625.721 Probability Course Project
Foundations of Tail Risk and GANs

Toshiaki Yamanaka

November 11, 2025

1 Introduction

As pointed out by Cont et al. [1], with the development of deep learning techniques, *generative models* based on machine learning have emerged to simulate patterns extracted from complex, high-dimensional datasets. In particular, Goodfellow et al. [2] introduced *Generative Adversarial Networks* (GANs), an adversarial framework for estimating generative models in which two models are trained simultaneously: a generator G that captures the data distribution, and a discriminator D that estimates the probability that a sample originates from the training data rather than from G. GANs have been used in various studies, including the simulation of financial markets (*e.g.*, Takahashi et al. [3], Wiese et al. [4], Vuletić et al. [5], and Vuletić and Cont [6]).

In this paper, we focus on the work of Cont et al. [1] on tail risk and GANs, providing alternative proofs for some of their results. Specifically, we present a calculus-based proof of the strict consistency of the Acerbi–Székely [7] score (Section 2) and offer proofs of a *simplified* version of Theorem 3.8 of Cont et al. [1] from both optimization and measure-theoretic perspectives (Section 3).

2 Strict consistency of the Acerbi–Székely score

Value-at-Risk (VaR) and Expected Shortfall (ES) are fundamental tail risk measures in financial mathematics. VaR_α represents the α-quantile of the loss distribution, while ES_α captures the expected loss conditional on exceeding VaR_α. Accurate estimation and backtesting of these measures require proper scoring rules. A key challenge is that, while VaR is elicitable, ES is not individually elicitable (Gneiting [8]). However, Fissler and Ziegel [9] demonstrated that the pair (VaR_α, ES_α) is jointly elicitable, enabling consistent joint estimation and backtesting.

In this section, we outline the Fissler–Ziegel framework for joint elicitability, specialize it to the Acerbi–Székely [7] score, and provide a calculus-based proof of its strict consistency. This score serves as the foundation for the discriminator loss in Tail-GAN's adversarial training objective (Cont et al. [1]).

Definition 2.1 (Gneiting [8], p. 749)**.** Let the interval I be the potential range of the outcomes, such as $I = \mathbb{R}$ for a real-valued quantity, or $I = (0, \infty)$ for a strictly positive quantity. Let the probability distribution F be concentrated on I. Then,

1. a *scoring function* is any mapping $S : I \times I \to [0, \infty)$, and

2. a *functional* is a potentially set-valued mapping $F \mapsto T(F) \subseteq I$.

3. The scoring function S is called *consistent* for the functional T if,

$$\mathbb{E}_F[S(t, Y)] \leq \mathbb{E}_F[S(x, Y)]$$

for F, all $t \in T(F)$, and all $x \in I$.

4. S is *strictly consistent* if it is consistent and equality of the expectations implies that $x \in T(F)$.

Definition 2.2 (Elicitability, Gneiting [8], Definition 2)**.** A functional T is *elicitable* relative to the class of probability distributions $\mathcal{F}$ if there exists a scoring function S that is strictly consistent for T relative to $\mathcal{F}$.

Fissler and Ziegel [9] characterized all strictly consistent scoring functions for $(\mathrm{VaR}_\alpha,\ \mathrm{ES}_\alpha)$.

Proposition 2.3 (Fissler–Ziegel [9], Theorem 5.2; Cont et al. [1], Proposition 2.1)**.** *Assume* $\int |x|\mu(dx) < \infty$*. If* $H_2 : \mathbb{R} \to \mathbb{R}$ *is strictly convex and* $H_1 : \mathbb{R} \to \mathbb{R}$ *is such that*

$$v \mapsto R_\alpha(v, e) := \frac{1}{\alpha} v H_2'(e) + H_1(v)$$

is strictly increasing for each $e \in \mathbb{R}$*, then the score function*

$$S_\alpha(v, e, x) = (\mathbb{1}_{\{x \leq v\}} - \alpha)(H_1(v) - H_1(x)) + \frac{1}{\alpha} H_2'(e) \mathbb{1}_{\{x \leq v\}}(v - x) + H_2'(e)(e - v) - H_2(e) \quad (1)$$

is strictly consistent for $(\mathrm{VaR}_\alpha(\mu), \mathrm{ES}_\alpha(\mu))$*; that is,*

$$(\mathrm{VaR}_\alpha(\mu), \mathrm{ES}_\alpha(\mu)) = \arg \min_{(v,e) \in \mathbb{R}^2} \int S_\alpha(v, e, x)\mu(dx).$$

