Where Does MEV Really Come From? Revisiting CEX–DEX Arbitrage on Ethereum

Bence Ladóczki[1], Miklós Rásonyi[2], and János Tapolcai[3]

[1] BME and HUN-REN Information Systems Research Group, Budapest, Hungary
[2] HUN-REN Alfréd Rényi Institute of Mathematics and Eötvös Loránd University, Budapest, Hungary, rasonyi.miklos@renyi.hu
[3] Budapest University of Technology and Economics (BME), Budapest, Hungary**

Abstract. A central question of the Ethereum ecosystem is where Maximal Extractable Value (MEV) revenue originates and to what extent it stems from harming unsuspecting users. It is acceptable if MEV arises from arbitrages between centralised and decentralised exchanges (CEX–DEX). Yet theoretical models have significantly underestimated the scale of these arbitrages, while empirical studies have highlighted their importance – though these remain conservative estimates, constrained by numerous debatable heuristic assumptions. Revisiting the theoretical model, we found that CEX–DEX arbitrages require trading volumes on the order of the total activity of major liquidity pools and yield profits comparable to MEV. Most prior AMM models utilised the Black–Scholes (BS) stochastic differential equation (SDE) – i.e., geometric Brownian motion – and assumed continuous price trajectories where asset prices move in small increments only. We argue that BS underestimates arbitrage profits by ignoring price jumps, which are precisely the points at which arbitrage opportunities tend to arise. To address this gap, we present an extended discrete-time AMM model in which the price process is the sum of a diffusive component and stochastic jumps that can have arbitrary noise distributions. Although mathematically more involved this framework allows us to employ a general discrete-time SDE and compute the stationary probability distribution via function iteration with geometric convergence. We further prove that the resulting mispricing process is an ergodic Markov chain. We implement our model in C++, collect spot prices and AMM exchange data from the Ethereum blockchain and fit the model parameters to the observed prices. The estimates derived from our model closely match empirical observations and provide a natural theoretical explanation for several fundamental questions in the blockchain ecosystem.

Keywords: Automatic Market Makers · Ethereum · CEX-DEX arbitrage · Stochastic Price Model · Constant Product Market Maker

** Dept. of Telecommunications and Artificial Intelligence, Faculty of Electrical Engineering and Informatics, Műegyetem rkp. 3., H-1111 Budapest, Hungary. ladoczki.bence@vik.bme.hu, tapolcai@tmit.bme.hu

1 Introduction

In the current implementation of the proof of stake (PoS) consensus protocol of Ethereum each slot is assigned to a single validator to propose a block, but in practice validators often resell this right at surprisingly high prices to so-called block builders. Therefore, block building and proposing are distinct processes in PoS Ethereum, hence the name proposal–builder separation (PBS). Today, one of the central questions in the Ethereum ecosystem is why block builders are willing to pay so much for this right, and more importantly, where this revenue originates and how 'dark' it is, that is, to what extent it derives from harming unsuspecting users. MEV (Maximal Extractable Value) denotes the profit that can be captured by including, excluding, or reordering transactions within a block. The most commonly observed forms of MEV are sandwich and liquidations attacks and arbitrage trades. Among these, only arbitrage is considered benign. It is regarded as a necessary mechanism for automated market makers (AMMs) [15] to keep exchange rates aligned with the external market.

Initially, MEV was believed to consist primarily of sandwich attacks, atomic (DEX–DEX) arbitrage, and back-running attacks – strategies that are relatively easy to detect on-chain and often yield conspicuously large profits. However, as these events are rare, their aggregate contribution is limited; see rows 3 and 4 of Table 1. In contrast, in the case of a CEX–DEX arbitrage only the DEX leg is on-chain, making it non-trivial to distinguish genuine arbitrage from noise trades. This perception aligned with theoretical models [23,22,25], which predicted only modest CEX–DEX arbitrage profits and volumes (see row 5 of Table 1). More recent empirical studies have demonstrated that CEX–DEX arbitrage plays a much larger role in MEV than previously assumed, with transaction volumes exceeding those of atomic arbitrage and sandwich attacks by several orders of magnitude [31,17,26] (see rows 2–4 of Table 1). These results suggest that the true scale of CEX–DEX arbitrage is substantially greater than predicted by existing theoretical models. With this work, we aim to correct and explain this discrepancy.

Table 1 compares empirical measurements with theoretical estimates of swap volumes and counts for a Uniswap V2 ETH–USDT pool (one of the largest by total value locked, TVL). The theoretical estimates are derived from a stochastic model describing the price process of AMMs and the arbitrageurs' response to CEX-DEX mispricing. Since such models can only be applied to a specific trading pair, the empirical data from [30] was filtered to this pool. Roughly speaking, the theoretical estimates represent the expected frequency and volume of arbitrage transactions implied by the SDE on an infinite-horizon. We argue that accurate stochastic models should yield estimates consistent with empirical observations.

Most previous theoretical work tackled the problem with simplified tools and derived closed-form equations to describe losses in AMMs and arbitrage profits [23,22,25]. The trade-off for this computational simplicity lies in that these AMM models are essentially all based on the assumption that price trajectories are continuous and that the spot price of the asset under consideration can only change by a small amount in a small interval. The backbone of these frameworks

		A Uniswap V2 pool in Sept. 2025	USDT → WETH		WETH → USDT		MEV
		0x0d4a11d5eeaac28ec3f61d100daf4d40471f1852	#swaps	volume	#swaps	volume	/profit
1	Empirical	All swap transactions	2724	$3 369 073	1907	$3 072 624	∼ $2 030
2		CEX-DEX arb. (heur. in [30])	496	$970 436	241	$1 002 020	
3		Sandwich (avg. in [30])	0.2	$25 345	0.2	$64 443	$414
4		Atomic arb. (avg. [30])	3	$10 515	2	$34 562	$136
5	SDE	Black–Scholes (LVR [23,22,25])	86	$208 844	75	$177 961	$48
6		Our model	256	$1 881 277	251	$1 838 985	$1 454

Table 1: Average daily swaps and volumes, comparing all transactions with arbitrage transactions between 13 − 29 September 2025 in one of the largest Uniswap V2 pools by TVL ($86M). For context, the total DEX trading volume during this period was roughly $462B, with Uniswap alone accounting for $106B. The upper part of the table reports empirical measurements, while the lower part presents theoretical estimates based on various stochastic differential equation (SDE) models for CEX–DEX arbitrage. Our model explicitly involves price jumps, with parameters fitted from the exchange rate time series (see Section 4.1). Empirical MEV was estimated by considering only single-swap transactions (ignoring bundles), summing their priority fees and direct coinbase transfers, and then scaling the result proportionally to account for bundle transactions.

is the Black–Scholes (BS) price model developed in the 1970s. While BS is a commonly used workhorse in financial mathematics, many believe that arbitrage activity tends to emerge precisely at price jumps. However, incorporating price jumps into AMM models requires more sophisticated approaches than a two-parameter SDE can provide.

The main contribution of this paper is an extended discrete-time AMM model that integrates both a diffusive component and stochastic jumps, while accommodating arbitrary noise distributions. Specifically, we employ a discrete-time general SDE and show that the stationary probability distribution (SPD) can be computed via function iteration that converges with geometric speed. We prove that the state space of the process under consideration is a small set, which implies that the misprice process is an ergodic Markov chain. Our extension can be reduced to the BS model, the Merton jump diffusion SDE [20] or the Ornstein-Uhlenbeck SDE [29] when the model parameters are chosen accordingly.

After calibrating the model parameters to match the statistical properties of real-life data, the SDE produces estimates similar to the observed values. In essence, by determining seven parameters and running multiple rounds of numerical integration, we obtained an estimated arbitrage volume that was roughly twice than that observed on-chain. Note that our model (as well as prior ones [23,22,25]) excludes noise traders and assumes the arbitrageur trades with high probability whenever the CEX–DEX price gap is sufficiently large. In other words, a significant proportion of the 24-hour trading volume appears to

be driven by the pool's need to track the CEX price[4]. Thus, in terms of volume, noise traders—agents who swap irrespective of the CEX price—play only a minor role (they are likewise omitted in the models of [23,22,25])

In empirical studies, swaps classified as CEX–DEX arbitrage by strict heuristics, e.g. MEV bots run by searchers. These heuristics exclude transactions seen in the public mempool, require that the trade be the first swap in its direction within a given pool and a block, and filter out atomic MEV strategies such as sandwiches, DEX-DEX arbitrage and liquidation attacks. Additional filters remove order flow auction (OFA) backruns, transactions routed through known smart contracts or labeled bots, and swaps involving NFTs [17,30].

Our theoretical results suggest that a significant portion of the remaining swap volume is also executed by (hidden) arbitrageurs. However, when looking at transaction counts, the observed number of swaps is higher than predicted by our theoretical model, which assumes that each arbitrage opportunity is exploited by a single arbitrageur. Note that the specific actor may differ across instances; from the perspective of the model, the identity of the arbitrageur is irrelevant. This discrepancy can be explained by the presence of many arbitrageurs in the network–each utilising slightly different CEX prices–as well as by the widening of the CEX bid–ask spread following price jumps. As a result, the CEX–DEX price gap is sometimes closed not by a single transaction, but through a small sequence of transactions in practice.

The paper is organised as follows. First, Section 2 provides a brief overview of related work, then the considered price dynamics is discussed. Next, in Section 3 the invariant probability distribution of the misprice process is given in terms of a function iteration formula. A numerical solver to determine the invariant distribution is described in Section 4, where we also present measurements for cross-validation of our findings. We conclude our work in Section 5.

2 Stochastic Price Models

First, let us review some of the SDEs utilised in previous works. [5] derives its result relying on an n-dimensional Brownian motion, i.e. $\frac{dS}{S} = \mu_t dt + dW_t$ for n assets, using the usual definition for the model parameters. [23] assumes that the CEX price is observable, and evolves exogenously according to a geometric Brownian motion (GBM) that is a continuous Q-martingale. Their SDE is $\frac{dP_t}{P_t} = \sigma_t dB_t^{\mathbb{Q}}$, with $B_t^{\mathbb{Q}}$ being the Brownian motion. In financial mathematics, a Q-martingale often refers to processes defined under the risk-neutral measure (Q) [10] and its future expected value equals its current value, given the past information. The case without a price trend is particularly convenient (notice that μ_t is missing from the aforementioned SDE), as it allows for the analysis of the system's asymptotic behavior.

