Market Making in TradFi, CeFi, and DEXs

A Quantitative Research Note

Amy O. Khaldoun

November 2025

Executive summary

Market making has one objective, provide liquidity in exchange for a spread, yet the mechanisms, constraints, and risks differ across three venues. In TradFi, queue position and inventory risk sit at the core of dealer quoting models. In CeFi, crypto order books inherit the same math but add extreme volatility, venue fragmentation, and operational frictions. In DEXs, automated market makers replace quotes with an invariant, which moves inventory risk into a continuous function of reserves and prices, introducing impermanent loss, just-in-time liquidity, and MEV exposure. The theory of payoffs and inventory utility largely survives, the microstructure and trust model change.

1 TradFi market making in limit order books

A standard reference is Avellaneda–Stoikov [1]. A dealer chooses bid and ask quotes to maximize expected utility of terminal wealth with an inventory penalty. Let S_t be the mid, δ^b, δ^a the distances from mid, and $\lambda^{b,a}(\delta)$ the Poisson arrival rates that decay with distance. With exponential utility one obtains closed form optimal quotes

$$\delta_t^{a,b} = \frac{1}{\gamma} \ln\Big(1 + \frac{\gamma}{k}\Big) \pm \frac{\gamma\sigma^2(T-t)}{2}\, q_t,$$

where q_t is inventory, σ is mid volatility, k is the order arrival slope, and γ is risk aversion. Inventory loads widen one side and tighten the other, queue priority rewards narrower quotes, and the spread trades off adverse selection against inventory risk. See extensions with general mid dynamics and explicit inventory penalties.

Market Making under Stochastic Volatility

Beyond the classical Avellaneda-Stoikov framework, which assumes constant volatility and Poisson order arrivals, more recent work extends the optimal quoting problem to stochastic volatility regimes. Aydoğan et al. [3] formulate a market-making optimization problem in a limit order book where the midprice follows a Heston process with potential jumps in both price and volatility.

They derive Hamilton–Jacobi–Bellman (HJB) equations under both quadratic and exponential utility functions, characterizing optimal bid–ask spreads as feedback controls depending on volatility state, inventory, and market order intensity. Numerical results show that higher volatility-of-volatility and jump components widen optimal spreads, while faster mean reversion in volatility narrows them.

This work bridges classical LOB control with stochastic volatility modeling, providing a quantitative baseline for extending inventory control to on-chain liquidity environments where volatility itself is protocol-driven rather than purely exogenous.

2 CeFi crypto market making on centralized exchanges

Crypto CLOBs reuse the same dealer logic, but three features change the calibration problem.

Fragmentation and depth Liquidity splits across many venues and pairs, depth varies through the day, and the removal of depth during shocks increases adverse selection and slippage relative to stationary backtests. Empirical work documents strong variation in book liquidity across crypto exchanges.

Volatility and inventory management Higher spot volatility widens optimal spreads and accelerates inventory mean reversion, which reduces realized fill rates at tight quotes. Evidence points to wider spreads on CEXs driven by volatility conditions.

Operational and data frictions Matching is off chain, funding and collateral are on chain, which creates timing mismatches. Tick data and order book snapshots lack uniform standards, which complicates model validation and backtesting.

3 Do CEX Market Makers Use Advanced Quantitative Models?

From a theoretical perspective, most academic models of market making such as Avellaneda–Stoikov, Guéant–Lehalle–Fernandez, or HJB-based stochastic control formulations offer mathematically elegant frameworks for inventory and spread optimization. However, their full implementation in live centralized exchange (CEX) environments is rare.

Theory versus production reality

In practice, CEX market makers do not solve partial differential equations or stochastic control problems in real time. Instead, they rely on simplified, data-driven approximations of the same principles. A typical spread or quoting rule may be expressed empirically as

$$\text{spread}_t = \alpha + \beta\,\sigma_t + \gamma\,q_t,$$

where σ_t is realized volatility and q_t denotes current inventory. Such adaptive linear or nonlinear models capture the intuition of optimal control without the computational burden.

