High-Frequency Price Formation in Fragmented Equity Markets[a]

Tino Cestonaro[b], Jonas De Paolis[c], Sven Panz[d]

January 13, 2023

Abstract

We analyze the price formation process in fragmented equity markets using deep learning-based sequence models for short-term price prediction. We use limit order book information from EURO STOXX 50 constituents to examine the increase in prediction power that models based on multiple trading venues yield over those based on a single trading venue. We find that accounting for limit order book information from multiple trading venues substantially improves prediction accuracy with regard to the market-wide mid-price, but that it does not provide additional predictive power with regard to the mid-price at an individual trading venue. Moreover, our results demonstrate that the relevance and predictive power of limit order book information depends not only on a trading venue's market share, but also on other aspects such as volatility, liquidity, and algorithmic trading activity.

Keywords— Price formation, limit order book, market microstructure, market fragmentation, deep learning, forecasting

JEL classification codes— C45, G17

[a]We thank the participants at the 27[th] International Conference on Forecasting Financial Markets for their insightful and helpful suggestions.

[b]Goethe University Frankfurt, cestonaro@wiwi.uni-frankfurt.de, corresponding author

[c]Goethe University Frankfurt, depaolis@wiwi.uni-frankfurt.de

[d]Goethe University Frankfurt, panz@wiwi.uni-frankfurt.de

1 Introduction

Over the last two decades, financial markets have undergone a substantial amount of structural change as a result of market fragmentation. Before 2000, equity markets in both the United States and the European Union used to be concentrated markets with a low degree of competition. Between 2000 and 2010, additional trading venues entered the market and thereby increased competition among the established trading venues (Petrella, 2010; Gomber et al., 2017). This development can be traced back mainly to regulatory initiatives such as the Regulation National Market System (RegNMS) in the United States and the Markets in Financial Instruments Directive (MiFID I) in the European Union, both aiming to encourage innovation and competition among individual trading venues (Haslag and Ringgenberg, 2022).[1] Moreover, the electronification of financial markets has been an additional catalyst for market fragmentation, and it has allowed for market participants to adopt new business models in the realm of algorithmic and high-frequency trading. Today, market participants submit and cancel orders within microseconds and across multiple trading venues, and the corresponding trading strategies, typically acting on pattern-based signals, have generally become far less fundamentally oriented (O'Hara, 2015).

In light of these changes in both market structure and market activity, the market environment has grown increasingly complex, and the same applies to the price formation process that connects the corresponding trading venues. In theory, price formation describes how investors' information is translated into demand and supply, and then finally into prices (Madhavan, 2000). In practice, the price formation process has traditionally been modeled in restriction to a single limit order book (LOB) representing a single trading venue, thereby implicitly assuming that there exists no relevant interconnection between trading venues (e.g., Cont et al., 2010). While not also mechanically interconnected as in the United States, trading venues in the European Union are indirectly interconnected through market participants engaging in multi-market trading activity across multiple trading venues. Consequently, price formation can be characterized as a global process that depends on a set of local processes, one per trading venue, that are at least indirectly interconnected through trading activity.

With regard to the complexity that underlies the price formation process, it is virtually impossible to express the functional relationship in terms of a closed-form solution approach. Instead, the large amounts of high-dimensional market data available today make it possible to approximate the functional relationship in the form of a deep learning based model. In theory, a neural network is able to approximate an arbitrarily complex and non-linear functional relationship[2] and, consequently, it is able to learn an optimal mapping from market data to a target variable of interest. Ultimately, although we do not entirely understand the price formation process, we can learn the price formation process from large amounts of high-dimensional market data.

[1] See SEC Exchange Act Release No. 34-51808 (9 June 2005) and Directive 2004/39/EC of the European Parliament and of the Council (21 April 2004). For a detailed view on fragmentation development in European equity markets, please see Figure A.1 in the Appendix.

[2] According to the Universal Approximation Theorem (Hornik et al., 1989).

In this paper, we use high-resolution LOB information from multiple trading venues to investigate the price formation process both across and within the respective trading venues by way of short-term price prediction. On this basis, we analyze the relevance of LOB information from individual trading venues in the cross-section, which is of particular interest as it provides insight into the patterns that arise from multi-market trading activity. So far, there does not exist a prior study that uses LOB information from multiple trading venues for price prediction.

In terms of predicting the market-wide mid-price, we find that LOB information from multiple trading venues provides additional predictive power beyond LOB information from a single trading venue (e.g., the primary exchange) only. In this regard, the individual contribution of a given trading venue significantly depends on market share, algorithmic trading activity, and liquidity. In turn, however, our results indicate that LOB information from multiple trading venues is not necessarily helpful in terms of predicting the mid-price on a single trading venue. The implication is that price formation in a fragmented market environment may indeed be characterized as a decentralized process that, in the short term, depends more on the trading venues' individual liquidity rather than on market-wide liquidity. In general, we believe that both researchers and market participants need to carefully evaluate the relevance of individual trading venues that they include in their models, given that it depends not only on market share but also on other factors such as algorithmic trading activity and liquidity of the LOB.

