WHAT'S A CREDIT WORTH? A MARKET FRAMEWORK FOR ATTRIBUTION-AWARE COMPENSATION IN GENERATIVE MUSIC

Anonymous authors
Paper under double-blind review

ABSTRACT

Advances in generative AI are rapidly increasing the quality and commercial value of generated music, and this progress depends on large catalogs of creators' recordings. This raises a central question for platform design: how should creators be compensated when their work is used to train generative AI models that in turn produce commercial outputs? We develop a framework for Attribution-Aware compensation in generative-music markets, where each creator's payment depends on an attribution signal estimating their contribution to model outputs. Compared to past compensation frameworks, our framework has two unique considerations: (1) attribution is traced to entire creator catalogs, not individual songs, and (2) the *informativeness* (signal-to-noise ratio) of the attribution score is an input to the payment mechanism. The framework yields a closed-form payment rule per creator and measures the welfare cost of inaccurate attribution for both creators and the platform. Whether the welfare-optimal contract is royalty-based or takes the form of fixed-fee licensing depends on how informative attribution is for that creator's catalog. We show that better attribution translates directly into welfare gains for both creators and the platform, yet under multi-platform competition only the most precise signal in the market captures those gains. Empirically, we train acoustic and symbolic music generation models and measure the informativeness of scalable attribution techniques against a leave-one-catalog-out ground truth. Our experiments reveal that noisy attribution signals push payment toward fixed-fee licensing and diminish welfare for both creators and the platform, providing an economic motivation for further research on improved attribution.

1 INTRODUCTION

Generative music models trained on copyrighted recordings are reshaping how music is produced and consumed. As these models improve, AI music is crowding music streaming services (Khouwes, 2026), copyright disputes have emerged, including the major labels' 2024 lawsuits against Suno and Udio, and regulators have responded with disclosure mandates (European Parliament and Council of the European Union, 2024), suggesting the absence of a reliable and Attribution-Aware compensation rule. Training-data provenance is rarely disclosed, and licensing, where it exists, is negotiated at the catalog or genre level for flat fees that ignore two key factors: (1) that some data contributes more than others for a given generated output (Deng et al., 2025; Barnett et al., 2024), and (2) that flat licensing fees do not pay creators proportionally to generated-output revenue. This raises a central question: how should creators be compensated when their work trains commercial generative AI models?

Answering this question requires combining two technical fields. *Data attribution* (Ghorbani & Zou, 2019; Koh & Liang, 2017; Ilyas et al., 2022) estimates how much each creator's catalog contributed to a model output, but its noise varies across creators and methods, and work applying it to music AI (Barnett et al., 2024; Choi et al., 2025; Deng et al., 2025) does not factor that noise into payment. *Payment design* (Holmström, 1979) supplies tools for payment under noisy signals, but has rarely been combined with the methods that produce the signal. We unify these isolated literatures.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

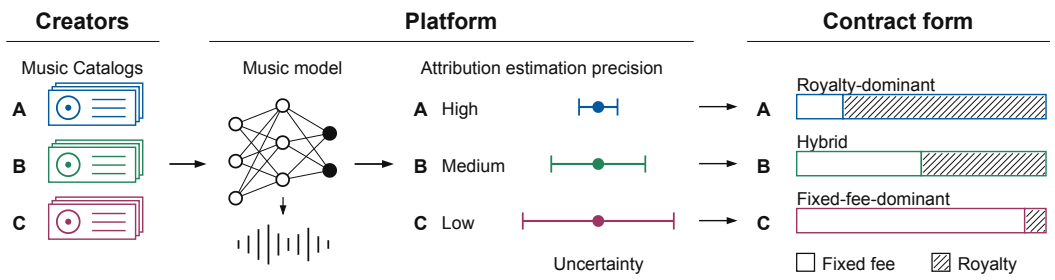

Figure 1: Creators license catalogs to a platform that estimates each creator's contribution. Each creator's mix of fixed fee (one-time licensing fee) and royalty (recurring payments proportional to attribution estimates) follows from how precisely the platform can measure their contribution.

To this end, we develop a framework for Attribution-Aware compensation in generative-music markets, where each creator's payment depends on an attribution signal estimating their contribution to model outputs. The framework's central concept is *informativeness*, a measure of how well the noisy attribution signal tracks the true contribution. We track influence *per-catalog* rather than per song, since catalogs group training data by distinct payee, and a *creator* here may be an independent artist or a record label. We derive a closed-form payment rule per creator that balances a fixed *licensing* fee and *royalties* proportional to the catalog attribution score, along with a closed-form expression for the welfare loss caused by imperfect informativeness.

The framework yields several implications about how attribution informativeness shapes contracts, welfare, and creator incentives. For music AI platforms, royalty-based payment is welfare-optimal when attribution adds precise information about a creator's catalog, and fixed-fee licensing when it does not. This does not depend on restricting attention to fixed fees and royalties, because no payment rule, however it is structured, can tie a creator's pay to their contribution more closely than the attribution signal itself allows. **We show that both creators and the AI platform can benefit from more precise attribution.** Under multi-platform competition, only the leading platform profits from improving attribution, since the gains accrue to whichever signal is most precise. For creators, uniform compensation can drive out distinctive catalogs, and the attribution scores that set their pay should not be self-reported.

Under this framework, we propose a leave-one-catalog-out measurement pipeline and study it empirically across two relevant settings: audio and symbolic music generation. We fine-tune Stable Audio Open (Evans et al., 2025) on 50 Creative Commons catalogs (Bogdanov et al., 2019) and evaluate EKFAC (Mlodozeniec et al., 2025) and D-TRAK (Zheng et al., 2024), and a symbolic music language model (Thickstun et al., 2024) on 100 TheoryTab catalogs (Donahue et al., 2022) with LoGra (Choe et al., 2025) and TRAK (Park et al., 2023). We observe that selecting attribution methods by rank agreement and by calibrated precision can favor different methods. In the audio setting, current attribution noise favors fixed-fee-dominant contracts under the framework at the risk-aversion levels considered. Platforms deploying Attribution-Aware compensation should therefore benchmark multiple methods rather than commit to one.

Our contributions are threefold:

- **Framework.** The first principal-agent framework bridging data attribution and contract design for generative music markets. It yields closed-form expressions for the welfare-optimal contract per creator (royalty-based or fixed-fee licensing) and for the welfare cost of imperfect attribution.
- **Measurement.** The first empirical benchmark of attribution informativeness in generative music models, with experiments that reflect the model paradigms and content modalities of today's commercial music AI systems under tractable attribution methods.
- **Implications.** Combined theoretical and empirical analysis yields concrete implications for the generative music industry, establishing attribution quality and market transparency as measurable economic primitives and providing a basis for industry policy and governance.

2 RELATED WORK

Data pricing and contract design. Research on data markets covers marketplace design, information pricing, and pricing data for ML (Agarwal et al., 2019; Bergemann et al., 2018; Chen et al., 2022). Related economic work studies data as a nonrival good (Jones & Tonetti, 2020; Cong et al., 2022; Acemoglu et al., 2022). Contract theory provides tools for noisy signals, heterogeneous participants, and computational contract design (Holmström, 1979; Mussa & Rosen, 1978; Guruganesh et al., 2021). Closer to our setting, Zhang et al. (2025) study LLM data pricing with risk-neutral creators and reliable attribution, while Wang et al. (2024) allocate royalties through counterfactual Shapley contributions under an assumed-reliable score. These studies address a question downstream of ours. They allocate payments from given signals, whereas we ask how strongly payments should depend on those signals. None of these works combine ML data-attribution methods with contract design, nor do they treat signal informativeness as a heterogeneous quantity across creators that the payment rule must explicitly accommodate.

Data attribution and music attribution. General data-attribution methods estimate per-datum contribution to model outputs. Data Shapley (Ghorbani & Zou, 2019; Jia et al., 2019) provides an axiomatic per-datum attribution; influence functions (Koh & Liang, 2017; Giordano et al., 2019) and TracIn (Pruthi et al., 2020) trace predictions to training examples; datamodels (Ilyas et al., 2022) and TRAK (Park et al., 2023) fit linear surrogates for the data-to-output mapping. Recent extensions scale these methods to approximate unrolled differentiation (Bae et al., 2024), datamodel-based selection (Engstrom et al., 2024), LLM-scale influence (Grosse et al., 2023), LoRA-tuned models (Kwon et al., 2024), and diffusion models (Georgiev et al., 2023). In music, fingerprinting (Wang, 2003) catches near-exact matches, embedding methods (Spijkervet & Burgoyne, 2021) measure distributional similarity, and Barnett et al. (2024); Choi et al. (2025); Deng et al. (2025) apply attribution to generative music for royalty assignment. Beyond music, Wang et al. (2023) evaluate attribution for text-to-image models, and Wang et al. (2025) study adversarial attacks on data attribution. Prior work improves attribution informativeness; we use it to design music-data payments.

Generative music models. Generative music models are trained on large catalogs of audio recordings, either as codec LMs over tokenized music (van den Oord et al., 2017; Dieleman et al., 2018; Dhariwal et al., 2020; Agostinelli et al., 2023; Copet et al., 2023) or as diffusion models over latent audio (Ho et al., 2020; Rombach et al., 2022; Forsgren & Martiros, 2022; Evans et al., 2025), with open-weight releases broadening access beyond major labs (Caillon et al., 2025; Gong et al., 2026). Output quality now supports substantial commercial value, with Suno reporting $300m in annual recurring revenue (Silberling, 2026), which motivates the payment mechanisms we consider here.

3 A FRAMEWORK FOR ATTRIBUTION-AWARE COMPENSATION

We formalize how a music AI platform can use attribution to set creator payments. Beyond characterizing the equilibrium payment rule, the framework reveals how attribution shapes contracts, welfare, and competition across the market.

3.1 FRAMEWORK FOR MUSICAL MARKET AND DATA ATTRIBUTION

We model a music AI platform licensing creator catalogs and setting payments according to attribution quality. The setup has four primitives: the market and its creators, the data-attribution layers, an informativeness measure, and the contract space with creator preferences over fixed-fee and royalty bundles.

Market and Creators. The market has three participants: N creators, an AI platform, and its users. We focus on the platform-creator interaction, taking user demand as exogenous (observable from platform logs). The platform operates a generative music service. A user request r arrives, and the platform earns gross utility $u(x, r)$ where $x \in \{0, 1\}^N$ indicates the catalogs active in the training set. User requests r are drawn from a distribution $\mathcal{D}$ that the platform can estimate empirically from request logs; we make no iid assumption across requests.

Each creator $j \in \{1, \ldots, N\}$ controls a distinct, non-overlapping catalog C_j (e.g., an artist's discography, or a label's collection) to be licensed wholesale, on an all-or-nothing basis. We assume that creators and the platform share the same assessment of attribution informativeness. Each creator

has a risk-aversion parameter $\alpha_j > 0$, which measures how strongly they prefer steady payments to volatile ones (Pratt, 1964). The platform's utility u is non-additive across creators, reflecting the non-additive nature of data attribution (two creators with similar styles add less joint value than two distinct ones).

Data attribution. The central informational primitive is attribution: how much each creator's catalog contributed to a generated output. The latent true attribution $a_j^* \in \mathbb{R}$ is the counterfactual contribution of catalog C_j to the platform's outputs over a billing period (e.g., monthly in streaming). The measured attribution $\hat{a}_j$ is a noisy estimate produced by an ML method (e.g., data Shapley (Ghorbani & Zou, 2019), influence functions (Koh & Liang, 2017), audio fingerprinting), and is what the platform pays on. Unlike most data-pricing settings (Agarwal et al., 2019; Chen et al., 2022; Zhang et al., 2025), music has a perceptual property: listeners can recognize an artist's style in generated output, so a_j^* may be semantically meaningful, not a marginal predictive effect. In practice, creators may lack access to the model parameters, training data, or training records needed to reproduce the platform's attribution estimates.

Informativeness. The quantity that determines the welfare-optimal royalty (Theorem 1) is the *informativeness* $\mathcal{I}_j$ of the payment signal, measuring how well the noisy measurement tracks the catalog's true value to the platform. In plain terms, $\mathcal{I}_j$ is the OLS slope of true contribution on measured attribution. Formally, $\mathcal{I}_j := \mathrm{Cov}(\hat{a}_j, a_j^*)/\mathrm{Var}(\hat{a}_j)$, which depends only on the joint distribution of attribution scores $(\hat{a}_j, a_j^*)$, not on platform revenue. The definition requires only finite second moments and $\mathrm{Var}(\hat{a}_j) > 0$.

Under classical measurement error ($\hat{a}_j = a_j^* + \xi_j$, $\xi_j \perp a_j^*$, $\mathbb{E}[\xi_j] = 0$), $\mathcal{I}_j \in [0, 1]$, taking the value 0 when the signal carries no information about value and 1 when it is a perfect linear predictor. Under additive Gaussian measurement noise $\hat{a}_j = a_j^* + \xi_j$ with $\xi_j \sim \mathcal{N}(0, \sigma_{\xi,j}^2)$ independent of a_j^*, $\mathcal{I}_j$ reduces to the signal-to-noise-ratio (SNR) form $\mathcal{I}_j = \mathrm{SNR}_j\,/(1 + \mathrm{SNR}_j)$, where $\mathrm{SNR}_j := \mathrm{Var}(a_j^*)/\sigma_{\xi,j}^2$ (Scharf & Demeure, 1991). SNR measures signal quality in signal processing and information theory (Shannon, 1948; Cover & Thomas, 2006). We use this Gaussian parametrization for examples and simulations; the structural results (Theorem 1 and Proposition 1) use the covariance definition above and do not require Gaussian noise. Multiplicative scale bias or sign-flipped calibration can yield $\widehat{\mathcal{I}}_j \notin [0, 1]$; we treat such cases as calibration flags rather than as violations of the definition.

Period-total revenue π is the dollar revenue base on which royalties are calculated. We consider π as $u(x, \cdot)$ summed over the period, taking u as the platform's gross per-request revenue. We treat $\pi > 0$ as known when the contract is written; Appendix C lets revenue be random. The catalog's expected contribution is $\bar{a}_j^* := \mathbb{E}[a_j^*] > 0$, so $\mathbb{E}[\pi a_j^*] = \pi \bar{a}_j^*$ is its expected gross value. The framework also covers cases where the platform values non-monetary outcomes such as user retention or market position, in which case π is the revenue component within $u(x, \cdot)$.

Contracts and preferences. The platform offers each creator a linear contract $\theta_j = (f_j, \rho_j)$, comprising a fixed fee $f_j \geq 0$ and a royalty rate $\rho_j \in [0, 1]$, with total payment

$$w_j(\theta_j) = x_j\,[f_j + \rho_j\, v_j], \qquad v_j = \hat{a}_j\,\pi, \tag{1}$$

nesting three canonical industry forms (pure fixed fee with $\rho_j = 0$, pure royalty with $f_j = 0$, and hybrid with $f_j > 0, \rho_j > 0$).

Creator j evaluates contracts using mean-variance utility: the creator values expected payment minus a multiple of payment variance (Holmström, 1979),

$$U_j(\theta_j) = f_j + \rho_j\,\mathbb{E}[v_j] - (\alpha_j/2)\rho_j^2\,\mathrm{Var}(v_j). \tag{2}$$

This preference primitive matches a music-industry fact: liquidity-constrained creators value payment-stream stability, not just expected payment (Markowitz, 1952).

Creator j participates if $U_j(\theta_j) \geq \bar{U}_j$, where $\bar{U}_j$ is the utility the creator would obtain by not licensing to the platform (e.g., distributing independently). As more catalogs are licensed and model quality improves, AI-generated music substitutes for the creator's standalone work. The outside option therefore shrinks, $\bar{U}_j = \bar{U}_j^0 - \delta_j(x_{-j}, Q)$, where $\bar{U}_j^0$ is the creator's standalone revenue, Q is the platform's model quality, and δ_j captures the displacement loss.

3.2 EQUILIBRIUM CONTRACT AND WELFARE COST

We characterize the equilibrium contract for a single creator and quantify the welfare cost of attribution noise. The platform pays based on the noisy attributed revenue $v_j = \hat{a}_j \pi$ but receives value from $a_j^* \pi$. The platform maximizes

$$\max_{\theta_j=(f_j,\rho_j)} \ \mathbb{E}[\pi\, a_j^*] - f_j - \rho_j\, \mathbb{E}[v_j] \quad \text{s.t.} \quad U_j(\theta_j) \geq \bar{U}_j. \tag{3}$$

To accept a risky royalty instead of a fixed fee, a risk-averse creator requires a higher expected payment. If the platform offers only enough to secure participation, that additional payment reduces its profit (Lemma 1). We therefore characterize the *welfare-optimal royalty*, balancing contribution-based payment against the creator's payment risk. We call the value of criterion (4) the alignment value and measure welfare in these units. This royalty is relevant when the platform, an auditor, or a regulator values alignment between pay and contribution in addition to expected profit. Appendix B.3 also studies how royalties affect creators' incentives to improve quality. Lemma 1 formalizes the case where the platform maximizes profit alone.

Lemma 1 (Static profit-max baseline)**.** *Under* (3) *with a binding participation constraint, substituting the fee* $f_j = \bar{U}_j - \rho_j \mathbb{E}[v_j] + (\alpha_j/2)\rho_j^2 \operatorname{Var}(v_j)$ *yields* $\Pi_j(\rho_j) = \mathbb{E}[\pi\, a_j^*] - \bar{U}_j - (\alpha_j/2)\rho_j^2 \operatorname{Var}(v_j)$*, which strictly decreases in* ρ_j^2 *whenever* $\operatorname{Var}(v_j) > 0$*. The unique maximizer is the pure fixed-fee contract* $(\rho_j, f_j) = (0, \bar{U}_j)$*, feasible when* $\bar{U}_j \geq 0$*.*

Theorem 1 (Welfare-optimal royalty)**.** *Whenever* $\operatorname{Var}(v_j) > 0$*, the* risk-adjusted alignment criterion

$$\max_{\rho_j \in [0,1]} \ \rho_j \ \operatorname{Cov}(v_j, a_j^* \pi) \ - \ (\alpha_j/2)\rho_j^2 \operatorname{Var}(v_j) \tag{4}$$

admits the unique maximizer

$$\rho_j^* = \operatorname{clip}\big(\mathcal{I}_j/\alpha_j,\ [0,1]\big). \tag{5}$$

The interior solution reduces to $\rho_j^* = \mathcal{I}_j/\alpha_j$*; under additive Gaussian noise it further reduces to* $\rho_j^* = \mathrm{SNR}_j\,/[\alpha_j(1+\mathrm{SNR}_j)]$*. Proof in Appendix B.*

The welfare-optimal royalty balances alignment against risk transfer. $\rho_j^* \to \min\{1, 1/\alpha_j\}$ as $\mathcal{I}_j \to 1$ (royalty-dominant when attribution is precise), and $\rho_j^* \to 0$ as $\mathcal{I}_j \to 0$ (fixed-fee-dominant when attribution is too noisy), with hybrid contracts in between. Section 4 calibrates these regimes on measured creator pools.

Attribution precision determines the royalty rate; the creator's outside option sets the required utility. At the welfare-optimal rate, the fee that meets the participation constraint is $f_j^* = \bar{U}_j - \rho_j^* \mathbb{E}[v_j] + (\alpha_j/2)(\rho_j^*)^2 \operatorname{Var}(v_j)$. The royalty contributes $\rho_j^* \pi \bar{a}_j$ to expected payment, with $\bar{a}_j := \mathbb{E}[\hat{a}_j]$. This fee is nonnegative when $\bar{U}_j + R_j^* \geq \rho_j^* \pi \bar{a}_j$, where $R_j^* := (\alpha_j/2)(\rho_j^*)^2 \operatorname{Var}(v_j)$ is the risk premium. If this formula gives a negative fee, setting the fee to zero leaves the creator better off than their outside option. The platform may then prefer a different royalty rate. Because the fixed fee implements the transfer level, surplus-splitting variants that share the same alignment criterion leave ρ_j^* unchanged. Noisy attribution limits how closely payment can track true contribution. The next result measures the loss relative to perfect attribution.