The Acerbi–Székely specialization (Cont et al. [1], Eq. (7)). Acerbi and Székely [7] proposed a specific choice of H_1 and H_2 that leads to a numerically stable score function:

$$H_1(v) = -\frac{W_\alpha}{2} v^2, \quad H_2(e) = \frac{\alpha}{2} e^2, \quad \text{with } \frac{\mathrm{ES}_\alpha(\mu)}{\mathrm{VaR}_\alpha(\mu)} \geq W_\alpha \geq 1.$$

Substituting into Eq. (1) yields the *Acerbi–Székely score*:

$$S_\alpha(v, e, x) = \frac{W_\alpha}{2}(\mathbb{1}_{\{x \leq v\}} - \alpha)(x^2 - v^2) + \mathbb{1}_{\{x \leq v\}} e(v - x) + \alpha e \left(\frac{e}{2} - v\right).$$

The score satisfies the conditions in Proposition 2.3 on $\{(v, e) \in \mathbb{R}^2 : W_\alpha v \leq e \leq v \leq 0\}$, ensuring strict consistency (Cont et al. [1]).

We provide our own proof of the strict consistency of S_α for $(\mathrm{VaR}_\alpha,\ \mathrm{ES}_\alpha)$, based on calculus and viewed from an optimization perspective.

Theorem 2.4 (Cont et al. [1], Eq. (7))**.** *The Acerbi–Székely score*

$$S_\alpha(\hat{v}, \hat{e}; x) = \frac{W_\alpha}{2}(\mathbb{1}_{\{x \leq \hat{v}\}} - \alpha)(x^2 - \hat{v}^2) + \mathbb{1}_{\{x \leq \hat{v}\}} \hat{e}(\hat{v} - x) + \alpha \hat{e} \left(\frac{\hat{e}}{2} - \hat{v}\right)$$

is strictly consistent for $(\mathrm{VaR}_\alpha(F), \mathrm{ES}_\alpha(F))$ *on the domain* $\mathcal{D} = \{(\hat{v}, \hat{e}) \in \mathbb{R}^2 : W_\alpha \hat{v} \leq \hat{e} \leq \hat{v} \leq 0\}$*, where* $\frac{\mathrm{ES}_\alpha(F)}{\mathrm{VaR}_\alpha(F)} \geq W_\alpha \geq 1$.

Proof. Let $X \sim F$ with continuous density f. Define the expected score: $\bar{S}(\hat{v}, \hat{e}) := \mathbb{E}_F[S_\alpha(\hat{v}, \hat{e}; X)]$.

Computing the expected score. We decompose the expectation based on the indicator function $\mathbb{1}_{\{X \le \hat{v}\}}$:

$$\bar{S}(\hat{v}, \hat{e}) = \int_{-\infty}^{\hat{v}} \left[\frac{W_\alpha}{2}(1-\alpha)(x^2 - \hat{v}^2) + \hat{e}(\hat{v} - x) \right] f(x)\, dx \\ + \int_{\hat{v}}^{\infty} \left[-\frac{W_\alpha \alpha}{2}(x^2 - \hat{v}^2) \right] f(x)\, dx + \alpha \hat{e} \left(\frac{\hat{e}}{2} - \hat{v} \right).$$

Let $p := F(\hat{v})$, $M_1(\hat{v}) := \int_{-\infty}^{\hat{v}} x f(x)\, dx$, and $M_2(\hat{v}) := \int_{-\infty}^{\hat{v}} x^2 f(x)\, dx$. By expanding and collecting terms, we have:

$$\bar{S}(\hat{v}, \hat{e}) = \frac{W_\alpha}{2} M_2(\hat{v}) - \frac{W_\alpha}{2} \mathbb{E}[X^2]\alpha + \frac{W_\alpha}{2} \hat{v}^2(\alpha - p) + \hat{e}[p\hat{v} - M_1(\hat{v})] + \frac{\alpha \hat{e}^2}{2} - \alpha \hat{e} \hat{v}.$$

First-order conditions (FOCs). Taking the partial derivative with respect to $\hat{e}$, we have $\frac{\partial \bar{S}}{\partial \hat{e}} = p\hat{v} - M_1(\hat{v}) + \alpha \hat{e} - \alpha \hat{v}$. Setting $\frac{\partial \bar{S}}{\partial \hat{e}} = 0$, we have:

$$\alpha \hat{e} = \alpha \hat{v} - p\hat{v} + M_1(\hat{v}) = \hat{v}(\alpha - p) + M_1(\hat{v}). \tag{2}$$

