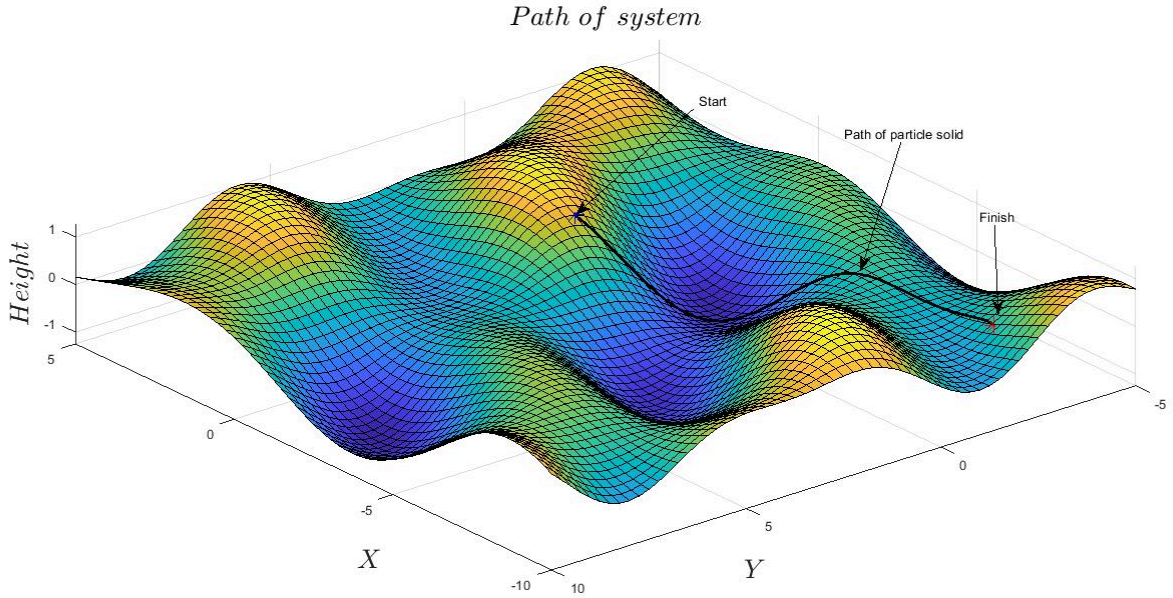
The Dissipative Structure of Value: Information Geometry and the Thermodynamics of NYC Real Estate

On Modelling with Madelung Objects

Kai Cobbs



12/18/2025

Introduction:

In August 2025 David Krakauer, president of the Santa Fe Institute, the world's leading research center on the study of complex systems said this on Neil deGrasse Tyson's *Star Talk* podcast:

"Warren Weaver, who was a mathematician, wanted to classify all of regularity or irregularity in the universe. He said there are simple phenomena; classical physics. It doesn't mean it's easy to understand, but it's simple. There are beautiful laws, elegant mathematical formalisms that explain it.

Then there's the world of what he called disorganized complexity. That's the study of gases, what we would now study with thermodynamics and statistical mechanics; irregular things that have beautiful descriptions.

And then in the middle, there's organized complexity. It's not a gas. It's an organism. It's an ant. It's a brain. It's a city. And in that space, what we realized really at the end of the 19th century is that we don't have good theories for it. We can do irregularity beautifully. We can do simplicity beautifully. But then you move to everything that we actually care about, in some sense as human beings, as animals, as living beings, and we kind of under-theorize it."

In many ways this describes my line of inquiry, the broad understanding of complexity through a novel framework descended from information theory, information geometry, thermodynamics, econometrics, and quantum hydrodynamics. Information theory, having originated from Claude Shannon, is the field of mathematics of quantifying information itself as a measurable physical quantity, at its core seeking to answer; "How much "surprise" or "uncertainty" is contained in a signal, and how efficiently can we encode it?". Information Geometry is the study of probability distributions as physical shapes, a fusion between differential geometry (the calculus of curved spaces, called manifolds, that look flat when zoomed in on, like the Earth) and statistics. Thermodynamics being the study of energy and entropy. Econometrics as the general science of causal modelling through hypothesis testing and regressions.Quantum mechanics models the microscopic word of particles as waves of probability rather than discrete points, breaking the traditional determinism of classical physics. Core to this is Schordinger's equation describing particles as wave packets through complex numbers. Erwin Madelung in 1926, the same year Schrodinger published his equation, released a version of it rewritten in hydrodynamic variables, effectively describing the quantum world as a flowing probability fluid rather than Schrodinger's amorphous and hard to conceive complex valued wave packets. My work combines these fields, often siloed into one continuous field-like theory, fields being a mathematical object that encodes mathematical properties over a space/manifold. When applied to economic data, it constitutes a new theory of the economy based on these principles. 2025 marks the 100th anniversary of the study of quantum mechanics. It excites me to present contributions related to this field that can be applied to macro phenomena.

The Data:

To establish a comprehensive dataset of New York City property transactions, I developed a custom automated web crawler to harvest historical Rolling Sales data directly from the NYC Department of Finance. The script utilized a "shotgun" request pattern to handle inconsistent URL structures and file formats (both .xls and .xlsx) across the target period of 2015 to the end of 2024. This process iteratively retrieved annualized sales files for all five boroughs, standardizing them into a unified structure despite varying header formats over the decade.

Once acquired, I subjected the raw data to a rigorous cleaning and harmonization pipeline to ensure the integrity of the thermodynamic variables. Non-numeric characters such as currency symbols and

commas were stripped from the Price, BuildingArea, and LandArea fields, which I then coerced into numeric types. To isolate genuine market activity from non-arms-length transfers, such as familial deed transfers or administrative corrections, I applied a strict filter excluding records with a sale price below $1,000 or a gross building area under 100 square feet. I subsequently concatenated the cleaned extracts into a single master dataframe, nyc_real_estate_history(all).csv, providing a continuous and granular field of valid market transactions for the subsequent geometric analysis. The data set was also adjusted for inflation (September 2025 dollars).

The Madelung Object Model:

The cornerstone of this analysis is what I term the "Madelung object", a continuous-time kernel density estimator (KDE) that constructs a smooth, differentiable probability field over the space of log-price, log-building-size, and time. Concretely, for any query time t, the algorithm estimates the joint probability density $\rho(p, s, t)$ of observing a transaction at price level p and building size s.

This is not a static histogram; it is a temporally weighted estimation. The class 'ContinuousTimeKDE' implements a sliding temporal Gaussian kernel that weighs historical transactions based on their proximity to t. This allows the probability distribution to evolve smoothly, "flowing" from one market regime to another without the jitter of discrete windowing.

The "Madelung" naming reflects the mathematical utility of this object. Just as Erwin Madelung reformulated the Schrödinger equation into hydrodynamic variables, expressing quantum mechanics as a probability fluid, this estimator reframes market dynamics as a fluid of transaction probability. The density ρ represents market liquidity, while its evolution encodes the changing preferences of market actors.

Crucially, because the KDE is built from Gaussian basis functions, we can calculate its derivatives analytically rather than numerically. My implementation explicitly computes the spatial gradients ($\nabla\rho$) and the time derivative ($\frac{\partial\rho}{\partial t}$) at every point in the grid. These raw kinematic variables allow us to invert the Fokker-Planck equation:

$$\frac{\partial\rho}{\partial t} + \nabla \cdot (\rho\mathbf{v}) = \nabla \cdot (\mathbf{D}\nabla\rho)$$

Instead of assuming a model and simulating forward, we take the observed ρ and its derivatives (the "Madelung object") as the ground truth. We then solve this continuity equation to extract the hidden dynamic variables: $\mathbf{v}$ represents the drift velocity (systematic flow of transaction probability toward certain price-size combinations), and $\mathbf{D}$ is the diffusion tensor (stochastic fluctuations in market microstructure). Economically, the drift component captures systematic market behavior, how prices adjust and building heterogeneity evolves, while the diffusion component quantifies market friction and uncertainty. The elegance of this formulation is that it requires no traditional assumptions about demand curves, equilibrium conditions, or behavioral hypotheses: it describes what the market *actually does* through its revealed transaction pattern, letting the data speak through its own geometric structure.

Computational Implementation

To make this mathematically tractable over a dataset of this magnitude, I developed two parallel implementations of the estimator:

CPU Implementation (NumPy/SciPy):
For smaller subsets or rapid prototyping, this version utilizes scipy.stats.gaussian_kde with custom temporal weighting. It leverages standard linear algebra libraries to compute covariance matrices and handle the weighted kernel summation efficiently.

GPU Acceleration (PyTorch):
To analyze the full high-resolution field (100×100 spatial grids over 3,650 days), the computational cost of evaluating Gaussian kernels scales quadratically ($N_{grid} \times N_{data}$). I implemented a hardware-accelerated version using PyTorch that treats the KDE summation as a massive tensor operation. This allows the pairwise distance computations and exponentiations to be broadcast across thousands of CUDA cores simultaneously, reducing the computation time for the full 10-year "Madelung object" generation from hours to minutes.

For reference all of the computation times in this research are manageable without GPU assistance, with the exception of the bandwidth optimization that I am about to cover. The GPU assistance was specifically added because that calculation would take hours, maybe even more than a day to complete with the CPU alone.

On Eliminating Look-Ahead Bias, The Causal Kernel

A critical structural refinement was necessary to ensure the validity of predictive experiments. The standard Continuous-Time KDE utilizes a symmetric Gaussian kernel in the temporal dimension, which effectively smooths the manifold for geometric visualization by weighting contributions from both past and future transactions. While this bidirectional smoothing is ideal for mapping the static structure of the market phase space and descriptive analysis of dynamics, it introduces a fatal look-ahead bias for predictive analysis. If the density estimate at a given time includes information from future price shocks, the resulting metrics would implicitly contain the very signal they purport to predict, rendering any causality tests invalid.

To resolve this, the Madelung object was re-engineered to support a strict causal mode for the time-series analysis. This adaptation forces the estimator to respect the arrow of time by reconstructing the probability density at any historical date using only the transaction data available up to that specific moment. This distinction allows us to retain the high-fidelity geometric visualizations for structural analysis while guaranteeing the integrity of the econometric validation.

Bandwidth Optimization
The central challenge in constructing a continuous-time probability field is determining the optimal temporal bandwidth (σ_t). This parameter effectively sets the "shutter speed" of the observation. If σ_t is too short, the model captures transient microstructure noise (high variance); if too long, it oversmooths the rapid phase transitions characteristic of market regime shifts (high bias). Note that because the underlying price data was already inflation-adjusted, this smoothing concern pertains to structural volatility rather than monetary inflation itself.

To rigorously select this parameter, I developed an optimization framework rooted in Information Geometry. Rather than relying on heuristic rules of thumb, I sought the bandwidth that maximized the

structural resolution of the phase space, specifically, the clarity of the relationship between Market Liquidity (ρ) and Kinematic Volatility ($|\mathbf{v}|$).