What production systems actually optimize

Live CEX strategies focus primarily on execution efficiency and real-time risk aggregation. Key state variables include:

- Inventory deltas per symbol and per venue,
- Cross-exchange arbitrage spreads and funding costs,
- Order book imbalance and realized volatility filters,
- Latency and slippage metrics.

Rather than solving Hamilton–Jacobi–Bellman equations, firms use adaptive algorithms and reinforcement-learning heuristics to update quoting behavior.

Why full models rarely survive in production

Several frictions limit the direct use of continuous-time models:

1. **Computational constraints:** HJB or stochastic volatility models are slow to evaluate and unsuitable for millisecond reaction times.
2. **Parameter instability:** Estimated arrival intensities and volatilities vary rapidly across market regimes.
3. **Microstructure noise:** Order arrivals on CEXs are non-Poissonian and heavily influenced by bots and arbitrageurs.
4. **Empirical pragmatism:** Machine-learning models or rule-based systems perform more robustly under uncertainty.

What remains from the academic models

Even though the mathematics is simplified, the conceptual grammar persists:

- Utility maximization translates into adaptive risk control.
- Arrival intensities become empirical fill probabilities.
- Inventory aversion is implemented as penalized exposure or position caps.
- Optimal spread formulas evolve into volatility-scaled quoting rules.

All in all

Crypto market makers rarely run Avellaneda-Stoikov equations in production, but they operate under its spirit; continuously balancing fill probability, volatility, and inventory risk through empirical control rules. The equations define the theoretical language of liquidity provision, even when expressed operationally as data pipelines, regression models, or reinforcement-learning agents.

4 DEX market making with CFMMs and concentrated liquidity

On DEXs (like Uniswap), the market maker is an automated function. A constant function market maker admits reserves x, y that satisfy a trading function $\varphi(x, y) = k$. The constant product case is $xy = k$. Price is the marginal rate of substitution $p = \partial_x \varphi / \partial_y \varphi$. General CFMM geometry and multi asset formulations show that many invariants share common convex analytic structure.

Note: Most DEXs follow Uniswap's CFMM logic, but alter the invariant, fee dynamics, or liquidity geometry to specialize for use cases. Uniswap is the model, others are parameterizations of the same functional family

Uniswap v3 concentrated liquidity

Liquidity L is provided only on a price band, which creates option like exposures. Outside the band, the position is fully converted to one asset, inside the band the delta equals the slope of the reserve curve for the active ticks. Empirical and theoretical studies analyze fees, returns, and liquidity surface dynamics under concentration. [9]

Impermanent Loss and Fee Compensation

Impermanent loss (IL) represents the opportunity cost faced by a liquidity provider (LP) in a constant product automated market maker (AMM) relative to a passive "buy-and-hold" portfolio of the same two assets. It arises because the AMM continuously rebalances inventory to maintain the invariant

$$x\,y = k,$$

where x and y denote the reserves of the two tokens, and k is the invariant constant. This rebalancing changes the LP's asset composition as market prices move, causing the portfolio value to diverge from that of simply holding the assets.

Let:

- P_{old} = initial price of the risky asset (in units of the quote asset),
- P_{new} = price after a price change,
- $r = P_{\text{new}}/P_{\text{old}}$ = relative price change (price ratio),
- $V_{\text{AMM}}(r)$ = LP portfolio value after rebalancing at ratio r,
- $V_{\text{HODL}}(r)$ = value of the initial "buy-and-hold" portfolio at ratio r.

For a two-asset constant product AMM without trading fees, the relative value difference (impermanent loss) is

$$\mathrm{IL}(r) = \frac{V_{\mathrm{AMM}}(r)}{V_{\mathrm{HODL}}(r)} - 1 = \frac{2\sqrt{r}}{1+r} - 1.$$

This expression is negative whenever $r \neq 1$, meaning that any price deviation from the entry price produces a loss relative to holding the assets separately. The loss magnitude increases with price volatility, reflecting the LP's *short convexity* (short gamma) exposure.