The remainder of this paper is organized as follows. In section 2, we review existing literature on price formation as it is reflected in LOB information, and we derive our research objective that is to model price formation across multiple trading venues in a fragmented market environment. In section 3, we introduce our approach to evaluate based on a set of data-driven models the predictive power of LOB information from multiple trading venues. In section 4, we describe our experimental protocol and the particular data that we use to train our models. In section 5, we describe the experimental results and provide context in relation to the theoretical background. In section 6, we discuss the experimental results, the limitations of our approach, and possible areas of future research. In section 7, we summarize our findings.

2 Theoretical Background

In theory, price formation assumes that the entirety of relevant information be available and thereby abstracts from the complexities underlying information dissemination in financial markets. Given that one can never observe the entirety of relevant information, the notion of a price formation process is therefore purely theoretical in nature. In general, the price formation process of a security can be formulated as

$$p_{t+1} = F\left(D_t, S_t, \epsilon_{t+1}\right), \tag{1}$$

where p_{t+1} denotes the future price of the security, where D_t and S_t denote respectively the demand and supply for the security at time t, representing the available information aggregated over all market participants, and where ϵ_{t+1} denotes an unanticipated innovation in the price, attributable to other factors that are not captured by the model.

In an order-driven market, securities are traded based on a LOB, providing pre-trade transparency and enabling market participants to observe in real-time and in great detail the supply and demand for a given security. Today, most exchanges rely on some variant of a LOB-based market mechanism, the ten largest stock exchanges in the world in terms of market capitalization use a LOB to some degree during their trading hours (Tripathi et al., 2020). The LOB is a wide-spread market mechanism that accumulates liquidity, supply and demand, and that respectively mediates between sellers and buyers based on a matching algorithm. The state of the LOB both captures the current trading intentions and reflects the previous trading decisions of all market participants up until a particular point in time, that is, it displays the historically accumulated and yet to be executed supply and demand for a given security. Consequently, the price formation process can be observed entirely through the history of the LOB, therefore making it an important subject in market microstructure research.

The LOB is organized in bid side and ask side that represent the market participants' buying and selling interest, respectively. Each side encompasses a set of price levels that, according to the principle of price-priority, are sorted based on the most favorable price limit from the perspective of the other side, best bid and best ask. Each price level itself represents a queue of orders that, according to the principle of time-priority, are sorted based on their arrival timestamp.

In an order-driven market, a market participant can submit a limit order or a market order to buy or to sell a security. Given that a market order or a marketable limit order are executed immediately, the LOB comprises only non-marketable limit orders that, depending on the side of the LOB, are sitting at a price level worse than best bid or best ask. Each change in LOB state is a result of an order submission, cancellation, or amendment, and therefore each LOB update fully explains the transition between two corresponding LOB states.

The price formation process is studied by market participants and researchers through (semi-)structured market data that itself is derived from the available LOB information. In general, market data is disseminated in different levels of granularity, also referred as level 1-3, that address different use-cases on the part of the market participant. Level 3 will typically display the LOB in the highest possible resolution and in full depth, that is, containing all orders per price level, and all price levels per side. Level 2 will typically display the LOB in lower resolution and in limited depth, containing only the aggregated quantity across all sitting limit orders per price level, and the n best price levels per side. Level 1 displays only the top-of-book information, that is, the same as level 2 with $n = 1$ best price level(s) per side. Ultimately, the decision between these data formats is a trade-off between informational value and processing cost.

Following Harris (1991), limit orders can be classified into value-driven and precommitted. On the one hand, a value-driven limit order is submitted by an informed investor, reflecting their estimate of the true asset value and thereby revealing private information. If their estimate of the true asset value changes in response to new information, this will lead the informed investor to amend or cancel his limit order. Consequently, both order amendments and order cancellations can be considered as updated information about the true asset value. On the other hand, a pre-committed limit order is submitted by the uninformed investor so as to reduce implicit trading costs, and it will eventually be updated in terms of a more favorable limit, or even replaced by a market order, should it remain unfilled. If their trading intention becomes more aggressive or changes otherwise, this will lead the uninformed investor to cancel or amend (cancel and resubmit) the limit order. Ultimately, either event (submission, amendment, or cancellation) may contain information about the true asset value reflected in the decisions of the informed investor, or the trading intentions of the uninformed investor. The informational value of the LOB is also empirically supported by the studies of Beber and Caglio (2005) and Brogaard et al. (2019) that respectively find that informed traders prefer to hide their information advantage by submitting passive limit orders, and that limit orders sitting in the LOB contribute most to price discovery.