Taking the partial derivative with respect to $\hat{v}$ (using Leibniz rule for $M_1(\hat{v})$, $M_2(\hat{v})$, and $p = F(\hat{v})$),

$$\begin{aligned} \frac{\partial \bar{S}}{\partial \hat{v}} &= \frac{\partial}{\partial \hat{v}} \left(\frac{W_\alpha}{2} M_2(\hat{v}) + \frac{W_\alpha}{2} \hat{v}^2(\alpha - p) + \hat{e}[p\hat{v} - M_1(\hat{v})] - \alpha \hat{e} \hat{v} \right) \\ &= \left(\frac{W_\alpha}{2} \hat{v}^2 f(\hat{v}) \right) + \left(\frac{W_\alpha}{2} (2\hat{v}(\alpha - p) + \hat{v}^2(-f(\hat{v}))) \right) + (\hat{e}(f(\hat{v})\hat{v} + p - \hat{v} f(\hat{v}))) - \alpha \hat{e} \\ &= \left(\frac{W_\alpha}{2} \hat{v}^2 f(\hat{v}) \right) + \left(W_\alpha \hat{v}(\alpha - p) - \frac{W_\alpha}{2} \hat{v}^2 f(\hat{v}) \right) + (\hat{e} p) - \alpha \hat{e} \\ &= W_\alpha \hat{v}(\alpha - p) + \hat{e}(p - \alpha) \\ &= (p - \alpha)(\hat{e} - W_\alpha \hat{v}). \end{aligned}$$

Setting $\frac{\partial \bar{S}}{\partial \hat{v}} = 0$, we have:

$$(p - \alpha)(\hat{e} - W_\alpha \hat{v}) = 0. \tag{3}$$

Solving the FOCs. From Eq. (3), a critical point must satisfy $p = \alpha$ (*i.e.*, $F(\hat{v}) = \alpha$) or $\hat{e} = W_\alpha \hat{v}$. We are seeking the general critical point corresponding to the tail measures. We set $p = \alpha$, which implies $\hat{v} = \mathrm{VaR}_\alpha(F)$.

Substituting $p = \alpha$ into Eq. (2), we have:

$$\alpha \hat{e} = M_1(\hat{v}) = \int_{-\infty}^{\mathrm{VaR}_\alpha} x f(x)\, dx.$$

Therefore,

$$\hat{e} = \frac{1}{\alpha} \int_{-\infty}^{\mathrm{VaR}_\alpha} x f(x)\, dx = \mathrm{ES}_\alpha(F).$$

This confirms that the critical point is $(\hat{v}, \hat{e}) = (\mathrm{VaR}_\alpha(F), \mathrm{ES}_\alpha(F))$.

Second-order conditions (SOCs). The Hessian matrix is $H = \begin{pmatrix} \frac{\partial^2 \bar{S}}{\partial \hat{v}^2} & \frac{\partial^2 \bar{S}}{\partial \hat{v} \partial \hat{e}} \\ \frac{\partial^2 \bar{S}}{\partial \hat{e} \partial \hat{v}} & \frac{\partial^2 \bar{S}}{\partial \hat{e}^2} \end{pmatrix}$. We compute the second derivatives:

$$\frac{\partial^2 \bar{S}}{\partial \hat{e}^2} = \alpha > 0, \quad \frac{\partial^2 \bar{S}}{\partial \hat{v} \partial \hat{e}} = p - \alpha + f(\hat{v})\hat{v}, \quad \frac{\partial^2 \bar{S}}{\partial \hat{v}^2} = 2W_\alpha \hat{v} f(\hat{v}) + W_\alpha \hat{v}^2 f'(\hat{v}) + W_\alpha f(\hat{v}) + \hat{e} f(\hat{v}).$$

At the critical point where $p = \alpha$ and $(\hat{v}, \hat{e}) = (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$,

$$\left. \frac{\partial^2 \bar{S}}{\partial \hat{v} \partial \hat{e}} \right|_{(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)} = f(\mathrm{VaR}_\alpha)\mathrm{VaR}_\alpha < 0,$$

since $\mathrm{VaR}_\alpha < 0$. For the Hessian to be positive definite, we need a) $\frac{\partial^2 \bar{S}}{\partial \hat{e}^2} = \alpha > 0$ (satisfied) and b) $\det(H) > 0$. The determinant condition requires:

$$\alpha \cdot \frac{\partial^2 \bar{S}}{\partial \hat{v}^2} - \left(\frac{\partial^2 \bar{S}}{\partial \hat{v} \partial \hat{e}} \right)^2 > 0.$$

This is satisfied when $W_\alpha \geq 1$ and $\mathrm{ES}_\alpha / \mathrm{VaR}_\alpha \geq W_\alpha$, ensuring strict convexity at the critical point.