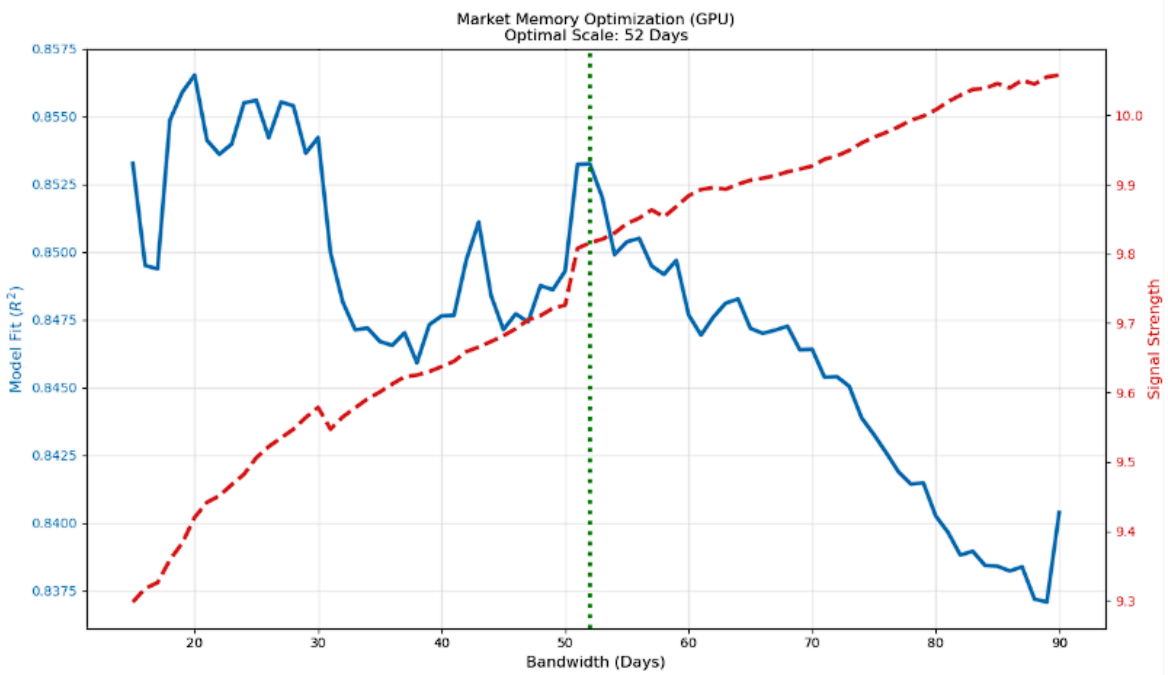
The "Frontier Extraction" Method

I implemented a GPU-accelerated sweep across a bandwidth spectrum of 15 to 90 days. For each candidate σ_t, I generated the full 4-dimensional phase space and extracted the "Constraint Frontier", defined as the 99th percentile of velocity magnitude within each density bin.

The logic of this optimization rests on the Cramér-Rao lower bound (from information geometry) and offers a direct economic analogue to Heisenberg's Uncertainty Principle. In quantum mechanics, one cannot simultaneously know a particle's position and momentum with infinite precision ($\Delta x \Delta p \geq \hbar/2$). Similarly, in market dynamics, we face a fundamental trade-off: as we narrow the temporal window ($\Delta t \to 0$) to capture instantaneous velocity ("momentum"), the statistical sample vanishes, and the uncertainty in the density manifold ("position") explodes. Conversely, expanding the window minimizes statistical error but blurs the velocity vector, losing the signal of the regime shift. By maximizing the clarity of the velocity-density frontier, I am effectively finding the "minimum uncertainty state" for this economic system, the scale where the trade-off between temporal resolution and statistical precision is optimized.

The optimization utilized a composite objective function:

$$\text{Score} = \text{Normalized}(R^2) + \text{Normalized}(\text{Signal Strength})$$

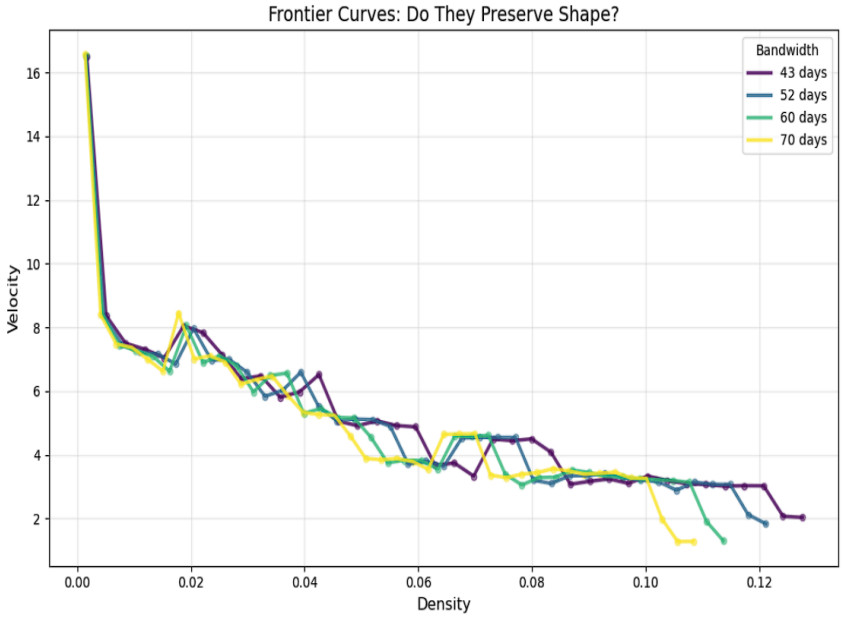


Here, R^2 measures how well the frontier fits a theoretical exponential decay model (validating that the observed structure obeys thermodynamic constraints), while "Signal Strength" measures the raw magnitude of the detected velocity. The algorithm converged convexly on a global optimum of **52 Days**. This value represents the information-theoretic coherence time of the NYC housing market. In other words the average shelf-life of price information; the duration for which a single transaction remains a valid signal for other buyers before fading into 'stale' historical noise. This likely corresponds to contract close latency.

Robustness and Invariance Checks

To ensure that this 52-day scale reflected a genuine property of the market rather than an artifact of the estimator, I performed a "Directional Consistency" test. I generated independent vector fields using bandwidths of 43, 52, 60, and 70 days and calculated the Cosine Similarity of the resulting drift vectors $\mathbf{v}$ at every point in the spatial grid (50×50).

The results demonstrated extraordinary geometric invariance:




```
================================================
DIRECTIONAL AGREEMENT (Cosine Similarity)
================================================
43 vs 52 days:
  Mean cosine similarity: 0.990
  Fraction agreeing (cos > 0.5): 99.3%

43 vs 60 days:
  Mean cosine similarity: 0.976
  Fraction agreeing (cos > 0.5): 99.1%

43 vs 70 days:
  Mean cosine similarity: 0.958
  Fraction agreeing (cos > 0.5): 98.4%

52 vs 60 days:
  Mean cosine similarity: 0.996
  Fraction agreeing (cos > 0.5): 100.0%

52 vs 70 days:
  Mean cosine similarity: 0.982
  Fraction agreeing (cos > 0.5): 99.5%

60 vs 70 days:
  Mean cosine similarity: 0.994
  Fraction agreeing (cos > 0.5): 99.8%
```

The fact that the vector fields remain directionally locked (mean cosine similarities all exceed 0.95) despite significant changes in bandwidth confirms that the "Madelung Flow" is a robust, physically real feature of the data. The 52-day optimization does not "create" the pattern; it simply brings an existing, invariant geometric structure into maximum focus. This exercise beautifully highlights the parallels between the Cramer-Rao Bound, Heisenberg's Uncertainty, and the bias-variance trade off.

<u>The Heisenberg Limits of Real Estate</u>

Having calibrated the thermodynamic instrument, I turned to the analysis of the phase space itself. The Madelung object provides two fundamental scalar fields at every coordinate (p, s) in the market:

Market Certainty (Z): The probability density, representing the consensus on value. High density indicates a "crowded" trade where buyers and sellers agree on price.

Liquidity Velocity ($|\mathbf{v}|$): The magnitude of the drift vector, representing the speed of price discovery or the "momentum" of the asset class.

The Uncertainty Principle of Markets

When visualizing these two variables against each other for any given date, a striking geometric constraint emerges as the economic analogue to Heisenberg's Uncertainty Principle.

In quantum mechanics, the precision of position and momentum are inversely related ($\sigma_x \sigma_p \geq \frac{\hbar}{2}$). In the NYC real estate market, we observe a similar fundamental trade-off between Consensus and Velocity.

The Attractor State: Regions of high density (high certainty) effectively "trap" prices. When market participants broadly agree on a value, the "mass" of the consensus creates inertia, preventing rapid price movement.
The Rapid State: Conversely, high velocity is only permissible in regions of low density (low certainty). Rapid price discovery requires a lack of established consensus.

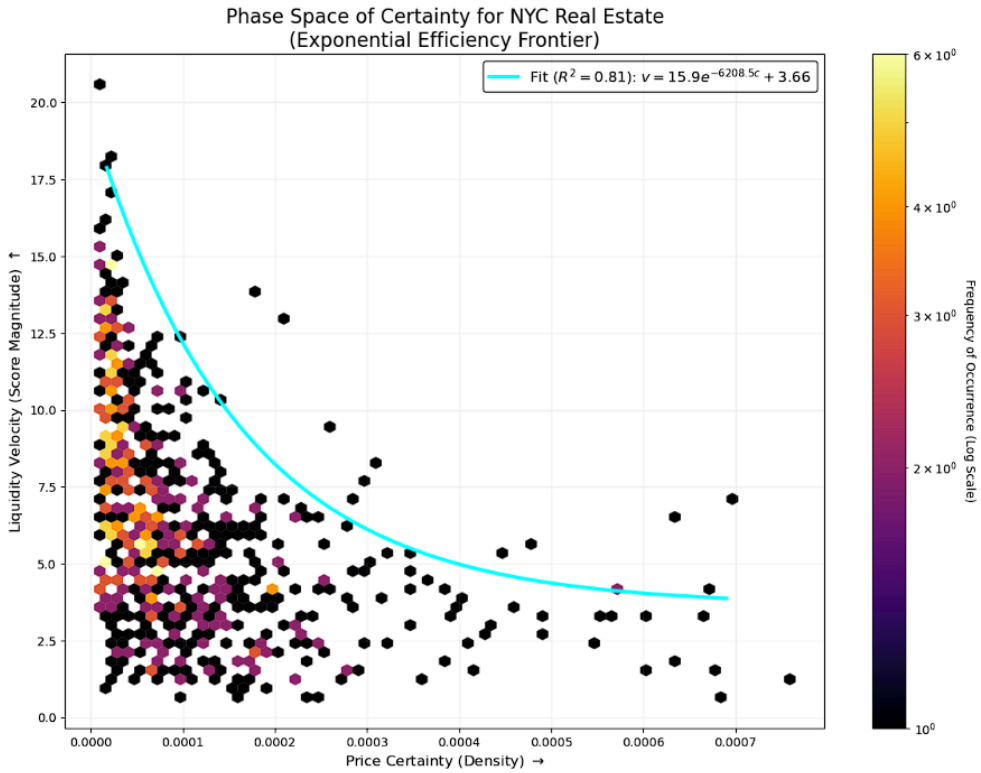
This is not merely a correlation; it is a "forbidden zone" geometry. While outliers exist, the vast majority of transactions adhere to a characteristic efficiency frontier. Beyond this curve, the market cannot physically sustain velocity given the weight of the consensus.

The Aggregated Efficiency Frontier

To quantify this limit, I aggregated the phase space data across 40 timestamps (spanning 2015–2024), effectively "superpositioning" the market's history to outline the empirical hard edge of the distribution. I extracted the 99th percentile of velocity within each density bin and fitted an exponential decay model to this frontier:

$$v_{max} = A \cdot e^{-B \cdot Z} + D$$

The global fit yielded a robust coefficient of determination ($R^2 = 0.8125$), confirming that this trade-off is a structural law of the market rather than transient noise.



Global Frontier Metrics:

- **Equation:** $v = 15.87 \cdot e^{-6208.47 \cdot Z} + 3.66$
- **RMSE:** 1.8369

This equation implies a "terminal velocity" for real estate. Even in the most illiquid, chaotic corners of the market (where $Z \to 0$), the speed of price change is capped at approximately **19.5 units**. Conversely,

as consensus forms (Z increases), the maximum possible velocity decays exponentially toward a friction floor of 3.66.

Temporal Stability and Parametric Drift

To test the invariance of this law, I analyzed the frontier during distinct macroeconomic regimes. While the functional form (exponential decay) remained robust, the specific parameters exhibited a measurable "wobble," reflecting the changing rigidity of the market.

