Blockchain Research Lab

Working Paper Series, No. 2
www.blockchainresearchlab.org

Discovering market prices: Which price formation model best predicts the next trade?

17 June 2019
This paper has not been peer reviewed. Please do not copy or cite without author's permission.

André Meyer [a, *], Ingo Fiedler [a, b]

Abstract

For most purposes of technical analysis, valuation metrics and many other relevant financial methods, the price of the last transaction is considered representative of the market price. The straightforward argument is that at this price, supply and demand have last met. However, on closer examination, the question arises as to why a past event should be relevant to the future, and why other, potentially more recent information should not be used to discover a future price. Building on this question, we apply a range of new price formation models to current data available on crypto currency exchanges that depict level II market data, and compare their short-term forecast accuracy against the common-used ticker price and mid-price. Data on crypto currencies is used as the closest example to free markets, since crypto currency trading is continuous, markets never close, and interferences through oversight is extremely rare. We find that two of the five price formation models investigated outperform the widely used ticker as a price indicator for the next trade. We conclude that the volume-limited clearing price best predicts the price of subsequent trades. Its usage can thus enhance the explanatory power of various financial analyses.

Keywords: Price Discovery; High-Frequency Trading; Short-Term Price Prediction, Limit Order Book; Mid-Price; Micro-Price, Clearing Price

1. Introduction

The concept of price has existed ever since man began to trade. From very early on, economists such as Smith (1776), Ricardo (1817) and Stackelberg (1934) have examined the significance of prices and their origin. While most economists agree that a price marks the equilibrium between supply and demand, there are different definitions of prices and different forms of markets that influence the discovery of prices.

The interplay of supply and demand is perhaps best illustrated by exchanges with the characteristics of an order-driven market and a situation of perfect competition, where the order books reflect the supply and demand curves. For this reason the Limit Order Book (LOB) is an important field of research and has been studied in various ways (Glosten 1994; Biais et al. 1995; Foucault et al. 2005; Roşu 2009; Gould et al. 2013). If sales and purchase orders meet or if a market order is executed, a trade is completed. The price at which such a trade takes place is displayed as the so-called "ticker" and represents the current value of a share or a good. This suggests that the ticker is a relevant reference point for the present and the near future. The more recent the trade and thus the ticker, the more convincing this assumption.

A widely used alternative to the last price is the mid-price (or midpoint price), the midpoint between the best ask and the best bid offer (Laruelle et al. 2013; O'Hara 2015; Cont and Kukanov 2017; Ntakaris et al. 2018). However, this indicator is associated with a number of weaknesses, as it fails to consider not only the volume of the last trade but also other important factors that can have a significant influence on market prices, such as the price depth of the bid and ask side (Kempf and Korn 1999; Ahn et al. 2001) or the tick size (Darley et al. 2000). Therefore, other measures have already been developed and discussed in the literature to discover more representative prices.

Besides asset pricing models (Sharpe 1963; Merton 1973; Bollerslev et al. 1988) and models that use fundamental and technical methods (Graham and Dodd 1940; Damodaran 2012; Taylor and Allen 1992; Damodaran 2012) to calculate future prices, many models have been created for the increasingly important high-frequency trading (Cont 2011; Agarwal 2012; O'Hara 2015; Avellaneda and Stoikov 2008). High-

[a] Blockchain Research Lab gGmbH, Hamburg, Germany
[b] Universität Hamburg, Faculty of Business, Economics & Social Sciences, Hamburg, Germany
* Correspondence: meyer@blockchainresearchlab.org

frequency trading poses a challenge especially for market makers. They not only have to fulfil their primary task of providing liquidity but must also defend themselves against possibly better-informed traders (Menkveld 2013). Taking into account the often-discussed diffusion of the price (Tóth et al. 2011; Mastromatteo et al. 2014), price criteria such as order book imbalance (OBI) or order flow imbalance (OFI) have been established in the literature in order to provide market makers with an important indication of price discovery (Eisler et al. 2012; Cont et al. 2014).

On the other hand, the literature is investigating strategies to liquidate large positions without a significant price impact (Easley and O'hara 1987; Lin et al. 1995). To measure the success of such strategies, the volume-weighted average price (VWAP) was introduced as a benchmark (Konishi 2002; Madhavan 2002; Guéant and Royer 2014; Frei and Westray 2015). This measure already takes into account the volume, albeit ex post.

The weighted mid-price in turn integrates volume in the form of order book imbalance into the discovery of a future representative price. However, this method is of limited use for high frequency trading as it only takes into account the best bid and ask prices and their respective volumes, which are susceptible to constant cancellations in fractions of a second, as is common in HFT (Gatheral and Oomen 2010; Robert and Rosenbaum 2012). Therefore, a number of new models based on the weighted mid-price were developed. Bonart and Lillo (2016) adapted the Madhavan et al. (1997) price formation model, taking into account quote discretization and liquidity rebates to introduce price definitions for large tick stocks. The approach by Jaisson (2015) incorporates conditional expectations and was adopted by Lehalle and Mounjid (2017) to form the so-called micro-price as the expected future mid-price under the condition of the current mid-price and the degree of order book imbalance. Finally, Stoikov (2017) tests mid-prices, weighted mid-prices and micro-prices for their informative value and short-term forecast accuracy. He finds that the micro-price yields the most accurate results.

Building on the methods described above, we define new price formation models and test their accuracy in short-term price prediction in relation to the commonly used ticker price. The null hypothesis of this paper is therefore that none of our price formation models is superior to the ticker price as an indicator of the next trade.

We do not only use current data provided in every level II order book, we also test a model that uses the trade history as an indicator for pricing. The use of the trade history is inspired by the ticker price, which is itself a past-related value.

We test the forecast accuracy of our models using crypto currency market data. Ghysels and Nguyen (2018) already use data from crypto currency exchanges in their work for new insights into price discovery. Our choice of crypto market data is primarily motivated by the non-stop 24/7 trading and the absence of distorting stabilizing price mechanisms or regulatory intervention. For example, the auctions that are used on traditional stock exchanges after trade disruption, for pre- or post-trade, but also in case of excessive price volatility, have a significant influence on price discovery. Madhavan and Panchapagesan (2000) find that while the opening auction of the NYSE increases price efficiency, it leads to stagnation of prices and thus reduces their flexibility. By contrast, the market for crypto currencies is cleaner in this regard and can thus be seen as prototype closer model of a free market. Furthermore, its data richness makes it the perfect environment to analyse price discovery.

Yet crypto markets also have disadvantages, specifically their lack of regulation entails a risk of manipulation. Many crypto exchanges are suspected of flaunting wrong volumes, making it difficult for researchers to obtain unspoiled results. We therefore use data from exchanges that according to the Blockchain Transparency Institute (2018) report unspoiled volumes (Fusaro and Hougan 2019).

2. Price formation models

In the following, we will test five different price formation models, two of which are tested with three different parameter weights each, yielding a total of nine separate calculations. The results will be compared to the ticker price.

The first price formation model we study is the mid-price. As already mentioned, the mid-price is often used both in research and in practice for short-term price predictions. In this simple calculation method, the reference price is obtained as the mid-point between the highest buy and the lowest sell offer:

$$mp = \frac{1}{2}(p_0^a + p_0^b) \tag{1}$$

where mp is the mid-price, p_0^b is the highest bid and p_0^a is the lowest ask offer in the LOB. These prices are always in the first position of the respective order book side. Their distance to the best order of each order book side is therefore zero, as indicated by the index 0. Though not relevant for the calculation of the mid-price, the index is all the more important for the subsequent model, the weighted mid-price (wmp):

$$wmp = ibp_0^a + (1 - ib)p_0^b \tag{2}$$

wmp is calculated using the imbalance (ib), which depends on the volume of the best bid (q_0^b) and ask (q_0^a) offer:

$$ib = \frac{q_0^b}{q_0^a + q_0^b} \tag{3}$$

Both the mid-price and the weighted mid-price are susceptible to frequent order changes or low volume orders, which merely serve to price discovery, especially in less liquid markets.

