Robust Pricing using Martingale Optimal Transport

SPDE-2024, IIT-Madras

Purba Banerjee

Department of Mathematics
Indian Institute of Science, Bangalore

June 3, 2024

Introduction

- Derivative pricing in finance can be divided into two major categories:
  1. Model-based pricing.
  2. Robust/model-independent pricing.
- Model based pricing, as the name suggests, involves pricing of a derivative, given certain assumptions on the underlying asset following some model. Examples- *Black-Scholes* model, *Heston* model, *Merton Jump-Diffusion* model.
- The quantity of interest in any form of pricing is the valuation of the underlying risk-neutral densities, under no-arbitrage conditions.
- One of the most pioneering works in this regard was by Breeden and Litzenberger (1978).

Breeden and Litzenberger's result

- Let $C(S, t, K, T)$ denote the time-t price of a European call with strike K and maturity T.
- The probability density function of the asset price under a risk-neutral measure $\mathbb{Q}$, evaluated at the future price level K and the future time T, conditional on the stock price starting at level S at an earlier time t, is denoted by $q(S, t, K, T)$.
- Breeden and Litzenberger (1978) proved that the risk-neutral density is related to the second strike derivative of the call pricing function as follows,

$$q(S, t, K, T) = e^{r(T-t)} \frac{\partial^2 C}{\partial K^2}(S, t, K, T). \tag{1}$$

Optimal Transport

- The *optimal transport* (OT) problem is concerned with transferring mass from one location to another in such a way as to optimize a given criterion.
- Rephrased mathematically, and for simplicity considering the one-dimensional case, we are given two probability distributions μ and ν on $\mathbb{R}$ and seek to minimize

$$\int_{\mathbb{R}^2} c(x,y)\mathbb{P}(dx,dy), \tag{2}$$

among all probability measures $\mathbb{P}$, also known as *transport plans*, such that

$$\mathbb{P}[E \times \mathbb{R}] = \mu[E] \text{ and } \mathbb{P}[\mathbb{R} \times E] = \nu[E], \text{ for all } E \in \mathcal{B}(\mathbb{R}), \tag{3}$$

where $c : \mathbb{R}^2 \to \mathbb{R}$ is a measurable cost function.

- Example: An *Asian* option with pay-off $c(x,y) = (\frac{1}{2}(x+y) - K)^+$, with K being the strike price at maturity T, and x, y denoting the underlying asset prices at times $0 < T_1 < T_2 = T$, respectively.

Optimal Transport

- In the absolutely continuous case, i.e., $\mu(dx) = \rho(x)dx$ and $\nu(dy) = \sigma(y)dy$, Benamou and Brenier (2000) proposed a numerical scheme for the quadratic distance function $c(x, y) = (x - y)^2$ using an equivalent formulation arising from fluid mechanics.
- In the purely discrete case, i.e. $\mu(dx) = \sum_{i=1}^{m} \alpha_i \delta_{x_i}(dx)$ and $\nu(dy) = \sum_{j=1}^{n} \beta_j \delta_{y_j}(dy)$, the OT problem reduces to a linear programming (LP) problem. It can be computed using the iterative Bregman projection as shown in Benamou et al. (2015).
- In the semi-discrete case, i.e. $\mu(dx) = \rho(x)dx$ and $\nu(dy) = \sum_{j=1}^{n} \beta_j \delta_{y_j}(dy)$, Lévy et al. (2015) adopted a computational geometry approach to the cost $c(x, y) = (x - y)^2$ and solved the OT problem utilizing Laguerre's tessellations.