Period A: Pre-Pandemic Stability (2017–2018)

- **Equation:** $v = 16.05 \cdot e^{-23462.00 \cdot Z} + 3.48$
- **Max Speed:** ~19.53
- **Decay:** -23,462

Period B: Pandemic Shock (2020–2021)

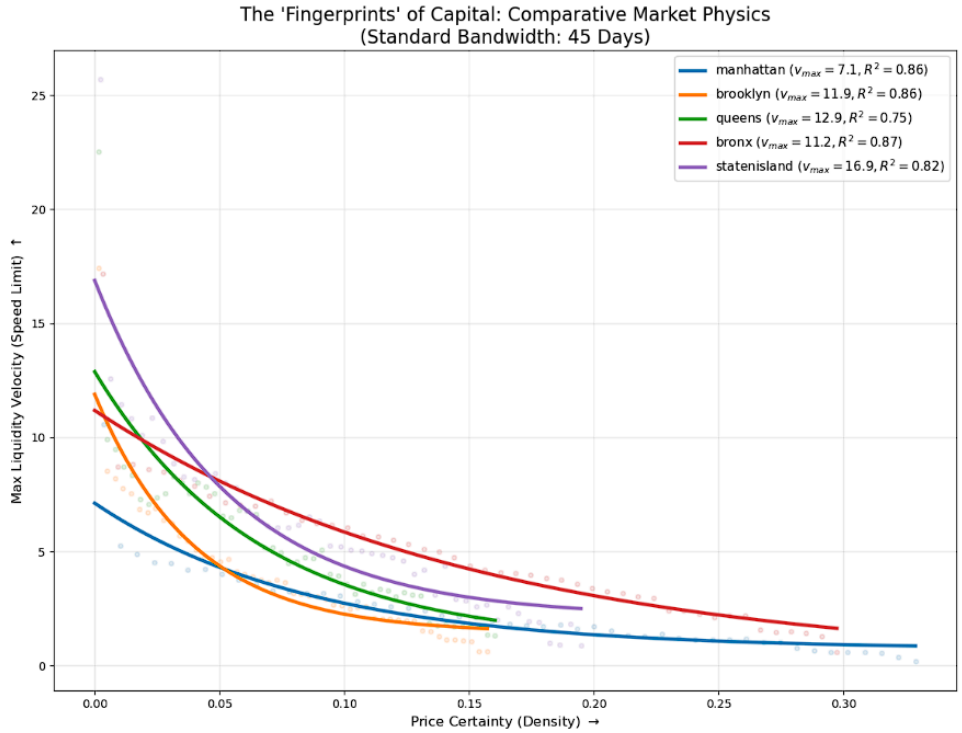
- **Equation:** $v = 11.09 \cdot e^{-8048.87 \cdot Z} + 4.10$
- **Max Speed:** ~15.19
- **Decay:** -8,049

This comparison reveals a fascinating physical transformation. During the stable period (2017–2018), the market was "stiff": the decay coefficient was massive (-23,462), meaning velocity evaporated almost instantly the moment consensus began to form. During the pandemic (2020–2021), the market became "mushier." The maximum speed dropped to 15.19, but the decay was far shallower (-8,049), suggesting that even as consensus formed, higher volatility persisted longer than usual. This parametric drift implies that while the form of the law is universal, its scale is not. Unlike the fixed Planck constant of quantum mechanics, the market's 'minimum unit of action' is a dynamic variable, expanding and contracting with the macroeconomic temperature.

The Fingerprints of Capital: Borough Analysis

While the *shape* of this law is universal, the *parameters* differ by geography. By isolating the dataset by borough and fitting the same Heisenberg frontier to each, I extracted the distinct "physical constants" of these sub-markets.

- **Max Speed ($A + D$):** The theoretical speed limit of the market in a vacuum (zero consensus).
- **Fragility (Decay B):** How quickly liquidity freezes as consensus forms. High fragility means the market calcifies rapidly.
- **Friction (D):** The base level of resistance that persists even at high velocity.



BOROUGH	MAX SPEED	FRAGILITY (Decay)	FRICTION (Floor)	R^2
manhattan	7.11	11.6	0.71	0.861
brooklyn	11.89	25.4	1.42	0.860
queens	12.88	15.2	0.95	0.752
bronx	11.18	6.5	0.00	0.869
statenisland	16.88	18.9	2.12	0.819

Manhattan: The Heavy Fluid (High Inertia)
Manhattan acts as a heavy fluid with the lowest Max Speed (7.11), indicating extreme inertial mass due to capital density. Even in a vacuum of consensus, the sheer value of assets creates a viscosity that physically prevents rapid acceleration. It appears stable not through efficiency, but through the immense energy required to move prices in either direction.

Brooklyn: The "Snap-Freeze" Phase Transition
Defined by extreme Fragility (25.4), Brooklyn mimics a system prone to instantaneous phase transitions where liquidity collapses the moment consensus forms. This "snap-freeze" behavior reflects a market driven by intense trend following. Unlike the gradual cooling of other boroughs, it locks into rigid price structures immediately upon establishing a baseline.

The Bronx: The Frictionless Anomaly
The Bronx behaves like a superconductor with a unique Friction Floor of 0.00 and the lowest Fragility (6.5). This laminar flow allows momentum to persist longer than anywhere else before settling into a perfect standstill. It stands as the most thermodynamically efficient borough where consensus does not impede flow as drastically as in capital-dense regions.

Staten Island: High-Energy Chaos
As the energetic opposite of Manhattan, Staten Island acts as a "light" fluid capable of explosive price discovery ($v_{max} = 16.88$). However, this comes with high Friction (2.12), reflecting a market of "wares" (unique single-family homes) rather than "shares" (fungible condos). This asset heterogeneity generates

irreducible thermal noise, ensuring that even under consensus, the thermodynamic temperature never reaches absolute zero.

Queens: The Hybrid State

Queens presents the lowest goodness-of-fit ($R^2 = 0.752$) because it does not behave as a single thermodynamic material. It likely represents a superposition of two distinct phases of matter, mixing high-density urban centers like Long Island City with vast suburban tracts. This duality muddies the geometric signal, creating a hybrid profile with moderate speed and fragility.

Configurational Entropy:

The Heisenberg analysis in the previous section quantified the kinematics of the boroughs; how fast prices move under varying degrees of consensus. However, to fully understand the material properties of these sub-markets, I also needed to quantify their static structure. To do this, I calculated the Shannon Entropy (S) of the probability density field for each borough. This metric serves as a proxy for Configurational Entropy, measuring the diversity and disorder of the housing stock itself.

Methodology

I utilized the same continuous-time KDE model to generate a probability density surface $\rho(x, y)$ for each borough. To ensure thermodynamic consistency and comparability across regions, I evaluated every borough on the Universal Phase Space grid defined in the global analysis.

The continuous density field was discretized into a probability mass function $P(x_i, y_j)$ such that $\sum P = 1$. The structural entropy was then calculated using the standard Shannon formulation:

$$S = -\sum_{i,j} P(x_i, y_j) \ln P(x_i, y_j)$$

A high entropy value indicates a "disordered" market occupying a vast region of the phase space (e.g., a mix of micro-studios, luxury penthouses, and brownstones). A low entropy value indicates a "crystalline" market, where transactions are tightly clustered around a specific asset type (e.g., standardized single-family homes). The analysis revealed a stark hierarchy of disorder across the city. Manhattan exists in a state of extreme entropic disarray, while Staten Island represents a highly ordered lattice.

Configurational Entropy by Borough:

	Borough	Entropy (S)	Peak Density
0	manhattan	5.656414	0.177895
1	brooklyn	4.364269	0.631871
2	bronx	4.002923	1.146864
3	queens	3.760683	1.016924
4	statenisland	3.544893	1.238102

This structural data provides the critical context needed to interpret the Heisenberg results. It resolves the apparent paradox of why Manhattan (High Entropy) moves slowly, while Staten Island (Low Entropy) moves fast.

Manhattan (The Viscous Glass): Manhattan combines high structural disorder ($S = 5.66$) with high inertia. Physically, this describes a Viscous Glass. Like a glass, it lacks a defined crystalline structure (diverse assets), but its viscosity prevents it from flowing freely. It is frozen by its own complexity.

Staten Island (The Hot Crystal): Staten Island combines high structural order ($S = 3.54$) with high kinetic velocity. Physically, this describes a Hot Crystal. The lattice is uniform (standardized housing stock), but the individual particles are vibrating violently (high price volatility).

The Bronx (The Superfluid): The Bronx exhibits the ordered structure of a crystal ($S = 4.00$) but the flow of a superfluid (0.00 friction). This suggests a Laminar Flow regime where standardization allows for frictionless price discovery.

Brooklyn (The Supercooled Liquid): Brooklyn sits in the middle of the entropy spectrum ($S = 4.36$) but has the highest Fragility (25.4). It behaves like a Supercooled Liquid, a state that is technically fluid but prone to "snap-freezing" into a rigid solid structure the moment a trend appears. The moderate asset diversity allows for both fluid price discovery and sudden rigid consensus.

Queens (The Polycrystalline Hybrid): Queens remains the anomaly with the lowest goodness-of-fit ($R^2 = 0.75$). Its low entropy ($S = 3.76$) suggests order, but its mechanics are inconsistent. It likely behaves as a Polycrystalline Solid, a material composed of many small, distinct ordered domains (e.g., rigid Long Island City condos vs. fluid Jamaica estates) that do not align along a single thermodynamic grain, scattering the geometric signal.

Visualization of the Probability Field

With the mathematical framework established and calibrated, the next step of the pipeline involves rendering the Madelung Object $M(x)$ into a human-interpretable format. I developed a suite of visualization scripts using Plotly (for interactive HTML exploration) and Matplotlib (for high-resolution GIFs) to map the evolution of the probability fluid over the decade.

The Flowing Market (4D Surface Animation)

The primary visualization renders the scalar field $\rho(p, s, t)$ as a dynamic 3D surface, effectively visualizing the real-time flows of liquidity through the market.

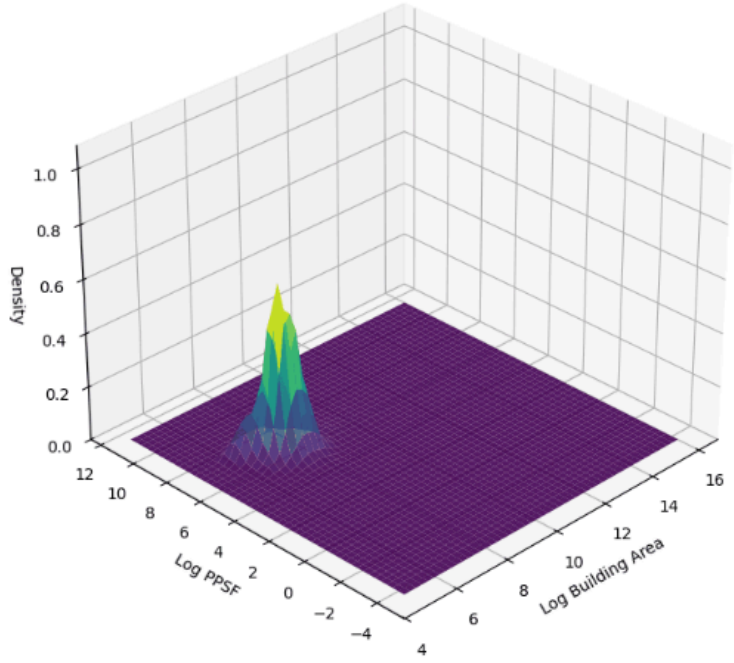
Grid Construction: I defined a 50×50 mesh grid spanning the log-space of Building Area (X) and Price Per Square Foot (Y).

Time-Stepping: The script iterates through the dataset in 14-day increments. For each time step, the 'ContinuousTimeKDE' object (tuned to $\sigma_t = 52$ days) calculates the density Z at every grid point using the weighted Gaussian summation.

Rendering: The Z-axis represents Market Consensus (Transaction Density).

- Peaks (High Z): Correspond to "Attractor" prices, regions of high liquidity where the volume of transactions is dense. These are the "popular" trades (e.g., standard condos in mid-tier neighborhoods).
- Valleys (Low Z): Represent illiquid or forbidden zones where capital rarely flows.