The clearing price is often defined as the price at which the market settles a commodity or security, i.e.

where the quantity delivered equals the quantity demanded. In practice, the ticker price is often assumed to be the clearing price because it is the price at which the last units of an asset were traded. However, as the above-mentioned studies on the liquidation of large positions show (e.g. Cartea and Jaimungal 2016), this is not always consistent with the required volume. The following price formation model calculates the mid-price for a given volume based on the original definition of the clearing price. We call it the volume-limited clearing price ($vlcp$) and define it as follows:

$$vlcp = \frac{1}{2}\left(\frac{sf\left(p_i^a, q_i^a, vl\right) + sf\left(p_i^b, q_i^b, vl\right)}{vl}\right) \tag{4}$$

where the total price of a market buy order for a fixed volume (volume limit) vl is given by the function sf, which is defined as:

$$sf = \begin{cases} \sum_{i=0}^{n} p_i^a q_i^a & , for \ \sum_{i=0}^{n} q_i^a \leq vl \\ 0 & , for \ \sum_{i=0}^{n} q_i^a > vl \end{cases} \tag{5}$$

The total price of a market sell order for a fixed volume vl is denoted as sf and defined analogously using the bid-side offers. The volume-weighted average bid and ask prices are then formed by dividing the total prices of each side by the given volume. The average of these values then yields the volume-limited clearing price as the middle between the volume-weighted average prices of each side of the LOB.

In calculating the market price, this model includes the volume up to a fixed amount, thus excluding non-representative orders located lower in the LOB that are often placed on crypto markets by speculators in the hope of a so-called fat-finger error, where an order is placed of a far greater size or price than intended, or in the wrong currency. The model adapts to market conditions. If the specified volume (vl) does not exceed the volume of the best bid and ask offer, the results of the model are equal to the mid-price. Conversely, if vl does exceed the volume of the best bid or ask orders, the order book imbalance is also indirectly included in the calculation, not directly by the volume itself, but by the volume-weighting of prices. In markets with low liquidity near the spread or in markets with frequent placement and cancellation of low volume orders at the top of the order book, the model produces more stable and consistent results, while it delivers the same results as the mid-price in liquid markets with high volumes close to the spread.

Our next model is related to the $vlcp$ but uses a price limit to calculate the price. We therefore call it the price-limited clearing price ($plcp$). The prices are weighted by volume and the distance to a reference price. The further away an order price is from the reference price, the lower its weight. The $plcp$ is defined as:

$$plcp = 2rp - \frac{\sum_{i=0}^{p_i^a \leq pl} p_i^a dw_i^a + \sum_{i=0}^{p_i^b \geq pl} p_i^b dw_i^b}{\sum_{i=0}^{p_i^a \leq pl} dw_i^a + \sum_{i=0}^{p_i^b \geq pl} dw_i^b} \tag{6}$$

where rp is the reference price, pl is the price limit, and dw_i^a and dw_i^b are the distance weights of the ask and the buy side, respectively. We define the price limit (pl) as:

$$pl = rp \times \left(1 \pm \frac{pd}{100}\right) \tag{7}$$

Starting from the reference price (rp), depending on the order book side, the price limit (pl) is augmented (ask side) or reduced (buy side) by a distance that is calculated using percentage depth (pd). pd is an external parameter that we assumed to be either 1, 2 or 3 in our sample. The parameter indicates up to which price level of the respective order book page orders are included in the calculation. The larger the value the greater the price range within the orders will be included in the calculation. While this can mean to incorporate more information in form of more orders, it also increases the likelihood of including a bias from noise, for example, in the form of an order book that is asymmetric by chance. Such risk is especially likely for illiquid order books. We use the $vlcp$ as the reference price, but the mid-price or any other reference value serve just as well. To calculate $plcp$, we must furthermore define the distance weights (dw):

$$dw = \frac{q_i}{\left(10{,}000 \times df(p_i, rp)\right)^{de}} \tag{8}$$

The distance weights are calculated by discounting the volume of an order (q_i) using the distance function ($df(p_i, rp)$) and the distance exponent (de). We set the external parameter de to 0.75 in our sample. The distance function is defined as follows:

$$df = \begin{cases} 0 & , if \ p_i = rp \\ \frac{p_i}{rp} - 1 & , if \ p_i > rp \\ \frac{rp}{p_i} - 1 & , if \ p_i < rp \end{cases} \tag{9}$$

If the price of an order is equal to the reference price, the function assumes the value 0, which would result in an error in the calculation of the weights. Therefore, the reference price must be chosen so as to make this impossible. Therefore, we use the $vlcp$ as the reference price (rp) in our sample. $vlcp$ cannot be reached by any buy or ask price, since its value is within the spread. Hence, it is impossible that the condition $p_i = rp$ occurs. It should be noted that even if the tick size does not allow a lower price scaling, this does not affect the hypothetical value of the $vlcp$, since it is infinitely scalable. For ask side prices, the function assumes the value $\frac{p_i}{rp} - 1$ since here the order prices exceed the reference price. For the bid side, $\frac{rp}{p_i} - 1$ applies, as the order prices are below the reference price.

An advantage of this model is that it takes into account that orders located towards the bottom of the order book are less relevant than those at the top. In addition, it allows users to specify to which price depth the LOB is taken into the calculation. The model also takes into account the OBI since the volumes are considered but discounted according to their distance to the reference price. By incorporating these discounted volumes, the model also considers any imbalance of the order book, which may provide a valuable signal of excess demand or supply. Due to the weighting function of the model, imbalances closer to the reference price are weighted more heavily than more distant ones. However, there is a risk that large orders will be placed within reach of the reference price in order to manipulate $plcp$ and affect trading strategies based on it.

The next model uses the $vlcp$ as the reference price as well and is furthermore using the price limit (pl) to adjust the reference price by the factor $\left(1 + cf(af, ma)\right)$. This model, which we call the adjusted reference price, is defined as follows:

$$arp = rp \times \left(1 + cf(af, ma)\right) \tag{10}$$

cf is the cap function that limits the results of the adjustment function (af) to a maximum adjustment (ma):

$$cf = \begin{cases} 0 & , if\ af = 0 \\ \min(af, ma) & , if\ af > 0 \\ \max(af, -ma) & , if\ af < 0 \end{cases} \tag{11}$$

The adjustment function (af) is determined by the adjusted weights (aw) and defined as follows:

$$af = \frac{1}{1000} \times \left(\frac{\sum_{i=0}^{p_i^b \geq pl} aw_i^b}{\sum_{i=0}^{p_i^a \leq p} aw_i^a} - 1 \right) \tag{12}$$

Within af the adjusted weights (aw) are taken into account up to the price limit (pl) given by equation (7) and are defined as:

$$aw = q_i \times b^{100 \times df(p_i, rp)} \tag{13}$$

Where q_i is the volume of order i and b is an external parameter whose exponent is 100 times the distance function (9). This model uses the different order volumes up to the predefined price limit of each order book side to create an imbalance correction factor by which the reference price is adjusted. The adjustment is limited by the external parameter ma. For our sample, we assumed 0.003 as the value for ma. While the price limit for the $plcp$ is determined by the prices of the respective orders, in the arp, the price limit is determined by the volume of the respective orders.

The last price formation model we shall test is completely different from the previous ones. It does not use current information but calculates a price based on historical transactions. The model thus builds on the ticker itself. While the ticker only uses the price of the last trade and thus has the disadvantage of being quite volatile, our model uses the prices and volumes all past transactions up to a certain age, though with decreasing weights. We call this price formation model the trade history model (th) and define it as follows:

$$th = \frac{\sum_{i=0}^{ag \leq al} tw_i p_i}{\sum_{i=0}^{ag \leq am} tw_i} \tag{14}$$

where the time weights (tw) are defined as:

$$tw = \frac{q_i^{qe}}{(ag_i + 1)^{ae}} \tag{15}$$

The volume of trade i (q_i) with the quantity exponent (qe) is divided by the age of trade i (ag_i) plus 1 raised to the age exponent (ae).The exponents allow us to adjust the impact that age and volume shall have. We selected three quarters for both exponents and 180 seconds for the age limit (al). The advantage of this model is that smaller fluctuations or price spikes that only correspond to small volumes have only a small influence towards the calculation of the representative price. The age limit allows us to determine how long a period should be considered relevant for the representative price. While the weight of older trades would eventually be discounted to zero, the age limit allows to cut off trades with a very low weight that do not add much information but, when included, would cost computing power and thus delay the result.