Martingale Optimal Transport

- Recently, an additional constraint has been taken into account, which leads to the so-called martingale optimal transport (MOT) problem.
- More precisely, the two given measures μ and ν describe the initial and final distributions of stock prices.
- These distributions can be recovered from market prices of traded call/put options.
- Calibrated market models are then identified by martingales with these prescribed marginals, i.e. transport plans $\mathbb{P}$ which further satisfy

$$\mathbb{E}_{\mathbb{P}}[Y|X] = X \tag{4}$$

- The MOT problem aims at maximizing the integral (2) overall $\mathbb{P}$, still named transport plans, satisfying the constraints (3) and (4), and it corresponds to the model-independent price for option c.
- This methodology was pioneered by Beiglböck et al. (2013).

LP formulation for MOT problems

- When the underlying marginal distributions are discrete, i.e. $\mu(dx) = \sum_{i=1}^{m} \alpha_i \delta_{x_i}(dx)$ and $\nu(dy) = \sum_{j=1}^{n} \beta_j \delta_{y_j}(dy)$, the MOT problem is equivalent to the following LP problem:

$$\max_{(p_{i,j})_{1\leq i\leq m, 1\leq j\leq n} \in \mathbb{R}_+^{mn}} \sum_{i=1}^{m} \sum_{j=1}^{n} p_{i,j} c(x_i, y_j) \text{ s.t. } \begin{aligned} &\sum_{j=1}^{n} p_{i,j} = \alpha_i, \quad \text{for } \; i = 1, .., m, \\ &\sum_{i=1}^{m} p_{i,j} = \beta_j, \quad \text{for } \; j = 1, .., n, \\ &\sum_{j=1}^{n} p_{i,j} y_j = \alpha_i x_i, \quad \text{for } \; i = 1, .., m. \end{aligned} \tag{5}$$

- Davis et al. (2014) developed such LP formulation, where instead of the marginal constraint ν, only a finite number of expectation constraints are given.
- For a convex reward function, this leads to optimizers with finite support.
- In order to generalize this approach, a natural direction would be to try approximating the MOT problem for (μ, ν) with the LP problem mentioned above, for finitely supported (μ^n, ν^n) which are 'close' to (μ, ν).
- One would encounter two main obstacles while working in that direction:
    1. General continuity results of the MOT problem are difficult to establish.
    2. Even if (μ, ν) admits a martingale transport plan, in dimensions $d > 1$, the construction of a discrete approximation (μ^n, ν^n) which also satisfies this, may be quite involved.
- Both of these issues were addressed in Guo and Obłój (2019), and forms the basis of this talk.

Outline of their results

- The authors provide an approximation approach for solving systematically $N-$period MOT problems on $\mathbb{R}^d$, with $N \geq 2$ and $d \geq 1$.
- Their approximation of the original problem relies on a discretization of the marginal distributions combined with a suitable relaxation of the martingale constraint leading to a sequence of LP problems.
- A proof of the convergence of this sequence is given.
- Results for the convergence speed are obtained when restricted to $N = 2$ and $d = 1$.

Preliminaries

- For a given set E, we denote by E^k its $k-$ fold product.
- If E is Polish, then $\mathcal{B}(E)$ denotes its Borel $\sigma-$field and $\mathcal{P}(E)$ is the set of probability measures on $(E, \mathcal{B}(E))$ which admit a finite first moment.
- Let $\Omega := \mathbb{R}^d$ with its elements denoted by $\mathbf{x} = (x_1, x_2, ..x_d)$ and $\mathcal{P} := \mathcal{P}(\Omega)$. Throughout, the Euclidean space $\mathbb{R}^d$ is endowed with the l_1 norm $|\cdot|$, i.e. $|\mathbf{x}| := \sum_{i=1}^{d} |x_i|$.
- Define Λ to be the space of Lipschitz functions on $\mathbb{R}^d$ and, given $f \in \Lambda$, denote by $\text{Lip}(f)$ its Lipschitz constant on $\mathbb{R}^d$.
- For each $L > 0$, let $\Lambda_L \subset \Lambda$ be the subspace of functions f with $\text{Lip}(f) \leq L$.