- Motion: As the animation plays, we see the probability mass "slosh" and migrate. This visualization directly confirms the "Parametric Drift", showing the market physically expanding and contracting; showing the flows of liquidity throughout the market over time.

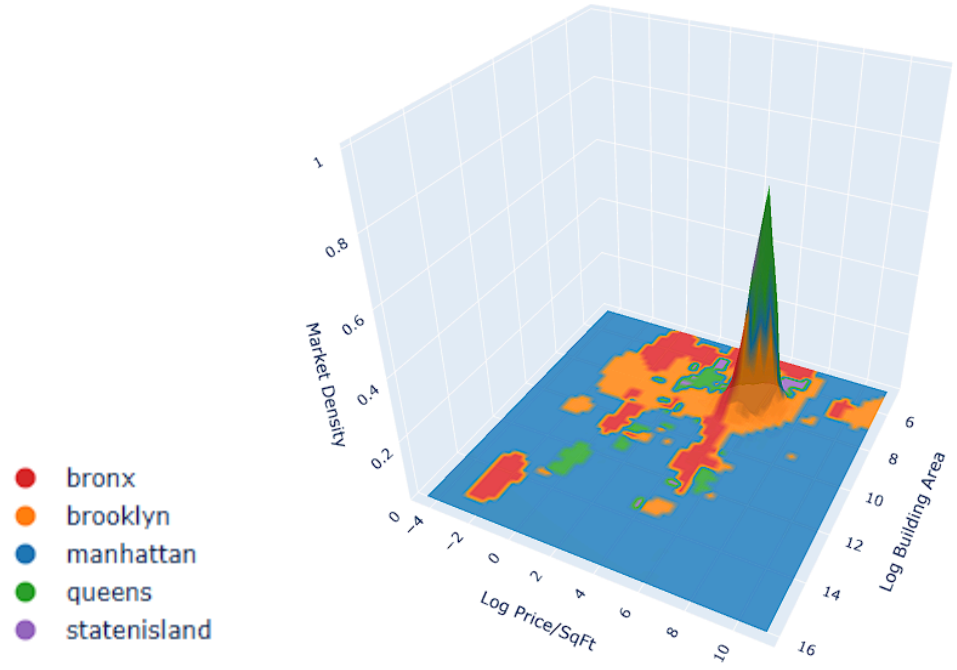


Borough Territories (Multi-Class Topology)

To visualize the structural heterogeneity, I implemented a Multi-Class KDE engine. This script creates a topological map of "Borough Dominance," showing which geographic sub-market controls which region of the price/size phase space.

- The Global Manifold: First, the script calculates a global density surface Z_{total} using all transactions, regardless of location. This defines the overall shape of the liquidity mountain range.
- The Winner-Takes-All Color Map: For every point on that grid, the script calculates the localized density of each individual borough. It then applies a np.argmax function to determine which borough possesses the highest probability density at that specific price/size coordinate.
- Surface Mapping: The visual output plots the global Z surface but colors the surface of the manifold based on the dominant borough.

The result is a 3D territorial map. We can see how all boroughs share the and swap pieces of the peak, with Manhattan dominating the areas outside of it with low transaction occurrence, confirming the high entropy measured in that borough earlier.



The Visualization of Forces

While the probability density ρ tells us where the market is, it is the hydrodynamic derivative fields that tell us what the market is doing. In the Madelung picture, the evolution of the market is governed by a continuity equation:

$$\frac{\partial \rho}{\partial t} + \nabla \cdot (\rho \mathbf{v}) = 0$$

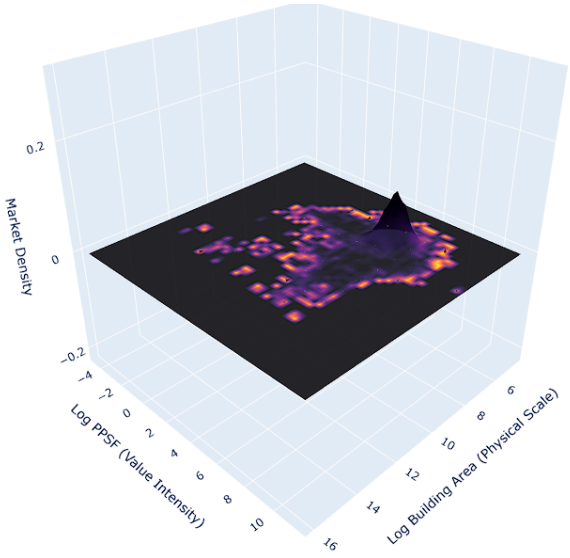
Here, the velocity field $\mathbf{v} = \nabla \log \rho$ represents the "drift" of the probability mass. I developed a series of visualizations to isolate and interpret the specific components of this drift.

The Metric Instability Field ($|\mathbf{v}|^2$)

The first derived quantity is the squared magnitude of the velocity vector, $|\mathbf{v}|^2$. In information geometry, this is proportional to the trace of the Spatial Fisher Information Metric. Think of Fisher Information as reaching into a bag of mixed marbles: if the mix is 50/50, pulling out a red marble tells you very little (low information, flat gradient); but if the bag is 99% blue, finding a red marble is a massive signal (high information, steep gradient). Unlike traditional FIM, which measures how sensitive the data is to hidden *parameters* (the mix ratio), the Spatial FIM measures how sensitive the probability is to location, specifically, how rapidly the market regime shifts as you move through price/size space. Because our velocity vector is defined as the gradient of the log-density ($\mathbf{v} = \nabla \log \rho$), squaring this vector ($|\mathbf{v}|^2$) naturally recovers the Fisher Information metric, confirming that 'speed' in this model is literally the speed of information update.

Code Implementation: I calculated the analytic derivatives U and V from the KDE and computed their squared sum at every grid point. To ensure visual clarity, I clipped the values at the 98th percentile; without this, a single extreme outlier (a massive price jump) would "wash out" the color scale, hiding the structural details of the rest of the market.

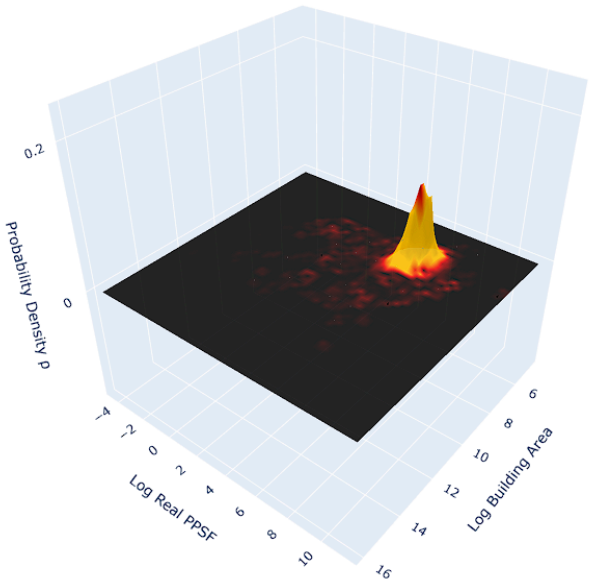
Physical Interpretation: I termed this quantity "Metric Instability." It measures the local speed of the log-likelihood surface. Regions with high instability are zones of Rapid Information Update. Mathematically, these are areas where small perturbations in position result in massive changes in probability; effectively, the "event horizons" of price discovery.



Kinetic Energy Density ($\rho|\mathbf{v}|^2$)

While Metric Instability tells us where the potential for movement is high, it does not tell us where the economic work is actually being done. For that, we need the Kinetic Energy Density, calculated by weighting the instability by the mass of the market:

$$K(x) = \rho(x)|\mathbf{v}(x)|^2$$

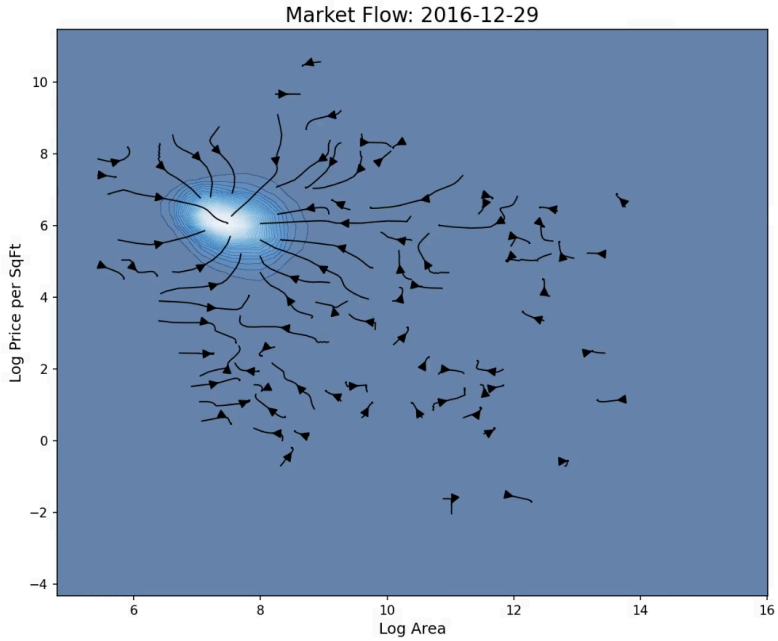


The "Ring of Fire": The visualization revealed a fascinating structural phenomenon.
The Core (Pale Peak): The absolute center of the density mountain (high ρ) actually has low kinetic energy. This is the "dead center" of consensus; transactions are frequent, but they occur at a stable, agreed-upon price.

The Periphery (Dark): The far edges have high velocity (instability) but zero density. No work is done here because there are no transactions to carry the energy.

The Shoulder (Bright Ring): The Kinetic Energy peaks in a distinct ring surrounding the core. This is the Active Transport Zone. It is the boundary where substantial transaction volume meets substantial price disagreement. This visualization confirms that the "thermodynamic work" of the market, the actual effort of price discovery, occurs not at the stable center, but at the negotiating fringe.

Streamlines of the Spatial FIM



To visualize the directionality of these forces, I mapped the gradient field $\nabla \log \rho$ using 2D streamlines.

Code Implementation: Using matplotlib.streamplot, I integrated the velocity vectors (U, V) to trace the "currents" of the probability fluid.

Geometric Interpretation: They show the natural "downhill" paths of the market. In a perfectly efficient market, these lines would converge radially on the attractor. In the NYC data, however, we see distinct vortices and "shear zones" where the flow of capital is turbulent, particularly in the transition zones between boroughs (e.g., the boundary between high-end Brooklyn and low-end Manhattan).

The Quantum Potential as Information Curvature

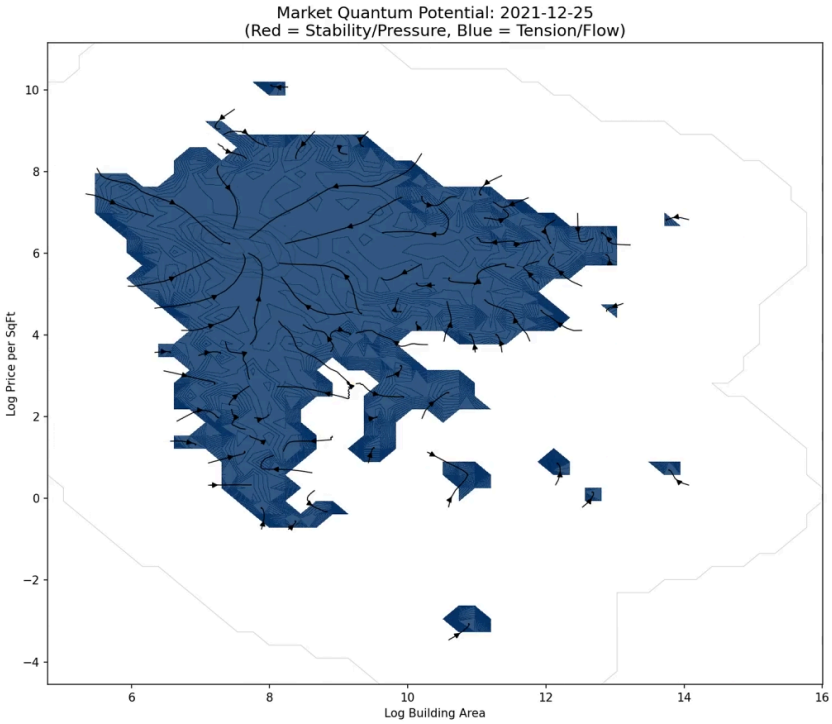
Finally, I visualized the **Quantum Potential (Q)**, revealing its explicit identity as a geometric object. In the Madelung formulation, Q is not a mysterious force, but an algebraic functional of the score function $\mathbf{s} = \nabla \log \rho$:

$$Q(x) \propto -\left(\frac{1}{2}|\mathbf{s}|^2 + \nabla \cdot \mathbf{s} \right)$$

This derivation makes the connection to Information Geometry explicit:

1. The first term, $|\mathbf{s}|^2$, is exactly the Spatial Fisher Information Metric (FIM) density.
2. The second term, $\nabla \cdot \mathbf{s}$, is the divergence of the score (the Laplacian of the log-density).