Incorporating trading fees. Most AMMs compensate LPs with trading fees proportional to traded volume. Let feeRate denote the per-trade fee and flow(t) the traded notional per unit time. The cumulative fee income over an interval $[0, T]$ is

$$\mathrm{Fees}(T) = \int_0^T \mathrm{feeRate} \times \mathrm{flow}(t)\, \mathrm{d}t.$$

The realized LP profit-and-loss (PnL) becomes

$$\mathrm{PnL}(T) = \mathrm{Fees}(T) + V_{\mathrm{AMM}}(r_T) - V_{\mathrm{HODL}}(r_T),$$

where the second term captures impermanent loss. Profitability depends on whether fee income offsets this curvature-induced loss. Empirically, this balance hinges on trade frequency, volatility, and the cross-sectional mix of flow directions.

Recent quantitative results. Analytical and empirical studies [11] show that expected LP returns can be decomposed as

$$\mathbb{E}[\mathrm{PnL}] \approx \mathbb{E}[\mathrm{Fees}] - \frac{1}{2}\,\Gamma_{\mathrm{eff}}\,\sigma^2 T,$$

where Γ_{eff} denotes the AMM's effective curvature exposure and σ^2 the variance of the underlying asset price. This formulation parallels the gamma exposure of short-option positions: LPs earn fee income akin to option premium but incur convexity loss as prices move away from the entry ratio. The profitability threshold occurs when fee-induced drift compensates the volatility-driven convexity cost.

Just in time liquidity and MEV

Liquidity can arrive for a single swap then leave, which transfers fees from passive LPs to strategic LPs and changes fee incidence. Recent studies formalize JIT behavior and its impact on passive LP profits. MEV further alters expected execution for traders and LPs, through reordering and sandwich effects.

Strategic Liquidity Provision in Uniswap v3

Recent quantitative research has formalized the dynamic problem faced by LPs in Uniswap v3 as one of optimal stochastic control under reallocation costs. Fan et al. [8] model liquidity providers as agents choosing interval widths and rebalancing frequencies to maximize expected utility of terminal wealth subject to trading and gas costs. The authors define families of τ-reset strategies, where liquidity is concentrated in a price band and reallocated only when prices exit that range.

Empirical simulations based on Ethereum price paths show that dynamic, context-aware reallocation strategies can substantially outperform static or uniform allocations. Risk-averse LPs tend to spread liquidity over wider intervals, while low-cost or high-volatility environments favor narrow, frequently rebalanced ranges. The results confirm that Uniswap v3 behaves as a *continuous-time stochastic optimization environment*, where gas costs, volatility, and flow composition jointly determine optimal liquidity curvature.

From a quant perspective, this reframes AMM provision as an optimal control problem analogous to inventory models in market making, but with endogenous execution costs and algorithmic clearing.

Liquidity Fragmentation and Economies of Scale in Decentralized Exchanges

While Fan et al. [8] model liquidity provision as an optimal stochastic control problem, Lehar, Parlour, and Zoican [7] provide a complementary perspective by documenting the equilibrium implications of fixed transaction costs on decentralized exchanges.

Their model demonstrates that gas fees create economies of scale, driving heterogeneity among liquidity providers. Large, capital-rich LPs optimally concentrate in low-fee pools where trading is frequent, liquidity cycles are short, and fee income compensates for high adjustment costs. Smaller, retail LPs migrate toward high-fee pools where liquidity updates are infrequent and gas costs represent a smaller share of expected profits.

Using data from over 13 million Uniswap v3 transactions between May 2021 and September 2022, the authors find that high-fee pools attract about 56% of total liquidity but execute only 35% of volume. In contrast, low-fee pools handle most trading flow but involve fewer, larger LPs. Gas price shocks exacerbate fragmentation: a one standard deviation increase in gas prices reduces low-fee pool market share by roughly 2.3 percentage points and cuts liquidity inflows by more than a third.

From a quantitative standpoint, their results illustrate that liquidity provision in Uniswap v3 exhibits a structural segmentation similar to that between institutional and retail market makers in traditional limit order books, with blockchain transaction costs acting as a novel form of inventory friction.