Preliminaries

- We consider the coordinate process $(S_k)_{1\leq k\leq N}$, i.e. $S_k(x_1, x_2, ..., x_N) \in \Omega_N$ and its natural filtration $(\mathcal{F}_k)_{1\leq k\leq N}$, i.e. $\mathcal{F}_k := \sigma(S_1, ...S_k)$.
- From a financial viewpoint, Ω^N models the collection of all possible trajectories for the price evolution of d stocks, where N is the number of trading dates.
- Given a vector of probability measures $\boldsymbol{\mu} = (\mu_k)_{1\leq k\leq N} \in \mathcal{P}^N$, define the set of transport plans with the marginal distributions $\mu_1, .., \mu_N$ by

$$\mathcal{P}(\boldsymbol{\mu}) := \{\mathbb{P} \in \mathcal{P}(\Omega^N) : \mathbb{P} \circ S_k^{-1} = \mu_k, \text{ for } k = 1, .., N\},$$

where $\mathbb{P} \circ S_k^{-1}$ denotes the push forward of $\mathbb{P}$ via the map $S_k : \Omega^N \to \Omega$.

Wasserstein distance

- The Wasserstein distance in terms S_k is given by

$$\mathcal{W}(\mu,\nu) := \inf_{\mathbb{P}\in\mathcal{P}(\mu,\nu)} \mathbb{E}_{\mathbb{P}}[|S_1 - S_2|] = \sup_{f\in\Lambda_1} \left\{ \int_{\mathbb{R}^d} f(x)\mu(x)dx - \int_{\mathbb{R}^d} f(x)\nu(x)dx \right\}, \tag{6}$$

- The probability space $\mathcal{P}$, equipped with the metric $\mathcal{W}$, is a Polish space.
- Further, for any $(\mu^n)_{n\geq 1} \subset \mathcal{P}$ and $\mu \in \mathcal{P}, \mathcal{W}(\mu^n,\mu) \to 0$ holds if and only if

$$\mu^n \xrightarrow{\mathcal{L}} \mu \quad \text{and} \quad \int_{\mathbb{R}^d} |\boldsymbol{x}|\mu^n(dx) \to \int_{\mathbb{R}^d} |\boldsymbol{x}|\mu(dx),$$

  where $\mathcal{L}$ represents the weak convergence of probability measures.
- The space $\mathcal{W}$ is endowed with the product metric $\mathcal{W}^{\oplus}(\boldsymbol{\mu},\boldsymbol{\nu}) := \sum_{k=1}^{N} \mathcal{W}(\mu_k,\nu_k)$, for all $\boldsymbol{\mu},\boldsymbol{\nu} \in \mathcal{P}^N$.
- Then $\mathcal{P}^N$ is Polish with respect to $\mathcal{W}^{\oplus}$.

$\epsilon-$approximating measure

We first define an $\epsilon-$approximating measure to introduce the main results.

Definition

For any $\epsilon \geq 0$, a probability measure $\mathbb{P} \in \mathcal{P}(\Omega^N)$ is said to be an $\epsilon-$approximating martingale measure if for each $k = 1, ..N-1$

$$\mathbb{E}_{\mathbb{P}}\left[\left|\mathbb{E}_{\mathbb{P}}[S_{k+1}|\mathcal{F}_k] - S_k\right|\right] \leq \epsilon, \tag{7}$$

Relaxed MOT

- Given $\epsilon \geq 0$, let $\mathcal{M}_\epsilon(\boldsymbol{\mu}) \subset \mathcal{P}(\boldsymbol{\mu})$ be the subset containing all $\epsilon-$ approximating martingale measures.
- For a measurable function $c : \Omega^N \to \mathbb{R}$, the relaxed MOT problem is defined by

$$P_\epsilon(\boldsymbol{\mu}) := \sup_{\mathbb{P} \in \mathcal{M}_\epsilon(\boldsymbol{\mu})} \mathbb{E}_{\mathbb{P}}[c(S_1, ..., S_N)], \tag{8}$$