The reduced weights of older trades are in line with the common sense notion that events in the more distant past should have less bearing on the future. Table 1 lists the values we selected for the external parameters:

Table 1: Parameters used in our price formation models

Parameter	**Value**
volume limit (vl)	0.5
percentage depth (pd)	1 / 2 / 3
distance exponent (de)	0.75
base parameter (b)	0.75
Maximum adjustment (ma)	0.003
quantity exponent (qe)	0.75
age exponent (ae)	0.75
age limit (al)	180 s

3. Methodology

3.1 Dataset

To assess the performance of our price formation models, we recorded their results in 44,640 minute-data points from 01.12.2018 00:00 to 31.12.2018 23:59 for the prices of 3 crypto currencies – Bitcoin (BTC), Litecoin (LTC), and Ethereum (ETH), each expressed in USD. These are the three largest-cap crypto currencies that use the proof-of-work mechanism which implies a natural price, as production of the cryptocurrencies entails hardware and electricity costs. These mining costs make such crypto currencies comparable to commodities that must be extracted before they can be traded and that therefore also have a natural price.

Table 2: Summary statistics

		Exchange: Bitstamp						Exchange: Coinbase Pro					
Variable*	Pair	Obs.	Mean	Std. Dev.	Min	Max	Gaps	Obs.	Mean	Std. Dev.	Min	Max	Gaps
average price (ap)	BTC/USD	42,397	3,672.419	284.56030	3,124.043	4,258.983	1,977	44,515	3,670.546	286.29840	3,130.007	4,262.333	7
cumulated volume	BTC/USD	42,397	8.452792	19.13613	0.00000001	747.4297	1,977	44,515	10.95885	26.00548	0.0033852	780.322	7
ticker	BTC/USD	43,599	3,669.414	284.82250	3,124.450	4,259.000	745	43,544	3,669.533	285.37100	3,130.000	4,260.990	747
mid-price (mp)	BTC/USD	43,599	3,669.120	284.75700	3,124.395	4,256.485	745	43,544	3,669.463	285.35080	3,130.005	4,260.995	747
vlcp	BTC/USD	43,599	3,669.120	284.75700	3,124.395	4,256.485	745	43,544	3,669.463	285.35090	3,130.005	4,260.995	747
plcp (pd = 1)	BTC/USD	43,599	3,670.005	284.02260	3,130.696	4,257.087	745	43,544	3,669.625	285.31030	3,130.295	4,260.977	747
plcp (pd = 2)	BTC/USD	43,599	3,671.592	283.57370	3,134.293	4,253.387	745	43,544	3,669.987	285.18640	3,130.366	4,260.812	747
plcp (pd = 3)	BTC/USD	43,599	3,673.874	282.50840	3,137.189	4,252.351	745	43,544	3,670.423	284.89480	3,130.801	4,260.651	747
arp (pd = 1)	BTC/USD	43,599	3,669.876	284.28040	3,133.768	4,256.876	745	43,544	3,669.933	285.24630	3,134.686	4,261.971	747
arp (pd = 2)	BTC/USD	43,599	3,670.135	284.20140	3,133.768	4,255.414	745	43,544	3,670.212	285.07940	3,135.707	4,260.744	747
arp (pd = 3)	BTC/USD	43,599	3,670.396	284.00800	3,133.768	4,255.279	745	43,544	3,670.484	284.85500	3,138.276	4,260.265	747
th	BTC/USD	43,593	3,669.238	284.75250	3,123.971	4,256.771	751	43,481	3,669.153	285.42910	3,131.589	4,258.585	751
average price (ap)	ETH/USD	32,828	108.0186	18.64506	81.00368	158.7521	6,500	44,206	107.3039	18.60715	81.05930	158.9046	382
cumulated volume	ETH/USD	32,828	72.88672	195.2765	0.00000001	10,535.9	6,500	44,206	161.7086	368.4197	0.00000908	9,751.561	382
ticker	ETH/USD	43,598	107.3014	18.6538	80.90000	159.0000	745	43,573	107.3293	18.66950	81.02000	158.9000	753
mid-price (mp)	ETH/USD	43,598	107.2962	18.65477	81.02500	159.1000	745	43,573	107.3252	18.66823	81.01500	158.9200	753
vlcp	ETH/USD	43,598	107.2973	18.65509	81.02500	159.1151	745	43,573	107.3244	18.66851	81.01500	158.9200	753
plcp (pd = 1)	ETH/USD	43,598	107.3134	18.67433	81.10455	158.7652	745	43,573	107.3615	18.68454	81.02133	158.9345	753
plcp (pd = 2)	ETH/USD	43,598	107.3066	18.67263	81.16977	158.6250	745	43,573	107.3729	18.67781	81.20864	158.7655	753
plcp (pd = 3)	ETH/USD	43,598	107.3309	18.67615	81.19224	158.9792	745	43,573	107.4017	18.67125	81.22263	158.8996	753
arp (pd = 1)	ETH/USD	43,598	107.3364	18.67386	81.18047	158.9958	745	43,573	107.3657	18.68673	81.00033	158.9191	753
arp (pd = 2)	ETH/USD	43,598	107.3087	18.66048	81.16103	159.0185	745	43,573	107.3603	18.67706	81.12344	158.8850	753
arp (pd = 3)	ETH/USD	43,598	107.3103	18.65994	81.15567	159.0649	745	43,573	107.3652	18.67346	81.11615	158.9086	753
th	ETH/USD	42,038	107.5064	18.63124	81.00939	158.6411	1,391	43,566	107.3254	18.66939	81.06059	158.7901	754
average price (ap)	LTC/USD	14,860	28.72921	3.472321	22.28312	36.54796	7,653	43,484	28.68160	3.515606	22.24804	36.60000	1,019
cumulated volume	LTC/USD	14,860	87.71006	215.8864	0.0000195	7,016.141	7,653	43,484	208.45730	518.9125	0.0000686	16,635.98	1,019
ticker	LTC/USD	43,598	28.67737	3.500075	22.26000	36.50000	745	43,599	28.68564	3.504076	22.29000	36.66000	745
mid-price (mp)	LTC/USD	43,598	28.67417	3.501161	22.29000	36.48000	745	43,599	28.68342	3.504247	22.29000	36.65500	745
vlcp	LTC/USD	43,598	28.67396	3.50122	22.29000	36.45334	745	43,599	28.68335	3.504342	22.29000	36.65500	745
plcp (pd = 1)	LTC/USD	43,598	28.70234	3.502903	22.30774	36.67040	745	43,599	28.69354	3.502409	22.28312	36.68698	745
plcp (pd = 2)	LTC/USD	43,598	28.70207	3.502335	22.35563	36.57466	745	43,599	28.69981	3.500840	22.29339	36.68510	745
plcp (pd = 3)	LTC/USD	43,598	28.71796	3.500369	22.43377	36.63261	745	43,599	28.70578	3.497388	22.28369	36.75701	745
arp (pd = 1)	LTC/USD	43,598	28.69513	3.502843	22.29490	36.56270	745	43,599	28.69688	3.503968	22.29043	36.67239	745
arp (pd = 2)	LTC/USD	43,598	28.68579	3.501306	22.30589	36.49267	745	43,599	28.69210	3.503236	22.29227	36.66402	745
arp (pd = 3)	LTC/USD	43,598	28.68722	3.500684	22.32009	36.50307	745	43,599	28.69255	3.502550	22.29079	36.68007	745
th	LTC/USD	29,872	28.74320	3.468519	22.31970	36.52623	3,219	43,579	28.68353	3.504506	22.25870	36.55967	756

**The listed variables were recorded in the period from 01.12.2018 00:00 until 31.12.2018 23:59 using the application programming interface (API) of the respective crypto exchanges.*

We continuously calculated and recorded all results of the price formation and discarded any raw data due to data storage limitations. To test the forecast accuracy of the models, we recorded all the trades that took place during this period and formed a minute-by-minute volume-weighted average price (ap). The aim of the models is to predict the next minute's ap.

The data was recorded using the application programming interface (API) of the crypto exchanges Bitstamp and Coinbase Pro, which we chose because they support trading against US-Dollar pairs rather than against a cryptocurrency that depicts the US-Dollar, such as Tether, which supposedly represents exactly one US-Dollar but in practice often diverges by at least a few basis points. Coinbase Pro features greater trading volumes than Bitstamp. For example, the average cumulative trading volume per minute during the sample period for the BTC/USD pair was 10.95885 BTC for Coinbase Pro and 8.452792 BTC for Bitstamp.

