

Discussion Paper Series

IZA DP No. 18511

March 2026

The Geometry of Heterogeneous Extremes: Optimal Transport and Entropic Design

Ioan-Sebastian Buhai

SOFI, Stockholm University, UC Chile,
NIPE, University of Minho and IZA@LISER

The IZA Discussion Paper Series (ISSN: 2365-9793) ("Series") is the primary platform for disseminating research produced within the framework of the IZA@LISER Network, an unincorporated international network of labour economists coordinated by the Luxembourg Institute of Socio-Economic Research (LISER). The Series is operated by LISER, a Luxembourg public establishment (établissement public) registered with the Luxembourg Business Registers under number J57, with its registered office at 11, Porte des Sciences, 4366 Esch-sur-Alzette, Grand Duchy of Luxembourg.

Any opinions expressed in this Series are solely those of the author(s). LISER accepts no responsibility or liability for the content of the contributions published herein. LISER adheres to the European Code of Conduct for Research Integrity. Contributions published in this Series present preliminary work intended to foster academic debate. They may be revised, are not definitive, and should be cited accordingly. Copyright remains with the author(s) unless otherwise indicated.



The Geometry of Heterogeneous Extremes: Optimal Transport and Entropic Design*

Abstract

Extreme outcomes depend not only on shock tails but also on heterogeneity in how many opportunities agents get to sample. In the mixed-Poisson search framework, a randomly drawn agent's normalized maximum converges to $H_{v,F}(x) = P_o(v_v(x))$, a Laplace-transform mixture of a classical extreme-value law, with F the mean-one distribution of opportunity intensities.

Treating F as primitive, we study the operator $F \rightarrow H_{v,F}$. That operator representation organizes the paper: convex-order comparisons, the homogeneous benchmark, and pointwise cdf bounds are Laplace-analytic consequences. Its main quantitative payoff is geometric. Via the canonical coupling representation $Z_{v,F} = w_v(E/X)$, optimal transport on transformed types yields explicit Wasserstein stability for the entire induced law, integrated control of the corresponding quantile schedule, canonical ambient interpolation paths, and an explicit renormalization bridge back to the mean-one economic slice. We also give a second-order expansion that separates extreme value theory (EVT) approximation error from the heterogeneity kernel.

As a complementary contribution, we study a Kullback-Leibler regularized design problem that chooses F subject to a mean constraint relative to a baseline. For objectives linear in F , including expected utility of normalized extremes under canonical representation, the solution is the corresponding exponential tilt, with the heterogeneous-EVT kernel supplying the score. A stylized labor market network application interprets F as the cross-sectional distribution of access to job opportunities, shows how the adapted geometry controls counterfactual movements in the full top wage distribution, and renders explicit that the operational robustness claims are conditional on the maintained metric: the main linear theorem lives in the adapted distance $d_{v,p}$, while the bridge from raw space Wasserstein error can be only Hölder in the economically-relevant Fréchet regime. We also distinguish ambient transport geometry from mean-preserving economic counterfactuals and separate finite horizon robustness from asymptotic approximation.

JEL classification

C46, C61, D83, D85, J31, J64

Keywords

extreme value theory, heterogeneous search, optimal transport, Wasserstein distance, entropic design, labor market networks

Corresponding author

Ioan-Sebastian Buhai

ebastian.buhai@sofi.su.se

* This paper grew out of many long conversations about how extremes arise in economics and why they matter. I am grateful to several colleagues in economics, and especially to a few in adjacent disciplines, who convinced me that the argument needed more structure to carry conviction. I also gratefully acknowledge Refine (<https://www.refine.ink/>) for assistance in checking proofs in a preliminary draft. All remaining errors, extreme or otherwise, are mine.

1 Introduction

1.1 Motivation and overall contribution

Extreme outcomes at the agent or unit level often shape broader inequality and efficiency patterns. Examples include a worker’s best wage offer, a firm’s best sourcing or productivity opportunity, and the best match generated through a network. In a broad class of models, the realized payoff is the maximum of a set of opportunities. The size of that set can depend, *inter alia*, on search, attention, or network position.

Standard extreme value theory (EVT) studies maxima of i.i.d. sequences under deterministic sample sizes. Economic environments often feature heterogeneous opportunity sets: different agents sample different numbers of opportunities, and the draw count is itself random. A leading economic formulation is the mixed Poisson framework of Mangin (2026), under which the distribution of a randomly drawn agent’s normalized maximum admits a Laplace-transform representation. If $X \sim F$ indexes an agent’s draw intensity and $P_0(z) = \mathbb{E}[\exp(-zX)]$ denotes the Laplace transform of F , then the distribution of normalized maxima converges to

$$H_{\gamma,F}(x) = P_0(v_\gamma(x)), \quad v_\gamma(x) = -\log H_\gamma(x),$$

where H_γ is the classical extreme-value limit associated with the underlying opportunity distribution. This Laplace-mixture form is the analytical backbone of the paper.

Our contribution is to take the heterogeneity distribution itself as the primitive object and to analyze the induced operator $F \mapsto H_{\gamma,F}$ on the space of type distributions. The novelty claim is deliberately sharp. Once the Laplace-mixture form is established, convex-order comparisons, benchmark comparisons with the homogeneous economy, and several pointwise statements follow from standard one-dimensional Laplace and Wasserstein arguments. Those results are useful, but they are not the paper’s main quantitative novelty.

The operator representation has a geometric payoff. Once the canonical coupling representation $Z_{\gamma,F} = w_\gamma(E/X)$ is available, Theorem 1 turns optimal transport on transformed types into explicit Wasserstein control of the entire induced law of extremes. Corollary 2 packages the same geometry into canonical ambient interpolation paths, Proposition 10 gives an explicit bridge back to the mean-one economic slice, Corollary 3 translates the same bound into control of the entire quantile schedule, and Theorem 2 separates second-order EVT approximation error from the heterogeneity kernel. These are the results that deliver robustness to misspecification or estimation error in F and quantitative comparative statics for whole laws, rather than only pointwise cdf comparisons. The main linear statement is formulated in the adapted metric $d_{\gamma,p}$; the bridge back to raw space Wasserstein error can be weaker and is made explicit later.

We also study a complementary normative problem on the same type space. There, the object is not a full two-marginal transport problem, and its solution does not rely on the Wasserstein geometry developed for the positive results. Instead, the planner chooses a new heterogeneity distribution F subject to a mean constraint and a Kullback-Leibler (KL) penalty relative to a baseline F_0 . This is a one-marginal entropy projection. The Gibbs form of the solution is standard in that class of problems; our contribution is to show how the same Laplace kernel that defines $F \mapsto H_{\gamma,F}$ generates the score functions entering the planner’s first-order conditions for cdf criteria and for expected utility under the canonical stochastic representation. Being explicit about that division of labor sharpens both the mathematics and the interpretation.

Throughout, the formal object is the cross-sectional law of a randomly drawn agent’s normalized maximum under a common offer distribution. Aggregate maxima across agents, as well as settings in which heterogeneity also shifts offer quality or induces dependence between opportunity access and offer values, lie outside the present reduced-form scope.

1.2 Main results

The paper contributes four sets of results.

A structural decomposition of heterogeneous extremes. We formalize the operator view $T : F \mapsto H_{\gamma,F}$ and separate what is purely (Laplace-)analytic from what is genuinely geometric. Order comparisons and several pointwise inequalities follow from the convexity of the Laplace kernel. By contrast, the quantitative comparison of entire laws requires an adapted metric on the type space. This distinction matters for both interpretation and application.

Wasserstein geometry and quantitative robustness. The paper’s main new theorem is Theorem 1: under the regime-specific moment conditions intrinsic to extreme-value theory, the map $F \mapsto H_{\gamma,F}$ is Lipschitz from the ambient transformed type space to the space of extreme laws equipped with Wasserstein distance. The linear bound is therefore stated in the adapted metric $d_{\gamma,p}$, which is itself an ordinary Wasserstein distance after a monotone transform of types. The same construction also furnishes canonical constant-speed geodesics in that ambient geometry, summarized in Corollary 2. When the transform nonlinear, those adapted geodesics need not preserve the economic normalization $\mathbb{E}[X] = 1$ at intermediate times; Proposition 10 therefore provides an explicit quantitative bridge from the ambient path back to the mean-one economic slice. The same whole law control also yields integrated bounds for the full quantile schedule of normalized extremes, not only fixed threshold cdf comparisons. Operationally, this means that one either controls $d_{\gamma,p}$ directly or uses a separate bridge from raw space error, which can be only Hölder in Fréchet settings, unless extra support restrictions make the transform Lipschitz.¹

A complementary entropy-regularized design problem on the same type space. For objectives that are linear in F , including expected utility of normalized extremes under the canonical stochastic representation, we solve the planner’s problem with a KL penalty and a mean constraint. The optimizer is the corresponding exponential tilt of the baseline. This problem is best understood as a one-marginal entropy projection. Its tight connection to the positive analysis is kernel-based rather than transport-based: the score functions entering the tilt are generated by the same heterogeneous-EVT primitives that determine the operator $F \mapsto H_{\gamma,F}$ and its directional derivative.

Application to labor market networks. We apply the framework to a reduced-form labor market environment in which network position governs offer arrival while the offer distribution remains common across workers. The opportunity distribution induced by degree, weighted degree, or other validated proxies for access to search translates into a heterogeneous extreme-value law for a randomly drawn worker’s normalized top wage. The positive results quantify how network inequality changes that cross-sectional right tail distribution, give a concrete whole law counterfactual for the full schedule of top wage quantiles, and make explicit the conditions under which measurement error in network heterogeneity propagates into tail prediction error.

¹In plain terms, the theorem controls not just one tail probability but the whole distribution of normalized extremes, and hence the whole quantile curve, in an integrated sense. In applications, the most direct route is to measure heterogeneity in the adapted metric $d_{\gamma,p}$. If one instead starts from error in the original type space, an additional comparison is needed. In Fréchet settings that comparison may be weaker than linear unless support restrictions keep the relevant transform Lipschitz, for instance by keeping the support away from zero.

1.3 Empirical interface and identification of heterogeneity

The empirical content of the framework is promising, though one should be careful about what is and is not identified. If a cross section of offer counts is observed at a fixed horizon, then the data identify the count distribution and its pgf. Under the mixed-Poisson structure, that pgf identified from data pins down the Laplace transform P_0 on the interval $z \in [0, \theta]$. Because P_0 is a Laplace transform, knowledge of that transform on any open interval identifies the underlying mixing distribution F at the population level. The real difficulty is statistical rather than logical: analytic continuation and Laplace inversion are severely ill-posed, so richer variation, e.g., multiple horizons, repeated exposures, or parametric/semi-parametric structure, remains highly valuable for stable estimation rather than for point identification per se.

A second empirical route uses direct proxies for opportunity intensity, such as normalized degree, referral exposure, centrality, or estimated arrival rates. That route identifies an empirical approximation to F after one specifies a modeling map from the observed proxy into the latent intensity entering the mixed-Poisson law. Likewise, network data by themselves do not identify the adapted metric $d_{\gamma,p}$; that metric becomes operational once the researcher has committed to a tail index γ and to a maintained measurement model linking the observed proxy to the latent type entering the relevant stability theorem.

The same caution applies to inference from extremes. Using the heterogeneous extreme-value limit to infer F from tail behavior requires the tail-limit ingredients, i.e. the index γ , the normalizations (a_θ, b_θ) , and, for second-order refinements, the corresponding second-order objects, either to be known or to be estimated in a first step. The full parent law G need not be known for the EVT approximation itself. Accordingly, our paper emphasizes robustness and error propagation more than stable nonparametric recovery. The main operational message is therefore conditional: once the analyst has an estimate \hat{F} and justified control of the adapted metric $d_{\gamma,p}$ around \hat{F} , the stability results convert uncertainty about heterogeneity into explicit uncertainty about tail predictions. If one starts instead from raw space Wasserstein error, Proposition 11 supplies the bridge. In particular, when $0 < \gamma < 1$ and the assumptions of Theorem 1 hold with $p\gamma < 1$, that bridge is only Hölder of order γ , so Fréchet-side error propagation can be materially slower than linear.

1.4 Related literature

Heterogeneous extreme value theory. We build on Mangin (2026); an earlier circulation is Becker and Mangin (2023). On the probability side, the mixed-Poisson law $H_{\gamma,F}$ sits in the classical lineage on maxima under random sample size or random indexing (Barndorff-Nielsen, 1964; Galambos, 1973; Silvestrov and Teugels, 1998), with related max-geometric and max-semistable notions in, e.g., Rachev and Resnick (1991) and Megyesi (2002). Our departure from that literature is to treat the heterogeneity distribution itself as the primitive object and to study the induced operator $F \mapsto H_{\gamma,F}$ through order, metric, and variational tools. Adjacent econometric work by Einmahl and He (2023) studies extreme-value estimation under heterogeneous marginals rather than heterogeneity in the intensity of opportunity arrival. For classical EVT background, see de Haan and Ferreira (2006) and Resnick (2008); for complementary peaks-over-threshold foundations, see Balkema and de Haan (1974) and Pickands (1975).

Optimal transport, coupling, and entropy projection. Our positive results use optimal transport as a geometry on distributions (Villani, 2009; Galichon, 2016) and are close in spirit to the coupling approach to extremes in Bobbia et al. (2021); complementary Wasserstein-type control of Fréchet approximation via Stein methods is developed in Mansanarez et al. (2025). On the normative side, our planner problem is a one-marginal entropy projection with a linear moment

constraint. The duality and exponential-tilt characterization connect to the Gibbs variational principle (Donsker and Varadhan, 1975) and, more loosely, to modern Schrödinger and entropic optimal transport treatments (Cuturi, 2013; Léonard, 2014; Peyré and Cuturi, 2019; Ghosal et al., 2022; Nutz, 2022). We therefore use optimal transport and entropy regularization in related but distinct ways in the positive and normative parts of the paper.

Labor market networks. Our application draws on the literature on job information and referrals in networks, including Granovetter (1973), Montgomery (1991), Topa (2001), Calvó-Armengol and Jackson (2004), and Ioannides and Loury (2004), as well as more recent work on referral inequality and segregation such as Buhai and van der Leij (2023) and Bolte et al. (2024). Relative to that literature, we emphasize a tail-sensitive reduced form: how the distribution of opportunities maps into the distribution of worker-level extremes, how robust that mapping is to perturbations in heterogeneity, and how policy might reshape the opportunity distribution when the right tail is economically consequential.

1.5 Roadmap and regime map

Structure of the paper. Section 2 introduces the heterogeneous extreme-value environment. Section 3 develops order properties, misallocation indices, and geodesic structure on the type space. Section 4 contains the main metric stability results, finite horizon robustness bounds, and the second-order expansion. Section 5 studies the complementary entropy-regularized design problem. Section 6 presents the labor market network application. Section 7 concludes. Appendix A contains proofs and Appendix B records technical tools.

Before moving on to the main sections, we already include Table 1 here in the Introduction, as a pragmatic overview of the paper’s regime-specific scope. A recurring theme is that lower-support restrictions are needed only when logarithmic, negative-power, or inverse-power transforms of X enter the argument. Another is that the adapted geodesics of Section 4 live in an ambient transformed type space; when s_γ is nonlinear, they need not preserve the mean-one normalization at intermediate times. Proposition 10 quantifies the discrepancy between the ambient path and its mean-one renormalization, while the raw space geodesics and pointwise comparisons in Section 3 remain available when exact mean preservation is required throughout.

Table 1: Scope of the main analytical results across extreme-value regimes

Result	Regime	Additional conditions
Proposition 1, Proposition 2, Corollary 4, Proposition 13, Theorem 2	all γ	unified baseline Assumptions 1–2; no positive lower support bound is imposed
Proposition 4 and Lemma 3	Fréchet / $\gamma > 0$	product representation specific to the Fréchet case; no support bound away from zero is required there
Theorem 1	all γ	$d_{\gamma,p}(F_1, F_2) < \infty$ is the primitive domain condition; equivalently, one needs finite transformed p th moments, namely $X^{\gamma p}$ for $\gamma > 0$, $ \log X ^p$ for $\gamma = 0$, and $X^{- \gamma p}$ for $\gamma < 0$
Proposition 11	$\gamma \in (0, 1]$ or $\gamma \leq 0$	for $\gamma \in (0, 1]$, no lower support bound is needed; for $\gamma \leq 0$, support on $[a, \infty)$ is imposed explicitly
Section 5.5 power and inverse-power tilts	all regimes in principle	inverse-power scores require baseline support bounded away from zero; positive-power scores require the baseline tail to be light enough for the exponential tilt to be integrable

2 Environment: heterogeneous extreme value primitives

2.1 Baseline extreme value theory

Let $\{Y_j\}_{j \geq 1}$ be i.i.d. real valued random variables with common cdf G . Let $M_n := \max_{1 \leq j \leq n} Y_j$ denote the sample maximum.

We impose a standard maximum domain of attraction condition. Equivalently, after affine normalization, the law of M_n converges to a generalized extreme value distribution.

Assumption 1 (Domain of attraction). There exist sequences $a_n > 0$ and $b_n \in \mathbb{R}$ and a parameter $\gamma \in \mathbb{R}$ such that

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) \rightarrow H_\gamma(x) \quad \text{as } n \rightarrow \infty,$$

for all continuity points x of H_γ .

We use the canonical generalized extreme value family

$$H_\gamma(x) = \begin{cases} \exp(-(1 + \gamma x)^{-1/\gamma}), & \gamma \neq 0, 1 + \gamma x > 0, \\ \exp(-e^{-x}), & \gamma = 0, \end{cases}$$

with the conventions $H_\gamma(x) = 0$ for $\gamma > 0$ and $x \leq -1/\gamma$, and $H_\gamma(x) = 1$ for $\gamma < 0$ and $x \geq -1/\gamma$. Define the tail transform

$$v_\gamma(x) := -\log H_\gamma(x),$$

so that $v_\gamma(x) = (1 + \gamma x)^{-1/\gamma}$ for $\gamma \neq 0$ on $\{1 + \gamma x > 0\}$ and $v_0(x) = e^{-x}$ on \mathbb{R} .