  where we set by convention $P_\epsilon(\boldsymbol{\mu}) := -\infty$ whenever $\mathcal{M}_\mu(\boldsymbol{\mu}) = \emptyset$.
- We denote $\mathcal{P}_\epsilon^{\preceq} \subset \mathcal{P}^N$ the collection of measures $\boldsymbol{\mu}$ such that $\mathcal{M}(\boldsymbol{\mu}) \neq \emptyset$. For $\epsilon = 0$, we drop the subscript and denote by $\mathcal{P}^{\preceq} \equiv \mathcal{P}_\epsilon^{\preceq}, \mathcal{M}(\boldsymbol{\mu}) \equiv \mathcal{P}_0(\boldsymbol{\mu})$, etc.

Main Result

Theorem

Fix $\boldsymbol{\mu} \in \mathcal{P}^{\preceq}$. Let $(\boldsymbol{\mu}^n)_{n\geq 1} \subset \mathcal{P}^N$ be a sequence converging to $\boldsymbol{\mu}$ under $\mathcal{W}^{\oplus}$. Then, for all $n \geq 1$, $\boldsymbol{\mu}^n \in \mathcal{P}^{\preceq}_{r_n}$ with $r_n := \mathcal{W}^{\oplus}(\boldsymbol{\mu}^n, \boldsymbol{\mu})$. Assume further c is Lipschitz.

1. For any sequence $(\epsilon_n)_{n\geq 1}$ converging to zero such that $\epsilon_n \geq r_n$ for all $n \geq 1$, one has
$$\lim_{n\to\infty} P_{\epsilon_n}(\boldsymbol{\mu}^n) = P(\boldsymbol{\mu}).$$
2. For each $n \geq 1, P_{\epsilon_n}(\boldsymbol{\mu}^n)$ admits an optimizer $\mathbb{P}_n \in \mathcal{M}_{\epsilon_n}(\boldsymbol{\mu}^n)$, i.e., $P_{\epsilon_n}(\boldsymbol{\mu}^n) = \mathbb{E}_{\mathbb{P}_n}[c]$. The sequence $(\mathbb{P}_n)_{n\geq 1}$ is tight, and every limit point must be an optimizer for $P(\boldsymbol{\mu})$. In particular, $(\mathbb{P}_n)_{n\geq 1}$ converges weakly whenever $P(\boldsymbol{\mu})$ has a unique optimizer.

Remark

1. By a theorem in Strassen (1965) we have $\boldsymbol{\mu} \in \mathcal{P}^{\preceq}$ if and only if $\mu_k \preceq \mu_{k+1}$ for $k = 1, .., N-1$, or namely, $\int f d\mu_k \leq \int f d\mu_{k+1}$ holds for all convex functions $f \in \Lambda$ and $k = 1, .., N-1,$. In addition, it follows by definition that $\mathcal{P}^{\preceq}_{r_n} \subset \mathcal{P}^N$ is convex and closed under $\mathcal{W}^{\oplus}$, and $\mathcal{M}(\boldsymbol{\mu}) \subset \mathcal{M}_\epsilon(\boldsymbol{\mu})$ for all $\epsilon \geq 0$.
2. **As mentioned earlier, we would like to approximate $P(\boldsymbol{\mu})$ by $P(\boldsymbol{\mu}^n)$ with finitely supported measures $\mu_1^n, ..., \mu_N^n$, since the latter reduces to an LP problem.**
3. The Lipschitz assumption can be slightly weakened. Let $E \subseteq \mathbb{R}^d$ be a closed subset such that $\text{supp}(\mu_k^n) \subseteq E$ for all $n \geq 1$ and $k = 1, .., N$. Then it suffices to assume in Theorem 1 that c, restricted to E^N, is Lipschitz.