Uniqueness. By strict convexity on $\mathcal{D}$, the critical point $(\mathrm{VaR}_\alpha(F), \mathrm{ES}_\alpha(F))$ is the unique global minimizer. Any $(\hat{v}, \hat{e}) \neq (\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$ violates the FOCs, yielding $\bar{S}(\hat{v}, \hat{e}) > \bar{S}(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$.

Therefore, the score is strictly consistent for the pair $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$. □

Remark 2.5. Our proof establishes strict consistency of the Acerbi–Székely score for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$. The parameter W_α controls tail sensitivity: larger values place more weight on tail deviations, improving the optimization landscape in the tail region. The score's strict convexity on $\mathcal{D}$ and interpretability make it particularly suitable for adversarial training frameworks such as Tail-GAN, where the generator is trained to minimize tail risk as measured by this score function.

At the time of writing, Cont et al. [1] is in the "Articles in Advance" stage, and we could not locate the online appendix containing their detailed proofs. Therefore, we refer to their preprint (Cont et al. [10]) for these proofs.

Remark 2.6 (Comparison with Cont et al. [10])*.* Our proof of Theorem 2.4 employs a direct calculus-based approach, computing FOCs and verifying SOCs via the Hessian matrix to establish strict consistency. This classical optimization methodology provides a transparent demonstration that $(\mathrm{VaR}_\alpha(F), \mathrm{ES}_\alpha(F))$ is the unique global minimizer of the expected score $\bar{S}(\hat{v}, \hat{e})$.

In contrast, Cont et al. [10] adopt a different strategy tailored to their machine learning application (Tail-GAN). Their proof (Appendix B.1) focuses on verifying the elicitability conditions and establishing positive semi-definiteness of the Hessian on specific regions B and $\tilde{B}_2$. This approach characterizes the optimization landscape of the score function, which is crucial for gradient-based training of generative adversarial networks.

The key differences are as follows.

1. We solve the FOCs to identify the critical point, while Cont et al. [10] verify elicitability and convexity on specific regions.

2. Our proof establishes strict consistency on the natural domain $\mathcal{D} = \{(\hat{v}, \hat{e}) \in \mathbb{R}^2 : W_\alpha \hat{v} \leq \hat{e} \leq \hat{v} \leq 0\}$, whereas Cont et al. [10] analyze finer geometric properties of the score function.

Both proofs are valid and complementary: ours provides a direct verification of strict consistency, while Cont et al. [10] offer additional insights into the computational properties of the Acerbi–Székely score.

3 Bilevel–penalized equivalence of minimizers

In this section, we consider Theorem 3.8 of Cont et al. [1], which establishes the equivalence between the bilevel risk minimization problem and the adversarial max-min game formulation underlying Tail-GAN's training procedure. We present our own proofs of a simplified version of this theorem from both optimization and measure-theoretic perspectives.

3.1 Simplified settings

We begin by introducing our simplified settings based on those in Cont et al. [1].

Let $S_\alpha : \mathbb{R}^2 \times \mathbb{R} \to \mathbb{R}$ be a strictly consistent scoring function for $(\mathrm{VaR}_\alpha, \mathrm{ES}_\alpha)$. For any distribution μ, let $C(\mu) := \min_{(v,e)} \mathbb{E}_{X\sim\mu}[S_\alpha((v,e),X)]$.

Definition 3.1 (Loss functions)**.** For any $G \in \mathcal{G}$ and $D \in \mathcal{D}$, we define as follows.