2.2 Mixed Poisson heterogeneity in draw counts

Each agent draws a random number $N(\theta)$ of i.i.d. opportunities, where $\theta > 0$ scales the average arrival rate. Following Mangin (2026), the draw count is mixed Poisson.

Assumption 2 (Mixed Poisson heterogeneity). Let $X \sim F$ on $[0, \infty)$ with $\mathbb{E}[X] = 1$ and $\Pr(X = 0) = 0$. Conditional on X , the draw count is Poisson:

$$N(\theta) \mid X \sim \text{Poisson}(\theta X), \quad \theta > 0.$$

The sequence $\{Y_j\}_{j \geq 1}$ is independent of $(X, N(\theta))$.

Remark 1 (Why we exclude an atom at zero). The restriction $\Pr(X = 0) = 0$ is mainly a unified domain convention rather than a substantive claim that near-zero opportunity types are unimportant. Mass arbitrarily close to zero is allowed throughout the baseline model. The reason to exclude an atom exactly at zero is that, for $\gamma \leq 0$, it produces an atom at $-\infty$ in the normalized limit law, while several metric constructions below use $\log x$ or x^γ and are therefore naturally formulated on $(0, \infty)$. When a positive lower support bound is genuinely needed, we impose it explicitly in the corresponding proposition.

The distribution of $N(\theta)$ is mixed Poisson with mean $\mathbb{E}[N(\theta)] = \theta$. Let

$$P_0(z) := \mathbb{E}[e^{-zX}], \quad z \geq 0,$$

denote the Laplace transform of F . The next lemma summarizes the key generating function identity that will be used repeatedly.

Lemma 1 (Probability generating function of mixed Poisson). *Under Assumption 2, for any $y \in [0, 1]$ and any $\theta > 0$,*

$$\mathbb{E}\left[y^{N(\theta)}\right] = P_0(\theta(1 - y)).$$

Remark 2 (Search technology versus type heterogeneity). Assumption 2 admits two complementary interpretations. One may take the mixed Poisson law of $N(\theta)$ as the primitive “search technology”, in which case F is an equivalent representation. Alternatively, one may take F as the primitive distribution of types X , where higher X implies a higher mean draw count θX . The Laplace transform P_0 summarizes the search technology and is the only object from F that enters the distribution of extremes. The mixed Poisson specification is analytically valuable precisely because the conditional *distribution function* of the maximum takes the exponential form

$$\Pr(M_\theta \leq x \mid X) = \exp\{-\theta X(1 - G(x))\},$$

so heterogeneity enters through a Laplace transform rather than through a more complicated count generating function. Its complement gives the corresponding conditional tail probability. This special role of Poisson search is also emphasized in Becker and Mangin (2023), which contrasts the Poisson benchmark with more general count technologies.

2.3 Heterogeneous maxima and their distribution

Given $N(\theta)$ draws, the realized outcome is the maximum

$$M_\theta := \sup_{1 \leq j \leq N(\theta)} Y_j,$$

with the convention $\sup \emptyset = -\infty$.

For any $x \in \mathbb{R}$,

$$\Pr(M_\theta \leq x \mid X) = \mathbb{E}\left[G(x)^{N(\theta)} \mid X\right] = \exp(-\theta X(1 - G(x))),$$

and hence, by using Lemma 1,

$$\Pr(M_\theta \leq x) = P_0(\theta(1 - G(x))). \tag{1}$$

2.4 Heterogeneous extreme value limit law

We now combine the domain of attraction condition with the mixed Poisson structure. Define $a_\theta := a_{\lfloor \theta \rfloor}$ and $b_\theta := b_{\lfloor \theta \rfloor}$ for $\theta \geq 1$. This step extension is convenient for the first-order limit theorem below. In the second-order subsection later on, we revert to continuous normalizing functions, as is standard in extreme value theory, to avoid discretization artifacts.

Proposition 1 (Heterogeneous extreme value law). *Suppose Assumptions 1 and 2 hold. Then for all continuity points x of $H_{\gamma,F}$,*

$$\Pr\left(\frac{M_\theta - b_\theta}{a_\theta} \leq x\right) \rightarrow H_{\gamma,F}(x) := P_0(v_\gamma(x)) \quad \text{as } \theta \rightarrow \infty.$$

Moreover, if $F = \delta_1$, then $P_0(z) = e^{-z}$ and $H_{\gamma,F} = H_\gamma$.

Remark 3 (Formal scope of the limit law). The object $H_{\gamma,F}$ is the limit distribution of the normalized maximum for a randomly drawn agent under a common offer distribution G . It is therefore a cross-sectional agent-level object. The paper does not analyze the economy-wide maximum $\max_i M_{i,\theta}$, nor settings in which network position also changes the distribution of the offer quality or induces dependence between X and the Y_{ij} . Those extensions are potentially important, but they require additional structure beyond the present mixed-Poisson layer.

Remark 4 (Connection to random-sample-size maxima). Proposition 1 is a mixed-Poisson instance of the classical literature on maxima with random sample size or random indexing; see, e.g., Barndorff-Nielsen (1964), Galambos (1973), and Silvestrov and Teugels (1998). In that literature, random indexing changes the form of the limit law while often preserving the underlying extreme-value type. The value added here is not another domain-of-attraction result, but the decision to treat the mean one heterogeneity distribution F and its Laplace transform P_0 as primitives for the order, metric, and design analysis developed below.

Comment. The representation (1) reduces the asymptotics of M_θ to the behavior of $\theta(1 - G(b_\theta + a_\theta x))$. Assumption 1 implies that $G(b_\theta + a_\theta x)^{[\theta]} \rightarrow H_\gamma(x) = e^{-v_\gamma(x)}$ and thus $\theta(1 - G(b_\theta + a_\theta x)) \rightarrow v_\gamma(x)$ at continuity points. The formal argument is given in Appendix A. Our focus on block maxima is natural in search environments where the economic object is the best offer over a horizon. Complementary peaks-over-threshold formulations are classical in extreme value theory (see, e.g., Pickands (1975) and Balkema and de Haan (1974)).

2.5 A canonical product representation in the Fréchet case

A particularly useful representation emerges under Fréchet normalization, which is natural for heavy-tailed primitives. Suppose $\gamma > 0$ and work with the standard Fréchet cdf

$$H_\gamma^{\text{Fr}}(z) = \exp\left(-z^{-1/\gamma}\right), \quad z > 0.$$

If Z_γ is a Fréchet random variable with cdf H_γ^{Fr} and Z_γ is independent of X , then

$$\Pr(X^\gamma Z_\gamma \leq z) = \mathbb{E}\left[\exp\left(-X z^{-1/\gamma}\right)\right] = P_0\left(z^{-1/\gamma}\right), \quad z > 0.$$

Thus the heterogeneous Fréchet limit admits the product representation

$$Z \stackrel{d}{=} X^\gamma Z_\gamma, \tag{2}$$

where Z has cdf $z \mapsto P_0(z^{-1/\gamma})$ on $(0, \infty)$. Equivalently, the canonically normalized GEV variable

$$Z_{\text{GEV}} := \frac{Z - 1}{\gamma}$$

has cdf $H_{\gamma,F}$. We shall use (2) to construct couplings and to obtain sharp Wasserstein bounds in Section 4.

2.6 Extreme outcome functionals

The paper studies tail-sensitive functionals of the distribution of extremes. Let $\zeta : \mathbb{R} \rightarrow \mathbb{R}$ be a measurable payoff function. For finite θ , define

$$\Phi_\theta(F) := \mathbb{E}[\zeta(M_\theta)].$$

For normalized extremes, define

$$Z_\theta := \frac{M_\theta - b_\theta}{a_\theta}, \quad \Psi_\theta(F) := \mathbb{E}[\psi(Z_\theta)]$$

for measurable ψ such that the expectation is well defined. Section 4 establishes pointwise finite horizon stability for the cdf of Z_θ , Wasserstein stability for the limit law $F \mapsto \mathcal{L}(Z)$, and derives Lipschitz-type bounds for classes of functionals ψ that are natural for tail analysis.

Section 5 uses these objects to define and solve an entropy-regularized design problem for F .

3 Geometry of heterogeneous extreme value laws

This section develops structural and geometric properties of the heterogeneous extreme value operator introduced in Section 2. Throughout, F denotes a probability measure on $[0, \infty)$ with $\mathbb{E}_F[X] = 1$ and $\Pr_F(X = 0) = 0$, and we write

$$P_0(z) := \mathbb{E}_F[e^{-zX}] = \int_0^\infty e^{-zx} F(dx), \quad z \geq 0,$$

for the Laplace transform of F . For a fixed tail index $\gamma \in \mathbb{R}$ and tail transform $v_\gamma(x) := -\log H_\gamma(x)$ (defined on $\{x : H_\gamma(x) \in (0, 1)\}$), the heterogeneous extreme value law is

$$H_{\gamma, F}(x) = P_0(v_\gamma(x)).$$

We view the mapping

$$T : F \mapsto H_{\gamma, F}$$

as an operator from a space of type distributions to a space of extreme value laws. The section has two roles. Subsection 3.1 records order and benchmark consequences of the Laplace-mixture structure. Those results are useful for interpretation and for the application, but they are not the core geometric innovation. The later subsections then isolate the coupling and variational ingredients needed for the genuinely geometric stability results in Section 4 and, more indirectly, for the complementary design problem in Section 5.

3.1 Order comparisons and extreme misallocation

A central economic question is how dispersion in opportunities changes extreme outcomes. In our setting, dispersion enters through the distribution F of draw intensities X . A natural formalization is convex order, which coincides with mean-preserving spreads.

Definition 1 (Convex order). Let F_1, F_2 be probability measures on $[0, \infty)$ with finite first moment and common mean. We write $F_2 \succeq_{\text{cx}} F_1$ if

$$\int \varphi(x) F_2(dx) \geq \int \varphi(x) F_1(dx)$$

for every convex function $\varphi : [0, \infty) \rightarrow \mathbb{R}$ for which both expectations are finite. When $F_2 \succeq_{\text{cx}} F_1$, F_2 is a mean preserving spread of F_1 in the sense of Rothschild and Stiglitz (1970).

Convex order has an immediate implication for Laplace transforms because $x \mapsto e^{-zx}$ is convex on $[0, \infty)$ for every $z \geq 0$.

Proposition 2 (Convex order implies Laplace order). *Let F_1, F_2 be probability measures on $[0, \infty)$ with $\mathbb{E}[X] = 1$. If $F_2 \succeq_{\text{cx}} F_1$, then for every $z \geq 0$,*

$$P_0^{(2)}(z) \geq P_0^{(1)}(z).$$

Consequently, for every x such that $v_\gamma(x) < \infty$,

$$H_{\gamma, F_2}(x) \geq H_{\gamma, F_1}(x).$$

Equivalently, if Z_i has cdf H_{γ, F_i} , then $\Pr(Z_2 > t) \leq \Pr(Z_1 > t)$ for all $t \in \mathbb{R}$.

Corollary 1 (Heterogeneity lowers extremes relative to the homogeneous benchmark). *Let F be a probability measure on $[0, \infty)$ with $\mathbb{E}[X] = 1$. Then for every $z \geq 0$,*

$$P_0(z) \geq e^{-z},$$

with strict inequality for all $z > 0$ whenever $F \neq \delta_1$. Consequently, for every x such that $v_\gamma(x) < \infty$,

$$H_{\gamma, F}(x) \geq H_\gamma(x),$$

so the heterogeneous extreme value limit is stochastically smaller than the homogeneous limit.

Remark 5 (Tail index versus level effects). Order comparisons such as Proposition 2 operate at the level of the entire distribution function. They are consistent with the tail index invariance emphasized by Mangin (2026): heterogeneity in the mixed Poisson draw counts changes the level of extreme outcomes but does not change the limiting tail index γ . Under the mean-one normalization, one can actually say more: since $0 \leq (1 - e^{-zX})/z \leq X$ for $z > 0$ and $\mathbb{E}[X] = 1$, dominated convergence gives $P_0(z) = 1 - z + o(z)$ as $z \downarrow 0$. Consequently, $1 - H_{\gamma, F}(x) = 1 - P_0(v_\gamma(x)) \sim v_\gamma(x) \sim 1 - H_\gamma(x)$ whenever $v_\gamma(x) \downarrow 0$. Thus heterogeneity changes distributional levels and high quantiles, but it does not change the first-order ultra-tail mass.

The previous results motivate quantitative measures of how far F is from the egalitarian benchmark δ_1 .

Definition 2 (Extreme misallocation indices). Fix the benchmark $F^{\text{eq}} := \delta_1$.

1. For $p \geq 1$ and $F \in \mathcal{P}_p([0, \infty))$, define the metric misallocation index

$$\mathcal{M}_p(F) := W_p(F, \delta_1).$$

2. Further define the Laplace misallocation curve

$$\Delta_F(z) := P_0(z) - e^{-z}, \quad z \geq 0.$$

Remark 6 (Interpretation). $\mathcal{M}_p(F)$ measures opportunity dispersion in a transportation geometry and will be directly compatible with the stability bounds in Section 4. Because the benchmark is a point mass, the first two indices are especially transparent: $\mathcal{M}_1(F) = \mathbb{E}[|X - 1|]$ and, under $\mathbb{E}[X] = 1$, $\mathcal{M}_2(F) = \sqrt{\mathbb{E}[(X - 1)^2]} = \sqrt{\text{Var}(X)}$. The Laplace curve $z \mapsto \Delta_F(z)$ measures the induced distortion in extreme value laws at the Laplace scale. By Corollary 1, $\Delta_F(z) \geq 0$ for all $z \geq 0$, and $\Delta_F \equiv 0$ if and only if $F = \delta_1$. In particular, these indices quantify how far the opportunity distribution is from the homogeneous benchmark before Section 4 translates that gap into explicit bounds on the induced law of extremes.

3.2 Wasserstein geometry on the type space

Now let $\mathcal{P}_p([0, \infty))$ denote the set of probability measures on $[0, \infty)$ with finite p th moment. For $\mu, \nu \in \mathcal{P}_p([0, \infty))$, the Wasserstein p distance is

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Gamma(\mu, \nu)} \int_{[0, \infty)^2} |x - y|^p \pi(dx, dy) \right)^{1/p},$$

where $\Gamma(\mu, \nu)$ is the set of couplings with marginals μ and ν . We refer to Villani (2009) and Galichon (2016) for background.

In one dimension, the geometry is particularly explicit.

Lemma 2 (Quantile representation and canonical geodesics in one dimension). *Let $\mu, \nu \in \mathcal{P}_p([0, \infty))$ and let Q_μ, Q_ν denote their quantile functions. Then*

$$W_p^p(\mu, \nu) = \int_0^1 |Q_\mu(u) - Q_\nu(u)|^p du.$$

Moreover, the path $(\mu_t)_{t \in [0, 1]}$ defined by

$$Q_{\mu_t}(u) = (1 - t)Q_\mu(u) + tQ_\nu(u), \quad u \in (0, 1),$$

is a constant-speed W_p geodesic from μ to ν . For $p > 1$ it is the unique geodesic, while for $p = 1$ it is the canonical monotone geodesic and uniqueness need not hold.

This one-dimensional characterization has a direct implication for convexity of expectations along the canonical monotone geodesic.

Proposition 3 (Convexity of expectations along type geodesics). *Let $\mu, \nu \in \mathcal{P}_1([0, \infty))$ and let $(\mu_t)_{t \in [0, 1]}$ be the canonical monotone Wasserstein geodesic from μ to ν . Let $\varphi : [0, \infty) \rightarrow \mathbb{R}$ be convex and such that $\int |\varphi| d\mu_t < \infty$ for all $t \in [0, 1]$. Then $t \mapsto \int \varphi(x) \mu_t(dx)$ is convex on $[0, 1]$.*

3.3 The extreme value map as a pushforward and coupling transform

The map T admits a useful representation in cases where the limit law can be written as a measurable image of (X, Ξ) for an auxiliary variable Ξ independent of X . This representation is the basis for the coupling constructions in Section 4 and for the variational arguments in Section 5.

Proposition 4 (Pushforward representations in Fréchet and Gumbel cases). *1. Fréchet case ($\gamma > 0$). Let Z_γ have the standard Fréchet law on $(0, \infty)$. Its cdf is $z \mapsto \exp(-z^{-1/\gamma})$, and Z_γ is independent of $X \sim F$. Then the heterogeneous Fréchet limit Z satisfies*

$$Z \stackrel{d}{=} X^\gamma Z_\gamma,$$

and its cdf is $z \mapsto P_0(z^{-1/\gamma})$ on $(0, \infty)$. Equivalently, the canonically normalized variable

$$Z_{\text{GEV}} := \frac{Z - 1}{\gamma}$$

is distributed according to $H_{\gamma, F}$.

2. Gumbel case ($\gamma = 0$). Let E be exponential with mean one, independent of $X \sim F$. Define $Z := \log(X/E)$. Then Z has cdf $x \mapsto P_0(e^{-x})$, hence Z is distributed according to $H_{0, F}$.

A key implication is that many extreme outcome functionals are linear in the mixing distribution F once the auxiliary variable is integrated out.

Lemma 3 (Kernel representation of expectations in the Fréchet case). *Assume $\gamma > 0$ and let $Z = X^\gamma Z_\gamma$ be as in Proposition 4. Let $\psi : (0, \infty) \rightarrow \mathbb{R}$ be measurable and assume $\mathbb{E}[|\psi(Z)|] < \infty$. Define the kernel*

$$\kappa_\psi(x) := \mathbb{E}[\psi(x^\gamma Z_\gamma)], \quad x > 0.$$

Then

$$\mathbb{E}[\psi(Z)] = \int_0^\infty \kappa_\psi(x) F(dx).$$

Lemma 3 is conceptually important for design: whenever the planner objective can be written as $\mathbb{E}[\psi(Z)]$ with Z in product form, the dependence on F is through a linear functional with kernel κ_ψ .

3.4 Differentiating the map $F \mapsto H_{\gamma,F}$

Because $H_{\gamma,F}(x)$ is an integral of a fixed kernel against F , the operator T has an explicit variational derivative. Let $\mathcal{M}([0, \infty))$ denote the space of finite signed measures on $[0, \infty)$.