[4] This implies that the 1-day vol/TVL metric greatly depends on the exchange-rate dynamics of the token pair and the DEX fee. Interestingly, for Uniswap V2 pools between a given currency and various stablecoins, this metric takes similar values regardless of the pool's TVL (see https://app.uniswap.org/explore/pools).

Studies [6,7,22] abandon the martingale assumption and extends the model by using the SDE $\frac{dP_t}{P_t} = \mu dt + \sigma dB_t$ to allow for the presence of price trends (μ), deriving formulas for this more general case. Their SDE is not governed by a driftless geometric Brownian motion, but by the full Black–Scholes model [4]. It is important to note that the Black–Scholes model has been a cornerstone of quantitative finance since the 1970s. However, more sophisticated models have since emerged, offering better alignment with observed market behaviour. In particular, the Black–Scholes model fails to capture several important stylised facts about log-price processes, such as the presence of heavy tails [8].

As emphasised in [20], the Black-Scholes model is only valid when trading takes place in continuous time and the price process has continuous paths. In other words, the exchange rate has a "local" Markov property, namely that it cannot change significantly in a short time interval. The conjecture of the current work is that the assumption that trading takes place in continuous time and that the price trajectories are continuous is misleading. To underpin our suspicion, we analysed two months of trading data provided by the Binance Spot price API between November and December 2024 and found that in some cases, the amplitude of price jumps in the ETH price can be two orders of magnitude larger than the instantaneous volatility. For a concrete example, we point out the jump that occurred at 2024-12-09 22:06:29, when the spot price changed from 3542 to 3640 in a second. Note that this was not the only occurrence of a huge jump in the spot price of ETH. This clearly indicates that prior works relying on the continuous path assumption cannot accurately describe the interaction between the arbitrageurs and the DEXs. In this work, we fill this research gap and investigate the effects of discontinuous price trajectories. In a somewhat similar fashion to our objectives, Dewey et al. in [11] use a Merton jump diffusion SDE in the following form: $\frac{dp_t}{p_t} = (\mu_D - \lambda_J k)\, dt + \sigma_D\, dB_t + (y_t - 1)\, dN_t$ (refer to the original paper for the definitions) and investigate AMM dynamics with this more comprehensive price model. Regarding the generality of the proposed mathematical framework, our model covers the two-parameter BS model, the Merton jump diffusion SDE and the widely used mean-reverting Ornstein-Uhlenbeck [29] SDE as well.

2.1 Price dynamics

Next, we introduce the equations to be then used to describe and analyse arbitrage opportunities and arbitrageur profits while considering a large class of price models. Our notational convention is presented in Table 2. We choose seconds as a unit to measure time, mostly because the block time of Ethereum is 12s, therefore it is more convenient to convert all the model parameters to seconds and use them accordingly. The results in the tables are converted back to daily values.

First, let us set the conditions of our mathematical model. At time $n \in \mathbb{N}$ the exchange rate of a risky asset (e.g. ETH), against a stablecoin (e.g. USDT or USDC) is $P_n > 0$ at a CEX. We define $X_n := \ln(P_n)$. The DEX price is usually different from this due to the passive nature of traditional on-chain AMMs.

Notation	Explanation
X_n	CEX price at time n
σ, μ	Volatility and drift
ϵ	Noise
Z_n	Jump indicator
U_n	Jump amplitude
q	Jump probability
p	Arbitrage trade probability

Notation	Explanation
$\hat{X}_n$	Auxillary process
$\tilde{X}_n$	AMM dynamics
γ_-, γ_+	Fee levels
$u()$	Jump amplitude distribution
$f()$	Noise density
$f_k()$	k-th iterated density
$f^*()$	Invariant density

Table 2: Notations used in this work.

At time n it is expressed as $\tilde{P}_n > 0$. The (logarithmic) mispricing [22] at time n is then defined as $\tilde{X}_n := \ln(P_n/\tilde{P}_n)$, $n \in \mathbb{N}$. We treat $X_n, \tilde{X}_n$ as a stochastic processes, i.e. X_n, $n \in \mathbb{N}$ are a sequence of random variables on a fixed probability space $(\Omega, \mathcal{F}, P)$ with the usual definitions. Expectation of a random variable Y will be denoted as $E[Y]$.

Remark 1. Throughout this work we work under the objective (statistical) probability P, and *not* under a risk-neutral probability. Risk-neutral measures are convenient tools for derivative pricing, but they are artifacts without statistical meaning, and hence not suitable for empirical analysis based on exchange rate data. Moreover, risk-neutral probabilities are typically well-defined only on finite horizons, whereas our results concern asymptotic properties (e.g., the limiting distribution of mispricing) as the time horizon tends to infinity. In this context, the objective probability is the only meaningful choice.

2.2 Recursive Description

We now describe the random movement of the CEX price process X_n, $n \in \mathbb{N}$. We choose to describe our stochastic processes in a discrete fashion, because computer simulations proceed in discrete time steps anyway, so the following recursive description is employed:

$$X_{n+1} - X_n = \mu(X_n) + \sigma(X_n)\varepsilon_{n+1} + Z_{n+1}U_{n+1}, \ n \in \mathbb{N}. \tag{1}$$

In other words, the CEX price process is governed by both diffusive and jump terms. The first two terms on the right-hand side of (1) are the *diffusive* terms: the function $\mu : \mathbb{R} \to \mathbb{R}$ is the *drift* and $\sigma : \mathbb{R} \to \mathbb{R}_+$ is the *volatility*. ε_n, $n \in \mathbb{N}$ is an i.i.d. noise term. We assume that the functions μ, σ are (Borel)-measurable and the common law of the ε_n, $n \in \mathbb{N}$ is absolutely continuous with density function $f(x)$.

The third term on the right-hand side of (1) represents random *jumps*. We assume that Z_n is an i.i.d. $\{0,1\}$-valued sequence, indicating whether a jump occurs at time n (that is, $Z_n = 1$) or not ($Z_n = 0$). We assume that $q := P(Z_0 = 1) < 1$. The *jump sizes* U_n, $n \in \mathbb{N}$ are i.i.d. with absolutely continuous law and the corresponding density function is $u(x)$.

The initial value X_0 can be either deterministic or random, but we always assume that X_0, $(\varepsilon_n)_{n\in\mathbb{N}}$, $(Z_n)_{n\in\mathbb{N}}$ and $(U_n)_{n\in\mathbb{N}}$ are independent. This immediately guarantees that X_n is a homogeneous Markov chain: from the current state $X_n = x$ the next state X_{n+1} is calculated as $x + \mu(x) + \sigma(x)\varepsilon_{n+1} + Z_{n+1}U_{n+1}$, using the i.i.d. noise sequence $(\varepsilon_n, Z_n, U_n)$ and no information is carried over from the past.

Remark 2. The interpretation of (1) is simple: the process has small, "diffusive" movements, but unexpected jumps are also allowed to occur. We believe that this extension enables one to describe highly volatile cryptocurrencies more accurately, see for example [11]. The term $Z_{n+1}U_{n+1}$ is the discrete-time analogue of a compound Poisson process (or, more generally, of a Lévy process). Indeed, let V be the first time a jump occurs (that is, the first n such that $Z_n = 1$). Clearly, the random variable V has a geometric distribution, which is the discrete-time analogue of the exponential distribution that appears between two consecutive jump times of a Poisson process.

Remark 3. The model (1) is the discrete-time counterpart of the continuous-time jump-diffusion model (considered e.g. in [20,14]) which satisfies the following stochastic integral equation:

$$X_t = X_0 + \int_0^t \mu(X_s)\, ds + \int_0^t \sigma(X_s) dW_s + \sum_{j=1}^{N_t} \xi_j, \tag{2}$$

where W_t is the Brownian motion, N_t is a Poisson process and ξ_j are i.i.d. jump amplitudes. $(W_t)_{t\in\mathbb{R}_+}, (N_t)_{t\in\mathbb{R}_+}, (\xi_n)_{n\in\mathbb{N}}$ are independent. Note, however, that the discretisation of (2) would result in Gaussian ε_n while in (1) we do not restrict the noise to be of Gaussian distribution.

Example 1. When μ, σ are constants, ε_i are standard Gaussian, $P(Z_0 = 0) = 1$ (that is, no jumps) then we obtain a *random walk* model [2], the discrete-time version of geometric Brownian motion that was meticulously analysed in [22].

When $\mu(x) := -\alpha(x - \theta)$ for some $0 < \alpha < 2$ and $\theta \in \mathbb{R}$, σ is constant and $P(Z_0 = 0) = 1$, we are working with an *autoregressive* model, the discrete-time analogue of an Ornstein-Uhlenbeck process [29]. One can use more complex μ, σ functions as it usually happens in the field of quantitative finance, see the monograph [13]. We stress once more that ε_n may be non-Gaussian. In fact, empirical evidence shows that ε_n has (much) fatter tails than a regular Gaussian distribution, see [8], hence realistic models *are* non-Gaussian. In Section F we will evaluate the logarithmic returns of several cryptocurrency pairs and perform statistical tests to demonstrate that the distribution of ε_n is far from Gaussian when considering real-life exchange rates.

If $0 < P(Z_n = 1) < 1$ then there is a nontrivial jump component in the discrete price process. We do not consider the case $P(Z_n = 1)$ (when there is always a jump), which slightly simplifies the ensuing proof.

In continuous time, such models (more generally, Lévy-process based models) enjoy popularity, see [9,13]. Adding jumps is a possible way of reconciling

empirical evidence with theoretical models in quantitative finance, consult [8] again and refer to Table 6 below for statistical tests on cryptocurrencies.