Proposition 1 (Welfare loss from noisy attribution)**.** *Let* $\Lambda_j := \pi^2 \operatorname{Var}(a_j^*)/(2\alpha_j)$ *be the alignment value under perfect attribution and* $R_j^2 := \operatorname{Corr}(\hat{a}_j, a_j^*)^2$*. If* $\alpha_j \geq 1$ *and* $0 \leq \mathcal{I}_j \leq \alpha_j$*, the loss relative to perfect attribution is* $\mathcal{L}_j = \Lambda_j\,(1 - R_j^2)$*. Under classical measurement error (*$\hat{a}_j = a_j^* + \xi_j$ *with* $\xi_j \perp a_j^*$*),* $R_j^2 = \mathcal{I}_j$*, so* $\mathcal{L}_j = \Lambda_j(1 - \mathcal{I}_j)$ *and, holding* $\operatorname{Var}(a_j^*)$ *and* α_j *fixed, the marginal welfare gains are*

$$\left|\frac{\partial \mathcal{L}_j}{\partial \mathcal{I}_j}\right| = \Lambda_j, \qquad \left|\frac{\partial \mathcal{L}_j}{\partial\, \mathrm{SNR}_j}\right| = \Lambda_j/(1+\mathrm{SNR}_j)^2. \tag{6}$$

Per unit of SNR, the gain is largest for the least precisely measured creators. Proof and the clipped-rate case are in Appendix B.

The loss measures how much noisy attribution reduces payment-contribution alignment, not extra compensation for risk. It compares methods beyond creator rankings and does not depend on score units. It also guides investment in more precise attribution (Appendix G.5). Royalty rates still depend on how scores are scaled (Appendix A).

Relation to classical contract theory. The signal-informativeness logic behind Theorem 1 is classical (Holmström, 1979; Mussa & Rosen, 1978; Maskin & Riley, 1984). Here, the signal comes from an ML method that the platform can improve and creators can post-process. Linearity is justified under mean-variance utility with Gaussian noise (Holmström & Milgrom, 1987); alternative preferences are discussed in Appendix F and nonlinear contracts in Appendix K.

3.3 Market and policy implications

We next examine how attribution quality affects platform competition, creator participation, and the use of scores to set payments.

Platforms. Better attribution can help a platform win exclusive catalog licenses. We consider at least two platforms offering contracts simultaneously. Each platform values alignment one-for-one alongside financial profit, and its payoffs add across catalogs. For each catalog, joint surplus is the sum of creator utility and platform payoff.

Theorem 2 (Multi-platform Bertrand equilibrium)**.** *Under the conditions in Appendix E.2, there exists a Bertrand-Nash equilibrium in which each bidding platform uses the royalty rule of Theorem 1. The platform with the highest attainable joint surplus wins the catalog if that surplus covers the creator's outside option. The creator receives the greater of the outside option and the second-highest surplus, and the winner keeps the remainder. If no platform can cover the outside option, no licensing occurs. The full assumptions, formulas, and proof are in Appendix E.2.*

Competition changes the fixed fee, not the relationship between attribution quality and the royalty rate. With equal base catalog values, the most informative platform wins whenever licensing occurs.

Proposition 2 (Attribution moat)**.** *Consider the equilibrium of Theorem 2 with equal base catalog values across platforms. Assume the runner-up's surplus initially covers the creator's outside option. A trailing platform improves its attribution while rivals' informativeness stays fixed and the rate and fee conditions continue to hold. It earns zero rent until its attribution informativeness exceeds the leader's; after overtaking, it wins the catalog and earns positive rent. The full statement and proof are in Appendix E.4.*

Because catalogs are licensed as a whole, a trailing platform cannot win part of a catalog. When the runner-up's offer determines creator utility, improvements below the threshold transfer rent to the creator without increasing total surplus (Appendix E.4).

Creators. Uniform compensation can drive out creators with valuable alternatives to licensing. The loss is especially costly when those catalogs also contribute styles that complement the remaining catalog. Proposition 3 states when this happens.

Proposition 3 (Selective exit under uniform contracts)**.** *Under a uniform contract, creators with higher outside options exit first when expected attributed revenue, its variance, and risk aversion are held fixed. If the exiting catalogs are complements, the loss of gross platform value is at least the sum of their individual contributions to the remaining catalog. The inequality is strict when at least two catalogs exit and their complementarities are strict. The full statement and proof are in Appendix D.1.*

Thus, distinctive creators are especially costly to lose when they have both high outside options and high complementary value. Their joint exit removes combinations that the remaining catalogs cannot supply. Even within personalized contracts, a higher royalty need not induce more effort. At sufficiently high rates, volatility risk outweighs the higher expected reward (Appendix B.3).

AI researchers. Methods used to set creator payments should be evaluated on welfare, not just rank correlation. Two methods that rank creators identically can still differ in how closely their scores track true contribution. The welfare loss $\mathcal{L}_j$ (Proposition 1) measures this difference through the Pearson correlation. Setting royalty rates also requires the calibration scale, which can differ by orders of magnitude across methods. Per unit of SNR improvement, the largest gains go to the creators measured least precisely today, holding contribution variance and risk aversion fixed. The investment rule uses these gains to decide which creators' attribution estimates to improve (Corollary 4 in Appendix G.5).

Regulation and institutional design. Creators who report their own attribution scores can reduce their spread without changing their mean. This lowers payment risk without reducing expected pay.

Proposition 4 (Score-smoothing manipulation cannot be deterred)**.** *Fix a linear contract with a positive royalty and a risk-averse creator. Assume scores have positive variance, revenue is independent of the scores, and both have finite second moments with* $\mathbb{E}[\pi^2] > 0$*. Any nontrivial, costless shrinkage of reported scores toward their mean strictly increases the creator's utility. No fixed fee and positive royalty rate can deter this manipulation; the full statement and proof are in Appendix C.1.*

Payment should therefore rest on attribution scores computed by the platform or an independent auditor rather than on unverified creator reports. A variance test can also detect smoothing under the conditions in Appendix C.4.

When creators and platforms hold different beliefs about attribution quality, creators may misjudge their royalty risk, which can distort their effort and participation. We do not model these beliefs, and transparency about the platform's measured $\widehat{\mathcal{I}}_j$ is a natural remedy.

Other licensing forms. A platform that already pays on usage can incorporate attribution without discarding the usage signal. The multiple-signal extension assigns each signal weight according to the information it adds about true contribution (Proposition 12, Appendix K). Usage counts can index contingent payments, while a flat-fee blanket license uses the fixed-fee component alone. Blanket licensing specifies the repertoire covered, not necessarily the payment schedule. Richer contract forms cannot recover contribution information absent from their payment signals. Appendix K bounds payment-contribution covariance for a given payment variance, allowing nonlinear as well as linear contracts.

4 Measuring attribution informativeness

We now apply the framework to ask whether current attribution is informative enough to support royalty payments. We measure attribution with our leave-one-catalog-out pipeline in two settings, audio latent diffusion (Stable Audio Open (Evans et al., 2025)) with EKFAC (Mlodozeniec et al., 2025) and D-TRAK (Zheng et al., 2024), and a symbolic language model (Anticipatory Music Transformer (Thickstun et al., 2024)) with LoGra (Choe et al., 2025) and TRAK (Park et al., 2023). A market-simulation companion in Appendix G.6 draws priors from the same measurement.

4.1 Setting 1: Latent-diffusion audio attribution

Setup. We fine-tune Stable Audio Open (Evans et al., 2025) on 50 artist catalogs from MTG-Jamendo (Bogdanov et al., 2019), retraining 50 leave-catalog-out variants alongside one full-catalog model, and evaluate on 200 generated outputs. Our reference is each catalog's leave-out change in model loss, evaluated on the same full-model generations (formal definition and sanity checks in Appendix G.1). For each method (EKFAC (Mlodozeniec et al., 2025) or D-TRAK (Zheng et al., 2024)) and creator, we report informativeness $\widehat{\mathcal{I}}_j$ from §3, calibration scale $\widehat{K}_j$ (regression slope of measured attribution on truth), and calibrated SNR $\widehat{\mathrm{SNR}}_j^{\mathrm{cal}} = \widehat{\mathrm{Var}}(a_j^*)/\widehat{\mathrm{Var}}(\hat{a}_j/\widehat{K}_j - a_j^*)$. Headline calibration is fitted and evaluated on the same queries in both settings.

Rank metrics miss signal that matters for payment. Attribution methods are usually judged by how well they rank training data, but payment depends on how closely scores track the size of each creator's contribution. Both audio methods have higher mean Pearson than Spearman correlations with the ground truth on the 50-catalog experiment, 0.109 versus 0.007 for EKFAC and 0.125 versus 0.031 for D-TRAK (Table 2). Rank-based evaluation (Park et al., 2023; Ilyas et al., 2022) measures ordering, not score calibration.

Rank correlation and calibrated precision favor different methods. D-TRAK has higher mean Spearman correlation than EKFAC, but EKFAC has higher median calibrated SNR (0.023 versus 0.009). Spearman compares catalogs within each query; calibrated SNR measures each catalog across queries. The raw scores of the two methods differ in scale by roughly $10^3\times$. We calibrate the scores before comparing their precision. These two summaries favor different methods.

Current attribution favors fixed-fee-dominant contracts. Median calibrated SNR is low under both audio methods. At an illustrative risk aversion of $\alpha = 2$, the median creator's welfare-optimal contract is therefore close to a pure fixed fee. On held-out audio queries, the non-random methods

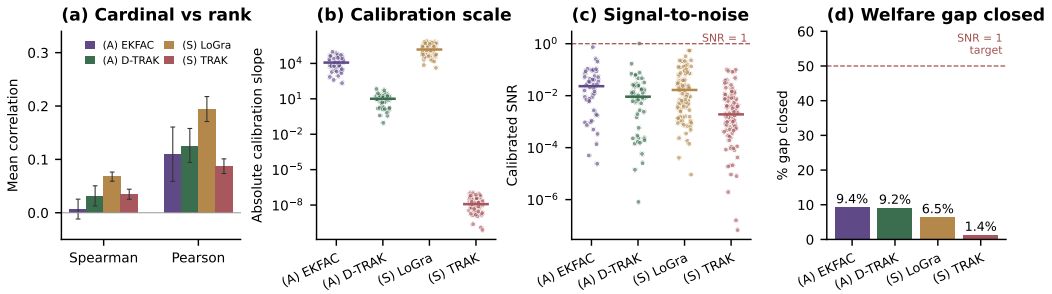

Figure 2: Attribution on Stable Audio Open (50 catalogs, EKFAC/D-TRAK) and Anticipatory Music Transformer (100 artists, LoGra/TRAK). **(a)** Mean per-query Spearman and Pearson correlations with the leave-catalog-out reference, with 95% query-bootstrap intervals. **(b)** Creator calibration scales (log axis). **(c)** Calibrated SNR distributions; dashed line at 1. **(d)** Welfare gap closed; dashed line at 50%, corresponding to $\mathrm{SNR} = 1$ for every creator.

retain positive aggregate prediction gains. Their query-bootstrap intervals include zero, and the median catalog's gain is negative for every method (Appendix G.3).

4.2 Setting 2: Symbolic music attribution

We also conduct pilot measurements in the symbolic domain, consistent with parameters of a commercially deployed music co-creation product (Donahue et al., 2025; Kim et al., 2025). The pilot fine-tunes an Anticipatory Music Transformer (Thickstun et al., 2024) on 100 TheoryTab artists, evaluates 500 generated outputs under the same leave-catalog-out protocol, and measures the same three quantities using LoGra (Choe et al., 2025) and TRAK (Park et al., 2023).

Symbolic attribution also has low precision. On symbolic data, median calibrated SNR is 0.017 for LoGra and 0.002 for TRAK. LoGra outperforms TRAK on mean Spearman and Pearson correlations and median calibrated SNR (Table 2).

4.3 Takeaways

Low precision favors fixed-fee-dominant contracts. Both settings show low calibrated attribution precision, so the shortfall is not confined to audio diffusion. At the illustrative risk-aversion levels considered here, this makes the welfare-optimal contract fixed-fee-dominant.

Low precision extends beyond gradient-based attribution. Random, CLAP, and Fingerprint also have low median calibrated SNR in both settings. These baselines use no model gradients and share the same leave-one-catalog-out reference and metrics (Table 2). Random draws independent $\mathrm{Uniform}[0, 1]$ scores for training-item-query pairs and sums them within catalogs. CLAP measures audio-audio embedding similarity, with symbolic outputs rendered to audio first. Fingerprint measures near-copy similarity through Chromaprint.

Better rankings do not imply more precise attribution. On the expanded audio setting, CLAP has higher mean Spearman correlation than EKFAC (0.046 versus 0.007), while EKFAC has higher median calibrated SNR (0.023 versus 0.016). For comparison, Table 6 reports these baselines on the original 22-catalog audio pilot.

4.4 Musical market simulation

We apply the welfare-optimal payment rule to the 50-catalog measurements. The resulting contracts are model-implied, not observed licensing outcomes. We then estimate how much attribution precision must improve to support higher royalty rates.

Most creators have low royalty rates at $\alpha = 2$. At $\alpha = 2$, 48 of 50 creators under EKFAC and 49 under D-TRAK have $\rho_j^* < 0.1$ (Figure 3(a,b)). The median coefficients are 0.011 and 0.0045, respectively. These coefficients favor fixed-fee-dominant contracts at $\alpha = 2$. The Setting 2 (symbolic)

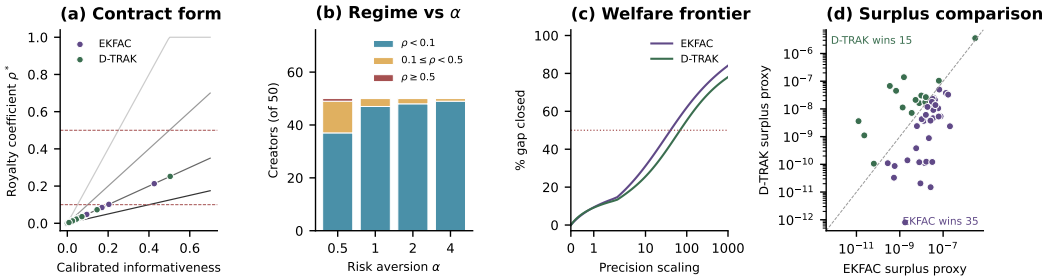

Figure 3: Welfare-optimal contracts for 50 catalogs under EKFAC and D-TRAK. **(a)** Royalty versus calibrated informativeness at $\alpha = 2$; grey curves use $\alpha \in \{0.5, 1, 2, 4\}$. Dashed lines at 0.1 and 0.5 define coefficient bands, not optimality thresholds. **(b)** EKFAC catalog counts by band and risk aversion. **(c)** Welfare gap closed as calibrated SNR increases. **(d)** Creator surplus proxies under both methods (log-log axes).

population yields similarly fixed-fee-dominant welfare-optimal contracts. These rates apply to sign-calibrated attribution scores. For each creator j and risk-aversion level α the welfare-optimal royalty is $\rho_j^* = \widehat{\mathcal{I}}_j^{\text{cal}}/\alpha$ after clipping to $[0, 1]$ (regime classification in Figure 3(a)). For the 19 EKFAC and 16 D-TRAK creators with $\widehat{K}_j < 0$, payment uses the sign-reversed raw score, with the sign estimated in-sample. Restricting the contract to the raw score sets their rates to zero and, at $\alpha = 2$, lowers the gap closed from 9.36% to 8.65% (EKFAC) and from 9.18% to 8.84% (D-TRAK).

Table 1: Music payment models in the contract space $\theta_j = (f_j, \rho_j)$. Fixed fees and royalties occupy its boundaries; AI training licenses can combine both.

Industry model	Payment form	$\theta = (f, \rho)$	Attribution signal	SNR regime
Blanket license (ASCAP/BMI)	Fixed fee	$(f, 0)$	None	N/A
Pro-rata streaming (Spotify)	Royalty	$(0, \rho)$	Stream counts	Low (coarse)
Sync license (film/TV)	Fixed fee	$(f, 0)$	Negotiated	N/A
User-centric (SoundCloud)	Royalty	$(0, \rho)$	Per-user streams	Medium
AI training license (proposed)	Hybrid	(f, ρ)	Attribution method	Varies

Closing the welfare gap requires large precision gains. We ask how much SNR improvement would be needed to close more of the gap (Figure 3(c)). The median creator has calibrated SNR 0.023 under EKFAC and 0.009 under D-TRAK. Raising the median SNR to 1 takes approximately 43× and 109× improvements, respectively. Under the plotted gap-closed measure, reaching 50% requires approximately 40× and 69× improvements. These targets concern the median creator and the aggregate, respectively.

The winning platform differs across creators. The EKFAC platform wins 35 of 50 catalogs and the D-TRAK platform wins 15 (Figure 3(d)). These outcomes use the Bertrand model (Theorem 2) and the surplus proxy in Appendix G.6. The per-creator surplus gap illustrates Proposition 2's dead zone on a measured pool rather than a hypothetical one. Full per-creator analysis is in Appendix G.6.

5 Conclusion

We propose a framework for compensating music creators whose work trains generative AI, where each creator's contract form follows from how accurately the platform can trace generated outputs back to their catalog. Attribution technology shapes who gets paid. Our framework helps platforms, creators, and policymakers assess that choice. Future work could extend it to catalogs that grow over time, generative models spanning music and other media, and institutional setups for attribution oversight.

AI USE STATEMENT

We used AI tools to polish the writing and help identify potentially missing related work. They also assisted with drafting parts of the related-work section and selected experimental subsections. The authors verified the suggested references and reviewed and rewrote all AI-assisted text. After the authors completed the mathematical proofs, AI tools were used to proofread the arguments and flag possible errors. The authors checked these suggestions and take full responsibility for the final content.

ETHICS STATEMENT

This work follows the ICLR Code of Ethics. It involves no human subjects or personally identifiable information, and all models and datasets we use are publicly released and used under their respective licenses. Our aim is to support Attribution-Aware compensation for music creators whose work trains generative models. Because attribution estimates are noisy, payments based on them can misallocate income, so attribution used for payment should be open to independent verification.

REPRODUCIBILITY STATEMENT

Section 3 states each theoretical result with its assumptions, and the appendix gives complete proofs (Appendices B to E, H, I and K). Section 4 and Appendix G.1 describe the datasets, catalog selection, model fine-tuning, attribution methods, and leave-one-catalog-out ground truth, and Appendices G.2, G.3 and G.6 give further experimental details. All models and datasets are publicly available, and we will release our code upon publication.

REFERENCES

Daron Acemoglu, Ali Makhdoumi, Azarakhsh Malekian, and Asu Ozdaglar. Too much data: Prices and inefficiencies in data markets. *American Economic Journal: Microeconomics*, 14 (4):218–56, November 2022. doi: 10.1257/mic.20200200. URL https://www.aeaweb.org/articles?id=10.1257/mic.20200200.

Anish Agarwal, Munther Dahleh, and Tuhin Sarkar. A marketplace for data: An algorithmic solution. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, EC '19, pp. 701–726, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450367929. doi: 10.1145/3328526.3329589. URL https://doi.org/10.1145/3328526.3329589.

Andrea Agostinelli, Timo I. Denk, Zalán Borsos, Jesse Engel, Mauro Verzetti, Antoine Caillon, Qingqing Huang, Aren Jansen, Adam Roberts, Marco Tagliasacchi, Matt Sharifi, Neil Zeghidour, and Christian Frank. MusicLM: Generating music from text, 2023. URL https://arxiv.org/abs/2301.11325.

Juhan Bae, Wu Lin, Jonathan Lorraine, and Roger Baker Grosse. Training data attribution via approximate unrolling. In *The Thirty-eighth Annual Conference on Neural Information Processing Systems*, 2024. URL https://openreview.net/forum?id=3NaqGg92KZ.