Proposition 5 (Gâteaux derivative of the extreme value operator). *Fix γ and consider the map $T : F \mapsto H_{\gamma,F}$. Let $\nu \in \mathcal{M}([0, \infty))$ satisfy $\nu([0, \infty)) = 0$ and $\int_0^\infty u \nu(du) = 0$, and let $F_\varepsilon := F + \varepsilon \nu$ be a perturbation for ε in a neighborhood of 0 such that F_ε remains a probability measure. Then for every x with $v_\gamma(x) < \infty$,*

$$\left. \frac{d}{d\varepsilon} H_{\gamma,F_\varepsilon}(x) \right|_{\varepsilon=0} = \int_0^\infty e^{-v_\gamma(x)u} \nu(du).$$

Remark 7 (Why the mean constraint is stated here). The derivative formula itself is algebraically valid without the restriction $\int u \nu(du) = 0$. We impose it in the proposition because the economic domain studied in the paper fixes $\mathbb{E}[X] = 1$, so admissible perturbations should remain on that slice.

Remark 8 (Link to entropic design). Proposition 5 identifies the kernel that governs marginal changes in the distribution of extremes when the type distribution is perturbed. In Section 5, the same kernel reappears in the entropy-regularized design problem whenever the planner objective is expressed through cdf levels or their smooth linear combinations.

3.5 Geodesics and convexity

We now combine the one-dimensional canonical Wasserstein geodesics with convexity of the Laplace kernel.

Proposition 6 (Convexity of heterogeneous extreme value laws along type geodesics). *Let $F^0, F^1 \in \mathcal{P}_1([0, \infty))$ with $\mathbb{E}[X] = 1$ and let $(F^t)_{t \in [0,1]}$ be the canonical monotone Wasserstein geodesic between them. Then for every $z \geq 0$, the map $t \mapsto P_0^t(z) := \int e^{-zx} F^t(dx)$ is convex on $[0, 1]$. Consequently, for every x such that $v_\gamma(x) < \infty$, the map*

$$t \mapsto H_{\gamma,F^t}(x) = P_0^t(v_\gamma(x))$$

is convex on $[0, 1]$.

Proposition 6 provides a disciplined way to compare extremes along canonical paths in the space of heterogeneity distributions. It is also useful for identification and policy discussions: when F is estimated, geodesic perturbations provide a structured local sensitivity analysis.

Remark 9 (Raw-space geodesics preserve the mean one slice). If F^0 and F^1 both satisfy $\int x F^k(dx) = 1$, then the canonical monotone Wasserstein geodesic $(F^t)_{t \in [0,1]}$ in Proposition 6 also satisfies $\int x F^t(dx) = 1$ for every t . Indeed, in one dimension the mean equals the integral of the quantile function, and quantiles interpolate linearly along this path. This mean-preserving property is specific to the raw type geometry used in this section and will contrast with the adapted geodesics of Section 4.

We also record a moment functional that will recur in heavy tailed applications.

Proposition 7 (Geodesic convexity of negative power moments). *Fix $\rho > 0$. Let F^0, F^1 be supported on $[a, \infty)$ for some $a > 0$, and let $(F^t)_{t \in [0,1]}$ be their canonical monotone Wasserstein geodesic. Then the map*

$$t \mapsto \int x^{-\rho} F^t(dx)$$

is convex on $[0, 1]$.

Remark 10 (Where this enters later). In Fréchet settings, many tail sensitive functionals can be written in terms of power moments of X through the product representation in Proposition 4 and the kernel representation in Lemma 3. The convexity statements above allow one to control these objects along structured perturbations of F .

4 Wasserstein stability and coupling bounds

This section contains the genuinely geometric layer of the paper. The Laplace-mixture representation already gives several pointwise and order-theoretic statements. What transport adds is different: by coupling transformed types with a common exponential shock, we obtain quantitative control of whole induced laws, canonical ambient interpolation paths, and robustness statements with explicit moduli. The key object remains the operator $T : F \mapsto H_{\gamma, F}$. The substantive content is not merely that one may rewrite the type space after a monotone transform, but that the operator representation and the canonical coupling make this transformed geometry propagate into law-level Wasserstein bounds for heterogeneous extremes. Complementary quantitative control of extreme-value approximation in Wasserstein-type distances has also recently been developed with Stein-type methods for Fréchet laws; see Mansanarez et al. (2025).

4.1 Wasserstein preliminaries and duality

We work on \mathbb{R} equipped with the metric $d(x, y) = |x - y|$. For $p \geq 1$ and $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$,

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Gamma(\mu, \nu)} \int_{\mathbb{R}^2} |x - y|^p \pi(dx, dy) \right)^{1/p},$$

where $\Gamma(\mu, \nu)$ is the set of couplings with marginals (μ, ν) . In one dimension, W_p admits the quantile representation, and for $p = 1$ it admits the Kantorovich-Rubinstein dual formulation. We collect the statements used below in Appendix B.1.

4.2 A canonical coupling representation

Recall the canonical GEV parametrization from Section 2.1 and define

$$v_\gamma(x) := -\log H_\gamma(x).$$

Let w_γ denote the inverse of v_γ on its range. For the canonical GEV family,

$$w_\gamma(t) = \begin{cases} \frac{t^{-\gamma} - 1}{\gamma}, & \gamma \neq 0, \\ -\log t, & \gamma = 0, \end{cases} \quad t > 0.$$

Proposition 8 (Canonical representation of $H_{\gamma,F}$). *Let $X \sim F$ and let $E \sim \text{Exp}(1)$ be independent of X . Define*

$$Z_{\gamma,F} := w_\gamma\left(\frac{E}{X}\right).$$

Then $Z_{\gamma,F}$ has cdf $H_{\gamma,F}$.

Equivalently,

$$Z_{\gamma,F} = \begin{cases} \frac{X^\gamma E^{-\gamma} - 1}{\gamma}, & \gamma \neq 0, \\ \log X - \log E, & \gamma = 0, \end{cases}$$

and this random variable has cdf $H_{\gamma,F}$.

Proposition 8 reduces the construction of couplings for $H_{\gamma,F}$ to couplings on the type space. Given any coupling of $X_1 \sim F_1$ and $X_2 \sim F_2$, one obtains a coupling of the corresponding extreme laws by using a common exponential variable E .

4.3 Stability of the extreme value operator

The canonical representation depends on X through the transform $x \mapsto x^\gamma$ when $\gamma \neq 0$, and through $x \mapsto \log x$ when $\gamma = 0$. This motivates metrics on the type space that are adapted to the index γ . For the geometric results in this section, the mean one normalization from Assumption 2 is not needed. We therefore temporarily enlarge the domain from the normalized economic type space to the ambient space of probability measures on $(0, \infty)$ for which the relevant transformed p th moment is finite. That ambient domain is the natural one for the adapted metric and, crucially, the one on which the geodesic construction is closed.

Definition 3 (Type metrics adapted to γ). Fix $p \geq 1$ and $\gamma \in \mathbb{R}$. Let $s_\gamma : (0, \infty) \rightarrow \mathbb{R}$ be defined by

$$s_\gamma(x) := \begin{cases} x^\gamma, & \gamma \neq 0, \\ \log x, & \gamma = 0. \end{cases}$$

For F_1, F_2 on $(0, \infty)$ such that the pushforwards $F_i \circ s_\gamma^{-1}$ belong to $\mathcal{P}_p(\mathbb{R})$, define

$$d_{\gamma,p}(F_1, F_2) := W_p(F_1 \circ s_\gamma^{-1}, F_2 \circ s_\gamma^{-1}).$$

Proposition 9 (Domain of the adapted metric and behavior near zero). *Fix $p \geq 1$ and $\gamma \in \mathbb{R}$. For probability measures F_1, F_2 on $(0, \infty)$, the distance $d_{\gamma,p}(F_1, F_2)$ is finite if and only if*

$$\int |s_\gamma(x)|^p F_i(dx) < \infty, \quad i \in \{1, 2\}.$$

Equivalently:

1. if $\gamma > 0$, one needs $\int x^{\gamma p} F_i(dx) < \infty$ for $i = 1, 2$;

2. if $\gamma = 0$, one needs $\int |\log x|^p F_i(dx) < \infty$ for $i = 1, 2$;
3. if $\gamma < 0$, one needs $\int x^{-|\gamma|p} F_i(dx) < \infty$ for $i = 1, 2$.

In particular, mass near zero is harmless for $\gamma > 0$ but can make $d_{\gamma,p}$ infinite for $\gamma \leq 0$ even when an ordinary Wasserstein distance on the raw type space is finite.

Remark 11 (Low-access types and the limits of adapted geometry). The transport theorem therefore remains available for very small access types as long as the relevant transformed moments are finite. For $\gamma = 0$ this is a logarithmic integrability requirement; for $\gamma < 0$ it is an inverse-moment requirement of order $|\gamma|p$. If an application permits heavier mass near zero than those conditions allow, the adapted metric is no longer the right comparison tool. One should then revert to the Laplace-order statements, the raw space mean-preserving geodesics of Section 3, or the finite horizon and pointwise bounds in this section, none of which require $d_{\gamma,p} < \infty$.

We now state the main stability bound. The constants are explicit and reflect moment restrictions intrinsic to extreme-value tails.

Theorem 1 (Wasserstein stability of $F \mapsto H_{\gamma,F}$). *Fix $\gamma \in \mathbb{R}$ and $p \geq 1$. Let F_1, F_2 be probability measures on $(0, \infty)$ such that $d_{\gamma,p}(F_1, F_2) < \infty$.*

If $\gamma > 0$, assume in addition that $p\gamma < 1$. If $\gamma \neq 0$, let $E \sim \text{Exp}(1)$ and define $\Xi_\gamma := E^{-\gamma}$. Then

$$W_p(H_{\gamma,F_1}, H_{\gamma,F_2}) \leq \frac{(\mathbb{E}[|\Xi_\gamma|^p])^{1/p}}{|\gamma|} d_{\gamma,p}(F_1, F_2).$$

If $\gamma = 0$, then

$$W_p(H_{0,F_1}, H_{0,F_2}) \leq d_{0,p}(F_1, F_2).$$

Remark 12 (What the transport layer adds). The transform s_γ by itself is only a reparameterization of types. The substantive content of Theorem 1 is that, under the canonical coupling $Z_{\gamma,F} = w_\gamma(E/X)$, optimal transport on $s_\gamma(X)$ yields an explicit Wasserstein bound on the entire induced extreme law, with constant determined by the tail index through $\mathbb{E}[E^{-p\gamma}]$ in the Fréchet regime. Corollary 2 then propagates this to canonical ambient paths in law space. These quantitative whole law statements do not follow from the Laplace mixture alone.

Remark 13 (Moment restrictions are intrinsic). For $\gamma > 0$ one has $\mathbb{E}[E^{-p\gamma}] = \Gamma(1 - p\gamma)$, so $\mathbb{E}[|\Xi_\gamma|^p] < \infty$ holds if and only if $p\gamma < 1$. For $\gamma < 0$, by contrast, $\Xi_\gamma = E^{|\gamma|}$ has finite moments of every order. The heavy-tail restriction is therefore not an artifact of the coupling argument. It reflects that the Fréchet-type limit law has finite p th moment only when $p < 1/\gamma$, and therefore $W_p(H_{\gamma,F_1}, H_{\gamma,F_2})$ is only defined in the usual Wasserstein sense in that regime.

Theorem 1 is proved by applying the random Lipschitz contraction Lemma 9 to the representation in Proposition 8, using a coupling of $s_\gamma(X_1)$ and $s_\gamma(X_2)$ and a common exponential variable. The construction is in the same spirit as the coupling arguments emphasized by Bobbia et al. (2021), but here the comparison is between heterogeneous extreme laws generated by different opportunity distributions.

The theorem becomes most useful once the adapted metric is treated as a geometry in its own right. Because $d_{\gamma,p}$ is an ordinary Wasserstein distance after the transform s_γ , it automatically induces canonical constant-speed geodesics in the ambient transformed type space. Along those paths, the entire law of extremes moves in a Lipschitz way. This is the main result in the paper that specifically requires the Wasserstein geometry rather than only the Laplace form.

Corollary 2 (Geodesic control in the adapted type geometry). *Fix $\gamma \in \mathbb{R}$ and $p \geq 1$. Let F^0, F^1 be two probability measures on $(0, \infty)$ for which the assumptions of Theorem 1 hold. Let $G^0 := s_\gamma \# F^0$ and $G^1 := s_\gamma \# F^1$, and let $(G^t)_{t \in [0,1]}$ be the canonical monotone constant-speed W_p geodesic from G^0 to G^1 . Define $F^t := (s_\gamma^{-1}) \# G^t$. Then, for all $s, t \in [0, 1]$,*

$$d_{\gamma,p}(F^s, F^t) = |t - s| d_{\gamma,p}(F^0, F^1).$$

Moreover,

$$W_p(H_{\gamma,F^s}, H_{\gamma,F^t}) \leq C_{\gamma,p} |t - s| d_{\gamma,p}(F^0, F^1),$$

where $C_{\gamma,p}$ denotes the constant from Theorem 1.

Remark 14 (Ambient geodesics versus the mean one normalization). The path in Corollary 2 is a geodesic in the ambient transformed type space, not necessarily in the mean one slice used in the economic model. Indeed, $Q_{G^t}(u)$ is affine in t in transformed coordinates, but $\int s_\gamma^{-1}(Q_{G^t}(u)) du$ is generally not constant unless s_γ^{-1} is affine, which occurs only when $\gamma = 1$. Thus the adapted path should be read as a geometric comparison device, not as a literal mean-preserving policy path. If one instead wants exact mean preservation at every step, the raw space geodesics of Section 3.5 remain available, but they need not inherit the constant-speed statement for $d_{\gamma,p}$.

Proposition 10 (Renormalization bridge back to the mean one slice). *Fix $\gamma \in \mathbb{R}$ and $p \geq 1$. Let F be a probability measure on $(0, \infty)$ with finite mean $m(F) := \int x F(dx) \in (0, \infty)$ and finite transformed p th moment. Define the mean one renormalization*

$$\tilde{F} := (x \mapsto x/m(F)) \# F.$$

Then

$$d_{\gamma,p}(F, \tilde{F}) = \begin{cases} |1 - m(F)^{-\gamma}| (\int x^{\gamma p} F(dx))^{1/p}, & \gamma \neq 0, \\ |\log m(F)|, & \gamma = 0. \end{cases}$$

If, in addition, the assumptions of Theorem 1 hold for F and \tilde{F} , then

$$W_p(H_{\gamma,F}, H_{\gamma,\tilde{F}}) \leq C_{\gamma,p} d_{\gamma,p}(F, \tilde{F}),$$

where $C_{\gamma,p}$ is the modulus from Theorem 1. In particular, if $m_t := \int x F^t(dx)$ and $\tilde{F}^t := (x \mapsto x/m_t) \# F^t$ along the ambient geodesic from Corollary 2, the same bound applies pointwise in t .

Corollary 3 (Quantile-schedule control of the induced extreme law). *Under the assumptions of Theorem 1, let $Q_{\gamma,F_i} : (0, 1) \rightarrow \mathbb{R}$ denote the quantile function of the law H_{γ,F_i} for $i \in \{1, 2\}$. Then*

$$\left(\int_0^1 |Q_{\gamma,F_1}(u) - Q_{\gamma,F_2}(u)|^p du \right)^{1/p} = W_p(H_{\gamma,F_1}, H_{\gamma,F_2}) \leq C_{\gamma,p} d_{\gamma,p}(F_1, F_2).$$

Along the adapted geodesic from Corollary 2, this yields

$$\|Q_{\gamma,F^s} - Q_{\gamma,F^t}\|_{L^p(0,1)} \leq C_{\gamma,p} |t - s| d_{\gamma,p}(F^0, F^1).$$

An immediate benchmark implication is obtained by comparing F to δ_1 . Whenever the conditions of Theorem 1 hold,

$$W_p(H_{\gamma,F}, H_\gamma) \leq C_{\gamma,p} d_{\gamma,p}(F, \delta_1),$$

where $C_{\gamma,p}$ denotes the constant from Theorem 1. Combined with Proposition 11 in the regimes covered there, this translates metric misallocation on the type space into an explicit distortion bound for extreme outcomes relative to the homogeneous benchmark.

Corollary 4 (Pointwise stability of the cdf). *Fix $\gamma \in \mathbb{R}$ and let $F_1, F_2 \in \mathcal{P}_1([0, \infty))$. Then for every x in the interior of the support of H_γ ,*

$$|H_{\gamma, F_1}(x) - H_{\gamma, F_2}(x)| \leq v_\gamma(x) W_1(F_1, F_2).$$

Remark 15 (Pointwise versus geometric bounds). Corollary 4 follows directly from the Laplace representation $H_{\gamma, F}(x) = \int e^{-v_\gamma(x)z} F(dz)$ and the fact that $z \mapsto e^{-v_\gamma(x)z}$ is $v_\gamma(x)$ -Lipschitz on $[0, \infty)$. It is therefore not itself a geometric theorem. Theorem 1, Proposition 10, and Corollary 3 are the results that genuinely use the adapted Wasserstein geometry. Those statements control whole induced laws, the bridge back to the mean one slice, and the full quantile schedule. They are not recovered from fixed threshold cdf bounds without additional density information.

In many applications one also wants conditions under which $d_{\gamma, p}$ is controlled by a standard Wasserstein distance on the type space. We record sufficient conditions that isolate the role of behavior near 0.

Proposition 11 (Relating adapted type metrics to standard Wasserstein distances). *Let $p \geq 1$ and let $F_1, F_2 \in \mathcal{P}_p([0, \infty))$.*

Case 1: $0 < \gamma \leq 1$. *Then*

$$d_{\gamma, p}(F_1, F_2) \leq W_p(F_1, F_2)^\gamma.$$

Case 2: $\gamma < 0$. *Assume F_1 and F_2 are supported on $[a, \infty)$ for some $a > 0$. Then*

$$d_{\gamma, p}(F_1, F_2) \leq |\gamma| a^{\gamma-1} W_p(F_1, F_2).$$

Case 3: $\gamma = 0$. *Assume F_1 and F_2 are supported on $[a, \infty)$ for some $a > 0$. Then*

$$d_{0, p}(F_1, F_2) \leq a^{-1} W_p(F_1, F_2).$$

Remark 16 (Why we keep $d_{\gamma, p}$ in the main bound). In the positive-tail regime covered by Theorem 1, namely $0 < \gamma < 1/p$, the map $x \mapsto x^\gamma$ is not globally Lipschitz on $[0, \infty)$. For $0 < \gamma < 1$, the obstruction is behavior near zero. Consequently, a linear stability bound stated directly in terms of $W_p(F_1, F_2)$ would either require extra support restrictions or lose the natural linear geometry of the canonical coupling. The adapted metric $d_{\gamma, p}$ is therefore the correct primitive quantity in the main bound, while Proposition 11 records conditions under which it can be controlled by a standard Wasserstein distance.