We will now shift our focus to examining the effects of arbitrageurs' presence in the market. Let $\gamma_+, \gamma_- > 0$ denote the exchange fee in units of log price for buying and selling tokens in the AMM following the notations of [22]. The arbitrageur trades against the AMM as follows. At each moment n, a myopic arbitrageur (maximising short-term profits) arrives at the AMM with probability $0 < p \leq 1$. When the logarithm of the misprice is outside $[-\gamma_-, \gamma_+]$ then the arbitrageur sets it back to the closest boundary point (that is, either to γ_+ or to $-\gamma_-$) thus making a profit. Otherwise no action is taken. Note that when there is no arbitrage activity against the AMM, the misprice process is identical to the price process. Consequently, the logarithm of the misprice continues to evolve according to (1).

Rigorously this can be formulated as follows: let A_n, $n \in \mathbb{N}$ be an i.i.d. sequence of $\{0,1\}$-valued random variables with $0 < P(A_0 = 1) = p \leq 1$. We assume that X_0, U_n, ε_n, Z_n, A_n are independent. $A_n = 1$ means that there is an arbitrageur arriving at discrete time n. Let us introduce the stochastic process $\tilde{X}_n$, $n \in \mathbb{N}$ to describe the dynamics of the mispricing between the CEX and DEX prices, with arbitrageurs taken into account. We fix $\tilde{X}_0$ (for instance $\tilde{X}_0 = 0$) and recursively define, for $n \in \mathbb{N}$,

$$\begin{aligned}\hat{X}_{n+1} &:= 1_{\{\tilde{X}_n > \gamma_+\}} A_{n+1}\gamma_+ - 1_{\{\tilde{X}_n < -\gamma_-\}} A_{n+1}\gamma_- \\ &\quad + (1 - A_{n+1})\tilde{X}_n 1_{\{\tilde{X}_n \notin [-\gamma_-, \gamma_+]\}} + \tilde{X}_n 1_{\{\tilde{X}_n \in [-\gamma_-, \gamma_+]\}}, \qquad (3)\\ \tilde{X}_{n+1} &:= \hat{X}_n + \mu(X_n) + \sigma(X_n)\varepsilon_{n+1} + Z_{n+1}U_{n+1}. \qquad (4)\end{aligned}$$

The expressions above can be explained as follows: when an arbitrageur arrives at time n and the current price $\tilde{X}_n$ is above γ_+ then the arbitrageur sets it back to γ_+; if the current price is below $-\gamma_-$ then the price gets set back to $-\gamma_-$; if no arbitrageur arrives then no action is taken even if the process $\tilde{X}_n$ is outside $[-\gamma_-, \gamma_+]$. When the process is *inside* $[-\gamma_-, \gamma_+]$ then nothing happens no matter what. In such a way we arrive at the value $\hat{X}_n$ as a result of arbitrage trades. After this, the mispricing process evolves following the stochastic dynamics of the CEX price, (1). That is, $\tilde{X}_{n+1}$, the value of the new misprice at time $n+1$, can be calculated from $\hat{X}_n$ in the same way as X_{n+1} was calculated from X_n in (1).

One can readily check that when $\{A_1 = 0, \ldots, A_n = 0\}$ and $X_0 = \tilde{X}_0$, one has $X_k = \tilde{X}_k$, $0 \leq k \leq n$, that is, without the intervention of arbitrageurs (3) reduces to the original CEX price evolution (1). Owing to the fact $(X_n, \tilde{X}_n)$ is being generated from an i.i.d. noise sequence $(A_n, Z_n, U_n, \varepsilon_n)$ with a recursion, $(X_n, \tilde{X}_n)$ is, again, a (homogeneous) Markov chain.

3 Invariant Probability Distribution

Now we make some technical assumptions about various components of the mispricing model $\tilde{X}_n$ described in the previous section.

Assumption 1 *$\mu(x) = \mu \in \mathbb{R}$ and $\sigma(x) = \sigma > 0$ are constants.*

This assumption guarantees that the incoming noise always enters the system with a non-vanishing coefficient σ, in other words, there is enough randomness to make the system ergodic. Under this Assumption, $\tilde{X}$ itself is a Markov process, as easily seen.

Finally, some regularity of the density f is required. The latter assumption holds, in particular, if f is continuous and positive everywhere, which is, again, a mild requirement that is fulfilled in all imaginable models in practice.

Assumption 2 *Let $K := |\mu| + \sigma$. Let $H := (2\max\{\gamma_+, \gamma_-\} + K)/c_0$. The density f is bounded away from 0 in the following sense:*

$$c_1 := \inf_{|x| \le H} f(x) > 0.$$

Now let us calculate the conditional density of the CEX price increments, expressed as the convolution of a rescaled density and the corresponding jump term.

Lemma 1. *The law of $X_{n+1} - X_n$ conditionally to $X_n = x$ is absolutely continuous with (conditional) density*

$$h(x, y) := (1 - q)\frac{1}{\sigma} f\left(\frac{y - \mu}{\sigma}\right) + q \int_{\mathbb{R}} u(y - z)\frac{1}{\sigma} f\left(\frac{z - \mu}{\sigma}\right) dz,$$

that is, for all Borel sets A,

$$P(X_{n+1} - X_n \in A | X_n = x) = \int_A h(x, y)\, dy.$$

Next, we turn our attention to the Markov process $\tilde{X}_n$ and determine its transition density.

Lemma 2. *For each Borel set A,*

$$P(\tilde{X}_{n+1} \in A | \tilde{X}_n = x) = \int_A q(x, y)\, dy,$$

where the transition density $q(x, y)$ is given by

$$q(x, y) = h(x, y - x)(1 - p) + h(\tilde{x}, y - \tilde{x})p.$$

Here we use the notation

$$\tilde{x} := x 1_{\{x \in [-\gamma_-, \gamma_+]\}} + \gamma_+ 1_{\{x > \gamma_+\}} - \gamma_- 1_{\{x < -\gamma_-\}}.$$

Let μ_0 denote the law of $\tilde{X}_0$. For instance, if $\tilde{X}_0$ is a constant w then $\mu_0 = \delta_w$, the probability concentrated on the one-point set $\{w\}$.

Corollary 1. *Whatever $\tilde{X}_0$ is, the random variables $\tilde{X}_k$, $k \geq 1$ have laws that admit density functions $f_k(x)$ given by*

$$f_1(x) = \int_{\mathbb{R}} q(x_0, x)\mu_0(dx_0)$$

and by the recursion

$$f_{k+1}(x) = \int_{\mathbb{R}} q(x_0, x) f_k(x_0)\, dx_0, \;\; k \geq 1. \tag{5}$$

Now we are in a position to state our main theoretical result: for the misprice process in our model a SPD **exists**, the system under consideration tends to this invariant state from an arbitrary initialisation (at a geometric speed) and the law of large numbers also holds.

Theorem 3. *Denote by f_n the density function of $\tilde{X}_n$ for $n \geq 1$. Let Assumptions 1 and 2 be in force. Then there is a density function f^* such that such that*

$$\int_{\mathbb{R}} |f^*(t) - f_n(t)|\, dt \leq C\rho^n, \;\; n \in \mathbb{N}, \tag{6}$$

where f^* is the invariant density of the Markov process $\tilde{X}_n$. Here $C > 0$, $\rho < 1$ are constants.

For an arbitrary bounded, measurable function $\phi : \mathbb{R} \to \mathbb{R}$ one has almost sure convergence

$$\frac{\phi(\tilde{X}_1) + \ldots + \phi(\tilde{X}_n)}{n} \to \int_{\mathbb{R}} \phi(x) f^*(x)\, dx, \quad n \to \infty. \tag{7}$$

Furthermore, one can approximate f^* by the function iteration (5) *whose convergence rate is estimated by* (6).

As an application of statement (7), we determine the long-term relative frequency of the mispricing process being outside the interval $[-\gamma_-, \gamma_+]$. (This quantity was denoted by $\mathtt{P}_{trade}$ in [22].)

Corollary 2. *Almost surely, as $n \to \infty$,*

$$\frac{\sum_{j=1}^n 1_{\{\tilde{X}_j \notin [-\gamma_-, \gamma_+]\}}}{n} \to \mu_*(\mathbb{R} \setminus [-\gamma_-, \gamma_+]) = \int_{\mathbb{R}\setminus[-\gamma_-, \gamma_+]} f^*(t)\, dt.$$

Here μ_* denotes the invariant measure whose density is f^*, to which the system converges as time tends to infinity, see (6) above.

In [22] the arbitrage profit *rate* was calculated, that is, the profit earned in an infinitesimally small interval. In the present discrete-time setting, we can also determine the arbitrageur's profit in one time step. Assume now that the system is in stationary state (that is, $\tilde{X}_n$ has law μ_* for all n).

We consider a CPMM with total value locked $\mathtt{TVL}$ and we assume that $\gamma_+ = \gamma_- =: \gamma$ (see [23] for a more detailed discussion).

Corollary 3. *For a CPMM in stationary state, the expected profit of the arbitrageur at any given instant can be expressed using the SPD f^* as follows:*

$$\mathtt{TVL}\cdot p\left[e^{\gamma/2}\int_{\gamma}^{\infty}\left(\cosh\left(\tfrac{t-\gamma}{2}\right)-1\right)f^*(t)\,dt+e^{-\gamma/2}\int_{-\infty}^{-\gamma}\left(\cosh\left(\tfrac{t+\gamma}{2}\right)-1\right)f^*(t)\,dt\right].$$

Corollary 4. *For a CPMM in stationary state, the expected arbitrage trading volume at any given instant can be expressed as*

$$\mathtt{TVL}\cdot p\left[\int_{\gamma}^{\infty}\left(e^{\gamma/2}-e^{-t/2+\gamma}\right)f^*(t)\,dt+\int_{-\infty}^{-\gamma}\left(e^{-\gamma/2}-e^{-t/2-\gamma}\right)f^*(t)\,dt\right].$$

Remark 4. The parameter choice $p = 1$ corresponds to the situation where arbitrageurs arrive at *every* time instant. This is the discrete-time version of the "fast-block regime" treated in [22], our results above apply directly to this situation too.