Julia Barnett, Hugo Flores García, and Bryan Pardo. Exploring musical roots: Applying audio embeddings to empower influence attribution for a generative music model. In Blair Kaneshiro, Gautham J. Mysore, Oriol Nieto, Chris Donahue, Cheng-Zhi Anna Huang, Jin Ha Lee, Brian McFee, and Matthew C. McCallum (eds.), *Proceedings of the 25th International Society for Music Information Retrieval Conference, ISMIR 2024, San Francisco, California, USA and Online, November 10-14, 2024*, pp. 360–368, 2024. doi: 10.5281/ZENODO.14877350. URL https://doi.org/10.5281/zenodo.14877350.

Dirk Bergemann, Alessandro Bonatti, and Alex Smolin. The design and price of information. *American Economic Review*, 108(1):1–48, 2018. doi: 10.1257/aer.20161079.

Dmitry Bogdanov, Minz Won, Philip Tovstogan, Alastair Porter, and Xavier Serra. The MTG-Jamendo dataset for automatic music tagging. In *Machine Learning for Music Discovery Workshop, International Conference on Machine Learning (ICML 2019)*, Long Beach, CA, United States, 2019. URL http://hdl.handle.net/10230/42015.

Antoine Caillon, Brian McWilliams, Cassie Tarakajian, Ian Simon, Ilaria Manco, Jesse Engel, Noah Constant, Yunpeng Li, Timo I. Denk, Alberto Lalama, Andrea Agostinelli, Cheng-Zhi Anna Huang, Ethan Manilow, George Brower, Hakan Erdogan, Heidi Lei, Itai Rolnick, Ivan Grishchenko, Manu Orsini, Matej Kastelic, Mauricio Zuluaga, Mauro Verzetti, Michael Dooley, Ondrej Skopek, Rafael Ferrer, Zalán Borsos, Aaron van den Oord, Douglas Eck, Eli Collins, Jason Michael Baldridge, Tom Hume, Chris Donahue, Kehang Han, and Adam Roberts. Live music models. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems Creative AI Track: Humanity*, 2025. URL https://openreview.net/forum?id=SB3XbPgvpl.

Junjie Chen, Minming Li, and Haifeng Xu. Selling data to a machine learner: Pricing via costly signaling. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 3336–3359. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/chen22m.html.

Sang Keun Choe, Hwijeen Ahn, Juhan Bae, Kewen Zhao, Youngseog Chung, Adithya Pratapa, Willie Neiswanger, Emma Strubell, Teruko Mitamura, Jeff Schneider, Eduard Hovy, Roger Baker Grosse, and Eric P. Xing. What is your data worth to GPT? LLM-scale data valuation with influence functions. In *The Thirty-ninth Annual Conference on Neural Information Processing Systems*, 2025. URL https://openreview.net/forum?id=zPKeJAEo27.

Woosung Choi, Junghyun Koo, Kin Wai Cheuk, Joan Serrà, Marco Martínez-Ramírez, Yukara Ikemiya, Naoki Murata, Yuhta Takida, Wei-Hsiang Liao, and Yuki Mitsufuji. Large-scale training data attribution for music generative models via unlearning. In D. Belgrave, C. Zhang, M. Coelho, L. Elliott, Y. Tian, and P. Prakash (eds.), *Advances*

in Neural Information Processing Systems, volume 38, Creative AI, pp. 327–335, 2025. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/5392d643eccfce4025f6dd567af05a16-Paper-Creative_AI_Track.pdf.

Lin William Cong, Wenshi Wei, Danxia Xie, and Longtian Zhang. Endogenous growth under multiple uses of data. *Journal of Economic Dynamics and Control*, 141:104395, 2022. ISSN 0165-1889. doi: 10.1016/j.jedc.2022.104395. URL https://www.sciencedirect.com/science/article/pii/S0165188922001014. Markets and Economies with Information Frictions.

Jade Copet, Felix Kreuk, Itai Gat, Tal Remez, David Kant, Gabriel Synnaeve, Yossi Adi, and Alexandre Défossez. Simple and controllable music generation. In *Proceedings of the 37th International Conference on Neural Information Processing Systems*, NIPS '23, Red Hook, NY, USA, 2023. Curran Associates Inc.

Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory*. Wiley, second edition, 2006. doi: 10.1002/047174882x.

Junwei Deng, Xirui Jiang, Shiyuan Zhang, Shichang Zhang, Himabindu Lakkaraju, Ruijiang Gao, Chris Donahue, and Jiaqi W. Ma. Computational copyright: Towards a royalty model for music generative AI, 2025. URL https://arxiv.org/abs/2312.06646.

Prafulla Dhariwal, Heewoo Jun, Christine Payne, Jong Wook Kim, Alec Radford, and Ilya Sutskever. Jukebox: A generative model for music, 2020. URL https://arxiv.org/abs/2005.00341.

Sander Dieleman, Aäron van den Oord, and Karen Simonyan. The challenge of realistic music generation: modelling raw audio at scale. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, pp. 8000–8010, Red Hook, NY, USA, 2018. Curran Associates Inc.

Chris Donahue, John Thickstun, and Percy Liang. Melody transcription via generative pre-training. In Preeti Rao, Hema A. Murthy, Ajay Srinivasamurthy, Rachel M. Bittner, Rafael Caro Repetto, Masataka Goto, Xavier Serra, and Marius Miron (eds.), *Proceedings of the 23rd International Society for Music Information Retrieval Conference, ISMIR 2022, Bengaluru, India, December 4-8, 2022*, pp. 485–492, 2022. URL https://archives.ismir.net/ismir2022/paper/000058.pdf.

Chris Donahue, Shih-Lun Wu, Yewon Kim, Dave Carlton, Ryan Miyakawa, and John Thickstun. Hookpad Aria: A copilot for songwriters, 2025. URL https://arxiv.org/abs/2502.08122.

Logan Engstrom, Axel Feldmann, and Aleksander Madry. DsDm: Model-Aware dataset selection with datamodels. In *Proceedings of the 41st International Conference on Machine Learning*, volume 235, pp. 12491–12526, 2024. URL https://proceedings.mlr.press/v235/engstrom24a.html.

European Parliament and Council of the European Union. Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence (AI Act). Technical report, European Parliament and Council of the European Union, 2024. URL https://eur-lex.europa.eu/eli/reg/2024/1689/oj. Official Journal of the European Union.

Zach Evans, Julian Parker, CJ Carr, Zack Zukowski, Josiah Taylor, and Jordi Pons. Stable Audio Open. In *ICASSP 2025 - 2025 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 1–5, 2025. URL https://api.semanticscholar.org/CorpusID:271310050.

Seth Forsgren and Hayk Martiros. Riffusion - Stable diffusion for real-time music generation, 2022. URL https://riffusion.com/about.

Kristian Georgiev, Joshua Vendrow, Hadi Salman, Sung Min Park, and Aleksander Madry. The journey, not the destination: How data guides diffusion models, 2023. URL https://arxiv.org/abs/2312.06205.

Amirata Ghorbani and James Zou. Data Shapley: Equitable valuation of data for machine learning. In *Proceedings of the 36th International Conference on Machine Learning*, volume 97, pp. 2242–2251, 2019. URL https://proceedings.mlr.press/v97/ghorbani19c.html.

Ryan Giordano, William Stephenson, Runjing Liu, Michael Jordan, and Tamara Broderick. A Swiss army infinitesimal jackknife. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1139–1147. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/giordano19a.html.

Junmin Gong, Yulin Song, Wenxiao Zhao, Sen Wang, Shengyuan Xu, Jing Guo, and Xuerui Yang. ACE-Step 1.5: Pushing the boundaries of Open-Source music generation, 2026. URL https://arxiv.org/abs/2602.00744.

Roger Grosse, Juhan Bae, Cem Anil, Nelson Elhage, Alex Tamkin, Amirhossein Tajdini, Benoit Steiner, Dustin Li, Esin Durmus, Ethan Perez, Evan Hubinger, Kamilė Lukošiūtė, Karina Nguyen, Nicholas Joseph, Sam McCandlish, Jared Kaplan, and Samuel R. Bowman. Studying large language model generalization with influence functions, 2023. URL https://arxiv.org/abs/2308.03296.

Guru Guruganesh, Jon Schneider, and Joshua R. Wang. Contracts under moral hazard and adverse selection. In *Proceedings of the 22nd ACM Conference on Economics and Computation*, EC '21, pp. 563–582, New York, NY, USA, 2021. Association for Computing Machinery. ISBN 9781450385541. doi: 10.1145/3465456.3467637. URL https://doi.org/10.1145/3465456.3467637.

Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Bengt Holmström. Moral hazard and observability. *The Bell Journal of Economics*, 10(1):74–91, 1979. doi: 10.2307/3003320.

Bengt Holmström and Paul Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55(2):303–328, 1987. doi: 10.2307/1913238.

Andrew Ilyas, Sung Min Park, Logan Engstrom, Guillaume Leclerc, and Aleksander Madry. Datamodels: Understanding predictions with data and data with predictions. In Kamalika Chaudhuri, Stefanie Jegelka, Le Song, Csaba Szepesvari, Gang Niu, and Sivan Sabato (eds.), *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pp. 9525–9587. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/ilyas22a.html.

Ruoxi Jia, David Dao, Boxin Wang, Frances Ann Hubis, Nick Hynes, Nezihe Merve Gürel, Bo Li, Ce Zhang, Dawn Song, and Costas J. Spanos. Towards efficient data valuation based on the Shapley value. In Kamalika Chaudhuri and Masashi Sugiyama (eds.), *Proceedings of the Twenty-Second International Conference on Artificial Intelligence and Statistics*, volume 89 of *Proceedings of Machine Learning Research*, pp. 1167–1176. PMLR, 16–18 Apr 2019. URL https://proceedings.mlr.press/v89/jia19a.html.

Charles I. Jones and Christopher Tonetti. Nonrivalry and the economics of data. *American Economic Review*, 110(9):2819–2858, 2020. doi: 10.1257/aer.20191330.

Maribel Khouwes. Deezer: AI-generated tracks now represent 44% of all new uploaded music. Deezer Newsroom, 2026. URL https://newsroom-deezer.com/2026/04/ai-generated-tracks-represent-44-of-new-uploaded-music/. Published April 20, 2026.

Yewon Kim, Sung-Ju Lee, and Chris Donahue. Amuse: Human-AI collaborative songwriting with multimodal inspirations. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems*, CHI '25, New York, NY, USA, 2025. Association for Computing Machinery. ISBN 9798400713941. doi: 10.1145/3706598.3713818. URL https://doi.org/10.1145/3706598.3713818.

Pang Wei Koh and Percy Liang. Understanding black-box predictions via influence functions. In Doina Precup and Yee Whye Teh (eds.), *Proceedings of the 34th International Conference on Machine Learning*, volume 70 of *Proceedings of Machine Learning Research*, pp. 1885–1894. PMLR, 06–11 Aug 2017. URL https://proceedings.mlr.press/v70/koh17a.html.

Yongchan Kwon, Eric Wu, Kevin Wu, and James Zou. DataInf: Efficiently estimating data influence in LoRA-tuned LLMs and diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=9m02ib92Wz.

Harry Markowitz. Portfolio selection. *The Journal of Finance*, 7(1):77–91, 1952. doi: 10.1111/j.1540-6261.1952.tb01525.x.

Eric Maskin and John Riley. Monopoly with incomplete information. *The RAND Journal of Economics*, 15(2):171–196, 1984. doi: 10.2307/2555674.

Bruno Kacper Mlodozeniec, Runa Eschenhagen, Juhan Bae, Alexander Immer, David Krueger, and Richard E. Turner. Influence functions for scalable data attribution in diffusion models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=esYrEndGsr.

Michael Mussa and Sherwin Rosen. Monopoly and product quality. *Journal of Economic Theory*, 18 (2):301–317, 1978. doi: 10.1016/0022-0531(78)90085-6.

Sung Min Park, Kristian Georgiev, Andrew Ilyas, Guillaume Leclerc, and Aleksander Mądry. TRAK: Attributing model behavior at scale. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

John W. Pratt. Risk aversion in the small and in the large. *Econometrica*, 32(1/2):122–136, 1964. doi: 10.2307/1913738.

Garima Pruthi, Frederick Liu, Satyen Kale, and Mukund Sundararajan. Estimating training data influence by tracing gradient descent. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*, NIPS '20, Red Hook, NY, USA, 2020. Curran Associates Inc. ISBN 9781713829546.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-Resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 10684–10695, 2022.

Louis L. Scharf and Cédric Demeure. *Statistical signal processing: detection, estimation, and time series analysis*. Addison-Wesley, 1991.

C. E. Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27(3): 379–423, 1948. doi: 10.1002/j.1538-7305.1948.tb01338.x.

Amanda Silberling. AI music generator Suno hits 2M paid subscribers and $300M in annual recurring revenue. TechCrunch, 2026. URL https://techcrunch.com/2026/02/27/ai-music-generator-suno-hits-2-million-paid-subscribers-and-300m-in-annual-recurring-revenue/. Published February 27, 2026.

Janne Spijkervet and John Ashley Burgoyne. Contrastive learning of musical representations. In Jin Ha Lee, Alexander Lerch, Zhiyao Duan, Juhan Nam, Preeti Rao, Peter van Kranenburg, and Ajay Srinivasamurthy (eds.), *Proceedings of the 22nd International Society for Music Information Retrieval Conference, ISMIR 2021, Online, November 7-12, 2021*, pp. 673–681, 2021. URL https://archives.ismir.net/ismir2021/paper/000084.pdf.

John Thickstun, David Leo Wright Hall, Chris Donahue, and Percy Liang. Anticipatory Music Transformer. *Transactions on Machine Learning Research*, 2024. ISSN 2835-8856. URL https://openreview.net/forum?id=EBNJ33Fcrl.

Aaron van den Oord, Oriol Vinyals, and Koray Kavukcuoglu. Neural discrete representation learning. In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, NIPS'17, pp. 6309–6318, Red Hook, NY, USA, 2017. Curran Associates Inc. ISBN 9781510860964.

Avery Wang. An industrial strength audio search algorithm. In *ISMIR 2003, 4th International Conference on Music Information Retrieval, Baltimore, Maryland, USA, October 27-30, 2003, Proceedings*, 2003.

Jiachen T. Wang, Zhun Deng, Hiroaki Chiba-Okabe, Boaz Barak, and Weijie J. Su. An economic solution to copyright challenges of generative AI, 2024. URL https://arxiv.org/abs/2404.13964.

Sheng-Yu Wang, Alexei A. Efros, Jun-Yan Zhu, and Richard Zhang. Evaluating data attribution for text-to-image models. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 7158–7169, 2023. doi: 10.1109/ICCV51070.2023.00661.

Xinhe Wang, Pingbang Hu, Junwei Deng, and Jiaqi W. Ma. Adversarial attacks on data attribution. In *The Thirteenth International Conference on Learning Representations*, 2025. URL https://openreview.net/forum?id=oJgIRwkIUB.

Luyang Zhang, Cathy Jiao, Beibei Li, and Chenyan Xiong. Fairshare data pricing via data valuation for large language models. In *Advances in Neural Information Processing Systems*, volume 38, Main Conference, pp. 33306–33345, 2025. doi: 10.52202/085713-1119. URL https://proceedings.neurips.cc/paper_files/paper/2025/file/2fed719c7af739e5a8615170629ed7fc-Paper-Conference.pdf.

Xiaosen Zheng, Tianyu Pang, Chao Du, Jing Jiang, and Min Lin. Intriguing properties of data attribution on diffusion models. In *The Twelfth International Conference on Learning Representations*, 2024. URL https://openreview.net/forum?id=vKViCoKGcB.

A GAUSSIAN SPECIAL CASE AND SIMPLIFYING CONDITIONS

The covariance formulas use second moments of $(\hat{a}_j, a_j^*)$ under the assumptions of the corresponding results. The closed-form SNR expressions used in examples and simulations rely on the additive Gaussian noise model below, and the remaining paragraphs record the calibration regime, the random-revenue extension, and the non-additive settings to which the selective-exit result applies.

Additive Gaussian noise. $\hat{a}_j = a_j^* + \xi_j$, with $\xi_j \sim \mathcal{N}(0, \sigma^2_{\xi,j})$ independent of a_j^* and π. This yields $\mathcal{I}_j = \mathrm{SNR}_j / (1 + \mathrm{SNR}_j)$ and the interior rate $\rho_j^* = \frac{1}{\alpha_j} \cdot \frac{\mathrm{SNR}_j}{1+\mathrm{SNR}_j}$. Under this model, the welfare loss in Proposition 1 reduces to the Gaussian-shortcut form

$$\mathcal{L}_j = \frac{\pi^2 \ \mathrm{Var}(a_j^*)}{2\alpha_j \, (1 + \mathrm{SNR}_j)}, \tag{7}$$

and the marginal welfare gain from improving SNR_j is

$$\left| \frac{\partial \mathcal{L}_j}{\partial \, \mathrm{SNR}_j} \right| = \frac{\pi^2 \ \mathrm{Var}(a_j^*)}{2\alpha_j \, (1 + \mathrm{SNR}_j)^2}, \tag{8}$$

which approaches its maximum Λ_j as $\mathrm{SNR}_j \to 0$.

Multiplicative-bias regime and calibration. A common generalization is $\hat{a}_j = K_j \cdot a_j^* + \xi_j$ with $\xi_j \perp a_j^*$ and $\xi_j \sim \mathcal{N}(0, \sigma^2_{\xi,j})$, which arises naturally for gradient-based attribution methods that estimate scores up to a method-specific scale (Section 3.1). Under this model, the calibrated estimator $\tilde{a}_j = \hat{a}_j / K_j$ satisfies $\tilde{a}_j = a_j^* + \xi_j / K_j$, returning the additive Gaussian regime with $\mathrm{SNR}_j^{\mathrm{cal}} = K_j^2 \cdot \mathrm{SNR}_j$. The welfare-optimal royalty should be evaluated on the calibrated signal when the Gaussian shortcut is used: $\mathcal{I}_j^{\mathrm{cal}} = \mathrm{SNR}_j^{\mathrm{cal}} / (1 + \mathrm{SNR}_j^{\mathrm{cal}})$, while the raw uncalibrated regression slope satisfies $\mathcal{I}_j^{\mathrm{raw}} = \mathrm{SNR}_j^{\mathrm{cal}} / [K_j(1 + \mathrm{SNR}_j^{\mathrm{cal}})]$. Thus the welfare-loss formula in Proposition 1 requires the calibrated $\widehat{\mathrm{SNR}}^{\mathrm{cal}}$ when expressed via the Gaussian shortcut. The proof is a one-line algebra exercise on $\mathrm{Cov}(\hat{a}, a^*) = K_j \, \mathrm{Var}(a^*)$ and $\mathrm{Var}(\hat{a}) = K_j^2 \, \mathrm{Var}(a^*) + \mathrm{Var}(\xi)$.

Negative calibration estimates ($\widehat{K}_j < 0$, equivalently $\widehat{\mathcal{I}}_j < 0$; Table 5 reports how often they occur) mean that the raw score moves against contribution, so a contract on the raw score sets $\rho_j^* = 0$, while a contract on the sign-calibrated score $\tilde{a}_j = \hat{a}_j / \widehat{K}_j$ has nonnegative informativeness, equal to the squared correlation between $\hat{a}_j$ and a_j^*.

Random revenue. If period revenue is random with $\pi \perp (\hat{a}_j, a_j^*)$, Theorem 1 holds exactly when the royalty is paid on centered attribution times revenue, $v_j = (\hat{a}_j - \bar{a}_j)\pi$, since then $\mathrm{Cov}(v_j, a_j^* \pi) = \mathbb{E}[\pi^2] \, \mathrm{Cov}(\hat{a}_j, a_j^*)$ and $\mathrm{Var}(v_j) = \mathbb{E}[\pi^2] \, \mathrm{Var}(\hat{a}_j)$. Centering attributed revenue instead, or leaving attribution uncentered, adds terms proportional to $\mathrm{Var}(\pi)$ to both moments whenever $\bar{a}_j \neq 0$ (Lemma 2), and the welfare-optimal royalty is then no longer $\mathcal{I}_j / \alpha_j$.

Non-additive utility. The selective-exit result (Proposition 3) and the cross-creator substitution argument apply to settings where $u(x, r)$ is non-additive across creators: some creators are complements (superadditive), others substitutes (submodular). This is a property of the platform's value function, not a distributional assumption on the attribution signal.