Thus, the Fisher Information is effectively the unscaled, divergence-free component of the Quantum Potential. Think of Q as the 'cost of certainty'; just as you cannot squeeze a balloon into a single point without it pushing back, Q is the repulsive force that engages whenever the market tries to form a perfect consensus, preventing the probability cloud from collapsing into a mathematical singularity.



This analysis reveals the New York City housing market is always in a flow state.

Note on Scaling: While the geometry is precise, interpreting the absolute magnitude of Q requires a coupling constant (an effective Planck parameter $\hbar_{eff}$). In a socio-economic system, this parameter is not a universal constant but a local, time-dependent scalar representing the market's "volatility temperature." Because computing this variable coupling is beyond the scope of this initial study, Q is presented here to demonstrate the theoretical link between Bohmian mechanics and Information Geometry, rather than as a calibrated signal.

Theoretical Extension: Lie Derivatives and Noether Symmetries

The construction of a differentiable manifold theoretically permits a rigorous search for hidden conservation laws using Lie Derivatives ($\mathcal{L}_{\mathbf{v}}$). By computing how the Fisher Metric g distorts as it flows along the velocity field, one can hunt for Noether Symmetries; mathematical proofs that a quantity (like "market energy") is conserved over time. Specifically, a vanishing Lie Derivative ($\mathcal{L}_{\mathbf{v}} g = 0$) would identify a "Killing Vector Field," implying that the market's information geometry is invariant under evolution. Conversely, a non-zero result quantifies the system's dissipation or entropy production. While preliminary experiments were conducted to test these geometric symmetries, they remain incomplete and were excluded from this analysis for the sake of brevity. However, this line of inquiry is of an extreme significance; these tests can reveal the fundamental rules a system follows, or has followed. The following article explains the concept quite well in an accessible way: https://www.quantamagazine.org/what-are-lie-groups-20251203/

The Bataillean Decomposition

The final experiment of this study bridges the gap between thermodynamic physics and heterodox economics. I would consider it the crown jewel of what I have so far. It is inspired by the work of French philosopher Georges Bataille, whose 1949 treatise *The Accursed Share* proposed a “solar economy” rooted in the physical reality that the sun floods the earth with a constant, unreciprocated superabundance of energy, implying that the fundamental drive of any system, biological or economic, is not the accumulation of scarce resources, but the necessary dissipation of inevitable excess. He categorized this expenditure into two forms:

Growth (Structure): Energy used to expand the system or maintain its form.

Waste (Luxury): The "accursed share" of energy that cannot be absorbed by growth and must be catastrophically expended (through luxury, war, or non-productive consumption) to prevent the system from exploding.

While dismissed by mainstream economists of his day, Bataille's intuition aligns remarkably with modern Stochastic Thermodynamics. In the Fokker-Planck framework, the total entropy production of a system can be mathematically decomposed into two distinct components: the "Housekeeping Heat" (energy dissipated to maintain non-equilibrium steady states) and the "Excess Heat" (energy dissipated during transitions between states).

Mathematical Formulation

I formalized Bataille's philosophy into a rigorous "Flux Decomposition" of the Madelung probability current. Using the extracted diffusion coefficient D, I separated the total velocity field $\mathbf{v}$ into two orthogonal vectors; flow driven by density gradients (like water flowing downhill) versus circular flow that maintains disequilibrium (like water swirling in an eddy). Crucially, the vector sum of these components reconstitutes the total observed flow, meaning the system's total kinetic energy is simply the additive sum of its structural work and its circulatory waste. Mathematically, this requires constructing a scalar potential field φ from the observed probability current $\mathbf{J} = \rho\mathbf{v}$. It is important to note that the diffusion estimation utilizes full historical context during training to establish a stable baseline, but enforces a strict 'past-data only' constraint during testing to prevent data leakage. To ensure the decomposition captures the market's true central tendency, I applied a 1%–99% outlier truncation to the Area and Price domains. By excluding these extreme tails, we isolate the structural core of the market distribution from idiosyncratic boundary events.

The challenge is solving the Poisson equation $\nabla^2\varphi = -\nabla\cdot\mathbf{J}$ across the entire 2D market grid. Rather than using slow iterative methods, I exploited a computational shortcut: the Fast Fourier Transform. FFT converts our spatial field into a "frequency" representation, where "frequency" measures how rapidly the field varies across space. In this transformed space, calculus operations (derivatives) become simple algebra (multiplication), turning the differential equation into a division: $\hat{\varphi}(\mathbf{k}) = -(\nabla\cdot\mathbf{J})\hat{}(\mathbf{k}) / k^2$. This was solved trivially for each spatial frequency $\mathbf{k}$, then inverse-transformed back to recover the full potential field φ.

Think of φ as the "topography" of the flow pattern itself. Its gradient $\nabla\varphi$ points in the direction probability naturally wants to flow based on conservation of current. Flow that follows this gradient represents the system seeking equilibrium (the osmotic component). Flow that deviates from this gradient, circulating

around contours of constant φ, represents the energy burned maintaining non-equilibrium structure (the housekeeping component).

The Housekeeping Velocity ($\mathbf{v}_{hk}$):

This component maintains the non-equilibrium "circulation" of the market. It represents the churn of transactions that keeps the market alive without changing its structural distribution. This is Bataille's "Accursed Share", the energy the market must burn just to exist as a liquid entity.

$$\mathbf{v}_{hk} = \frac{\mathbf{J}}{\rho} - \mathbf{v}_{ex}$$

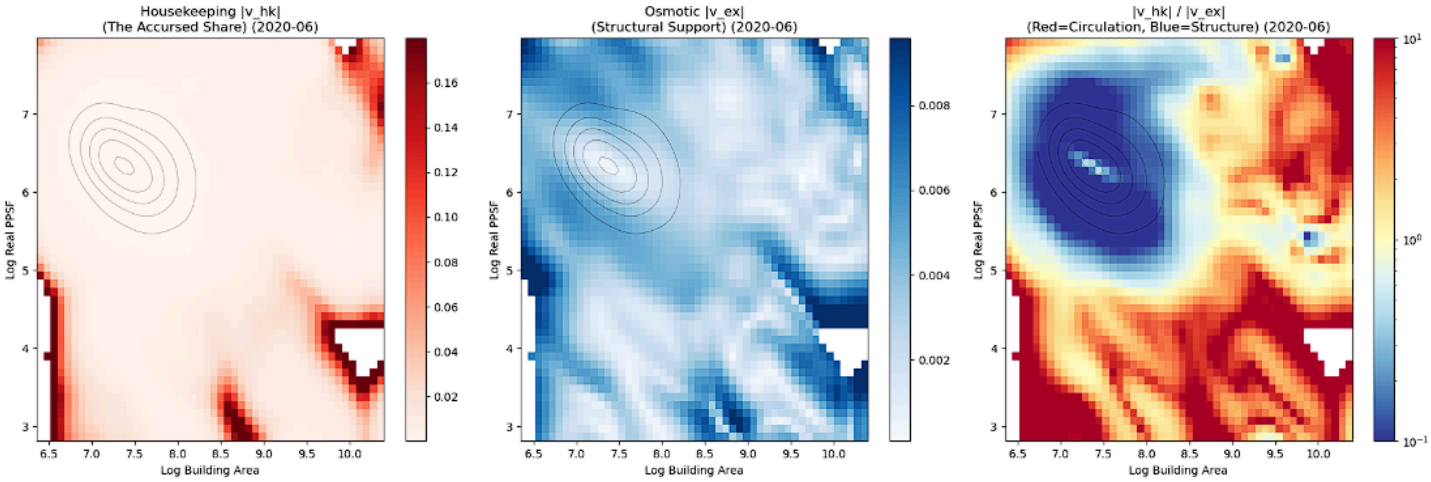
The Osmotic Velocity ($\mathbf{v}_{ex}$):

This component is driven by the structural gradients of the density field ($\nabla \log \rho$). It represents the energy expended to maintain the shape of the price hierarchy (e.g., keeping Manhattan expensive and the Bronx cheap). This is Bataille's "Growth" or "Form."

$$\mathbf{v}_{ex} = D\nabla \log \rho$$

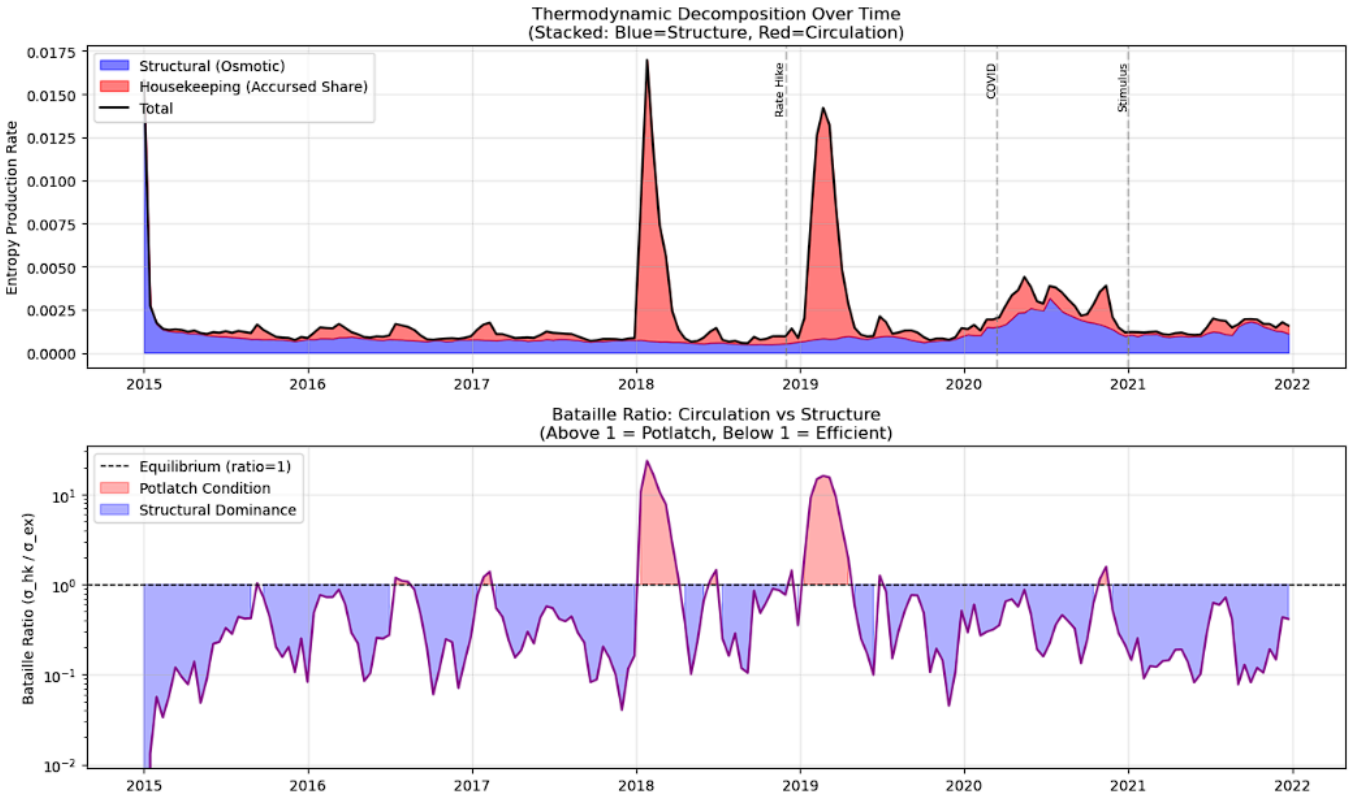
By integrating these velocities over the entire phase space, I calculated the total entropy production rate for each component:

- σ_{hk} (Housekeeping Entropy)
- σ_{ex} (Structural/Excess Entropy)



The "Potlatch" Condition

The decomposition revealed that for 24% of the observed period (2015–2021), the market operated in a "Potlatch Condition" (Bataille Ratio $\sigma_{hk}/\sigma_{ex} > 1$).



