Efficient or not? Price measures in market microstructure

Tino Cestonaro[a], Niklas Trimpe[a]

[a] *Goethe University Frankfurt, Theodor-W.-Adorno-Platz 4, 60323 Frankfurt, Germany*

Abstract

We systematically evaluate the qualitative and quantitative properties of five price measures commonly used in market microstructure and assess their informational efficiency in approximating an asset's true value using quote and trade data from German stocks. By using return predictability as an inverse measure of efficiency, we find that more sophisticated measures, such as the micro-price, reflect public information within two seconds. In contrast, established price measures like the transaction price and midpoint require at least 30 seconds to incorporate public information. Relying on inefficient proxies for an asset's true value can introduce significant and systematic biases in study results and trading outcomes, leading to overestimated transaction costs and unfair dark pool execution prices. Our findings offer practical guidance for selecting efficient true value estimators, informing both the design of future research and investor decision-making.

Keywords: Price efficiency, Price measures, True value, Market microstructure, High-frequency trading
JEL: G14, G17

⋆We gratefully acknowledge research support from efl – the Data Science Institute.

Email addresses: `cestonaro@wiwi.uni-frankfurt.de` (Tino Cestonaro), `trimpe@wiwi.uni-frankfurt.de` (Niklas Trimpe)

March 26, 2025

1. Introduction

According to the semi-strong form of the efficient market hypothesis, the price of an asset fully reflects all publicly available information (Fama, 1970). In market microstructure theory, this efficient price is equal to the true value of an asset and is typically modeled as a continuous and unpredictable variable following a random walk with martingale properties. The true value of an asset is a central element in numerous market microstructure models (e.g., Kyle, 1985) and serves as input for important concepts such as liquidity or volatility measures.

However, the true value of an asset is inherently unobservable, as its discounted future cash flows are subject to uncertainty. In reality, researchers and practitioners rely on observable trading outcomes, such as transaction prices or the current best bid and offer prices, to approximate the true value of an asset. Despite their widespread application as true value proxies, price measures like the midpoint - the average of the best bid and offer price - are known to be inefficient true value proxies, because of several microstructure frictions. For instance, the midpoint is a discrete variable, constrained by the tick size, the minimum incremental price change, whereas the true value is by definition a continuous variable. In addition, theory suggests that liquidity suppliers set best bid and offer prices depending on their inventory, potentially resulting in quotes whose average significantly differs from the true value (Hendershott and Menkveld, 2014). Empirical studies further demonstrate the predictability of future midpoint returns contradicting the martingale assumption and questioning the midpoint's validity as a proxy for an asset's true value (Chordia et al., 2005; Rösch et al., 2017; Goldstein et al., 2023).

The application of inefficient price measures, such as the midpoint, can significantly affect both research findings and investors' decision-making. For instance, Hagströmer (2021) demonstrates that using the midpoint to approximate the true value can lead to substantial overestimation of transaction costs. This finding calls into question the conclusions of numerous prior studies on asset liquidity and highlights the potential for investors to make suboptimal trading decisions based on miscalculated costs. Consequently, developing more efficient proxies for the true value of an asset remains a critical challenge in market microstructure research, especially as modern financial markets demand proxies that update at the high-frequency level.

Against this background, the goal of this research is to compare the qualitative and quantitative properties of various price measures used in market microstructure and to systematically evaluate their efficiency. Furthermore, we aim to analyze the determinants of price measures' efficiency and demonstrate the consequences for empirical research and practitioners of using inefficient true value proxies. Specifically, we aim to provide insight into the following three research questions: (1)

How do different price measures compare in their ability and speed of reflecting past information?, (2) Does the ability to reflect information depend on the extent of market frictions, and do the key drivers differ across various price measures?, and (3) To what extent can the use of inefficient price measures impact researchers' findings and investors' trading outcomes? In our analysis, we examine five distinct price measures: transaction price, midpoint, quantity-weighted midpoint, tick-size constrained quantity-weighted midpoint, and the micro-price according to Stoikov (2018).

Previous research assessing the efficiency of low frequency transaction price returns, using weekly or daily prices, predominantly find no return predictability in developed or advanced emerging markets (Chordia et al., 2002; Kim and Shamsuddin, 2008). Recent literature focuses on intraday midpoint returns and identifies inefficiencies over short time intervals, demonstrating the predictability of five-minute up to five-second returns (Cao et al., 2009; Rösch et al., 2017; Aït-Sahalia et al., 2022). The majority of empirical studies evaluate the extent to which a price time series follows a random walk by analyzing the autocorrelation or predictability of returns using econometric approaches such as variance ratio (Lo and MacKinlay, 1989) or various regression models (Hendershott and Jones, 2005).

However, previous studies on price efficiency often overlook relevant aspects. In modern financial markets, high-frequency traders react to new information within microseconds, submitting and canceling orders at an extraordinary pace. Yet, their fast trading activity primarily consists of limit orders at the best bid and offer or lower price levels and, hence, does not affect rigid price measures such as the transaction price or midpoint (Brogaard et al., 2019). Consequently, these conventional price measures may not fully reflect prior information and may be unsuitable for approximating an asset's true value, especially for assets heavily traded by high-frequency traders. Nevertheless, the existing literature has predominantly focused on either transaction prices or midpoints when investigating price efficiency. To date, no study has systematically analyzed and compared the efficiency of different price measures.

Our study seeks to address this existing research gap by evaluating the qualitative and quantitative properties of both established and alternative price measures, as well as analyzing their degrees of efficiency. Our results are designed to assist researchers in selecting the most appropriate price measure based on their research objectives and data availability. By doing so, we contribute to reducing systematic biases in empirical microstructure research, which are prevalent in many published studies, as highlighted by Hagströmer (2021).

To address our research questions, we evaluate the efficiency of the five mentioned price measures using limit order book data from all DAX40 constituents between

January 2 and June 30, 2023. The DAX40 represents the 40 largest companies by market capitalization, listed on Germany's main stock exchange, Xetra. We employ the predictability of future stock returns, calculated using the five different price measures, as our inverse measure of price efficiency. In our empirical analysis, we estimate various regression models to predict these returns across different prediction horizons based on prior public information, including lagged values of the market return, the asset return, and imbalances in the asset's order book. For each of the five price measures, we estimate regression models based on a single trading day, and evaluate them out-of-sample on the next trading day. The larger the out-of-sample predictability across all asset-days, the less efficient is the price measure. To account for high-frequency dynamics, we use lags and prediction horizons starting at 100 milliseconds (ms). However, rigid price measures, such as transaction price or midpoint, predominantly yield zero returns when sampled at this frequency, making it difficult to identify correlations and potential inefficiencies. To overcome this issue, we propose a regression approach with lagged independent variables calculated over progressively larger intervals. Thereby, our approach can effectively capture fast information dissemination in financial markets fostered by high-frequency traders, while also addressing the problem of zero-inflated returns caused by high sampling frequencies. In contrast, traditional efficiency metrics, such as the variance ratio and autocorrelation, cannot be applied effectively at high frequencies, because their estimation requires time series with sufficient variation (Conrad et al., 2015).

Our results indicate that all price measures exhibit significant predictability based on lagged public information, demonstrating that they are not perfectly efficient in the sense of the semi-strong market efficiency hypothesis. Among all evaluated price measures, the transaction price is the least efficient, exhibiting the highest level of predictability. Specifically, we find that - on average - it takes more than 30 seconds for the transaction price to fully reflect public information. In contrast, the more sophisticated price measures, the micro-price and the tick-size constrained quantity-weighted midpoint, are the most efficient. These measures exhibit the lowest predictability in our sample and take less than 5 seconds to fully incorporate public information. Specifically, we find that a larger share of passive informed order flow, measured by the average order book imbalance, increases the predictability of returns. Asynchronous trading and insufficient activity in the order book also increase the inefficiency of the price measures in the short-term. Moreover, when a stock's incremental price change is constrained by a relatively large tick size compared to its price level, true value estimators are less efficient. In general, the price measures are affected similarly by the examined frictions, except for the transaction price, indicating a substantial difference between trade-driven and order-driven price measures.

Demonstrating the implications of our results for researchers and practitioners, we show that inefficient price measures deviate, on average, between 1.47 and 2.68 basis points from the most efficient price measure, highlighting the economic relevance of selecting an appropriate true value proxy. For instance, when calculating transaction costs using the effective spread, we show that inefficient price measures introduce a bias of up to 38% compared to the most efficient measure. Additionally, we demonstrate that dark pool executions are more likely to occur when the reference price deviates more strongly from the true value. On average, execution prices deviate by 1.83 basis points from the true value, leading to implicit trading costs for the disadvantaged side in the transaction.

This paper makes three key contributions to the literature, advancing the understanding of price efficiency and true value proxies in market microstructure. First, we provide a comprehensive overview of existing true value proxies in market microstructure, analyzing their different properties and revealing significant variations in price efficiency across these measures. To the best of our knowledge, this is the first study to comprehensively evaluate these differences. Second, we show that the use of inefficient price measures can have substantial implications for both researchers and investors, potentially leading to biased empirical results and suboptimal decision-making. We find similar results as Hagströmer (2021) regarding biases in transaction cost estimation, however, our analysis enhances these findings by quantifying the deviations from each price measure to the best true value proxy and examining the underlying determinants of these deviations. Moreover, our results suggest that price discovery mechanisms relying on a reference price, such as dark pools, are systematically used to exploit inefficiencies in the reference price. The fairness in dark pool executions can be improved by adopting more efficient price measures as the reference price. Third, we advance the research on statistical tests of price efficiency by presenting a robust empirical approach for assessing price efficiency in high-frequency trading environments. Compared to traditional measures of price efficiency (e.g. Lo and MacKinlay, 1989; Hendershott and Jones, 2005; Hou and Moskowitz, 2005), our approach yields reliable estimates regarding return predictability, even at high sampling frequencies with minimal price variation and zero-inflated return series. While most closely related to the price delay measure of Hou and Moskowitz (2005), our regression models incorporate multiple sources of past information across increasing time intervals to predict out-of-sample returns at various prediction horizons.

In summary, our results guide researchers in selecting suitable price measures based on their research questions and data restrictions, while highlighting the implications of inefficiencies in these measures. We also offer practical guidelines for investors to more accurately approximate an asset's true value, enabling more precise assessments of execution quality and transaction costs. Lastly, our findings have implications for market design, particularly in environments such as dark pools, where accurate true value approximations are crucial.

The remainder of this paper is structured as follows: Section 2 provides an overview of the price measures under investigation, describes the dataset, and presents key descriptive statistics. Section 3 outlines the empirical framework used to assess price efficiency. Section 4 presents the main results. Section 5 discusses the optimal selection of price measures considering data and computational constraints. Finally, Section 6 concludes.

2. Background and Data

2.1. Price Measures

In this chapter, we provide an overview of the price measures analyzed in this study and conduct a qualitative comparison of their individual characteristics.

The *transaction price* is a widely used price measure in financial research, especially when examining lower frequencies such as daily or monthly. The transaction price is the actual price at which a trade is executed between a buyer and a seller. As the transaction price is only observable at the time of a trade, we define its time series representation as

$$p_t^{tp} = TP_{t_i}, \tag{1}$$

where p_t^{tp} is the transaction price at time t and TP_{t_i} is the price of the last executed trade at discrete time t_i, with $t_i \leq t$. While requiring no calculations and the least granular data (transaction price data) among the five price measures investigated in this study, the transaction price has significant limitations. Its observations are tied to the occurrence of trades, making them sporadic, especially for less heavily traded assets. Infrequent and asynchronous trading can result in time series dominated by zero returns when sampled at higher frequencies, reducing their usefulness for high-frequency analysis. Additionally, transaction price changes are constrained to discrete jumps between the best bid and ask prices. This constraint limits the price measure's granularity to the tick size and can cause the price series to bounce between these two values, potentially introducing serial correlation (Glosten and Harris, 1988).

The *midpoint* is arguably the most commonly used price measure in studies analyzing financial markets at an intraday level. It is defined as the average of the best available bid and ask prices. In a time series context it is defined as

$$p_t^{mid} = 0.5 \left(p_t^{bid} + p_t^{ask} \right), \tag{2}$$

with p_t^{mid} being the midpoint, and p_t^{bid} and p_t^{ask} representing the best bid and ask prices at the top of the order book at time t. Unlike the transaction price, the midpoint is continuously observable, as the best bid and ask prices remain valid until the corresponding limit orders are executed or canceled. Additionally, the midpoint updates with every change in the best bid or ask, rather than only when a trade occurs, making it a more dynamic and responsive price measure. It also avoids strictly bouncing between the bid and ask prices, providing a smoother representation of price movements. However, it is also subject to tick-size constraints, as the minimum changes in the best bid and ask prices are restricted by the asset's tick size. As a result, the smallest possible increment of the midpoint is half of the asset's tick size.

While the midpoint assumes that the true value lies symmetrically between the best bid and ask prices, this assumption may not always hold in practice. Asymmetric quoting by liquidity providers, driven by discrete price levels (Anshuman and Kalay, 1998) or inventory risks (Hendershott and Menkveld, 2014), can distort this approximation. Additionally, the midpoint, like the transaction price, only reflects the current state of the market and does not incorporate expectations of future price movements (Hasbrouck, 2002; Hagströmer, 2021).

The *quantity-weighted (QW) midpoint* addresses some of the shortcomings of the midpoint. Instead of weighting the bid and ask prices equally, the QW midpoint adjusts the weighting of the best bid and ask prices based on their respective volumes. The QW midpoint is defined as

$$p_t^{qw} = \frac{q_t^{ask} p_t^{bid} + q_t^{bid} p_t^{ask}}{q_t^{bid} + q_t^{ask}}, \tag{3}$$

where p_t^{qw} is the quantity-weighted midpoint at time t. Moreover, q_t^{ask} and q_t^{bid} represent the number of shares available at the best ask and bid prices, respectively. By incorporating the quantity imbalance at the top of the order book, it provides insight into potential future price movements, aiming to offer a more accurate approximation of an asset's true value.

This approach is supported by both theoretical and empirical literature. Glosten (1994) demonstrates that an asset's true value is closer to the bid price when the quantity at the best bid is significantly lower than the quantity at the best ask, and vice versa. Theoretical research suggests that the informational value of order book imbalance can arise from two factors. First, imbalances can result from uninformed, liquidity-driven traders, whose orders become more aggressive if their limit order remain unexecuted, thereby signaling future price movements (Harris, 1990). Second, imbalances can arise from (better) informed investors submitting limit orders trying to minimize implicit transaction costs, partially revealing their information in the order book (Kaniel and Liu, 2006; Riccó et al., 2020). Several empirical studies confirm the predictive power of order book imbalances, emphasizing the ability of the QW midpoint to reflect future price dynamics (Cao et al., 2009; Rösch et al., 2017; Aït-Sahalia et al., 2022). Furthermore, unlike the transaction price and midpoint, the QW midpoint is not constrained by tick size, allowing it to take on a continuous range of values, which better aligns with the assumptions about an asset's true value.

However, the QW midpoint also has its drawbacks. For instance, its adjustment of the midpoint can become extreme in the case of large order book imbalances, particularly when the bid-ask spread is large. Furthermore, some changes in the price measure do not align with fundamental economic principles (Stoikov, 2018). For instance, an increase in demand, through the submission of a buy-order that improves the current best bid, can result in price measure's decrease.[1]

Overcoming these issues regarding the QW midpoint, practitioners further rely on the *constrained quantity-weighted (CQW) midpoint*, which adjusts the midpoint in a more conservative manner. The price measure is defined as

$$p_t^{cqw} = p_t^{mid} + 0.5ts \, \frac{q_t^{bid} - q_t^{ask}}{q_t^{bid} + q_t^{ask}}, \tag{4}$$

where p_t^{cqw} is the CQW midpoint at t and ts is the tick size of the asset. In contrast to the QW midpoint, the correction is capped at half the tick size, ensuring that the measure remains a reasonable approximation of the true value, even under extreme imbalances and spreads larger than one tick. This property makes the CQW midpoint particularly robust in markets with higher volatility or less liquidity. Furthermore, the CQW midpoint is also continuous variable aligning with the characteristics of the true value.