B FULL PROOFS AND THE QUALITY-CHOICE EXTENSION

We give the proofs of Theorem 1 and Proposition 1, followed by an extension to endogenous effort.

B.1 PROOF OF THEOREM 1

Proof of Theorem 1. The theorem maximizes the risk-adjusted alignment criterion (4), which is distinct from the platform-profit objective (3). The latter is handled by Lemma 1 and gives the pure fixed-fee baseline.

With period revenue π known (Section 3.1), for any means of $(\hat{a}_j, a_j^*)$,

$$\mathrm{Cov}(v_j, a_j^* \pi) = \pi^2 \, \mathrm{Cov}(\hat{a}_j, a_j^*), \qquad \mathrm{Var}(v_j) = \pi^2 \, \mathrm{Var}(\hat{a}_j).$$

The first-order condition for (4) gives the interior maximizer

$$\rho_j^{\text{int}} = \frac{\text{Cov}(v_j, a_j^*\pi)}{\alpha_j \,\text{Var}(v_j)} = \frac{\text{Cov}(\hat{a}_j, a_j^*)}{\alpha_j \,\text{Var}(\hat{a}_j)} = \frac{\mathcal{I}_j}{\alpha_j}.$$

The second-order condition is $-\alpha_j \,\text{Var}(v_j) < 0$, so the criterion is strictly concave whenever $\text{Var}(v_j) > 0$. Imposing the contract-space constraint $\rho_j \in [0,1]$ projects the interior maximizer onto the feasible interval, yielding $\rho_j^* = \text{clip}(\mathcal{I}_j/\alpha_j, [0,1])$.

For the Gaussian special case, $\hat{a}_j = a_j^* + \xi_j$ with ξ_j independent of a_j^* gives $\mathcal{I}_j = \text{SNR}_j\,/(1+\text{SNR}_j)$, so the interior welfare-optimal royalty is $\rho_j^* = \text{SNR}_j\,/[\alpha_j(1+\text{SNR}_j)]$. □

B.2 Proof of Proposition 1

Proof of Proposition 1. For the welfare-optimal royalty, the reduced criterion as a function of ρ_j is

$$J_j(\rho_j) = -\frac{\alpha_j}{2}\rho_j^2 \,\text{Var}(v_j) + \rho_j \,\text{Cov}(v_j, a_j^*\pi),$$

up to terms independent of ρ_j. With π known, the covariance identity $\text{Cov}(v_j, a_j^*\pi) = \mathcal{I}_j \,\text{Var}(v_j)$ holds. Completing the square around $\rho_j^* = \mathcal{I}_j/\alpha_j$ gives

$$-\frac{\alpha_j}{2}\rho_j^2 \,\text{Var}(v_j) + \rho_j \mathcal{I}_j \,\text{Var}(v_j) = -\frac{\alpha_j \,\text{Var}(v_j)}{2}\big(\rho_j - \rho_j^*\big)^2 + \frac{\text{Var}(v_j)}{2\alpha_j}\mathcal{I}_j^2.$$

Thus

$$J_j(\rho_j) = \frac{\text{Var}(v_j)}{2\alpha_j}\mathcal{I}_j^2 - \frac{\alpha_j \,\text{Var}(v_j)}{2}\big(\rho_j - \rho_j^*\big)^2. \tag{9}$$

Its maximum is $J_j^* = \text{Cov}(v_j, a_j^*\pi)^2/(2\alpha_j \,\text{Var}(v_j)) = \Lambda_j R_j^2$ when $\mathcal{I}_j/\alpha_j \in [0,1]$, using $\text{Cov}(v_j, a_j^*\pi) = \pi^2 \,\text{Cov}(\hat{a}_j, a_j^*)$ and $\text{Var}(v_j) = \pi^2 \,\text{Var}(\hat{a}_j)$. With perfect attribution, $\hat{a}_j = a_j^*$ gives $R_j^2 = 1$ and, for $\alpha_j \geq 1$, the value Λ_j. Subtracting gives $\mathcal{L}_j = \Lambda_j(1 - R_j^2)$. Outside this case the loss is $J_j^{\text{P}}(r^{\text{P}}) - J_j(r_j)$, with $r^{\text{P}} = \text{clip}(1/\alpha_j, [0,1])$, $r_j = \text{clip}(\mathcal{I}_j/\alpha_j, [0,1])$, and J_j^{P} the criterion evaluated on the perfect signal. Under classical measurement error, $\text{Cov}(\hat{a}_j, a_j^*) = \text{Var}(a_j^*)$, so $R_j^2 = \text{Var}(a_j^*)/\,\text{Var}(\hat{a}_j) = \mathcal{I}_j$, and the marginals follow by differentiating $\Lambda_j(1 - \mathcal{I}_j) = \Lambda_j/(1 + \text{SNR}_j)$.

Cost of ignoring noise. A different quantity follows from (9). A platform that pays the perfect-attribution rate $1/\alpha_j$ on the noisy signal loses $J_j(\rho_j^*) - J_j(1/\alpha_j) = \text{Var}(v_j)(1-\mathcal{I}_j)^2/(2\alpha_j)$ relative to the welfare-optimal royalty. This coefficient-misspecification cost compares two rates on the same noisy signal, so it is not the loss from noisy attribution itself and can exceed Λ_j.

In the Gaussian special case, substituting $\mathcal{I}_j = \text{SNR}_j\,/(1 + \text{SNR}_j)$ gives the SNR forms stated in Appendix A. □

B.3 Endogenous quality choice

The static framework of Section 3 takes the creator's contribution a_j^* as exogenous. We now allow the creator to choose effort q_j that scales their contribution multiplicatively, so platform value from creator j is $q_j \cdot a_j^*\pi$.

Setup. Adopt the following extension of the main-text assumptions:

(Q1) Multiplicative scaling. Quality enters multiplicatively: platform value from creator j is $q_j \cdot a_j^*\pi$, and the measured score scales in the same way, $q_j\hat{a}_j$. Under multiplicative scaling, $\mathcal{I}_j$ is invariant to q_j (the q_j^2 cancels in Cov over Var).

(Q2) Participation. Outside option $\bar{U}_j$ does not depend on ρ_j; the IR constraint binds at the optimum.

(Q3) Sequential timing. The platform commits to $\theta_j = (f_j, \rho_j)$, after which the creator chooses effort q_j to maximize

$$U_j(\theta_j, q) = f_j + \rho_j q \mu_j - \frac{A_j}{2}\rho_j^2 q^2 - \frac{\gamma_j}{2}q^2,$$

where $\mu_j := \mathbb{E}[v_j]$, $\sigma^2_{v,j} := \mathrm{Var}(v_j)$, $A_j := \alpha_j \sigma^2_{v,j}$, $D_j(\rho) := A_j\rho^2 + \gamma_j$, and $\gamma_j > 0$ is the effort-cost coefficient. Payment is $f_j + \rho_j q_j \hat{a}_j \pi$, so effort scales the expected royalty base to $q_j \mu_j$, where $\mu_j = \pi \bar{a}_j = \pi \bar{a}^*_j > 0$ by (Q4) and $\bar{a}^*_j > 0$ (Section 3.1).

(Q4) Unbiased attribution. $\mathbb{E}[\hat{a}_j \pi] = \mathbb{E}[a^*_j \pi]$, so the platform's expected gross benefit from effort is $q_j \mu_j$.

Proof of the endogenous-quality-choice result. The creator's objective is strictly concave in q ($\partial^2 U_j / \partial q^2 = -D_j < 0$). The corner $q = 0$ is excluded since $\partial U_j / \partial q\big|_{q=0} = \rho\mu > 0$. The interior FOC gives $q^{\mathrm{int}} = \rho\mu / D$. Differentiating: $dq^*/d\rho = \mu(\gamma - A\rho^2)/D^2$, positive iff $\rho < \sqrt{\gamma/A}$. The peak is obtained by substituting $\rho^{\max} = \sqrt{\gamma/A}$ into q^{int}. Comparative statics in γ_j and α_j follow from D being strictly increasing in both. □

Welfare-maximizing royalty. The hump-shape implies that the welfare-maximizing royalty, which maximizes the effort surplus $W_j(\rho) = \mu_j^2 \rho(2-\rho)/[2D_j(\rho)]$, exceeds the alignment-only optimum $\rho^*_j = \mathcal{I}_j/\alpha_j$ whenever $\alpha_j > (\mathcal{I}_j/2)(1 + \sqrt{1 + 4\beta_j})$ with $\beta_j := A_j/\gamma_j$ (Proposition 6).

Proposition 5 (Effort-only optimal royalty)**.** *The unique maximizer of W_j on $(0,1)$ is*

$$\rho^{**}_j = \frac{-\gamma_j + \sqrt{\gamma_j^2 + 4A_j\gamma_j}}{2A_j}, \qquad 0 < \rho^{**}_j < 1.$$

Proof. $W_j'(\rho) = \mu_j^2 h(\rho)/D_j(\rho)^2$ where $h(\rho) = -A_j\rho^2 - \gamma_j\rho + \gamma_j$. Setting $h = 0$ and applying the quadratic formula yields the positive root. The intermediate-value theorem confirms the root lies in $(0,1)$ since $h(0) = \gamma_j > 0$ and $h(1) = -A_j < 0$. □

Proposition 6 (Royalty hierarchy)**.** *Under (Q1) through (Q4), write $\beta_j := A_j/\gamma_j$, so that $\rho^{**}_j = 2/(1 + \sqrt{1+4\beta_j})$. If $\mathcal{I}_j \geq 0$ and $\rho^*_j < \rho^{**}_j$, equivalently $\alpha_j > (\mathcal{I}_j/2)(1 + \sqrt{1+4\beta_j})$, then $W_j'(\rho^*_j) > 0$, so raising the royalty slightly above ρ^*_j increases effort surplus at no first-order cost in the alignment criterion. When $0 \leq \mathcal{I}_j \leq 1$ and $\beta_j < 1$, $\alpha_j \geq (1+\sqrt{5})/2$ is sufficient. As $\gamma_j \to \infty$, $\rho^{**}_j \to 1$ and $W_j \to 0$ uniformly, so the effort channel vanishes.*

Proof. Since W_j is strictly increasing on $(0, \rho^{**}_j)$ (from $W_j' > 0$ when $h > 0$) and $\rho^*_j < \rho^{**}_j$ by hypothesis, $W_j'(\rho^*_j) > 0$, so raising the royalty above ρ^*_j improves effort welfare. Because $\mathcal{I}_j \geq 0$ and $\rho^*_j < \rho^{**}_j < 1$, the rate $\rho^*_j = \mathcal{I}_j/\alpha_j$ is unclipped, so the derivative of the alignment criterion vanishes there and there is no first-order alignment cost. The equivalence follows from $\rho^{**}_j = 2/(1+\sqrt{1+4\beta_j})$, obtained by rationalizing the root in Proposition 5, and $\rho^{**}_j < 1$ rules out the upper clip. Since $W_j(\rho) \leq \mu_j^2/(2\gamma_j)$, $W_j \to 0$ uniformly as $\gamma_j \to \infty$. □

The effort and alignment objectives are optimized separately here. At pilot parameters ($\alpha = 2$, $\beta \approx 0.02$), $\rho^{**}_j \approx 0.98$ for all three creators, 2.6 to 9.4× the alignment-only $\rho^*_j \leq 0.378$.

C Incentive compatibility: full analysis

This appendix develops the full incentive-compatibility machinery underlying the score-smoothing analysis (Proposition 4). We give the proof of the three-tier hierarchy, characterize when monitoring frictions deter Tier 2 (mean inflation) bias, work out the screening problem when risk aversion α_j is private information, derive a complementary statistical detection test for Tier 3 (variance shrinkage), and show that the IC hierarchy is invariant to the competitive environment under exclusive licensing. Throughout we use the notation of Section 3.1: $v_j = \hat{a}_j \pi$, $\sigma^2_{v,j} = \mathrm{Var}(v_j)$, $\rho^*_j = \mathcal{I}_j/\alpha_j$, and the creator's mean-variance utility (2). The single-platform variance $\sigma^2_{v,j}$ is the $K = 1$ specialization of the per-platform variance $V_j^{(k)} = \mathrm{Var}(v_j^{(k)})$ used in Appendix E. This appendix lets period revenue be random. The score-smoothing result holds for any fixed positive royalty; the subsequent hierarchy evaluates the rate of Theorem 1.

C.1 Full statement and proof of Proposition 4

Score smoothing is profitable even without changing the expected payment. Fix a contract (f_j, ρ_j) with $\rho_j > 0$ and a creator with $\alpha_j > 0$. Assume $\hat{a}_j \perp \pi$, finite second moments, $\mathrm{Var}(\hat{a}_j) > 0$, and $\mathbb{E}[\pi^2] > 0$. Known positive revenue is a special case. The creator can costlessly replace the reported score by $g_\epsilon(\hat{a}_j) = (1-\epsilon)\hat{a}_j + \epsilon\bar{a}_j$, where $\bar{a}_j = \mathbb{E}[\hat{a}_j]$ and $0 < \epsilon \leq 1$. Holding the contract fixed, the utility gain is

$$\Delta U_j^{\mathrm{shrink}}(\epsilon) = \frac{\alpha_j}{2}\rho_j^2\epsilon(2-\epsilon)\,\mathrm{Var}(\hat{a}_j)\mathbb{E}[\pi^2] > 0. \tag{10}$$

Thus, no fixed fee and positive royalty rate can deter this form of score manipulation.

Proof. Independence implies $\mathbb{E}[g_\epsilon(\hat{a}_j)\pi] = \bar{a}_j\mathbb{E}[\pi] = \mathbb{E}[\hat{a}_j\pi]$, so expected payment is unchanged. By Lemma 2, the variance of the reported attribution-based revenue is

$$\mathrm{Var}(g_\epsilon(\hat{a}_j)\pi) = (1-\epsilon)^2\,\mathrm{Var}(\hat{a}_j)\mathbb{E}[\pi^2] + \bar{a}_j^2\,\mathrm{Var}(\pi).$$

Subtracting this from $\mathrm{Var}(\hat{a}_j\pi)$ gives the variance reduction $\epsilon(2-\epsilon)\,\mathrm{Var}(\hat{a}_j)\mathbb{E}[\pi^2]$. Mean-variance utility multiplies this reduction by $\alpha_j\rho_j^2/2$, yielding (10). Every factor is positive under the stated conditions, and the fixed fee cancels. □

C.2 Proof of the three-tier hierarchy

We give the formal version of the three-tier hierarchy whose Tier 3 conclusion appears as Proposition 4 in the main text.

Proposition 7 (Three-tier creator manipulation hierarchy)**.** *Fix $\rho_j^* = \mathcal{I}_j/\alpha_j$ from Theorem 1, and write $\bar{\pi} := \mathbb{E}[\pi]$, $\sigma_\pi^2 := \mathrm{Var}(\pi)$, and $\bar{a} := \mathbb{E}[\hat{a}_j]$. Consider three post-contractual distortions of the reported $\hat{a}_j$:*

(T1) Noise inflation. *The creator adds independent noise $\eta \perp (\hat{a}, a^*, \pi)$ with $\mathbb{E}[\eta] = 0$, $\mathrm{Var}(\eta) = \Sigma_\eta > 0$ at private cost $c \geq 0$. The creator's utility change is*

$$\Delta U_j^{\mathrm{noise}} = -\frac{\alpha_j}{2}(\rho_j^*)^2\Sigma_\eta\mathbb{E}[\pi^2] - c < 0.$$

Tier 1 is automatically deterred for every $\alpha_j > 0$ and $\rho_j^* > 0$.

(T2) Mean inflation. *The creator adds bias $b \geq 0$ at quadratic cost $(\kappa/2)b^2$, $\kappa > 0$. The interior-bias maximizer is*

$$b_j^* = \max\Big(0,\ \frac{\rho_j^*\bar{\pi} - \alpha_j(\rho_j^*)^2\,\mathrm{Cov}(v_j, \pi)}{\kappa + \alpha_j(\rho_j^*)^2\sigma_\pi^2}\Big).$$

Under $\pi \equiv 1$ this simplifies to $b_j^* = \mathcal{I}_j/(\alpha_j\kappa) > 0$. Tier 2 is not self-policing: external monitoring (κ large enough) is required to enforce truthfulness.

(T3) Variance shrinkage. *Assume $\hat{a}_j \perp \pi$ or $\pi \equiv 1$. For any mean-preserving shrinkage $g_\epsilon(\hat{a}) = (1-\epsilon)\hat{a} + \epsilon\bar{a}$ with $\epsilon \in (0, 1]$,*

$$\Delta U_j^{\mathrm{shrink}}(\epsilon) = \frac{\alpha_j}{2}(\rho_j^*)^2\,\epsilon(2-\epsilon)\,\mathrm{Var}(\hat{a}_j)\,\mathbb{E}[\pi^2] > 0.$$

Tier 3 is impossible to deter under any linear contract with $\rho_j > 0$.

Proof. **(T1).** Let $\tilde{v} = v + \eta\pi$. By independence and $\mathbb{E}[\eta] = 0$, $\mathbb{E}[\eta\pi] = \mathbb{E}[\eta]\mathbb{E}[\pi] = 0$, so the mean is preserved. By the tower property, $\mathrm{Cov}(v, \eta\pi) = \mathbb{E}[\hat{a}\pi^2\mathbb{E}[\eta \mid \hat{a}, a^*, \pi]] = 0$. Independence gives $\mathrm{Var}(\eta\pi) = \mathbb{E}[\eta^2]\mathbb{E}[\pi^2] = \Sigma_\eta\mathbb{E}[\pi^2]$. Hence $\mathrm{Var}(\tilde{v}) = \mathrm{Var}(v) + \Sigma_\eta\mathbb{E}[\pi^2]$, and the creator's utility change is $-(\alpha_j/2)(\rho_j^*)^2\Sigma_\eta\mathbb{E}[\pi^2] - c$, which is strictly negative.

(T2). Adding b replaces v with $v + b\pi$. The creator solves $\max_{b\geq 0}\rho_j^*\bar{\pi}b - (\alpha_j/2)(\rho_j^*)^2(b^2\sigma_\pi^2 + 2b\,\mathrm{Cov}(v, \pi)) - (\kappa/2)b^2$. The objective is strictly concave (SOC: $-\kappa - \alpha_j(\rho_j^*)^2\sigma_\pi^2 < 0$); the interior FOC and the $b \geq 0$ constraint give the stated b_j^*.

(T3). Under $\hat{a}_j \perp \pi$, mean preservation: $\mathbb{E}[g_\epsilon(\hat{a})\pi] = \bar{a}\bar{\pi} = \mathbb{E}[\hat{a}\pi]$. Apply the product-of-independents variance identity (Lemma 2): $\mathrm{Var}(g_\epsilon(\hat{a})\pi) = (1-\epsilon)^2\,\mathrm{Var}(\hat{a})\mathbb{E}[\pi^2] + \bar{a}^2\,\mathrm{Var}(\pi)$ and

$\mathrm{Var}(\hat{a}\pi) = \mathrm{Var}(\hat{a})\mathbb{E}[\pi^2] + \bar{a}^2\,\mathrm{Var}(\pi)$. The variance reduction is $\epsilon(2-\epsilon)\,\mathrm{Var}(\hat{a})\mathbb{E}[\pi^2]$; the $\bar{a}^2\,\mathrm{Var}(\pi)$ term cancels exactly. Mean payment is unchanged, so the utility change is $(\alpha_j/2)(\rho_j^*)^2\,\epsilon(2-\epsilon)\,\mathrm{Var}(\hat{a})\mathbb{E}[\pi^2] > 0$. □

Lemma 2 (Product of independents)**.** *For $X \perp Y$ with finite second moments,* $\mathrm{Var}(XY) = \mathrm{Var}(X)\mathbb{E}[Y^2] + \mathbb{E}[X]^2\,\mathrm{Var}(Y)$.