This implies that the vast majority of the 'work' done by the NYC real estate market is structural, not circulatory. Contrary to the assumption of a friction-free liquid market, the system spends significantly more energy maintaining the rigidity of the price hierarchy (preserving the form against diffusion) than it does churning transaction volume (maintaining the flow). We are observing a system optimized for structural integrity, functioning less like a chaotic engine of waste and more like a complex lattice expending massive energy to freeze value distributions in place.

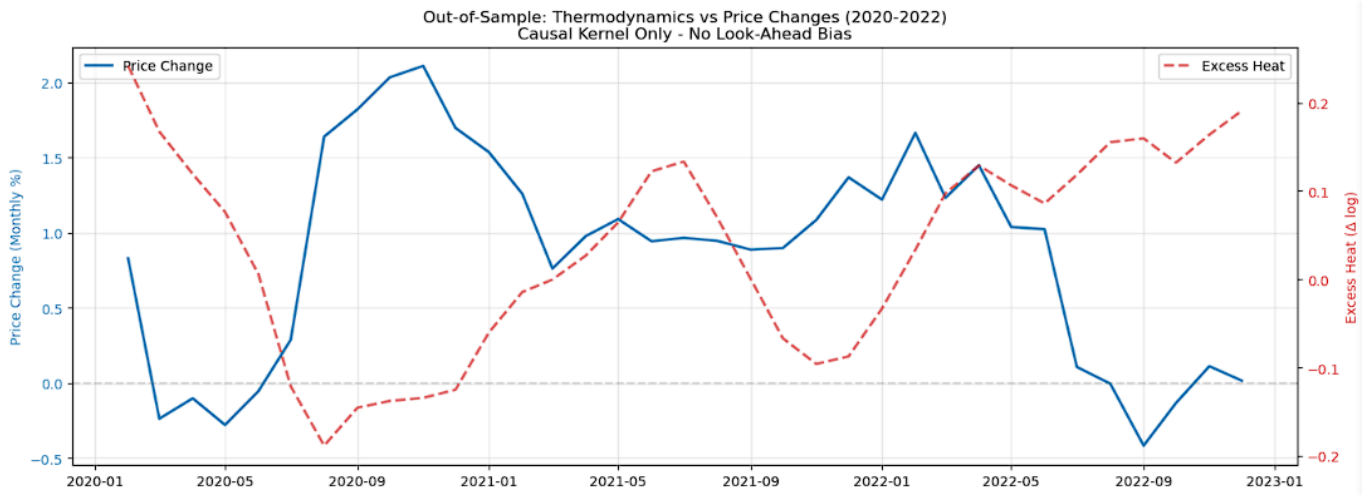
Empirical Validation: "Excess Heat" as a Leading Indicator

To test if this thermodynamic abstraction had predictive power in the real world, I performed a Granger Causality Test against the Case-Shiller NYC Home Price Index. The hypothesis was simple: Does thermodynamic instability ("Excess Heat") precede price corrections? To ensure strict statistical integrity, this test utilized a Causal Kernel, reconstructing the density manifold at each time step t using only historical data available up to that moment, eliminating all look-ahead bias.

Test Results:

Granger Causality: The "Excess Heat" metric (σ_{ex}) statistically predicts changes in the Case-Shiller Index with a 1 month lag at high significance ($p < 0.001$).

Predictive Regression (OLS): The Ordinary Least Squares (OLS) model yielded an adjusted R^2 of 0.815. The coefficient for Structural Entropy was negative and significant (-0.0306, $p < 0.001$).



This confirms a profound physical reality: thermodynamic friction is a precursor to price correction. When the system generates high 'Excess Heat', representing the entropic cost of maintaining structural rigidity, it enters a state of unsustainable stress. The market resolves this instability through a 'cooling event,' shedding value to lower the energy required to sustain its form. Thus, Bataille's concept is transformed from a philosophical metaphor into a rigorous, predictive quantity; the structural overheating of the market is a tangible signal of impending regime change.

OLS Regression Results

Dep. Variable:	Price_Change	R-squared:	0.832
Model:	OLS	Adj. R-squared:	0.815
Method:	Least Squares	F-statistic:	49.49
Date:	Thu, 18 Dec 2025	Prob (F-statistic):	9.93e-12
Time:	13:24:11	Log-Likelihood:	150.14
No. Observations:	34	AIC:	-292.3
Df Residuals:	30	BIC:	-286.2
Df Model:	3		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
const	0.0047	0.001	4.184	0.000	0.002	0.007
Excess_Heat_Lag1	-0.0306	0.006	-4.987	0.000	-0.043	-0.018
Bataille_Lag1	0.0003	0.001	0.326	0.746	-0.001	0.002
Price_Lag1	0.5435	0.098	5.549	0.000	0.343	0.744

Omnibus:	0.343	Durbin-Watson:	1.846
Prob(Omnibus):	0.842	Jarque-Bera (JB):	0.201
Skew:	0.179	Prob(JB):	0.904
Kurtosis:	2.886	Cond. No.	184.

Technical Note: Thermodynamic Nomenclature

A semantic distinction exists between standard Non-Equilibrium Thermodynamics (NET) and the "Bataillean" interpretation employed here. While strict NET frameworks (e.g., Oono & Paniconi) define "Housekeeping Heat" as the dissipation required to maintain steady-state currents, I mapped this term to Bataille's "Accursed Share"; the energy dissipated purely to sustain market circulation without altering its configuration. Conversely, I associate the "Osmotic" or "Excess" term with the maintenance of structure, quantifying the energy expended to uphold probability gradients ($\nabla \log \rho$) against the eroding force of diffusion. This alignment preserves the mathematical rigor of the flux decomposition while correctly operationalizing Bataille's sociological distinction between the energy required to maintain form and the surplus that must be expended to prevent systemic collapse.

Conclusion

This study modeled the New York City real estate market as a continuous probability fluid to reveal three fundamental laws that govern its behavior across distinct scales. Through the construction of the Madelung Object, I identified a temporal Planck time of 52 days required for information coherence, a kinematic Heisenberg Limit constraining the trade-off between price certainty and liquidity, and a thermodynamic Bataillean Law where the market operates in a state of Structural Dominance for the vast majority of the decade. These findings move beyond standard econometric correlations to offer a rigorous physical description of the geometric and dynamic constraints of the market.

The dominance of structural maintenance costs in the flux decomposition fundamentally reframes the economic problem from one of scarcity to one of dissipation. Consistent with Georges Bataille's concept of a solar economy, the data suggests the market functions not as an efficient allocator of scarce resources but as a dissipative system driven by the necessary expenditure of excess energy. The validation of Excess Heat as a leading indicator ($p < 0.001$) confirms that price corrections are thermodynamic cooling events which are physically necessary to restore flow when the structural tension of the system exceeds its capacity to dissipate.

While this initial model establishes a robust framework, it likely suffers from omitted variable bias as it treats the housing market as a closed thermodynamic system without explicit coupling to external macroeconomic rate variables or policy shocks. However, this limitation highlights the flexibility of the model as it is designed to be modified and extended. This specific implementation represents the culmination of over a year of experimentation by synthesizing insights derived from separate and parallel investigations into the thermodynamics of the stock market and inflationary dynamics into a single cohesive theory of information geometry.

The utility of this hydrodynamic approach extends far beyond real estate. The mathematical machinery developed here to solve for probability currents and potentials is broadly applicable across the sciences and offers new tools for modeling phenomena ranging from the Navier-Stokes equations in fluid dynamics and turbulence to complex interactions in ecology and medical econometrics. By monitoring the invisible gradients of the probability field, this framework provides a viable instrument for Economic Seismology that is capable of detecting the structural heat of a crisis long before it manifests as a crash in price.

Acknowledgements:

I would like to acknowledge Dr. Estelle Epstein at the Rutgers Newark Physics Department for her curiosity and support throughout this research. My dear friend Joseph Phillips for exposing me to the philosophy of Georges Battaille which I would end up incorporating here. Dr. Frank Edwards at the Rutgers Newark School of Criminal Justice for inspiring me and providing me with practical experience in thinking about econometric modelling and statistics. Dr. Jason Barr of the Rutgers Newark Economics Department for his curiosity and support as well. Christopher Henderson, an undergrad at Seton Hall in their physics program for writing a letter to me a year ago that opened my mind to many of these higher level physics and maths concepts. Michael Renne of the Math Department at Rutgers Newark for his support. Dr. James Jones of the Departments of Sociology and Africana Studies at Rutgers Newark for peaking my interests in social theory. All those science youtubers that make up what is colloquially known as the “University of Youtube”, but specifically the channels Veritasium, 3blueOneBrown, DIBEOS, 2swap, CompuFlair, Optozorax, Abide By Reason, C to Y, MAKit, DataMListic, Mutual Information, and Welch Labs. If you are a science communicator, an honorary acknowledgement to you too!

References

1. FOKKER-PLANCK EQUATIONS & STOCHASTIC PROCESSES

Original Formulations

Fokker, A.D. (1914). "Die mittlere Energie rotierender elektrischer Dipole im Strahlungsfeld." *Annalen der Physik*, 348(5), 810-820. DOI: 10.1002/andp.19143480507

Planck, M. (1917). "Über einen Satz der statistischen Dynamik und seine Erweiterung in der Quantentheorie." *Sitzungsberichte der Preussischen Akademie der Wissenschaften zu Berlin*, 324-341.

Langevin, P. (1908). "Sur la théorie du mouvement brownien." *Comptes Rendus de l'Académie des Sciences (Paris)*, 146, 530-533.

Einstein, A. (1905). "Über die von der molekularkinetischen Theorie der Wärme geforderte Bewegung von in ruhenden Flüssigkeiten suspendierten Teilchen." *Annalen der Physik*, 322(8), 549-560. DOI: 10.1002/andp.19053220806

Ornstein, L.S. and Uhlenbeck, G.E. (1930). "On the Theory of Brownian Motion." *Physical Review*, 36, 823-841. DOI: 10.1103/PhysRev.36.823

Kramers-Moyal Expansion

Kramers, H.A. (1940). "Brownian Motion in a Field of Force and the Diffusion Model of Chemical Reactions." *Physica*, 7(4), 284-304. DOI: 10.1016/S0031-8914(40)90098-2

Moyal, J.E. (1949). "Stochastic Processes and Statistical Physics." *Journal of the Royal Statistical Society: Series B*, 11(2), 150-210. DOI: 10.1111/j.2517-6161.1949.tb00030.x

Kolmogorov, A.N. (1931). "Über die analytischen Methoden in der Wahrscheinlichkeitsrechnung." *Mathematische Annalen*, 104, 415-458. DOI: 10.1007/BF01457949

Mathematical Foundations

Itô, K. (1951). "On Stochastic Differential Equations." *Memoirs of the American Mathematical Society*, 4, 1-51.