Overall, previous studies suggest that current and historical LOB states may contain valuable information with regard to future price movements, reflecting private information and the trading intentions of uninformed investors, and thereby also the direction of future trades. It is worth noting that the predictive power of lagged LOB information does not necessarily contradict market efficiency - if an investor wants to make use of and trade according to the relationship between LOB information and future price behavior, the price impact of his own orders and market frictions such as trading costs might diminish positive expected trading profits (Huang and Stoll, 1994). In fact, these theoretical assumptions about the predictive power of LOB information are also supported by empirical studies that are presented in more detail below.

In one of the first studies to investigate the predictive power of LOB information with regard to future price development, Huang and Stoll (1994) estimate a set of econometric models that make use of level 1 market data sampled on a five-minute interval. The regression results show that the historical mid-prices as well as the historical differences between best-ask and best-bid volume can partially explain the variation of future mid-prices. Cao et al. (2009) use a similar regression setup that aims to predict five-minute mid-price stock returns, but they also include the imbalance in quantity across the ten best price levels on each side of the LOB, as well as an additional measure to capture the difference among price levels. Again, the results show that quantity and price imbalances in the LOB are related to future short-term returns of the mid-price. Furthermore, Härdle et al. (2012) also finds that order book imbalance can predict future mid-price returns on a five- and one-minute interval. This also holds for imbalance measures capturing price levels deeper in the LOB. In addition to the predictive power of LOB information with regard to future short-term price returns, several studies demonstrate their predictive power with regard to other variables of interest such as future price volatility (Foucault et al., 2007), order choice (Pascual and Veredas, 2009), and liquidity supply (Härdle et al., 2012).

Although these studies provide evidence for the significant predictive power of lagged

LOB information with regard to various target variables of interest, the forecast quality of these models is usually rather low and the models barely outperform their naïve benchmarks. The poor performance of the corresponding models is attributable to the price efficiency of equity markets where equity prices reflect information within short time intervals (Chordia et al., 2005), but also to the inability of econometric models to consider complex and non-linear market dynamics (Sirignano and Cont, 2019).

In recent years, increasingly many studies have tried to address the above-mentioned challenges by employing deep learning techniques so as to predict non-linear price dynamics in equity markets using lagged LOB information.[3] In contrast to linear econometric models, non-linear deep learning models are able to capture the non-linear dynamics of price formation from (semi-)structured market data (De Prado, 2018), using higher-resolution data and yielding substantially higher predictive power (Tsantekidis et al., 2017a; Sirignano and Cont, 2019).

Generally, deep learning uses a neural network to approximate some arbitrarily complex and non-linear function f^* by means of a model $y = f(x; \theta)$ in that it learns the best set of model parameters θ with respect to input samples x and output labels y. More specifically, a neural network is a composition of specialized functions, also referred to as layers, that are able to learn latent feature representations on different levels of abstraction. Ultimately, the neural network based model is able to learn the optimal mapping from input space to output space with respect to a given cost function.

In terms of price prediction based on LOB information, the input space will usually comprise level 2 market data that, although becoming increasingly less important in light of recent advancements in architectural design, may also be further augmented by way of feature engineering. The output space will usually employ a directional label in combination with a classification learning task, that aims to reduce the complexity underlying the true label in combination with a regression learning task. Usually, the authors formulate the prediction problem as a binary (up, down) or ternary classification task (up, down, neutral) to predict the next mid-price change. Given the vast amount of neural network design options, the described models vary in terms of input features, output labels, and especially neural network architecture.

In all of the previously mentioned work, the price formation process has been modeled in restriction to a single LOB from a single trading venue. This is straight-forward as the authors thereby implicitly assume that the LOB represents the self-contained scope across which the price formation process unfolds. In the single-venue setting, as also described by Sirignano and Cont (2019), the price formation process can be characterized as follows:

$$p_{t+1} = F\left(X_t, \epsilon_{t+1}\right), \tag{2}$$

where $X_t \in \mathbb{R}^{m \times n}$ denotes a matrix encompassing a history of the most recent m LOB states, each a vector of size n, which are relevant for the price in $t + 1$. In terms of outcome, there is only a single target variable of interest, that is, p_{t+1} or the local

[3] For an overview see, Zaznov et al. (2022).

best bid and offer (BBO) on that particular trading venue.

Most existing studies that follow the single-venue setting focus on improving existing short-term price prediction models by identifying more suitable neural network architectures and corresponding hyperparameters (e.g., Tsantekidis et al., 2017a or Zhang et al., 2019). While identifying more suitable neural network architectures is a legitimate concern, these studies do not pose an adequate solution from a perspective of market microstructure, given that they do not take into account the effects that market fragmentation has on the structure of information dissemination in financial markets. Despite price prediction and price formation being two sides of the same coin, we want to emphasize the price formation side by ensuring that the effects of market fragmentation are reflected in our model, with regard to all input, output, and architecture.