Proof. $\mathrm{Var}(XY) = \mathbb{E}[X^2Y^2] - \mathbb{E}[XY]^2 = \mathbb{E}[X^2]\mathbb{E}[Y^2] - \mathbb{E}[X]^2\mathbb{E}[Y]^2 = (\mathrm{Var}(X) + \mathbb{E}[X]^2)\mathbb{E}[Y^2] - \mathbb{E}[X]^2\mathbb{E}[Y]^2 = \mathrm{Var}(X)\mathbb{E}[Y^2] + \mathbb{E}[X]^2\,\mathrm{Var}(Y)$. □

Tightness. Each hypothesis binds. Tier 1: at $\alpha_j = 0$ the variance penalty vanishes and Tier 1 fails weakly when $c = 0$. Tier 2: $\kappa = 0 \Rightarrow b^* \to \infty$ when $\sigma_\pi^2 = 0$. Tier 3: $\mathrm{Var}(\hat{a}_j) = 0$ removes any variance to reduce, and $\rho_j^* = 0$ disconnects payment from $\hat{a}$. The condition $\hat{a}_j \perp \pi$ (or $\pi \equiv 1$) is used in Tier 3; the general-π analysis (Remark 1) shows that the SNR ratio carries over to random revenue when the paid attribution is centered.

C.3 SNR MANIPULATION-WELFARE RATIO

Corollary 1 (SNR manipulation-welfare ratio)**.** *Under $\pi \equiv 1$ and classical measurement error, with $\alpha_j \geq 1$ and $0 < \mathcal{I}_j < 1$ as in Proposition 1, the ratio of the maximum Tier 3 manipulation gain (at $\epsilon = 1$) to the welfare loss from Proposition 1 is*

$$\frac{\Delta U_j^{\max}}{\mathcal{L}_j} = \frac{\mathcal{I}_j}{1-\mathcal{I}_j} = \mathrm{SNR}_j\,.$$

Proof. From Proposition 7, $\Delta U_j^{\max} = (\alpha_j/2)(\rho_j^*)^2\,\mathrm{Var}(\hat{a}_j) = \mathcal{I}_j^2\,\mathrm{Var}(\hat{a}_j)/(2\alpha_j)$ at $\mathbb{E}[\pi^2] = 1$. Under classical measurement error $\mathrm{Var}(\hat{a}_j) = \mathrm{Var}(a_j^*)/\mathcal{I}_j$, so $\Delta U_j^{\max} = \Lambda_j\mathcal{I}_j$ with $\Lambda_j = \mathrm{Var}(a_j^*)/(2\alpha_j)$, while Proposition 1 gives $\mathcal{L}_j = \Lambda_j(1-\mathcal{I}_j)$. The ratio is $\mathcal{I}_j/(1-\mathcal{I}_j) = \mathrm{SNR}_j$. □

Remark 1 (General-π case)**.** *When π is random with $\pi \perp (\hat{a}_j, a_j^*)$ and the royalty is paid on centered attribution times revenue (Appendix A), the Tier 3 gain and the loss both scale by $\mathbb{E}[\pi^2]$, so the ratio remains SNR_j. With uncentered attribution and random revenue, both payment moments acquire terms in $\mathrm{Var}(\pi)$ (Lemma 2), and the ratio must be recomputed.*

C.4 STATISTICAL DETECTION OF TIER-3 MANIPULATION

The following test compares reported variance with its calibration value.

Remark 2 (Variance-ratio test)**.** *Under a Gaussian model, a variance-ratio test on the empirical distribution of reported $\hat{a}$ over T periods has statistic*

$$T_{\mathrm{stat}} = (T-1)\widehat{\mathrm{Var}}(\hat{a}^{\mathrm{rep}})/\,\mathrm{Var}(\hat{a}^{\mathrm{cal}}),$$

distributed χ^2_{T-1} under truthful reporting, where $\mathrm{Var}(\hat{a}^{\mathrm{cal}})$ is the calibration-time variance. For a one-sided (lower-tail) level-0.05 test, power ≥ 0.95 at $T \geq 50$ for shrinkage $\epsilon \geq 0.3$.

C.5 MUSSA-ROSEN SCREENING WITH PRIVATE RISK AVERSION

Suppose risk aversion α_j is private information, drawn from a distribution G on $[\underline{\alpha}, \bar{\alpha}]$ with density g satisfying the regularity condition that $\varphi(\alpha) := \alpha + G(\alpha)/g(\alpha)$ is strictly increasing.

Proposition 8 (Optimal contract menu under private types)**.** *Suppose attribution is platform-computed, so creators cannot bias or shrink scores after contracting. Assume $\mathcal{I}_j \geq 0$, $\mathrm{Var}(v_j) > 0$, a type-independent outside option $\bar{U}_j$, and nonnegative fees for the menu defined below. Under this regularity condition, the IC menu that maximizes the platform's alignment-adjusted payoff, as in (C0) of Appendix E, offers*

$$\rho_j^{\mathrm{SB}}(\alpha) = \min\Big(1,\ \frac{\mathcal{I}_j}{\varphi(\alpha)}\Big),$$

implementable in dominant strategies without ironing. Per-type deadweight loss from screening (for types with $\rho_j^* < 1$ and $\rho_j^{\mathrm{SB}} < 1$) is

$$\mathrm{DWL}_j(\alpha) = \frac{\alpha \operatorname{Var}(v_j)}{2}\big(\rho_j^* - \rho_j^{\mathrm{SB}}\big)^2 = \frac{\mathcal{I}_j^2 \operatorname{Var}(v_j)\,(\varphi(\alpha)-\alpha)^2}{2\alpha\,\varphi(\alpha)^2}.$$

Proof. Normalize the highest type's utility to $\bar{U}_j$ and define the menu transfers by

$$U(\alpha) = \bar{U}_j + \frac{\operatorname{Var}(v_j)}{2}\int_\alpha^{\bar{\alpha}} \rho_j^{\mathrm{SB}}(t)^2\,dt, \qquad f(\alpha) = U(\alpha) - \rho_j^{\mathrm{SB}}(\alpha)\mathbb{E}[v_j] + \frac{\alpha}{2}\rho_j^{\mathrm{SB}}(\alpha)^2 \operatorname{Var}(v_j).$$

The hypothesis requires $f(\alpha) \geq 0$ for every type; if it fails, the constrained menu must be solved separately. Single-crossing: $\partial^2 U/\partial\rho\partial\alpha = -\rho \operatorname{Var}(v_j) < 0$ for $\rho > 0$. Envelope: $(U^*)'(\alpha) = -\rho(\alpha)^2 \operatorname{Var}(v_j)/2 \leq 0$. Substituting and integrating by parts gives the pointwise problem $\max_\rho[\rho\mathcal{I}_j \operatorname{Var}(v_j) - (\varphi(\alpha)/2)\rho^2 \operatorname{Var}(v_j)]$, with FOC $\rho^{\mathrm{SB}} = \mathcal{I}_j/\varphi(\alpha)$. Regularity gives $\varphi' > 0$, so monotonicity holds without ironing. The cap $\min(1,\cdot)$ enforces $\rho \leq 1$, binding for $\varphi(\alpha) < \mathcal{I}_j$. The DWL formula follows by substituting $\rho^* = \mathcal{I}_j/\alpha$ and $\rho^{\mathrm{SB}} = \mathcal{I}_j/\varphi(\alpha)$. □

Screening and manipulation interact. If creators could still manipulate after choosing from the menu, the rate $\rho(\hat{\alpha})$ they select would also set the return to manipulation, because bias adds a term proportional to $\rho(\hat{\alpha})$ and shrinkage removes risk in proportion to $\rho(\hat{\alpha})^2$. The type report and the manipulation choice would then interact, and Tier 3 shrinkage would remove the very risk that makes types separate. Proposition 8 therefore assumes platform-computed attribution, which is also the institutional remedy of Section 3.3.

C.6 Dynamic self-policing and competition invariance

Suppose detected Tier 2 bias ends the creator's active status, and let $H_j \geq 0$ be the creator's own continuation gain from staying active. Under binding participation the creator receives $\bar{U}_j$ whether active or not, so $H_j = 0$ unless the platform pays a retention rent, which happens in the thick-margin regime of Proposition 11, where H_j is the discounted value of those rents. Write $\mathcal{A}_j := \rho_j^*\bar{\pi} - \alpha_j(\rho_j^*)^2 \operatorname{Cov}(v_j,\pi)$ and $\mathcal{D}_j := \kappa + \alpha_j(\rho_j^*)^2\sigma_\pi^2$, so that the Tier 2 bias of Proposition 7 maximizes the static gain $\mathcal{A}_j b - \mathcal{D}_j b^2/2$. Under constant detection $\Pr(\text{detect} \mid b > 0) = \mathrm{Pr}_0$, truthful reporting is optimal if and only if $\delta\,\mathrm{Pr}_0\,H_j \geq (\max\{\mathcal{A}_j, 0\})^2/(2\mathcal{D}_j)$, and otherwise the creator chooses the static bias. Under proportional detection $\Pr(\text{detect} \mid b) = \min\{1, \phi_d b\}$, the best response among biases $b \leq 1/\phi_d$ is $b^*_{j,\mathrm{dyn}} = \max\{0, (\mathcal{A}_j - \delta\phi_d H_j)/\mathcal{D}_j\}$ whenever this lies below $1/\phi_d$. A larger bias is detected for sure and yields at most $(\max\{\mathcal{A}_j,0\})^2/(2\mathcal{D}_j) - \delta H_j$, so truthful reporting is optimal whenever $\delta\phi_d H_j \geq \mathcal{A}_j$ and $\delta H_j \geq (\max\{\mathcal{A}_j,0\})^2/(2\mathcal{D}_j)$. Equality can leave a nonzero bias equally attractive; strict inequalities ensure strict deterrence. In the thin-margin regime $H_j = 0$, detection alone cannot deter bias when $\mathcal{A}_j > 0$.

Proposition 9 (Manipulation incentives are unchanged by competition)**.** *Under exclusive licensing and Bertrand competition with $K \geq 2$ platforms, the three-tier IC hierarchy of Proposition 7 is invariant to the number of competitors K and to $\bar{U}_j$.*

Proof. By Theorem 2 (multi-platform Bertrand equilibrium), $\rho_j^{*(k)} = \mathcal{I}_j^{(k)}/\alpha_j$ is independent of K and $\bar{U}_j$ because the effective outside option $\hat{U}_j$ enters platform k's objective additively through $f_j^{(k)}$ and cancels in the ρ-derivative. The Tier 1, Tier 2, and Tier 3 manipulation utilities depend only on $(\rho_j^{*(k)}, \alpha_j, \operatorname{Var}(\hat{a}_j), \bar{\pi}, \mathbb{E}[\pi^2], \kappa, \sigma_\pi^2, \operatorname{Cov}(v,\pi))$, none of which depend on K or $\bar{U}_j$ under exclusive licensing. □

Under non-exclusive licensing the variance of combined per-period payment includes a cross-platform term $2\rho^A\rho^B \operatorname{Cov}(v_j^A, v_j^B)$, which depends on ρ^B and through it on the rival's market structure; IC conditions then become K-dependent.

C.7 FAILURE MODES

Two boundary cases warrant comment. *Risk-neutral limit:* the cap $\rho_j \leq 1$ binds at $\alpha_j = \mathcal{I}_j$; for $\alpha_j < \mathcal{I}_j$, $\rho_j^* = 1$. At $\alpha_j = 0$ the Tier 1 variance penalty vanishes, and Tier 1 IC fails weakly when $c = 0$. *Pairwise additive pooling:* if creator i pools their attribution report with creator j's, reporting $\hat{a}_i + \hat{a}_j$, creator i's payoff variance changes by $(\rho_i^*)^2 \mathbb{E}[\pi^2][\mathrm{Var}(\hat{a}_j) + 2\,\mathrm{Cov}(\hat{a}_i, \hat{a}_j)]$, so additive pooling is unprofitable iff $\mathrm{Cov}(\hat{a}_i, \hat{a}_j) > -\mathrm{Var}(\hat{a}_j)/2$. In music markets, where stylistic similarity drives positive cross-catalog correlation in measured attribution, this condition is typically satisfied and pooling is deterred.

D WELFARE COST OF UNIFORM CONTRACTS

D.1 FULL STATEMENT AND PROOF OF PROPOSITION 3

Suppose the platform offers a uniform contract $(\bar{f}, \bar{\rho})$ to every creator. Let $\mathcal{S}^{\mathrm{exit}} = \{j : U_j(\bar{f}, \bar{\rho}) < \bar{U}_j\}$ be the exit set. Holding $\mathbb{E}[v_j]$, α_j, and $\mathrm{Var}(v_j)$ fixed, creators with higher outside options exit first. If distinctive creators also have high complementary value, their exit can be particularly costly to the platform. Write $T = x \setminus \mathcal{S}^{\mathrm{exit}}$ for the remaining catalog. When u is supermodular in catalog participation,

$$u(x, r) - u(T, r) \geq \sum_{j \in \mathcal{S}^{\mathrm{exit}}} \big[u(T \cup \{j\}, r) - u(T, r)\big]. \tag{11}$$

The inequality is strict if at least two creators exit and their complementarities are strict. This compares lost gross value with individual contributions to the remaining catalog, not with individual removal losses from the original catalog.

Proof of Proposition 3. We prove each claim in turn.

Claim 1 (Selective exit of high-outside-option creators). Under a uniform contract $(\bar{f}, \bar{\rho})$, creator j participates if and only if $U_j(\bar{f}, \bar{\rho}) \geq \bar{U}_j$. Expanding:

$$\bar{f} + \bar{\rho}\,\mathbb{E}[v_j] - \frac{\alpha_j}{2}\bar{\rho}^2\ \mathrm{Var}(v_j) \geq \bar{U}_j.$$

Rearranging, creator j exits if $\bar{U}_j > \bar{f} + \bar{\rho}\,\mathbb{E}[v_j] - \frac{\alpha_j}{2}\bar{\rho}^2\ \mathrm{Var}(v_j)$. The right-hand side is the utility the uniform contract provides. Holding $\mathbb{E}[v_j]$, α_j, and $\mathrm{Var}(v_j)$ fixed, the right-hand side is common across creators, so creators with higher $\bar{U}_j$ are the first for whom this inequality is violated. Without that qualifier the ordering can fail, since a creator with a lower outside option but higher royalty risk can exit while one with a higher outside option stays. In the music setting, if distinctive creators have both high $\bar{U}_j$ (their original recordings command high streaming revenue and sync fees) and high complementary value in u (their unique styles cannot be replicated by combining others), then $\mathcal{S}^{\mathrm{exit}}$ is disproportionately composed of the most valuable creators.

Moreover, under the alignment criterion (4), the alignment value of a creator who stays under the uniform rate falls by $J_j(\rho_j^*) - J_j(\bar{\rho}) = (\alpha_j\,\mathrm{Var}(v_j)/2)(\bar{\rho} - \rho_j^*)^2$ for interior ρ_j^*, by (9). This loss is symmetric in the direction of the mismatch, so creators whose welfare-optimal royalties lie far above or far below the uniform rate both lose alignment value, in proportion to $\alpha_j\,\mathrm{Var}(v_j)$, even though each still receives at least $\bar{U}_j$.

Claim 2 (Super-additive loss from joint exit). For any set $A \subseteq \mathcal{S}^{\mathrm{exit}}$ with $|A| \geq 2$, write $T := x \setminus A$, so the loss from removing A simultaneously is

$$u(x, r) - u(x \setminus A, r).$$

If u is supermodular over the creators in A (these creators are complements), then

$$u(x, r) - u(T, r) \geq \sum_{j \in A} \left[u(T \cup \{j\}, r) - u(T, r)\right].$$

To see this, add the creators in A to T one at a time; under supermodularity each one's marginal contribution is at least its stand-alone value $u(T \cup \{j\}, r) - u(T, r)$. The inequality is strict when complementarities are strict. The comparison uses stand-alone contributions to the post-exit catalog;

marginal contributions to the full catalog would reverse the inequality for complements, because each complement's marginal value to the full catalog already includes the joint value. Since $\mathcal{S}^{\text{exit}}$ is concentrated among distinctive creators under the conditions of Claim 1, and distinctive creators tend to be complements (a reggae producer and a classical composer together enable cross-genre generation that neither supports alone), their joint exit costs more than the sum of their stand-alone contributions to the remaining catalog.

Claim 2 concerns gross platform value, not net payoff after licensing payments. An aggregate payoff comparison must also account for payments saved on exiting catalogs and changes in the remaining catalogs' contributions. □

E Multi-platform Bertrand competition

This appendix gives the formal results behind the attribution-moat analysis (Section 3.3), including an equilibrium construction, the moat result with both homogeneous and heterogeneous demand, and the welfare comparisons. We use the notation of Section 3.1 extended across $K \geq 2$ platforms: platform k has creator-specific informativeness $\mathcal{I}_j^{(k)} \in (0,1)$ and signal variance $V_j^{(k)} := \mathrm{Var}(v_j^{(k)})$, the multi-platform analogue of $\mathrm{Var}(v_j)$ from §3.1 with $v_j^{(k)} = \hat{a}_j^{(k)}\pi$. Throughout this appendix, $\alpha_j \geq 1$ and $V_j^{(k)} > 0$, so the displayed royalty rates satisfy the cap.

E.1 Setting and surplus decomposition

We assume **(C0)** each platform values payment-contribution alignment one-for-one, so its payoff from creator j is financial profit plus the alignment term $\rho\ \mathrm{Cov}(v_j^{(k)}, a_j^*\pi)$ of criterion (4); **(C1)** exclusive licensing (each creator licenses to at most one platform per period); **(C2)** heterogeneous informativeness $\mathcal{I}_j^{(k)}$ across platforms; **(C3)** additive separability of platform k's total surplus across the creators it acquires; and, where indicated, **(C4)** additive Gaussian noise $\hat{a}_j^{(k)} = a_j^* + \xi_j^{(k)}$ with $\xi_j^{(k)} \perp a_j^*$.

Under (C0), the per-creator alignment-adjusted surplus under royalty ρ is

$$\Sigma_j^{(k)}(\rho) = \mu_j^{(k)} + \rho\mathcal{I}_j^{(k)}V_j^{(k)} - \frac{\alpha_j}{2}\rho^2 V_j^{(k)},$$

where $\mu_j^{(k)}$ is the base value from including creator j's catalog. The maximizer is $\rho_j^{*(k)} = \mathcal{I}_j^{(k)}/\alpha_j$ with maximized surplus

$$\Sigma_j^{*(k)} = \mu_j^{(k)} + \frac{(\mathcal{I}_j^{(k)})^2 V_j^{(k)}}{2\alpha_j}.$$

Under (C4), the V-cancellation identity $V_j^{(k)} = \sigma_{a^*}^2\pi^2/\mathcal{I}_j^{(k)}$ gives the simpler form

$$\Sigma_j^{*(k)} - \mu_j^{(k)} = \frac{\mathcal{I}_j^{(k)}\sigma_{a^*}^2\pi^2}{2\alpha_j},$$

linear in $\mathcal{I}_j^{(k)}$.

Remark 3 (Financial profit alone)**.** *Without (C0), Lemma 1 gives a pure fixed-fee optimum at any feasible promised creator utility, and the construction below uses $\rho = 0$ and $\Sigma_j^{*(k)} = \mu_j^{(k)}$. Under homogeneous demand, participating platforms then tie and earn no rent in this equilibrium, so attribution precision affects competition only through (C0). If platforms weight the alignment term by $\lambda > 0$ instead, then $\rho_j^{*(k)} = \mathrm{clip}(\lambda\mathcal{I}_j^{(k)}/\alpha_j, [0,1])$ and, while this rate is interior, the precision terms below scale by λ^2.*

E.2 Full statement and proof of Theorem 2

We restate the equilibrium result with the fee condition written explicitly. Under (C0) through (C4), assume $\bar{U}_j \geq 0$ and that each platform can implement its welfare-optimal royalty at the outside option

with a nonnegative fee

$$f_j^{(k)}(\bar{U}_j) := \bar{U}_j - \rho_j^{*(k)}\mathbb{E}[v_j^{(k)}] + \frac{\alpha_j}{2}(\rho_j^{*(k)})^2 V_j^{(k)} \geq 0.$$

Platforms may abstain; ties favor higher $\Sigma_j^{*(k)}$, with a fixed priority for equal values, and the creator accepts an offer equal to the outside option. There exists a simultaneous Bertrand-Nash equilibrium with the following structure. **(i) Royalty.** Every bidding platform offers $\rho_j^{*(k)} = \mathcal{I}_j^{(k)}/\alpha_j$, independent of K and rivals' offers. **(ii) Allocation.** Creator j joins $k_j^* = \arg\max_k \Sigma_j^{*(k)}$ if $\Sigma_j^{*(k_j^*)} \geq \bar{U}_j$, and otherwise exits. **(iii) Rents.** When licensing occurs, the winning platform earns $\Pi_j = \Sigma_j^{*(k_j^*)} - \hat{U}_j$, where $\hat{U}_j = \max(\bar{U}_j, \Sigma_j^{*(2)})$ and $\Sigma_j^{*(2)}$ is the second-highest maximized surplus; the creator earns $U_j^* = \hat{U}_j$. If the creator exits, all platforms earn zero and the creator receives $\bar{U}_j$.