4.4 Stability of extreme outcome functionals

The Wasserstein bounds above immediately imply stability of a large class of tail-sensitive functionals.

Proposition 12 (Lipschitz functionals of the limit law). *Let $\psi : \mathbb{R} \rightarrow \mathbb{R}$ be L -Lipschitz. Let $Z_i \sim H_{\gamma, F_i}$ for $i \in \{1, 2\}$ and assume $W_1(H_{\gamma, F_1}, H_{\gamma, F_2}) < \infty$. Then*

$$|\mathbb{E}[\psi(Z_1)] - \mathbb{E}[\psi(Z_2)]| \leq L W_1(H_{\gamma, F_1}, H_{\gamma, F_2}).$$

Combining with Theorem 1 yields explicit bounds in terms of $d_{\gamma, p}(F_1, F_2)$ whenever the conditions of that theorem hold.

4.5 Finite horizon maxima and tail probability stability

Fix $\theta > 0$ and let M_θ be the maximum of N_θ i.i.d. draws from G , where $N_\theta | X \sim \text{Poisson}(\theta X)$ as in Section 2. Conditional on X , one has

$$\Pr(M_\theta \leq x | X) = \exp(-\theta X(1 - G(x))),$$

and therefore $\Pr(M_\theta \leq x) = \mathbb{E}[\exp(-\theta X(1 - G(x)))]$.

Proposition 13 (Pointwise stability for finite θ). *Let $F_1, F_2 \in \mathcal{P}_1([0, \infty))$. Then for every $x \in \mathbb{R}$,*

$$\left| \Pr_{F_1}(M_\theta \leq x) - \Pr_{F_2}(M_\theta \leq x) \right| \leq \theta(1 - G(x)) W_1(F_1, F_2).$$

Remark 17 (Finite-horizon versus asymptotic robustness). Proposition 13 is a finite horizon robustness statement. It does not use the extreme-value limit. By contrast, Theorem 2 below is an asymptotic expansion around the limit law. Keeping these two layers separate is important in empirical work: finite-sample prediction error and EVT approximation error are different objects.

4.6 Rates under second-order tail conditions

We conclude with a second-order expansion for the prelimit cdf of normalized maxima. The proof is given in Appendix A. In this subsection, unlike the first-order limit theorem in Section 2.4, we work with continuous normalizing functions $a : (1, \infty) \rightarrow (0, \infty)$ and $b : (1, \infty) \rightarrow \mathbb{R}$. This is the standard continuous-parameter formulation in extreme value theory and avoids the artificial $O(1/\theta)$ oscillation created by the step extension $a_{\lfloor \theta \rfloor}, b_{\lfloor \theta \rfloor}$. Assumption 3 is a standard second-order regularity condition; see, for example, de Haan and Ferreira (2006, Chapters 2.10 and 9).

Assumption 3 (Second-order tail condition). In addition to Assumption 1, suppose there exist continuous normalizing functions $a : (1, \infty) \rightarrow (0, \infty)$ and $b : (1, \infty) \rightarrow \mathbb{R}$, a scalar function $A(\theta) \rightarrow 0$, and a locally bounded function h_γ such that

$$\theta(1 - G(b(\theta) + a(\theta)x)) = v_\gamma(x) + A(\theta) h_\gamma(x) + o(A(\theta))$$

as $\theta \rightarrow \infty$, locally uniformly in x on the interior of the support of H_γ .

Theorem 2 (Second-order expansion for heterogeneous extremes). *Suppose Assumptions 2 and 3 hold and let*

$$Z_\theta := \frac{M_\theta - b(\theta)}{a(\theta)}.$$

Fix a compact set K contained in the interior of the support of H_γ . Then there exists a remainder $r_{\theta,K}$ on K such that

$$\sup_{x \in K} \frac{|r_{\theta,K}(x)|}{|A(\theta)|} \rightarrow 0 \quad \text{as } \theta \rightarrow \infty,$$

and, uniformly for $x \in K$,

$$\Pr(Z_\theta \leq x) = H_{\gamma,F}(x) - A(\theta) h_\gamma(x) \mathbb{E}[X e^{-v_\gamma(x)X}] + r_{\theta,K}(x).$$

Remark 18 (Interpretation). The leading correction term is multiplicatively separable. The function h_γ captures the second-order tail behavior of G , while $x \mapsto \mathbb{E}[X e^{-v_\gamma(x)X}]$ is the heterogeneity kernel inherited from the Laplace mixture. This separation is what allows one to distinguish EVT approximation error from heterogeneity-driven sensitivity. Converting the expansion into global W_1 rates require additional tail-envelope conditions and we therefore left outside the present theorem.

5 A complementary entropy projection for heterogeneity

In this section we study a tractable normative problem on the same type space used in the positive analysis. A planner reallocates opportunities across agents to improve a tail-sensitive criterion while penalizing deviations from a baseline distribution by relative entropy. The object is deliberately modest: it is a one-marginal entropy projection, not a full two-marginal transport problem, and it does not rely on the Wasserstein geometry developed earlier. It is useful in our framework, but the connection should be described accurately: this section is complementary rather than foundational. The structural link to the positive analysis is that the same Laplace kernel that defines $F \mapsto H_{\gamma,F}$ also generates the marginal score functions that enter the planner's first-order conditions for cdf criteria and, through the stochastic representation below, for expected utility of normalized extremes.

5.1 A variational planner problem

Fix a baseline distribution F_0 on $(0, \infty)$ such that $\mathbb{E}_{F_0}[X] = 1$. We interpret F_0 as the status quo distribution of effective search intensities, network degrees, or other shifters of arrival rates. A policy induces a new distribution F , and we write $X \sim F$.

For $F \ll F_0$, define the relative entropy

$$D_{\text{KL}}(F \| F_0) := \int \log \left(\frac{dF}{dF_0} \right) dF,$$

with the convention $D_{\text{KL}}(F \| F_0) = +\infty$ if F is not absolutely continuous with respect to F_0 .

Let $H_{\gamma,F}$ denote the heterogeneous extreme-value law from Proposition 1. We focus on tail-sensitive objectives of the form

$$U(F) = \mathbb{E}[u(Z)], \quad Z \sim H_{\gamma,F},$$

for a measurable function u such that the expectation is well defined under the candidate distributions considered below. This class includes expected utility of normalized extremes and many smoothed tail criteria. We comment on high-quantile objectives at the end of the section.

The planner solves

$$\sup_{F: \mathbb{E}_F[X]=1} \left\{ U(F) - \lambda D_{\text{KL}}(F \| F_0) \right\}, \quad (3)$$

where $\lambda > 0$ governs the marginal cost of reshaping opportunities away from the baseline.

Remark 19 (Interpretation). The constraint $\mathbb{E}_F[X] = 1$ normalizes the aggregate scale of opportunities. In a network interpretation it fixes total expected offer arrival in the population and allows the planner to reallocate the cross-sectional distribution. The entropy term penalizes concentrated reallocations and yields a strictly concave problem whenever U is linear in F , which is the main case treated below.

Remark 20 (Normalized welfare and prelimit maxima). The criterion $U(F) = \mathbb{E}[u(Z)]$ should be read as the asymptotic counterpart of prelimit preferences over maxima. Indeed, if u is bounded and continuous, then Proposition 1 implies

$$\mathbb{E} \left[u \left(\frac{M_\theta - b_\theta}{a_\theta} \right) \right] \rightarrow U(F) \quad \text{as } \theta \rightarrow \infty.$$

Equivalently, the planner may be viewed as evaluating a sequence of prelimit utilities $u_\theta(m) := u((m - b_\theta)/a_\theta)$ that preserves the affine normalization used by extreme-value theory.

5.2 A stochastic representation and linearization of expected utility

The canonical representation from Section 4.2 immediately linearizes expected-utility objectives in the heterogeneity distribution. Let $w_\gamma := v_\gamma^{-1}$ denote the inverse map introduced there.

Lemma 4 (Linearization of expected utility). *Let $X \sim F$ and let $E \sim \text{Exp}(1)$ be independent. Define*

$$Z := w_\gamma\left(\frac{E}{X}\right),$$

so that $Z \sim H_{\gamma,F}$ by Proposition 8. If $u : \mathbb{R} \rightarrow \mathbb{R}$ is measurable and $\mathbb{E}[|u(Z)|] < \infty$, then

$$U(F) = \mathbb{E}[u(Z)] = \int \psi_u(x) F(dx), \quad \psi_u(x) := \mathbb{E}\left[u\left(w_\gamma\left(\frac{E}{x}\right)\right)\right].$$

Proof. The distributional representation of Z is Proposition 8. The formula for $U(F)$ then follows by iterated expectation. \square

Remark 21 (Fréchet specialization). When $\gamma > 0$, Proposition 8 reduces to the product representation from Section 2.5: if $\Xi_\gamma := E^{-\gamma}$, then

$$1 + \gamma Z = X^\gamma \Xi_\gamma.$$

5.3 Duality and exponential tilting

We now solve (3) in the canonical case where the objective is linear in F . This case is both the baseline design problem and the relevant reduction for expected-utility objectives by Lemma 4. The exponential-tilt form itself is the standard entropy-projection conclusion. What is specific here is that the heterogeneous-EVT operator supplies the score ψ and reduces the mean constraint to a one-dimensional dual variable. The duality argument is recorded in Appendix B.2; see also Donsker and Varadhan (1975), Léonard (2014), Ghosal et al. (2022), and Nutz (2022), for background on the broader Schrödinger and entropic-transport literature.

Assumption 4 (Linear objective, local integrability, and interiority). There exists a measurable function $\psi : (0, \infty) \rightarrow \mathbb{R}$ such that

$$U(F) = \int \psi(x) F(dx) \quad \text{for all } F \text{ with } \mathbb{E}_F[X] = 1 \text{ and } D_{\text{KL}}(F||F_0) < \infty.$$

For $\eta \in \mathbb{R}$, define

$$Z(\eta) := \int \exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) F_0(dx)$$

and, for $k \in \{1, 2\}$,

$$M_k(\eta) := \int x^k \exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) F_0(dx).$$

Assume there exists a nonempty open interval $\mathcal{D} \subset \mathbb{R}$ such that, for all $\eta \in \mathcal{D}$,

$$Z(\eta) < \infty, \quad M_1(\eta) < \infty, \quad M_2(\eta) < \infty, \quad \int |\psi(x)| \exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) F_0(dx) < \infty.$$

Define

$$m(\eta) := \frac{M_1(\eta)}{Z(\eta)}.$$

Assume moreover that there exist $\eta_-, \eta_+ \in \mathcal{D}$ with

$$m(\eta_-) < 1 < m(\eta_+).$$

Theorem 3 (Strong duality and exponential tilt). *Suppose Assumption 4 holds. Then there exists a unique scalar $\eta^* \in \mathcal{D}$ satisfying $m(\eta^*) = 1$. The design problem (3) admits a unique optimizer F^* , and F^* is given by the exponential tilt*

$$\frac{dF^*}{dF_0}(x) = \frac{\exp\left(\frac{\psi(x) + \eta^* x}{\lambda}\right)}{\int \exp\left(\frac{\psi(t) + \eta^* t}{\lambda}\right) F_0(dt)}. \quad (4)$$

Moreover, $\mathbb{E}_{F^*}[X] = 1$.

The optimal value of (3) admits the dual representation

$$\sup_{F: \mathbb{E}_F[X]=1} \left\{ U(F) - \lambda D_{\text{KL}}(F \| F_0) \right\} = \inf_{\eta \in \mathcal{D}} \left\{ \lambda \log \int \exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) F_0(dx) - \eta \right\}. \quad (5)$$

Comment. Equation (4) is the key structural conclusion of the section. Relative to F_0 , the planner reweights types with higher score $\psi(x)$, with the magnitude of reweighting disciplined by λ . The scalar η^* is the Lagrange multiplier on the mean one constraint and is pinned down by a single scalar equation $m(\eta) = 1$. With the sign convention in (5), the shadow value of relaxing the target mean upward is $-\eta^*$. The novelty here is not the Gibbs form by itself, but the identification of the heterogeneous-EVT score ψ and the explicit one-dimensional implementation of the mean constraint. The interiority assumption in Assumption 4 is the precise condition that guarantees existence of a finite dual optimizer. In many applications it is verified through a standard steepness argument for the tilted mean $m(\eta)$.

A complete proof of Theorem 3 is given in Appendix A. It applies the entropy-duality tools in Appendix B.2, but the proof is written directly for the one-dimensional mean constraint used here so that the existence of the dual optimizer is explicit rather than implicit.

5.4 First-order conditions and the link to the heterogeneous EVT kernel

The design problem inherits its structure from the heterogeneous extreme-value operator

$$H_{\gamma, F}(x) = \int e^{-v_\gamma(x)z} F(dz).$$

Section 3.4 records the corresponding directional derivative,

$$\delta H_{\gamma, F}(x) = \int e^{-v_\gamma(x)z} \delta F(dz),$$

which makes explicit that the Laplace kernel $z \mapsto e^{-v_\gamma(x)z}$ is the marginal channel through which perturbations of F affect the distribution of extremes.

In the linear class treated in Theorem 3, the score function ψ is the marginal value of shifting mass toward type x . For expected-utility objectives $U(F) = \mathbb{E}[u(Z)]$, Lemma 4 identifies this marginal value explicitly as $\psi_u(x) = \mathbb{E}[u(w_\gamma(E/x))]$. This is the direct bridge to Section 4: the same stochastic representation that yields linearity for design also provides the canonical coupling device for stability.

Remark 22 (Direct cdf criteria use the EVT kernel literally). For any fixed threshold y , the criterion

$$U_y(F) := H_{\gamma, F}(y)$$

is already linear in F , with score

$$\psi_y(x) = e^{-v_\gamma(y)x}.$$

More generally, any weighted average of cdf levels of the form $\int H_{\gamma,F}(y) \nu(dy)$ has score $x \mapsto \int e^{-v_\gamma(y)x} \nu(dy)$ whenever the integral is finite. This is the cleanest sense in which the same kernel governs both the heterogeneous extreme-value operator and a nontrivial class of normative objectives.

5.5 Closed-form solutions for canonical tail objectives

Closed-form solutions are especially transparent when ψ is a simple transform of x . This class is relevant economically because several extreme outcome statistics in heterogeneous EVT reduces to moments or inverse moments of X under Fréchet-type normalization, as in Mangin (2026). The formulas below should be read together with their admissibility conditions: the exponential tilt is meaningful only when Assumption 4 is satisfied.

Power and inverse-power objectives. Let $\rho > 0$ and consider objectives of the form

$$U(F) = C \cdot \mathbb{E}_F[X^{-\rho}] \quad \text{or} \quad U(F) = C \cdot \mathbb{E}_F[X^\rho],$$

for a constant $C \in \mathbb{R}$. These correspond to $\psi(x) = Cx^{-\rho}$ or $\psi(x) = Cx^\rho$ in Assumption 4. When the relevant integrability conditions hold, Theorem 3 yields

$$\frac{dF^*}{dF_0}(x) \propto \exp\left(\frac{Cx^{-\rho} + \eta^*x}{\lambda}\right) \quad \text{or} \quad \frac{dF^*}{dF_0}(x) \propto \exp\left(\frac{Cx^\rho + \eta^*x}{\lambda}\right),$$

with η^* chosen so that $\mathbb{E}_{F^*}[X] = 1$.

Admissibility conditions. For inverse-power scores $x \mapsto Cx^{-\rho}$, a sufficient condition is that the baseline support be bounded away from zero, say $\text{supp}(F_0) \subseteq [a, \infty)$ with $a > 0$; without such a condition, the case of the positive coefficient $C > 0$ typically makes the normalizing integral diverge near zero. For positive-power scores $x \mapsto Cx^\rho$, admissibility depends on the right tail of F_0 . A sufficient condition is that F_0 be light enough so that

$$\int \exp\left(\frac{Cx^\rho + \eta x}{\lambda}\right) F_0(dx) < \infty$$

for all η in some open interval containing the optimizer. When $\rho > 1$ and $C > 0$, this excludes exponential or heavier baseline tails. These support and tail restrictions are not technical decoration; they are exactly what ensures the closed-form tilt is well defined.

Implementation. In applications, η^* is obtained by solving the scalar equation $m(\eta) = 1$. The value function is then obtained from the dual expression (5). The entire computation is one-dimensional once the score ψ is known.

5.6 Interpretation: policy as reallocation of opportunities

The entropy penalty $D_{\text{KL}}(F||F_0)$ imposes a reduced-form cost of reallocating opportunities across types. In a labor market network interpretation, F can for instance represent the distribution of effective meeting rates or centralities, and policy changes referral frictions, search subsidies, or platform rules that alter the induced F . The entropy term summarizes implementation frictions that render highly concentrated reallocations costly. In richer environments, one can interpret

the exponential tilt as the reduced-form allocation induced by subsidy, tax, or referral policies that change the private return to search intensity. It is beyond the scope of this paper to study decentralized implementation, but we note that the one-dimensional form of (4) makes that question especially tractable.

A natural extension, which we keep separate from the core analysis to preserve the one-dimensional tractability of (3), replaces the marginal penalty by a genuine entropic transport problem on a richer type space with cost $c(x_0, x)$. That extension would produce a pair of dual potentials on the source and target type spaces. The present section should therefore be read as an entropy projection on the target marginal, not as a full Schrödinger bridge.

Quantile objectives. If the objective is a high quantile, $U(F) = Q_{1-\alpha}(H_{\gamma,F})$, then U is typically not linear in F . Two tractable approaches are to work with smoothed tail criteria or tail expectations that admit linear representations via Lemma 4, or to characterize optimizers through first-order conditions using the directional derivative of $F \mapsto H_{\gamma,F}$ from Section 3.4. We implement the linear class in the core text, leaving exact quantile design as a promising extension.