In a fragmented market environment, as stated previously, the price formation process is decentralized across multiple trading venues that are more or less interconnected, depending i.a. on factors such as market structure. Consequently, a single security is subject to multiple price formation processes that are more or less interconnected and that together form the actual price formation process of which we aim to improve our understanding. In the single-venue setting, we treat a single LOB in isolation and thus do not consider interconnections between trading venues. For this reason, arguably, the single-venue setting accounts for only part of the true price formation process, and we hypothesize that other trading venues are predictive of a given trading venue. In the multi-venue setting, the price formation process can be characterized as a decentralized process between multiple trading venues that are to some extent interconnected:

$$P_{t+1} = \left(p_{t+1}^{(1)}, p_{t+1}^{(2)}, ..., p_{t+1}^{(V)}\right) = F\left(X_t^{(1)}, X_t^{(2)}, ..., X_t^{(V)}, \mathcal{E}_{t+1}\right) \tag{3}$$

where P_{t+1} is a vector of prices in $t+1$, one for each venue $v \in V$ at which a given security is traded. Similarly, $\mathcal{E}_{t+1}$ represents a vector of unanticipated innovations in $t+1$, one for each trading venue $v \in V$, which are attributable to other factors that cannot be captured by the model. Most importantly, $X_t^{(v)}$ is a matrix that represents a history of the most recent LOB states at trading venue v up to time t, similar to equation 2. In terms of outcome, there are two target variables of interest, that is, the global BBO observed across multiple trading venues on the one hand, and the local BBO observed at a single trading venue on the other hand.

In our opinion, accounting for LOB information from multiple trading venues is prerequisite to a good approximation of the decentralized nature of the price formation process in today's fragmented market environment. Several studies show that other trading venues have a minor but significant contribution to price discovery at a given trading venue (Hasbrouck, 1995; Ozturk et al., 2017; Wu et al., 2021), but there does not exist a study that investigates the predictive power of LOB information from multiple trading venues with respect to short-term price changes at one or multiple trading venues. To the best of our knowledge, we are the first research group to take a data-driven approach in order to investigate the interdependencies across multiple trading venues, but within a given security. Rather than simply improving existing forecasting models by identifying more suitable architectures or optimizing hyperparameters of existing models, we aim to answer the following research question:

> *RQ: To what extent does LOB information from multiple trading venues provide additional predictive power to predict short-term mid-price changes?*

3 Methodology

To better understand the price formation process both within a single trading venue and across multiple trading venues, we evaluate the predictive power of single-venue and multi-venue input (features) with regard to single-venue and multi-venue output (labels). We achieve this by means of deep learning-based sequence models that are able to approximate the true price formation process based on large amounts of high-dimensional level 2 market data. To ensure consistency across our models, we keep a fixed architecture (including the same pre-processing steps and hyperparameters) with placeholders for different input and output variants. To account for non-deterministic outcomes in training neural network models, each model is trained and evaluated three times to ensure adequate robustness of our results, and we report the average accuracy across all runs. In this section, we describe the fixed architecture with input and output placeholders on the one hand and, on the other hand, the different input and output variants that are specified later in section 4.

In terms of original data structure, we use level 2 market data where each observation represents a LOB state for a given stock traded on a given venue. In the original data structure, each LOB state is characterized by a set of features that comprises limit price, aggregated quantity, and the number of underlying orders for the $n = 10$ best price levels on either side of the LOB, resulting in a total of 60 features. Additionally, each observation includes a high-resolution timestamp that allows us to determine the order of LOB states across multiple trading venues.

We approach the representation of multi-venue LOB information from a perspective of LOB consolidation, similar in notion to a consolidated tape as it is prevalent in the United States market environment. In general, the consolidated view is achieved by aggregating all LOB states in a way that, for the same price level, both aggregated quantity and the number of underlying orders are added together across a set of trading venues V, resulting in a set of price levels. For each side of the LOB, we limit this set to the ten best price levels. We decide on this representation because it allows us to consolidate information on market-wide liquidity with fixed dimensionality, that is, the number of features per LOB state remaining the same, irrespective of the number of underlying LOB, so that single-venue setting and multi-venue setting remain comparable. This representation serves as a basis for both model input and model output that are outlined in more detail below.

Nevertheless, using this representation comes at the cost of information loss. On the one hand, that is information about the distribution of liquidity across trading venues and, on the other hand, about liquidity on an individual trading venue that is beyond the best n price levels market-wide. It is important to note that there are other solutions to representing LOB information from multiple trading venues[4], but none of them yield better results in terms of out-of-sample prediction accuracy.