Depending on the crypto currency and exchange, the data set contained between 7 (0.02%) and 7,653 (17.43%) gaps. These gaps in the reporting can be due either to server problems or to the fact that no trading took place and therefore no prices and volumes could be recorded. Since the results of the mid-price, $vlcp$, $plcp$ and arp models are based on order book data, the gaps of these models are exclusively due to server connectivity problems. In this case the order book was not accessible via web sockets and therefore no data could be recorded. The number of recording gaps due to server connectivity problems is between 745 (1.67%) and 753 (1.69%). In addition to these gaps, the trade history model (th) also has gaps that result from a lack of trading over a period that exceeded the external parameter al, so that no result could be determined. Therefore, this model has between 751 (1.69%) and 3,219 (7.21%) gaps, depending on the currency pair and the exchange. At 6,500 (14.56%) to 7,653 (17.43%), the currency pairs ETH/USD and LTC/USD as traded on Bitstamp featured the most gaps.

So as not to distort the forecast performance of the models, the gaps were not filled by interpolation or similar methods. Rather, it was assumed that no trade would have been possible even at other prices for the gaps. Depending on the type of accuracy measure and the crypto currency and exchange, varying numbers of observations were used to calculate the measures of forecast accuracy. As a result, some of the forecast measures are more meaningful than others, which is why we provide the number of observations used for the calculation for each forecast measure, crypto currency and exchange. For example, only 6,947 observations were available to calculate the Mean Directional Accuracy (MDA) for LTC/USD on Bitstamp, as compared to 43,472 observation for other measurements regarding BTC/USD on Coinbase Pro.

During the observation period (December 2018), the crypto market experienced a phase of sideways movement and relative price stability. During that month, the Bitcoin price on Bitstamp fell from 3,973.253 USD to 3,690.607 USD. The prices on Coinbase Pro and the prices of LTC/USD and ETH/USD show a similar pattern, though ETH/USD rose slightly over the period. The lowest price for a Bitcoin traded on Bitstamp was 3,124.043 USD while the highest price was 4,258.983 USD, a range of 1,134.94 USD. Despite this large range, the standard deviation around the mean of 3,672.419 USD was only 284.5603 USD, or 7.75% of the mean. For ETH/USD and LTC/USD, the standard deviation amounted to 17.26% and 12.07% or the respective means. Thus, the price of Bitcoin was more stable than the other two currencies, yet all three crypto assets were more volatile than most stock prices. Table 2 provides an overview of the dataset.

3.1 Measures of forecast quality

To assess the forecast quality of the price formation models, we first looked at the mean errors (ME) between the weighted average price at minute t and the prediction of each price formation model at time $t-1$. The mean error thus expresses how well the price formation model (x_{t-1}) can determine the price of the next minute (ap_t).

$$ME = \frac{1}{n}\sum_{t=1}^{t=n}(ap_t - x_{t-1}) \quad (16)$$

The mean error itself gives little information about the quality of a forecast, which is why it should be interpreted in conjunction with other measures. But it can give an indication of potential systematic distortions, which can be confirmed by the distribution of the forecast errors. The mean absolute error (MAE), on the other hand, uses the absolute values of the forecast errors and thus provides information about the quality of the price formation models, specifically the average absolute difference between the realized volume-weighted average price and the forecast result of the respective price formation model. The smaller the MAE the better. It is defined as:

$$MAE = \frac{1}{n}\sum_{t=1}^{t=n}|ap_t - x_{t-1}| \quad (17)$$

Next, the root-mean-square error ($RMSE$) is often mistakenly interpreted as MAE but it is the square root of the average quadratic error, so any error has a squared impact on the $RMSE$. Therefore, larger errors have a stronger effect on the $RMSE$ than on the MAE. In conjunction with the MAE, the $RMSE$ can provide information on the size and frequency of outliers of the price formation models. The $RMSE$ is:

$$RMSE = \sqrt{\frac{1}{n}\sum_{t=1}^{t=n}(ap_t - x_{t-1})^2} \quad (18)$$

All of the above forecast measures are scale-dependent, so their results are not comparable across different scales and thus different crypto currency pairs. We therefore also draw upon the mean absolute percentage error ($MAPE$), which is defined as:

$$MAPE = \frac{1}{n}\sum_{t=1}^{t=n}\left|\frac{ap_t - x_{t-1}}{ap_t}\right| x\ 100 \quad (19)$$

Each deviation between the actual volume-weighted price at time t (ap_t) and the price calculated by a price formation model at time $t-1$ (x_{t-1}) is scaled by ap_t. The resulting absolute percentage errors are added up and divided by their number. The result is multiplied by 100 for an easily interpretable percentage.

Another scale-independent measure of forecast quality is the Mean Directional Accuracy (MDA). It compares the direction of movement of the actual volume-weighted price between times $t-1$ and t to the corresponding price movements predicted by each price formation model and counts the number of matches. MDA is defined as:

$$MDA = \frac{1}{n}\sum_{t=1}^{t=n} 1\ sign(ap_t - ap_{t-1}) == sign(x_{t-1} - ap_{t-1}) \quad (20)$$

The signum function $sign(ap_t - ap_{t-1})$ extracts the sign of the result and is an indicator function which delivers the value 1 if the Boolean expression is true and the value 0 if it is false. In the following, these five forecast measures will be used to assess the accuracy of the price formation models introduced above.

4. Results

4.1 Descriptive results

Table 3 provides an overview of the distribution of forecast errors. While the number of forecast errors on Coinbase Pro is between 42,446 and 43,472 for all three crypto currency pairs, the number of observations on Bitstamp varies considerably depending on the crypto currency pair. For example, the number of forecast errors on Bitstamp ranges from 11,872 (LTC) to 41,400 (BTC), depending on the currency pair. This is due to the lower number of trades. Increased trading and higher volumes could have a stabilizing effect on the forecasting power of the price formation models. This is also reflected in the standard deviation of the forecast errors, which is lower on Coinbase Pro for all three currency pairs than on Bitstamp. Furthermore, the forecast errors of the mid-price and the $vlcp$ have the lowest standard deviation across all currency pairs and exchanges.

The mean error (ME) is closest to the ideal value of 0 for the price formation model th, while we often find the greatest distance for $plcp$ with a percentage depth (pd) of 3. The forecast errors can cancel each other out, which is why this information only permits conclusions on any systematic bias in the forecast. Therefore, it makes sense to also interpret skewness and kurtosis. The ideal distribution of the prediction errors should have the bulk of its mass around 0 and be symmetrical, i.e. free of systematic bias. Therefore, a skewness of 0 is desirable. None of the price formation models we tested conform with these expectations for the present data set; each model is skewed either left or right, depending on the exchange and currency pair. However, the data set does not allow a definitive conclusion on a systematic bias of the models, as the sample size is insufficient. Furthermore, factors such as trading volume and trend can influence the skewness of the models. When considering the kurtosis, all models have a leptokurtic distribution, meaning that outlier forecast errors cause a higher kurtosis compared to a normal distribution. It should be noted that the crypto market is highly volatile and numerous and large outliers are to be expected.

The range of forecast errors is depicted by the range between the minimum and maximum values of the distributions. Based on the mean of the volume-weighted average prices per minute, $plcp$ (pd=2) produces the largest outlier, at -11.65193 USD (-10.79%) for ETH/USD on Bitstamp, which gives an impression of in what percentage range the maximum forecast error of this sample moves. Taking this approximation into account, the $vlcp$ features the lowest negative and positive outliers for the currency pair LTC/USD on Coinbase Pro, at -0.4422455 USD (-1.54% off mean *ap*) and 0.4973602 USD (1.73% off mean *ap*).

