–91.

Merton, R. C. (1971). Optimum consumption and portfolio rules in a continuous-time model. *Journal of Economic Theory* 3(4), 373–413.

Moallemi, C. C. and K. Yuan (2016). A Model for Queue Position Valuation in a Limit Order Book. Working paper, Columbia Business School Research Paper No. 17-70.

Pratt, J. W. (1964). Risk Aversion in the Small and in the Large. *Econometrica* 32(1–2), 122–136.

Rosenbaum, M. and J. Zhang (2022). Multi-asset market making under the quadratic rough Heston. *arXiv preprint* arXiv:2212.10164.

Shen, Y., Z. Van Oosten, and R. Wang (2026). Partial Law Invariance and Risk Measures. *Management Science* 72(7), 6251–6266. DOI: 10.1287/mnsc.2024.06518.

Strotz, R. H. (1955). Myopia and Inconsistency in Dynamic Utility Maximization. *Review of Economic Studies* 23(3), 165–180.

Weber, S. (2006). Distribution-Invariant Risk Measures, Information, and Dynamic Consistency. *Mathematical Finance* 16(2), 419–441.