Model architecture

Our model architecture employs different types of neural network layers to map the input space onto the output space, given that it needs to first turn time series of high-dimensional market data into a temporal embedding, and then use that temporal embedding to estimate a one-dimensional output label. Similar to a Markov process, a traditional fully-connected neural network layer is incapable of capturing path-dependent information in time series data. Contrarily, a recurrent layer allows for a sequence model of form $(y_t, h_t) = f(x_t, h_{t-1}; \theta)$ where sample x_t and label y_t are observed at time t, and where the internal state h_t reflects information from all previous time-steps $x_{0:t}$ in that it implements a recurrent feedback loop to memorize historical information (Goodfellow et al., 2016). Consequently, the main building block in our deep learning architecture is the LSTM layer (Hochreiter and Schmidhuber, 1997), a type of recurrent layer that we use to compute in latent feature space a temporal embedding that represents the market history provided as model input.

The neural network used in this study is similar to the architecture proposed in Sirignano and Cont (2019) and consists of an LSTM block, a dense block, and an output layer. First, the *LSTM block* comprises three layers, each with 150 units and tanh activation, turning a time series of input vectors into a single embedding vector. Next, the *dense block* comprises a single layer, using 150 units and ReLU activation, adding depth while keeping the latent feature space 150-dimensional. Last, the *output layer* reflects the binary classification learning task, using a single unit and sigmoid activation together with a binary cross-entropy loss function. In previous studies, LSTM-based model architectures trained on LOB information have shown substantial forecasting ability in regard to short-term price prediction (Tsantekidis et al., 2017b; Sirignano and Cont, 2019; Zhang et al., 2019).

Model training

In our model training, we follow a universal approach that is similar to the work of Sirignano and Cont (2019), and according to which each model is trained on all stocks included in the training data. Not only does this approach improve prediction accuracy, but this form of data pooling also allows for a more general view on price formation in fragmented equity markets. The model of a universal price formation process combines all stock-specific price formation processes and therefore both abstracts from stock-specific idiosyncrasies and generalizes to the entire universe reflected in the training data. From the data pool, we randomly sample without replacement individual time series and feed them as input to the model. We use a heuristic approach to deal with larger-than-memory time series data, a detailed explanation of which can be found in

[4] For example, some of those solutions are concatenating or one-hot-encoding LOB states from multiple trading venues, or consolidating LOB states under consideration of the original distribution of liquidity across multiple trading venues.

the Appendix A.1

It is important to mention that the notion of a traditional epoch does not apply in our case, given that we do not actually use every available sample due to the size of our universal data pool. We define an epoch as 100,000 training examples (3,125 mini-batches with 32 samples each), and we run up to 200 epochs per training job, that is, up to 20,000,000 samples in total. After each epoch the model is evaluated out-of-sample based on 2,000 mini-batches. Moreover, we perform early stopping after 20 consecutive epochs without improvement in the target metric that is the validation loss.

Model input

We use as model input time series that are based on different variants of market data, single-venue and multi-venue, that are outlined in more detail below. Irrespective of a particular variant, the model input can be formally described as a time series

$$X_t^k = X_{(t-k-1):t} = (x_{(t-k-1)}, ..., x_{(t-1)}, x_t) \tag{4}$$

where x_t denotes the LOB state at time t, and where k denotes the overall length of the time series. We set the length of the time series to $k = 100$ steps in all experiments, meaning that the model learns from the last 100 LOB states, similar to what has been used in previous studies (e.g., Tsantekidis et al., 2017b; Zhang et al., 2019). With regard to level 2 market data, we generally prefer to avoid feature engineering[5] altogether as, provided a sufficiently capable sequence model architecture, a neural network should theoretically itself be able to extract the most meaningful features (with regard to a given learning task) in latent feature space (Goodfellow et al., 2016), across both spatial and temporal dimension. Consequently, each model receives as input time series of observations $X_{t-99:t}$, that is, the last 100 LOB states each consisting of 60 features.

In general, using a data pool results in high variability across different stocks, in terms of all price, quantity, and the number of underlying orders. In light of our universal approach, however, we benefit from the corresponding heterogeneity in our training data, and we employ several techniques in our pre-processing that address this challenge. Similar to the approach proposed by Tashiro et al. (2019), we normalize prices by their absolute distance to the mid-price (mid) divided by the tick-size (ts).[6]

$$x_{i,t}^{norm} = \frac{x_{i,t} - mid_{i,t}}{ts_{i,t}} \tag{5}$$

Thereafter we clip the prices at a maximum distance of 25 ticks. While, regarding the ten best price levels on either side of the LOB, adjacent price levels will typically be only one or two ticks apart, we use the clipping to control for potential outliers. Similar

[5] Note that we make a distinction between data engineering, pre-processing, and feature engineering. Data engineering refers to the process that transforms the original data into a suitable data representation. Pre-processing refers to the process that transforms the data representation into suitable model input. And feature engineering refers to the introduction of additional features in order to boost model performance.