1. The *real loss* is D's score on the true distribution P_r:

$$L_r(D) := \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{X\sim\Pi^k_{\#}P_r}[S_\alpha(D(\Pi^k_{\#}P_r), X)].$$

2. The *generated loss* is D's score on the generated distribution $P_G = G_{\#}P_z$:

$$L_g(G,D) := \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{X\sim\Pi^k_{\#}P_G}[S_\alpha(D(\Pi^k_{\#}P_G), X)].$$

Definition 3.2 (Optimal discriminator set)**.** We define the set of discriminators that are optimal on the true distribution: $\mathcal{D}_0 := \arg\min_{D\in\mathcal{D}} L_r(D)$. By strict consistency, $D \in \mathcal{D}_0$ if and only if $D(\Pi^k_{\#}P_r) = (\mathrm{VaR}_\alpha(\Pi^k_{\#}P_r), \mathrm{ES}_\alpha(\Pi^k_{\#}P_r))$ for all k.

Definition 3.3 (Bilevel problem)**.** We find the generator G whose distribution P_G is scored most favorably by a discriminator that is optimal on the real data.

$$V_B := \min_{G\in\mathcal{G}} V_B(G), \quad \text{where} \quad V_B(G) = \min_{D\in\mathcal{D}_0} L_g(G,D).$$

Remark 3.4. Since $D \in \mathcal{D}_0$ only constrains the behavior of D on P_r, the inner minimization $\min_{D\in\mathcal{D}_0} L_g(G,D)$ is achieved by a $D \in \mathcal{D}_0$ that also happens to be optimal for P_G. Therefore,

$$V_B(G) = \min_{D\in\mathcal{D}_0} \frac{1}{K}\sum_{k=1}^{K} \mathbb{E}_{\Pi^k_{\#}P_G}[S_\alpha(D(\Pi^k_{\#}P_G), X)] = \frac{1}{K}\sum_{k=1}^{K} C(\Pi^k_{\#}P_G).$$

The bilevel problem is $V_B = \min_{G\in\mathcal{G}} C(\Pi_{\#}P_G)$, which seeks the generator that produces the distribution with the lowest possible optimal score.

Definition 3.5 (Penalized problem)**.** Find the generator G and discriminator D that jointly minimize the generated loss, subject to a penalty for D being sub-optimal on the real data.

$$V_P(\lambda) := \min_{G\in\mathcal{G}} V_P(G,\lambda), \quad \text{where} \quad V_P(G,\lambda) = \min_{D\in\mathcal{D}} \left[L_g(G,D) + \lambda L_r(D)\right].$$

3.2 Optimization perspective

Theorem 3.6 (Bilevel–penalized equivalence, simplified version of Cont et al. [1], Theorem 3.8)**.** *Let S_α be strictly consistent and assume $\mathcal{G}$ is compact and L_g is continuous in G. Let $C_r := \min_{D\in\mathcal{D}} L_r(D)$ be the minimal score on the true distribution. Then, the optimal generator G^* is the same for both problems, and their optimal values are related by a constant:* $\lim_{\lambda\to\infty} (V_P(\lambda) - \lambda C_r) = V_B$*. For a sufficiently large but finite λ, the solution (G^*, D^*) to the penalized problem satisfies $G^* \in \arg\min_G V_B(G)$ and $D^* \in \mathcal{D}_0$.*

Proof. In the sense of epi-convergence (see, *e.g.,* Kall [11]), $\lim_{\lambda\to\infty} (V_P(\lambda) - \lambda C_r) = V_B$ implies convergence of minimizers. Let $G \in \mathcal{G}$ be fixed. We analyze the inner minimization problem for $V_P(G, \lambda)$:

$$V_P(G,\lambda) = \min_{D\in\mathcal{D}} \left[L_g(G,D) + \lambda L_r(D)\right].$$

Let $C_r = \min_{D\in\mathcal{D}} L_r(D)$. We rewrite the objective by adding and subtracting λC_r:

$$V_P(G,\lambda) = \min_{D\in\mathcal{D}} \left[L_g(G,D) + \lambda(L_r(D) - C_r)\right] + \lambda C_r.$$

Define the non-negative penalty function $\Psi(D) := L_r(D) - C_r$. By definition of C_r and $\mathcal{D}_0$, we have a) $\Psi(D) \geq 0$ for all $D \in \mathcal{D}$ and b) $\Psi(D) = 0$ if and only if $D \in \mathcal{D}_0$. The objective for the inner minimization is:

$$\min_{D\in\mathcal{D}} \left[L_g(G,D) + \lambda\Psi(D)\right] + \lambda C_r.$$

The term $\lambda\Psi(D)$ is an exact penalty function for the constraint $D \in \mathcal{D}_0$. As $\lambda \to \infty$, the minimization is forced to select D such that $\Psi(D) = 0$, *i.e.*, $D \in \mathcal{D}_0$.