4.2 Forecast accuracy results

Tables 4 and 5 summarise the performance of the price forecast models according to the accuracy measures. We find that only the mid-price and the $vlcp$ outperform the ticker. The $vlcp$ delivers the best results in all settings except BTC/USD on Coinbase Pro, where the mid-price fares best in terms of MAE, RMSE and MAPE. The arp outperforms the ticker regarding MAE, RMSE and MAPE only for the currency pairs ETH/USD (except pd=1) and LTC/USD on Bitstamp. In terms of the MDA, the arp outperforms the ticker for all currency pairs on Bitstamp. This also applies to LTC/USD on Coinbase Pro. All of the models perform well in terms of MDA Only the ticker and the th predicted the right direction for LTC/USD on Bitstamp in less than 50% of the cases. In all other cases, the MDA was higher than 50%, reaching almost 80% using the $vlcp$ for the BTC/USD currency pair on Coinbase Pro. Therefore, the MDA is also the only measure in which the $plcp$ performs better than the ticker, but only on the exchange Bitstamp. If you take the other measures into account in addition to the MDA, one of the three variants of the $plcp$, with the exception of the BTC/USD currency pair on Coinbase Pro, always yields the worst results. The th, with one exception, has the worst MDA scores.

Th also has high error values in MAE, RMSE and MAPE that make this price formation model the most inaccurate one regarding BTC/USD on Coinbase Pro. Percentage depth (pd) affects the two price formation models in a different way. While the results of the $plcp$ get worse with increasing pd except one case, the results of the arp are very different for each currency pair and crypto exchange. The arp delivers the same MAE, RMSE and MAPE for the pd of 2 and 3. This is not the case for the MDA values where all 3 variants always deliver different results.

Table 3: Descriptive results of forecast errors

		Exchange: Bitstamp							**Exchange: Coinbase Pro**						
Variable	Pair	Obs.	Mean***	Std. Dev.	Min	Max	Skewness	Kurtosis	Obs.	Mean***	Std. Dev.	Min	Max	Skewness	Kurtosis
ticker	BTC/USD	41,400	-0.1428145	3.853002	-65.27563	113.3840	0.5221805	33.05087	43,472	-0.1071414	2.853945	-53.65112	115.0898	1.6709690	109.88060
mid-price	BTC/USD	41,400	0.1363966	3.661642	-68.14551	113.3792	0.5204996	42.12715	43,472	-0.0373140	2.797878	-52.85620	115.0950	1.7704400	117.73320
vlcp	BTC/USD	41,400	0.1363927	3.661639	-68.14551	113.3792	0.5205011	42.12726	43,472	-0.0373310	2.798049	-52.85620	115.0950	1.7716970	117.72060
plcp (pd = 1)	BTC/USD	41,400	-0.7454101	4.492013	-63.77490	113.3997	0.0792365	20.84293	43,472	-0.1995917	2.896663	-52.78857	114.9985	1.4810550	102.64520
plcp (pd = 2)	BTC/USD	41,400	-2.3389820	5.904117	-64.78003	113.1602	-0.6820557	11.27913	43,472	-0.5624896	3.209623	-56.88379	115.0166	0.9109057	73.07854
plcp (pd = 3)	BTC/USD	41,400	-4.6120950**	7.555653	-70.60889	112.1841	-0.8569101	7.487913	43,472	-0.9989331	3.714471	-59.09766	115.0649	0.1397506	45.25375
arp (pd = 1)	BTC/USD	41,400	-0.6231517	4.033147	-66.44995	113.5842	0.1667691	29.84744	43,472	-0.5087502	2.942000	-55.24170	112.2522	1.1721450	91.37984
arp (pd = 2)	BTC/USD	41,400	-0.8835018	4.120588	-67.18335	113.0457	0.0886651	27.82384	43,472	-0.7879328	3.020007	-53.66357	113.8369	1.2799050	86.92400
arp (pd = 3)	BTC/USD	41,400	-1.1451200	4.187862	-68.07227	112.1870	0.0516573	26.02866	43,472	-1.0607610**	3.114840	-53.96118	114.1516	1.2116100	78.19570
th	BTC/USD	41,394	0.0622466*	4.739802	-68.08594	112.0938	0.6705497	28.64003	43,466	0.0031333*	4.025582	-68.02002	114.1594	0.9714150	50.66674
ticker	ETH/USD	32,071	-0.0015450	0.2234428	-10.25539	8.878510	-0.5448075	237.89000	43,172	-0.0032428	0.1378413	-3.693253	5.338250	0.9488835	89.52297
mid-price	ETH/USD	32,071	0.0028443	0.2032276	-10.51039	4.964989	-3.1962300	261.88960	43,172	0.0007962	0.1320679	-3.698250	5.333250	1.0750760	104.93300
vlcp	ETH/USD	32,071	0.0018532	0.2026191	-10.51039	4.904999	-3.2249620	264.64150	43,172	0.0016475	0.1302262	-3.698250	5.321250	1.1032770	109.74290
plcp (pd = 1)	ETH/USD	32,071	-0.0182400	0.2906091	-11.63182	4.652527	-1.8394470	89.61051	43,172	-0.0356000	0.1480548	-3.709229	5.197601	0.2204351	65.61776
plcp (pd = 2)	ETH/USD	32,071	-0.0092176	0.2512048	-11.65193	4.777916	-2.5553030	161.71860	43,172	-0.0469610	0.1633088	-3.736038	5.226929	0.1239929	48.54863
plcp (pd = 3)	ETH/USD	32,071	-0.0337831	0.2557950	-10.59348	4.619133	-1.7099730	107.96230	43,172	-0.0758313**	0.1959872	-3.785088	5.044724	-0.1422650	24.32324
arp (pd = 1)	ETH/USD	32,071	-0.0402868**	0.2311842	-10.93855	4.676468	-2.8361330	176.58790	43,172	-0.0397929	0.1468128	-3.734612	5.047510	0.0103373	63.90995
arp (pd = 2)	ETH/USD	32,071	-0.0107153	0.2081569	-10.93855	4.866119	-3.6262080	270.22660	43,172	-0.0343764	0.1410405	-3.737450	5.167580	0.3573516	79.22394
arp (pd = 3)	ETH/USD	32,071	-0.0120872	0.2066320	-10.60657	4.848122	-3.2271050	250.13630	43,172	-0.0392539	0.1430195	-3.759537	5.079120	0.3465085	71.63035
Th	ETH/USD	31,421	0.0009207*	0.2677621	-10.84866	10.444200	0.2909025	179.65070	43,166	0.0003678*	0.1902531	-3.781166	5.334140	0.3902895	48.78142
Ticker	LTC/USD	14,472	-0.0007101	0.0873739	-0.9375744	0.4965343	-0.0522049	6.171695	42,461	-0.0010371	0.0354229	-0.489542	0.5215492	0.130377	23.95535
mid-price	LTC/USD	14,472	0.0021091	0.0654932	-0.7525730	0.5035858	-0.1073891	7.103322	42,461	0.0011830	0.0336209	-0.4554214	0.4973602	0.1797124	27.66747
vlcp	LTC/USD	14,472	0.0022034	0.0650917	-0.7431812	0.5035858	-0.1137581	6.892412	42,461	0.0012623	0.0327919	-0.4422455	0.4973602	0.2108542	29.67061
plcp (pd = 1)	LTC/USD	14,472	-0.0270455	0.0871657	-0.7736092	0.4435349	-0.0592737	4.727339	42,461	-0.0089732	0.0471586	-0.4731712	0.5649624	-0.0401711	10.18452
plcp (pd = 2)	LTC/USD	14,472	-0.0312074	0.0890187	-0.7848682	0.5145149	-0.3758366	5.199633	42,461	-0.0151752	0.04744	-0.5008621	0.6017933	-0.1066015	10.83832
plcp (pd = 3)	LTC/USD	14,472	-0.0486829**	0.1011004	-0.7873383	0.4550667	-0.7049779	5.567241	42,461	-0.0211312**	0.0557252	-0.5280437	0.5517101	-0.1043430	7.30047
arp (pd = 1)	LTC/USD	14,472	-0.0199254	0.0717577	-0.7540188	0.4739609	-0.1902120	5.690813	42,461	-0.0122715	0.0418143	-0.5015144	0.5211658	-0.3046029	13.74064
arp (pd = 2)	LTC/USD	14,472	-0.0122200	0.0688298	-0.7552490	0.4984360	-0.2236288	6.258894	42,461	-0.0074566	0.0358036	-0.4715919	0.5216732	-0.0406260	22.48294
arp (pd = 3)	LTC/USD	14,472	-0.0136467	0.069117	-0.7555485	0.4974041	-0.1970743	6.085492	42,461	-0.0079102	0.0358172	-0.4814987	0.5126400	-0.0059778	22.22126
th	LTC/USD	11,872	-0.0004761*	0.0920367	-0.9375744	0.7291298	0.0664221	7.979197	42,446	0.0002627*	0.0473474	-0.5510941	0.7700900	0.4660955	21.76047

** Smallest absolute difference between the realized value and the forecast value among the models per exchange and currency.*
*** Largest absolute difference between the realized value and the forecast value among the models per exchange and currency.*
**** The mean of the forecast errors is also called mean error (ME) and corresponds to the measure presented as equation (16) above.*

5. Discussion

For the purpose of short-term price prediction, we tested various price formation models to determine the best possible reference price for next minute transactions. Our point of comparison was the widely used reference price, the price of the last trade, or ticker. To evaluate the results, we discussed, compared and applied five different measures of forecast accuracy.