[6] Note that a tick represents the smallest possible increment between two adjacent price levels.

to many related studies (e.g., Ntakaris et al., 2019), we normalize both quantity and the number of underlying orders by first clipping at the 1% and 99% percentile and thereafter applying z-score standardization. Both the percentiles as well as mean and standard deviation are calculated stock-wise on the basis of the last five trading days.

$$x_{i,t}^{norm} = \frac{x_{i,t} - \hat{x}_{i,t}}{\sigma_{x_{i,t}}} \tag{6}$$

We explicitly refrain from using z-score standardization on prices so as to avoid capturing weekly trends.

In the *single-venue setting*, we include LOB information from only a single trading venue, therefore each observation conforms to only a single LOB state as contained in the original data structure. For $\|V\| = 1$ trading venue, the consolidation procedure is obsolete. In the *multi-venue setting*, we include LOB information from multiple trading venues, therefore each observation conforms to a consolidated LOB state and requires upfront data engineering. For $\|V\| > 1$ trading venues, we use the previously described consolidation procedure to process our data. It is important to note that, keeping the length k of the time series fixed, each additional trading venue will result in the shortening of the time period that is reflected in the time series, in a way that is approximately proportional to its market share in a given stock. In the *multi-venue leave-one-out setting*, we include LOB information from all except one trading venues, following the same approach as in the multi-venue setting. We use this setting only to evaluate the contribution of the left-out trading venue in the multi-venue setting.

Model output

We use as model output a directional label that reflects the next mid-price change in different variants, single-venue and multi-venue, that are outlined in more detail below. The model output conforms to the target variable that we intend to predict and, analogous to the model input, it is processed based on an underlying set V of trading venues. The directional label can be formally described as

$$y_t = \begin{cases} 1, & P_{\tau_{t+1}} - P_{\tau_{t-1}} > 0 \\ 0, & otherwise \end{cases} \tag{7}$$

where the values 0 and 1 denote a negative and a positive mid-price change, respectively, where P denotes the mid-price, and where τ_{t+1} and τ_{t-1} denote the timestamp of both next and previous mid-price change, respectively. The model output reflects the binary classification learning task and can be formally described as

$$\hat{y}_t = f(X_t^k) = \mathbb{P}\left[y_t > 0 \mid X_t^k\right] = \mathbb{P}\left[(P_{\tau_{t+1}} - P_{\tau_{t-1}}) > 0 \mid X_{(t-k-1):t}\right] \tag{8}$$

where $\hat{y}_t$ denotes a probabilistic value. We utilize this form of binary label as the distribution of this label has favorable properties with regard to its use in a universal model. Primarily, using a binary label (up, down) to reflect the overall next mid-price change approximately results in an equal distribution between observations of upward and downward price movement, whereas using a ternary label (up, down, neutral) to reflect a potential mid-price change at the next event results in most observations falling into the neutral class. Also, using a label that implicitly abstracts from the underlying time horizon (in terms of both clock-time and event-time) results in different markets remaining comparable, whereas using a fixed-time label results in divergent distributions for different stocks on different trading venues.

In the *single-venue setting*, we consider LOB information from only a single trading venue, therefore the label is computed directly on a single LOB state as contained in the original data structure, the single-venue input. The single-venue output displays the direction of the next mid-price change with regard to the individual LOB of a given trading venue. In the *multi-venue setting*, we consider LOB information from multiple trading venues, therefore the label is computed on a consolidated LOB state, the multi-venue input. The multi-venue output displays the direction of the next mid-price change with regard to the consolidated LOB across multiple trading venues.

Following our approach, it is important to note that model input and model output each use a different sampling frequency. On the one hand, the model input encompasses all available observations, given that those carry the most informational value and therefore yield better model performance than for instance only subsampling on each mid-point change. On the other hand, the model output considers only each mid-point change, given that binary classification of the next mid-point change (up, down) yields a more balanced class distribution than ternary classification of the next overall change (up, down, neutral). Having introduced the architecture and defined both input and output variants, we will outline our experimental setup built atop these in the next section.

4 Experimental Setup and Data

To assess the informational value of LOB information from multiple trading venues in a fragmented market environment, we train multiple neural network based models that share the same architecture, but differ with respect to model input and model output. Each model is trained on the entire stock universe included in our dataset, given that this universal approach produces much more robust and accurate results than a stock-specific approach (Sirignano and Cont, 2019) and only differs with regard to the included trading venues. We address our research question by comparing the predictive power of both multi-venue and single-venue input with regard to both multi-venue (Experiment 1) and single-venue output (Experiment 2).

Experiment 1

To assess the additional predictive power that multi-venue LOB information has over single-venue LOB information with respect to a multi-venue label, we train three types of models.