While the CQW midpoint overcomes or reduces some of the issues from the QW midpoint, it still has its limitations. Stoikov (2018) challenges the use of the QW (and implicitly CQW) midpoint as true value approximators, citing their susceptibility to noise and the lack of theoretical justification, as it does not necessarily behave as a

[1]For instance, assume the current best bid is \$98 with a corresponding quantity of 90 shares, while the current ask is \$102 with a corresponding quantity of 10 shares. The resulting QW midpoint is \$101.60. Now, a trader submits a buy at \$99 and a quantity of 10 shares, improving the best bid. Despite a demand increase, the QW midpoint decreases to \$100.50.

martingale. As alternative price measure, Stoikov (2018) proposes the *micro-price*, which incorporates expected future price movements and is defined as

$$p_t^{micro} = p_t^{mid} + f\left(p_t^{ask} - p_t^{bid}, \frac{q_t^{bid}}{q_t^{bid} + q_t^{ask}}\right), \tag{5}$$

where p_t^{micro} is the micro-price at time t, which is a function f of the current bid-ask spread and the order book imbalance. This function is estimated based on historical order book data to reflect expected future price movements of the midpoint. This makes the micro-price a martingale by construction, incorporating the expected movement of the future midpoint given current public information, here the bid-ask spread and order book imbalance. While the micro-price offers a sophisticated and dynamic view of the market, the estimation of the adjustment function f requires significant computational effort.

To estimate the micro-price, we draw on Stoikov's (2018) guidelines. In this study, the function f is estimated using a rolling window of the previous 10 trading days individually for each asset. Like Stoikov (2018), we discretize the states of the order book imbalance based on the deciles in the historical data. The bid-ask spread and future midpoint changes are measured in ticks. We consider only those states of spread and midpoint changes that account for at least 1% of all observations during the past 10 trading days. If there are tick-size changes within these days, we exclude the present asset-day from the analysis, because the micro-price cannot be reliably estimated.[2] In total, we exclude 12.4% of all asset-days from our sample, because of this issue. The estimation of the function is based on order book data sampled on a 1s-frequency. For detailed information on the estimation of the micro-price, we refer readers to the original paper by Stoikov (2018).

Figure 1 illustrates the behavior of all five price measures over a very short time period of approximately 6ms on January 5, 2023, for the stock BMW. While the dotted grey lines represent the best bid and ask prices, the colored lines depict the different price measures. Observations within the measure are marked with symbols such as squares or circles. The figure shows that the transaction price is measured three times within the time period and observed exactly at the best bid or ask price. By contrast, measures derived from order book data (midpoint, QW midpoint, CQW midpoint, and micro-price) are observed each time the order book

[2]There are multiple ways to circumvent the problem with changing tick sizes, like discretizing states based on relative spreads. However, we decided to follow the original definition of Stoikov (2018).

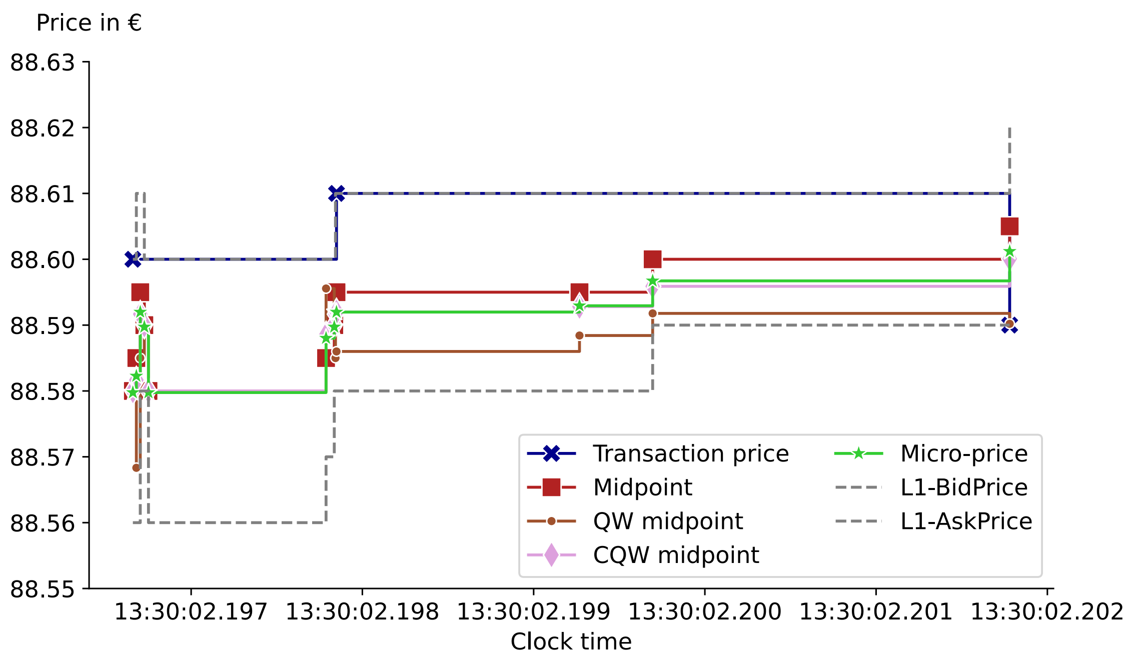


Figure 1: Exemplary lineplot showing a short intraday excerpt of all five used price measures
This figure shows a short time frame of intraday data for the BMW stock on the 5th January 2023 covering roughly 6ms. All five price measures used in this study are shown in colored lines. The best bid and ask prices are plotted in dashed gray lines.

changes. As a result, they are updated more frequently and fluctuate within the bid-ask range. Moreover, the figure demonstrates that the QW midpoint can skew closer to the best bid or ask under significant imbalances, while the CQW midpoint and the micro-price provide a more moderated adjustment.

The list of price measures considered in this paper is not exhaustive. The literature includes other true value estimators, such as the price measure proposed by Bonart and Lillo (2018), which incorporates information about exchange pricing models, like liquidity rebates. However, to maintain a manageable scope, we focus on the selected set of five price measures.

2.2. Data

To analyze the efficiency of different price measures, we rely on a dataset sourced from the Deutsche Börse's A7 Analytics Platform. The dataset comprises order book and trade data spanning a six-month period from January to June 2023, covering all constituents of the DAX40 during that time period. The DAX40 represents the 40 largest German companies by market capitalization, all traded on Germany's main stock exchange Xetra. In total, the dataset encompasses 42 stocks, as it accounts for both index leavers and joiners over the full sample period.[3]

In the order book data, each observation corresponds to the current state of the limit order book (LOB). Alongside a timestamp with nanosecond precision, each LOB state includes the price and quantity for the ten best price levels on both the bid and ask sides. A new observation is generated whenever there is a change in the price or quantity at any of these ten best price levels. In addition to the DAX40 stocks, the dataset incorporates order book and trade data from the DAX40 futures, which are traded in the same period on Eurex. For each trading day, we select the futures contract with the highest trading activity, as measured by the number of trades.

The trading hours on Xetra are from 08:00 to 16:30 UTC. Our analysis focuses exclusively on continuous trading phases, excluding periods influenced by auctions. To minimize the impact of auctions on liquidity, we exclude the 15 minutes before and after each scheduled auction. Xetra conducts three scheduled auctions: the opening auction (07:50 to 08:02 UTC), the midday auction (12:00 to 12:02 UTC), and the closing auction (16:30 to 16:35 UTC). Consistent with this schedule, each asset-day in our sample consists of two trading sessions: the morning session, spanning 08:17 to 11:45 UTC, and the afternoon session, from 12:17 to 16:15 UTC. Taking both sessions together, they constitute one asset-day in our filtered data sample.

Table A.1 in the Appendix summarizes key descriptive statistics on daily trading and order book activity for all stocks in our dataset. On average, each stock has a trading volume of approximately €40 million, with around 4,700 trades across the two trading sessions, i.e., per filtered trading day. Order book activity is notably high, with nearly 340,000 updates per asset-day, reflecting the high liquidity and activity in these markets.

To better understand the variation of the five price measures during a trading day, Table 1 provides descriptive statistics on the frequency of changes for each measure. The transaction price is the most rigid measure of all price measures. Its value changes roughly 1,300 times per asset-day, or once every 20.58 seconds on average. By contrast, the other measures, which can update with every change in the order book, exhibit significantly higher update rates. The midpoint changes the least frequently among these order book measures, as it depends only on the best bid and ask prices, changing approximately every 1.72 seconds. In contrast, the

[3]The only exception to this is Linde PLC. The company's stock is included in our sample only until February 28, as it is traded exclusively on the NYSE since then and is therefore no longer available on Xetra.

Table 1: Descriptive statistics regarding the number of non-zero changes in each price measure
The table shows average daily descriptive statistics for each of the used price measures. The first column shows the number of non-zero changes of the price measure in thousands per trading day. The interchange duration represents the average time it takes until the next non-zero change occurs. This time is expressed in seconds. To obtain the values in the last two columns, we sample each price measure at a frequency of one-minute and one-second, respectively. The proportion of changes is the number of non-zero changes in that time series divided by the total number of observations.

	No. changes per day (in thousands)	Interchange duration (in seconds)	Share non-zero changes (freq. = 1min)	Share non-zero changes (freq. = 1s)
Trans. price	1.30	20.58	0.67	0.04
Midpoint	15.52	1.72	0.81	0.14
QW midpoint	165.48	0.16	0.99	0.47
CQW midpoint	165.19	0.16	0.99	0.47
Micro-price	86.60	0.31	0.96	0.35

QW and CQW midpoint change the most frequently as both measures continuously adjust to changes in both the best prices or corresponding quantities. On average, these measures change approximately 165,000 times during a trading day. Although the micro-price depends on three factors, prices, quantities, and bid-ask spread, it changes less frequently than the QW and CQW midpoint with approximately 86,000 changes per trading day. This arises from the discretization of order book imbalance and spread, which prevents the micro-price from responding to marginal input changes.

Even though the price measures change multiple times per minute, their variability is limited when sampled at similar but fixed time intervals. As shown in the last two columns of Table 1, the proportion of non-zero returns for each price measure varies significantly when sampled at one-minute and one-second intervals, respectively. At lower frequencies (one-minute intervals), variability is sufficient for empirical analysis: 67% of transaction price returns and 81% of midpoint returns are non-zero, while more sophisticated measures achieve ratios of 96% or higher. However, at higher sampling frequencies (one-second intervals), zero returns dominate across all measures, with ratios dropping below 50%. The midpoint and transaction price, in particular, show non-zero return ratios of 14% and 4%, respectively. This high prevalence of zero returns arises from asynchronous trading, where price measure changes cluster within short bursts of trading activity, followed by periods of inactivity. This zero-return inflation poses significant challenges for conventional price efficiency tests, as discussed in the next section.

3. Evaluation of price efficiency and its determinants

To systematically evaluate and compare price efficiency across different price measures, we propose an empirical approach to assess return predictability for each price measure. According to the semi-strong market efficiency hypothesis, efficient prices should fully reflect all available public information (Fama, 1970). Thus, the predictability of returns can be interpreted as an inverse measure of price efficiency (Hasbrouck, 1993; Hou and Moskowitz, 2005). Common tests leverage this principle by examining whether price series follow a random walk (e.g. Lo and MacKinlay, 1989), exhibit no serial autocorrelation (e.g. Hendershott and Jones, 2005) or are unpredictable based on public information (Rösch et al., 2017).

However, these traditional methods require sufficient variation in returns to provide meaningful results. When increasing the sampling frequency, efficiency metrics such as the variance ratio and autocorrelation of returns diminish toward zero, primarily due to the effects of asynchronous trading and the dominance of microstructural noise. Consequently, these methods are unsuitable in modern high-frequency trading environments, where trading speed is high, but price variation is low when observed at a high frequency. This is especially true for conventional price measures such as the transaction price and midpoint, which result in zero-inflated return time series when sampled at high frequencies, as demonstrated in Table 1. This highlights the necessity for more robust efficiency tests that account for the characteristics of high-frequency financial data, enabling more accurate analysis and modeling in such contexts.

To address this issue, we evaluate the predictability of price measures using (i) multiple prediction horizons and based on (ii) independent variables that represent publicly available information. Specifically, we predict the returns of each price measure from time t to $t+h$, where h is equal to one of the following prediction horizons: $h = \{$100ms, 1s, 2s, 5s, 15s, 30s, 1min, 2.5min, 5min$\}$. By incorporating varying prediction horizons, we capture the reaction of each price measure over different time intervals and ensure that more rigid measures such as transaction prices have enough time to react, mitigating the effect of irregular trading.

The lagged variables in our prediction model capture prior changes in both stock-specific and market-wide information, measured over progressively larger, non-overlapping time intervals. The changes in the independent variables are measured using the corresponding intervals as the prediction horizons. Specifically, the lagged changes are measured between: $L = \{(t-100ms, t],\ (t-1s, t-100ms],\ (t-2s, t-1s],\ ..., (t-2.5min, t-1min],\ (t-5min, t-2.5min]\}$. This means that the largest model in our sample includes $k = 9$ lagged values of each indepdendent variable. The non-overlapping and increasing time intervals ensure that we capture the high- and low-frequency changes in the independent variables and avoid multicollinearity issues.

Regarding the regressors, we focus on three variables that have been shown to carry relevant information with respect to an asset's future price movements: the market return (r^m), asset's own return (r^s), and changes in the imbalance at the top of the assets' order book (Δib^s). As demonstrated, for example, by Hou and Moskowitz (2005), the market return is often a strong predictor of an asset's future price movements as it represents market-wide information and systematic risk relevant for all assets. Therefore, we include lagged market returns in our prediction model, specifically using returns from the DAX40 futures. As mentioned before, we use the futures contract with the largest number of trades on the respective trading day. To address potential autocorrelation, we also include past returns of the asset itself to examine whether historical asset-level information is fully reflected in the price. Both market and asset returns are calculated based on the logarithmic returns of the corresponding micro-price over a specific time interval. We use the micro-price for return calculation because it is theoretically the most informative price measure. It reflects key information, including prices and quantities at the best bid and ask sides, as well as the bid-ask spread. Furthermore, it is a martingale by construction (Stoikov, 2018). Lastly, as shown by Cao et al. (2009), Cont et al. (2014), and other researchers, the order book imbalance at the best bid and ask prices provides crucial information about future price movements. Translating order book imbalance into an interval-based measure, we define it as the sum of volume changes at the best bid minus those at the best ask levels over the lagged time interval.

To summarize, we propose the following OLS regression model to assess the predictability of future returns across different price measures:

$$r_{t+h} = \beta_0 + \sum_{j=1}^{k} \beta_{1,j}\ r^m_{L_j} + \beta_{2,j}\ r^s_{L_j} + \beta_{3,j}\ \Delta ib^s_{L_j} + \epsilon_{t+h} \tag{6}$$

This model is estimated for each price measure m, each prediction horizon h, each asset s, and each trading day td. This results in 45 model estimations per asset-day.[4] For each asset-day, we sample the dependent and independent variables at a 100-ms frequency and calculate the changes over longer horizons and lags by aggregating the respective changes. In the regression model, r_{t+h} represents the asset return calculated using the log difference of price measure m between time $t+h$ and t. The independent variables include all $k = 9$ lagged values of $r^m_{L_j}, r^s_{L_j}, \Delta ib^s_{L_j}$.

The models are then used to predict the returns of the price measures out-of-sample on the subsequent trading day. To assess the predictability of the price measure, we compare the resulting mean squared error (MSE) of the proposed model, to the MSE of a no-change benchmark that constantly predicts a zero return. This comparison is appropriate both theoretically and empirically, as the expected change in the efficient price is zero, and most high-frequency returns are indeed zero.

We define the ratio of predictability, $\bar{e}_{m,h,k}$, for a given price measure m, prediction horizon h, and lag length k, as follows:

$$\bar{e}_{m,h,k} = \frac{1}{|TD| \cdot |S|} \cdot \sum_{td \in TD} \sum_{s \in S} e_{m,h,k,s,td} \quad , \qquad \text{with } e_{m,h,k,s,td} = \frac{MSE^{nc}_{m,h,k,s,td}}{MSE^{ols}_{m,h,k,s,td}} \tag{7}$$

Here, TD represents all trading days and S represents all securities in the sample. $MSE^{ols}_{m,h,k,s,td}$ denotes the MSE of the OLS regression model as described in (6), given price measure m, the security s, trading day td, prediction horizon h, and lag length k. Similarly, $MSE^{nc}_{m,h,k,s,td}$ is the MSE of the no-change benchmark. The metric $\bar{e}_{m,h,k}$ thus provides the average predictability of a price measure, compared to the no-change benchmark.

If $\bar{e}_{m,h,k} > 1$, it implies that the return from t to $t+h$ for price measure m is predictable to some extend using k lagged values of public information. This indicates that prior information is not fully incorporated into price measure m at time t, and all or part of this information is incorporated within the prediction horizon h.