4 Numerical Results

To ascertain the validity of our mathematical model, we developed a numerical integrator in C++ to calculate the SPD of the misprice process for different model parameters[5]. Downstream, the logarithmic misprice process, taken from [22], is understood as:

$$z_t \triangleq \log P_t/\tilde{P}_t, \tag{8}$$

where $\tilde{P}_t$ is the DEX price and P_t is the external market price. Our C++ driver calculates the invariant probability distribution of this process.

Appendix D analyses the convergence properties of the stationary distribution. Our results indicate that after approximately 100 iterations, the function iteration yields curves that are nearly indistinguishable from the limiting distribution. Thus in the rest of the section the function iteration was run for 100 steps on a grid of 201 points.

In Appendix E, we demonstrate that our model can reproduce the numerical values reported in Table 1. of [22] by computing the SPD to arbitrary order and accuracy.

Appendix F analyses hourly log-returns of major cryptocurrencies (Table 6). The results show heavy tails, skewness, and uniformly low Kolmogorov–Smirnov p-values, confirming that Gaussian-based models such as Black–Scholes cannot capture the empirical dynamics. This motivates the need for richer model families, including jump-augmented processes.

4.1 Fitting a model on empirical data

To shed light on the applicability and the accuracy of our mathematical framework, we calculate the SPD of Eq. 8 with and without jumps, see Fig 1. We

[5] Source code: https://gitlab.com/jtapolcai/sde_amm

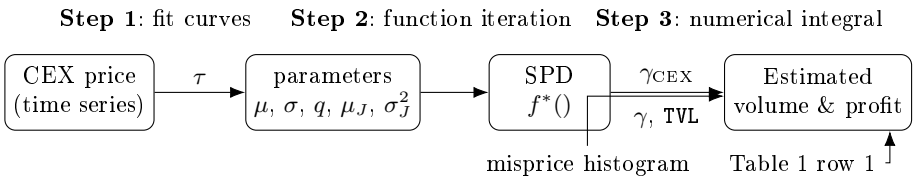


Fig. 1: Flowchart of stochastic model evaluation on real data from an AMM.

extract blockchain data using our own execution client, collecting transactions from the Uniswap V2 ETH–USDT AMM over the period 13–29 September 2025.

The values of $\tilde{P}_t$ are computed as follows. In parallel, we collect exchange-rate data at one-second resolution from multiple CEXs[6]. We assume competition among arbitrageurs: the arbitrageur who can source the asset at the best external price ultimately "wins," as they can afford to pay the highest transaction fee and therefore have their transaction included on-chain. Accordingly, we compute the mispricing by comparing the best bid (or best ask) observed on the AMM to the best prices available across all monitored CEXs. We define a 12-second time window and assume that if, at any moment within this window, any CEX offers a better price, then some arbitrageur captures the opportunity and submits a transaction. We estimate that blocks arrive at our node with an approximately 8-second delay; the start of the 12-second window is aligned accordingly. The intuition behind this choice is that blocks are typically proposed within the first four seconds of a slot, while measuring this delay precisely would be extremely difficult. Finally, $\tilde{P}_t$ is inferred from the swap events' amount-in and amount-out values.

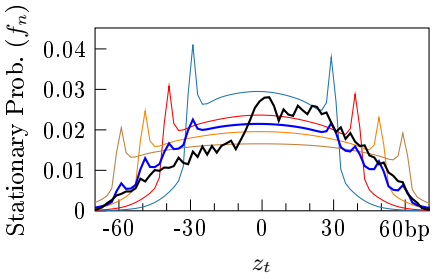
Our stochastic model computes the expected number of arbitrages, their total volume, and the expected profit using seven parameters (inputs): TVL ($\texttt{TVL}= 2L\sqrt{W}$), exchange fee (γ), daily price volatility (σ) and drift (μ), and jump parameters — per-step jump probability (q), and jump-size distribution $U \sim \mathcal{N}(\mu_J, \sigma_J^2)$ with mean μ_J and variance σ_J^2.

In Step 1, we estimate μ, σ, q, μ_J, and σ_J^2 from spot price data. Recall that in our stochastic model, price movements arise from either a geometric Brownian motion (GBM) or discrete price jumps. In practice, these two sources cannot be perfectly distinguished. Therefore, a fixed jump-threshold parameter τ is commonly used: if the absolute return is smaller than τ times the price volatility, the movement is classified as GBM; otherwise, it is classified as a jump. If τ is chosen too small, the jump fitting becomes inaccurate because many observations that actually belong to the GBM process are incorrectly classified as jumps. Conversely, if τ is chosen too large, too few jump observations remain for reliable parameter estimation. The measurements in the observed time period were not particularly sensitive to the choice of τ; however, as a rule of thumb, we suggest setting $\tau = 2.0$. Under a Gaussian assumption, this

[6] Best ask price originate from Binance (84%), KuCoin (9%), and Gate.io (7%), while best bids originate from Binance (88%), KuCoin (5%), and Gate.io (7%).

threshold implies that approximately 4.55% of GBM observations (two-sided) are incorrectly included in the jump sample, which represents a reasonable trade-off between bias and variance in the jump-parameter estimation.

We then fit Gaussian curves to the empirical mispricing histograms: one to the GBM component, yielding estimates of μ and σ, and another to the jump component, providing μ_J and σ_J for the jump-size distribution U (these estimates were rescaled to match the time scale of our model). The jump probability q is inferred using the threshold τ, which was set to attain higher log-likelihood against the observed time series.



Model	Volume	Profit	Vol/P
$\gamma_{\text{CEX}} = 0$ bp	75%	93%	1312
$\gamma_{\text{CEX}} = 10$ bp	58%	72%	1264
$\gamma_{\text{CEX}} = 20$ bp	46%	57%	1264
$\gamma_{\text{CEX}} = 30$ bp	39%	48%	1267
Linear fit	51%	63%	1274
Histogram of misprice			

Fig. 2: Comparison of the SPD for five different values of γ_{CEX} with the empirical distribution. A linear combination fit was also introduced to approximate the empirical mispricing histogram, yielding weights $w_0 = 19\%$, $w_{10} = 19\%$, $w_{20} = 28\%$, and $w_{30} = 34\%$. A histogram of the empirical mispricing is also shown. The legend further reports the share of the computed arbitrage volume relative to the empirically measured total pool volume, as well as the share of profit relative to the MEV-based estimate (see row 1 of Table 1). The last column of the legend reports the volume-to-profit ratio, which is only marginally affected by γ_{CEX}.

With $\tau = 2.0$, the estimated daily volatility is 1.9%, the drift is 0.023, and the jump probability is $q = 0.045$ per one-second step, with mean -24.9% and standard deviation 15.7% in log-return units. For comparison, when fitting the same parameters under a pure diffusion model (i.e., setting $q = 0$), we obtained a daily volatility of 2.1% and a drift of 0.046.

Next, in Step 2, we evaluate the function iteration taking a few minutes with 200 steps. Finally, in Step 3, we compute trade volumes using the formula in Corollary 4, profits using Corollary 3, and transaction counts corresponding to the lower- and upper-tail integrals of the SPD. To this end, we needed pool parameters (TVL and γ are directly observable). In practice, however, when solving the SDE we opt to use effective γ that also includes the CEX leg. This is denoted by γ_{CEX}, i.e., $\gamma_- = \gamma_+ = \gamma + \gamma_{\text{CEX}}$. Note that this essentially corresponds

to the CEX bid–ask spread, which is not precisely known, and, as Fig. 2 shows, our calculations are highly sensitive to this value.

Fig. 2 compares the observed misprice histogram with the stationary distribution f obtained from the SDE under different values of the γ_{CEX} parameter. In our model, γ_{CEX} is constant, which results in two sharp discontinuities of the distribution at γ_- and γ_+. In practice, however, the boundary is blurred. This can be explained either by temporal variation in γ_{CEX}, which shifts the effective cut-off points, or by differences across CEX exchange rates, whose superposition results in smoother curves. As a further extension, the mid-price SDE can be made more sophisticated with a spread process, likely mean-reverting (e.g., spreads widening in volatile periods and narrowing in calm markets).

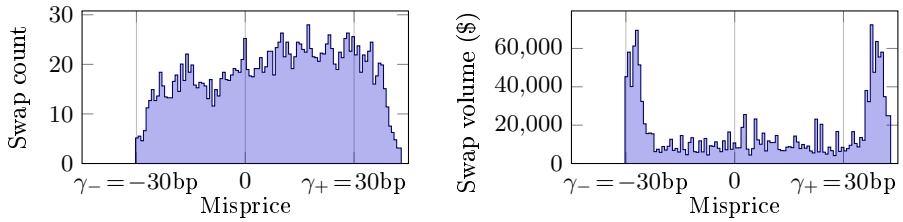
To approximate this effect, we use a linear combination over spreads $\gamma_{\mathrm{CEX}} \in 0, 10, 20, 30$bp. Let $f(x \mid \gamma_{\mathrm{CEX}} = g)$ denote the stationary density at spread g. We choose nonnegative weights w_g (summing to one) to obtain the best fit of the empirical histogram via the linear combination $f_{\mathrm{comb}}(x) = \sum_g w_g\, f(x \mid \gamma_{\mathrm{CEX}} = g)$. The resulting weights w_g are shown in the caption of Fig. 2. This modelling approach is consistent with the assumption that the CEX spread is random and can take multiple values. This interpretation aligns well with our view of arbitrageurs' activity in the market: at any given moment, different participants have access to different CEX venues, each characterised by its own bid–ask spread level.