In order to predict the next market-wide mid-price (multi-venue output), we first train a model based on LOB information from multiple trading venues (multi-venue input), hereinafter referred to as *multi-venue model* (model 1.1). This model takes a consolidated view of the market, similar in notion to a consolidated tape. The learning task remaining the same, we then train a model based on LOB information from only one of the trading venues (single-venue input), hereinafter referred to as *single-venue model* (model 1.2). This model takes the traditional view restricted to a single trading venue. In comparison between the multi-venue model and all single-venue models, we evaluate (venue- and stock-wise) the difference in out-of-sample prediction accuracy. In regard to the results, a positive difference indicates that information from additional trading venues provides additional predictive power in forecasting the market-wide mid-price.

The learning task remaining the same, we last train a model based on LOB information from all except one of the trading venues (multi-venue leave-one-out input), hereinafter referred to as *multi-venue leave-one-out model* (model 1.3). This model takes a reduced consolidated view based on only a subset of trading venues. In comparison between the multi-venue model and all multi-venue leave-one-out models, we evaluate (venue- and stock-wise) the difference in out-of-sample prediction accuracy. In doing so, we measure the contribution of LOB information from the left-out trading venue with regard to the overall predictive power when forecasting the market-wide mid-price.

In order to assess stock- or venue-specific effects, we lastly estimate cross-sectional regressions to explain the difference in out-of-sample prediction accuracy based on the trading venues' market share (for the respective stock), the level of algorithmic trading activity, stock price volatility and liquidity.

Experiment 2

To assess the additional predictive power that multi-venue LOB information has over single-venue LOB information with respect to a single-venue label, we train two types of models.

To predict the next mid-price on a particular trading venue (single-venue output), we train a model based on LOB information from multiple trading venues (multi-venue input), hereinafter referred to as *multi-venue model* (model 2.1). This model takes a consolidated view of the entire market. The learning task remaining the same, we then train a model based on LOB information from only one of the trading venues (single-venue input), hereinafter referred to as *single-venue model* (model 2.2). This model takes the traditional view restricted to a single trading venue. In comparison between the multi-venue model and all single-venue models, we evaluate (venue- and stock-wise) the difference in out-of-sample prediction accuracy. In regard to the results, a positive difference indicates that information from additional trading venues provides additional predictive power in forecasting the mid-price on a particular trading venue.

Again, we lastly estimate cross-sectional regressions to explain the difference in out-of-sample forecasting accuracy based on the trading venues' market share (for the respective stock), the level of algorithmic trading activity, stock price volatility and liquidity.

Data set

In this study, we rely on level 2 market data from Thomson Reuters Tick History, the data structure of which is described in more detail in section 2. Besides a millisecond-based timestamp, the feature set comprises price, quantity, and the number of underlying orders for the $n = 10$ best price levels on both bid and ask side. Given that we use the timestamp only for data engineering purposes[7] and do not include it in the model itself, this results in a total of 60 features per LOB state. Our dataset comprises data for all constituents of the EURO STOXX 50 index, for all trading days between January 1 and March 31 2019. This amounts to a total of 50 stocks that used to constitute the index during that period. However, we ultimately include only 46 stocks, given that 4 stocks (CRH, Linde, Nokia, and Unilever) are subject to issues such as deviating trading currencies and availability restrictions.

To investigate the price formation process in a fragmented market environment and to assess the relevance of multiple trading venues, we use LOB data from the five most relevant trading venues per stock in terms of continuous trading volume. Despite our selected equity universe being multi-national, the fragmentation patterns across all stocks are identical, meaning that each stock exhibits the highest lit trading volume on its primary stock exchange (36% - 70%) while the majority of the remaining lit volume is traded on Aquis (AQX), BATS (BS), Chi-X (CHI) and Turquoise (TQ). The included main trading venues are Euronext Amsterdam, Brussels, Milan and Paris, XETRA (Frankfurt) and Bolsa de Madrid. On average, the main trading venue and the four alternative trading venues together account for more than 97% of the lit trading volume of the included stocks. The stocks in the sample and their corresponding market shares in continuous trading across all five trading venues are shown in table A.1 in the Appendix.

At each trading venue in our sample, equities are traded from 08:00 until 16:30 UTC. Given that we focus on continuous trading exclusively, we exclude the first and last 30 minutes of each trading day in order to diminish the impact that opening and closing auction have on market liquidity. To further set a lower bound on market liquidity in our experiments, we exclude LOB states with three price levels or less on each side of the LOB. In the prediction of mid-price changes across multiple trading venues based on high-resolution market data, timestamp precision is paramount. We use a heuristic approach to ensure accurate and synchronized timestamps across all trading venues, a detailed explanation of which can be found in the Appendix A.2. Overall, our dataset consists of more than 1.3 billion LOB states after filtering, the distribution of which (per trading venue and across stocks) is shown in Table 1. The summary statistics highlight the variation in the number of observations across different stocks.