5 Comparison: objective same, state variables differ

Venue	**Control Variable**	**Dominant Risks**
TradFi CLOB	Quote distances, inventory target	Adverse selection, queue position, inventory variance.
CeFi CLOB	Same as TradFi, plus venue selection	Depth evaporation, latency, data quality, venue fragmentation.
DEX CFMM	Fee and range placement, liquidity size	Impermanent loss, JIT competition, oracle quality, MEV exposure.

Table 1: Comparison of market making mechanics and dominant risks across venues.

6 Calibration and validation checklist

- TradFi and CeFi: estimate arrival rate elasticity k, volatility σ, and inventory penalty from fills versus distance to mid, with venue specific depth models. Benchmarks include Avellaneda–Stoikov and its extensions.
- DEX: estimate fee income conditional on flow and price path, compute IL path functionals, and perform counterfactuals with and without JIT arrivals. Use CFMM geometry for sensitivity and bounds.
- Stress tests: apply volatility shocks, depth withdrawal in CeFi, and fee or oracle latency shocks in DEXs. Consider MEV reordering scenarios for DEX execution quality.

Takeaways

The dealer utility framework still guides spread and inventory decisions. In crypto, the inputs are regime dependent. CeFi inherits limit order book behavior with larger volatility and fragmented depth. DEXs replace quotes with an invariant, which shifts the edge calculus from queue priority to curve selection, fee design, and protection against MEV and JIT behavior. Model success depends less on closed form elegance and more on microstructure aware calibration and stress design.

References

[1] Avellaneda, M. and Stoikov, S. (2008). *High-frequency trading in a limit order book*. Quantitative Finance, 8(3), 217–224.

[2] Menkveld, A. J. (2013). *High-frequency trading and the new market makers*. Journal of Financial Markets, 16(4), 712–740.

[3] Aydoğan, B., Uğur, Ö., & Aksoy, Ü. (2022). *Optimal Limit Order Book Trading Strategies with Stochastic Volatility in the Underlying Asset.* Computational Economics, 62, 289–324. https://doi.org/10.1007/s10614-022-10272-4

[4] Easley, D., Kiefer, N. M., and O'Hara, M. (2012). *Inventory and volatility in limit order markets.* Journal of Financial Economics, 106(3), 547–568.

[5] Biais, B., Foucault, T., and Moinas, S. (2015). *High-frequency trading and price discovery.* Review of Financial Studies, 28(12), 3387–3446.

[6] Angeris, G., Kao, A., Chiang, C.-K., and Chitra, T. (2021). *Optimal liquidity provision in Uniswap v3.* arXiv preprint arXiv:2106.14404.

[7] Lehar, A., Parlour, C., & Zoican, M. (2023). *Liquidity Fragmentation on Decentralized Exchanges.* Working Paper, University of Calgary, UC Berkeley, and University of Toronto. Available at SSRN: `https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4267429`.

[8] Fan, Z., Marmolejo-Cossio, F. J., Moroz, D., Neuder, M., Rao, R., & Parkes, D. C. (2024). *Strategic Liquidity Provision in Uniswap v3.* Harvard University and IOG. arXiv preprint arXiv:2106.12033.

[9] Adams, H., Zinsmeister, N., Salem, M., Keefer, R., and Robinson, D. (2021). *Uniswap v3 Core.* Uniswap Labs Whitepaper, March 2021. `https://uniswap.org/whitepaper-v3.pdf`.

[10] Angeris, G., Kao, A., Chiang, C.-K., Chitra, T., and Roughgarden, T. (2022). *Constant Function Market Makers: Multi-Asset Trades via Convex Optimization.* arXiv preprint arXiv:2203.11352. `https://arxiv.org/abs/2203.11352`.

[11] Cartea, Á., Jaimungal, S., and Riccardi, A. (2023). *Decentralized Market Making: Impermanent Loss and Liquidity Provision.* Working Paper, University of Oxford and University of Toronto. `https://arxiv.org/pdf/2309.08431` .