Especially the $vlcp$ produced satisfactory results in terms of MAE, RMSE, MAPE and MDA. With few exceptions, the $vlcp$ yielded the smallest forecast errors. However, the mid-price, one of the simplest methods, performed similarly well, while the performance of the more complex models such as $plcp$ and arp fell short of the ticker.

The model th, while comparable to the ticker in its use of past data, also scored worse than the ticker, in particular in terms of the MDA. As mentioned before, the RMSE weighs outliers more heavily than the MAE. Joint consideration of these two errors can therefore also provide information on how often a price formation model delivers outliers. However, the results presented in Table 4 do not provide any indication that a price formation model has extreme outliers as there are no significant differences between the RMSE and the MAE.

Table 4: Prediction accuracy statistics

		Exchange: Bitstamp				**Exchange: Coinbase Pro**			
Variable	Currency Pair	Obs.	MAE	RMSE	MAPE (in %)	Obs.	MAE	RMSE	MAPE (in %)
ticker	BTC/USD	41,400	2.5008290	3.8556010	0.0676906	43,472	1.3942840	2.8559230	0.0377010
mid-price	BTC/USD	41,400	2.3837850	3.6641370	0.0644824*	43,472	1.3368460*	2.7980950*	0.0361454*
vlcp	BTC/USD	41,400	2.3837830*	3.6641350*	0.0644824*	43,472	1.3368700	2.7982660	0.0361462
plcp (pd = 1)	BTC/USD	41,400	3.1394030	4.5533870	0.0858069	43,472	1.4802030	2.9034980	0.0400555
plcp (pd = 2)	BTC/USD	41,400	4.3306660	6.3504800	0.1197455	43,472	1.7814480	3.2585020	0.0484595
plcp (pd = 3)	BTC/USD	41,400	6.2953940**	8.8520020**	0.1760370**	43,472	2.1920540	3.8464070	0.0601219
arp (pd = 1)	BTC/USD	41,400	2.7277360	4.0809560	0.0744427	43,472	1.7306120	2.9856310	0.0470424
arp (pd = 2)	BTC/USD	41,400	2.8097420	4.2141910	0.0769994	43,472	1.8593170	3.1210680	0.0508788
arp (pd = 3)	BTC/USD	41,400	2.8097420	4.2141910	0.0769994	43,472	1.8593170	3.1210680	0.0508788
th	BTC/USD	41,394	3.0625170	4.7401540	0.0829009	43,466	2.3309770**	4.0255370**	0.0629825**
ticker	ETH/USD	32,071	0.1350433	0.2234446	0.1218861	43,172	0.0993632	0.1599703	0.0897102
mid-price	ETH/USD	32,071	0.1257658	0.2032443	0.1134796	43,172	0.0918924	0.1532304	0.0830182
vlcp	ETH/USD	32,071	0.1253738*	0.2026244*	0.1131290*	43,172	0.0903942*	0.1511029*	0.0816293*
plcp (pd = 1)	ETH/USD	32,071	0.2069371**	0.2911764**	0.1875569**	43,172	0.1233823	0.1766720	0.1115929
plcp (pd = 2)	ETH/USD	32,071	0.1696992	0.2513699	0.1562302	43,172	0.1474271	0.1971524	0.1356303
plcp (pd = 3)	ETH/USD	32,071	0.1777428	0.2580123	0.1642171	43,172	0.1934527**	0.2438157**	0.1809024**
arp (pd = 1)	ETH/USD	32,071	0.1546615	0.2346647	0.1399154	43,172	0.1219117	0.1764811	0.1099843
arp (pd = 2)	ETH/USD	32,071	0.1311313	0.2084292	0.1188359	43,172	0.1163638	0.1684283	0.1061308
arp (pd = 3)	ETH/USD	32,071	0.1311313	0.2084292	0.1188359	43,172	0.1163638	0.1684283	0.1061308
th	ETH/USD	31,421	0.1600286	0.2677594	0.1441450	43,166	0.1499057	0.2229915	0.1357589
ticker	LTC/USD	14,472	0.0611823	0.0873738	0.2128796	42,461	0.0201459	0.0354377	0.0702414
mid-price	LTC/USD	14,472	0.0511455	0.0655249	0.1780156	42,461	0.0184600	0.0336414	0.0643586
vlcp	LTC/USD	14,472	0.0511193*	0.0651268*	0.1779636*	42,461	0.0179407*	0.0328158*	0.0625378*
plcp (pd = 1)	LTC/USD	14,472	0.0698912	0.0912622	0.2434180	42,461	0.0338185	0.0480042	0.1185498
plcp (pd = 2)	LTC/USD	14,472	0.0713118	0.0943276	0.2500376	42,461	0.0353329	0.0498075	0.1244850
plcp (pd = 3)	LTC/USD	14,472	0.0830481**	0.1122079**	0.2935278**	42,461	0.0441307**	0.0595965**	0.1570750**
arp (pd = 1)	LTC/USD	14,472	0.0570743	0.0744704	0.1986870	42,461	0.0288751	0.0435773	0.1010922
arp (pd = 2)	LTC/USD	14,472	0.0541281	0.0699038	0.1887111	42,461	0.0225699	0.0365714	0.0790098
arp (pd = 3)	LTC/USD	14,472	0.0541281	0.0699038	0.1887111	42,461	0.0225699	0.0365714	0.0790098
th	LTC/USD	11,872	0.0636875	0.0920341	0.2215854	42,446	0.0284269	0.0473476	0.0988673

** lowest value (highest degree of accuracy)*
*** highest value (lowest degree of accuracy)*

All 4 error measures can be compared across the exchanges for a given currency pair. A uniform picture for all measures and currency pairs emerges; on Coinbase Pro, which has the highest trading volume for all 3 currency pairs, all price formation models perform better than they do on the less busy exchange Bitstamp. The MAPE allows a comparison of the price formation models beyond currency pairs and crypto exchanges. This confirms the influence of trading volume, as described above. The greater the volume, the smaller the MAPEs, which supports the argument that liquidity fosters market efficiency. Conversely, this finding enables the detection of wash trading, that is, trades in which the buyer and the seller are the same entity. Such trades do not add to price discovery and are considered manipulative. As Coinbase Pro has more trading volume than Bitstamp in all currency pairs, the MAPEs are also lower in the former. However, considering the BTC/USD pair on Bitstamp and ETH/USD on Coinbase Pro, it appears that regardless of the crypto exchange, the high-volume currency pair (BTC/USD) leads to lower MAPEs.

The ME of the price formation models points to systematic bias in the forecasts. The mid-price and $vlcp$ are positive for some currency pairs and exchanges while they are negative for others. It is therefore unclear whether these two models are biased. For the remaining models the MEs are negative, which could be an indication of a general overestimation. However, there is also the possibility that individual extreme overestimates of the models may overcompensate those of the underestimates and thus merely give the impression of systemic distortion.