To further evaluate which intervals of past information are reflected in the price measure at time t, we iteratively estimate the regression model in Equation 6 while progressively including additional lags of past information. This means that we estimate all 45 regression models per asset-day using all possible lag lengths with $k = 1, 2, ..., 9$. Simply speaking, in the first iteration the regression model only includes lagged values of the independent variables capturing changes between t-100ms and t, while the next iteration of regression models incorporates changes in the regressors

[4] We have $|m| = 5$ different price measures and $|h| = 9$ prediction horizons, resulting in 45 regressions per asset-day.

between t-100ms and t and t-1s and t-100ms. In this way, we progressively increase the time interval of previous information until the regression models capture all prior information up to the last five minutes ($k = 9$). In each iteration, we calculate the gain in predictability achieved by incorporating the additional lag compared to the best-performing model with fewer lags. If the inclusion of additional past information improves predictability, it implies that this information was not fully reflected in the price measure at time t.

The gain in predictability is then defined as

$$\bar{g}_{m,h,k} = \bar{e}_{m,h,k} - \bar{e}_{m,h,k^*}, \tag{8}$$

where k^* is the lag where the ratio of predictability is locally maximized for $k^* < k$. Thus, if $\bar{g}_{m,h,k} > 0$, it indicates that public information from lag k was not fully incorporated in price measure m at time t, and that this information gets subsequently incorporated within the prediction horizon h.

Finally, we test whether the ratio of predictability ($\bar{e}_{m,h,k}$) and the gain in predictability ($\bar{g}_{m,h,k}$) are statistically significantly greater than 1 and 0, respectively. To account for multiple hypothesis testing, we apply the Bonferroni correction, setting the adjusted significance level as $\alpha_{corr} = \alpha/n$, where n is the number of related hypotheses. In our case, we set $n = 9$ for both tests: assessing whether the ratio of predictability is significantly greater than 1 and evaluating whether the gain in predictability is significantly greater than 0, to account for the multiple prediction horizons and different lag lengths.

After analyzing the ability and speed of the price measures in reflecting past information, we finally aim to understand what the determinants of their (in)efficiencies are. The literature has a clear yet multifaceted answer to this: market frictions. Market frictions refer to factors that prevent financial markets from operating perfectly efficiently and result in prices that do not fully reflect all available information. In market microstructure, common types of frictions include: information asymmetry, inventory risk, transaction costs, order execution delays, tick size, and other regulatory constraints (Hou and Moskowitz, 2005). Depending on the examined time interval, these frictions can have different impacts on price efficiency. For instance, frictions, such as delays caused by latency, might prevent price measures from incorporating information for a few milli- or microseconds, but become negligible over longer time frames.

As our study focuses on the efficiency of price measures at the high-frequency level, we assess the impact of market frictions that potentially reduce the ability of price measures to reflect information in the short term. According to theory, one of these central frictions is information asymmetry, meaning that some traders possess (private) information about an asset that is not yet reflected in market prices (Kyle, 1985). These informed investors will submit buy orders when prices are lower than the true value and sell orders when prices are higher. This leads to a one-sided order flow and, consequently, an imbalance in the order book (Chordia et al., 2008). Thus, we hypothesize that the stronger the order book imbalance, the more predictable the returns of price measures will be in the short term, because the prediction model can observe informed signals implicitly through the imbalance.[5] Moreover, the price measures are affected by the activity in the limit order book. Asynchronous trading and insufficient activity in the order book can lead to rigid price measures that do not adequately update in the short term. While the transaction price is only updated through executions, other price measures can be updated via passive order flow tackling or changing the best bid or ask price. Consequently, we presume that both the number of trades and the number of order book updates are negatively correlated with the predictability of price measures' returns. Another central market friction that is relevant at the high-frequency level is the tick size as it defines the minimum profit margin for liquidity providers. The higher the tick size (relative to the asset price), the stronger are price movements, constrained. Consequently, we hypothesize that the larger the (relative) tick size of an asset, the better the predictability of price measures' returns. Several studies have demonstrated that the tick size is negatively associated with price efficiency (Chordia et al., 2008; Chung and Chuwonganant, 2023). However, the majority of studies have focused on exogenous changes of the absolute tick size rather than considering the continuous variation of the relative tick size of an asset.

To empirically test which frictions impact the efficiency of price measures, we estimate panel regressions to explain the predictability ratios for each asset s and trading day td ($e_{m,h,k,s,td}$ in Equation 7). To reduce dimensionality, we focus on a specific prediction horizon and lags for the predictability ratio. We only consider the predictability ratios for $h = 5s$ and $k = 5$, i.e., we include five lagged values measuring changes in the independent variables within the last 15 seconds, which is a compromise between short- and long-term predictability. Specifically, we estimate a pooled panel regression explaining these predictability ratios for all price measures. In addition, we estimate an individual regression for each price measure

[5] As mentioned before, the imbalances are not necessarily the result of informed flow. They can also arise from uninformed, liquidity-driven trading. Nevertheless, this order flow also provides information about future price movements as their orders become more aggressive if their liquidity demand is not fulfilled (Harris, 1990).

and their corresponding predictability ratios. Thereby, we try to investigate whether the determinants of predictability differ across the price measures.

As suggested by theory, we explain the predictability ratios using (1) the absolute order book imbalance being the absolute difference between volumes at the best ask and bid, scaled by their sum, (2) the number of transactions, (3) the number of limit order book updates, and (4) the relative tick size, which is the tick size divided by the average midpoint. As control variables, we include the asset volatility, calculated using 5-min CQW midpoint returns, and the order-to-trade ratio, which is the number of limit order book updates divided by the number of trades. All independent variables represent daily averages for asset s on trading day td, the model's evaluation day. Only for volatility, we consider both the average on the evaluation day and the average on the training day. Specifically, we include the absolute value of the relative difference between the volatility on the evaluation day and estimation day in percentage points. Thereby, we aim to consider differences between the asset-days where the model is estimated and evaluated, as large differences may decrease the model's generalization ability, resulting in a decrease in out-of-sample prediction accuracy. To focus on specific market frictions, we control for market-wide developments and constant asset attributes by using time- and asset-fixed effects. While time-fixed effects control for factors such as overall market volatility, asset-fixed effects carry information on constant or more rigid asset attributes like industry or market capitalization, which varies little over a few months for most stocks. We estimate the following regression model $|m|+1$ times to explain the predictability ratios for each price measure m and use a pooled regression to capture overall trends.

$$e_{s,td} = \gamma_0 + \gamma_1 x_{1,s,td} + \gamma_2 x_{2,s,td} + ... + \gamma_7 x_{7,s,td} + \nu_s + \nu_{td} + \varepsilon_{s,td} \tag{9}$$

The independent variables $x_{1,s,td}, x_{2,s,td}, ..., x_{7,s,td}$ are the aforementioned variables, measured as daily averages for each asset s on trading day td. ν_s and ν_{td} represent stock and time fixed effects, respectively. Note that we also include $|m|-1$ dummy variables in the pooled regression to account for the corresponding price measure and control for general differences in their predictability. Coefficients that are statistically significantly greater than zero in the regression indicate that the corresponding market friction increases return predictability, meaning it reduces the efficiency of the price measure.

4. Results

In this section, we apply the empirical approach proposed in Section 3 to assess the predictability of each price measure. We tackle our first two research questions

"How do different price measures compare in their speed of (fully) reflecting past information?" and "Does the ability to reflect information depend on the extent of market frictions, and do the key drivers differ across various price measures?" in Section 4.1. In Section 4.2, we shed light into the third research question "To what extent can the application of inefficient price measures impact researchers' study results and investors' trading outcomes?".

4.1. Efficiency of different price measures

To explore how different price measures vary in their ability to efficiently reflect past information, we first examine their general predictability using all lagged values of market returns, the asset's return, and the asset's order book imbalance. To quantify predictability, we compute the ratio of predictability $\bar{e}_{m,h,k=9}$, as defined in Equation 7, for each price measure m, across all prediction horizons h considering all lags of independent variables ($k = 9$), which capture changes in public information over the previous 5 minutes.[6] A value of $\bar{e}_{m,h,k=9} > 1$ indicates that price measure m is on average better predictable using past information compared to the benchmark, indicating that at least some of the changes in public information within the last five minutes remain unreflected at observation time t. We test the following hypothesis:

$$H_0 : \bar{e}_{m,h,k=9} \leq 1, H_1 : \bar{e}_{m,h,k=9} > 1 \tag{10}$$

Rejection of H_0 implies inefficiency in the price measure, as future returns are on average predictable using past information.

The ratio of predictability and the corresponding test results are visualized in Figure 2. The figure demonstrates for each price measure and prediction horizon, displayed on the horizontal axis, the average ratio of predictability across all asset-days. If the ratio is statistically significantly larger than 1 at the corrected 5% significance level, the data point is marked by a star. Otherwise, it is represented by a dot. The test results presented in Figure 2 provide evidence for significant predictability across all price measures. This finding underscores the notion that none of the examined price measures are perfectly efficient. However, the extent of predictability and its temporal characteristics differ substantially among the measures. For example, the transaction price demonstrates significant predictability up to a one-minute horizon, while also exhibiting the highest predictability ratio ($\bar{e}_{p^{tp},15s,k=9} = 1.033$) among all measures. This result implies that, on average, changes in public infor-

[6] We further report the results for all other lag lengths in Figure B.1 and Figure B.2 in the Appendix.

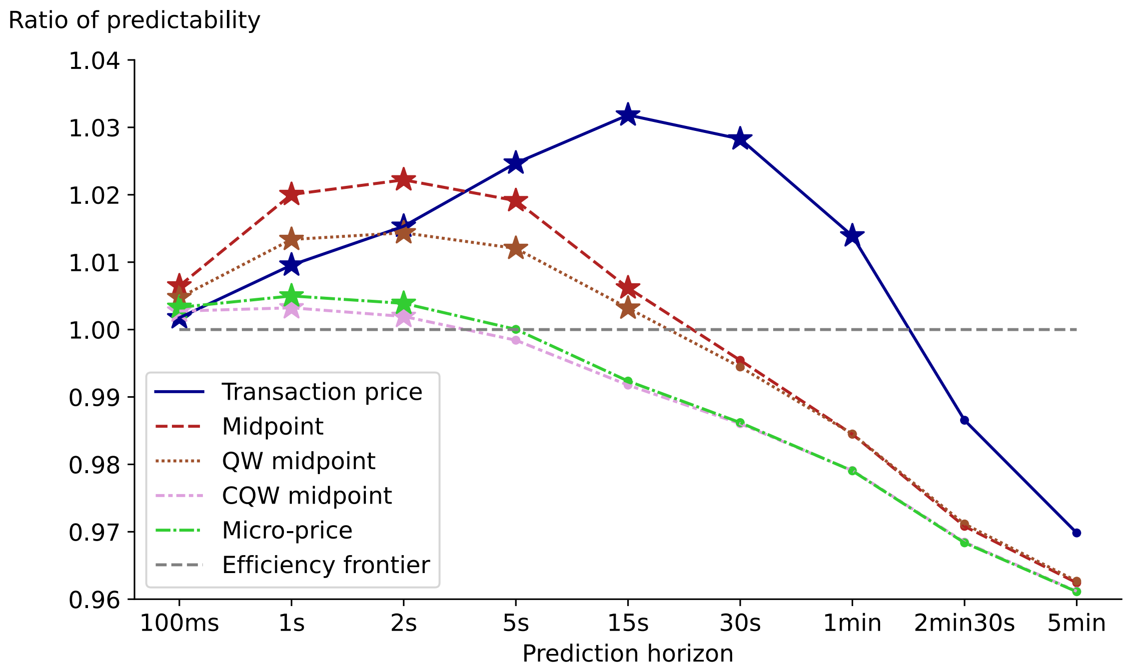


Figure 2: Ratio of predictability of all five price measures at various prediction horizons given five minutes of past information

In this figure a star indicates that the ratio of predictability is statistically significantly larger than 1 at the corrected 5% significance level after applying the Bonferroni correction.

mation within the previous five minutes provide predictive power that reduces the model's prediction error compared to the no-change benchmark. On average, the no-change benchmark predicts with a mean-squared error that is 3.3% greater than the prediction error of the regression model.

In contrast to the transaction price, the CQW midpoint is significantly predictable only for up to a two-second horizon, with a predictability ratio close to 1 ($\bar{e}_{p^{cqw},1s,k=9} = 1.003$), suggesting higher efficiency. Similarly, the micro-price has a predictability ratio comparable to that of the CQW midpoint ($\bar{e}_{p^{mp},1s,k=9} = 1.005$). Furthermore, the midpoint and QW midpoint are significantly predictable for up to 15-second horizons, indicating that they need more time than the CQW midpoint or the micro-price to reflect public information. Additionally, the QW midpoint has a lower ratio of predictability than the midpoint across all horizons, highlighting efficiency improvements from incorporating order book imbalance information. In general, these findings highlight that while all price measures exhibit some degree of inefficiency, their ability to incorporate past information varies considerably. We find that the transaction price is the most predictable price measure, followed by the midpoint and then by the QW midpoint. The micro-price and CQW midpoint are the most efficient price measures, as their returns are only marginally more predictable based on prior information.

Building on this, we analyze the speed at which past information is reflected in each price measure. Specifically, we compute the gain in predictability, $\bar{g}_{m,h,k}$, as defined in Equation 8, which captures the additional predictive power obtained by incorporating the additional lag k. A significant positive value of $\bar{g}_{m,h,k}$ indicates that lagged information becoming public during interval k is not fully incorporated at time t, but is (at least partially) reflected in the price measure between t and $t+h$. This motivates the following hypothesis:

$$H_0 : \bar{g}_{m,h,k} \leq 0, H_1 : \bar{g}_{m,h,k} > 0 \tag{11}$$

Rejection of H_0 implies that lag k contributes to the predictability of future returns, indicating that the price measure m does not fully reflect this information.

Figure 3 demonstrates the results of this analysis for m being the transaction price. The figure shows the gain in predictability (vertical axis) for each additional lag (horizontal axis) and each prediction horizon (indicated by the different bars). A colored bar indicates a statistically significant gain in predictability at the corrected 5% significance level. Moreover, bold x-axis labels suggest a significant gain in the predictability ratio for at least one prediction horizon.

The analysis reveals that lagged information up to 30 seconds prior to t provides statistically significant predictive power for future transaction price returns. For a 15-second prediction horizon, the gain in predictability ranges from 0.38 percentage points for the interval $(t - 30s, t - 15s]$ to 1.02 percentage points for the interval $(t - 15s, t - 5s]$. However, information from intervals beyond 30 seconds do not significantly enhance predictability. This indicates that it takes more than 30 seconds but less than one minute for the transaction price to incorporate public information. Moreover, this suggests that time series of transaction price returns sampled at frequencies of one minute or lower yield, on average, an informationally efficient series of returns. Hence, researchers can utilize this price measure when not analyzing high-frequency dynamics.