4.2 Cross verification of the results

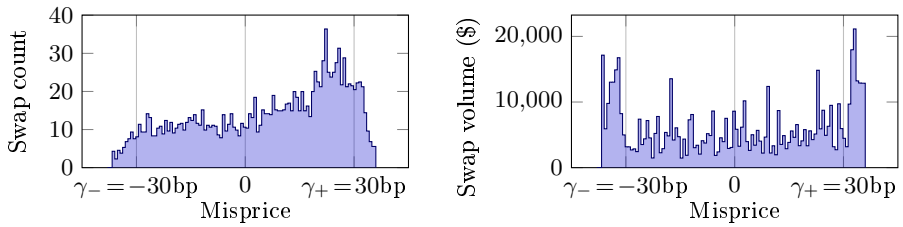
The measurement in the previous section was fairly complex and relied on several assumptions; therefore, we conducted a second, much simpler measurement to verify the validity of our results.

The first question we examined was whether the volume produced by noise traders is indeed smaller compared to the arbitrage volume. Fig. 3 presents the daily aggregate number of swaps (left) and their aggregate volume in USDT (right) as functions of the mispricing in Uniswap V2 pools (with $\gamma = 30$ bp). We observe that trade counts are distributed uniformly across the $[-\gamma, \gamma]$ mispricing interval, but volumes are significantly higher near the boundaries, where arbitrage opportunities arise. The figures suggest that arbitrage accounts for approximately half of the volume and that γ_{CEX} is typically very small. Interestingly, the figure shows that arbitrage swaps can also occur for slightly negative γ_{CEX}. This may be related to the fact that arbitrageurs may anticipate imminent price movements (as the examined time interval exhibits a mild price drift), or that they may receive information about price jumps slightly earlier, for example due to minor latencies in price data obtained via free CEX APIs.

Finally, we note that the arbitrage profit estimation in our model exhibits only limited sensitivity to γ_{CEX}, suggesting that arbitrage volume can serve as a reasonably accurate proxy for arbitrage profits and the associated MEV. The last column of the legend on Fig. 2 shows that the ratio of volume to profit computed from the SDE is not particularly sensitive to γ_{CEX}. Intuitively, higher arbitrage profits necessarily require higher total volume. The resulting upper



(a) Daily swap count and volume histograms for Uniswap V2 ETH-USDT pool



(b) Daily swap count and volume histograms for Uniswap V2 ETH-USDC pool

Fig. 3: Cross-validation measurements on two ETH–stablecoin Uniswap V2 pools confirming the small volume share of noise traders during September 13–29, 2025. For each swap, we computed the misprice before execution (CEX–DEX difference). The chart also shows the histogram, weighted by the trade volume.

and lower bounds lie fairly close to each other, indicating that the MEV paid out from swaps is only slightly lower than the profit generated when the AMM is used exclusively by arbitrageurs in the examined AMM. Based on the analysis of price fluctuations in the ETH–USD currency pair, we conjecture that the arbitrage market is saturated: block builders can choose among many competing arbitrageurs, which forces them to rebate a large fraction of their profits.

5 Conclusion

In this work we developed an extended stochastic framework for modeling the price differences between CEXs and DEXs, treating the mispricing process as a Markov chain with state-dependent volatility, drift, and jumps. We proved that under mild assumptions the process admits a stationary distribution, and we provided constructive formulas—via function iteration—for calculating this invariant density. Building on this result, we derived closed-form expressions for arbitrage volumes and profits in a CPMM setting. Our estimates indicate that arbitrage profits are orders of magnitude larger than those predicted by earlier GBM-based models.

Beyond theoretical analysis, we carried out extensive empirical work: we extracted DEX data from Ethereum, combined it with CEX price feeds, and subjected the resulting mispricing series to statistical tests. An open-source C++

implementation accompanies this research to aid reproducibility and further exploration.

We evaluated the estimates of arbitrage volume derived from our model on the Uniswap V2 USDT–ETH pool and found them to be of comparable magnitude, though slightly larger, than the measured CEX–DEX arbitrage volume. This strengthens our confidence that the model's assumptions are reasonably close to reality and that it can provide meaningful insights into several important practical questions, such as:

- The revenue sustaining PBS arises naturally from the inherent dynamics of financial markets through CEX–DEX arbitrages in AMMs.
- A large amount of AMM volume may be attributable to CEX–DEX arbitrage transactions.
- CEX–DEX arbitrage profits are comparable to the measured MEV, suggesting that the arbitrage market may be saturated.
- The heuristics employed in empirical studies to identify CEX–DEX arbitrage transactions may be overly restrictive.
- Shorter block times are unlikely to reduce MEV, as arbitrage opportunities primarily emerge from rapid price jumps.

Acknowledgements

The project was partly supported by Project no. K23 146347 of National Research, Development and Innovation Fund of Hungary.

References

1. Adams, A., Chan, B.Y., Markovich, S., Wan, X.: The Costs of Swapping on the Uniswap Protocol. Papers 2309.13648, arXiv.org (Sep 2023)
2. Bachelier, L.: Théorie de la Spéculation. Annales Scientifiques de L'Ecole Normale Supérieure **17**, 21–88 (1900)
3. Bichuch, M., Feinstein, Z.: DeFi Arbitrage in Hedged Liquidity Tokens. Papers 2409.11339, arXiv.org (Sep 2024), https://ideas.repec.org/p/arx/papers/2409.11339.html
4. Black, F., Scholes, M.: The Pricing of Options and Corporate Liabilities. Journal of Political Economy **81**(3), 637–654 (1973)
5. Boueri, N.: G3M Impermanent Loss Dynamics. Papers 2108.06593, arXiv.org (Aug 2021), https://ideas.repec.org/p/arx/papers/2108.06593.html
6. Cartea, Á., Drissi, F., Monga, M.: Decentralised Finance and Automated Market Making: Predictable Loss and Optimal Liquidity Provision (November 2022). https://doi.org/10.2139/ssrn.4273989, https://ssrn.com/abstract=4273989, forthcoming in SIAM Journal on Financial Mathematics
7. Cartea, Á., Drissi, F., Monga, M.: Predictable losses of liquidity provision in constant function markets and concentrated liquidity markets. Applied Mathematical Finance **30**(2), 69–93 (2023). https://doi.org/10.1080/1350486X.2023.2277957
8. Cont, R.: Empirical properties of asset returns: Stylized facts and statistical issues. Quantitative Finance **1**, 1–14 (2005)

9. Cont, R.: P. tankov financial modelling with jump processes (2004)
10. Delbaen, F., Schachermayer, W.: The Mathematics of Arbitrage. Springer (2006), https://link.springer.com/book/10.1007/3-540-33589-1
11. Dewey, R., Newbold, C.: The pricing and hedging of constant function market makers. SSRN (2023). https://doi.org/10.2139/ssrn.4455835, https://ssrn.com/abstract=4455835, available at SSRN: https://ssrn.com/abstract=4455835
12. Douc, R., Moulines, E., Priouret, P., Soulier, P.: Markov chains. Springer, Berlin (2018)
13. Eberlein, E., Kallsen, J.: Mathematical finance. Springer (2019)
14. Hanson, F.B., Westman, J.J.: Stochastic analysis of jump-diffusions for financial log-return processes. In: Pasik-Duncan, B. (ed.) Stochastic Theory and Control, pp. 169–183. Springer, Berlin Heidelberg (2002)
15. Hanson, R.: Logarithmic markets coring rules for modular combinatorial information aggregation. The Journal of Prediction Markets **1**(1), 3–15 (2007)
16. Hasbrouck, J., Rivera, T.J., Saleh, F.: An economic model of a decentralized exchange with concentrated liquidity (Aug 2023). https://doi.org/10.2139/ssrn.4529513, https://ssrn.com/abstract=4529513, available at SSRN: https://ssrn.com/abstract=4529513
17. Heimbach, L., Pahari, V., Schertenleib, E.: Non-atomic arbitrage in decentralized finance. In: IEEE Symposium on Security and Privacy (SP). pp. 3866–3884. IEEE (2024)
18. J. Milionis, C. C. Moallemi and T. Roughgarden: A Myersonian Framework for Optimal Liquidity Provision in Automated Market Makers. In: 15th Innovations in Theoretical Computer Science Conference (ITCS 2024). LIPIcs, vol. 287, pp. 81:1–81:19. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl, Germany (2024)
19. J. Milionis, C. C. Moallemi and T. Roughgarden: Complexity-Approximation Trade-Offs in Exchange Mechanisms: AMMs vs. LOBs. In: Financial Cryptography and Data Security. pp. 326–343. Springer Nature Switzerland, Cham (2024)
20. Merton, R.C.: Option pricing when underlying stock returns are discontinuous. Journal of Financial Economics **3**(1-2), 125–144 (1976)
21. Meyn, S., Tweedie, R.L.: Markov chains and stochastic stability. Cambridge University Press, 2nd edn. (2009)
22. Milionis, J., Moallemi, C.C., Roughgarden, T.: Automated market making and arbitrage profits in the presence of fees. arXiv (2023)
23. Milionis, J., Moallemi, C.C., Roughgarden, T., Zhang, A.L.: Automated market making and loss-versus-rebalancing. arXiv (2024)
24. Milionis, J., Wan, X., Adams, A.: FLAIR: A Metric for Liquidity Provider Competitiveness in Automated Market Makers (2023)
25. Nezlobin, A., Tassy, M.: Loss-versus-rebalancing under deterministic and generalized block-times. arXiv preprint arXiv:2505.05113 (2025)
26. Öz, B., Sui, D., Thiery, T., Matthes, F.: Who wins ethereum block building auctions and why? In: Advances in Financial Technologies (AFT). p. 1 (2024)
27. Park, A.: The conceptual flaws of constant product automated market making. SSRN Electronic Journal (01 2021). https://doi.org/10.2139/ssrn.3805750
28. Rao, R., Shah, N.: Triangle fees (2023), https://arxiv.org/abs/2306.17316
29. Uhlenbeck, G.E., Ornstein, L.S.: On the Theory of the Brownian Motion. Phys. Rev. **36**, 823–841 (Sep 1930). https://doi.org/10.1103/PhysRev.36.823
30. Wu, F., Sui, D., Thiery, T., Pai, M.: Measuring cex-dex extracted value and searcher profitability: The darkest of the mev dark forest. In: Advances in Financial Technologies (AFT) (2025), arXiv preprint arXiv:2507.13023

31. Yang, S., Nayak, K., Zhang, F.: Decentralization of ethereum's builder market. In: IEEE Symposium on Security and Privacy (SP). pp. 1512–1530. IEEE (2025)

A Proofs

Proof (Proof of Lemma 1). If $Z_{n+1} = 0$ then the density of $\mu+\sigma\varepsilon_{n+1}$ is $\frac{1}{\sigma}f\left(\frac{y-\mu}{\sigma}\right)$ as easily seen. If $Z_{n+1} = 1$ then we need to calculate the convolution of $u(z)$ with $\frac{1}{\sigma}f\left(\frac{z-\mu}{\sigma}\right)$, which equals

$$\int_{\mathbb{R}} u(y-z)\frac{1}{\sigma}f\left(\frac{z-\mu}{\sigma}\right)$$

for $y \in \mathbb{R}$. The statement now follows trivially.