LP formulation for finitely supported marginals

- The following Corollary shows that $P_{\epsilon_n}(\boldsymbol{\mu}^n)$ is equivalent to an LP problem.
- Henceforth, $P_{\epsilon_n}(\boldsymbol{\mu}^n)$ will denote the approximating LP problem of $P(\boldsymbol{\mu})$.

Corollary

Let $\boldsymbol{\mu}^n = (\mu_k^n)_{1 \leq k \leq N}$ be chosen such that each μ_k^n has finite support, i.e.

$$\mu_k^n(dx) = \sum_{i_k \in I_k} \alpha_{i_k}^k \delta_{x_{i_k}^k}(dx),$$

where $I_k = \{1, .., n(k)\}$ labels the support $\text{supp}(\mu_k^n)$. Denote by $p = (p_{i_1,..i_N})_{i_1 \in I_1,..,i_N \in I_N}$ the elements of $\mathbb{R}_+^D$ with $D := \prod_{k=1}^N n(k)$, then $P_{\epsilon_n}(\boldsymbol{\mu}^n)$ can be rewritten as an LP problem.

LP formulation for finitely supported marginals

PROOF. By assumption, every element $\mathbb{P} \in \mathcal{M}_{\epsilon_n}(\boldsymbol{\mu}^n)$ can be identified by some $p \in \mathbb{R}_+^D$. Therefore, $P_{\epsilon_n}(\boldsymbol{\mu}^n)$ turns to be the optimization problem below

$$\begin{aligned} &\max_{p \in \mathbb{R}_+^D} \sum_{i_1,..,i_N} p_{i_1,..,i_N} c(x^1_{i_1}, ..., X^N_{i_N}) \\ &\sum_{i_1,..,i_{k-1},i_{k+1},..,i_N} p_{i_1,..,i_N} = \alpha^k_{i_k}, \text{ for } i_k \in I_k \text{ and } k = 1, ..., N, \\ &\sum_{i_1,..,i_k} \left| \sum_{i_{k+1},..,i_N} p_{i_1,..,i_N}(x^{k+1}_{i_{k+1}} - x^k_{i_k}) \right| \leq \epsilon_n, \text{ for } k = 1, ..., N. \end{aligned} \tag{9}$$

The optimization problem (9) is not an LP formulation. However, by adding slack variables $(\delta^k_{i_1,i..,i_k,j})_{i_1 \in I_1,...,i_k \in I_k, j \in J} \in \mathbb{R}_+^{D_k}$ with $J := \{1, .., d\}$ and $D_k := d \prod_{r=1}^k n(r)$, (9) is equivalent to the following LP problem.

LP formulation for finitely supported marginals

$$\max_{p\in\mathbb{R}_+^D,\delta^1\in\mathbb{R}_+^{D_1},\delta^{N-1}\in\mathbb{R}_+^{D_{N-1}}} \sum_{i_1,..,i_N} p_{i_1,..,i_N} c(x^1_{i_1},...,X^N_{i_N})$$

$$\sum_{i_1,..,i_{k-1},i_{k+1},..,i_N} p_{i_1,..,i_N} = \alpha^k_{i_k}, \text{ for } i_k \in I_k \text{ and } k = 1,...,N,$$

$$-\delta^k_{i_1,..,i_k,j} \leq \sum_{i_{k+1},..,i_N} p_{i_1,..,i_N}(x^{k+1}_{i_{k+1}} - x^k_{i_k}) \leq \delta^k_{i_1,..,i_k,j}, \text{ for } i_k \in I_k, j \in J \text{ and } k = 1,..,N,$$

$$\sum_{i_1,..,i_k,j} \delta^k_{i_1,..,i_k,j} \leq \epsilon_n, \text{for } k = 1,...,N-1.$$

where we recall $x^k_{i_k} = (x^k_{i_k,1},..,x^k_{i_k,d})$. □

Convergence of the finitely supported marginals

The following theorem gives the convergence rate for $N = 2$ and $d = 1$.