The gains in predictability for the remaining price measures are displayed in Figure 4. The figure suggests that the patterns of predictability gains substantially differ across the price measures. For the midpoint, lagged intervals up to one minute contribute significantly to return predictability over short horizons, such as 100ms, 1s, and 2s. However, these gains are small in magnitude, with for example $\bar{g}_{p^{mp},5s,k=6} =$

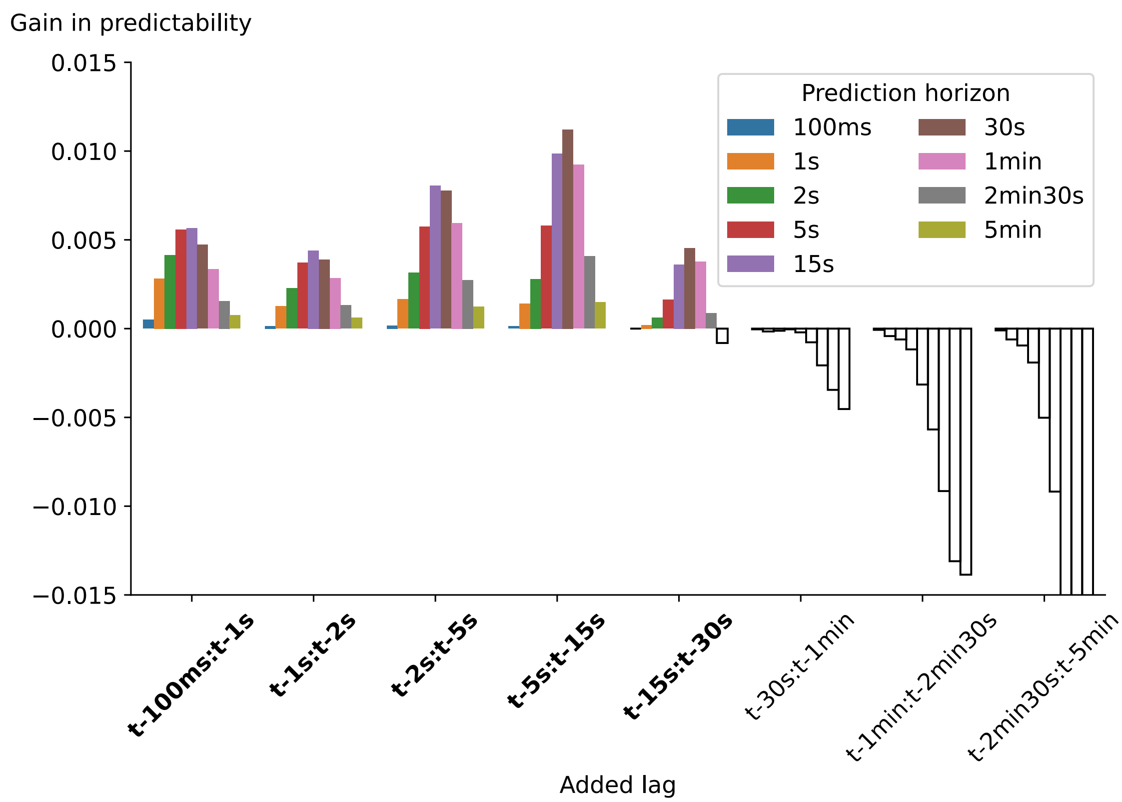


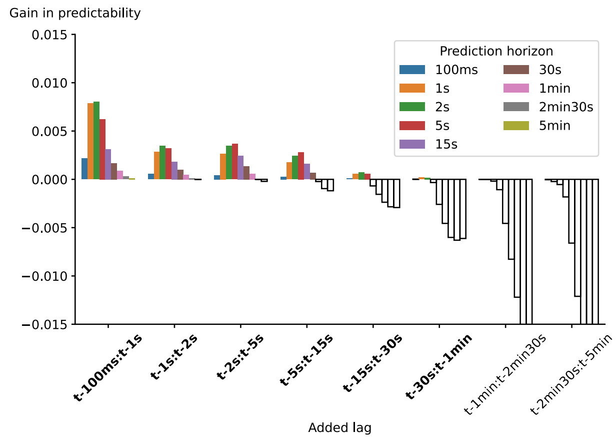
Figure 3: Gain in predictability for the transaction price with varying levels of past information
In this figure the gain in predictability for the transaction price for various prediction horizons is shown. The x-axis shows the level of past information included. A colored bar indicates that the gain in predictability is statistically significantly larger than 0 at the corrected 5% significance level. If the x-axis label is bold, we find a significant gain in the ratio of predictability for at least one prediction horizon.

0.00058, underscoring a rather marginal gain in predictability for these intervals relative to the transaction price. Furthermore, the QW midpoint exhibits significant predictability gains for information from intervals capturing the previous 30 seconds. For example, the interval $(t-30s, t-15s]$ contributes a marginal gain of $\bar{g}_{p^{qw},2s,k=6} = 0.00019$.

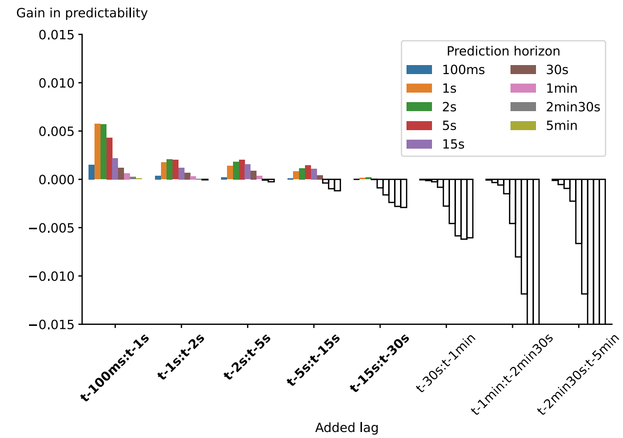
The micro-price, by contrast, shows a markedly different pattern. Gains in predictability are only significant for information up to two seconds before t, regardless of the prediction horizon. This suggests that the micro-price fully reflects past information within two seconds. Moreover, the magnitude of its predictability gains is notably smaller than those of transaction price, midpoint, and QW midpoint, affirming its higher efficiency. These findings align with theoretical implications (e.g. Stoikov, 2018; Hagströmer, 2021) that the micro-price serves as a robust proxy for an asset's true value in high-frequency trading environments.

Interestingly, the CQW midpoint exhibits efficiency characteristics similar to those of the micro-price. Despite being less sophisticated and resource-intensive, the CQW midpoint incorporates past information also within two seconds. In addition, the price measure has the lowest predictability gains of all analyzed measures. This efficiency likely arises from its design, which adjusts the midpoint for volume imbalances at the top of the order book while constraining adjustments to half a tick size. Unlike the micro-price, the CQW midpoint does not directly account for expected future price movements or incorporate spread information. However, in liquid markets such as German blue-chip stocks, where the spread is often near the tick size, the CQW midpoint indirectly reflects this information, resulting in an efficiency level similar to, and even marginally higher than, the efficiency of the micro-price. However, it is worth noting that the results might change when analyzing less liquid financial assets. Overall, the results highlight that different price measures vary significantly in how quickly they reflect past information. While the transaction price, midpoint and QW midpoint take a longer period to incorporate information, the micro-price and CQW midpoint adjust within a much shorter time frame, making them the most efficient.

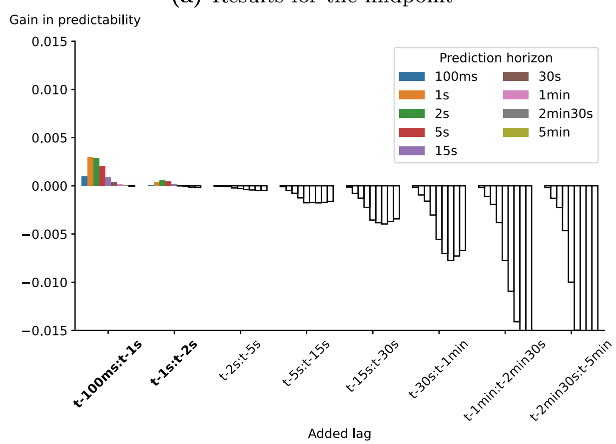
Our results clearly show that even the returns of more complex price measures are predictable over time horizons that exceed the reaction speeds of institutional investors, particularly high-frequency traders. Thus, the question remains: what prevents price measures from incorporating past information or, put differently, institutional investors from leveraging this predictability? A key reason is the presence of market frictions that impede or slow down the incorporation of information into price measures. Factors such as asymmetric information, transaction costs, order book constraints, regulatory limitations, or risk management considerations can create barriers preventing institutional investors from fully exploiting predictable patterns.



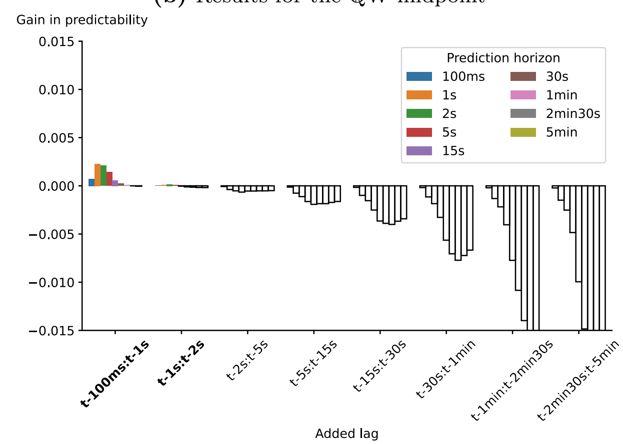
(a) Results for the midpoint



(b) Results for the QW midpoint



(c) Results for the micro-price



(d) Results for the CQW midpoint

Figure 4: Gain in predictability for four price measures with varying levels of past information

In this figure the gain in predictability for (a) the midpoint, (b) the QW midpoint, (c) the micro-price, and (d) the CQW midpoint for various prediction horizons is shown. The x-axis shows the level of past information included. A colored bar indicates that the gain in predictability is statistically significantly larger than 0 at the corrected 5% significance level. If the x-axis label is bold, we find a significant gain in the ratio of predictability for at least one prediction horizon.

As mentioned in the methodology section, we estimate several regression models to examine the impact of market frictions on price measures' inefficiencies. Specifically, we estimate both a pooled panel regression model and separate panel regression models for each price measure to explain the ratios of predictability both collectively and individually. To reduce dimensionality, we only consider the predictability ratios for $h = 5s$ and $k = 5$, as these values represent a combination of prediction horizon and included lags for which all measures exhibit significant predictability, as shown in Figure B.2 in the Appendix.

The results of these regression models are displayed in Table 2. The coefficients in the second column refer to the pooled regression, while the subsequent columns refer to the individual regression for each price measure.

In general, larger coefficients indicate greater return predictability and thus lower efficiency. The dependent variable is scaled by a factor of 100 to enhance the readability of the coefficients. The results show that absolute order book imbalance is significantly positively correlated with predictability, supporting the hypothesis that a higher share of passive informed order flow is associated with larger deviations from the true value, implying short-term return predictability. This passive informed order flow is directly observable through the order book imbalance allowing the predictability of future price movements. Interestingly, this is true for all price measures, including those explicitly considering the order book imbalance, except for the micro-price. This result indicates that the micro-price is the most effective at incorporating order book imbalance information, potentially because it also takes into account the bid-ask spread and does not change with each update in the order book imbalance due to its discretization. On the other hand, the QW midpoint exhibits the largest coefficient for the order book imbalance, which is more than twice as large as the coefficient in the model including all price measures. The QW midpoint is therefore increasingly inefficient when a high order book imbalance exists. This is likely due to the fact that the QW midpoint is too close to the best bid and ask prices when the imbalance and bid-ask spread are high, causing it to deviate substantially from the true value.

Table 2: Determinants of price measures' predictability

This table presents panel regression estimates for the determinants of predictability. The dependent variable is the ratio of predictability, measured as the ratio of the model vs. benchmark prediction performance per asset-day multiplied by 100 for scaling. A higher ratio indicates superior model performance relative to the benchmark. We examine predictability using various market friction indicators. *Relative tick size* is the tick size divided by the asset-day's average midpoint. *Abs. imbalance* is the absolute difference between volumes at the best ask and bid, scaled by their sum. *No. trades* and *No. LOB updates* represent the number of trades and order book changes per asset-day in thousands and millions, respectively. *Order-to-trade ratio* is their reciprocal. *Volatility* captures 5-min CQW midpoint return volatility. All measures represent daily averages for asset s on trading day td, the model's evaluation day. The exception is *diff. in volatility*, which refers to the absolute value of the relative difference between the volatility on the model evaluation day and estimation day in percentage points. Asterisks denote that the coefficient is different from zero at the 90% (*), 95% (**), and 99% (***) confidence levels, respectively.

	Dep. variable: Ratio of predictability: $\bar{e}_{m,s,td} \cdot 100$					
	Overall	Transaction price	Midpoint	QW midpoint	CQW midpoint	Micro-price
Intercept	98.93***	101.14***	100.55***	98.12***	99.40***	100.14***
Abs. imbalance	1.82***	1.62***	0.99***	5.10***	1.01***	0.13
No. trades	−0.04***	−0.20***	−0.03***	0.01	−0.01	−0.01
No. LOB updates	−0.38***	3.45***	−1.45***	−1.52***	−0.37***	−0.52***
Relative tick size	0.19***	−0.05	0.62***	0.12***	0.17***	0.17***
Order-to-trade ratio	0.00***	−0.01***	0.01***	0.01***	0.00***	0.00***
Volatility	−0.01***	0.06***	−0.06***	−0.02***	−0.01***	−0.02***
Diff. in volatility	−0.00**	−0.00**	−0.00**	−0.00	−0.00	−0.00
R^2 (in perc. pts.)	46.54	14.26	29.33	32.30	9.36	10.78
Observations	21790	4358	4358	4358	4358	4358
Fixed effects	Asset-day	Asset-day	Asset-day	Asset-day	Asset-day	Asset-day
Additional dummies	Price measures	None	None	None	None	None

Conversely, the number of trades has a negative and significant coefficient in the panel regression, suggesting that a larger number of transactions reduce return predictability as price measures are more frequently updated, thereby, reflecting the current level of information more accurately. Differentiating between the price measures, the results show that the number of transactions impacts transaction price efficiency most strongly, whereas the more complex price measures are not significantly affected by the number of transactions. In contrast, the number of LOB updates is positively associated with the predictability of transaction price returns, i.e. having a negative impact on efficiency, whereas the remaining price measures become more efficient with more order book updates. The deviations with respect to significance and sign of these coefficients highlight the substantial differences between trade-driven and order-driven price measures. In modern financial markets, the majority of information is transferred via passive order flow rather than trades (Brogaard et al., 2019). As a result, future transaction prices are highly predictable using prior order book information. This predictability increases when the number of order book updates increases, as order book imbalances reflect current information in a more timely manner. On the other hand, more transactions per asset-day do not necessarily affect the efficiency of complex price measures as the order book and the passive orders are already the central source of information.

Our findings also reveal that a larger relative tick size, which is constraining price movements, significantly increases predictability. This holds true in the pooled regression and all individual regressions except for the transaction price where the coefficient is not significantly different from zero. It may surprise that the QW midpoint, CQW midpoint, and micro-price are better predictable with an increase in relative tick size, even though their values are explicitly not restricted by the tick size as they depend on the best bid and ask prices and sizes. However, if there are investors willing to buy or sell at price limits that are invalid due to tick size constraints, this information may not be transferred into these price measures. Nevertheless, we would also expect that the transaction price to suffer from the same limitation, potentially even more than the other price measures. The insignificant and even negative coefficient can potentially be traced back to a higher rate of aggressive informed trading. When the relative tick size increases, implicit trading cost increases, especially when the bid-ask spread is equal to the tick size, as this artificially constrains the spread. This in turn reduces trading in small volumes as well as noise trading leaving a higher share of informed trading (Chung et al., 2020). This aggressive informed trading directly induces information at the time of the trade and is therefore not observable prior to the trade and consequently not an input of our prediction model. We interpret the insignificant coefficient for the transaction price as a balanced level of both opposing effects, the higher level of aggressive informed trading and the higher constraining effect of the tick size.

Regarding the other control variables, we observe a positive and significant impact of the order-to-trade ratio on predictability for all regression models. This suggests that the positive effect of the total number of order book updates is being reduced when the ratio of order book updates to the number of trades increases. This indicates that the informativeness of passive order flow decreases, when it is less likely to result in trades. Moreover, volatility on the evaluation day has a negative and significant effect in all regression models except for the transaction price. Again, the sign of the coefficient in the transaction price regression is flipped, further underlining the difference between trade-driven and order-driven price measures. For the relative difference between volatility on the evaluation and estimation day, we find that overall, a higher volatility difference reduces predictability on the evaluation day. This is in line with our priors, suggesting that a greater difference between the estimation and evaluation days, as measured by volatility, reduces the prediction model's generalization ability. However, the coefficient is not significant in all individual regression models.

In summary, the regression results highlight the significant impact of market frictions on the short-term predictability of price measures. In particular, the emergence of informed trading, observable through the order book imbalance, enhances the prediction of subsequent short term price developments. Frequent and synchronized trading activity, in turn, ensures that price measures are regularly updated, thereby reflecting all available information in a timely manner. In general, the price measures are affected by the examined frictions similarly, except for the transaction price which is the only price measure that is not affected by tick size constraints. These results underscore the role of frictions for price efficiency and highlight the need for future research to develop price measures that are less affected by such frictions in order to obtain a more accurate estimator of the true value.

Overall, our results underscore significant differences in the efficiency of price measures. The transaction price, midpoint, and QW midpoint reflect past information at a much slower rate than the micro-price or CQW midpoint. Although the midpoint and QW midpoint process more information than the transaction price due to their reliance on order book data, they require similar time periods to fully incorporate public information. The micro-price stands out as the most sophisticated measure, incorporating past information within a few seconds. Surprisingly, the CQW midpoint achieves an even slightly higher level of efficiency, providing a practical and computationally simpler alternative to the micro-price in high-frequency trading contexts. Our results remain consistent when we calculate the ratio of predictability, as defined in Equation 7, using an alternative benchmark instead of the no-change benchmark. Appendix C.1 presents the results of our analyses using the mean return of the previous asset-day as alternative naive benchmark, which yield similar outcomes.

4.2. Implications of inefficient price measures

The previous results indicate significant differences in the efficiency of various price measures. Discrepancies between an asset's true value and the price measure approximating it can distort empirical findings. If these deviations from the true value follow systematic patterns, this can bias study results in a specific direction.

In this section, we investigate the extent to which the five price measures deviate from an asset's true value and assess whether these deviations exhibit systematic patterns. Afterwards, we discuss the implications of these differences for research and practice. As in Hagströmer (2021), we demonstrate that less efficient price measures introduce biases in trading cost estimates. We extend these findings by quantifying the deviations of each price measure to the true asset value and analyzing their determinants. Additionally, we show how the utilization of inefficient reference prices can lead to suboptimal trading outcomes in dark pools. These outcomes can be mitigated by adopting more efficient price measures, creating fairer trading conditions in dark pool markets.