6 Application: labor market networks and the distribution of top wages

This section illustrates how heterogeneous extreme-value theory can be used as a reduced-form device in labor market settings where access to job opportunities is mediated by social networks. The key object is the cross-sectional distribution F of effective offer-arrival intensities induced by heterogeneity in network position, referrals, and related opportunity channels. Once F is specified or measured, the general results in Sections 3, 4, and 5 translate into three types of statements: (i) comparative statics for the right tail of wages as a function of network inequality, (ii) robustness bounds for tail predictions under measurement or estimation error in F , and (iii) an entropy-regularized policy problem that reallocates opportunities subject to implementation frictions.

While the broader labor network motivation was discussed in Section 1.4, here our goal is to translate the general results into a familiar reduced-form environment for tail wage outcomes. Throughout, the object remains the cross-sectional law of a randomly drawn worker’s normalized maximum under a common offer distribution. Remark that we do not seek to model herein equilibrium wage setting, endogenous search, or welfare. Accordingly, statements below about segregation, homophily, or policy should be read as claims about how those forces reshape the reduced-form access distribution F , not as standalone equilibrium or welfare conclusions.

6.1 A stylized network-based search environment

Types and offer arrivals. Consider a large population of workers indexed by i . Each worker has a network position summarized by a scalar type $X_i > 0$ capturing effective access to job opportunities. The interpretation of X_i is as a shifter of job offer arrival rates, generated by a referral network, local contacts, or platform-mediated matching. We impose the normalization $\mathbb{E}[X_i] = 1$, so the average scale of opportunities is fixed and the object of interest is the cross-sectional distribution of access.

Fix a market thickness or horizon parameter $\theta > 0$. Conditional on X_i , the number of offers received by worker i follows the mixed-Poisson specification from Section 2:

$$N_{i,\theta} \mid X_i \sim \text{Poisson}(\theta X_i).$$

Let F denote the cross-sectional distribution of X_i in the population, so $X \sim F$ for a randomly drawn worker.

Offer values and realized wages. Let $\{Y_{ij}\}_{j \geq 1}$ be i.i.d. wage offers with common distribution G , independent of $(X_i, N_{i,\theta})$. Worker i accepts the best offer over the horizon,

$$M_{i,\theta} := \sup_{1 \leq j \leq N_{i,\theta}} Y_{ij},$$

with the convention $\sup \emptyset = -\infty$. The economic content is that heterogeneity enters through access to opportunities, while the distribution of offer values is common across workers.

From network primitives to F . The distribution F can be linked to standard network objects. As a benchmark, suppose workers are nodes in an undirected contact network and each neighbor generates job leads at approximately independent random times. Conditional on degree D_i , the number of leads over a horizon is then well approximated by a Poisson random variable with mean proportional to θD_i . After normalization, one can set $X_i := D_i/\mathbb{E}[D]$, so that $\mathbb{E}[X_i] = 1$ and F is the distribution of normalized degrees. With weighted links, one can take X_i as normalized weighted degree. If the arrival mechanism aggregates information over longer paths, X_i can represent a normalized centrality index. The analysis that follows uses F as the reduced-form sufficient statistic for the cross-sectional opportunity structure.

Tail normalization and the heterogeneous extreme-value law. Assume G is in the max domain of attraction of the generalized extreme-value law H_γ with index γ . Under the normalization in Section 2.1, the distribution of normalized outcomes for a randomly drawn worker converges to the heterogeneous extreme-value law from Proposition 1:

$$H_{\gamma,F}(x) = \mathbb{E}[\exp(-X v_\gamma(x))], \quad X \sim F.$$

Thus, once F is specified or estimated, the model delivers a tractable mapping from network heterogeneity into the cross-sectional distribution of workers' top wages in the sense of maxima over offers.

6.2 Positive implications: network heterogeneity and the right tail

Network inequality and tail losses under convex order. A natural notion of increased network inequality is a mean-preserving spread of F holding $\mathbb{E}[X] = 1$ fixed, equivalently an increase in the convex order; see, e.g., Rothschild and Stiglitz (1970) and Shaked and Shanthikumar (2007). In the present setting, this order has a direct implication for extremes because, for each $z > 0$, the kernel $x \mapsto \exp(-zx)$ is convex and decreasing. This is precisely the mechanism behind the Jensen and Laplace-ordering results in Section 3.1.

Corollary 5 (Network inequality and normalized top wages). *Fix γ and let F_1, F_2 satisfy $\mathbb{E}[X] = 1$. If F_2 is a mean-preserving spread of F_1 , then*

$$H_{\gamma,F_2}(x) \geq H_{\gamma,F_1}(x) \quad \text{for all } x,$$

so the corresponding limit distribution of normalized top wages under F_2 is first-order stochastically smaller than under F_1 . Equivalently, for any increasing measurable φ for which expectations are finite,

$$\mathbb{E}[\varphi(Z_2)] \leq \mathbb{E}[\varphi(Z_1)], \quad Z_k \sim H_{\gamma,F_k}.$$

Corollary 5 isolates a clean reduced-form message about the right tail. Holding fixed the mean scale of opportunities and the common offer distribution, redistributing access more unequally lowers the distribution of maxima for a randomly drawn worker because the Laplace kernel is convex in types. In this sense, homophily, segregation, or referral frictions matter here through the extent to which they reshape the opportunity distribution F . At the same time, one should not overread the result at the furthest tail. By Remark 5, the mean one normalization implies $1 - H_{\gamma,F}(x) \sim v_\gamma(x) \sim 1 - H_\gamma(x)$ whenever $v_\gamma(x) \downarrow 0$, so greater network inequality changes distributional levels and high quantiles, but not the first-order ultra-tail mass. Translating that comparative static into welfare, equilibrium efficiency, or group incidence would require additional structure on search behavior, offer-quality heterogeneity, and market clearing.

A concrete rewiring counterfactual. Consider a referral-platform redesign, mentoring intervention, or local rewiring experiment that changes the empirical distribution of normalized degree from F^0 to F^1 while preserving the population mean. If F^1 is a mean-preserving contraction of F^0 , Corollary 5 implies an everywhere improvement in the normalized top wage distribution for a randomly drawn worker. If the change is not ordered by convex order, the adapted geometry below still provide a disciplined whole law comparison.

A simple numerical illustration. Consider a baseline network distribution

$$F_0 = 0.8 \delta_{0.5} + 0.2 \delta_3,$$

which can be read as a labor market with many peripheral workers and a smaller highly connected group, while preserving the normalization $\mathbb{E}[X] = 1$. At the centered threshold $x = 0$, one has $v_\gamma(0) = 1$ for every γ , so

$$H_{\gamma,F_0}(0) = P_0(1) = 0.8e^{-0.5} + 0.2e^{-3} \approx 0.495, \quad H_\gamma(0) = e^{-1} \approx 0.368.$$

Thus the unequal-opportunity economy places about 12.7 percentage points more mass below the centered threshold, equivalently about 12.7 percentage points less mass above it, than the homogeneous benchmark. The corresponding misallocation indices are $\mathcal{M}_1(F_0) = \mathbb{E}[|X - 1|] = 0.8$ and $\mathcal{M}_2(F_0) = \sqrt{\text{Var}(X)} = 1$. This simple example makes concrete how a small, highly connected minority can coexist with a materially weaker distribution of normalized top wages for a randomly drawn worker.

Geometric counterfactual paths. When two economies with type distributions F^0 and F^1 are not ordered by convex order, the adapted-Wasserstein geometry can still provide, quite remarkably, a disciplined comparison. Corollary 2 constructs a canonical interpolation path $(F^t)_{t \in [0,1]}$ in the ambient transformed type geometry such that the induced law of top wages moves at most linearly in path length. Corollary 3 turns the same result into an economically direct statement for the entire counterfactual schedule of top wage quantiles:

$$\|Q_{\gamma,F^s} - Q_{\gamma,F^t}\|_{L^p(0,1)} \leq C_{\gamma,p} |t - s| d_{\gamma,p}(F^0, F^1).$$

Thus the adapted metric is not only a pointwise cdf device. It controls the whole wage distribution in quantile space, which is the natural object for many counterfactual exercises. Because transformed coordinate geodesics need not preserve $\mathbb{E}[X] = 1$ at intermediate times when s_γ is nonlinear, this path should be read as a sensitivity device in ambient shape space, not as a literal mean-preserving policy path. If a counterfactual must keep aggregate opportunity scale fixed throughout, Proposition 10 quantifies the discrepancy between the ambient path and its mean one renormalization, while the raw space mean-preserving geodesics of Section 3 remain available when exact preservation is required throughout.

Robust tail prediction under measurement error in F . In practice, F can be estimated from network data or inferred from proxies for the intensity of job contact. The Wasserstein stability results in Section 4 then provide a direct way to translate estimation error in F into error bounds for predicted tail outcomes. The main linear bound is stated in the adapted metric $d_{\gamma,p}$. When the available statistical control is instead in raw space Wasserstein distance, Proposition 11 provides the additional bridge, which can be weaker in the Fréchet regime.

Corollary 6 (Robustness of predicted tail distributions). *Let \widehat{F} be an estimator of F with $\mathbb{E}[X] = 1$. Assume the conditions of Theorem 1 hold for some $p \geq 1$. Then the induced error on the predicted distribution of normalized top wages satisfies*

$$W_p\left(H_{\gamma,\widehat{F}}, H_{\gamma,F}\right) \leq C_{\gamma,p} d_{\gamma,p}\left(\widehat{F}, F\right),$$

for the modulus $C_{\gamma,p}$ characterized in Section 4. Under the support and tail-shape conditions of Proposition 11, the adapted distance on the right-hand side can in turn be controlled by a standard Wasserstein distance on the type space. Moreover, if ψ is Lipschitz on the support of $H_{\gamma,F}$ and the relevant moments exist, then

$$\left|\mathbb{E}[\psi(Z_{\widehat{F}})] - \mathbb{E}[\psi(Z_F)]\right| \leq \text{Lip}(\psi) W_1\left(H_{\gamma,\widehat{F}}, H_{\gamma,F}\right),$$

where $Z_{\widehat{F}} \sim H_{\gamma,\widehat{F}}$ and $Z_F \sim H_{\gamma,F}$.

Corollary 6 is the operational bridge between statistical or measurement uncertainty in network heterogeneity and uncertainty in predicted tail-wage outcomes, but only after the maintained comparison metric has been specified clearly. If the researcher controls $d_{\gamma,p}$ directly, the bound is linear. If the researcher begins from raw space Wasserstein error, the relevant conclusion is obtained only after applying Proposition 11, and in Fréchet settings that bridge can be merely Hölder. For peripheral or intermittently inactive workers, the relevant issue is behavior near $X = 0$. Proposition 9 makes the boundary explicit: the adapted transport bounds remain available as long as the transformed moments required by $d_{\gamma,p}$ are finite. When those conditions fail, one should fall back on the pointwise or finite horizon comparisons rather than on the adapted metric.

Group heterogeneity, segregation, and tail gaps. The mapping can of course be applied group by group. If two groups $g \in \{A, B\}$ face different induced opportunity distributions F_g due to segregation, homophily, or differential access to referrals, then the model implies group-specific limit laws H_{γ,F_g} . The geometric tools in Sections 3 and 4 then provide two complementary comparisons: order comparisons when F_A dominates F_B in convex order, and metric comparisons when F_A and F_B are close in Wasserstein distance but not ordered.

6.3 Normative implications: opportunity reallocation and policy design

Many interventions that affect labor market outcomes operate through changing access to opportunities, rather than directly changing offer values. Examples include referral programs, mentoring and placement initiatives, changes in search subsidies, and institutional designs that affect who meets whom. At an abstract level, such policies can all be represented as perturbations of the induced type distribution F .

Section 5 provides a tractable design problem in which a planner chooses F to optimize a tail-sensitive objective while controlling deviations from a baseline F_0 via relative entropy. In the present setting, a canonical objective is expected utility of normalized top wages,

$$U(F) = \mathbb{E}[u(Z)], \quad Z \sim H_{\gamma,F},$$

or a smoothed tail criterion that places higher marginal weight on large wage realizations. The planner problem takes the form

$$\sup_{F: \mathbb{E}_F[X]=1} \left\{ U(F) - \lambda D_{\text{KL}}(F \| F_0) \right\},$$

where λ governs the marginal cost of reshaping the cross-sectional distribution of access. When U is linear in F , including the expected-utility class identified via the stochastic representation in Section 5.2, Theorem 3 yields a unique optimizer F^* given by an exponential tilt of F_0 .

A discrete illustration with genuine degrees of freedom. To make the mean-one constraint non-vacuous, consider a baseline with three support points,

$$F_0 = \pi_1^0 \delta_{x_1} + \pi_2^0 \delta_{x_2} + \pi_3^0 \delta_{x_3}, \quad 0 < x_1 < x_2 < x_3,$$

with $\sum_j \pi_j^0 = 1$ and $\sum_j \pi_j^0 x_j = 1$. The optimal policy keeps the same support and reweights the masses according to

$$\pi_j^* \propto \pi_j^0 \exp\left(\frac{\psi(x_j) + \eta^* x_j}{\lambda}\right), \quad j = 1, 2, 3,$$

where η^* is chosen so that $\sum_j \pi_j^* x_j = 1$. Pairwise odds satisfy

$$\frac{\pi_i^* / \pi_j^*}{\pi_i^0 / \pi_j^0} = \exp\left(\frac{\psi(x_i) - \psi(x_j) + \eta^*(x_i - x_j)}{\lambda}\right).$$

Unlike the two-point case, the mean-one restriction doesn't pin down the feasible distribution uniquely, so the entropy penalty and the score produce a genuine tradeoff. This pairwise-odds equation is the discrete reduced-form analogue of the full optimal policy tradeoff in our network application: the tail score ψ , the implementation friction λ , and the shadow value η^* jointly determine how mass is reallocated across access types.

This upshot formalizes the idea that improving tail outcomes can require reallocating access across the population, but such reallocations are constrained by implementation frictions. It also complements richer normative analyses of network policy by isolating the tail-sensitive, reduced-form margin through which access is reweighted.

6.4 Microfoundations and discussion

The mixed-Poisson arrival structure can be grounded in standard network-based mechanisms of job search. For instance, if workers receive job leads from contacts and each contact generates opportunities at approximately independent random times, then conditional on network exposure the number of leads over a horizon is naturally approximated by a Poisson random variable with mean proportional to a network index. This is the reduced-form counterpart of the job search mechanisms discussed in the network-of-contacts literature overviewed in Section 1.4.

In modern settings, X can be interpreted more broadly than degree. For instance, it can incorporate platform-mediated matching intensity, differential search effort, or institution-specific access, while still entering the analysis only through its distribution F and the normalization $\mathbb{E}[X] = 1$. This interpretation is especially natural when referral processes, homophily, or segregation generate persistent differences in access to opportunities across workers or groups.

6.5 Empirical interface and identification caveats

The framework suggests a useful empirical interface, but the scope of identification should be delimited carefully.

Counts point-identify F in population, but inversion is ill-posed. If one observes offer counts at a single horizon θ , then the data identify the count distribution and its pgf. Under the mixed-Poisson structure, that data-identified pgf pins down the Laplace transform P_0 on the interval $z \in [0, \theta]$. Because P_0 is the Laplace transform of a positive measure, its values on any open interval determine the entire transform and therefore the mixing law F . Population point identification is thus not the issue. Rather, the difficulty is statistical: recovering F from one observed interval of the transform is a severely ill-posed inverse problem, so multiple horizons, repeated observations, or parametric/semi-parametric structure remain valuable for stable estimation and regularization.

Network proxies require a measurement model. Observed degree, weighted degree, referral exposure, or centrality can be informative proxies for opportunity intensity, but they do not mechanically equal the latent type X . Using them to estimate F requires a maintained mapping from observed network statistics to the shifter of arrival rates entering the mixed-Poisson specification. For the same reason, network proxies alone do not identify the adapted metric $d_{\gamma,p}$ without a maintained choice of tail index γ and a maintained measurement model for how the observed proxy maps into the latent type entering the extreme-value law.

Extreme-based inversion requires a first-step tail analysis. Inferring F from extreme outcomes requires the tail-limit ingredients, namely the index γ , the normalization (a_θ, b_θ) , and, for second-order refinements, the corresponding second-order objects, either to be known or to be estimated separately. The full parent law G need not be known for the EVT approximation itself. The asymptotic expansion in Theorem 2 useful precisely because it separates second-order EVT error from the heterogeneity kernel. But it does not, on its own, solve the statistical recovery problem for F .

Error propagation is conditional on the sampling scheme and can be Hölder in Fréchet settings. Once an empirical approximation \hat{F} is available, Corollary 6 converts metric error in \hat{F} into error in predicted tail laws. If $0 < \gamma < 1$ and the assumptions of Theorem 1 hold with $p\gamma < 1$, Proposition 11 yields

$$W_p(H_{\gamma, \hat{F}}, H_{\gamma, F}) \leq C_{\gamma,p} W_p(\hat{F}, F)^\gamma.$$

Combined with Corollary 3, the same argument gives

$$\|Q_{\gamma, \hat{F}} - Q_{\gamma, F}\|_{L^p(0,1)} \leq C_{\gamma,p} W_p(\hat{F}, F)^\gamma.$$

Thus the bridge from raw space estimation error in F to error in predicted extremes is generally only Hölder, not linear, in the economically important Fréchet regime. If $W_p(\hat{F}, F) = O_P(r_n)$, the induced law error is only $O_P(r_n^\gamma)$. Standard benchmark Wasserstein rates such as Fournier and Guillin (2015) therefore remain informative, but only after this nonlinear penalty is made explicit. Shared-network data, however, feature cross-sectional dependence, so those rates should be treated as benchmarks rather than as automatic guarantees. Additional graph-dependence assumptions would be needed for a full statistical theory.

Finite-horizon and asymptotic claims should be kept distinct. For prelimit predictions at a fixed horizon θ , Proposition 13 is the relevant robustness statement. For asymptotic tail approximations, Theorem 1 and Theorem 2 are the relevant tools. Keeping those objects separate helps prevent finite-sample robustness, asymptotic approximation, and statistical identification from being conflated.