Proof of Theorem 2. For a promised utility $U \geq \bar{U}_j$, the fee of the welfare-optimal contract is $f_j^{(k)}(U) = f_j^{(k)}(\bar{U}_j) + U - \bar{U}_j \geq 0$. An accepted contract with royalty ρ and creator utility U gives platform payoff $\Sigma_j^{(k)}(\rho) - U \leq \Sigma_j^{*(k)} - U$. Strict concavity of $\Sigma_j^{(k)}$ gives the stated welfare-optimal royalty, which attains this bound.

Rank platforms by $\Sigma_j^{*(k)}$, breaking equal values by the stated priority. If $\Sigma_j^{*(1)} \geq \bar{U}_j$, let the leader offer $\hat{U}_j$, each other platform with $\Sigma_j^{*(k)} \geq \bar{U}_j$ offer its own $\Sigma_j^{*(k)}$, and the remaining platforms abstain. All offers use the welfare-optimal royalty and are feasible by the fee condition. The leader wins by the tie rule and cannot lower its utility offer without losing the creator, while raising it or changing its rate cannot increase payoff. A losing platform must offer more than $\hat{U}_j \geq \Sigma_j^{*(k)}$ to win and would therefore earn a negative payoff. If $\Sigma_j^{*(1)} < \bar{U}_j$, all platforms abstain, since any accepted offer would have negative payoff. These arguments establish the stated equilibrium, not uniqueness. □

E.3 Vertical differentiation rent and the moat

Corollary 2 (Rent from being the attribution-precision leader)**.** *In the equilibrium of Theorem 2, with $\bar{U}_j \leq \Sigma_j^{*(2)}$ and homogeneous demand ($\mu_j^{(k)} = \mu_j$ for all k),*

$$\Pi_j = \frac{(\mathcal{I}_j^{(k^*)} - \mathcal{I}_j^{(2)})\sigma_{a^*}^2 \pi^2}{2\alpha_j}.$$

Proof. The Bertrand rent is $\Pi_j = \Sigma_j^{*(k^*)} - \Sigma_j^{*(2)}$. Evaluating both maximized surpluses under V-cancellation gives $\Sigma_j^{*(k)} = \mu_j + \mathcal{I}_j^{(k)}\sigma_{a^*}^2\pi^2/(2\alpha_j)$, and subtraction yields the stated formula. □

E.4 Full statement and proof of Proposition 2

Consider the equilibrium of Theorem 2 under (C0) through (C4) and homogeneous demand, $\mu_j^{(k)} = \mu_j$. Fix a creator and order platforms by initial informativeness, with $\mathcal{I}^{(1)} > \mathcal{I}^{(2)}$. Let $\Delta\mathcal{I}_j = \mathcal{I}^{(1)} - \mathcal{I}^{(2)}$ and $\sigma_{a^*}^2 = \mathrm{Var}(a_j^*) > 0$. Assume $\bar{U}_j \leq \Sigma_j^{*(2)}$ initially. Platform 2 improves its informativeness by $0 \leq \epsilon \leq 1 - \mathcal{I}^{(2)}$, while all other platforms' informativeness remains fixed. The rate and nonnegative-fee conditions of Theorem 2 must hold throughout the improvement.

Platform 2 earns zero rent for $\epsilon \leq \Delta\mathcal{I}_j$. For $\epsilon > \Delta\mathcal{I}_j$, it wins the catalog and earns

$$\Pi_j^{(2)}(\epsilon) = (\epsilon - \Delta\mathcal{I}_j)\pi^2\sigma_{a^*}^2/(2\alpha_j).$$

These rents are measured before any cost of improving attribution. Below the threshold, improvements transfer rent from the leader to the creator without changing total alignment-adjusted surplus. Above the threshold, the increase in total surplus equals Platform 2's rent. At the threshold, the leading platforms tie and each earns zero rent; allocation follows the tie rule.

Proof of Proposition 2. **(i) Dead zone.** Under homogeneous demand and (C4), $\Sigma_j^{*(k)} - \mu_j$ is linear in $\mathcal{I}_j^{(k)}$. For $\epsilon < \Delta\mathcal{I}_j$, Platform 2 remains the runner-up; its surplus improvement raises the creator's effective outside option but generates zero rent for itself. At equality the leading platforms tie, so either allocation gives Platform 2 zero rent. For $\epsilon > \Delta\mathcal{I}_j$, Platform 2 becomes the unique leader and wins the catalog.

(ii) Post-flip rent. Direct application of Corollary 2 with the updated informativeness $\mathcal{I}^{(2)} + \epsilon$.

(iii) Social return. In the dead zone, $\mathcal{W} = U_j^* + \Pi_j = \Sigma_j^{*(1)}$ is constant in ϵ, because the creator's outside-option gain exactly offsets the leader's rent reduction. After the flip, $\mathcal{W} = \Sigma_j^{*(2)}(\epsilon)$, so $\mathcal{W}(\epsilon) - \mathcal{W}(0) = \max\{0, \epsilon - \Delta\mathcal{I}_j\}\,\pi^2\sigma_{a^*}^2/(2\alpha_j) = \Pi_j^{(2)}(\epsilon)$, and the trailing platform's private return equals the social return. The initial condition $\bar{U}_j \le \Sigma_j^{*(2)}$ ensures that the outside option does not replace the runner-up's offer in this comparison. □

Remark 4 (Heterogeneous demand)**.** *With heterogeneous demand $\mu^{(k)}$, the flipping threshold becomes*

$$\epsilon^{\text{flip}} = \Delta\mathcal{I}_j + \frac{2\alpha_j(\mu^{(1)} - \mu^{(2)})}{\sigma_{a^*}^2\pi^2},$$

extending the dead zone when the leader has a demand advantage and shrinking it when the trailer does.

E.5 Welfare comparisons

Proposition 10 (Welfare loss under competition)**.** *Under Theorem 2, summing over participating creators,*

$$\mathcal{L}^{\text{comp}} = \sum_j \mathcal{L}_j^{(k_j^*)}, \qquad \mathcal{L}_j^{(k)} = \frac{\pi^2\sigma_{a^*}^2}{2\alpha_j}(1 - \mathcal{I}_j^{(k)}).$$

Under (C4) and homogeneous demand, the competitive allocation minimizes total welfare loss among allocations satisfying (C1) through (C3) that license the same creators.

Proof. Under Bertrand, each creator joins $k_j^* = \arg\max_k \Sigma_j^{*(k)}$. Under homogeneous demand, $\Sigma_j^{*(k)} - \mu$ is increasing in $\mathcal{I}_j^{(k)}$, while $\mathcal{L}_j^{(k)} = \pi^2\sigma_{a^*}^2(1 - \mathcal{I}_j^{(k)})/(2\alpha_j)$ is strictly decreasing in $\mathcal{I}_j^{(k)}$. Therefore competitive allocation assigns each creator to their max-$\mathcal{I}$ platform, which simultaneously minimizes loss; total loss is the sum, by (C3). □

Corollary 3 (Diminishing returns to entry)**.** *Under (C0), (C4), and homogeneous demand, let $W_K := \mu_j + c_j\,\mathbb{E}[\max_{k\le K}\mathcal{I}_j^{(k)}]$ with $c_j := \pi^2\sigma_{a^*}^2/(2\alpha_j)$ be the expected maximized surplus with K platforms, and $\Delta W_K := W_{K+1} - W_K$. For i.i.d. informativeness draws, ΔW_K is nonincreasing in K, and for $\mathcal{I} \sim \text{Uniform}[0,1]$ it equals $c_j/((K+1)(K+2))$, which is strictly decreasing.*

Proof. Write $M_K := \max_{k\le K}\mathcal{I}_j^{(k)}$. Because the new draw is independent with the common law, $\Delta W_K = c_j\,\mathbb{E}[(X - M_K)^+]$ for an independent copy X, which is nonincreasing in K because M_K is nondecreasing. For the uniform case, $\mathbb{E}[M_K] = K/(K+1)$, so $\Delta W_K = c_j/((K+1)(K+2))$. □

E.6 Royalty invariance under moral hazard

The royalty invariance of Theorem 2 extends to environments with endogenous quality choice (Appendix B.3). Under (Q1) through (Q4) and (C0) through (C3), both the static optimal royalty $\rho_j^{*(k)} = \mathcal{I}_j^{(k)}/\alpha_j$ and the equilibrium effort $q_j^*(\rho_j^{*(k)}) = \rho_j^{*(k)}\mu_j/D_j(\rho_j^{*(k)})$ depend only on platform k's own primitives $(\mathcal{I}_j^{(k)}, V_j^{(k)}, \alpha_j, \mu_j, \gamma_j)$, not on K or $\bar{U}_j$. The effort-channel welfare $W_j(\rho) = \mu_j^2\rho(2-\rho)/(2D_j(\rho))$ inherits the invariance, so the welfare-maximizing royalty ρ_j^{**} is also K-invariant. Contract structure (royalty, effort, per-creator welfare) is competition-invariant if the surplus function decomposes as $\Pi_j^{(k)}(\rho) = \Sigma_j^{(k)}(\rho) - \hat{U}_j$ with $\hat{U}_j$ not depending on ρ.

E.7 TENSION WITH NON-ADDITIVE UTILITY

The competition results assume additive values across catalogs under (C3), whereas the selective-exit analysis permits complementarities. For $u_t = \sum_j s_j + \epsilon \sum_{j<l} c_{jl}$, the competition results apply at $\epsilon = 0$. Equilibrium assignments with nonzero complementarities are not derived here.

F DISCUSSION OF MODELING CHOICES

Mean-variance vs. expected utility. The analysis uses mean-variance preferences; general expected-utility preferences are not analyzed. With Gaussian attribution the royalty term, and hence realized payment, can be negative in a bad period; as in the linear-normal model of Holmström & Milgrom (1987), the analysis imposes no ex post payment floor.

Independence of attribution noise. The per-creator analysis never uses independence of ξ_j across creators. Because contracts are separable and each creator bears risk only on their own payment, such correlation leaves the per-creator rate of Theorem 1 unchanged whenever the marginal second moments of $(\hat{a}_j, a_j^*, \pi)$ are held fixed. It binds instead under pooled budgets or jointly indexed payments, where the full covariance structure enters the contract, which we leave open.

Static vs. dynamic model. The main model studies one licensing decision. Appendix I adds repeated participation, exit, re-entry, and retention rents under a stationary contract. Renegotiation and learning about attribution quality are outside that extension.

Endogenous attribution cost. We treat the attribution technology (and hence $\sigma^2_{\xi,j}$) as a platform choice but do not fully solve for the joint optimization over contracts and attribution investment. The first-order condition for optimal investment equates the marginal cost of reducing $\sigma^2_{\xi,j}$ to the marginal welfare gain $|\partial \mathcal{L}_j / \partial \sigma^2_{\xi,j}|$, summed across all creators. Solving the joint problem requires a specified attribution-cost function C^{attr}.

G ADDITIONAL EMPIRICAL EXPERIMENTS AND THE ATTRIBUTION-INVESTMENT PORTFOLIO

We report the experimental protocol, additional measurements, held-out calibration, and attribution-investment calculations.

G.1 DETAILED EMPIRICAL SETUP

We describe the data, retraining reference, and calibration used in both settings.

Expanded audio experiment. The primary audio results in Table 2 use 50 catalogs, 200 generated queries, and one full-catalog plus 50 leave-catalog-out models. The training manifest contains 1,120 tracks; attribution scores are available for 1,117 examples after exclusions.

We retain the original 22-catalog pilot for the baseline comparison in Table 6 and the held-out comparison in Appendix G.3.

Table 2: Attribution measurements on Stable Audio Open (SAO, 50 catalogs) and the Anticipatory Music Transformer (AMT, 100 catalogs). Correlations are averaged over queries; calibrated SNR is the median across creators. Welfare gap closed is an aggregate across creators.

Model	Method	Spearman	Pearson	Calibrated SNR	Welfare gap closed
SAO	Random	0.011	0.053	0.003	0.53%
SAO	CLAP	0.046	0.148	0.016	4.96%
SAO	Fingerprint	0.029	0.150	0.006	5.80%
SAO	EKFAC	0.007	0.109	0.023	9.36%
SAO	D-TRAK	0.031	0.125	0.009	9.18%
AMT	Random	0.001	0.005	0.001	0.15%
AMT	CLAP	0.004	0.007	0.012	1.42%
AMT	Fingerprint	0.012	0.109	0.006	9.09%
AMT	LoGra	0.068	0.194	0.017	6.48%
AMT	TRAK	0.035	0.087	0.002	1.38%

Retraining reference. For each creator j, the loss-delta reference is

$$a_j^* \;=\; \ell\big(y^{\text{full}} \,\big|\, r; \phi^{-j}\big) \;-\; \ell\big(y^{\text{full}} \,\big|\, r; \phi^{\text{full}}\big)\,, \tag{12}$$

where ϕ^{full} are the model weights of the full-catalog fine-tune, ϕ^{-j} are the weights of the leave-creator-j-out fine-tune, r is the prompt, and y^{full} is the generation produced by ϕ^{full} at r. This follows the counterfactual-loss formulation in the data-attribution literature (Koh & Liang, 2017; Park et al., 2023; Ilyas et al., 2022), applied at the catalog level rather than the example level. Every catalog is evaluated on the same outputs. The reference measures changes in model loss, not listener judgments or revenue. The contract calculations use this reference as the operational measure of contribution.

Catalog selection. The $N = 50$ artist catalogs are drawn from MTG-Jamendo's autotagging split, with at least ten tracks per artist and at least 40% within-artist genre coherence. Catalogs span four prolificness bands by available track count: 18 low (10 to 15 tracks), 18 mid (16 to 25), 11 high (26 to 50), and 3 extra-high (> 50, capped at 50). They span 19 primary genres, including electronic, classical, hip-hop, pop, reggae, rock, and jazz. The total training set is 1,120 tracks; each leave-catalog-out fine-tune drops the corresponding catalog and retrains.

Model and fine-tuning. We fine-tune Stable Audio Open (Evans et al., 2025) on the curated training set with the VAE frozen, training only the diffusion transformer. Each of the $N + 1 = 51$ models (one full + 50 leave-catalog-out) is fine-tuned from the public checkpoint with matched hyperparameters; we use the same prompted-continuation set of $T = 200$ evaluation prompts across all models.

Symbolic data and training. The symbolic experiment selects 100 artists from 203 eligible artists with at least 25 training examples each. The selected subset contains 3,633 examples, with 25 to 85 per artist and a median of 30. Selection uses artist-count quartiles, drawing 36, 36, 23, and 5 artists from successive bands. The full-model training index contains these 3,633 examples; leave-out fits remove one artist's examples. There are 101 recorded fits and 500 evaluation outputs. Table 3 reports training settings for both experiments.

Table 3: Recorded training settings for the full and leave-catalog-out models. Batch size is per device; accumulation gives the number of batches per optimizer step.

Setting	Audio	Symbolic
Training examples	1,120 tracks	3,633 examples
Catalogs / outputs	50 / 200	100 / 500
Full + leave-out fits	$1 + 50$	$1 + 100$
Epochs	50	5
Batch / accumulation	4 / 2	8 / 1
Learning rate	5×10^{-5}	2×10^{-5}
Weight decay	0.001	0.01
Numerical precision	Mixed FP16	FP16
Recorded training seeds	Full 42; leave-out 43 to 92	42 in all fits

Reference computation. The audio loss records use ten uniformly spaced diffusion timesteps from 0 through 900 on a $1{,}000$-step grid, with conditioning dropout disabled. The symbolic scoring script evaluates each generated sequence separately using the language model's token prediction loss, with the model in evaluation mode.

Attribution-method implementation. For both EKFAC influence functions (Mlodozeniec et al., 2025) and D-TRAK (Zheng et al., 2024), we compute per-sample attribution scores against each of the T generated outputs and aggregate to the catalog level by summation. EKFAC uses block-diagonal Fisher approximation; D-TRAK uses random projections of training and query gradients followed by a kernel-based score.

Table 4: Additional attribution settings recorded for audio EKFAC and D-TRAK and symbolic LoGra.

Method	Settings
EKFAC	Corrected eigenvalues; damping 10^{-8}; last 89 of 178 indexed diffusion-model linear modules; four-bit query-gradient compression.
D-TRAK	One full-model checkpoint; $2{,}048$ projected features; regularization 0.1; ten uniformly spaced diffusion timesteps for training and query features.
LoGra	49 projected layers; $4{,}096$ features per layer; $200{,}704$ features in total.

Quantity definitions. For each method m and creator j we report four quantities: $\widehat{\mathcal{I}}_j^{(m)}$ is the regression slope of a^* on $\hat{a}^{(m)}$ across the T outputs; $\widehat{K}_j^{(m)} = \widehat{\mathrm{Cov}}(\hat{a}^{(m)}, a^*)/\widehat{\mathrm{Var}}(a^*)$ is the multiplicative-bias scale; $\widehat{\mathrm{SNR}}_j^{(m,\mathrm{cal})}$ is the Gaussian-shortcut SNR computed from the calibrated estimator $\tilde{a}_{j,t} = \hat{a}_{j,t}/\widehat{K}_j^{(m)}$; and $\widehat{r}_j^{(m)}$ is the per-creator Pearson correlation between $\hat{a}^{(m)}$ and a^* across the T outputs. Bootstrap 95% confidence intervals in the measurement table and per-creator figure are computed by resampling the T outputs with replacement, $2{,}000$ replicates per quantity.

Aggregation and calibration. The Spearman and Pearson columns in Table 2 first compare catalogs within each query, then average over queries. Calibrated SNR instead uses variation across queries for each catalog, followed by a median across catalogs. Welfare gap closed aggregates the model-implied gains across catalogs, weighted by their contribution variances when revenue and risk aversion are common.

Headline calibration uses the same queries to estimate and evaluate $\widehat{K}_j$, including its sign. Section G.3 separately tests fitted calibration on held-out audio queries.

Sanity checks. The ground truth a^* is positive in 94.3% of $10{,}000$ creator-output pairs (mean 0.00212, std 0.00197). The four-quantity tuple satisfies the identity $\widehat{\mathcal{I}}_j = \widehat{\mathrm{SNR}}_j^{\mathrm{cal}} / \big[\widehat{K}_j(1+\widehat{\mathrm{SNR}}_j^{\mathrm{cal}})\big]$ up to numerical precision on the exported data, an algebraic consistency check.

Expanded measurement statistics. Table 5 provides additional measurements for the 50-catalog experiment reported in Table 2.

Table 5: Headline measurement statistics on Stable Audio Open with $N = 50$ artist catalogs and $T = 200$ generated outputs, under EKFAC and D-TRAK. Each correlation row reports the per-clip cross-creator correlation averaged over clips, with 95% bootstrap CIs in brackets. The $|\widehat{K}_j|$ row reports the median across the 50 creators.