Differences between each price measure and the true value

We begin by calculating the differences between each price measure and an asset's true value at time t. As the true value is unobservable, this calculation requires a proxy. We decide to approximate the true value in t using the CQW midpoint in $t + 5s$. This choice is motivated by the CQW midpoint being the most efficient price measure among those considered. Moreover, using its realization five seconds after t ensures that, on average, all public information available up to t has been incorporated into the CQW midpoint, as demonstrated in the previous section (see Figure 4). Stoikov (2018) proceeds in a similar way when evaluating the micro-price. For robustness, we also repeat the analyses using the micro-price at $t + 5s$ as true value estimator at time t, yielding similar results as reported in Appendix C.2.

The differences between each price measure and the true value proxy are calculated based on each limit order book update in our sample. We calculate the differences for each asset-day and report the arithmetic mean across all asset-days.

To ensure that overall price levels do not influence our analysis, we calculate relative differences rather than absolute differences:

$$\delta_{m,s,t} = \frac{p^m_{s,t} - tv_{s,t+5s}}{tv_{s,t+5s}} \tag{12}$$

with $\delta_{m,s,t}$ being the relative difference between the price measure m ($p_{m,s,t}$) and the true value proxy ($tv_{m,s,t}$) at time t for stock s. Table 3 presents the average deviations of each price measure from the true value proxy, shown as both absolute values ($\overline{|\delta_{m,s,t}|}$) and non-transformed values ($\overline{\delta}_{m,s,t}$). While the values in the first column quantify the average absolute magnitude of the deviation, the second column indicates whether the price measures systematically under- or overestimate the true value. We conduct a t-test to test whether the average deviation differs significantly from zero.

Table 3: Average relative difference between the true value and each price measure

This table shows the average of the difference between the true value and each price measure per asset-day. As the true value proxy in t, we use the constrained quantity-weighted midpoint in $t+5s$. We use a t-test to test whether the average deviation in the third column differs significantly from zero. One star (*), two stars (**), and three stars (***) following the value indicate a rejection of the null hypothesis at the 90%, 95%, and 99% confidence levels, respectively.

Price measure	Average absolute deviation (bps)	Average deviation (bps)
Transaction price	2.68	−0.03***
Midpoint	1.63	−0.01***
QW midpoint	1.62	0.00
CQW midpoint	1.47	0.00**
Micro-price	1.51	−0.01***

The average absolute deviations reported in the first column range between 1.47 and 2.68 basis points, indicating that price measures differ on average from the true asset value. These deviations are large in magnitude and economically meaningful, considering that the average relative bid-ask spread of all assets is 4.70 basis points. Differentiating between the price measures, we find results in line with theory and our previous findings. The transaction price has the largest deviation from the true value proxy with an average difference of approximately 2.68 basis points. This corresponds to 57% of the average bid-ask spread. The remaining measures update more frequently and consequently exhibit smaller deviations. We find the same ranking as in our previous analyses, with the more informative measures, CQW midpoint and micro-price, being the ones with the least deviation from the true value.

To identify the existence of systematic over- or underestimations, we examine the average difference across all asset-days. The values in the second column show that most price measures, on average, underestimate the true value, with almost all deviations being significantly different from zero. However, these deviations are small in magnitude and close to zero. Most likely, the slight underestimation can be attributed to the overall upward trend observed across all DAX40 stocks during our sample period.[7] Overall, the average differences suggest that over- and underestimations of the true value are nearly balanced across all observations in our sample.

While these deviations are negligible across all observations, they may not be evenly distributed across different market conditions. Since both theoretical models and empirical findings indicate that order book imbalance contains valuable information about an asset's true value, three out of the five price measures consider this information. To further examine potential systematic biases in the price measures, we categorize observations based on the prevailing order imbalance. Specifically, we discretize the order book imbalance into equally sized deciles: $(0, 0.1], (0.1, 0.2], ...,$ $(0.9, 1.0]$.

Figure 5 illustrates how the average deviations from the true value vary with order book imbalance. The x-axis shows the mid of each of the ten equally spaced order book imbalance buckets, while the y-axis displays the difference between each price measure and the true value proxy. The plot demonstrates that the deviations differ systematically across price measures when controlling for order book imbalance. Specifically, the midpoint consistently overestimates the true value when order book imbalance is low (i.e., sell-pressure is high) and underestimates the true value when order book imbalance is high (i.e., buy-pressure is high) by up to 1.05 bps. The QW midpoint exhibits an inverse relationship compared to the midpoint, with deviations of up to 0.78 bps. This inverse relationship is attributable to the QW midpoint's tendency to shift excessively toward the best bid or ask price during times of large order book imbalances, especially when the bid-ask spread exceeds one tick. This is evident when comparing the QW with the CQW midpoint. When the bid-ask spread equals one tick, the QW equals the CQW midpoint, but in times of larger spreads the QW shifts stronger than the CQW towards the best bid or ask. In contrast to the QW midpoint, the CQW is only slightly attributing the differences to the extreme shifts when spreads are large. While the figure shows that the tick-size-constrained

[7] The DAX40 gained 14.8% between January and June 2023.

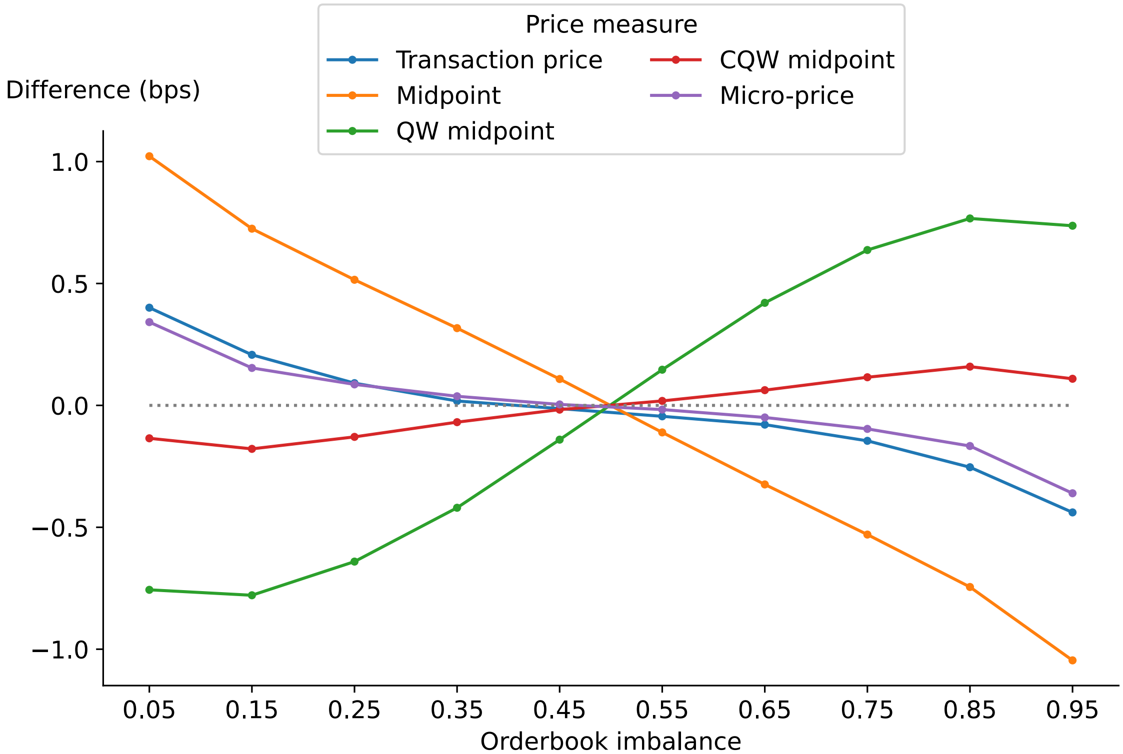


Figure 5: Average difference between the true value proxy and each price measure

This figure shows the average difference between the true value and each price measure per asset-day as a function of order book imbalance. The order book imbalance is divided into 10 equally sized deciles. The x-axis represents the mid of each respective decile. As the true value proxy in t, we use the constrained quantity-weighted midpoint in $t + 5s$.

adjustments of the CQW midpoint with respect to the order book imbalance are still too large, up to 0.18 bps, the bias is substantially lower than for the QW. The micro-price behaves similarly to the transaction price, with both exhibiting larger over- and underestimations during periods of large imbalances compared to the CQW midpoint, but to a lesser extent than the other two measures. However, if we use the micro-price in $t + 5s$ as true value proxy, the micro-price in t instead of the CQW midpoint has the least deviations from the true value proxy. All other findings remain as shown by Figure C.6 in the Appendix.

While these results provide evidence that the deviations of a price measure from the true value depend on order book imbalance, the difference between the QW and CQW midpoint further suggests that this may also be influenced by other market conditions, such as the bid-ask spread. Table B.3 in the Appendix lists the average absolute deviations given different levels of the bid-ask spread. The table indicates that, in general, a wider spread increases the difference to the true value across all price measures. However, for the midpoint, CQW midpoint, and micro-price, the deviations are slightly smaller for observations with two ticks than for those with one tick. This is likely due to the fact that as the spread widens, the midpoint of the bid-ask spread becomes a relatively reasonable approximation of the true value. However, this effect is eventually outweighed by generally larger deviations from the true value at higher spread levels.

Overall, the transaction price exhibits the largest bias in absolute terms because it captures information only at the moment a trade occurs and constantly oscillates between the best bid and ask prices, leading to persistent over- or underestimation regardless of the order book imbalance. In contrast, both the micro-price and the CQW midpoint display the smallest deviations from the true value, reaffirming their superiority in informational efficiency, as highlighted in Section 4.1. However, the two measures differ systematically in how they over- or underestimate the true value under varying conditions.

In summary, price measures systematically over- or underestimate the true value of an asset, with the direction of these deviations influenced by market dynamics, such as order book imbalance. The biases as well as their magnitude differ across price measures under varying conditions. Our results challenge the dominance of the midpoint as true value proxy among both researchers and practitioners.

Trading cost estimation

To examine how these biases can affect study results or investor trading outcomes, we calculate the effective spread using different price measures as proxies for the true value of an asset. The effective spread, along with its components — price impact and realized spread — is a widely used metric for liquidity and transaction cost estimation. Accurate computation of the effective spread and its components relies on selecting an appropriate true value estimator. The effective spread is defined as

$$es_{s,t} = 2q_{s,t} \, \frac{tp_{s,t} - tv_{s,t}}{tv_{s,t}}, \tag{13}$$

where $q_{s,t}$ represents the trade direction (+1 for a buy-initiated and -1 for a sell-initiated transaction), tp is the transaction price, and tv denotes the true asset value of asset s at time t. The effective spread reflects the implicit cost of trading by considering the difference between the actual execution price and the prevailing true value. In research and practice, the midpoint is commonly used as the true value estimator when calculating effective spreads. However, our previous results demonstrate that the midpoint is not the optimal true value proxy. In particular, when order book imbalance is low or high, the midpoint systemically over- or underestimates the true value, respectively. To examine how transaction cost estimates depend on the choice of the true value proxy, we calculate the effective spread for all transactions across all asset-days in our sample using each price measure. Since the CQW midpoint is the most accurate estimator of the true value according to our results, we consider it the optimal choice as a true value proxy for calculating effective spreads. Thus, we compare the effective spread calculated using the CQW midpoint to those based on other price measures.[8] We use a t-test to test whether the deviations between the different effective spread estimates are statistically significantly different from zero. Note that we exclude the transaction price as a true value proxy in this analysis because it will always result in an effective spread equal to zero.

Table 4 presents the average effective spread calculated using different price measures, alongside the average (relative) differences between the effective spread based on the CQW midpoint and those based on other price measures. The results indicate that the midpoint consistently overestimates the effective spread by nearly one basis point, while the micro-price overestimates by approximately half a basis point. Conversely, the QW midpoint underestimates the effective spread by about 0.2 basis points. All biases are both statistically significant and substantial in magnitude, with the relative bias reaching up to 38% in the case of the midpoint.

The substantial overestimation of transaction cost by the effective spread based on the midpoint can be traced back to the midpoint's inability to take into account the current order book imbalance. When buy-side pressure is high (i.e., when the order book imbalance is close to 1), the occurrence of a buy-initiated transaction is more likely, shifting the true value towards the best ask price, thus shifting it above the midpoint. Vice versa, in the presence of high sell-side pressure, the true value shifts towards the best bid, resulting in transaction cost overestimation. This finding aligns with the current literature. Hagströmer (2021) also finds that using the midpoint leads to systematic overestimation of effective spreads, though to a lesser extent, ranging between 13% and 18%. However, he does not make use of the CQW midpoint, and his results are based on a U.S. sample from 2015, making it difficult to compare both results.

A less intuitive finding from our results is that the less efficient QW midpoint estimates transaction costs more accurately than the more efficient micro-price. To further investigate this finding, Table 5 categorizes the effective spread biases by

[8] We have repeated the same analysis considering the effective spread based on the micro-price as the optimal choice. The results are presented in Table C.6, C.7, and C.8 in the Appendix.

Table 4: Effective spread bias of different price measures compared to the true effective spread
In this table we present the average effective spread across all asset-days when using different price measures as true value estimators. The average bias in the second column is calculated as the difference between the average effective spread in the first column and the true effective spread estimate. We use the effective spread calculated using the constrained quantity-weighted midpoint as the true estimate. All effective spread calculations are given in basis points. The relative average bias is then calculated as the average bias divided by the average effective spread and is given in percentage points. We use a t-test to test whether the average bias in the second column differs significantly from zero and report the t-statistics as well as the p-value in the last columns.

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
CQW midpoint	2.55				
Midpoint	3.52	−0.97	−38.18	−130.53	0.00
QW midpoint	2.34	0.21	8.06	82.29	0.00
Micro-price	3.03	−0.48	−18.93	−105.45	0.00

bid-ask spread size, grouping them into buckets corresponding to spreads of one, two, and three or more tick sizes. When the spread equals one tick, the CQW midpoint and the QW midpoint are identical, as the constraining effect of the CQW midpoint is irrelevant, and the bias of the QW midpoint is equal to zero. For spreads of two ticks, the midpoint overestimates the effective spread by an average of 0.65 basis points, while the QW midpoint underestimates it by the same magnitude. In contrast, the micro-price demonstrates the smallest bias, overestimating the effective spread by only 0.15 basis points. For spreads of three or more ticks, the QW midpoint shows the largest bias, underestimating the effective spread by 0.54 basis points. The midpoint continues to overestimate, with an average bias of 0.22 basis points, while the micro-price closely approximates the effective spread calculated using the CQW midpoint, with a minor overestimation of just 0.05 basis points. Overall, the bias in the effective spread is not only driven by order book imbalance but also by the bid-ask spread at the time of trade. Under high liquidity conditions, where the bid-ask spread is equal to one tick, the QW and CQW midpoints yield identical results. In our sample, almost two-thirds of all trades occur under these conditions, contributing to the small bias in the QW midpoint when estimating transaction costs. However, as spreads widen, the micro-price becomes increasingly accurate in approximating the effective spread, whereas the biases associated with the QW midpoint become more pronounced relative to those of the midpoint.

Table 5: Effective spread bias of different price measures given different spread sizes

In this table we present the average effective spread across all asset-days when using different price measures as true value estimators. The table divides observations of effective spread calculations by different absolute spread sizes: (a) one tick, (b) two ticks, and (c) three or more ticks. The average bias in the second column is calculated as the difference between the average effective spread in the first column and the true effective spread estimate. We use the effective spread calculated using the constrained quantity-weighted midpoint as the true estimate. All effective spread calculations are given in basis points. The relative average bias is then calculated as the average bias divided by the average effective spread and is given in percentage points. We use a t-test to test whether the average bias in the second column differs significantly from zero and report the t-statistics as well as the p-value in the last columns.