7 Discussion and conclusion

This paper studies extremes in environments where the number of opportunities is heterogeneous and random. Starting from the mixed-Poisson heterogeneous agent extreme value limits in Mangin (2026), we treat the heterogeneity distribution as the primitive object and analyze the induced map $F \mapsto H_{\gamma,F}$ with order, Wasserstein-geometric, and complementary entropy-projection tools.

7.1 Main takeaways

Four points structure the contribution.

First, heterogeneous extremes admit a compact Laplace-transform representation. Once the primitive offer distribution is in a classical domain of attraction, the effect of unequal access to opportunities is summarized by the mean-one distribution F of draw intensities and its Laplace transform. This makes heterogeneity analytically tractable without collapsing it to a scalar index.

Second, the paper’s main new quantitative theorem is geometric. Theorem 1 shows that, once the canonical coupling representation is in hand, perturbations in F propagate Lipschitzly into Wasserstein perturbations of the entire induced law of extremes. Corollary 2 packages the same structure into canonical ambient interpolation paths, while Theorem 2 cleanly separates second-order EVT approximation error from the heterogeneity kernel. By contrast, convex-order comparisons and several pointwise inequalities are consequences of the Laplace form and are included because they make the operator economically usable.

Third, the labor market network application shows how these ingredients fit together in a familiar environment, but only in reduced form. Network position maps into opportunity intensity, network inequality maps into tail distortions for a randomly drawn worker’s top wage, and the adapted geometry controls the whole counterfactual quantile schedule together with the renormalization error induced by returning to the mean one slice. Turning those objects into equilibrium or welfare claims would require additional structure beyond the scope of the present paper.

Fourth, the same type space supports a complementary design problem. Under KL regularization and a mean constraint, the planner’s problem is a one-marginal entropy projection with a unique exponential-tilt solution. The connection to the positive analysis is structural rather than geometric: the same Laplace kernel governs both the heterogeneous extreme-value operator and the score functions of a broad class of linear objectives, including cdf-based criteria and expected utility of normalized extremes.

7.2 Extensions

Several extensions seem natural.

- **Endogenous networks and strategic exposure.** In applications where network position is chosen or co-determined with outcomes, one could embed the mixed-Poisson intensity into an equilibrium model of link formation or effort choice. The present stability and design results can then be used either as reduced-form comparative statics or as primitives inside a fixed-point argument.
- **Dynamic designs and Schrödinger bridges.** Our entropy-regularized formulation in Section 5 is static. A dynamic extension would allow the planner to steer heterogeneity over time subject to intertemporal entropic costs, which naturally leads to Schrödinger-bridge-type problems. That extension would be genuinely richer than the one-marginal projection studied here.

- **Dynamic record processes and sequential search.** Because each new offer can be viewed as a potential record, one natural extension is to study the full record process generated by heterogeneous arrival rates rather than only the terminal maximum. Classical references in this context include Arnold et al. (1998). Such an extension would connect the present geometry of extremes to questions about the timing of record improvements and threshold-crossing times.
- **Beyond mixed Poisson counts and beyond independence.** The mixed-Poisson structure provides tractability through a Laplace transform. Other count models or dependence structures can be accommodated whenever one can represent the distribution of maxima by a tractable transform and obtain couplings that separate heterogeneity from idiosyncratic shocks. The classical random-sample-size maxima literature provides the natural benchmark for that extension; see Barndorff-Nielsen (1964), Galambos (1973), or Silvestrov and Teugels (1998).
- **Statistical identification and inference.** The rate expansion in Section 4.6 and the stability results in Section 4 suggest a path toward inference for F from tail observations or count data. A full treatment would have to combine second-order tail estimation for G , stable recovery or regularized inversion of F from observed count distributions or extremes, and a dependence-aware theory for network data. The present paper isolates the operator-theoretic ingredients needed for that future step rather than claiming to complete it.

Conclusion. Extremes are economically salient precisely because they aggregate rare opportunities and heterogeneity in access. The framework developed here provides a tractable way to quantify how heterogeneity shapes agent-level extremes, how robust those implications are to perturbations in heterogeneity, and how a planner can reshape heterogeneity when the objective is tail sensitive and policy changes are entropy penalized. The transport results should be read on their natural geometric domain, Proposition 10 as the explicit link back to the mean-one economic slice, and the labor-network application as a disciplined reduced-form rather than a complete equilibrium model.

A Proofs

A.1 Auxiliary limit for maxima

We repeatedly use the elementary implication that turns convergence of powers into a first-order tail approximation.

Lemma 5 (From powers to first-order tails). *Let $\{u_n\}_{n \geq 1} \subset [0, 1)$ and suppose there exists $v \in [0, \infty)$ such that*

$$(1 - u_n)^n \rightarrow e^{-v} \quad \text{as } n \rightarrow \infty.$$

Then $u_n \rightarrow 0$ and $nu_n \rightarrow v$.

Proof. Since $(1 - u_n)^n \rightarrow e^{-v} \in (0, 1]$, we must have $1 - u_n \rightarrow 1$, hence $u_n \rightarrow 0$.

Write $(1 - u_n)^n = \exp(n \log(1 - u_n))$. Taking logs yields $n \log(1 - u_n) \rightarrow -v$.

For $u \in [0, 1)$, the inequalities

$$-\frac{u}{1-u} \leq \log(1-u) \leq -u$$

hold. Applying them with $u = u_n$ and multiplying by $-n$ gives

$$\frac{nu_n}{1-u_n} \geq -n \log(1-u_n) \geq nu_n.$$

Letting $n \rightarrow \infty$ and using $u_n \rightarrow 0$ and $-n \log(1 - u_n) \rightarrow v$, we obtain

$$\limsup_{n \rightarrow \infty} nu_n \leq v \quad \text{and} \quad \liminf_{n \rightarrow \infty} nu_n \geq v,$$

so $nu_n \rightarrow v$. □

A.2 Proof of Lemma 1

Proof. Fix $y \in [0, 1]$ and $\theta > 0$. By conditioning on X and using the probability generating function of a Poisson random variable,

$$\mathbb{E}\left[y^{N(\theta)} \mid X\right] = \exp(-\theta X(1 - y)).$$

Taking expectations over X yields

$$\mathbb{E}\left[y^{N(\theta)}\right] = \mathbb{E}\left[\exp(-\theta X(1 - y))\right] = P_0(\theta(1 - y)),$$

which is the claimed identity. □

A.3 Proof of Proposition 1

Proof. Fix $x \in \mathbb{R}$. Let $n = \lfloor \theta \rfloor$ and recall that $a_\theta = a_n$ and $b_\theta = b_n$ by definition. Set

$$x_\theta := b_\theta + a_\theta x = b_n + a_n x.$$

Step 1: reduce to a Laplace argument. By (1),

$$\Pr\left(\frac{M_\theta - b_\theta}{a_\theta} \leq x\right) = \Pr(M_\theta \leq x_\theta) = P_0(\theta(1 - G(x_\theta))).$$

Since P_0 is the Laplace transform of a nonnegative random variable, it is continuous on $[0, \infty)$.

Step 2: identify the limit of $\theta(1 - G(x_\theta))$. Assumption 1 implies

$$\Pr\left(\frac{M_n - b_n}{a_n} \leq x\right) = \Pr(M_n \leq b_n + a_n x) = G(b_n + a_n x)^n \rightarrow H_\gamma(x) = e^{-v_\gamma(x)}$$

at all continuity points x of H_γ . Define $u_n(x) := 1 - G(b_n + a_n x) \in [0, 1]$. Then $G(b_n + a_n x)^n = (1 - u_n(x))^n$.

If $H_\gamma(x) > 0$, equivalently $v_\gamma(x) < \infty$, Lemma 5 gives

$$nu_n(x) \rightarrow v_\gamma(x) \quad \text{and} \quad u_n(x) \rightarrow 0.$$

Hence

$$|\theta u_n(x) - nu_n(x)| \leq |\theta - n| u_n(x) \leq u_n(x) \rightarrow 0,$$

since $|\theta - n| < 1$. Therefore

$$\theta(1 - G(b_n + a_n x)) = \theta u_n(x) \rightarrow v_\gamma(x).$$

If instead $H_\gamma(x) = 0$, then $(1 - u_n(x))^n \rightarrow 0$. We claim that $nu_n(x) \rightarrow \infty$. If not, there would exist a subsequence n_k and a constant $M < \infty$ such that $n_k u_{n_k}(x) \leq M$. Then $u_{n_k}(x) \leq M/n_k \rightarrow 0$.

Passing to a further subsequence if needed, we may assume $n_k u_{n_k}(x) \rightarrow L$ for some finite $L \in [0, M]$. Since $\log(1 - u) = -u + o(u)$ as $u \downarrow 0$, it follows that

$$n_k \log(1 - u_{n_k}(x)) \rightarrow -L,$$

and therefore

$$(1 - u_{n_k}(x))^{n_k} \rightarrow e^{-L} > 0,$$

contradicting $(1 - u_n(x))^n \rightarrow 0$. Thus $n u_n(x) \rightarrow \infty$. Since $\theta \geq n$, it follows that

$$\theta u_n(x) \geq n u_n(x) \rightarrow \infty.$$

Step 3: pass to the limit through P_0 . If $H_\gamma(x) > 0$, Step 2 and continuity of P_0 yield

$$P_0(\theta(1 - G(x_\theta))) \rightarrow P_0(v_\gamma(x)) = H_{\gamma,F}(x).$$

If $H_\gamma(x) = 0$, then $\theta(1 - G(x_\theta)) \rightarrow \infty$ by Step 2. Because $e^{-zX} \rightarrow 1_{\{X=0\}}$ as $z \rightarrow \infty$ and Assumption 2 imposes $\Pr(X = 0) = 0$, dominated convergence gives

$$P_0(z) = \mathbb{E}[e^{-zX}] \rightarrow 0 \quad \text{as } z \rightarrow \infty.$$

Hence

$$P_0(\theta(1 - G(x_\theta))) \rightarrow 0 = H_{\gamma,F}(x).$$

This proves the asserted convergence at all continuity points of $H_{\gamma,F}$.

Degenerate case. If $F = \delta_1$, then $X = 1$ almost surely, so $P_0(z) = \mathbb{E}[e^{-zX}] = e^{-z}$ and hence

$$H_{\gamma,F}(x) = P_0(v_\gamma(x)) = e^{-v_\gamma(x)} = H_\gamma(x).$$

□

A.4 Proofs for Section 3

Proof of Proposition 2. Let $X_1 \sim F_1$ and $X_2 \sim F_2$ with $\mathbb{E}[X_1] = \mathbb{E}[X_2] = 1$. If F_2 is a mean-preserving spread of F_1 , then X_2 dominates X_1 in convex order. Equivalently,

$$\mathbb{E}[\varphi(X_2)] \geq \mathbb{E}[\varphi(X_1)] \quad \text{for every convex } \varphi \text{ for which both expectations are finite.}$$

Fix $z \geq 0$ and consider $\varphi_z(x) := e^{-zx}$. Then φ_z is convex on $[0, \infty)$ (since $\varphi_z''(x) = z^2 e^{-zx} \geq 0$) and bounded by 1. Hence,

$$P_0^{(2)}(z) = \mathbb{E}[e^{-zX_2}] \geq \mathbb{E}[e^{-zX_1}] = P_0^{(1)}(z) \quad \text{for all } z \geq 0.$$

Composing with $z = v_\gamma(x) \geq 0$ yields

$$H_{\gamma,F_2}(x) = P_0^{(2)}(v_\gamma(x)) \geq P_0^{(1)}(v_\gamma(x)) = H_{\gamma,F_1}(x) \quad \text{for all } x.$$

Therefore, the distribution function under F_2 is everywhere larger, which is equivalent to first-order stochastic dominance in the direction that extremes are smaller under F_2 . □

Proof of Corollary 1. Apply Proposition 2 with $F_1 = \delta_1$ and $F_2 = F$. Since every mean one distribution on $[0, \infty)$ dominates δ_1 in convex order, one obtains

$$P_0(z) \geq e^{-z} \quad \text{for all } z \geq 0,$$

and therefore $H_{\gamma, F}(x) \geq H_\gamma(x)$ for every x with $v_\gamma(x) < \infty$. If $F \neq \delta_1$, then strict convexity of $x \mapsto e^{-zx}$ for every $z > 0$ and Jensen's inequality imply strict inequality for all $z > 0$. \square

Proof of Proposition 3. By Lemma 2, the canonical monotone geodesic $(\mu_t)_{t \in [0,1]}$ satisfies

$$Q_{\mu_t}(u) = (1-t)Q_\mu(u) + tQ_\nu(u), \quad u \in (0, 1).$$

Since φ is convex,

$$\varphi(Q_{\mu_t}(u)) \leq (1-t)\varphi(Q_\mu(u)) + t\varphi(Q_\nu(u)) \quad \text{for every } u \in (0, 1).$$

Integrating over $u \in (0, 1)$ and using the quantile representation of integrals gives

$$\int \varphi d\mu_t \leq (1-t) \int \varphi d\mu + t \int \varphi d\nu.$$

Applying the same argument to any pair of times $s, t \in [0, 1]$ in place of 0, 1 yields convexity of $t \mapsto \int \varphi d\mu_t$ on $[0, 1]$. \square

Proof of Proposition 4. For the Fréchet case, let Z_γ be independent of $X \sim F$ and satisfy

$$\Pr(Z_\gamma \leq z) = \exp(-z^{-1/\gamma}), \quad z > 0.$$

Then, for $z > 0$,

$$\Pr(X^\gamma Z_\gamma \leq z \mid X) = \Pr(Z_\gamma \leq zX^{-\gamma} \mid X) = \exp(-Xz^{-1/\gamma}).$$

Taking expectations over X gives

$$\Pr(X^\gamma Z_\gamma \leq z) = \mathbb{E}[e^{-Xz^{-1/\gamma}}] = P_0(z^{-1/\gamma}),$$

which is the cdf of the heterogeneous Fréchet limit. If $Z_{\text{GEV}} := (X^\gamma Z_\gamma - 1)/\gamma$, then

$$\Pr(Z_{\text{GEV}} \leq x) = \Pr(X^\gamma Z_\gamma \leq 1 + \gamma x) = P_0((1 + \gamma x)^{-1/\gamma}) = H_{\gamma, F}(x).$$

For the Gumbel case, let $E \sim \text{Exp}(1)$ be independent of $X \sim F$ and define $Z := \log(X/E)$. Then, for every $x \in \mathbb{R}$,

$$\Pr(Z \leq x \mid X) = \Pr(E \geq Xe^{-x} \mid X) = e^{-Xe^{-x}}.$$

Taking expectations yields

$$\Pr(Z \leq x) = \mathbb{E}[e^{-Xe^{-x}}] = P_0(e^{-x}) = H_{0, F}(x).$$

\square

Proof of Lemma 3. By Proposition 4, $Z = X^\gamma Z_\gamma$ with Z_γ independent of X . Hence, by iterated expectation,

$$\mathbb{E}[\psi(Z)] = \mathbb{E}[\mathbb{E}[\psi(X^\gamma Z_\gamma) \mid X]] = \mathbb{E}[\kappa_\psi(X)] = \int_0^\infty \kappa_\psi(x) F(dx).$$

□

Proof of Proposition 5. For every x with $v_\gamma(x) < \infty$,

$$H_{\gamma, F_\varepsilon}(x) = \int e^{-v_\gamma(x)u} F_\varepsilon(du) = \int e^{-v_\gamma(x)u} F(du) + \varepsilon \int e^{-v_\gamma(x)u} \nu(du).$$

Differentiating with respect to ε at 0 yields the claim. □

Proof of Proposition 6. Fix $z \geq 0$ and apply Proposition 3 with the convex function $\varphi_z(x) = e^{-zx}$. Then

$$t \mapsto \int e^{-zx} F^t(dx) = P_0^t(z)$$

is convex on $[0, 1]$. Setting $z = v_\gamma(x)$ gives the convexity of $t \mapsto H_{\gamma, F^t}(x)$. □

Proof of Proposition 7. On $(0, \infty)$, the function $\varphi(x) = x^{-\rho}$ is convex for every $\rho > 0$. Since the geodesic is supported on $[a, \infty)$, the function is also bounded on the support of every F^t . Applying Proposition 3 with this φ yields convexity of

$$t \mapsto \int x^{-\rho} F^t(dx).$$

□

A.5 Proofs for Section 4

Proof of Proposition 8. Let $X \sim F$ and $E \sim \text{Exp}(1)$ be independent, and define $Z_{\gamma, F} = w_\gamma(E/X)$. For any x such that $v_\gamma(x) \in (0, \infty)$,

$$\Pr(Z_{\gamma, F} \leq x \mid X) = \Pr(w_\gamma(E/X) \leq x \mid X) = \Pr(E/X \geq v_\gamma(x) \mid X),$$

since w_γ is decreasing. Therefore,

$$\Pr(Z_{\gamma, F} \leq x \mid X) = \exp(-Xv_\gamma(x)).$$

Taking expectations yields

$$\Pr(Z_{\gamma, F} \leq x) = \mathbb{E}\left[e^{-Xv_\gamma(x)}\right] = P_0(v_\gamma(x)) = H_{\gamma, F}(x).$$

The explicit formulas follow from $w_\gamma(t) = (t^{-\gamma} - 1)/\gamma$ for $\gamma \neq 0$ and $w_0(t) = -\log t$. □

Proof of Theorem 1. Case $\gamma \neq 0$. Let $\mu_i := s_\gamma \# F_i$ for $i \in \{1, 2\}$, so that $d_{\gamma, p}(F_1, F_2) = W_p(\mu_1, \mu_2)$. Let (U_1, U_2) be an optimal coupling of (μ_1, μ_2) , and let $E \sim \text{Exp}(1)$ be independent of (U_1, U_2) . Set $V := E^{-\gamma}$ and define

$$\phi(u, v) := \frac{uv - 1}{\gamma}.$$

If $X_i := s_\gamma^{-1}(U_i)$, then $X_i \sim F_i$ and

$$\phi(U_i, V) = \frac{X_i^\gamma E^{-\gamma} - 1}{\gamma} = w_\gamma(E/X_i).$$

By Proposition 8, $\mathcal{L}(\phi(U_i, V)) = H_{\gamma, F_i}$. For each fixed v , the map $u \mapsto \phi(u, v)$ is Lipschitz with constant $|v|/|\gamma|$. Hence Lemma 9 gives

$$W_p(H_{\gamma, F_1}, H_{\gamma, F_2}) \leq \frac{(\mathbb{E}[|E^{-\gamma}|^p])^{1/p}}{|\gamma|} W_p(\mu_1, \mu_2) = \frac{(\mathbb{E}[|E^{-\gamma}|^p])^{1/p}}{|\gamma|} d_{\gamma, p}(F_1, F_2).$$

This proves the bound for $\gamma \neq 0$.