Formally, we consider the limit of the inner objective function:

$$\lim_{\lambda\to\infty} \left[L_g(G,D) + \lambda\Psi(D)\right] = \begin{cases} L_g(G,D) & \text{if } D \in \mathcal{D}_0 \\ +\infty & \text{if } D \notin \mathcal{D}_0. \end{cases}$$

This is the definition of the indicator function $L_g(G,D) + \mathcal{I}_{\mathcal{D}_0}(D)$. Therefore, the limit of the inner minimization in the sense of epi-convergence is:

$$\begin{aligned} \lim_{\lambda\to\infty} V_P(G,\lambda) &= \left(\min_{D\in\mathcal{D}} \left[L_g(G,D) + \mathcal{I}_{\mathcal{D}_0}(D)\right]\right) + \lambda C_r \\ &= \left(\min_{D\in\mathcal{D}_0} L_g(G,D)\right) + \lambda C_r \\ &= V_B(G) + \lambda C_r. \end{aligned}$$

We take the minimum over $G \in \mathcal{G}$:

$$\min_{G\in\mathcal{G}} \left(\lim_{\lambda\to\infty} V_P(G,\lambda)\right) = \min_{G\in\mathcal{G}} \left(V_B(G) + \lambda C_r\right).$$

Since λC_r is a constant with respect to G, the optimal generator G^* that minimizes the left-hand side is the same G^* that minimizes $V_B(G)$ on the right-hand side. Therefore, we have $\lim_{\lambda\to\infty} V_P(\lambda) = V_B + \lambda C_r$, which completes the proof. □

Remark 3.7 (Comparison with Cont et al. [10]). Our equivalence result differs fundamentally from Theorem A.1 and Appendix B.4 in Cont et al. [10] in several aspects, as detailed below.

1. Cont et al. [10] establish equivalence between a bilevel problem and a max-min game. Our result establishes equivalence between bilevel and penalized formulations.

2. Cont et al. [10] employ measure-theoretic tools, specifically from optimal transport theory (Theorem 7.1 of Ambrosio and Pratelli [12]), to construct generators achieving optimal discriminator values. In contrast, our approach is optimization-theoretic, relying on penalty methods and epi-convergence.

3. Cont et al. [10] require that P_z is absolutely continuous with respect to the Lebesgue measure and that both P_z and measures in $\Sigma(\bar{D})$ have finite first moments. We require strict consistency of S_α, compactness of $\mathcal{G}$, and continuity of L_g in G, but we do not impose measure-theoretic regularity conditions on P_z.

4. Cont et al.'s [10] result establishes the theoretical foundation of the max-min game. Our penalized formulation provides a computationally tractable reformulation, as the problem $\min_{G,D}[L_g(G,D) + \lambda L_r(D)]$ can be solved via standard gradient-based methods without explicitly enforcing the constraint $D \in \mathcal{D}_0$.

In summary, while Cont et al. [10] offer the theoretical foundation using measure theory, we provide an optimization perspective that addresses the computational implementation via penalty methods.

3.3 Measure theoretic perspective

We now present our measure-theoretic proof of Theorem 3.6. For background on Γ-convergence, see Maso [13] and Braides [14].

Proof. In the sense of Γ-convergence, $\lim_{\lambda\to\infty}(V_P(\lambda) - \lambda C_r) = V_B$ implies convergence of minimizers. Let $G \in \mathcal{G}$ be fixed. We analyze the inner minimization problem for $V_P(G,\lambda)$:

$$V_P(G,\lambda) - \lambda C_r = \min_{D\in\mathcal{D}} \left[L_g(G,D) + \lambda(L_r(D) - C_r)\right].$$

Define a sequence of objective functionals $\Phi_\lambda : \mathcal{D} \to \mathbb{R} \cup \{+\infty\}$ by the following.

$$\Phi_\lambda(D) := L_g(G,D) + \lambda\Psi(D), \quad \text{where } \Psi(D) := L_r(D) - C_r.$$

Since $L_r(D) \geq C_r$, we have $\Psi(D) \geq 0$ for all $D \in \mathcal{D}$. The equality $\Psi(D) = 0$ holds if and only if $D \in \mathcal{D}_0$. The term $\lambda\Psi(D)$ is an exact penalty function for the constraint $D \in \mathcal{D}_0$.