(a) Effective spread (absolute spread is 1 tick; 65% of all observations)					
Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
CQW midpoint	1.61				
Midpoint	2.64	−1.03	−64.29	−126.55	0.00
QW midpoint	1.61	0.00	0.00	0.00	1.00
Micro-price	2.17	−0.57	−35.26	−111.84	0.00
(b) Effective spread (absolute spread is 2 ticks; 28% of all observations)					
Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
CQW midpoint	4.71				
Midpoint	5.35	−0.65	−13.74	−129.33	0.00
QW Midpoint	4.06	0.65	13.82	128.96	0.00
Micro-price	4.86	−0.15	−3.27	−68.75	0.00
(c) Effective spread (absolute spread is 3+ ticks; 7% of all observations)					
Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
CQW midpoint	8.60				
Midpoint	8.82	−0.22	−2.57	−23.24	0.00
QW Midpoint	8.06	0.54	6.29	16.09	0.00
Micro-price	8.65	−0.05	−0.53	−14.22	0.00

For the sake of completeness, we have also calculated the biases in the components of the effective spread, namely price impact and realized spread. The results are presented in Table B.2 in the Appendix. The results demonstrate that these biases persist in both components in the short term (one-second horizon), while in the long term, the bias is predominantly present in the price impact measure. These findings are also consistent with those of Hagströmer (2021).

Our analysis highlights that the choice of a true value proxy significantly impacts effective spread estimation. Biases in effective spread estimation arise not only from order book imbalances but also from larger bid-ask spreads. Overall, these biases are economically meaningful and may have introduced distortions in both existing studies and investors' decision-making. Using more efficient price measures, such as the micro-price or the CQW midpoint, can help mitigate these biases in the future.

Dark pool trading outcomes

Another critical context requiring an efficient proxy for the true value is in dark pools, where institutional investors can trade without publicly displaying their trading intentions. These trading venues have gained significance over the past decade and play a central role in global equity trading. According to Deutsche Börse Group (2024), dark pool trading accounted for approximately 11% of the total trading volume in DAX40 equities as of September 2024. Unlike traditional stock exchanges, dark pools function without publicly visible quotes, meaning there is no information about resting orders in the pool before execution. While trading mechanisms may vary across different dark pools, the arguably most common approach is matching incoming orders with resting orders at a specific reference price, most often the midpoint of the primary exchange. Advocates argue that, due to the lack of pre-trade transparency and execution at a reference price such as the midpoint, institutional investors can execute large trades without market impact and avoid implicit transaction costs by not paying half of the bid-ask spread.

However, as demonstrated earlier, the midpoint is not the most efficient estimator of the true value and introduces a bias of 38% when used for effective spread calculation. Under certain market conditions, such as high order book imbalance, the midpoint can deviate from the best true value estimate by up to one basis point. Consequently, we argue that implicit transaction costs still exist in dark pool transactions, when the true value deviates from the reference price. This deviation can lead to unfair executions, as one side of the trade incurs implicit transaction costs while the other benefits. Furthermore, it poses a risk of passive orders in the pool being adversely selected when the reference price deviates from the true value in favor of the execution-triggering dark pool order.

To assess the potential implicit costs of executions in dark pools, we calculate the difference between execution prices and the most efficient true value estimator. Our analysis is based on Xetra Midpoint trades, a dark pool service introduced by Deutsche Börse on December 9, 2024, as an addition to its continuous trading market model. The observation period spans from the service launch until January 29, 2025, covering all DAX40 stocks. Our sample includes 6.541 dark pool transactions across all asset-days in this sample. As in Table 3, we calculate the differences between each price measure and the true value at t, with the CQW midpoint in $t+5s$ serving as an approximation of that true value. However, in this analysis, we calculate these differences using two different sampling frequencies. As before, we sample at event-time and use each order book update to calculate this difference, but we also calculate it each time a dark pool trade occurs. This enables us to assess whether implicit transaction costs can be reduced by utilizing an alternative price measure as the reference price, potentially leading to fairer executions in dark pool transactions. Additionally, we analyze whether trades in dark pools tend to occur at times when the difference between the midpoint (and other price measures) and the true value is systematically different. Table 6 shows the average absolute relative difference between each price measure and the true value considering each dark pool transaction (first column) and limit order book update (second column).

Table 6: Average relative absolute difference between each price measure and the true value considering each dark pool trade and order book update

This table shows the average of the relative absolute difference between each price measure and the true value per asset-day. The difference is calculated at the time of each dark pool transaction (first column) and each limit order book update (second column). As the true value proxy in t, we use the constrained quantity-weighted midpoint in $t+5s$.

	Average absolute deviations (bps)	
Price measure	Dark pool transactions	All limit order book updates
Transaction price	2.27	2.59
Midpoint	1.83	1.59
QW midpoint	1.78	1.57
CQW midpoint	1.63	1.41
Micro-price	1.71	1.47

The first column shows that dark pool execution prices, with the midpoint as reference price, deviate on average by 1.83 bps from the best true value approximation. In contrast, if these trades were executed using the CQW midpoint as reference price, this would reduce the deviation to 1.63 bps, representing a relative improvement of nearly 11% over midpoint executions. Furthermore, substituting the midpoint with the QW midpoint or micro-price as the reference price would enhance the fairness of dark pool transactions by approximately 3% and 7% in relative terms, respectively.

Interestingly, the average deviation between an asset's true value and the midpoint is higher when focusing on the times of dark pool executions (1.83 bps) compared to considering all limit order book updates (1.59 bps) as shown in the second column of Table 6. This suggests that dark pool trades are more frequently executed under conditions in which the midpoint is particularly biased, such as when order book imbalance is high. To test this hypothesis, we apply the same discretization of the order book imbalance as before and compare its distribution at the time of all dark pool trades with the imbalance at the time of each order book update on the reference market. Figure 6 shows that the order book imbalance is approximately normally distributed when considering all order book updates. However, dark pool transactions are more likely to occur when the reference market's order book imbalance is particularly high or low, meaning the midpoint is especially biased. This pattern suggests that high-frequency traders may exploit inefficiencies in the reference price in times of low or high imbalances, effectively "picking off" less informed or slow traders who provide liquidity in the dark pool.

In summary, since dark pools rely on price measures from the main market to establish a reference price, they are particularly vulnerable to inefficiencies in these measures, which can affect execution fairness in these venues. The use of an inefficient reference price, such as the midpoint, introduces implicit transaction costs due to adverse selection, as high-frequency traders can exploit these inefficiencies. Adopting a more efficient price measure, such as the CQW midpoint, could reduce true value misestimation by 11%, enhancing fairness and execution quality in dark pools.

The use of less efficient price measures results in systematically biased approximations of an asset's true value. This bias is systematic in the sense that each price measure persistently over- or underestimates the true value under specific market conditions. The transaction price is a notable exception, as its misjudgment of the true value is unsystematic, yet it exhibits the largest bias among all measures.

These systematic biases directly affect empirical applications, such as measuring market illiquidity and estimating transaction costs through the effective spread. Depending on the chosen price measure, the bias in effective spread calculations can reach up to 38%. Additionally, execution prices in dark pools deviate on average, by 1.83 bps from the true value. These implicit trading costs could be reduced by 11% when a more efficient price measure is used as reference price. This underscores the importance of selecting appropriate price measures for both academic researchers as well as practitioners, such as market participants and market operators.

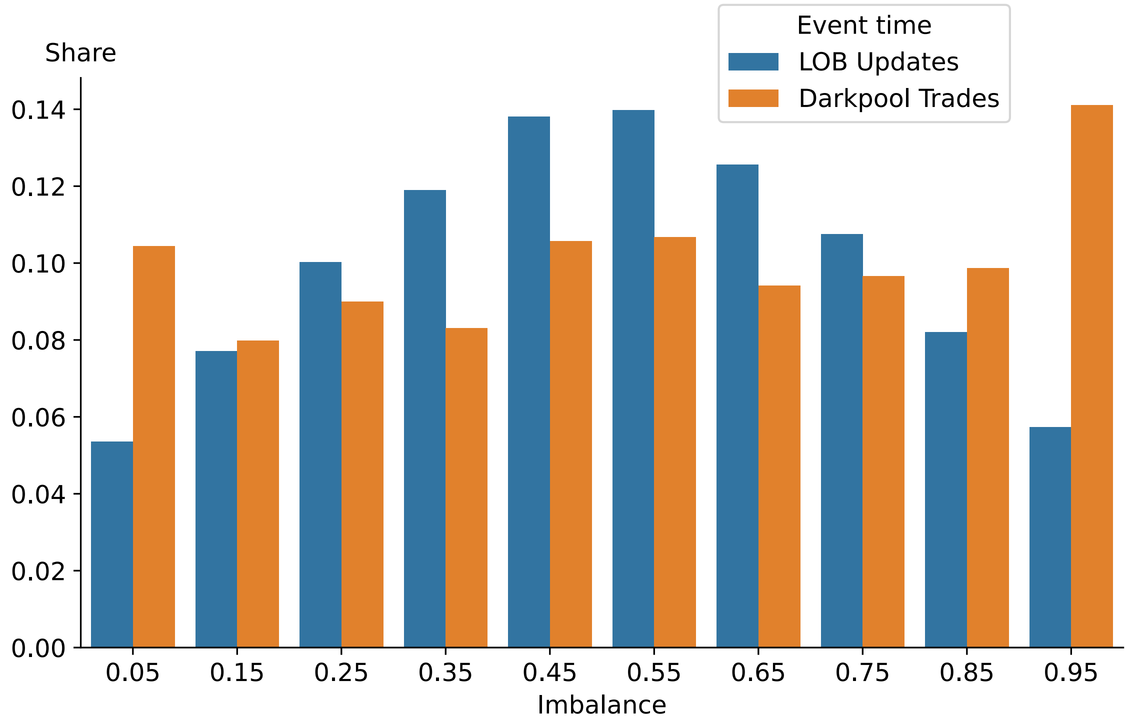


Figure 6: Distribution of order book imbalance

This figure shows the share of trades in the dark pool in orange and limit order updates on the reference market in blue given different levels of order book imbalance. The order book imbalance is divided into 10 equally sized deciles. The x-axis represents the mid of each respective decile.

5. True value estimator choice

The true value of an asset, while unobservable, is a critical input for both empirical research and investors' decision-making processes. To approximate this elusive true value, researchers have proposed and applied various price measures. However, our analyses demonstrate that established price measures can differ significantly from the true value, with deviations averaging up to 2.68 basis points. Such discrepancies can introduce biases into research findings or execution prices, underscoring the importance of evaluating each price measure's efficiency under varying conditions. This section provides a brief summary to guide researchers and practitioners in selecting an appropriate price measure. For an overview of the qualitative and quantitative characteristics of the analyzed price measures, we refer to Table B.4 in the Appendix.

From a theoretical perspective, the micro-price seems to be the best choice as a true value estimator among all analyzed price measures. Stoikov (2018) designed the price measure to meet several requirements for a true value proxy. In contrast to all other measures, it is constructed as a martingale and explicitly takes into account numerous future price movements. While the micro-price is strictly speaking not a continuous variable, the number of states for the discretized input variables can be increased, allowing the measure to converge to a nearly continuous variable.

Our empirical analysis, however, identifies the CQW midpoint as the most efficient proxy of an asset's true value. It is the least predictable price measure and takes less than five seconds to fully incorporate past information. More rigid measures such as the transaction price or the midpoint need more than 30 seconds to fully reflect historical information. However, the micro-price has a very similar level of efficiency. It exhibits almost the same degree of predictability and also reflects prior information within less than five seconds.

Nevertheless, the CQW midpoint appears to be an excellent choice for a true value approximation, as it is significantly easier to calculate compared to the similarly efficient micro-price. However, it should be noted that our results only apply to liquid assets. The micro-price might demonstrate its strengths in less liquid markets, as variations in the spread could have a significantly greater impact there than in liquid markets, where the bid-ask spreads of assets are predominantly one or two tick sizes wide.

All price measures only require trade and quote data. As a result, researcher's choice regarding the applied true value proxy should not be significantly affected by data restrictions, as most databases provide trades and quotes as a bundled dataset. Thus, the selection of a price measure predominantly depends on the research question, investigated sampling frequency, and computational resources.

In general, our findings demonstrate that the choice of a price measure becomes increasingly critical at higher frequencies, where order book-driven estimators are necessary to obtain reliable results. Conversely, at frequencies lower than one minute, all measures converge in efficiency, as they sufficiently capture past information within such time intervals. Hence, researchers addressing questions that require low-frequency intraday data, such as 5-minute returns or lower, can leverage simple price measures like transaction prices or midpoints when dealing with liquid assets. However, when investigating more granular frequencies, researchers should consider the CQW midpoint or the micro-price.

In conclusion, selecting an appropriate price measure requires careful consideration of the research context, computational resources, and sampling frequency. While the CQW midpoint emerges as a strong general-purpose choice due to its simplicity and efficiency, less efficient measures can also suffice when paired with appropriate adjustments to the sampling frequency. At least, researchers and practitioners should be aware of the potential consequences of using inefficient price measures when balancing the trade-off between the practicality of a study and the validity of results.

6. Conclusion

This paper provides a comprehensive analysis of the qualitative and quantitative properties of different price measures that are used in market microstructure as proxies for the true value of a financial asset. Specifically, we examine five distinct price measures - transaction price, midpoint, quantity-weighted (QW) midpoint, constrained quantity-weighted (CQW) midpoint, and the micro-price proposed by Stoikov (2018) - regarding their ability to incorporate public information efficiently. This evaluation is based on return predictability, which serves as an inverse measure of price efficiency.

Our findings reveal that all price measures exhibit significant levels of predictability, but substantially differ in their ability and speed of reflecting past information. The transaction price, midpoint, and QW midpoint exhibit higher inefficiencies, requiring over 30 seconds to fully reflect public information. In contrast, more sophisticated measures, such as the micro-price and the CQW midpoint, incorporate information within just a few seconds, making them superior proxies for an asset's true value.

Our findings have important implications for both researchers and market practitioners. The choice of an inefficient price measure can introduce systematic biases in empirical research, leading to misestimated transaction costs and suboptimal trading decisions. For instance, we show that using the midpoint as a proxy for the true value leads to an overestimation of effective spreads by up to 38%, significantly impacting liquidity measurement and execution cost analysis. Additionally, we demonstrate that execution prices in dark pools, where executions rely on reference prices, deviate on average by 1.83 basis points from the true value, leading to implicit trading costs and potential unfair execution outcomes. Using the CQW midpoint as the reference price could help mitigate this issue, enhancing execution fairness by 11%.

From a practical perspective, our results highlight the importance of selecting appropriate price measures based on research objectives and the underlying sampling frequency. While the micro-price theoretically provides the most robust true value approximation, our empirical results suggest that the CQW midpoint offers a similar level of efficiency with significantly lower computational complexity. Therefore, for high-frequency trading applications or studies requiring precise true value estimation, we recommend the adoption of the CQW midpoint or the micro-price over traditional measures such as the midpoint or transaction price. When using simpler price measures, such as the transaction price or midpoint, either for simplicity or due to the unavailability of volume data, researchers and practitioners must consider the potential impact of the measures' inefficiencies on the validity of their results.

Overall, our study advances the understanding of price efficiency in market microstructure and provides actionable guidance for improving research methodologies and trading strategies. Future research could explore the applicability of these findings across different market environments, including less liquid assets. Furthermore, future research can develop new price measures or enhance existing ones to be more resilient to market frictions, providing a more accurate and efficient representation of an asset's true value.

References

Aït-Sahalia, Y., Fan, J., Xue, L., Zhou, Y., 2022. How and When are High-Frequency Stock Returns Predictable? Technical Report. National Bureau of Economic Research.

Anshuman, V.R., Kalay, A., 1998. Market making with discrete prices. The Review of Financial Studies 11, 81–109.

Bonart, J., Lillo, F., 2018. A continuous and efficient fundamental price on the discrete order book grid. Physica A: Statistical Mechanics and its Applications 503, 698–713.

Brogaard, J., Hendershott, T., Riordan, R., 2019. Price discovery without trading: Evidence from limit orders. The Journal of Finance 74, 1621–1658.

Cao, C., Hansch, O., Wang, X., 2009. The information content of an open limit-order book. Journal of Futures Markets: Futures, Options, and Other Derivative Products 29, 16–41.

Chordia, T., Roll, R., Subrahmanyam, A., 2002. Order imbalance, liquidity, and market returns. Journal of Financial Economics 65, 111–130.

Chordia, T., Roll, R., Subrahmanyam, A., 2005. Evidence on the speed of convergence to market efficiency. Journal of Financial Economics 76, 271–292.

Chordia, T., Roll, R., Subrahmanyam, A., 2008. Liquidity and market efficiency. Journal of Financial Economics 87, 249–268.

Chung, K.H., Chuwonganant, C., 2023. Tick size and price efficiency: Further evidence from the tick size pilot program. Financial Management 52, 483–511.

Chung, K.H., Lee, A.J., Rösch, D., 2020. Tick size, liquidity for small and large orders, and price informativeness: Evidence from the tick size pilot program. Journal of Financial Economics 136, 879–899.

Conrad, J., Wahal, S., Xiang, J., 2015. High-frequency quoting, trading, and the efficiency of prices. Journal of Financial Economics 116, 271–291.

Cont, R., Kukanov, A., Stoikov, S., 2014. The price impact of order book events. Journal of Financial Econometrics 12, 47–88.