Proof (Proof of Lemma 2). This follows directly from Lemma 1 and from (3), (4).

Proof (Proof of Corollary 1). Denoting by μ_0 the law of X_0, we have

$$P(X_1 \in A) = \int_A \int_{\mathbb{R}} q(x_0, x)\mu_0(dx_0)dx,$$

so X_1 has density $f_1(x) = \int_{\mathbb{R}} q(x_0, x)\mu_0(dx_0)$. Repeating the same argument, we can calculate $f_k(x)$, $x \in \mathbb{R}$, for all $k \geq 2$.

We now state a general law of large numbers for functionals of Markov processes, in a form that is convenient for us for the proof of Theorem 3. For simplicity, we assume that the state space is $\mathbb{R}$ but the result is true for general state spaces. Let the Markov chain Y_n have transition kernel $R(y, \cdot)$, that is, let

$$P(Y_{n+1} \in A|Y_n = y) = R(y, A)$$

hold for all Borel sets A and for all $y \in \mathbb{R}$. For an arbitrary probability ν, the notation $R\nu$ refers to the probability $A \to \int_{\mathbb{R}} R(y, A)\nu(dy)$. Clearly, $R^n\nu$ (the nth iterate of the above mapping) equals the law of Y_n when the law of Y_0 is ν).

We recall the total variation distance of two probabilities ν_1, ν_2:

$$||\nu_1 - \nu_2||_{TV} := \sup_A |\nu_1(A) - \nu_2(A)|,$$

where A ranges over all the Borel sets in $\mathbb{R}$.

Proposition 1. *Let the Markov chain Y_n have a unique invariant measure ν_* and assume that, for all probabilities ν on $\mathbb{R}$,*

$$||R^n\nu - \nu_*||_{TV} \to 0, \ n \to \infty \tag{9}$$

holds (that is, the law of Y_n tends to ν_* whatever the initial distribution ν of Y_0 is). Then for all measurable $\phi : \mathbb{R} \to \mathbb{R}$ such that $\int_{\mathbb{R}} |\phi(x)|\nu_*(dx) < \infty$, one has almost surely

$$\frac{\phi(Y_1) + \ldots + \phi(Y_n)}{n} \to \int_{\mathbb{R}} \phi(x)\nu_*(dx), \ n \to \infty. \tag{10}$$

Proof. We rely on [12] and all the cited results are from that book. Note first that (9) implies that, for arbitrary probabilities ν_1, ν_2, one has

$$||R^n\nu_1 - R^n\nu_2||_{TV} \to 0 \tag{11}$$

as $n \to \infty$.

By Proposition 11.A.3, (11) implies that bounded harmonic functions are constant, so Corollary 5.2.4 ensures that the tail sigma-field is trivial. As the invariant probability is unique, Theorems 5.2.6 and 5.2.1 imply that the law of large numbers (10) holds in the stated form.

The next technical lemma is also necessary for the proof of Theorem 3.

Lemma 3.

$$\delta := \inf_{(x_0,x_1)\in\mathbb{R}\times[-\gamma_-,\gamma_+]} q(x_0, x_1) > 0.$$

Proof. We have that

$$\begin{aligned} q(x,y) &\geq p(1-q) \inf_{x\in\mathbb{R}, y\in[-\gamma_-,\gamma_+]} \frac{1}{\sigma} f\left(\frac{y-\mu-x}{\sigma}\right) \\ &\geq \frac{p(1-q)}{K} \inf_{x,y\in[-\gamma_-,\gamma_+]} f\left(\frac{y-\mu-x}{\sigma}\right) \geq \frac{p(1-q)}{K} c_1 \end{aligned}$$

Proof (Proof of Theorem 3). We rely on Chapter 16 of [21]. We define $\nu_1(A) := \mathrm{Leb}(A \cap [-\gamma_-, \gamma_+])/(\gamma_+ - \gamma_-)$ and first verify that the whole state space of the process $\tilde{X}$ (that is, $\mathbb{R}$) is a ν_1-small set (see Section 5.2 of [21] for the definition). Indeed, we have $q(x_0, x_1) \geq \delta > 0$ for all $x_0 \in \mathbb{R}$ and $x_1 \in [-\gamma_-, \gamma_+]$, hence

$$\inf_x Q(x, A) \geq \delta\nu_1(A)(\gamma_+ - \gamma_-).$$

Theorem 16.0.2 of [21] now implies that the Markov chain $\tilde{X}$ is *uniformly ergodic*: there is an invariant measure μ_* such that

$$||Q^n(x, A) - \mu_*||_{TV} \leq C'\rho^n \tag{12}$$

for suitable $C' > 0$ and $\rho < 1$. By invariance and Fubini's theorem,

$$\mu_*(A) = \int_{\mathbb{R}} \int_A q(x_0, x_1) dx_0 \mu_*(dx_1) = \int_A \int_{\mathbb{R}} q(x_0, x_1) \mu_*(dx_1) dx_0,$$

showing that μ_* is absolutely continuous w.r.t. the Lebesgue measure and hence admits a density f_*. It follows from (12) that

$$\int_{\mathbb{R}} |f^*(t) - f_n(t)|\, dt \leq 2C'\rho^n. \tag{13}$$

Proof (Proof of Corollary 2). Apply (7) with the choice $\phi(x) := 1_{\{x\notin[-\gamma_-,\gamma_+\}}$.

Proof (Proof of Corollary 3). It is explained and calculated in Lemma 2 and Corollary 2 of [22] (see also Example 2 of [23]) that, if the current price is W and the mispricing is z then the arbitrageur (if present) makes an amount

$$L\sqrt{W}e^{\gamma/2}[e^{(z-\gamma)/2} - 2 + e^{-(z-\gamma)/2}]1_{\{z>\gamma\}}$$

when mispricing is too high ($z > \gamma$) and

$$L\sqrt{W}e^{-\gamma/2}[e^{(z+\gamma)/2} - 2 + e^{-(z+\gamma)/2}]1_{\{z<-\gamma\}}$$

when mispricing is too low ($z < -\gamma$). In a stationary state the distribution of z is μ_*. The presence of an arbitrageur occurs with probability p, independently of what happened before. Consider the second term without $L\sqrt{W}$

$$p\int_{-\infty}^{-\gamma} e^{-\gamma/2}\Big[e^{(t+\gamma)/2} - 2 + e^{-(t+\gamma)/2}\Big]f^*(t)\,dt.$$

Using the identity

$$e^x + e^{-x} = 2\cosh(x),$$

the expression in brackets can be rewritten as

$$e^{(t+\gamma)/2} - 2 + e^{-(t+\gamma)/2} = 2\cosh\big(\tfrac{t+\gamma}{2}\big) - 2.$$

Then we have $\texttt{TVL} = 2L\sqrt{(W)}$, and the claimed formula now follows directly.

Proof (Proof of Corollary 4). By Lemma 2 and Corollary 2 of [22] (see also Example 2 of [23]), when the mid price is W and the log mispricing is t, the arbitrageur (if present) trades exactly the amount needed to bring the pool price back to the closest boundary $e^{\pm\gamma}$ of the no-arbitrage band.

For a CPMM with invariant $xy = L^2$, this rebalancing induces the following *gross notional volume* (denominated in the quote asset):

$$L\sqrt{W}\,p\left(e^{\gamma/2} - e^{-t/2+\gamma}\right)1_{\{t>\gamma\}} \quad \text{and} \quad L\sqrt{W}\,p\left(e^{-\gamma/2} - e^{-t/2-\gamma}\right)1_{\{t<-\gamma\}},$$

respectively for the overpricing ($t > \gamma$) and underpricing ($t < -\gamma$) cases. These follow from the CPMM algebra in [22], specialised to trade size (gross volume) rather than profit.

In the stationary regime, the distribution of t is μ_* with density f^*, and the arbitrageur is present independently with probability p. Taking expectations with respect to f^* yields the claimed expression.

Remark 5. For the general class of constant function market makers (CFMMs), the trading function $f(x, y) = x^\theta y^{1-\theta}$ for some $\theta \in (0, 1)$ and positions in the risky/riskless asset must satisfy $f(x, y) = L$ at all time. As it is calculated in the second example of [23], the holdings maximising pool value, at price W, are equal to

$$x^*(W) = L\left(\frac{\theta}{(1-\theta)W}\right)^{1-\theta}, \quad y^*(W) = L\left(\frac{(1-\theta)W}{\theta}\right)^{\theta}. \tag{14}$$

Let us now introduce the constants

$$c_1 = \left(\frac{\theta}{1-\theta}\right)^{1-\theta}, \quad c_2 = \left(\frac{1-\theta}{\theta}\right)^{\theta}.$$

Calculations analogous to those of Corollary 3 show that, in this case the expected profit of the arbitrageur at a given trading date equals

$$LW^{\theta}p \int_{\gamma}^{\infty} [c_1(e^{t(1-\theta)} - e^{\gamma(1-\theta)}) + c_2(e^{\gamma - t\theta} - e^{\gamma(1-\theta)})] f^*(t)\, dt \tag{15}$$

$$+ LW^{\theta}p \int_{-\infty}^{-\gamma} [c_1(e^{t(1-\theta)} - e^{\gamma(\theta-1)}) + c_2(e^{-t\theta-\gamma} - e^{\gamma(\theta-1)})] f^*(t)\, dt \ . \tag{16}$$

B Limitations

Notwithstanding the fact that an attempt has been made to derive mathematical methods for a larger class of stochastic models to arrive at a more complete picture, our approach has some limitations. In this work, only L1 blockchains are considered and we believe that our discussion can be extended to L2 chains, whereby transactions are executed in rollups. Such an approach would entail much lower transaction fees. Another limitation of our work is that we only investigate V2 AMMs and no analysis has been made on concentrated liquidity provision (Uniswap V3). Finally, let us draw attention to the fact that the stochastic effects of noise traders are also neglected in this work. This topic shall constitute new research directions.