Quantity	EKFAC influence	D-TRAK
Mean cross-artist Spearman	0.007 $[-0.012, 0.025]$	0.031 $[0.013, 0.051]$
Mean cross-artist Pearson	0.109 $[0.059, 0.161]$	0.125 $[0.094, 0.158]$
Median $\lvert\widehat{K}_j\rvert$	11 710	10.12
Creators with $\widehat{K}_j < 0$ (negative estimate)	$19/50$	$16/50$
Creators with $\widehat{\mathrm{SNR}}_j^{\mathrm{cal}} > 0.1$	$6/50$	$2/50$
Creators with $\widehat{\mathrm{SNR}}_j^{\mathrm{cal}} > 1$	$0/50$	$1/50$
Welfare gap closed	9.36%	9.18%

Additional baselines in the original audio pilot. We also evaluate Random, CLAP, and Fingerprint on the original 22-catalog, 200-query audio pilot. Table 6 reports these results alongside EKFAC and D-TRAK on the same pilot. These measurements are separate from the 50-catalog audio results in Table 2; the symbolic results in that table use the same 100 catalogs and 500 queries as the original symbolic experiment.

Table 6: Baseline comparison on the original Stable Audio Open pilot, 22 catalogs and 200 queries. Spearman and Pearson correlations are averaged over queries where defined. Calibrated SNR is the median across catalogs, and welfare gap closed is an aggregate across catalogs.

Method	Spearman	Pearson	Calibrated SNR	Welfare gap closed
Random	0.053	0.092	0.002	0.27%
CLAP	0.127	0.289	0.133	16.58%
Fingerprint	0.114	0.290	0.035	10.94%
EKFAC	0.050	0.188	0.092	10.59%
D-TRAK	0.054	0.177	0.021	9.77%

The pilot Fingerprint correlations use 199 queries because one query has identical scores across catalogs; its per-catalog SNR and welfare calculations use all 200 queries. All other correlations in Tables 2 and 6 are defined on every query.

None of the 22 audio-pilot catalogs has calibrated SNR above 1 under Random, CLAP, or Fingerprint. The same holds for all 50 catalogs in the expanded audio experiment. In the symbolic experiment, Fingerprint exceeds 1 for one of 100 catalogs; Random and CLAP do so for none. Here, SNR $= 1$ is a precision reference, not a universal threshold for choosing royalties.

The baseline comparison also shows that no tested method leads on every metric. In the audio pilot, CLAP closes 16.58% of the welfare gap, compared with 10.59% for EKFAC and 9.77% for D-TRAK. In the symbolic experiment, Fingerprint closes 9.09%, compared with 6.48% for LoGra (Table 2).

G.2 Per-creator measurement diagnostics

Figure 4 reports per-creator measurements that aggregate to the headline statistics in Table 5. At $T = 200$ the per-creator bootstrap intervals are wide and span zero for most creators on both methods. A small subset of creators have positive intervals, and the per-creator Pearson estimates are correlated across methods at Spearman $\rho = 0.503$ ($p = 0.0002$), even though the two methods' calibration scales $\widehat{K}_j$ differ by orders of magnitude.

Figure 5 reports two scaling diagnostics. The variance of the loss-delta ground truth is positively associated with catalog size on a log-log scale, with Pearson $r = 0.374$ ($p = 0.0075$). The right panel overlays the closed-form marginal welfare gain $|\partial\mathcal{L}/\partial\,\mathrm{SNR}\,|$ from Proposition 1 (computed at the median $\mathrm{Var}(a^*)$) with per-creator measured points. Per-creator points scatter above and below the median curve in proportion to each creator's own $\mathrm{Var}(a^*)$.

G.3 Held-out calibration and comparison with the pilot

Calibration direction in the expanded audio experiment. EKFAC has 19 negative slope estimates across the 50 creators. Of its nominal 95% bootstrap intervals for the slope $\widehat{\mathcal{I}}_j$, 12 are wholly positive, 2 are wholly negative, and 36 include zero. For D-TRAK, the corresponding counts are 7, 2, and 41. The slope intervals are not adjusted for multiple comparisons.

Held-out prediction gain. For each creator, let x_{jt} be the attribution score and $y_{jt} = a^*_{jt}$ the retraining reference. On each training fold we fit the linear predictor $\hat{y}_{jt} = \bar{y}_{j,\mathrm{train}} + b_{j,\mathrm{train}}(x_{jt} - \bar{x}_{j,\mathrm{train}})$. On the held-out fold we compare its squared error with that of the training-fold mean predictor. The reported aggregate gain is

$$G = \frac{\sum_j \mathrm{mean}_t\left[(y_{jt} - \bar{y}_{j,\mathrm{train}})^2 - (y_{jt} - \hat{y}_{jt})^2\right]}{\sum_j \widehat{\mathrm{Var}}_t(y_{jt})},$$

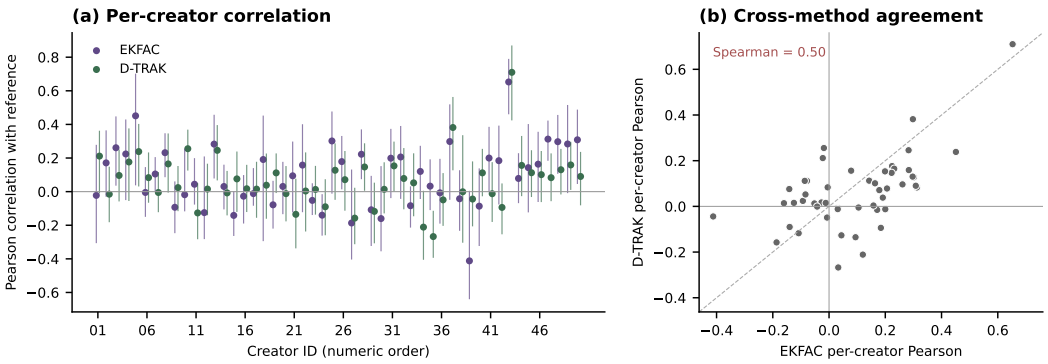

Figure 4: Per-creator detail behind the main-text headline numbers, 50 catalogs. **(a)** Per-creator Pearson with 95% bootstrap confidence intervals, in numeric ID order. Most intervals span zero at $T = 200$. **(b)** Cross-method agreement on per-creator Pearson, with the diagonal in dashed grey. The estimates have rank correlation $\rho = 0.503$ ($p = 0.0002$), even though the two methods' calibration scales $\widehat{K}_j$ differ by three orders of magnitude.

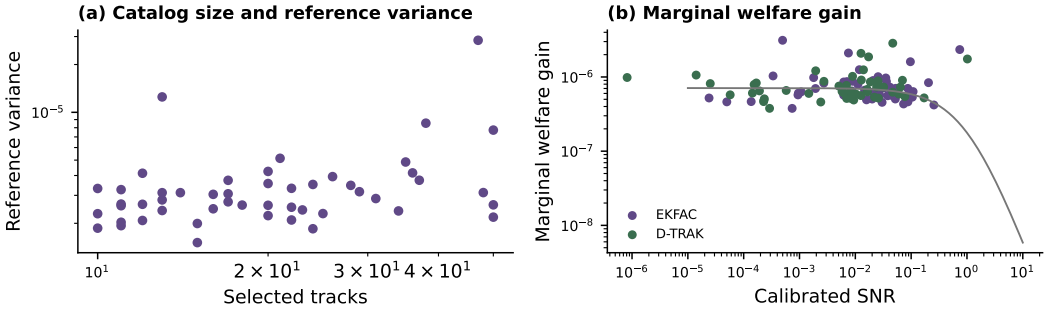

Figure 5: Scaling and marginal-welfare diagnostics for 50 catalogs. **(a)** Ground-truth signal variance versus catalog size, log-log axes (Pearson $r = 0.374$, $p = 0.0075$). **(b)** Closed-form marginal welfare gain from Proposition 1 (grey curve, evaluated at the median $\mathrm{Var}(a^*)$ and $\alpha = 2$) overlaid with per-creator measured points.

where each query is scored using the fold that held it out, and the denominator uses the full observed query set. Negative values indicate worse prediction than the fitted constant baseline, not observed financial losses. A constant attribution score is assigned zero slope.

We average 20 repetitions of 10-fold cross-validation, fitting on 180 and evaluating on 20 of the 200 queries in each fold. The reproduced analysis uses the same split sequence across methods for each creator, with separate partitions across creators.

Random scores are summed within catalogs, so this control preserves catalog-size differences.

Table 7: Supporting held-out calibration results for the 50-catalog audio experiment. Values are normalized prediction gains; under the loss of Proposition 1 the in-sample column equals the welfare gap closed in Figure 2(d), and no value is a realized payment. Held-out estimates average 20 repetitions of 10-fold cross-validation.

Method	In-sample gain (%)	Held-out gain (%)
Random	0.53	−0.63
CLAP	4.96	2.83
Fingerprint	5.80	3.55
EKFAC	9.36	5.08
D-TRAK	9.18	6.92

For the original pilot, EKFAC's in-sample aggregate of 10.59% becomes -2.91% held out; D-TRAK, CLAP, Fingerprint, and Random yield 7.34%, 14.20%, 7.44%, and -0.80%, respectively.

In the expanded setting, 27, 33, 26, 35, and 44 of the 50 creators have negative mean gains for EKFAC, D-TRAK, CLAP, Fingerprint, and Random, respectively. These counts use mean prediction gains across repetitions, without creator-level significance tests.

Query bootstrap. We separately resample the 200 prompt IDs with replacement in $5{,}000$ bootstrap replicates. Each replicate uses the same resampled prompts across creators and methods. All copies of a prompt stay in the same fold, and calibration is refitted on the training folds. We report percentile intervals for aggregate gain and its paired difference from Random.

Scope of uncertainty. The query-bootstrap intervals treat prompts as independent sampling units and condition on the observed catalogs, fitted models, attribution configurations, and Random scores. The audio full and leave-out fits use different recorded training seeds, so their loss differences may include training variation as well as catalog-removal effects. The intervals do not measure that variation, configuration-selection effects, or generalization to other catalogs. Repeated-fold variation is not a population confidence interval; the point estimates in Table 7 are the original-data estimates, not bootstrap averages.

Concentration. For the concentration check, we average 20 additional 10-fold partitions, shared across creators and methods. One catalog contributes 113.4%, 100.8%, 89.8%, and 95.7% of net aggregate gain for EKFAC, D-TRAK, CLAP, and Fingerprint, respectively. Shares above 100% mean the other catalogs have a negative combined gain. The remaining catalogs have a positive combined gain for CLAP and Fingerprint.

The two audio settings share 22 creator identities but only 37 unique prompt strings; their numerical prompt IDs do not identify matched queries. Across those creators, Spearman correlations between the calibrated-SNR profiles are 0.348 for EKFAC, 0.338 for D-TRAK, 0.513 for CLAP, 0.467 for Fingerprint, and -0.048 for Random. These comparisons describe different training pools and query sets, not the isolated causal effect of increasing catalog count.

Held-out gains are concentrated in a few catalogs.

In the 50-catalog audio experiment, all four non-random methods retain positive aggregate gains. However, the bootstrap intervals include zero for both these gains and their differences from Random. The median catalog has a negative gain under every method.

Figure 6 plots these gains.

G.4 HELD-OUT CALIBRATION FIGURE

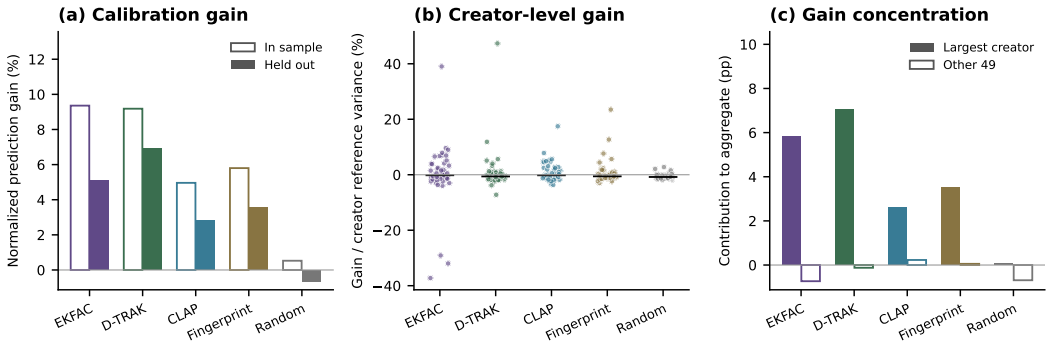

Figure 6: Held-out calibration on the 50-catalog audio setting. **(a)** Prediction gain in sample and held out, $20\times$ 10-fold. **(b)** Per-creator held-out gain, normalized by reference variance. **(c)** Largest single creator against the other 49, in points of the held-out bar. Gains are squared-error reductions, not payments.

G.5 Optimal attribution-investment portfolio

Suppose the platform has a budget $B > 0$ to invest in improving attribution precision and faces a per-creator improvement technology $g_j : [0, \bar{\beta}_j] \to \mathbb{R}_+$ that is strictly increasing, strictly concave, C^1, with $g_j(0) = 0$ and $g_j'(0) < \infty$. Investing β_j units of budget in creator j raises their SNR from S_j to $S_j + g_j(\beta_j)$. Adopt additive-Gaussian noise with known period revenue as in Section 3.1. The platform allocates the budget to maximize total welfare-loss reduction.

Corollary 4 (Optimal attribution-investment portfolio)**.** *The welfare-maximizing allocation* $\{\beta_j^*\}$ *solving*

$$\max_{\beta \geq 0, \sum_j \beta_j = B} \sum_{j=1}^{N} \left[\mathcal{L}_j(S_j) - \mathcal{L}_j(S_j + g_j(\beta_j)) \right]$$

satisfies the following:

1. Existence and uniqueness. β^* *is unique by strict concavity of the composed objective.*
2. Priority index. *Define* $\mathcal{M}_j := \phi_j \cdot g_j'(0)$ *where*

$$\phi_j = \frac{\Lambda_j}{(1 + S_j)^2}, \qquad \Lambda_j = \frac{\pi^2 \operatorname{Var}(a_j^*)}{2\alpha_j}$$

   is the marginal welfare gain from an infinitesimal SNR improvement. Creator j *receives investment iff* $\mathcal{M}_j$ *exceeds the shadow price* λ^* *of the budget constraint.*
3. Closed form under square-root technology. *If* $g_j(\beta) = \gamma_j \sqrt{\beta}$ *(which violates* $g_j'(0) < \infty$*, so every creator receives positive investment), the first-order approximation of the optimal allocation is*

$$\beta_j^{\text{approx}} = \frac{(\phi_j \gamma_j)^2}{\sum_k (\phi_k \gamma_k)^2} \cdot B.$$

   Budget shares follow a square law in marginal welfare gains.
4. Comparative statics. $\partial \beta_j / \partial S_j < 0$*,* $\partial \beta_j / \partial \alpha_j < 0$*,* $\partial \beta_j / \partial \operatorname{Var}(a_j^*) > 0$ *for interior allocations with at least two creators.*

Proof. (i) $\mathcal{L}(S) = \Lambda/(1+S)$ is strictly convex, with $\mathcal{L}'' = 2\Lambda/(1+S)^3 > 0$. The gain $G_j(s) := \mathcal{L}(S_j) - \mathcal{L}(S_j + s)$ is therefore strictly concave in s. Composition of strictly increasing concave G_j with strictly concave g_j is strictly concave, so $h_j(\beta_j) := G_j(g_j(\beta_j))$ is strictly concave on $[0, \bar{\beta}_j]$. The sum $\sum_j h_j(\beta_j)$ is strictly concave on the compact convex constraint set; existence and uniqueness follow from Weierstrass plus strict concavity.

(ii) KKT: $h_j'(\beta_j^*) = \lambda^*$ if $\beta_j^* > 0$, else $h_j'(0) \leq \lambda^*$. With $h_j'(0) = G_j'(0) g_j'(0) = \phi_j \cdot g_j'(0) = \mathcal{M}_j$, strict concavity gives the if-and-only-if characterization.

(iii) Linearize $G_j(s) \approx \phi_j s$ for small s. Then $h_j(\beta_j) \approx \phi_j \gamma_j \sqrt{\beta_j}$. The FOC $\phi_j \gamma_j / (2\sqrt{\beta_j}) = \lambda$ yields $\sqrt{\beta_j} = \phi_j \gamma_j / (2\lambda)$; squaring and enforcing the budget constraint gives the stated formula.

(iv) Direct differentiation of $\beta_j \propto \phi_j^2$. □

Pilot calibration. At the stylized calibration used for the theoretical examples, whose three creators are illustrative inputs rather than members of the measured pool, ($\alpha_j = 2$, $\pi = 1$, three creators with $\mathcal{I} \in \{0.617, 0.755, 0.209\}$, $\operatorname{Var}(a^*) \in \{0.013, 0.017, 0.006\}$, and $\gamma_j = 1$), the budget shares are 19.4% (pop), 5.5% (reggae), and 75.1% (classical).

Approximation guarantee. Let $\epsilon_a := \max_j g_j(\beta_j^{\text{approx}})/(1 + S_j)$. Under square-root technology, β^{approx} maximizes $\sum_j \phi_j g_j(\beta_j)$, while $G_j(s) \leq \phi_j s$ by concavity and $G_j(s) \geq \phi_j s/(1+\epsilon_a)$ for $s \leq g_j(\beta_j^{\text{approx}})$, so the closed-form allocation captures at least fraction $1/(1+\epsilon_a)$ of the optimal gain. In the pilot calibration with $B = 0.01$, the classical creator receives $\beta = 0.00751$, $\epsilon_a = 0.069$, and the guarantee is at least 93.5%.

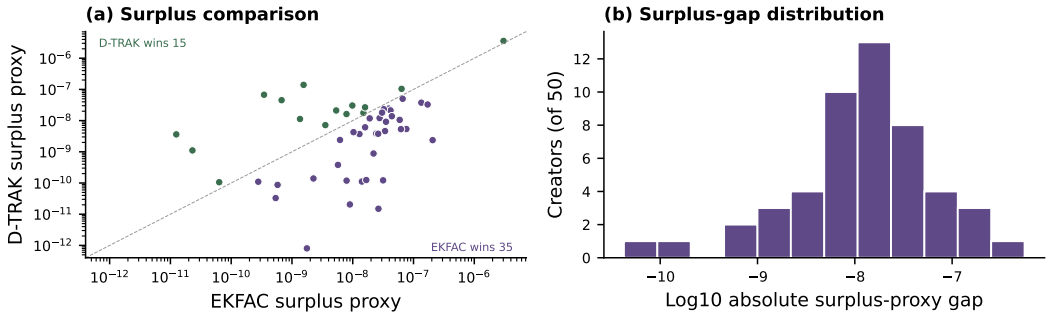

Figure 7: Surplus-proxy comparison between EKFAC and D-TRAK for 50 catalogs. **(a)** Per-creator surplus proxy on each platform; points above the diagonal favor D-TRAK, below favor EKFAC. **(b)** Distribution of absolute surplus-proxy gaps on a logarithmic axis, with median 1.52×10^{-8}.

Augmented water-filling. The priority index extends to incorporate quality dividends (from Appendix B.3), retention dividends (from the dynamic results in Appendix I), and IC costs (from Proposition 8): $\tilde{\phi}_j = \phi_j + \Delta_j^q + \Delta_j^r - \Delta_j^{\mathrm{IC}}$, with $\tilde{\beta}_j^{\mathrm{approx}} \propto (\tilde{\phi}_j \gamma_j)^2$. Its components are measured in different units (effort surplus, retention profit, and alignment value), so this extension is heuristic until they are put on a common scale.

G.6 Multi-platform Bertrand competition between attribution methods

We treat the two state-of-the-art audio attribution methods evaluated in Section 4 , EKFAC influence functions and D-TRAK, as two competing platforms in the Bertrand equilibrium of Theorem 2. Each platform offers each creator the platform-specific welfare-optimal contract from Theorem 1 computed using its own $\widehat{\mathcal{I}}_j^{\mathrm{cal},(m)}$. Under exclusive licensing, creator j joins the platform with the higher maximized surplus $\Sigma_j^{*(m)}$, and the trailing platform's private return to attribution improvement is zero up to the leader's lead.