Deutsche Börse Group, 2024. Deutsche Börse Cash Market launches new trading feature with Xetra Midpoint. Technical Report. Deutsche Boerse Group.

Fama, E.F., 1970. Efficient capital markets. The Journal of Finance 25, 383–417.

Glosten, L.R., 1994. Is the electronic open limit order book inevitable? The Journal of Finance 49, 1127–1161.

Glosten, L.R., Harris, L.E., 1988. Estimating the components of the bid/ask spread. Journal of Financial Economics 21, 123–142.

Goldstein, M., Kwan, A., Philip, R., 2023. High-frequency trading strategies. Management Science 69, 4413–4434.

Hagströmer, B., 2021. Bias in the effective bid-ask spread. Journal of Financial Economics 142, 314–337.

Harris, L., 1990. Liquidity, trading rules and electronic trading systems. Monograph Series in Finance and Economics 90-4.

Hasbrouck, J., 1993. Assessing the quality of a security market: A new approach to transaction-cost measurement. The Review of Financial Studies 6, 191–212.

Hasbrouck, J., 2002. Stalking the "efficient price" in market microstructure specifications: an overview. Journal of Financial Markets 5, 329–339.

Hendershott, T., Jones, C.M., 2005. Island goes dark: Transparency, fragmentation, and regulation. The Review of Financial Studies 18, 743–793.

Hendershott, T., Menkveld, A.J., 2014. Price pressures. Journal of Financial Economics 114, 405–423.

Hou, K., Moskowitz, T.J., 2005. Market frictions, price delay, and the cross-section of expected returns. The Review of Financial Studies 18, 981–1020.

Kaniel, R., Liu, H., 2006. So what orders do informed traders use? The Journal of Business 79, 1867–1913.

Kim, J.H., Shamsuddin, A., 2008. Are asian stock markets efficient? evidence from new multiple variance ratio tests. Journal of Empirical Finance 15, 518–532.

Kyle, A.S., 1985. Continuous auctions and insider trading. Econometrica: Journal of the Econometric Society , 1315–1335.

Lo, A.W., MacKinlay, A.C., 1989. The size and power of the variance ratio test in finite samples: A monte carlo investigation. Journal of Econometrics 40, 203–238.

Riccó, R., Rindi, B., Seppi, D.J., 2020. Information, liquidity, and dynamic limit order markets. IGIER, Università Bocconi.

Rösch, D.M., Subrahmanyam, A., Van Dijk, M.A., 2017. The dynamics of market efficiency. The Review of Financial Studies 30, 1151–1187.

Stoikov, S., 2018. The micro-price: a high-frequency estimator of future prices. Quantitative Finance 18, 1959–1966.

Appendix A. Descriptive statistics

Table A.1: Descriptive statistics

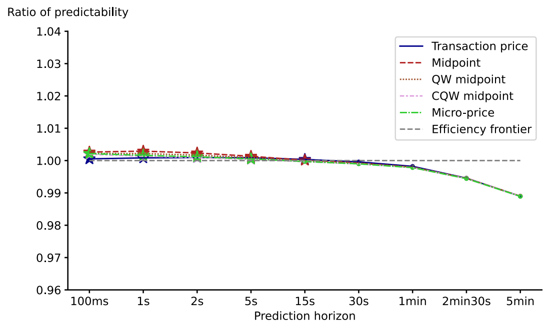
This table shows descriptive statistics for all stocks in our sample. The values are averages based on the daily activity during continuous trading from 08:17 to 11:45 and 12:17 to 16:15 UTC. The trading volume is given in million €. The volatility is calculated as the standard deviation of one minute midpoint returns. Both the relative spread and the volatility is given in basis points. The number of trades and the number of updates is given in thousands. The number of updates is the number of order book changes up to the ten best price levels.

Asset	Trading volume	Relative spread	Volatility	No. trades	No. updates
Adidas	39.87	3.63	7.43	6.68	713.44
Airbus	18.94	3.74	4.91	1.98	490.08
Allianz	113.08	3.19	3.65	5.12	287.31
BASF	60.27	2.83	5.18	7.14	602.25
BMW	47.92	2.60	4.36	6.46	644.57
Bayer	62.50	3.00	4.80	6.72	382.84
Beiersdorf	11.94	6.22	4.13	1.66	89.28
Brenntag	14.25	4.98	5.21	3.06	215.64
Commerzbank	36.40	6.10	7.90	4.60	261.20
Continental	14.52	5.89	7.43	3.16	292.19
Covestro	24.18	5.15	7.51	4.30	294.46
Daimler Truck	21.08	5.03	5.95	4.12	312.92
Deutsche Bank	63.70	3.32	6.11	8.17	660.93
Deutsche Boerse	27.18	4.67	4.60	3.23	164.28
Deutsche Post	44.14	2.82	5.20	6.53	611.80
Deutsche Telekom	75.12	3.15	3.85	5.26	284.81
E.ON	27.15	5.74	4.12	2.68	96.92
Fresenius	14.74	6.31	6.89	3.02	192.12
Fresenius Med. Care	11.05	6.75	7.09	2.72	180.75
Hannover Rueck	11.25	5.30	4.90	2.08	174.66
Heidelberg Cement	12.81	5.27	5.33	2.40	207.96
Henkel	14.21	5.07	4.10	3.09	133.39
Infineon	64.52	3.35	7.23	8.81	751.02
Linde	165.97	2.66	5.38	11.73	499.55
MTUAeroEngines	14.26	6.53	5.27	1.89	118.43
Mercedes Benz	90.77	2.64	5.23	8.55	666.21

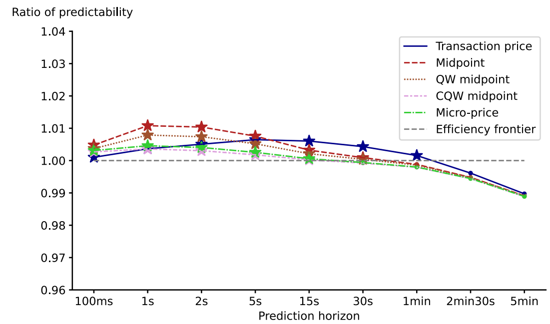
Continuation of table A.1

Asset	Trading volume	Relative spread	Volatility	No. trades	No. updates
Merck	22.17	5.05	5.53	3.27	186.93
Muenchener Rueck.	49.42	4.44	4.66	3.55	198.18
Porsche	27.27	5.88	6.02	3.21	213.32
Porsche Automobil	26.83	5.74	5.84	3.11	209.40
Qiagen	9.88	4.84	4.59	2.16	230.10
RWE	34.51	4.23	5.39	4.56	225.94
Rheinmetall	37.51	7.00	7.23	4.00	162.07
SAP	84.33	2.70	4.46	7.76	479.00
Sartorius	19.88	6.00	8.81	3.86	389.97
Siemens	82.58	2.46	5.13	8.50	671.40
Siemens Health.	21.19	5.60	5.29	2.62	159.45
SiemensEnergy	25.53	5.81	7.38	4.82	212.78
Symrise	14.12	6.13	5.36	2.36	205.69
Volkswagen	79.16	3.26	5.54	7.70	588.78
Vonovia	41.99	5.98	9.20	6.16	344.64
Zalando	20.69	6.22	10.10	4.71	406.07
Average	40.45	4.70	5.82	4.70	338.40

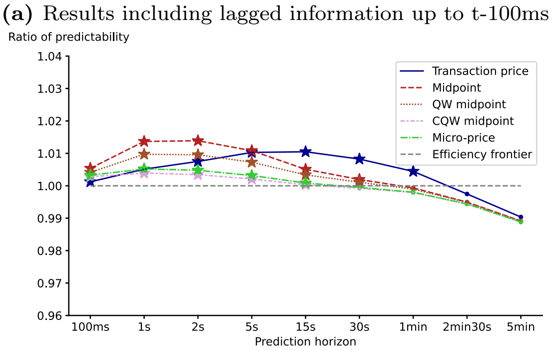
Appendix B. Additional results



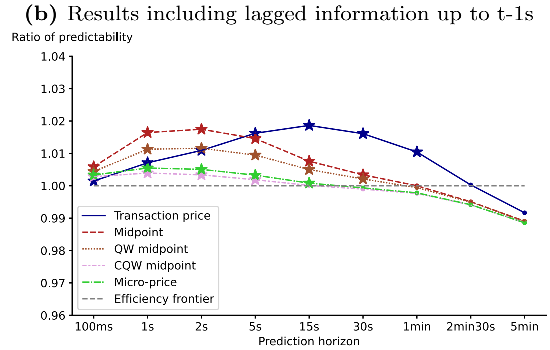
(a) Results including lagged information up to t-100ms



(b) Results including lagged information up to t-1s



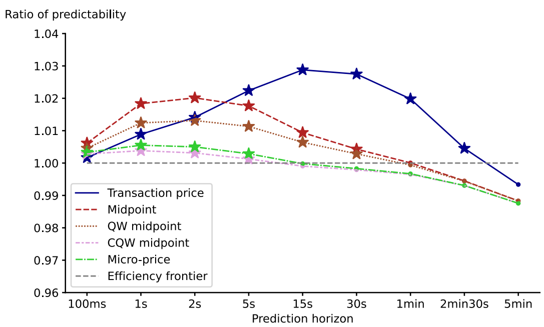
(c) Results including lagged information up to t-2s



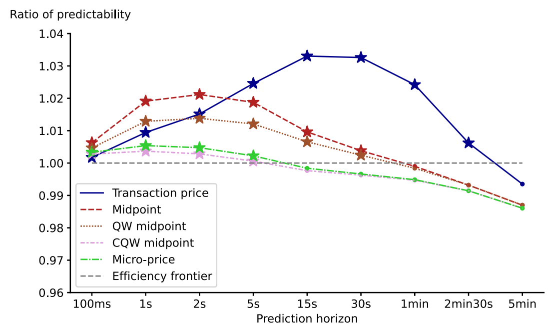
(d) Results including lagged information up to t-5s

Figure B.1: Ratio of predictability of all five price measures at various prediction horizons using past information up to $t - 5s$

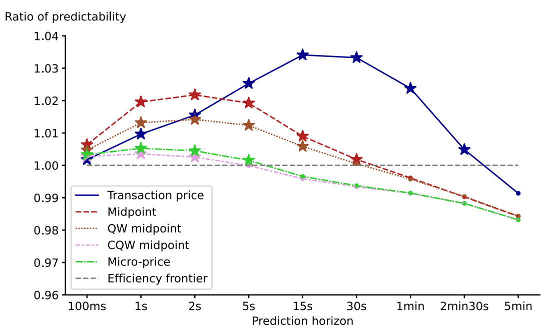
In this figure a star indicates that the ratio of predictability is statistically significantly larger than 1 at the corrected 5% significance level after applying the Bonferroni correction.



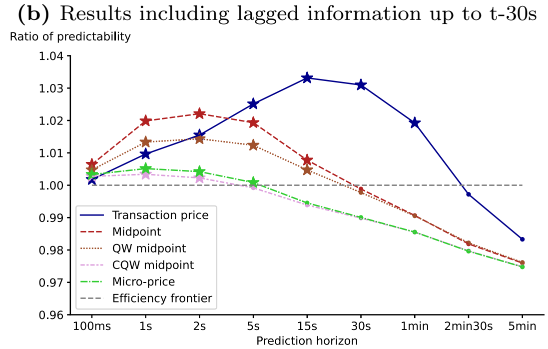
(a) Results including lagged information up to t-15s



(b) Results including lagged information up to t-30s



(c) Results including lagged information up to t-1min



(d) Results including lagged information up to t-2min30s

Figure B.2: Ratio of predictability of all five price measures at various prediction horizons using past information up to $t - 2.5min$

In this figure a star indicates that the ratio of predictability is statistically significantly larger than 1 at the corrected 5% significance level after applying the Bonferroni correction.

Table B.2: Price impact and realized spread bias of different price measures at 1s and 60s horizons
This table presents the average price impact and realized spread at 1s and 60s horizons using different true value estimators. Results include the price impact in (a) and (c) as well as realized spread in (b) and (d) at both horizons. The average bias is the difference between the first column and the true estimate, based on the constrained quantity-weighted midpoint. The relative bias is expressed in percentage points. A t-test assesses significance, with t-statistics and p-values in the last columns.

(a) Price impact 1s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
MidPriceCQW	2.80				
MidPrice	3.61	−0.81	−28.73	−124.19	0.00
MidPriceQW	2.67	0.13	4.78	61.58	0.00
MicroPrice	3.21	−0.41	−14.63	−102.72	0.00

(b) Realized spread 1s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
MidPriceCQW	−0.25				
MidPrice	−0.09	−0.17	−66.43	−88.51	0.00
MidPriceQW	−0.32	0.07	28.23	50.16	0.00
MicroPrice	−0.18	−0.07	−28.75	−77.27	0.00

(c) Price impact 60s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
MidPriceCQW	3.08				
MidPrice	4.06	−0.98	−31.95	−129.69	0.00
MidPriceQW	2.85	0.22	7.19	85.37	0.00
MicroPrice	3.56	−0.48	−15.69	−105.59	0.00

(d) Realized spread 60s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
MidPriceCQW	−0.53				
MidPrice	−0.53	0.01	1.73	9.01	0.00
MidPriceQW	−0.51	−0.02	−2.97	−18.53	0.00
MicroPrice	−0.53	0.00	−0.06	−0.66	0.51

Table B.3: Average absolute difference between the true value and each price measure given different bid-ask spread sizes

This table shows the average of the difference between the true value and each price measure per asset-day given different absolute bid-ask spread values in ticks. As the true value proxy in t, we use the constrained quantity-weighted midpoint in $t + 5s$.

Average absolute deviations (bps)			
Price measure	Bid-ask spread		
	1 tick	2 ticks	3+ ticks
Transaction price	2.47	2.56	3.62
Midpoint	1.65	1.55	2.17
QW midpoint	1.47	1.59	2.41
CQW midpoint	1.47	1.40	2.13
Micro-price	1.54	1.42	2.14

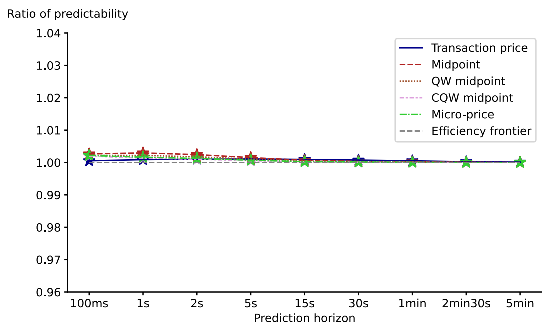
Table B.4: Overview of qualitative and quantitative characteristics of different price measures

	Transaction price	Midpoint	QW midpoint	CQW midpoint	Micro-price
Required data	Trade data	Prices at best bid and offer	Prices and quantities at best bid and offer	Prices and quantities at best bid and offer, asset's tick size	Prices and quantities at best bid and offer
Methodology	Observation	Arithmetic mean	Weighted mean	Weighted mean	Estimation of a Markov model based on past trading days
Discrete vs. continuous values	discrete with changes bounded to one tick	discrete with changes bounded to half a tick	continuous	continuous	discrete with changes bounded to state space in the Markov model
Max. ratio of predictability	1.034	1.022	1.014	1.004	1.005
Minimum time to incorporate past information	30 seconds	60 seconds	15 seconds	2 seconds	2 seconds
Average absolute difference to efficient price	2.68 bps	1.63 bps	1.62 bps	1.47 bps	1.51 bps
Average bias in effective spread	n.a.	-0.97 bps	0.21 bps	n.a.	-0.48 bps

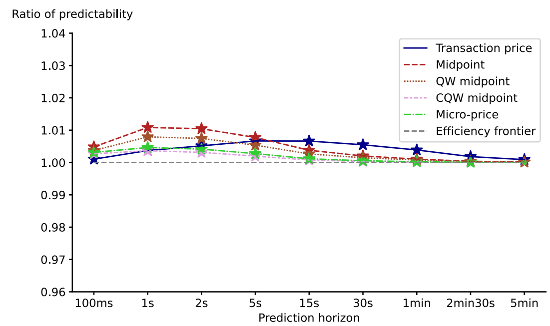
Appendix C. Robustness tests

Appendix C.1. Different benchmark

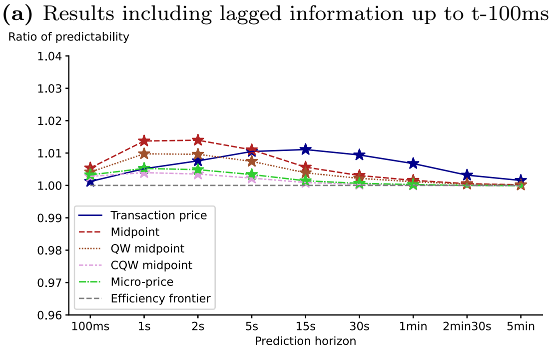
In this section we present the results of Section 4.1 when the ratio of predictability, as described in Equation 7, is calculated using the average return of the previous asset-day as benchmark instead of the no-change benchmark. Therefore, the following results are calculated using $MSE^{mean}_{m,h,k,s,td}$ instead of $MSE^{nc}_{m,h,k,s,td}$ in Equation 7.