We establish the Γ-convergence of Φ_λ to the limiting functional $\Phi_\infty : \mathcal{D} \to \mathbb{R} \cup \{+\infty\}$ defined by:

$$\Phi_\infty(D) := L_g(G,D) + \mathcal{I}_{\mathcal{D}_0}(D) = \begin{cases} L_g(G,D) & \text{if } D \in \mathcal{D}_0 \\ +\infty & \text{if } D \notin \mathcal{D}_0, \end{cases}$$

where $\mathcal{I}_{\mathcal{D}_0}$ is the indicator function of the set $\mathcal{D}_0$.

For any $D \in \mathcal{D}$ and any sequence $D_\lambda \to D$, we show $\liminf_{\lambda\to\infty} \Phi_\lambda(D_\lambda) \geq \Phi_\infty(D)$.

- If $D \in \mathcal{D}_0$, since $\Psi(D_\lambda) \geq 0$, $\liminf_{\lambda\to\infty} \Phi_\lambda(D_\lambda) \geq \lim_{\lambda\to\infty} L_g(G,D_\lambda) = L_g(G,D) = \Phi_\infty(D)$, by continuity of L_g.

- If $D \notin \mathcal{D}_0$, then $\Psi(D) > 0$. If $D_\lambda \to D$, then $\Psi(D_\lambda) \to \Psi(D) > 0$ by continuity of L_r. Thus, $\lim_{\lambda\to\infty} \lambda\Psi(D_\lambda) = +\infty$, so $\liminf_{\lambda\to\infty} \Phi_\lambda(D_\lambda) = +\infty = \Phi_\infty(D)$.

For any $D \in \mathcal{D}$, we show that there exists a sequence $D_\lambda \to D$ such that $\limsup_{\lambda\to\infty} \Phi_\lambda(D_\lambda) \leq \Phi_\infty(D)$.

- If $D \in \mathcal{D}_0$, choose the constant sequence $D_\lambda = D$. Since $\Psi(D) = 0$, $\limsup_{\lambda\to\infty} \Phi_\lambda(D) = L_g(G, D) = \Phi_\infty(D)$.
- If $D \notin \mathcal{D}_0$, the inequality $\limsup_{\lambda\to\infty} \Phi_\lambda(D_\lambda) \leq +\infty$ is trivially satisfied.

Since $\Phi_\lambda\Gamma$ converges to Φ_∞, by the fundamental property of Γ-convergence, the infima converge:

$$\lim_{\lambda\to\infty} \left(\min_{D\in\mathcal{D}} \Phi_\lambda(D) \right) = \min_{D\in\mathcal{D}} \Phi_\infty(D).$$

Substituting back the definitions, we obtain:

$$\begin{aligned} \lim_{\lambda\to\infty} (V_P(G, \lambda) - \lambda C_r) &= \min_{D\in\mathcal{D}} [L_g(G, D) + \mathcal{I}_{\mathcal{D}_0}(D)] \\ &= \min_{D\in\mathcal{D}_0} L_g(G, D) \\ &= V_B(G). \end{aligned}$$

Taking the minimum over $G \in \mathcal{G}$ and using the compactness of $\mathcal{G}$ and the continuity of L_g (which ensures the convergence of minimizers in Γ-convergence), we conclude $\lim_{\lambda\to\infty} (V_P(\lambda) - \lambda C_r) = V_B$. Moreover, the minimizers converge: $G^*_\lambda \to G^* \in \arg\min_G V_B(G)$ and $D^*_\lambda \to D^* \in \mathcal{D}_0$. □

Remark 3.8 (The role of Γ-convergence). The bilevel–penalized equivalence is established via the framework of Γ-convergence (also known as epi-convergence). This is the variational tool for analyzing the convergence of minimum values and minimizers for a sequence of optimization problems.

Standard measure-theoretic convergence (*e.g.*, pointwise or uniform) of the objective function $\Phi_\lambda(D)$ is not sufficient to guarantee the convergence of its minimum value $\min_D \Phi_\lambda(D)$ or its minimizers. Γ-convergence precisely formalizes the action of an exact penalty function. It ensures that, in the limit $\lambda \to \infty$, the sequence of penalized objectives $\Phi_\lambda(D) = L_g(G, D) + \lambda\Psi(D)$ effectively removes the space outside the feasible set $\mathcal{D}_0$ by mapping it to $+\infty$. The Γ-limit of the sequence Φ_λ is the indicator-augmented function $\Phi_\infty(D) = L_g(G, D) + \mathcal{I}_{\mathcal{D}_0}(D)$. This limit directly yields the constrained minimum:

$$\lim_{\lambda\to\infty} \left(\min_{D\in\mathcal{D}} \Phi_\lambda(D) \right) = \min_{D\in\mathcal{D}} \Phi_\infty(D) = \min_{D\in\mathcal{D}_0} L_g(G, D) = V_B(G).$$