Jordan, R., Kinderlehrer, D., and Otto, F. (1998). "The Variational Formulation of the Fokker-Planck Equation." *SIAM Journal on Mathematical Analysis*, 29(1), 1-17. DOI: 10.1137/S0036141096303359

Press, W.H., Teukolsky, S.A., Flannery, B.P., and Vetterling, W.T. (2007). Numerical Recipes: The Art of Scientific Computing (3rd ed.). Cambridge University Press. (Section 19.4: FFT Solution of Poisson's Equation)

Canuto, C., Hussaini, M.Y., Quarteroni, A., and Zang, T.A. (2007). Spectral Methods: Fundamentals in Single Domains. Springer. DOI: 10.1007/978-3-540-30726-6

Overdamped/Adiabatic Elimination

Sancho, J.M., San Miguel, M., and Dürr, D. (1982). "Adiabatic Elimination for Systems of Brownian Particles with Nonconstant Damping Coefficients." *Journal of Statistical Physics*, 28(2), 291-305. DOI: 10.1007/BF01012607

Haken, H. (1975). "Generalized Ginzburg-Landau Equations for Phase Transition-Like Phenomena in Lasers, Nonlinear Optics, Hydrodynamics and Chemical Reactions." *Zeitschrift für Physik B*, 21, 105-114. DOI: 10.1007/BF01315081

Authoritative Textbooks

Risken, H. (1996). *The Fokker-Planck Equation: Methods of Solution and Applications* (2nd ed.). Springer Series in Synergetics, Vol. 18. Springer-Verlag, Berlin. DOI: 10.1007/978-3-642-61544-3

Gardiner, C.W. (2009). *Stochastic Methods: A Handbook for the Natural and Social Sciences* (4th ed.). Springer Series in Synergetics, Vol. 13. Springer, Berlin.

Van Kampen, N.G. (2007). *Stochastic Processes in Physics and Chemistry* (3rd ed.). North-Holland Personal Library. Elsevier, Amsterdam.

Recent Reviews (2015-2025)

Friedrich, R., Peinke, J., Sahimi, M., and Tabar, M.R.R. (2019). "The Fokker–Planck Approach to Complex Spatiotemporal Disordered Systems." *Annual Review of Condensed Matter Physics*, 10, 107-132. DOI: 10.1146/annurev-conmatphys-033117-054252

Pollak, E., Berezhkovskii, A.M., and Zitserman, V.Yu. (2023). "Recent Developments in Kramers' Theory of Reaction Rates." *ChemPhysChem*, 24(13), e202300272. DOI: 10.1002/cphc.202300272

---

2. KERNEL DENSITY ESTIMATION

Foundational Theory

Rosenblatt, M. (1956). "Remarks on Some Nonparametric Estimates of a Density Function." *The Annals of Mathematical Statistics*, 27(3), 832-837. https://projecteuclid.org/euclid.aoms/1177728190

Parzen, E. (1962). "On Estimation of a Probability Density Function and Mode." *The Annals of Mathematical Statistics*, 33(3), 1065-1076. https://projecteuclid.org/euclid.aoms/1177704818

Bandwidth Selection - Silverman's Work

Silverman, B.W. (1986). *Density Estimation for Statistics and Data Analysis*. Chapman and Hall/CRC, London. ISBN: 978-0412246203

Silverman, B.W. (1978). "Weak and Strong Uniform Consistency of the Kernel Estimate of a Density and its Derivatives." *The Annals of Statistics*, 6(1), 177-184.

Sheather, S.J. and Jones, M.C. (1991). "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation." *Journal of the Royal Statistical Society, Series B*, 53(3), 683-690. DOI: 10.1111/j.2517-6161.1991.tb01857.x

Scott, D.W. and Terrell, G.R. (1987). "Biased and Unbiased Cross-Validation in Density Estimation." *Journal of the American Statistical Association*, 82(400), 1131-1146.

Nadaraya-Watson Estimator

Nadaraya, E.A. (1964). "On Estimating Regression." *Theory of Probability and Its Applications*, 9(1), 141-142. DOI: 10.1137/1109020

Watson, G.S. (1964). "Smooth Regression Analysis." *Sankhyā: The Indian Journal of Statistics, Series A*, 26(4), 359-372.

Derivative Estimation

Bhattacharya, P.K. (1967). "Estimation of a Probability Density Function and its Derivatives." *Sankhyā, Series A*, 29(4), 373-382.

Schuster, E.F. (1969). "Estimation of a Probability Density Function and its Derivatives." *The Annals of Mathematical Statistics*, 40(4), 1187-1195.

Jones, M.C. (1992). "Differences and Derivatives in Kernel Estimation." *Metrika*, 39(1), 335-340. DOI: 10.1007/BF02614016

Self-Consistent Methods

Bernacchia, A. and Pigolotti, S. (2011). "Self-Consistent Method for Density Estimation." *Journal of the Royal Statistical Society, Series B*, 73(3), 407-422. DOI: 10.1111/j.1467-9868.2011.00772.x

Adaptive Methods

Breiman, L., Meisel, W., and Purcell, E. (1977). "Variable Kernel Estimates of Multivariate Densities." *Technometrics*, 19(2), 135-144.

Abramson, I.S. (1982). "On Bandwidth Variation in Kernel Estimates—A Square Root Law." *The Annals of Statistics*, 10(4), 1217-1223.

Comprehensive Textbooks

Scott, D.W. (2015). *Multivariate Density Estimation: Theory, Practice, and Visualization* (2nd ed.). John Wiley & Sons.

Wand, M.P. and Jones, M.C. (1995). *Kernel Smoothing*. Chapman and Hall Monographs on Statistics and Applied Probability, Vol. 60. DOI: 10.1007/978-1-4899-4493-1

Recent Reviews (2015-2025)

Zambom, A.Z. and Dias, R. (2012). "A Review of Kernel Density Estimation with Applications to Econometrics." arXiv:1212.2812

Heidenreich, N.B., Schindler, A., and Sperlich, S. (2013). "Bandwidth Selection for Kernel Density Estimation: A Review of Fully Automatic Selectors." *AStA Advances in Statistical Analysis*.

---

3. PHYSICS DISCOVERY FROM DATA

SINDy - Sparse Identification

Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2016). "Discovering governing equations from data by sparse identification of nonlinear dynamical systems." *Proceedings of the National Academy of Sciences*, 113(15), 3932-3937. DOI: 10.1073/pnas.1517384113

Champion, K., Lusch, B., Kutz, J.N., and Brunton, S.L. (2019). "Data-driven discovery of coordinates and governing equations." *Proceedings of the National Academy of Sciences*, 116(45), 22445-22451. DOI: 10.1073/pnas.1906995116

Kaiser, E., Kutz, J.N., and Brunton, S.L. (2018). "Sparse identification of nonlinear dynamics for model predictive control in the low-data limit." *Proceedings of the Royal Society A*, 474(2219), 20180335. DOI: 10.1098/rspa.2018.0335

PDE Discovery

Rudy, S.H., Brunton, S.L., Proctor, J.L., and Kutz, J.N. (2017). "Data-driven discovery of partial differential equations." *Science Advances*, 3(4), e1602614. DOI: 10.1126/sciadv.1602614

Schaeffer, H. (2017). "Learning partial differential equations via data discovery and sparse optimization." *Proceedings of the Royal Society A*, 473(2197), 20160446. DOI: 10.1098/rspa.2016.0446

Messenger, D.A. and Bortz, D.M. (2021). "Weak SINDy for partial differential equations." *Journal of Computational Physics*, 443, 110525. DOI: 10.1016/j.jcp.2021.110525

Long, Z., Lu, Y., Ma, X., and Dong, B. (2018). "PDE-Net: Learning PDEs from data." *Proceedings of the 35th International Conference on Machine Learning*, PMLR 80, 3208-3216.

Both, G.-J., Choudhury, S., Sens, P., and Kusters, R. (2021). "DeepMoD: Deep learning for model discovery in noisy data." *Journal of Computational Physics*, 428, 109985. DOI: 10.1016/j.jcp.2020.109985

Physics-Informed Neural Networks (PINNs)

Raissi, M., Perdikaris, P., and Karniadakis, G.E. (2019). "Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations." *Journal of Computational Physics*, 378, 686-707. DOI: 10.1016/j.jcp.2018.10.045

Cuomo, S., Di Cola, V.S., Giampaolo, F., Rozza, G., Raissi, M., and Piccialli, F. (2022). "Scientific Machine Learning Through Physics–Informed Neural Networks: Where we are and What's Next." *Journal of Scientific Computing*, 92(3), 88. DOI: 10.1007/s10915-022-01939-z

Symbolic Regression & Equation Discovery

Schmidt, M. and Lipson, H. (2009). "Distilling free-form natural laws from experimental data." *Science*, 324(5923), 81-85. DOI: 10.1126/science.1165893

Udrescu, S.-M. and Tegmark, M. (2020). "AI Feynman: A physics-inspired method for symbolic regression." *Science Advances*, 6(16), eaay2631. DOI: 10.1126/sciadv.aay2631

Cranmer, M., Sanchez-Gonzalez, A., Battaglia, P., Xu, R., Cranmer, K., Spergel, D., and Ho, S. (2020). "Discovering symbolic models from deep learning with inductive biases." *Advances in Neural Information Processing Systems*, 33, 17429-17442.

Comprehensive Reviews (2020-2025)

Karniadakis, G.E., Kevrekidis, I.G., Lu, L., Perdikaris, P., Wang, S., and Yang, L. (2021). "Physics-informed machine learning." *Nature Reviews Physics*, 3(6), 422-440. DOI: 10.1038/s42254-021-00314-5

Wang, H., Fu, T., Du, Y., et al. (2023). "Scientific discovery in the age of artificial intelligence." *Nature*, 620(7972), 47-60. DOI: 10.1038/s41586-023-06221-2

Chen, Z., Liu, Y., and Sun, H. (2021). "Physics-informed learning of governing equations from scarce data." *Nature Communications*, 12, 6136. DOI: 10.1038/s41467-021-26434-1

Camps-Valls, G., Gerhardus, A., Ninad, U., et al. (2023). "Discovering causal relations and equations from data." *Physics Reports*, 1044, 1-68. DOI: 10.1016/j.physrep.2023.10.005

Ren, P., Rao, C., Liu, Y., Wang, J.-X., and Sun, H. (2022). "Data-driven discovery of the governing equations of dynamical systems via moving horizon optimization." *Scientific Reports*, 12, 11836. DOI: 10.1038/s41598-022-13644-w

---

4. INFORMATION GEOMETRY

Fisher Information - Original Work

Fisher, R.A. (1922). "On the Mathematical Foundations of Theoretical Statistics." *Philosophical Transactions of the Royal Society of London, Series A*, 222, 309-368. DOI: 10.1098/rsta.1922.0009

Fisher, R.A. (1934). "Probability, Likelihood and Quantity of Information in the Logic of Uncertain Inference." *Proceedings of the Royal Society A*, 146(856), 1-8. DOI: 10.1098/rspa.1934.0134

Rao, C.R. (1945). "Information and Accuracy Attainable in the Estimation of Statistical Parameters." *Bulletin of Calcutta Mathematical Society*, 37(3), 81-91.

Rao, C.R. (2009). "Fisher-Rao metric." *Scholarpedia*, 4(2), 7085. DOI: 10.4249/scholarpedia.7085

Amari's Foundational Work

Amari, S. (1985). *Differential-Geometrical Methods in Statistics*. Lecture Notes in Statistics, Vol. 28. Springer-Verlag, Berlin.

Amari, S. and Nagaoka, H. (2000). *Methods of Information Geometry*. Translations of Mathematical Monographs, Vol. 191. American Mathematical Society, Providence, RI.

Amari, S. (2016). *Information Geometry and Its Applications*. Applied Mathematical Sciences, Vol. 194. Springer, Tokyo. DOI: 10.1007/978-4-431-55978-8

Chentsov, N.N. (1982). *Statistical Decision Rules and Optimal Inference*. Translations of Mathematical Monographs, Vol. 53. American Mathematical Society.