Case $\gamma = 0$. Let $\mu_i := (\log)\#F_i$ for $i \in \{1, 2\}$, so that $d_{0, p}(F_1, F_2) = W_p(\mu_1, \mu_2)$. Let (U_1, U_2) be an optimal coupling of (μ_1, μ_2) , let $E \sim \text{Exp}(1)$ be independent, and define

$$\phi(u, v) := u - \log v.$$

If $X_i := e^{U_i}$, then $X_i \sim F_i$ and

$$\phi(U_i, E) = \log X_i - \log E = w_0(E/X_i).$$

By Proposition 8, $\mathcal{L}(\phi(U_i, E)) = H_{0, F_i}$. For each fixed $v > 0$, the map $u \mapsto \phi(u, v)$ is 1-Lipschitz, so Lemma 9 yields

$$W_p(H_{0, F_1}, H_{0, F_2}) \leq W_p(\mu_1, \mu_2) = d_{0, p}(F_1, F_2). \quad \square$$

Proof of Corollary 2. By construction,

$$d_{\gamma, p}(F^s, F^t) = W_p(G^s, G^t).$$

Since $(G^t)_{t \in [0, 1]}$ is the canonical monotone constant-speed W_p geodesic from G^0 to G^1 ,

$$W_p(G^s, G^t) = |t - s| W_p(G^0, G^1) = |t - s| d_{\gamma, p}(F^0, F^1).$$

Applying Theorem 1 to F^s and F^t yields the second claim. □

Proof of Proposition 10. Let $m := m(F)$ and let $Y := s_\gamma(X)$ for $X \sim F$. If $\gamma \neq 0$, then

$$s_\gamma\left(\frac{X}{m}\right) = \left(\frac{X}{m}\right)^\gamma = m^{-\gamma} X^\gamma = m^{-\gamma} Y.$$

Hence $s_\gamma\#\tilde{F}$ is the law of $m^{-\gamma}Y$. Because quantiles scale linearly under multiplication by a positive constant,

$$d_{\gamma, p}(F, \tilde{F})^p = W_p(\mathcal{L}(Y), \mathcal{L}(m^{-\gamma}Y))^p = \int_0^1 |Q_Y(u) - m^{-\gamma}Q_Y(u)|^p du.$$

Therefore

$$d_{\gamma, p}(F, \tilde{F}) = |1 - m^{-\gamma}| \left(\int_0^1 |Q_Y(u)|^p du \right)^{1/p} = |1 - m^{-\gamma}| (\mathbb{E}[|Y|^p])^{1/p},$$

which is exactly the stated formula for $\gamma \neq 0$. If $\gamma = 0$, then $s_0(X/m) = \log X - \log m$, so $s_0\#\tilde{F}$ is the translate of $s_0\#F$ by $-\log m$. Again by the quantile representation,

$$d_{0, p}(F, \tilde{F}) = W_p((\log)\#F, \tau_{-\log m}\#((\log)\#F)) = |\log m|,$$

where $\tau_c(y) := y + c$. The Wasserstein bound for the induced extreme laws is then immediate from Theorem 1. □

Proof of Corollary 3. The equality with $W_p(H_{\gamma, F_1}, H_{\gamma, F_2})$ is the one-dimensional quantile representation of Wasserstein distance. The first inequality is Theorem 1. The geodesic statement then follows from Corollary 2. \square

Proof of Corollary 4. Fix x in the interior of the support of H_γ and define

$$f_x(z) := e^{-v_\gamma(x)z}, \quad z \geq 0.$$

Then f_x is $v_\gamma(x)$ -Lipschitz on $[0, \infty)$ because

$$|f'_x(z)| = v_\gamma(x)e^{-v_\gamma(x)z} \leq v_\gamma(x).$$

Since

$$H_{\gamma, F_i}(x) = \int f_x(z) F_i(dz),$$

Corollary 7 implies

$$|H_{\gamma, F_1}(x) - H_{\gamma, F_2}(x)| \leq v_\gamma(x)W_1(F_1, F_2).$$

\square

Proof of Proposition 9. By definition,

$$d_{\gamma, p}(F_1, F_2) = W_p(F_1 \circ s_\gamma^{-1}, F_2 \circ s_\gamma^{-1}).$$

A Wasserstein distance on \mathbb{R} is finite if and only if both marginals belong to $\mathcal{P}_p(\mathbb{R})$. Hence $d_{\gamma, p}(F_1, F_2) < \infty$ if and only if

$$\int |y|^p (F_i \circ s_\gamma^{-1})(dy) < \infty, \quad i \in \{1, 2\}.$$

By change of variables this is equivalent to

$$\int |s_\gamma(x)|^p F_i(dx) < \infty, \quad i \in \{1, 2\}.$$

The three regime-specific statements follow by substituting $s_\gamma(x) = x^\gamma$ for $\gamma \neq 0$ and $s_0(x) = \log x$. \square

Proof of Proposition 11. Case 1: $0 < \gamma \leq 1$. Let (X, Y) be any coupling of (F_1, F_2) . Since $x \mapsto x^\gamma$ is γ -Hölder on $[0, \infty)$,

$$|X^\gamma - Y^\gamma| \leq |X - Y|^\gamma.$$

Therefore,

$$\mathbb{E}[|X^\gamma - Y^\gamma|^p] \leq \mathbb{E}[|X - Y|^{\gamma p}] \leq (\mathbb{E}[|X - Y|^p])^\gamma,$$

where the last step is Jensen's inequality applied to the concave map $t \mapsto t^\gamma$. Taking the infimum over couplings and then p th roots yields

$$d_{\gamma, p}(F_1, F_2) \leq W_p(F_1, F_2)^\gamma.$$

Case 2: $\gamma < 0$. On $[a, \infty)$, the derivative of $x \mapsto x^\gamma$ satisfies

$$\left| \frac{d}{dx} x^\gamma \right| = |\gamma| x^{\gamma-1} \leq |\gamma| a^{\gamma-1}.$$

Hence $x \mapsto x^\gamma$ is $|\gamma|a^{\gamma-1}$ -Lipschitz on $[a, \infty)$. Lemma 8 gives

$$d_{\gamma,p}(F_1, F_2) = W_p((x \mapsto x^\gamma)_\#F_1, (x \mapsto x^\gamma)_\#F_2) \leq |\gamma|a^{\gamma-1}W_p(F_1, F_2).$$

Case 3: $\gamma = 0$. On $[a, \infty)$, the derivative of $\log x$ is bounded by a^{-1} . Thus \log is a^{-1} -Lipschitz there, and Lemma 8 yields

$$d_{0,p}(F_1, F_2) = W_p((\log)_\#F_1, (\log)_\#F_2) \leq a^{-1}W_p(F_1, F_2).$$

□

Proof of Proposition 12. Apply Corollary 7 with $\mu = H_{\gamma, F_1}$, $\nu = H_{\gamma, F_2}$, and the L -Lipschitz function ψ . □

Proof of Proposition 13. Fix $x \in \mathbb{R}$ and define

$$f_x(z) := \exp(-\theta z(1 - G(x))), \quad z \geq 0.$$

Then

$$|f'_x(z)| = \theta(1 - G(x)) \exp(-\theta z(1 - G(x))) \leq \theta(1 - G(x)),$$

so f_x is $\theta(1 - G(x))$ -Lipschitz on $[0, \infty)$. Because

$$\Pr_F(M_\theta \leq x) = \int f_x(z) F(dz),$$

Corollary 7 gives

$$\left| \Pr_{F_1}(M_\theta \leq x) - \Pr_{F_2}(M_\theta \leq x) \right| \leq \theta(1 - G(x))W_1(F_1, F_2).$$

□

Proof of Theorem 2. Fix a compact set K contained in the interior of the support of H_γ . By Assumption 3, there exists a remainder $\delta_\theta(x)$ such that

$$\theta(1 - G(b(\theta) + a(\theta)x)) = v_\gamma(x) + A(\theta)h_\gamma(x) + \delta_\theta(x), \quad x \in K,$$

with

$$\sup_{x \in K} \frac{|\delta_\theta(x)|}{|A(\theta)|} \rightarrow 0.$$

Write

$$\Delta_\theta(x) := A(\theta)h_\gamma(x) + \delta_\theta(x).$$

Since h_γ is locally bounded and K is compact, there exists $B_K < \infty$ such that $|h_\gamma(x)| \leq B_K$ on K . Hence

$$\sup_{x \in K} |\Delta_\theta(x)| = O(|A(\theta)|).$$

By (1),

$$\Pr(Z_\theta \leq x) = P_0(v_\gamma(x) + \Delta_\theta(x)), \quad x \in K.$$

Set

$$m_K := \inf_{x \in K} v_\gamma(x) > 0, \quad M_K := \sup_{x \in K} v_\gamma(x) < \infty.$$

For $t \geq m_K/2$, differentiation under the integral sign gives

$$P'_0(t) = -\mathbb{E}[Xe^{-tX}], \quad P''_0(t) = \mathbb{E}[X^2e^{-tX}].$$

Moreover, for every $t > 0$,

$$X^2e^{-tX} \leq \sup_{y \geq 0} y^2 e^{-y} t^{-2} = 4e^{-2}t^{-2},$$

so

$$\sup_{t \geq m_K/2} |P''_0(t)| \leq \frac{16e^{-2}}{m_K^2} < \infty.$$

Therefore P_0 has uniformly bounded second derivative on a neighborhood of $\{v_\gamma(x) : x \in K\}$. A second-order Taylor expansion yields, uniformly for $x \in K$,

$$P_0(v_\gamma(x) + \Delta_\theta(x)) = P_0(v_\gamma(x)) + P'_0(v_\gamma(x)) \Delta_\theta(x) + R_\theta(x),$$

where

$$\sup_{x \in K} |R_\theta(x)| \leq C_K \sup_{x \in K} |\Delta_\theta(x)|^2 = O(A(\theta)^2) = o(A(\theta)).$$

Since $P_0(v_\gamma(x)) = H_{\gamma,F}(x)$ and $P'_0(v_\gamma(x)) = -\mathbb{E}[Xe^{-v_\gamma(x)X}]$, we obtain

$$\Pr(Z_\theta \leq x) = H_{\gamma,F}(x) - A(\theta)h_\gamma(x)\mathbb{E}[Xe^{-v_\gamma(x)X}] + r_{\theta,K}(x),$$

where

$$r_{\theta,K}(x) := -\delta_\theta(x)\mathbb{E}[Xe^{-v_\gamma(x)X}] + R_\theta(x).$$

Because $Xe^{-v_\gamma(x)X} \leq e^{-1}/m_K$ uniformly on K , the first term is $o(A(\theta))$ uniformly on K ; the second term is already $o(A(\theta))$ uniformly on K . Hence

$$\sup_{x \in K} \frac{|r_{\theta,K}(x)|}{|A(\theta)|} \rightarrow 0,$$

which proves the theorem. □

A.6 Proofs for Section 5

Proof of Theorem 3. We write the design problem (3) in the linear case as

$$\sup_{F: \mathbb{E}_F[X]=1} \left\{ \int \psi(x) F(dx) - \lambda D_{\text{KL}}(F \| F_0) \right\}.$$

Throughout, we restrict attention to $F \ll F_0$, since otherwise $D_{\text{KL}}(F \| F_0) = +\infty$.

Step 1: weak duality. Fix $\eta \in \mathcal{D}$. For any feasible F with $\mathbb{E}_F[X] = 1$,

$$\int \psi dF - \lambda D_{\text{KL}}(F \| F_0) = \int (\psi + \eta x) dF - \lambda D_{\text{KL}}(F \| F_0) - \eta.$$

Applying Theorem 5 with $P = F_0$ and $f = (\psi + \eta x)/\lambda$ gives

$$\int (\psi + \eta x) dF - \lambda D_{\text{KL}}(F \| F_0) \leq \lambda \log Z(\eta).$$

Hence, for every feasible F and every $\eta \in \mathcal{D}$,

$$\int \psi dF - \lambda D_{\text{KL}}(F \| F_0) \leq \lambda \log Z(\eta) - \eta.$$

Taking the supremum over feasible F and the the infimum over $\eta \in \mathcal{D}$ yields the weak-duality bound

$$\sup_{F: \mathbb{E}_F[X]=1} \left\{ \int \psi dF - \lambda D_{\text{KL}}(F \| F_0) \right\} \leq \inf_{\eta \in \mathcal{D}} \{ \lambda \log Z(\eta) - \eta \}.$$

Step 2: the dual objective is smooth and has a unique minimizer. Define

$$\mathcal{J}(\eta) := \lambda \log Z(\eta) - \eta, \quad \eta \in \mathcal{D}.$$

Let $I = [a, b] \subset \mathcal{D}$ be compact. For every $\eta \in I$ and every $x > 0$,

$$\exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) \leq \exp\left(\frac{\psi(x) + bx}{\lambda}\right),$$

and the latter is integrable by Assumption 4. Similarly,

$$x \exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) \leq x \exp\left(\frac{\psi(x) + bx}{\lambda}\right),$$

and

$$x^2 \exp\left(\frac{\psi(x) + \eta x}{\lambda}\right) \leq x^2 \exp\left(\frac{\psi(x) + bx}{\lambda}\right),$$

with both dominating functions integrable. Therefore dominated convergence implies that Z is twice continuously differentiable on \mathcal{D} , with

$$Z'(\eta) = \frac{1}{\lambda} M_1(\eta), \quad Z''(\eta) = \frac{1}{\lambda^2} M_2(\eta).$$

Hence

$$\mathcal{J}'(\eta) = \frac{M_1(\eta)}{Z(\eta)} - 1 = m(\eta) - 1,$$

and

$$\mathcal{J}''(\eta) = \frac{1}{\lambda} \text{Var}_{F_\eta}(X) \geq 0,$$

where F_η is the exponential tilt defined by

$$\frac{dF_\eta}{dF_0}(x) = \frac{\exp\left(\frac{\psi(x) + \eta x}{\lambda}\right)}{Z(\eta)}.$$

We claim that in fact $\text{Var}_{F_\eta}(X) > 0$ for every $\eta \in \mathcal{D}$. Indeed, if $\text{Var}_{F_\eta}(X) = 0$, then F_η would be a Dirac mass. But the density dF_η/dF_0 is strictly positive on the support of F_0 , so this can happen only if F_0 itself is a Dirac mass. In that case $m(\eta) \equiv 1$, contradicting the assumption that there exist $\eta_-, \eta_+ \in \mathcal{D}$ with $m(\eta_-) < 1 < m(\eta_+)$. Therefore

$$\mathcal{J}''(\eta) > 0 \quad \text{for all } \eta \in \mathcal{D},$$

so \mathcal{J} is strictly convex on \mathcal{D} and \mathcal{J}' is strictly increasing and continuous. By Assumption 4,

$$\mathcal{J}'(\eta_-) = m(\eta_-) - 1 < 0 < m(\eta_+) - 1 = \mathcal{J}'(\eta_+).$$

The intermediate value theorem therefore yields some $\eta^* \in (\eta_-, \eta_+)$ such that $\mathcal{J}'(\eta^*) = 0$, that is, $m(\eta^*) = 1$. Because \mathcal{J}' is strictly increasing, this root is unique. Since \mathcal{J} is strictly convex, the same η^* is the unique minimizer of \mathcal{J} on \mathcal{D} .

Step 3: construct the primal optimizer and prove strong duality. Let $F^* := F_{\eta^*}$. By construction,

$$\mathbb{E}_{F^*}[X] = m(\eta^*) = 1,$$

so F^* is feasible. Because Assumption 4 also gives

$$\int |\psi(x)| \exp\left(\frac{\psi(x) + \eta^* x}{\lambda}\right) F_0(dx) < \infty,$$

and because $\mathbb{E}_{F^*}[X] = 1$ implies

$$\int x \exp\left(\frac{\psi(x) + \eta^* x}{\lambda}\right) F_0(dx) = Z(\eta^*) < \infty,$$

one also has $\int |f| e^f dF_0 < \infty$ for $f = (\psi + \eta^* x)/\lambda$. Therefore Theorem 5 attains equality at F^* for that function f . Hence

$$\int (\psi + \eta^* x) dF^* - \lambda D_{\text{KL}}(F^* \| F_0) = \lambda \log Z(\eta^*).$$

Using $\mathbb{E}_{F^*}[X] = 1$, we obtain

$$\int \psi dF^* - \lambda D_{\text{KL}}(F^* \| F_0) = \lambda \log Z(\eta^*) - \eta^* = \mathcal{J}(\eta^*).$$

Because η^* minimizes \mathcal{J} , this value coincide with the weak-duality upper bound. Therefore strong duality holds, F^* is optimal, and (5) and (4) follow.

Step 4: uniqueness of the primal optimizer. The map $F \mapsto \int \psi dF$ is linear, and $F \mapsto D_{\text{KL}}(F \| F_0)$ is strictly convex on $\{F \ll F_0\}$. Hence

$$F \mapsto \int \psi dF - \lambda D_{\text{KL}}(F \| F_0)$$

is strictly concave on the convex feasible set $\{F : \mathbb{E}_F[X] = 1, F \ll F_0\}$. A strictly concave objective has at most one maximizer, so the optimal F^* is unique. \square

B Technical tools

B.1 Wasserstein geometry and duality

We collect standard facts about Wasserstein distance that are used throughout the paper. References include Villani (2009) and Santambrogio (2015).

Let (X, d) be a Polish metric space. For $p \geq 1$, write $\mathcal{P}_p(X)$ for the set of Borel probability measures on X with finite p th moment.