Surplus proxy. For this numerical illustration we use

$$\Sigma_j^{*(m)} - \mu_j \;\propto\; \widehat{\mathcal{I}}_j^{\mathrm{cal},(m)} \cdot \frac{\mathrm{Var}(a_j^*)}{2\alpha},$$

with $\alpha = 2$ and a unit proportionality constant. The following gap magnitudes use this normalization. This is the precision term of the maximized surplus in Appendix E under additive noise and homogeneous demand, which is linear in informativeness.

Per-creator equilibrium. Of the 50 creators, EKFAC wins 35 and D-TRAK wins 15 (Figure 7 (a)).

Dead-zone distribution. The absolute surplus-proxy gaps range from 4.24×10^{-11} to 5.55×10^{-7}, with median 1.52×10^{-8} (Figure 7 (b)). These normalized gaps illustrate heterogeneity across creators; their magnitudes do not identify monetary investment costs.

H Bargaining with large rights holders

We consider bilateral bargaining over the contract $\theta_j = (f_j, \rho_j)$. Let $\bar{\Pi}_0$ denote the platform's payoff if negotiation fails (it generates without creator j) and $\bar{U}_j$ the creator's outside option. Suppose the platform values payment-contribution alignment as in (C0) of Appendix E, while the creator keeps the mean-variance utility (2), so the fixed fee is a pure transfer and the joint surplus from agreement under royalty ρ_j is

$$S_j(\rho_j) = \mathbb{E}[\pi a_j^*] - \bar{\Pi}_0 - \bar{U}_j + J_j(\rho_j),$$

where J_j is criterion (4). Suppose $S_j(\rho_j^*) > 0$ and the fee implementing the Nash split is nonnegative

$$f_j^{\mathrm{Nash}} = \bar{U}_j + \beta_j S_j(\rho_j^*) - \rho_j^* \mathbb{E}[v_j] + R_j^* \geq 0,$$

where $\beta_j \in (0,1)$ is the creator's bargaining power. Then Nash bargaining selects ρ_j^* of Theorem 1, and creator j receives $\bar{U}_j + \beta_j S_j(\rho_j^*)$. This follows because the unconstrained Nash product at each rate is proportional to $S_j(\rho_j)$, and its maximum is feasible under the stated fee condition. Under classical measurement error and $\alpha_j \geq 1$, $J_j(\rho_j^*) = \Lambda_j \mathcal{I}_j$, so each unit of informativeness adds Λ_j to the joint surplus and the creator captures $\beta_j \Lambda_j$ while the fee condition continues to hold. Without the alignment preference the joint surplus is $\mathbb{E}[\pi a_j^*] - \bar{\Pi}_0 - \bar{U}_j - \frac{\alpha_j}{2}\rho_j^2 \operatorname{Var}(v_j)$, which is maximized at $\rho_j = 0$. If $\bar{U}_j \geq 0$ and this surplus is positive at $\rho_j = 0$, its Nash split has a nonnegative fee, so bargaining alone gives no reason to pay royalties.

I Dynamic participation

The static framework of Section 3 assumes a one-shot licensing decision. In practice, creators observe realized payments and decide whether to continue licensing. This appendix gives the formal dynamic extension: a Bellman equation for the platform's value from each active creator, the steady-state participation rate, and a two-regime characterization of when the platform optimally pays a retention rent.

I.1 Setup

(D1) Myopic exit rule. At the end of period t, creator j exits iff $w_{j,t} < \bar{U}_j$. This captures liquidity-constrained creators or behavioral anchoring on $\bar{U}_j$.

(D2) Re-entry. An exited creator returns the next period with probability $\lambda_j \in (0,1)$.

(D3) Stationary contract. The platform commits to $\theta_j = (f_j^*, \rho_j^*)$ from Theorem 1 each period, plus an optional retention rent $s_j \geq 0$ added to f_j. Assume $f_j^* = \bar{U}_j - \rho_j^* \mathbb{E}[v_j] + R_j^* \geq 0$, so the baseline contract implements binding participation without violating the fee floor.

(D4) Discount factor. The platform discounts at $\delta \in (0,1)$.

(D5) Gaussian payments. $w_{j,t} \mid x_{j,t} = 1 \sim \mathcal{N}(\mathbb{E}[w_j], (\rho_j^*)^2 \sigma_{v,j}^2)$, independently across periods, with $0 < \mathcal{I}_j \leq \alpha_j$ so that $\rho_j^* = \mathcal{I}_j / \alpha_j$.

Define net surplus $N_j := \mathbb{E}[a_{j,t}^* \pi_t] - \bar{U}_j > 0$ and net margin $\mu_j^{\text{net}} := N_j - R_j^* > 0$ (the platform's expected flow profit from an active creator under the welfare-optimal contract, where $R_j^* = \mathcal{I}_j^2 \sigma_{v,j}^2 / (2\alpha_j)$ is the risk premium carried by the fee). For brevity we write $\sigma_{v,j}^2 := \operatorname{Var}(v_{j,t})$ throughout this appendix.

I.2 Static exit probability and the α-cancellation

Theorem 3 (Dynamic participation)**.** *Under (D1) through (D5) and Theorem 1's payment rule:*

(a) Static exit probability. *With zero retention rent,*

$$p_j^{\text{static}} = \Phi(-d_j), \qquad d_j = \frac{\mathcal{I}_j \sigma_{v,j}}{2}.$$

The exit buffer d_j is independent of risk aversion α_j: $\rho_j^* \propto 1/\alpha_j$ exactly offsets the risk-aversion scaling in the variance penalty.

(b) Bellman equation. *The platform's value from an active creator with retention rent $s \geq 0$ is*

$$V_j^A(s) = \frac{\Pi_j(s)}{1 - \delta \tilde{\sigma}_j(s)},$$

where $\Pi_j(s) = \mu_j^{\text{net}} - s$, $\sigma_j(s) = 1 - \Phi(-d_j - s/(\rho_j^* \sigma_{v,j}))$ is the retention probability, $\tilde{\sigma}_j(s) = \omega_j + (1-\omega_j)\sigma_j(s)$ is the effective continuation probability, and $\omega_j = \delta\lambda_j / (1 - \delta(1-\lambda_j))$ is the re-entry discount factor.

(c) Steady-state participation. *In the ergodic steady state,*

$$n_j^* = \frac{\lambda_j}{\lambda_j + p_j(\delta)},$$

where $p_j(\delta) := 1 - \sigma_j(s_j^*(\delta))$ is the exit probability at the platform's optimal retention rent $s_j^*(\delta)$, which maximizes $V_j^A(s)$ over $s \geq 0$.

Proof. (a) Under binding IR, $\mathbb{E}[w_j] = \bar{U}_j + R_j^*$ with $R_j^* = (\alpha_j/2)(\rho_j^*)^2\sigma_{v,j}^2 = \mathcal{I}_j^2\sigma_{v,j}^2/(2\alpha_j)$. Payment standard deviation is $\rho_j^*\sigma_{v,j} = \mathcal{I}_j\sigma_{v,j}/\alpha_j$. Exit iff $w_{j,t} < \bar{U}_j$, so $p_j = \Phi(-R_j^*/(\rho_j^*\sigma_{v,j}))$. The ratio simplifies: $d_j = R_j^*/(\rho_j^*\sigma_{v,j}) = \mathcal{I}_j\sigma_{v,j}/2$, manifestly independent of α_j.

(b) Bellman: $V^A(s) = \Pi(s) + \delta[\sigma(s)V^A(s) + (1-\sigma(s))V^I]$, $V^I = \delta[\lambda V^A(s) + (1-\lambda)V^I]$. Solving the second for $V^I = \omega V^A$ and substituting into the first yields $V^A(s)[1 - \delta(\sigma + (1-\sigma)\omega)] = \Pi(s)$, i.e., $V^A(s) = \Pi(s)/[1 - \delta\tilde{\sigma}(s)]$.

(c) Flow balance of the two-state Markov chain: $p_j n_j = \lambda_j(1 - n_j)$, giving $n_j = \lambda_j/(\lambda_j + p_j)$. □

I.3 Two-regime characterization

Proposition 11 (Two-regime characterization)**.** *The platform optimally sets $s_j^* > 0$ (thick-margin regime) if and only if*

$$\mu_j^{\text{net}} > \mu_j^{\text{crit}} := \frac{p_j^{\text{static}} + \lambda_j/\omega_j}{\sigma_j'(0)}.$$

Otherwise $s_j^* = 0$ (thin-margin regime) and exit occurs at the static rate p_j^{static}.

Proof. Define $\Gamma(s) = F(s) - G(s)$ with $F = \delta(1-\omega)\Pi\sigma'$ and $G = 1 - \delta\tilde{\sigma}$. The first-order condition for the Bellman is $\Gamma(s^*) = 0$, and $(F-G)' = \delta(1-\omega)\Pi\sigma'' < 0$ on $[0, \mu^{\text{net}})$, since $\Pi > 0$ there and $\sigma''(s) = -(d + s/\tau)\phi(d + s/\tau)/\tau^2 < 0$ with $\tau := \rho^*\sigma_v$ and $d > 0$; for $s \geq \mu^{\text{net}}$, $V^A(s) \leq 0 < V^A(0)$, so such s are never optimal. Strictly decreasing Γ implies at most one zero. Evaluate $\Gamma(0) = \delta(1-\omega)\mu^{\text{net}}\sigma'(0) - (1 - \delta\tilde{\sigma}(0))$; the FOC has an interior root iff $\Gamma(0) > 0$, which rearranges to $\mu^{\text{net}} > \mu^{\text{crit}}$ using $1 - \delta\tilde{\sigma}(0) = 1 - \delta + \delta(1-\omega)p^{\text{static}}$ and $(1-\delta)/[\delta(1-\omega)] = \lambda/\omega$. Otherwise the corner solution $s^* = 0$ is optimal. □

Patient-limit behavior. Because $\mu_j^{\text{crit}} = (p_j^{\text{static}} + \lambda_j + (1-\delta)/\delta)/\sigma_j'(0)$ falls with δ, a creator with $\mu_j^{\text{net}} \leq (p_j^{\text{static}} + \lambda_j)/\sigma_j'(0)$ is in the thin-margin regime for every δ, where $p_j(\delta) = p_j^{\text{static}}$, while a creator above that bound becomes thick-margin once the platform is patient enough. In the thick-margin regime, $s_j^*(\delta)$ is strictly increasing in δ, with patient limit $s_j^\infty = \lim_{\delta\to 1} s_j^*(\delta)$ satisfying $\Pi_j(s^\infty)\sigma_j'(s^\infty) = \lambda_j + 1 - \sigma_j(s^\infty)$, and the patient-limit exit probability $p_j^\infty = 1 - \sigma_j(s^\infty) < p_j^{\text{static}}$. The retention rent paid by a patient platform is bounded above by the net margin.

Comparative statics. $\partial p_j^{\text{static}}/\partial\mathcal{I}_j = -(\sigma_{v,j}/2)\phi(d_j) < 0$ (higher informativeness reduces exit) and $\partial p_j^{\text{static}}/\partial\sigma_{v,j} = -(\mathcal{I}_j/2)\phi(d_j) < 0$. The latter is initially counterintuitive (higher signal variance *reduces* exit) and reflects the fact that the risk premium grows quadratically in σ_v while the payment standard deviation grows only linearly, so the buffer-to-std-dev ratio d_j scales with σ_v.

Retention dividend. In the thin-margin regime, improving $\mathcal{I}_j$ with $\sigma_{v,j}$ held fixed raises the steady-state participation rate and yields a marginal profit gain through participation

$$\Delta_j^{\text{ret}} = \frac{\lambda_j(N_j - R_j^*)\sigma_{v,j}\phi(d_j)}{2(\lambda_j + p_j)^2}.$$

This dividend feeds the augmented water-filling rule of Appendix G.5.

J Synthetic data value chains

The platform may use its generative model not only to serve end users but also to produce *synthetic music data* for downstream purposes: fine-tuning specialized models, licensing to third parties, or reducing dependence on future human-created music licensing. This creates a multi-stage value chain: the platform licenses original music → trains a model → generates synthetic data → extracts downstream value.

In stage 1, the platform generates output y with attributed value v_j for creator j. In stage 2, the platform uses its model to produce synthetic data $\tilde{D}$ and extracts downstream value $V^{\mathrm{syn}}(\tilde{D})$. The total value attributable to creator j's original data is not just $a_j^*\pi$ but $a_j^*\pi + \Delta v_j^{\mathrm{syn}}$, where $\Delta v_j^{\mathrm{syn}}$ captures j's marginal contribution to the synthetic data's downstream value, while the payment signal $v_j = \hat{a}_j\pi$ tracks only the first term.

This defines a broader contribution target that includes downstream synthetic-data value. Estimating $\Delta v_j^{\mathrm{syn}}$ and deriving a contract for that target remain outside the present analysis.

K Composite signals and an information bound

This appendix proves Proposition 12 and states the information bound referred to in Section 3.3. The signal bundle is $m_j \in \mathbb{R}^d$, and its variance matrix $\Sigma_j := \mathrm{Var}(m_j)$ is assumed positive definite. Throughout, $\pi \perp (m_j, a_j^*)$; the proof of Proposition 12 treats period revenue π as known, as in the main text, and the information bound below also allows random π.

K.1 Multiple payment signals

Proposition 12 (Any payment signal, the same rule)**.** *Let the platform pay on a bundle of signals $m_j \in \mathbb{R}^d$ rather than on attribution alone, through $w_j = x_j[f_j + \rho_j^\top m_j\,\pi]$, under the assumptions of Theorem 1 applied to m_j in place of $\hat{a}_j$, with $\mathrm{Var}(m_j)$ positive definite. The unconstrained maximizer of the vector form of criterion* (4) *is*

$$\rho_j^* = (1/\alpha_j)\ \mathrm{Var}(m_j)^{-1}\,\mathrm{Cov}(m_j, a_j^*), \tag{13}$$

the coefficient obtained by regressing true contribution on the signals, divided by risk aversion. The single signal $m_j = \hat{a}_j$ recovers the interior rate of Theorem 1. Adding a signal never lowers welfare, and a signal that adds no linear predictive power for a_j^* beyond the others receives zero weight.

Proof.

With π known, each entry of the covariance vector satisfies $\mathrm{Cov}(m_{j,k}\pi, a_j^*\pi) = \pi^2\,\mathrm{Cov}(m_{j,k}, a_j^*)$, and each entry of the variance matrix satisfies $\mathrm{Cov}(m_{j,k}\pi, m_{j,l}\pi) = \pi^2\,\mathrm{Cov}(m_{j,k}, m_{j,l})$. The criterion is therefore $\pi^2\big[\rho^\top c_j - \frac{\alpha_j}{2}\rho^\top\Sigma_j\rho\big]$ with $c_j := \mathrm{Cov}(m_j, a_j^*)$. It is strictly concave because $\Sigma_j \succ 0$, so the first-order condition $c_j - \alpha_j\Sigma_j\rho = 0$ identifies the unique maximizer $\rho_j^* = \alpha_j^{-1}\Sigma_j^{-1}c_j$, and π^2 cancels. Substituting back gives the surplus $\pi^2\, c_j^\top\Sigma_j^{-1}c_j/(2\alpha_j)$. The quadratic form $c_j^\top\Sigma_j^{-1}c_j$ is the variance of the best linear predictor of a_j^* from m_j, which is nondecreasing under extension of m_j by nested-regression monotonicity. The k-th entry of $\Sigma_j^{-1}c_j$ is the partial regression coefficient of a_j^* on $m_{j,k}$ given the other entries, which is zero when $m_{j,k}$ adds no explained variance beyond them. For $d = 1$ and $m_j = \hat{a}_j$, $\Sigma_j^{-1}c_j = \mathrm{Cov}(\hat{a}_j, a_j^*)/\mathrm{Var}(\hat{a}_j) = \mathcal{I}_j$. □

Feasible set. The scalar clip to $[0, 1]$ in Theorem 1 has no unique vector analogue. A per-signal box $\rho \in [0, 1]^d$ and a total-royalty budget $\mathbf{1}^\top\rho \le 1$ are both natural, and the formula above applies when the unconstrained solution is feasible. Which constraint binds in practice is a modeling choice we leave to the application.

K.2 An information bound for arbitrary contract forms

The main text restricts attention to contracts that are linear in the signal. The following bound applies to any measurable payment based on the observed signals and revenue.

Lemma 3 (Information bound on alignment)**.** *Let $w_j = W(m_j, \pi)$ be any measurable payment with finite variance, let $Y_j := a_j^*\pi$ be the contribution target of criterion* (4)*, and write $h_j := \mathbb{E}[Y_j \mid m_j, \pi]$ and $\mu_j(m) := \mathbb{E}[a_j^* \mid m_j = m]$. Then*

$$\big|\,\mathrm{Cov}(w_j, Y_j)\big| \ \le\ \sqrt{\mathrm{Var}(w_j)}\,\sqrt{\mathrm{Var}(h_j)}. \tag{14}$$

If $\mathrm{Var}(h_j) > 0$, equality holds if and only if w_j is a deterministic affine function of h_j. Under $\pi \perp (m_j, a_j^*)$, $h_j = \pi\,\mu_j(m_j)$.

Proof. Because w_j is a function of (m_j, π), the tower property gives $\mathbb{E}[w_j Y_j] = \mathbb{E}[w_j h_j]$ and $\mathbb{E}[Y_j] = \mathbb{E}[h_j]$, so $\mathrm{Cov}(w_j, Y_j) = \mathrm{Cov}(w_j, h_j)$. The Cauchy-Schwarz inequality then gives (14) and its equality case. □

The quantity $B_j := \mathrm{Var}(h_j)$ is the signal's value-explaining power. It depends only on the joint law of (m_j, a_j^*, π) and not on the contract. With π known, $B_j = \pi^2 \mathrm{Var}(\mu_j(m_j))$, and in the additive-noise model of Section 3.1 with $m_j = \hat{a}_j$ and a_j^* also Gaussian, μ_j is affine with slope $\mathcal{I}_j$, so $B_j = \pi^2 \mathcal{I}_j \mathrm{Var}(a_j^*)$ vanishes linearly as informativeness vanishes.

Corollary 5 (The fixed-fee tilt survives any contract form)**.** *Let a planner value payment-contribution alignment at marginal rate $\lambda \geq 0$ and bear risk cost $\frac{\alpha_j}{2} \mathrm{Var}(w_j)$. Write $C := \mathrm{Cov}(w_j, a_j^* \pi)$ and let $B_j := \mathrm{Var}(h_j)$ as above. Then for every measurable contract, $\lambda C - \frac{\alpha_j}{2} \mathrm{Var}(w_j) \leq \lambda C - \frac{\alpha_j}{2} C^2 / B_j$, the optimal alignment satisfies $C^* \leq 2\lambda B_j / \alpha_j$, and the maximum achievable net alignment value is $\lambda^2 B_j / (2\alpha_j)$. All three vanish as $B_j \to 0$.*

Proof. Lemma 3 gives $\mathrm{Var}(w_j) \geq C^2 / B_j$, which yields the first inequality. The upper bound $g(c) = \lambda c - \frac{\alpha_j}{2} c^2 / B_j$ is maximized at $c = \lambda B_j / \alpha_j$ with value $\lambda^2 B_j / (2\alpha_j)$. The fixed-fee contract attains objective 0 with $C = 0$, so any optimal contract has $g(C^*) \geq 0$, which forces $C^* \leq 2\lambda B_j / \alpha_j$. □

Scope of the bound. The bound is attained by a payment affine in h_j. With π known and (m_j, a_j^*) jointly Gaussian, h_j is affine in $m_j \pi$, so the linear royalty family attains the bound and loses nothing, which is the Holmström-Milgrom regime already invoked in Section 3.2. Outside joint Gaussianity the conditional expectation can be nonlinear, and the gap between B_j and its linear-projection counterpart is exactly the room a nonlinear contract has.

Outside the Gaussian model a small linear slope does not imply a small B_j, since a signal can predict contribution nonlinearly with zero linear slope. The bound concerns richer contract forms on a given signal. Learning about attribution quality over time is outside this bound and the dynamics of Appendix I. Minimum guarantees and tiered rates are measurable functions of (m_j, π) and are therefore covered by Lemma 3; recoupable advances, which depend on payment history across periods, are outside it.