C Related Works

The similarities between market scoring rules and automated market making were pointed out by R. Hanson [15] as early as 2002. The performance of a rebalancing strategy for various AMMs has been investigated in [23]. The concept of LVR, along with empirical studies, is also present in this work. The authors of [22] set out to find a closed-form solution of the AMM problem when fees are present. A comprehensive framework is provided in [19] for describing and analysing exchange mechanisms, enabling straightforward comparisons between limit order books and AMMs in terms of complexity and expressiveness. The work elucidates the notion of exchange complexity, clarifying the extent to which certain exchange methods are simpler compared to others. In [18], the authors analyse the profit-maximising strategy of a monopolist liquidity provider AMMs using a Bayesian-inspired belief framework. The authors in [1] evaluate the efficiency of DEX trading using slippage as a key metric; their study explores the components of transaction costs, highlighting the variation based on trade characteristics. [28] investigates new fee structures to make AMMs more robust. In another line of work, FLAIR [24] has been introduced as a novel metric for liquidity provider competitiveness, complementing LVR. [11] bears resemblance

to our work, notably through its empirical investigations and model extensions, however, unlike our work the authors present no rigorous mathematical discussion on the existence of the SPD. [3] focuses on developing a risk-neutral pricing and hedging framework for constant product market makers (CPMM) liquidity tokens aiming to integrate this DeFi product into the principles of modern financial theory. A fairly recent work [16] puts forth an economic model of a DEX with concentrated liquidity provision, examining a continuous-time model for a single DEX facilitating the trading of a risky asset. The authors demonstrate that the portfolio characteristics of concentrated liquidity provision can be directly linked to those of a covered call trading strategy. The findings of [27] suggest that when there are sufficiently many liquidity providers who are not overly risk-averse, linear pricing becomes much more cost-effective for investors.

D Convergence of the stationary distribution

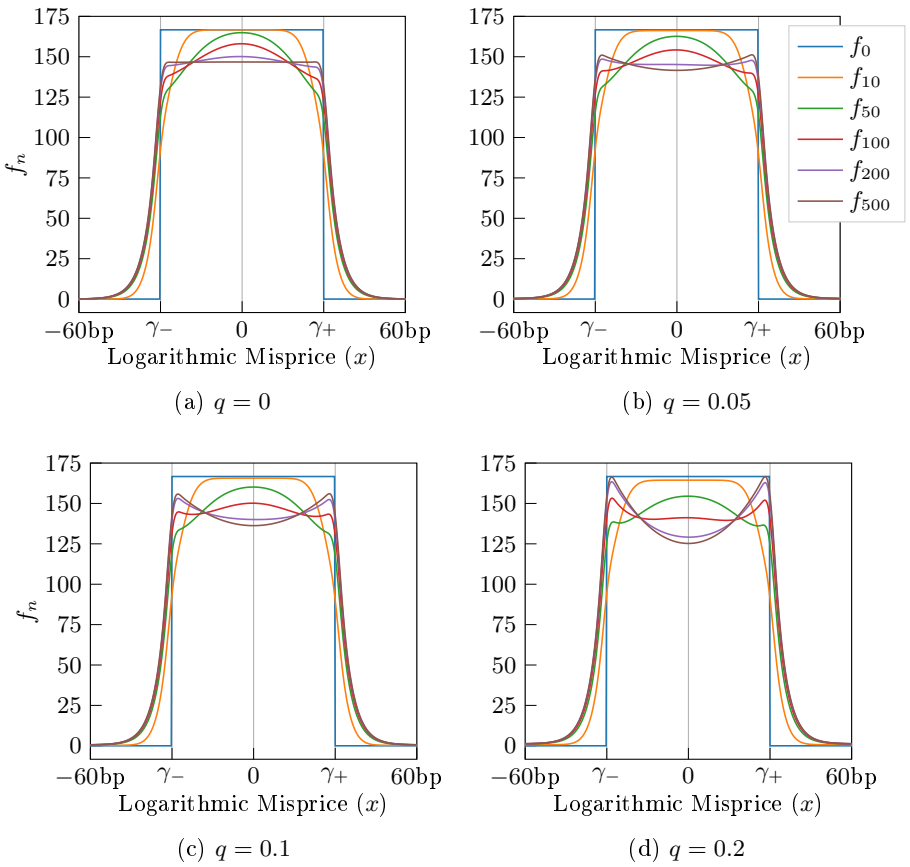
Now we turn to the investigation of the convergence properties of the SPD. The zeroth order density function $f_0(x, \gamma_+, \gamma_-)$ is defined as a window function:

$$f_0(x, \gamma_+, \gamma_-) = \begin{cases} 1 & \text{if } -\gamma_- \leq x \leq \gamma_+, \\ 0 & \text{otherwise.} \end{cases}$$

Note that other initial guesses are also possible (e.g. Dirac delta in 0), however for faster convergence it is expedient to use an initial distribution that is as close to the final one as possible. The higher order densities are calculated using Eq. 5 with a numerical integration formula. We normalise the density at each step and let the integrator run its course using the results of the previous iteration. The numerical integration based on Eq. 5 was performed on a discrete grid using the Gauss–Kronrod quadrature formula with 15 quadrature points. The convergence of the stationary distribution of the misprice process was investigated using a numerical integrator. Our results are presented in Fig. 4. For this calculation, we used the same parameter set as in [22], e.g. $\sigma = 5\%$, $\mu = \sigma^2/2$, $\Delta t = 12$s, $-\gamma = +\gamma = 0.003$ and the integration was performed on 801 grid points. In Fig. 4 one can observe the effect of jumps (i.e. when $q > 0$). With jumps, the SPD has a well in the middle, implying that it is rare for the misprice to be around 0 compared to the case with no jumps. This is in accord with our intuition, as jumps in the external market price lead to increased arbitrage activity and the arbitrageurs tend to push the AMM back to the boundary more often.

E Reproducing reference values

Our objective was first to reproduce the numerical values in Table 1. of [22] with our numerical solver that is capable of calculating the SPD up to arbitrary order and accuracy. The iteration order and the grid size are external parameters to the integrator module.



(a) $q = 0$ (b) $q = 0.05$

(c) $q = 0.1$ (d) $q = 0.2$

Fig. 4: Convergence of the SPD of the misprice process with ($q > 0$) and without jumps ($q = 0$). Observe that the stationary state the system converges onto with different q values. The number of grid points was set at 801. The density was iterated 1000 times and for clarity, only some of the functions are presented in the figures.

The size of the grid has been varied and the absolute error for different block times and fee levels with respect to the reference values taken from Table 1. of [22] are presented in Fig. 5 and the calculated values are presented in Table 3. The calculation of the fast block region presented serious difficulties for the numerical integrator. Therefore the data points of 50 msec have not been reproduced.

The model parameters are set according to that of [22] in the symmetric case. The daily volatility was set at 5%, the diffusion drift (μ) was set at $\sigma^2/2$ and the fee levels and block times were varied to reproduce the results of [22]. The size of the trade region was calculated on the discrete grid as the ratio of $\int_{-\infty}^{-\gamma} f^*(t)dt +$

Table 3: Comparison of reference trade region coverage (reference values are taken from [22]) with calculated results (in parentheses) for various block times (Δt) and fee levels (γ, expressed in basis points). Integrated results are rounded to one decimal place.

$\Delta t \setminus \gamma$	1 bp	5 bp	10 bp	30 bp	100 bp
10 min	96.7% (96.8%)	85.5% (85.5%)	74.7% (74.6%)	49.6% (49.4%)	22.8% (20.8%)
2 min	92.9% (93.1%)	72.5% (72.6%)	56.9% (56.9%)	30.5% (30.5%)	11.6% (11.6%)
12 sec	80.7% (81.9%)	45.6% (45.9%)	29.5% (29.6%)	12.3% (12.2%)	4.0% (4.0%)
2 sec	63.0% (67.2%)	25.4% (26.2%)	14.5% (14.8%)	5.4% (5.4%)	1.7% (1.7%)

$\int_{+\gamma}^{+\infty} f^*(t)dt$ compared to $\int_{-\infty}^{+\infty} f^*(t)dt$. Note that the probability of stochastic jumps in this case was set to zero, i.e. $q = 0$. At the end of each iteration, we normalise the densities, such that their integral from $-\infty$ to ∞ equals 1. When using our initial approach for the numerical integrator, we experienced numerical instabilities at the calculation of $h(x, y)$. We traced it to the term in which $\frac{1}{\sigma}$ is multiplied by $f\left(\frac{y-\mu(x)}{\sigma(x)}\right)$. When the volatility is small this operation leads to numerical instability, and we solved this issue by calculating the logarithm of these terms. The interested reader is referred to the open-source repository for the implementation details. Notice that in some cases there is some residual

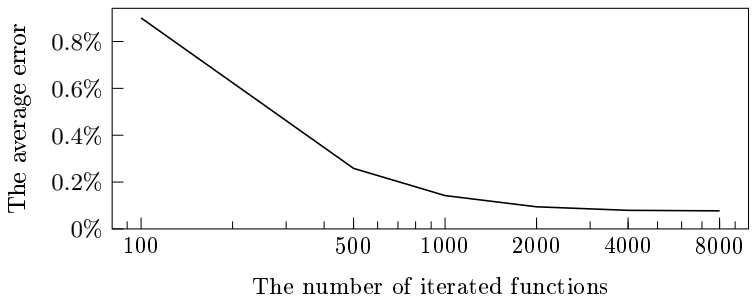


Fig. 5: The average absolute errors of our numerical integrator with respect to the values of Table 1. of [22] with the number of function iterations. The number of grid points was set at 801. The average of errors with respect to the values in Table 1. of [22] is shown.

error, but it can be decreased arbitrarily by increasing the number of grid points and the number of iterated functions. We observe the most discrepancy at $\Delta t = 2$ sec, $\gamma = 1$bp. For this data point the result converges very slowly to its reference value with an increased number of iterated functions and grid points. The reason behind this slow convergence is yet to be investigated. Regarding the runtime of the numerical integrator, it calculates the data points that were necessary to

generate Fig. 5 in a few minutes fully utilising an AMD Ryzen 9 3900X 12-Core CPU.