Remark 3.9 (Comparison with Cont et al. [10], Theorem A.1). Our measure-theoretic proof, based on Γ-convergence and the exact penalty method, establishes the equivalence between the bilevel and penalized formulations, as summarized below.

$$\min_{G\in\mathcal{G}} \min_{D\in\mathcal{D}_0} L_g(G, D) \quad \Longleftrightarrow \quad \lim_{\lambda\to\infty} \min_{G\in\mathcal{G}} \min_{D\in\mathcal{D}} \left[L_g(G, D) + \lambda L_r(D) \right].$$

This approach justifies the computational stability and enforcement of the constraint $D \in \mathcal{D}_0$ in a tractable optimization setting.

The proof of Theorem A.1 (Appendix B.4) in Cont et al. [10] takes a different measure-theoretic approach, including the following.

1. Their result establishes the equivalence between the bilevel problem and the max-min game (GAN formulation) via measure construction, whereas our result establishes the equivalence between the bilevel problem and the penalized formulation via Γ-convergence.

2. Their proof adopts the existence theorem from optimal transport (Ambrosio and Pratelli [12], Theorem 7.1).

Cont et al.'s [10] proof establishes the foundational measure-theoretic justification that the optimal generator's measure can be constructed. Our proof provides a justification that the constraint can be satisfied via a penalty function in a simplified setting.

References

[1] R. Cont, M. Cucuringu, R. Xu, and C. Zhang. Tail-GAN: Learning to simulate tail risk scenarios. *Management Science*, 2025. doi:10.1287/mnsc.2023.00936. Articles in Advance.

[2] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio. Generative adversarial nets. In Z. Ghahramani, M. Welling, C. Cortes, N. Lawrence, and K. Q. Weinberger, editors, *Advances in Neural Information Processing Systems*, pages 2672–2680. Curran Associates, Inc., 2014.

[3] S. Takahashi, Y. Chen, and K. Tanaka-Ishii. Modeling financial time-series with generative adversarial networks. *Physica A: Statistical Mechanics and its Applications*, 527:121261, 2019. doi:10.1016/j.physa.2019.121261.

[4] M. Wiese, R. Knobloch, R. Korn, and P. Kretschmer. Quant GANs: deep generation of financial time series. *Quantitative Finance*, 20:1419–1440, 2020. doi:10.1080/14697688.2020.1730426.

[5] M. Vuletić, F. Prenzel, and M. Cucuringu. Fin-GAN: forecasting and classifying financial time series via generative adversarial networks. *Quantitative Finance*, 24:175–199, 2024. doi:10.1080/14697688.2023.2299466.

[6] M. Vuletić and R. Cont. VolGAN: A generative model for arbitrage-free implied volatility surfaces. *Applied Mathematical Finance*, 31:203–238, 2024. doi:10.1080/1350486X.2025.2471317.

[7] C. Acerbi and B. Székely. Back-testing expected shortfall. *Risk*, 27:76–81, 2014.

[8] T. Gneiting. Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106:746–762, 2011. doi:10.1198/jasa.2011.r10138.

[9] T. Fissler and J. F. Ziegel. Higher order elicitability and Osband's principle. *Annals of Statistics*, 44:1680–1707, 2016. doi:10.1214/16-AOS1439.

[10] R. Cont, M. Cucuringu, R. Xu, and C. Zhang. Tail-GAN: Learning to simulate tail risk scenarios, 2025. preprint, arXiv:2203.01664v4 [q-fin.RM].

[11] P. Kall. Approximation to optimization problems: An elementary review. *Mathematics of Operations Research*, 11:9–18, 1986. doi:10.1287/moor.11.1.9.

[12] L. Ambrosio and A. Pratelli. Existence and stability results in the L^1 theory of optimal transportation. In *Optimal Transportation and Applications: Lectures given at the CIME Summer School held in Martina Franca, Italy, September 2-8, 2001*, pages 123–160. Springer, 2003. doi:10.1007/b12016.

[13] G. Maso. *An Introduction to* Γ*-Convergence*. Springer, 1993. doi:10.1007/978-1-4612-0327-8.

[14] A. Braides. Γ*-Convergence for Beginners*. Oxford University Press, 2002.