Table 5: Mean Directional Accuracy results

		Exchange: Bitstamp		**Exchange: Coinbase Pro**	
Variable	Currency Pair	Obs.	MDA (in %)	Obs.	MDA (in %)
ticker	BTC/USD	39,457	61.7507667	43,466	75.3485483
mid-price	BTC/USD	39,457	70.0737512**	43,466	79.1745272
vlcp	BTC/USD	39,457	70.0737512**	43,466	79.1791285**
plcp (pd = 1)	BTC/USD	39,457	63.4817650	43,466	74.3937790
plcp (pd = 2)	BTC/USD	39,457	60.6837823	43,466	71.4696544
plcp (pd = 3)	BTC/USD	39,457	57.1888385	43,466	69.3277504
arp (pd = 1)	BTC/USD	39,457	66.7359404	43,466	73.5632448
arp (pd = 2)	BTC/USD	39,457	67.0552754	43,466	71.8009479
arp (pd = 3)	BTC/USD	39,457	66.4419495	43,466	69.9765334
th	BTC/USD	39,456	58.0165247*	43,463	55.0698295*
ticker	ETH/USD	25,704	55.2443200	42,799	69.0623613
mid-price	ETH/USD	25,704	66.4332400	42,799	75.7751349
vlcp	ETH/USD	25,704	66.7289138**	42,799	76.5485175**
plcp (pd = 1)	ETH/USD	25,704	56.9366636	42,799	66.4992173
plcp (pd = 2)	ETH/USD	25,704	60.1190476	42,799	64.0902825
plcp (pd = 3)	ETH/USD	25,704	60.2668845	42,799	61.0598378
arp (pd = 1)	ETH/USD	25,704	62.2082166	42,799	67.8730811
arp (pd = 2)	ETH/USD	25,704	65.2699969	42,799	68.8660950
arp (pd = 3)	ETH/USD	25,704	65.5539994	42,799	68.0156078
th	ETH/USD	25,694	54.5964038*	42,795	55.0858745*
ticker	LTC/USD	6,986	44.9184082*	41,463	61.7827943
mid-price	LTC/USD	6,986	61.0506728	41,463	71.8399537
vlcp	LTC/USD	6,986	61.7377612**	41,463	73.2942624**
plcp (pd = 1)	LTC/USD	6,986	53.9650730	41,463	58.3894074
plcp (pd = 2)	LTC/USD	6,986	54.5949041	41,463	60.3164267
plcp (pd = 3)	LTC/USD	6,986	54.5949041	41,463	58.1892290
arp (pd = 1)	LTC/USD	6,986	58.4025193	41,463	62.7812749
arp (pd = 2)	LTC/USD	6,986	60.0772975	41,463	67.1128476
arp (pd = 3)	LTC/USD	6,986	59.9627827	41,463	66.6087837
th	LTC/USD	6,947	47.6608608	41,455	54.1888795*

** lowest value (worst performance).*
*** highest value (best performance).*

The examination of the different weights of the percentage depth parameter used in the $plcp$ and arp models has shown that more collected data in terms of higher parameter weighting does not enhance the predictive power for this data set. This could mean that the information value contained in orders that are placed deep in the book but intended to be executed is diluted and outweighed by the misleading information contained in orders that are placed solely to move the market. The measures of forecast quality suggest that the $plcp$ model fares worse with increasing percentage depths, so the benefit of recording the data at higher percentage depths may be doubted. However, the opposite often applies to the arp model. Although the arp model delivered roughly the same results for percentage depths of 2 and 3, raising the parameter from 1 to 2 improved the forecast quality, which justifies additional data recording effort.

The analyses allow us to reject our null hypothesis, which held that none of the price formation models we examined is superior to the ticker. The models $plcp$ and mid-price in fact delivered superior forecast quality. This holds implications for science and practice.

Other prediction models such as the ARIMA model could use the $vlcp$ instead of historical price data and possibly produce more accurate results. The $vlcp$ may indeed become a more relevant reference price than the ticker. This model should thus be taken into account in future research.

Our results indicate that the $vlcp$ is a good short-term prediction model and may therefore be used by investors to estimate the value of future transactions. Used as a signal for trading algorithms and deep-learning, it may facilitate more accurate results. Market makers can base their price discovery strategy on it or combine currently used methods with the results of $vlcp$. The results suggest that it might be useful for crypto and traditional exchanges to display the $vlcp$ n next to the ticker and the order book to provide traders with additional information.

Yet the present study also has a number of limitations which we will discuss briefly in the following. Firstly, it suffers from limited data quality and availability. Since the interfaces from which the data were obtained were not always accessible, the record has gaps, which reduce the representativeness of the results and necessitate further investigations on other datasets. Furthermore, it would be worthwhile to check whether the results we obtained from minute data also apply to shorter or longer intervals.

The crypto pairs analysed in this study were selected on the basis of their large market capitalisation; the crypto exchanges were picked because they list USD currency pairs. It remains unclear whether our results would hold when applied to other exchanges with their own sets of procedures and rules, and to other crypto currencies, or to altogether different classes of assets.

Some of the price formation models use externally specified parameters. However, with the exception of three different depth values we did not test different values for each parameter, so we cannot state the optimal values for each parameter in a given situation.

The $vlcp$ has emerged as the most accurate model, so further research in this area appears to be warranted. Determining the right volume limit is an interesting challenge for future research. Yet in the less successful models, too, different external parameters could be tested to improve prediction performance. Furthermore, the robustness of the model parameters should be examined more closely. Finally, we did not test a number of existing price formation models such as the micro-price.

The implications for further research immediately arise from these limitations. The robustness of the models should be checked for other crypto currency pairs and crypto exchanges. An analysis of the price formation models over different time series by means of the Diebold-Mariano test for differences in the mean square error or an encompassing test such as the Fair-Shiller test are promising tasks for further investigation in this area (Diebold and Mariano 1995; Mizon 1984; Mizon and Richard 1986; Fair and Shiller 1988).

The comparison of traditional financial data and crypto data is another interesting subject for further research. It would clarify whether the price formation models perform equally well in both markets. Another possible field of research would be the comparison with traditional financial market forecast models. It remains to be determined whether the price formation models, which are based primarily on current data, are superior to the traditional methods of technical analysis, which are mostly based on historical data or whether there can be a meaningful combination of both approaches.

Finally, as already mentioned, it would be worthwhile to test how well the price formation models perform when fed not with minute data but with longer or shorter time periods. More specifically, it would be interesting to see whether the $vlcp$ remains the most accurate model when applied to different time periods.

6. Conclusion

In the current system of finance and crypto currency exchanges, the price of the last transaction is issued as the authoritative price. Yet doubts arise as to whether the ticker, looking solely at the past and ignoring the information contained in trading volume, can be considered representative of future transactions.

We thus proposed a number of alternative market price formation models and assessed their ability to predict the price of the next trade. Various tests were conducted to determine whether any of these computational models could beat the ticker and can therefore be considered more representative of future trade. In a first step, different measures of forecast quality were created and discussed. Additionally, the price formation models were checked for systematic over- or underestimations using their mean errors. It turned out that, depending on the currency pair and crypto market, the bias produced by a price formation model may shift. An analysis of the distribution of the forecast errors revealed a leptokurtosis for all price formation models. We found that the mid-price and the volume-limited clearing price ($vlcp$) provided more accurate forecasts than the other models and the ticker.

In conclusion, this work has revealed a number of new findings that may prove very valuable in financial market practice and that furthermore provide a solid foundation for further research. The simple mid-price and $vlcp$ models produced the most accurate forecast results and are therefore most representative of future transactions. The null hypothesis of this work that none of the examined price formation models is superior to the ticker can thus be rejected.

References

Agarwal, Anuj (2012): High frequency trading. Evolution and the future. In *Capgemini, London, UK*, p. 20.

Ahn, Hee-Joon; Bae, Kee-Hong; Chan, Kalok (2001): Limit Orders, Depth, and Volatility. Evidence from the Stock Exchange of Hong Kong. In *The journal of finance* 56 (2), pp. 767–788. DOI: 10.1111/0022-1082.00345.

Avellaneda, Marco; Stoikov, Sasha (2008): High-frequency trading in a limit order book. In *Quantitative Finance* 8 (3), pp. 217–224.

Biais, Bruno; Hillion, Pierre; Spatt, Chester (1995): An empirical analysis of the limit order book and the order flow in the Paris Bourse. In *The journal of finance* 50 (5), pp. 1655–1689.

Blockchain Transparency Institute (2018): Exchange Volumes Report. December 2018. Available online at https://www.blockchaintransparency.org/, updated on 3/29/2019, checked on 3/29/2019.