E.1 Arbitrage Profit

Having calculated the SPD of the misprice process, we can calculate the arbitrage profits for a given trading interval. To this end, we rely on the above-introduced equation without the price (W^θ) and the liquidity (L) terms, i.e.:

$$\text{ARB} = \quad p \int_{\gamma}^{\infty} [c_1(e^{t(1-\theta)} - e^{\gamma(1-\theta)}) + c_2(e^{\gamma - t\theta} - e^{\gamma(1-\theta)})] f^*(t)\, dt \tag{17}$$

$$+ \, p \int_{-\infty}^{-\gamma} [c_1(e^{t(1-\theta)} - e^{\gamma(\theta-1)}) + c_2(e^{-t\theta-\gamma} - e^{\gamma(\theta-1)})] f^*(t)\, dt, \tag{18}$$

where p is the probability of an arbitrage trade, θ is a parameter of the AMM, γ is the symmetric fee rate, c_1 and c_2 are as introduced above and $f^*(t)$ is the SPD of the logarithmic misprice process z_t.

We calculate arbitrage profits in the symmetric setting, whereby $\mu = \sigma^2/2$ and present our results in Table 4. Notice that arbitrage profits increase significantly with increasing jump probability. This is in line with the common belief that arbitrage activity is generated by sudden price jumps. Observe for example the row where the block time is 12 seconds (Ethereum) and an AMM with 30 bp fee level (typical for Uniswap V2) and notice that jumps lead to orders of magnitude larger arbitrage profits.

E.2 Further Results

Using the blockchain data extracted from our execution client, we plot Figure 6 to demonstrate the daily volatility and the number of executed trades over an interval between November and December 2025. In the right subplot of this figure, a histogram of the exchanged amount of ETH is presented for the same interval. Notice that the number of trades per day can sometimes vary entirely independently of the daily volatility. This further supports the claim that volatility itself cannot be the only source of arbitrage trades on an AMM.

F Statistical Tests

To underpin the applicability of the above-introduced mathematical framework, we investigated cryptocurrency exchange rates and applied statistical tests. Market data of cryptocurrencies `fil`, `btc`, `ETH`, `dxdy`, `xch`, `1inch`, `aave`, `ltc`, `beam`, `cro`, `sol` and `bnb` has been collected. Some of these cryptocurrencies are also tradable on platforms like Uniswap. In Table 6. results of a Kolmogorov-Smirnov test, along with kurtosis and skewness values are presented for these cryptocurrencies. Numerical data has been obtained from an API provided by GateIO[7].

[7] https://www.gate.io/gate-api

Table 4: The proportion of $L\sqrt{W}$ accounting for arbitrage profits for different model parameters (block times and fee levels) without jumps and with jumps, with 4 different q values. In each cell, the upper number corresponds to $q = 0$ and the lower number corresponds to $q = 0.2$. ARB was calculated according to Eq. 18. The number of grid points used was 801, the number of iterated distributions was set to 1000. $\theta = 0.5$, $\sigma = 5\%$ (daily), $\mu = \sigma^2/2$.

Δt	q	1 bp	5 bp	10 bp	30 bp	100 bp
10 min	0	0.00202760	0.00175490	0.00149120	0.00087340	0.00023100
	0.05	0.00583060	0.00536270	0.00485070	0.00328570	0.00062900
	0.1	0.00855130	0.00794330	0.00725320	0.00500800	0.00091520
	0.2	0.01199680	0.01121090	0.01029490	0.00718960	0.00129170
2 min	0	0.00009700	0.00007500	0.00005900	0.00003200	0.00001100
	0.05	0.00038730	0.00034990	0.00031370	0.00021390	0.00004600
	0.1	0.00066180	0.00060930	0.00055420	0.00038560	0.00007800
	0.2	0.00116750	0.00108700	0.00099710	0.00070100	0.00013500
12 sec	0	0.00000084	0.00000047	0.00000030	0.00000013	0.00000004
	0.05	0.00000460	0.00000400	0.00000360	0.00000250	0.00000052
	0.1	0.00000860	0.00000780	0.00000710	0.00000500	0.00000100
	0.2	0.00001800	0.00001700	0.00001500	0.00001100	0.00000220
2 sec	0	0.00000002	0.00000001	0.00000000	0.00000000	0.00000000
	0.05	0.00000013	0.00000011	0.00000010	0.00000007	0.00000001
	0.1	0.00000025	0.00000022	0.00000020	0.00000015	0.00000003
	0.2	0.00000053	0.00000049	0.00000045	0.00000032	0.00000007

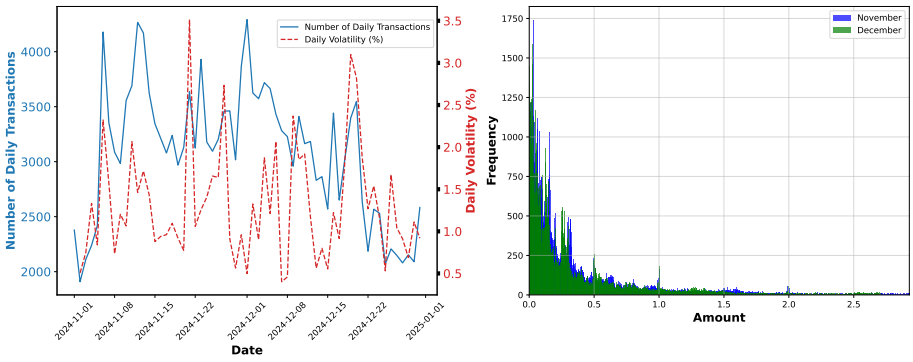


Fig. 6: The daily volatility in % vs. the number of DEX transactions per day for each day throughout November and December 2024 (left) and a histogram of the amount of exchanged ETH during the investigated interval (right).

The high kurtosis values in the tables indicate that the corresponding distributions have fatter tails, meaning that they contain more extreme values or outliers compared to a normal distribution as conjectured in the introduction. The positive and the negative skewness values imply that the distributions are

Table 5: Estimated parameters of the fitted model for different jump thresholds τ. Columns are the log-likelihood (LL), daily volatility parameter σ, daily mean parameter μ, the per-step (12 sec) jump probability (q), and the daily jump-size distribution mean μ_U and daily variance σ_U^2.

τ	LL	σ	μ	μ_J	σ_J	q
–	6.6192	0.0234	0.0234	–	–	0
1.5	7.6895	0.0234	0.0517	-0.0699	0.1649	0.0582
1.6	7.6901	0.0234	0.0550	-0.0911	0.1710	0.0541
1.7	7.6903	0.0234	0.0587	-0.1165	0.1767	0.0506
1.8	7.6901	0.0234	0.0597	-0.1321	0.1820	0.0476
1.9	7.6896	0.0234	0.0598	-0.1466	0.1883	0.0443
2.0	7.6886	0.0234	0.0677	-0.2035	0.1952	0.0411
2.1	7.6875	0.0234	0.0724	-0.2498	0.2018	0.0384
2.2	7.6862	0.0234	0.0798	-0.3135	0.2075	0.0362
2.3	7.6847	0.0234	0.0791	-0.3350	0.2138	0.0340
2.4	7.6828	0.0234	0.0778	-0.3571	0.2211	0.0317
2.5	7.6809	0.0234	0.0799	-0.4027	0.2281	0.0297
3.0	7.6709	0.0234	0.0587	-0.3345	0.2581	0.0228
3.5	7.6603	0.0234	0.0470	-0.2828	0.2868	0.0182
4.0	7.6498	0.0234	0.0349	-0.1623	0.3155	0.0148

not centered, but rather stretched to the right/left. Finally, the small p-values for the Kolmogorov-Smirnov test for Gaussian distribution also implies that our conjectures were right.

This analysis on empirical data shows such heterogeneity in the invariant densities that it cannot be reproduced by a simple two-parameter model family of price processes such as the Black-Scholes model. We emphasise our findings again. In order to have a better match of the data with model predictions, a much larger class of models should be considered. Section 3 of the present work guarantees that an invariant density exists and can be calculated in a remarkably wide class of models.

Table 6: Analysis results of hourly log-returns using the exchange rate of various cryptocurrencies from December 11, 2024, to January 15, 2025: kurtosis, skewness and Kolmogorov-Smirnov p-values.

Cryptocurrency	Kurtosis	Skewness	KSP
1inch	7.5280	-0.3181	0.0027
aave	5.7957	0.5147	0.0011
arb	6.3933	-0.1058	0.0021
beam	14.4068	-0.5742	0.0000
bnb	9.6333	-0.5604	0.0001
btc	6.6794	-0.4529	0.0000
cro	13.7036	0.9348	0.0000
dxdy	6.8077	0.2083	0.0008
ETH	8.7076	-0.5141	0.0000
fil	7.5472	-0.2352	0.0002
ltc	14.7859	0.7149	0.0001
sol	8.6612	-0.1319	0.0001
xch	7.8040	-0.1468	0.0004