Definition 4 (Wasserstein distance). For $\mu, \nu \in \mathcal{P}_p(X)$,

$$W_p(\mu, \nu) := \left(\inf_{\pi \in \Gamma(\mu, \nu)} \int d(x, y)^p \pi(dx, dy) \right)^{1/p},$$

where $\Gamma(\mu, \nu)$ denotes the set of couplings of μ and ν .

Lemma 6 (Upper bound from any coupling). *Let $p \geq 1$ and let U, V be random variables in X . If $\mathcal{L}(U) = \mu$ and $\mathcal{L}(V) = \nu$, then*

$$W_p(\mu, \nu) \leq (\mathbb{E}[d(U, V)^p])^{1/p}.$$

Proof. Let $\pi := \mathcal{L}(U, V) \in \Gamma(\mu, \nu)$. By Definition 4,

$$W_p^p(\mu, \nu) \leq \int d(x, y)^p \pi(dx, dy) = \mathbb{E}[d(U, V)^p].$$

□

Lemma 7 (One-dimensional quantile representation). *Let $p \geq 1$ and let $\mu, \nu \in \mathcal{P}_p(\mathbb{R})$, where \mathbb{R} is equipped with the metric $d(x, y) = |x - y|$. Let F_μ, F_ν denote the distribution functions and define the quantile functions*

$$Q_\mu(u) := \inf\{x \in \mathbb{R} : F_\mu(x) \geq u\}, \quad u \in (0, 1),$$

and similarly for Q_ν . Then

$$W_p^p(\mu, \nu) = \int_0^1 |Q_\mu(u) - Q_\nu(u)|^p du.$$

Theorem 4 (Kantorovich-Rubinstein duality for W_1). *Let $\mu, \nu \in \mathcal{P}_1(\mathsf{X})$. Then*

$$W_1(\mu, \nu) = \sup \left\{ \int f d\mu - \int f d\nu : f : \mathsf{X} \rightarrow \mathbb{R}, \|f\|_{\text{Lip}} \leq 1 \right\},$$

where $\|f\|_{\text{Lip}} := \sup_{x \neq y} \frac{|f(x) - f(y)|}{d(x, y)}$.

Corollary 7 (Lipschitz test functions). *Let $\mu, \nu \in \mathcal{P}_1(\mathsf{X})$ and let $f : \mathsf{X} \rightarrow \mathbb{R}$ be L -Lipschitz. Then*

$$\left| \int f d\mu - \int f d\nu \right| \leq L W_1(\mu, \nu).$$

Proof. If $L = 0$, the claim is immediate. Otherwise, f/L has Lipschitz seminorm at most 1, and Theorem 4 yields

$$\left| \int f d\mu - \int f d\nu \right| = L \left| \int (f/L) d\mu - \int (f/L) d\nu \right| \leq L W_1(\mu, \nu).$$

□

Lemma 8 (Deterministic Lipschitz pushforward). *Let $p \geq 1$, let $\mu, \nu \in \mathcal{P}_p(\mathsf{X})$, and let $g : \mathsf{X} \rightarrow \mathsf{Y}$ be L -Lipschitz between metric spaces $(\mathsf{X}, d_{\mathsf{X}})$ and $(\mathsf{Y}, d_{\mathsf{Y}})$. Then the pushforward measures $g_{\#}\mu$ and $g_{\#}\nu$ belong to $\mathcal{P}_p(\mathsf{Y})$ and satisfy*

$$W_p(g_{\#}\mu, g_{\#}\nu) \leq L W_p(\mu, \nu).$$

Proof. Let $\pi \in \Gamma(\mu, \nu)$ and let $(U, V) \sim \pi$. Then $(g(U), g(V))$ is a coupling of $(g_{\#}\mu, g_{\#}\nu)$ and

$$\mathbb{E}[d_{\mathsf{Y}}(g(U), g(V))^p] \leq L^p \mathbb{E}[d_{\mathsf{X}}(U, V)^p].$$

Taking the infimum over $\pi \in \Gamma(\mu, \nu)$ and then p th roots yields the result. □

A central step in Section 4 uses a coupling where the mapping is Lischitz conditional on a random environment. The following lemma isolates this idea.

Lemma 9 (Random Lipschitz contraction). *Let $p \geq 1$. Let U_1, U_2 be real-valued random variables with laws $\mu_1, \mu_2 \in \mathcal{P}_p(\mathbb{R})$. Let V be an auxiliary random variable taking values in a measurable space \mathbb{V} , independent of (U_1, U_2) . Let $\phi : \mathbb{R} \times \mathbb{V} \rightarrow \mathbb{R}$ be measurable and assume there exists a measurable function $L : \mathbb{V} \rightarrow [0, \infty)$ such that*

$$|\phi(u, v) - \phi(u', v)| \leq L(v) |u - u'| \quad \text{for all } u, u' \in \mathbb{R} \text{ and all } v \in \mathbb{V},$$

and $\mathbb{E}[L(V)^p] < \infty$. Then

$$W_p(\mathcal{L}(\phi(U_1, V)), \mathcal{L}(\phi(U_2, V))) \leq (\mathbb{E}[L(V)^p])^{1/p} W_p(\mu_1, \mu_2).$$

Proof. Let $(\tilde{U}_1, \tilde{U}_2)$ be an optimal coupling of μ_1 and μ_2 . Enlarge the probability space so that V is independent of $(\tilde{U}_1, \tilde{U}_2)$ and has the same law as in the statement. Define $\tilde{Z}_i := \phi(\tilde{U}_i, V)$ for $i \in \{1, 2\}$, so that $\mathcal{L}(\tilde{Z}_i) = \mathcal{L}(\phi(U_i, V))$. By Lemma 6 and the Lipschitz property of ϕ ,

$$W_p(\mathcal{L}(\phi(U_1, V)), \mathcal{L}(\phi(U_2, V))) \leq \mathbb{E}\left[|\tilde{Z}_1 - \tilde{Z}_2|^p\right] \leq \mathbb{E}[L(V)^p] \mathbb{E}\left[|\tilde{U}_1 - \tilde{U}_2|^p\right].$$

Since $(\tilde{U}_1, \tilde{U}_2)$ is optimal, $\mathbb{E}[|\tilde{U}_1 - \tilde{U}_2|^p] = W_p^p(\mu_1, \mu_2)$, and taking p th roots yields the claim. \square

Remark 23. Lemma 9 is applied in Section 4 with random transformations induced by the canonical coupling representation in Section 4.2. In particular, it is used with maps of the form $\phi(u, v) = \frac{uv-1}{\gamma}$ when $\gamma \neq 0$ and $\phi(u, v) = u - \log v$ when $\gamma = 0$.

B.2 Entropic variational identities

Let P be a probability measure on a measurable space $(\mathbb{X}, \mathcal{F})$. The basic variational identity below is the Donsker-Varadhan formula; see Donsker and Varadhan (1975). For $Q \ll P$, define the relative entropy

$$D(Q\|P) := \int \log\left(\frac{dQ}{dP}\right) dQ,$$

and set $D(Q\|P) = +\infty$ if Q is not absolutely continuous with respect to P .

Theorem 5 (Donsker-Varadhan variational formula). *Let $f : \mathbb{X} \rightarrow \mathbb{R}$ be measurable and assume $\int e^f dP < \infty$. Then*

$$\log \int e^f dP = \sup_{Q \in \mathcal{P}(\mathbb{X})} \left\{ \int f dQ - D(Q\|P) \right\}.$$

Define the Gibbs tilt Q^* by

$$\frac{dQ^*}{dP}(x) = \frac{e^{f(x)}}{\int e^f dP}.$$

If, in addition, $\int |f| e^f dP < \infty$ (equivalently, $\int |f| dQ^* < \infty$ and $D(Q^*\|P) < \infty$), then the supremum is attained at Q^* .

Corollary 8 (Entropy penalization in minimization form). *Let $g : \mathbb{X} \rightarrow \mathbb{R}$ be measurable and assume $\int e^{-g} dP < \infty$. Then*

$$-\log \int e^{-g} dP = \inf_{Q \in \mathcal{P}(\mathbb{X})} \left\{ \int g dQ + D(Q\|P) \right\}.$$

Define the Gibbs tilt Q^* by

$$\frac{dQ^*}{dP}(x) = \frac{e^{-g(x)}}{\int e^{-g} dP}.$$

If, in addition, $\int |g|e^{-g} dP < \infty$, then the infimum is attained at Q^* .

Lemma 10 (Entropy penalization with linear constraints). *Let $P \in \mathcal{P}(X)$ and let $g_0, g_1, \dots, g_m : X \rightarrow \mathbb{R}$ be measurable. Fix $\varepsilon > 0$ and constants $c_1, \dots, c_m \in \mathbb{R}$. Assume the feasible set*

$$\mathcal{Q} := \left\{ Q \in \mathcal{P}(X) : \int g_k dQ = c_k, k = 1, \dots, m \right\}$$

is nonempty and contains some $Q_0 \ll P$ with $D(Q_0 \| P) < \infty$ and $\int |g_0| dQ_0 < \infty$. Let

$$\mathcal{C} := \left\{ \left(\int g_1 dQ, \dots, \int g_m dQ \right) : Q \in \mathcal{P}(X), Q \ll P, D(Q \| P) < \infty, \int |g_0| dQ < \infty \right\} \subseteq \mathbb{R}^m.$$

Assume the target moment vector $c := (c_1, \dots, c_m)$ belongs to the relative interior $\text{ri}(\mathcal{C})$. Assume also that the value of the problem

$$\inf_{Q \in \mathcal{Q}} \left\{ \int g_0 dQ + \varepsilon D(Q \| P) \right\}$$

is finite and attained at some Q^* . Then there exist finite Lagrange multipliers $\theta_1, \dots, \theta_m \in \mathbb{R}$ such that Q^* admits the exponential-tilt representation

$$\frac{dQ^*}{dP}(x) = \frac{\exp\left(-\frac{1}{\varepsilon} (g_0(x) + \sum_{k=1}^m \theta_k g_k(x))\right)}{\int \exp\left(-\frac{1}{\varepsilon} (g_0 + \sum_{k=1}^m \theta_k g_k)\right) dP}.$$

Remark 24. Lemma 10 is in fact a general constrained-entropy projection result. Section 5 uses a sharper one-dimensional argument tailored to the mean constraint $\mathbb{E}_F[X] = 1$ so that the existence of a finite dual optimizer can be stated directly in terms of the tilted mean.

B.3 Background on entropic optimal transport

Let $\mu \in \mathcal{P}(X)$ and $\nu \in \mathcal{P}(Y)$ be Borel probability measures on Polish spaces and let $c : X \times Y \rightarrow \mathbb{R} \cup \{+\infty\}$ be a measurable cost. Let $\Gamma(\mu, \nu)$ denote the set of couplings with marginals (μ, ν) . For $\varepsilon > 0$, define the entropic optimal transport problem

$$\inf_{\pi \in \Gamma(\mu, \nu)} \left\{ \int c d\pi + \varepsilon D(\pi \| \mu \otimes \nu) \right\}.$$

We use only structural consequences. The relevant ingredients are the duality and the Gibbs form of optimizers. See, e.g., Léonard (2014), Peyré and Cuturi (2019), Ghosal et al. (2022), and Nutz (2022); and see also Cuturi (2013) for the computational Sinkhorn formulation that made entropy regularization central in applied optimal transport.

Theorem 6 (Entropic optimal transport duality and Gibbs form). *Assume the primal problem is finite for some $\varepsilon > 0$ and that an optimizer exists. Then there exist measurable functions $\varphi : X \rightarrow \mathbb{R}$ and $\psi : Y \rightarrow \mathbb{R}$ such that an optimal coupling π^* admits a density with respect to $\mu \otimes \nu$ of the form*

$$\frac{d\pi^*}{d(\mu \otimes \nu)}(x, y) = \exp\left(\frac{\varphi(x) + \psi(y) - c(x, y)}{\varepsilon}\right),$$

with (φ, ψ) chosen so that $\pi^* \in \Gamma(\mu, \nu)$. Moreover, (φ, ψ) solve a concave dual maximization problem obtained by Fenchel duality.

Remark 25. Note that Theorem 6 is stated at a level sufficient for our purposes. Section 5 uses the same convex-duality logic in a reduced setting where the penalty is relative entropy with respect to a baseline F_0 and the decision variable is the marginal distribution F . In that sense, the planner problem can be read as a one-marginal entropy projection derived from the broader Schrödinger and entropic-transport framework.

References

- ARNOLD, B. C., N. BALAKRISHNAN, AND H. N. NAGARAJA (1998): *Records*, New York: Wiley.
- BALKEMA, A. A. AND L. DE HAAN (1974): “Residual Life Time at Great Age,” *The Annals of Probability*, 2, 792–804.
- BARNDORFF-NIELSEN, O. (1964): “On the Limit Distribution of the Maximum of a Random Number of Independent Random Variables,” *Acta Mathematica Academiae Scientiarum Hungaricae*, 15, 399–403.
- BECKER, L. AND S. MANGIN (2023): “The Effect of Search Frictions on Extreme Outcomes,” Working paper, originally circulated in 2023.
- BOBBIA, B., C. DOMBRY, AND D. VARRON (2021): “The Coupling Method in Extreme Value Theory,” *Bernoulli*, 27, 1824–1850.
- BOLTE, L., N. IMMORLICA, AND M. O. JACKSON (2024): “The Role of Referrals in Immobility, Inequality, and Inefficiency in Labor Markets,” *Journal of Labor Economics*, advance online publication / forthcoming in *Journal of Labor Economics*.
- BUHAI, I. S. AND M. J. VAN DER LEIJ (2023): “A Social Network Analysis of Occupational Segregation,” *Journal of Economic Dynamics and Control*, 147, 104593.
- CALVÓ-ARMENGOL, A. AND M. O. JACKSON (2004): “The Effects of Social Networks on Employment and Inequality,” *American Economic Review*, 94, 426–454.
- CUTURI, M. (2013): “Sinkhorn Distances: Lightspeed Computation of Optimal Transport,” in *Advances in Neural Information Processing Systems 26*, 2292–2300.
- DE HAAN, L. AND A. FERREIRA (2006): *Extreme Value Theory: An Introduction*, New York: Springer.
- DONSKER, M. D. AND S. R. S. VARADHAN (1975): “Asymptotic Evaluation of Certain Markov Process Expectations for Large Time, I,” *Communications on Pure and Applied Mathematics*, 28, 1–47.
- EINMAHL, J. H. J. AND Y. HE (2023): “Extreme Value Estimation for Heterogeneous Data,” *Journal of Business & Economic Statistics*, 41, 255–269.
- FOURNIER, N. AND A. GUILLIN (2015): “On the Rate of Convergence in Wasserstein Distance of the Empirical Measure,” *Probability Theory and Related Fields*, 162, 707–738.
- GALAMBOS, J. (1973): “The Distribution of the Maximum of a Random Number of Random Variables with Applications,” *Journal of Applied Probability*, 10, 122–129.
- GALICHON, A. (2016): *Optimal Transport Methods in Economics*, Princeton, NJ: Princeton University Press.
- GHOSAL, P., M. NUTZ, AND E. BERNTON (2022): “Stability of Entropic Optimal Transport and Schrödinger Bridges,” *Journal of Functional Analysis*, 283, 109622.
- GRANOVETTER, M. S. (1973): “The Strength of Weak Ties,” *American Journal of Sociology*, 78, 1360–1380.
- IOANNIDES, Y. M. AND L. D. LOURY (2004): “Job Information Networks, Neighborhood Effects, and Inequality,” *Journal of Economic Literature*, 42, 1056–1093.
- LÉONARD, C. (2014): “A Survey of the Schrödinger Problem and Some of Its Connections with Optimal Transport,” *Discrete & Continuous Dynamical Systems - A*, 34, 1533–1574.
- MANGIN, S. (2026): “Extreme Value Theory with Heterogeneous Agents,” *Econometrica*, forthcoming.
- MANSANAREZ, P., G. POLY, AND Y. SWAN (2025): “Stein’s Method for Fréchet Approximation: A Regularly Varying Functions Approach,” arXiv preprint arXiv:2510.14016, preprint.
- MEGYESI, Z. (2002): “Domains of Geometric Partial Attraction of Max-Semistable Laws: Structure, Merge and Almost Sure Limit Theorems,” *Journal of Theoretical Probability*, 15, 973–1005.
- MONTGOMERY, J. D. (1991): “Social Networks and Labor-Market Outcomes: Toward an Economic Analysis,” *American Economic Review*, 81, 1408–1418.
- NUTZ, M. (2022): “Introduction to Entropic Optimal Transport,” Lecture notes, available online.
- PEYRÉ, G. AND M. CUTURI (2019): “Computational Optimal Transport,” *Foundations and Trends in Machine Learning*, 11, 355–607.
- PICKANDS, J. (1975): “Statistical Inference Using Extreme Order Statistics,” *The Annals of Statistics*, 3, 119–131.
- RACHEV, S. T. AND S. I. RESNICK (1991): “Max-Geometric Infinite Divisibility and Stability,” *Stochastic Models*, 7, 191–218.
- RESNICK, S. I. (2008): *Extreme Values, Regular Variation and Point Processes*, New York: Springer, 2 ed.
- ROTHSCHILD, M. AND J. E. STIGLITZ (1970): “Increasing Risk: I. A Definition,” *Journal of Economic Theory*, 2, 225–243.

- SANTAMBROGIO, F. (2015): *Optimal Transport for Applied Mathematicians: Calculus of Variations, PDEs, and Modeling*, vol. 87 of *Progress in Nonlinear Differential Equations and Their Applications*, Basel: Birkh" auser.
- SHAKED, M. AND J. G. SHANTHIKUMAR (2007): *Stochastic Orders*, Springer Series in Statistics, New York: Springer.
- SILVESTROV, D. AND J. L. TEUGELS (1998): "Limit Theorems for Extremes with Random Sample Size," *Advances in Applied Probability*, 30, 777–806.
- TOPA, G. (2001): "Social Interactions, Local Spillovers and Unemployment," *Review of Economic Studies*, 68, 261–295.
- VILLANI, C. (2009): *Optimal Transport: Old and New*, vol. 338 of *Grundlehren der mathematischen Wissenschaften*, Springer.