Bollerslev, Tim; Engle, Robert F.; Wooldridge, Jeffrey M. (1988): A capital asset pricing model with time-varying covariances. In *Journal of political Economy* 96 (1), pp. 116–131.

Bonart, Julius; Lillo, Fabrizio (2016): A Continuous and Efficient Fundamental Price on the Discrete Order Book Grid. In *SSRN Journal. DOI:* 10.2139/ssrn.2817279.

Cartea, Alvaro; Jaimungal, Sebastian (2016): A closed-form execution strategy to target volume weighted average price. In *SIAM Journal on Financial Mathematics* 7 (1), pp. 760–785.

Cont, Rama (2011): Statistical modeling of high-frequency financial data. In *IEEE Signal Processing Magazine* 28 (5), pp. 16–25.

Cont, Rama; Kukanov, Arseniy (2017): Optimal order placement in limit order markets. In *Quantitative Finance* 17 (1), pp. 21–39.

Cont, Rama; Kukanov, Arseniy; Stoikov, Sasha (2014): The price impact of order book events. In *Journal of financial econometrics* 12 (1), pp. 47–88.

Damodaran, Aswath (2012): Investment valuation. Tools and techniques for determining the value of any asset: John Wiley & Sons (666).

Darley, Vince; Outkin, Alexander; Plate, Tony; Gao, Frank (Eds.) (2000): Sixteenths or pennies? Observations from a simulation of the NASDAQ stock market: IEEE.

Diebold, Francis X.; Mariano, Roberto S. (1995): Comparing Predictive Accuracy. In *Journal of Business & Economic Statistics*, pp. 253–263.

Easley, David; O'hara, Maureen (1987): Price, trade size, and information in securities markets. In *Journal of financial economics* 19 (1), pp. 69–90.

Eisler, Zoltan; Bouchaud, Jean-Philippe; Kockelkoren, Julien (2012): The price impact of order book events. Market orders, limit orders and cancellations. In *Quantitative Finance* 12 (9), pp. 1395–1419.

Fair, Ray; Shiller, Robert (1988): The Informational Content of Ex Ante Forecasts. Cambridge, MA: National Bureau of Economic Research.

Foucault, Thierry; Kadan, Ohad; Kandel, Eugene (2005): Limit order book as a market for liquidity. In *The Review of Financial Studies* 18 (4), pp. 1171–1217.

Frei, Christoph; Westray, Nicholas (2015): Optimal execution of a VWAP order. A stochastic control approach. In *Mathematical Finance* 25 (3), pp. 612–639.

Fusaro, Teddy; Hougan, Matt (2019): Memorandum: Bitwise Asset Management Presentation to the U.S. Securities and Exchange Commission. Available online at www.sec.gov/comments/sr-nysearca-2019-01/srnysearca201901-5164833-183434.pdf, checked on 3/29/2019.

Gatheral, Jim; Oomen, Roel C. A. (2010): Zero-intelligence realized variance estimation. In *Finance Stoch* 14 (2), pp. 249–283.

Ghysels, Eric; Nguyen, Giang (2018): Price Discovery of a Speculative Asset. Evidence from a Bitcoin Exchange. In *SSRN Journal. DOI:* 10.2139/ssrn.3258508.

Glosten, Lawrence R. (1994): Is the electronic open limit order book inevitable? In *The journal of finance* 49 (4), pp. 1127–1161.

Gould, Martin D.; Porter, Mason A.; Williams, Stacy; McDonald, Mark; Fenn, Daniel J.; Howison, Sam D. (2013): Limit order books. In *Quantitative Finance* 13 (11), pp. 1709–1742.

Graham, Benjamin; Dodd, David L. (1940): Security analysis. Principles and techniques: McGraw-Hill Book Company.

Guéant, Olivier; Royer, Guillaume (2014): VWAP execution and guaranteed VWAP. In *SIAM Journal on Financial Mathematics* 5 (1), pp. 445–471.

Jaisson, Thibault (2015): Liquidity and impact in fair markets. In *Market Microstructure and Liquidity* 1 (02), p. 1550010.

Kempf, Alexander; Korn, Olaf (1999): Market depth and order size. In *Journal of Financial Markets* 2 (1), pp. 29–48. DOI: 10.1016/S1386-4181(98)00007-X.

Konishi, Hizuru (2002): Optimal slice of a VWAP trade. In *Journal of Financial Markets* 5 (2), pp. 197–221.

Laruelle, Sophie; Lehalle, Charles-Albert; Pagès, Gilles (2013): Optimal posting price of limit orders. Learning by trading. In *Math Finan Econ* 7 (3), pp. 359–403. DOI: 10.1007/s11579-013-0096-7.

Lehalle, Charles-Albert; Mounjid, Othmane (2017): Limit order strategic placement with adverse selection risk and the role of latency. In *Market Microstructure and Liquidity* 3 (01), p. 1750009.

Lin, Ji-Chai; Sanger, Gary C.; Booth, G. Geoffrey (1995): Trade size and components of the bid-ask spread. In *The Review of Financial Studies* 8 (4), pp. 1153–1183.

Madhavan, Ananth (2002): VWAP strategies. In *Trading* 1, pp. 32–39.

Madhavan, Ananth; Panchapagesan, Venkatesh (2000): Price discovery in auction markets. A look inside the black box. In *The Review of Financial Studies* 13 (3), pp. 627–658.

Madhavan, Ananth; Richardson, Matthew; Roomans, Mark (1997): Why do security prices change? A transaction-level analysis of NYSE stocks. In *The Review of Financial Studies* 10 (4), pp. 1035–1064.

Mastromatteo, Iacopo; Toth, Bence; Bouchaud, Jean-Philippe (2014): Agent-based models for latent liquidity and concave price impact. In *Physical Review E* 89 (4), p. 42805.

Menkveld, Albert J. (2013): High frequency trading and the new market makers. In *Journal of Financial Markets* 16 (4), pp. 712–740.

Merton, Robert C. (1973): An intertemporal capital asset pricing model. In *Econometrica* 41 (5), pp. 867–887.

Mizon, Grayham E. (1984): The encompassing approach in econometrics: Australian National University, Faculty of Economics and Research School of …

Mizon, Grayham E.; Richard, Jean-Francois (1986): The Encompassing Principle and its Application to Testing Non-Nested Hypotheses. In *Econometrica* 54 (3), p. 657. DOI: 10.2307/1911313.

Ntakaris, Adamantios; Magris, Martin; Kanniainen, Juho; Gabbouj, Moncef; Iosifidis, Alexandros (2018): Benchmark dataset for mid-price forecasting of limit order book data with machine learning methods. In *Journal of Forecasting* 37 (8), pp. 852–866. DOI: 10.1002/for.2543.

O'Hara, Maureen (2015): High frequency market microstructure. In *Journal of financial economics* 116 (2), pp. 257–270. DOI: 10.1016/j.jfineco.2015.01.003.

Ricardo, David (1817): On the principles of political economy and taxation: London: John Murray.

Robert, Christian Yann; Rosenbaum, Mathieu (2012): Volatility and covariation estimation when microstructure noise and trading times are endogenous. In *Mathematical Finance* 22 (1), pp. 133–164.

Roşu, Ioanid (2009): A dynamic model of the limit order book. In *The Review of Financial Studies* 22 (11), pp. 4601–4641.

Sharpe, William F. (1963): A simplified model for portfolio analysis. In *Management Science* 9 (2), pp. 277–293.

Smith, Adam (1776): An inquiry into the nature and causes of the wealth of nations. Volume One. In : London: printed for W. Strahan; and T. Cadell, 1776.

Stackelberg, Heinrich von (1934): Marktform und Gleichgewicht: J. springer.

Stoikov, Sasha (2017): The Micro-Price. In *SSRN Journal. DOI:* 10.2139/ssrn.2970694.

Taylor, Mark P.; Allen, Helen (1992): The use of technical analysis in the foreign exchange market. In *Journal of international Money and Finance* 11 (3), pp. 304–314.

Tóth, B.; Lempérière, Y.; Deremble, C.; Lataillade, J. de; Kockelkoren, J.; Bouchaud, J.-P. (2011): Anomalous Price Impact and the Critical Nature of Liquidity in Financial Markets. In *Phys. Rev. X* 1 (2), p. 383. DOI: 10.1103/PhysRevX.1.021006.