Theorem

Let $N = 2$ and $d = 1$, or equivalently, $\boldsymbol{\mu} = (\mu, \nu)$ and $c : \mathbb{R}^2 \to \mathbb{R}$. In addition to the conditions of Theorem 1, we assume that $\sup_{(x,y)\in\mathbb{R}^2} |\partial^2_{yy} c(x, y)| < \infty$ and ν has a finite second moment. Then there exists $C > 0$ such that

$$|P_{\epsilon_n}(\mu^n, \nu^n) - P(\mu, \nu)| \leq C \inf_{R \geq 0} \lambda_n(R), \text{ for all } n \geq 1,$$

where $\lambda_n : (0, \infty) \to \mathbb{R}$ is given by

$$\lambda_n(R) := (R + 1)\epsilon_n + \int_{(-\infty,-R)\cup(R,\infty)} (|y| - R)^2 \nu(dy).$$

In particular, the convergence rate is proportional to ϵ_n if $\text{supp}(\nu)$ is bounded.

Case for finitely many option prices

Remark

In general, the distributions $\mu_1, .., \mu_N$ will not be fully specified by the market when $d \geq 2$. For $k = 1, .., N$, let $S_k := (S_k^{(1)}, ...S_k^{(d)})$, where $S_k^{(i)}$ stands for the price of the i^{th} stock at time k. Then, in practice, only prices of call options $(S_k^{(i)} - K)^+$, or put options $(K - S_k^{(i)})^+$, for a finite set of strikes K are actively traded in the market. Even assuming call options are quotes for all possible strikes K only yields the distributions $\mu_{k,i}$ of S_k^i. Therefore, this leads to a modified optimization problem. Denote $\vec{\mu}_k := (\mu_{k,1}, ..., \mu_{k,d})$ and $\vec{\boldsymbol{\mu}} := (\vec{\mu}_k)_{1 \leq k \leq N}$, and let $\mathcal{M}_\epsilon(\vec{\boldsymbol{\mu}})$ be the set of ϵ- approximating martingale measures $\mathbb{P}$ satisfying $\mathbb{P}(S_k^i)^{-1} = \mu_{k,i}$, for $k = 1, .., N$ and $i = 1, ..., d$. Then, we define the optimization problem by

$$P_\epsilon(\vec{\boldsymbol{\mu}}) := \sup_{\mathbb{P} \in \mathcal{M}_\epsilon(\vec{\boldsymbol{\mu}})} \mathbb{E}_{\mathbb{P}}[c(S_1, .., S_N)]. \tag{10}$$

The problem (10), with $\epsilon = 0$, was first introduced in Lim (2024) and was called *multi-martingale optimal transport.*

References

Beiglböck, M., Henry-Labordere, P., and Penkner, F. (2013). Model-independent bounds for option prices—a mass transport approach. *Finance and Stochastics*, 17:477–501.

Benamou, J.-D. and Brenier, Y. (2000). A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393.

Benamou, J.-D., Carlier, G., Cuturi, M., Nenna, L., and Peyré, G. (2015). Iterative bregman projections for regularized transportation problems. *SIAM Journal on Scientific Computing*, 37(2):A1111–A1138.

Breeden, D. T. and Litzenberger, R. H. (1978). Prices of state-contingent claims implicit in option prices. *Journal of business*, pages 621–651.

Davis, M., Obłój, J., and Raval, V. (2014). Arbitrage bounds for prices of weighted variance swaps. *Mathematical Finance*, 24(4):821–854.

Guo, G. and Obłój, J. (2019). Computational methods for martingale optimal transport problems. *The Annals of Applied Probability*, 29(6):3311–3347.

Lévy, B., Bastien, F., and Magnien, J. (2015). B. lévy-a numerical algorithm for l2 semi-discrete optimal transport in 3d.

Lim, T. (2024). Geometry of vectorial martingale optimal transportations and duality. *Mathematical Programming*, 204(1):349–383.