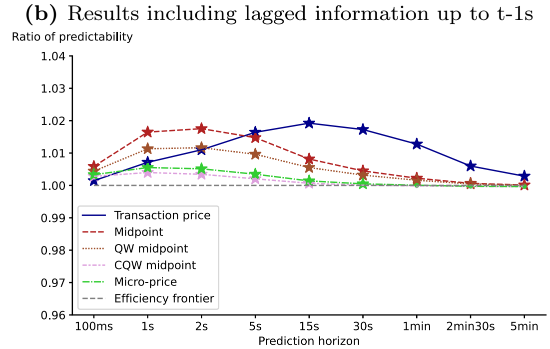
(a) Results including lagged information up to t-100ms



(b) Results including lagged information up to t-1s



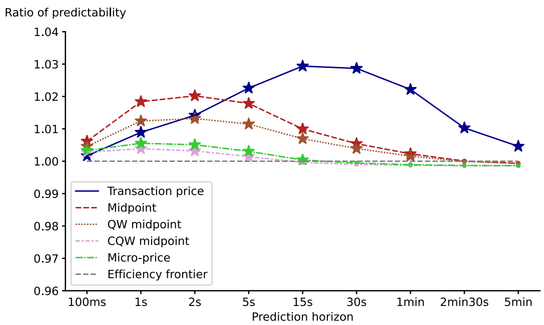
(c) Results including lagged information up to t-2s



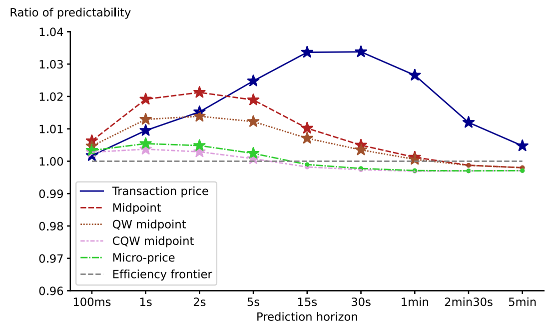
(d) Results including lagged information up to t-5s

Figure C.3: Ratio of predictability of all five price measures at various prediction horizons using past information up to $t - 5s$

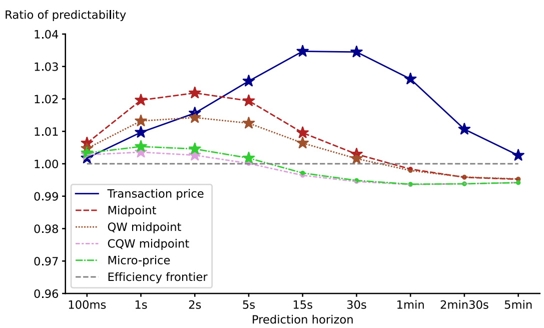
In this figure a star indicates that the ratio of predictability is statistically significantly larger than 1 at the corrected 5% significance level after applying the Bonferroni correction.



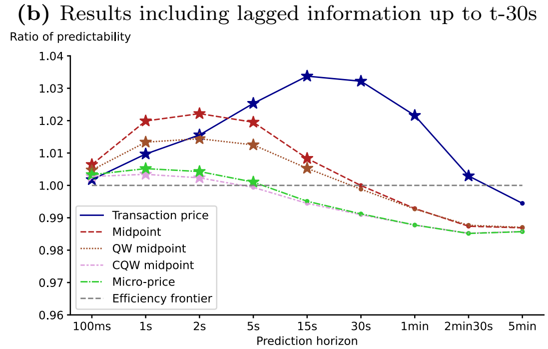
(a) Results including lagged information up to t-15s



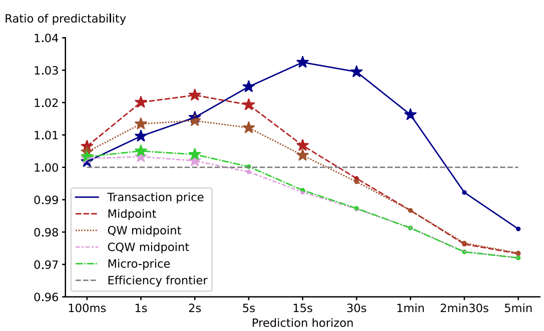
(b) Results including lagged information up to t-30s



(c) Results including lagged information up to t-1min



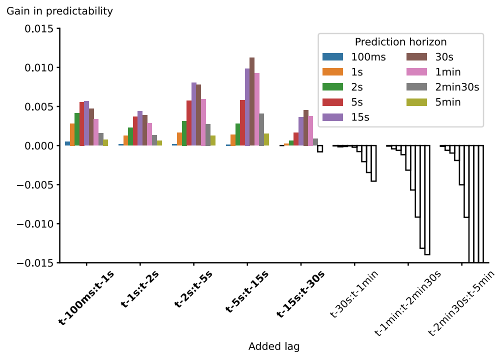
(d) Results including lagged information up to t-2min30s



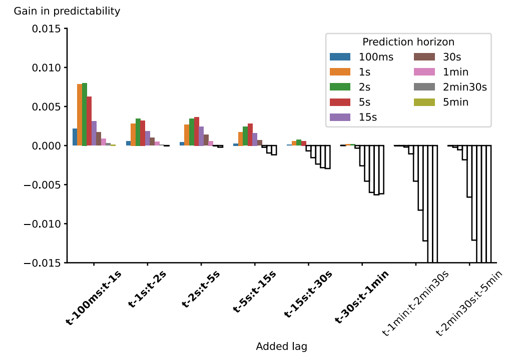
(e) Results including lagged information up to t-5min

Figure C.4: Ratio of predictability of all five price measures at various prediction horizons using past information up to $t - 5min$

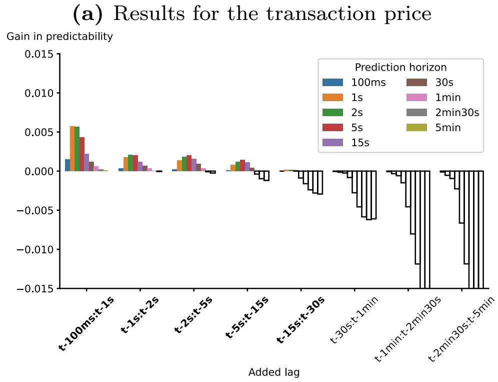
In this figure a star indicates that the ratio of predictability is statistically significantly larger than 1 at the corrected 5% significance level after applying the Bonferroni correction.



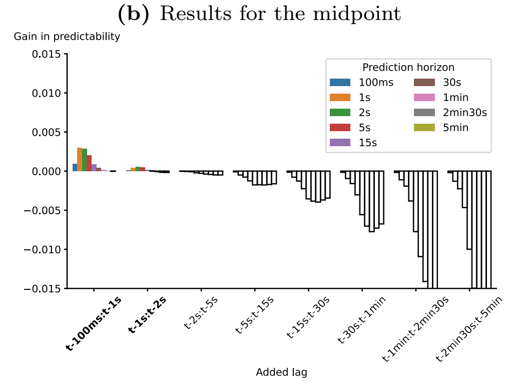
(a) Results for the transaction price



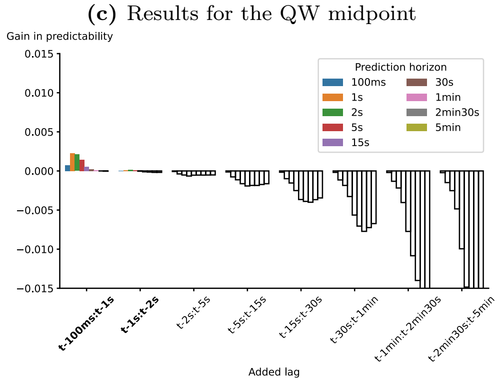
(b) Results for the midpoint



(c) Results for the QW midpoint



(d) Results for the micro-price



(e) Results for the CQW midpoint

Figure C.5: Gain in predictability for all five price measures with varying levels of past information
In this figure the gain in predictability for (a) the transaction price, (b) the midpoint, (c) the QW midpoint, (d) the micro-price, and (e) the CQW midpoint for various prediction horizons is shown. The x-axis shows the level of past information included. A colored bar indicates that the gain in predictability is statistically significantly larger than 0 at the corrected 5% significance level. If the x-axis label is bold, we find a significant gain in the ratio of predictability for at least one prediction horizon. In this figure we calculate the gain in predictability using the mean return of the previous asset-day as benchmark.

Appendix C.2. Micro-price as best true value estimator

In this section we present the results of Section 4.2 if we use the micro-price as the best true value proxy.

Table C.5: Average relative difference between the true value and each price measure

This table shows the average of the difference between the true value and each price measure per asset-day. As the true value proxy in t, we use the micro-price in $t + 5s$. We use a t-test to test whether the average deviation in the second column differs significantly from zero. One star (*), two stars (**), and three stars (***) following the value indicate rejected of the null hypothesis at the 90%, 95%, and 99% confidence levels, respectively.

Price measure	Average absolute deviation (bps)	Average deviation (bps)
Transaction price	2.67	−0.03***
Midpoint	1.54	−0.01***
QW midpoint	1.63	0.00
CQW midpoint	1.49	0.00**
Micro-price	1.43	−0.01***

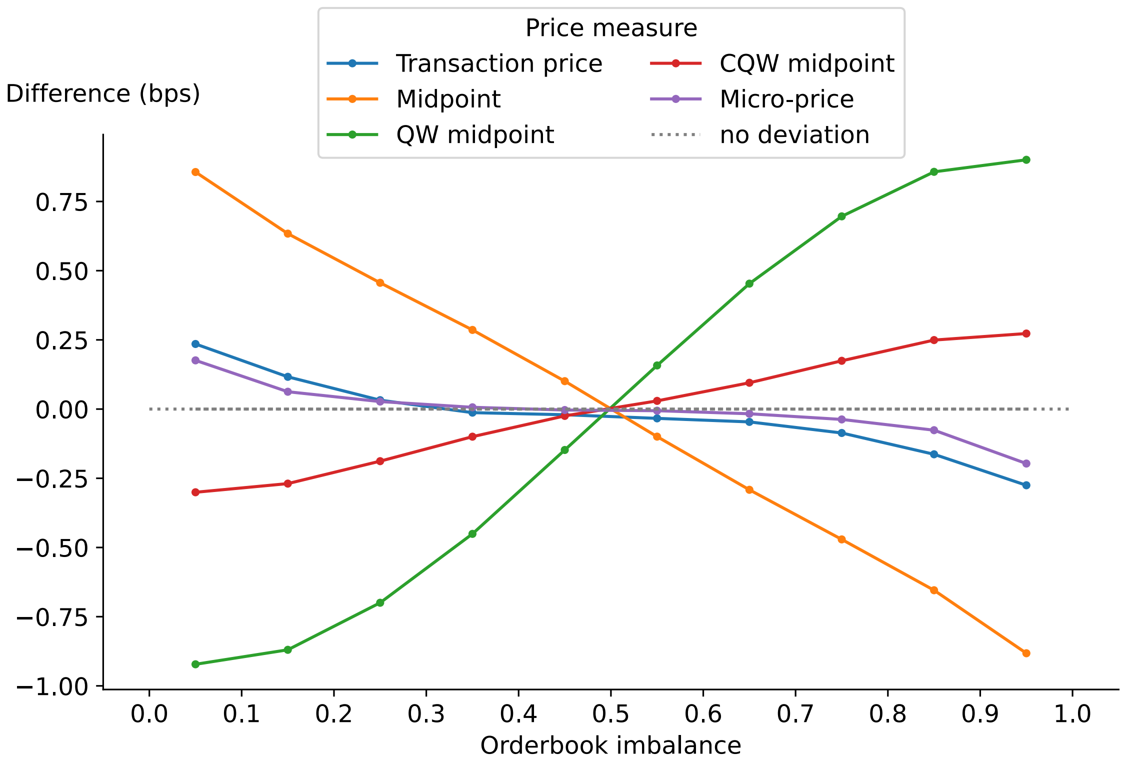


Figure C.6: Average differences between each price measure and the micro-price in $t + 5s$

This figure shows the average difference between the true value and each price measure per asset-day as a function of order book imbalance. The order book imbalance is divided into 10 equally sized deciles. The x-axis represents the mid of each respective decile. As the true value proxy in t, we use the micro-price in $t + 5s$.

Table C.6: Effective spread bias of different price measures compared to the true effective spread

In this table we present the average effective spread across all asset-days when using different price measures as true value estimators. The average bias in the second column is calculated as the difference between the average effective spread in the first column and the true effective spread estimate. We use the effective spread calculated using the micro-price as the true estimate. All effective spread calculations are given in basis points. The relative average bias is then calculated as the average bias divided by the average effective spread and is given in percentage points. We use a t-test to test whether the average bias in the second column differs significantly from zero and report the t-statistics as well as the p-value in the last columns.

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	3.03				
Midpoint	3.52	−0.49	−16.18	−151.64	0.00
QW midpoint	2.34	0.69	22.69	145.52	0.00
CQW midpoint	2.55	0.48	15.92	105.45	0.00

Table C.7: Effective spread bias of different price measures given different spread sizes

In this table we present the average effective spread across all asset-days when using different price measures as true value estimators. The table divides observations of effective spread calculations by different absolute spread sizes: (a) one tick, (b) two ticks, and (c) three or more ticks. The average bias in the second column is calculated as the difference between the average effective spread in the first column and the true effective spread estimate. We use the effective spread calculated using the micro-price as the true estimate. All effective spread calculations are given in basis points. The relative average bias is then calculated as the average bias divided by the average effective spread and is given in percentage points. We use a t-test to test whether the average bias in the second column differs significantly from zero and report the t-statistics as well as the p-value in the last columns.

(a) Effective spread (absolute spread is 1 tick; 65% of all observations)					
Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	2.17				
Midpoint	2.64	−0.47	−21.47	−131.33	0.00
QW midpoint	1.61	0.57	26.07	111.84	0.00
CQW midpoint	1.61	0.57	26.07	111.84	0.00
(b) Effective spread (absolute spread is 2 ticks; 28% of all observations)					
Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	4.86				
Midpoint	5.35	−0.49	−10.13	−126.54	0.00
QW midpoint	4.06	0.80	16.55	119.27	0.00
CQW midpoint	4.71	0.15	3.17	68.75	0.00
(c) Effective spread (absolute spread is 3+ ticks; 7% of all observations)					
Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	8.65				
Midpoint	8.82	−0.18	−2.02	−22.22	0.00
QW midpoint	8.06	0.59	6.79	16.52	0.00
CQW midpoint	8.60	0.05	0.53	14.22	0.00

Table C.8: Price impact and realized spread bias of different price measures at 1s and 60s horizons

This table presents the average price impact and realized spread at 1s and 60s horizons using different true value estimators. Results include the price impact in (a) and (c) as well as realized spread in (b) and (d) at both horizons. The average bias is the difference between the first column and the true estimate, based on the micro-price. The relative bias is expressed in percentage points. A t-test assesses significance, with t-statistics and p-values in the last columns.

(a) Price impact 1s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	3.21				
Midpoint	3.61	−0.40	−12.30	−140.28	0.00
QW midpoint	2.67	0.54	16.93	133.25	0.00
CQW midpoint	2.80	0.41	12.76	102.72	0.00

(b) Realized spread 1s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	−0.18				
Midpoint	−0.09	−0.10	−52.88	−84.66	0.00
QW midpoint	−0.32	0.14	79.95	73.90	0.00
CQW midpoint	−0.25	0.07	40.34	77.27	0.00

(c) Price impact 60s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	3.56				
Midpoint	4.06	−0.50	−14.06	−148.35	0.00
QW midpoint	2.85	0.70	19.77	143.78	0.00
CQW midpoint	3.08	0.48	13.56	105.59	0.00

(d) Realized spread 60s horizon

Price measure	Mean (bps)	Average bias (bps)	Relative average bias (%)	t-stat.	p-value
Micro-price	−0.53				
Midpoint	−0.53	0.01	1.79	14.35	0.00
QW midpoint	−0.51	−0.02	−2.91	−13.71	0.00
CQW midpoint	−0.53	0.00	0.06	0.66	0.51