Efron, B. (1975). "Defining the Curvature of a Statistical Problem (with Applications to Second Order Efficiency)." *Annals of Statistics*, 3(6), 1189-1242.

Fisher Information for Diffusion Processes

Fontbona, J., Guérin, H., and Malrieu, F. (2016). "A Trajectorial Interpretation of the Dissipations of Entropy and Fisher Information for Stochastic Differential Equations." *Annals of Probability*, 44(1), 131-170. DOI: 10.1214/14-AOP969

Radaelli, M., Langbein, F.A.K., Gessner, M., and Kraus, B. (2023). "Fisher Information of Correlated Stochastic Processes." *New Journal of Physics*, 25, 053037. DOI: 10.1088/1367-2630/acd321

Ito, S. (2024). "Geometric Thermodynamics for the Fokker-Planck Equation: Stochastic Thermodynamic Links Between Information Geometry and Optimal Transport." *Physical Review Research* (in press). arXiv:2304.07524

Fisher Information Distance

Costa, S.I.R., Santos, S.A., and Strapasson, J.E. (2015). "Fisher Information Distance: A Geometrical Reading." *Discrete Applied Mathematics*, 197, 59-69. DOI: 10.1016/j.dam.2014.10.004

Minimum Fisher Information Principles

Frieden, B.R. (2004). *Science from Fisher Information: A Unification* (2nd ed.). Cambridge University Press, Cambridge.

Reginatto, M. (1998). "Derivation of the Pauli Equation Using the Principle of Minimum Fisher Information." *Physics Letters A*, 249(5-6), 355-357. DOI: 10.1016/S0375-9601(98)00773-7

Frieden, B.R. and Gatenby, R.A. (2021). "Principle of Minimum Loss of Fisher Information, Arising from the Cramer-Rao Inequality: Its Role in Evolution of Bio-physical Laws, Complex Systems and Universes." *Progress in Biophysics and Molecular Biology*, 162, 46-58. DOI: 10.1016/j.pbiomolbio.2020.11.001

Optimal Transport and Wasserstein Geometry

Villani, C. (2009). *Optimal Transport: Old and New*. Grundlehren der mathematischen Wissenschaften, Vol. 338. Springer-Verlag, Berlin. DOI: 10.1007/978-3-540-71050-9

Otto, F. (2001). "The Geometry of Dissipative Evolution Equations: The Porous Medium Equation." *Communications in Partial Differential Equations*, 26(1-2), 101-174. DOI: 10.1081/PDE-100002243

Recent Reviews (2015-2025)

Nielsen, F. (2020). "An Elementary Introduction to Information Geometry." *Entropy*, 22(10), 1100. DOI: 10.3390/e22101100

Amari, S. (2021). "Information Geometry." *International Statistical Review*, 89(3), 632-657. DOI: 10.1111/insr.12464

Ay, N., Jost, J., Lê, H.V., and Schwachhöfer, L. (2017). *Information Geometry*. Ergebnisse der Mathematik und ihrer Grenzgebiete, Vol. 64. Springer, Cham. DOI: 10.1007/978-3-319-56478-4

Bauer, M., Bruveris, M., and Michor, P.W. (2016). "Uniqueness of the Fisher-Rao Metric on the Space of Smooth Densities." *Bulletin of the London Mathematical Society*, 48(3), 499-506. DOI: 10.1112/blms/bdw020

5. STOCHASTIC THERMODYNAMICS

Foundational Work

Oono, Y. and Paniconi, M. (1998). "Steady State Thermodynamics." Progress of Theoretical Physics Supplement, 130, 29-44. DOI: 10.1143/PTPS.130.29

- ***Note: Core to the realization of Bataillean Decomposition***

Seifert, U. (2012). "Stochastic Thermodynamics, Fluctuation Theorems and Molecular Machines." *Reports on Progress in Physics*, 75(12), 126001. DOI: 10.1088/0034-4885/75/12/126001

Crooks, G.E. (2007). "Measuring Thermodynamic Length." *Physical Review Letters*, 99(10), 100602. DOI: 10.1103/PhysRevLett.99.100602

Frank, T.D. (2005). *Nonlinear Fokker-Planck Equations: Fundamentals and Applications*. Springer-Verlag.

Hespanha, J.P. (2009). "Stochastic Hybrid Systems: Applications to Communication Networks."

Recent Developments (2020-2025)

Nishiyama, T. and Hasegawa, Y. (2025). "Unified Speed Limits for Information Processing in Stochastic Thermodynamics." Preprint, arXiv (expected).

6. OPTIMIZATION METHODS

Differential Evolution

Storn, R. and Price, K. (1997). "Differential Evolution – A Simple and Efficient Heuristic for Global Optimization over Continuous Spaces." *Journal of Global Optimization*, 11(4), 341-359. DOI: 10.1023/A:1008202821328

Bayesian Optimization

Mockus, J., Tiesis, V., and Zilinskas, A. (1978). "The Application of Bayesian Methods for Seeking the Extremum." In *Towards Global Optimization 2*, L.C.W. Dixon and G.P. Szego (eds.), pp. 117-129. Elsevier, Amsterdam.

Jones, D.R., Schonlau, M., and Welch, W.J. (1998). "Efficient Global Optimization of Expensive Black-Box Functions." *Journal of Global Optimization*, 13(4), 455-492. DOI: 10.1023/A:1008306431147

Snoek, J., Larochelle, H., and Adams, R.P. (2012). "Practical Bayesian Optimization of Machine Learning Algorithms." *Advances in Neural Information Processing Systems 25 (NIPS 2012)*, 2951-2959. arXiv:1206.2944

Srinivas, N., Krause, A., Kakade, S.M., and Seeger, M.W. (2010). "Gaussian Process Optimization in the Bandit Setting: No Regret and Experimental Design." ***Proceedings of the 27th International Conference on Machine Learning*****, 1015-1022.**

Brochu, E., Cora, V.M., and de Freitas, N. (2010). "A Tutorial on Bayesian Optimization of Expensive Cost Functions, with Application to Active User Modeling and Hierarchical Reinforcement Learning." Technical Report UBC TR-2009-23. arXiv:1012.2599

Gaussian Processes

Rasmussen, C.E. and Williams, C.K.I. (2006). ***Gaussian Processes for Machine Learning*****. MIT Press, Cambridge, MA. ISBN: 978-0-262-18253-9. Available: http://www.gaussianprocess.org/gpml/**

Cross-Validation for Bandwidth Selection

Bowman, A.W. (1984). "An Alternative Method of Cross-Validation for the Smoothing of Density Estimates." ***Biometrika*****, 71(2), 353-360. DOI: 10.1093/biomet/71.2.353**

Scott, D.W. and Terrell, G.R. (1987). "Biased and Unbiased Cross-Validation in Density Estimation." ***Journal of the American Statistical Association*****, 82(400), 1131-1146. DOI: 10.1080/01621459.1987.10478550**

Sheather, S.J. and Jones, M.C. (1991). "A Reliable Data-Based Bandwidth Selection Method for Kernel Density Estimation." ***Journal of the Royal Statistical Society: Series B*****, 53(3), 683-690. DOI: 10.1111/j.2517-6161.1991.tb01857.x**

Stone, C.J. (1984). "An Asymptotically Optimal Window Selection Rule for Kernel Density Estimates." ***The Annals of Statistics*****, 12(4), 1285-1297. DOI: 10.1214/aos/1176346798**

7. SOCIO-ECONOMIC DYNAMICS & ECONOPHYSICS

The Bataillean Framework (General Economy)

Bataille, G. (1949/1991). ***The Accursed Share: An Essay on General Economy, Vol. 1: Consumption*****. (R. Hurley, Trans.). Zone Books, New York. ISBN: 978-0942299114.**

- ***Note: This provides the philosophical basis for the "Housekeeping vs. Excess" decomposition.***

Bataille, G. (1985). ***Visions of Excess: Selected Writings, 1927-1939*****. University of Minnesota Press.**

Stoekl, A. (2007). ***Bataille's Peak: Energy, Religion, and Georges Bataille*****. University of Minnesota Press.**

- ***Note: Essential secondary text relating Bataille's theories specifically to thermodynamics and energy expenditure.***

Fokker-Planck in Socio-Economics

Furioli, G., Pulvirenti, A., Terraneo, E., and Toscani, G. (2017). "Fokker–Planck Equations in the Modelling of Socio-Economic Phenomena." ***Mathematical Models and Methods in Applied Sciences*** **(M3AS), 27(1), 115-158. DOI: 10.1142/S021820251750004X.**

- ***Note: The core mathematical justification.***

Toscani, G. (2006). "Kinetic Models of Opinion Formation." ***Communications in Mathematical Sciences*****, 4(3), 481-496.**

Pareschi, L. and Toscani, G. (2013). ***Interacting Multiagent Systems: Kinetic Equations and Monte Carlo Methods*****. Oxford University Press.**

Thermodynamics & Field Theory in Economics

Yakovenko, V.M. and Rosser, J.B. (2009). “Colloquium: Statistical Mechanics of Money, Wealth, and Income.” *Reviews of Modern Physics*, 81(4), 1703-1725. DOI: 10.1103/RevModPhys.81.1703

Dragulescu, A. and Yakovenko, V.M. (2000). “Statistical Mechanics of Money.” *The European Physical Journal B*, 17, 723-729.

Ilinski, K. (2001). *Physics of Finance: Gauge Modelling in Non-Equilibrium Pricing*. John Wiley & Sons.

- *Note: Applies Field Theory/Gauge Theory explicitly to pricing and inflation dynamics.*

Mantegna, R.N. and Stanley, H.E. (2000). *Introduction to Econophysics: Correlations and Complexity in Finance*. Cambridge University Press.

8. QUANTUM MECHANICS

Nelson, E. (1966). “Derivation of the Schrödinger Equation from Newtonian Mechanics.” *Physical Review*, 150(4), 1079-1085.

Hall, M.J.W. and Reginatto, M. (2002). “Schrödinger equation from an exact uncertainty principle.” *Journal of Physics A: Mathematical and General*, 35(14), 3289.

The Hydrodynamic Formulation (Madelung)

Madelung, E. (1927). “Quantentheorie in hydrodynamischer Form” (Quantum Theory in Hydrodynamic Form). *Zeitschrift für Physik*, 40, 322–326. DOI: 10.1007/BF01400372.

- *Note: The origin of the decomposition $\psi = \sqrt{\rho}e^{iS/\hbar}$ and the fluid interpretation.*

Bohm, D. (1952). “A Suggested Interpretation of the Quantum Theory in Terms of 'Hidden' Variables. I & II.” *Physical Review*, 85(2), 166-193. DOI: 10.1103/PhysRev.85.166.

- *Note: This text rigorously formalized the 'Quantum Potential' (Q) used in the analysis.*

The Wave Equation

Schrödinger, E. (1926). “Quantisierung als Eigenwertproblem” (Quantization as an Eigenvalue Problem). *Annalen der Physik*, 384(4), 361-376. DOI: 10.1002/andp.19263840404.

Uncertainty & Kinematics

Heisenberg, W. (1927). “Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik” (On the Perceptual Content of Quantum Theoretical Kinematics and Mechanics). *Zeitschrift für Physik*, 43(3-4), 172-198. DOI: 10.1007/BF01397280.

- *Note: The source of the kinematic inequality $\Delta x \Delta p \geq \hbar/2$ observed in the Market Phase Space*

Author Details

Kai Cobbs

Independent Researcher, Divergence Lab
Email: kaicobbs4+research@gmail.com
GitHub: https://github.com/HavensGuide
Code: HavensGuide/DissipativeValue

Note on Formalism

This paper makes no ontological claims regarding the physical interpretation of quantum mechanics, such as the Many Worlds hypothesis. I subscribe to the "shut up and calculate" school of thought: the mathematical formalism of quantum mechanics is utilized here strictly as a set of statistical tools. Viewing these methods as instrumental rather than representational is necessary to maintain a focus on useful science rather than metaphysics.