We train our models using the first nine trading weeks in our data set (January 2 to March 1 2019). The next two weeks (March 4 to March 15 2019) are used as validation set, while the remaining two trading weeks (March 18 to March 29) are used as test set. For each stock on every trading venue, the data set comprises between 343,000 and 35,500,000 LOB states (see table 1). Out of this immense amount of observations in our data set, we use only a fraction to train our models. Although the total number of

[7] See Appendix A.2.

Table 1: Number of observations

This table shows for each of the five trading venues the number of observations across stocks (in millions). The mean as well as the percentiles provide information about the variation across the stocks in our sample. MAIN refers to the primary exchange of the respective stock.

	Mean	Minimum	25%-percentile	50%-percentile	75%-percentile	Maximum	Sum
AQX	3.43	0.34	2.14	3.17	4.47	7.38	161.10
BS	3.44	0.95	2.08	2.91	4.30	9.37	161.91
CHI	4.63	1.51	2.85	3.81	6.43	11.02	217.39
MAIN	13.20	3.65	7.25	11.98	17.43	35.50	620.22
TQ	3.05	0.95	1.66	2.58	4.25	7.88	143.39
Sum							1,304.01

training samples thereby appears rather small compared to the total number of available observations, our models generally converge reliably within 200 epochs. This can be partly traced back to the training setup for which each mini-batch randomly selects observations from the large data pool as described in the Appendix A.1.

5 Results

In this section, we report the results of our experiments described in the previous chapter.

Prerequisites

Before we report the results, we describe both the scatter plots that we use for visualization, and the regressions that we use to make our results interpretable. The performance of each model is depicted as a point in a scatter plot with x-axis and y-axis representing the out-of-sample prediction accuracy of the single-venue model and the multi-venue model, respectively. While the color of the point represents the individual trading venue that serves as single-venue input, the size of the point is proportional to the respective trading venue's market share for a given stock in terms of trading volume during continuous trading. The diagonal in the scatter plot divides the scatter in half, observations above and below the diagonal indicate superior and inferior performance of the multi-venue model compared to the single-venue model, respectively.

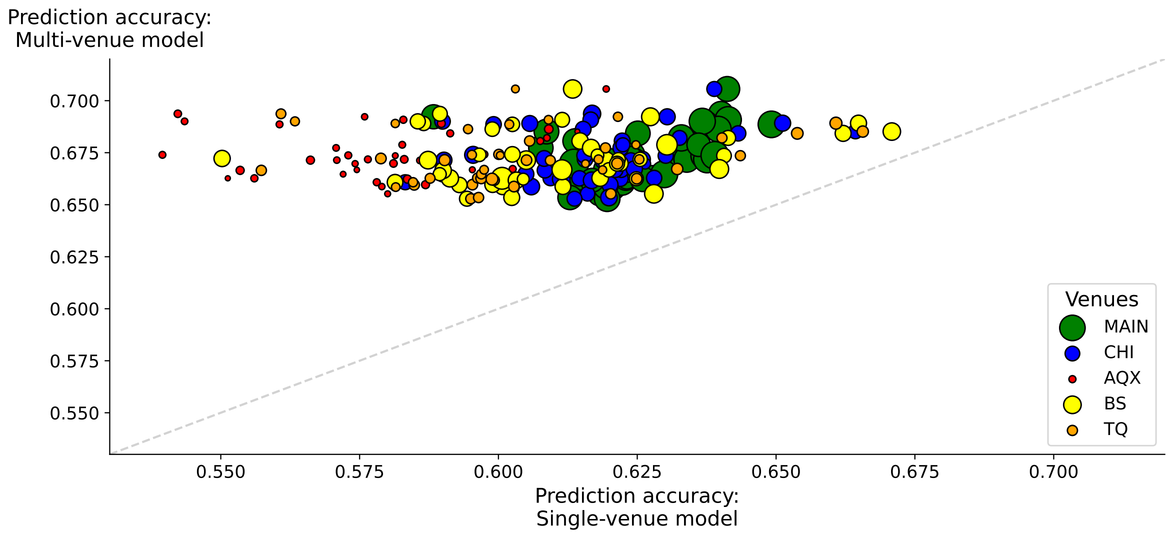
To make the results of our deep learning models more interpretable, we further estimate cross-sectional regressions to understand which factors mainly contribute to the predictive power of LOB information with regard to the next market-wide mid-price. In the regressions, we focus on the impact of fragmentation and algorithmic trading, but also consider factors such as volatility and liquidity. As the dependent variable, we use the difference in prediction accuracy between multi-venue model and each single-venue model in percentage points. As the set of independent variables that may potentially explain any disparity in model performance, we use a set of proxy variables for market share, algorithmic trading activity, volatility, and liquidity. More precisely, the variable MS is used to describe a given trading venue's market share for a given stock in terms of trading volume during continuous trading (in percentage points). The variable OTR is used to describe the algorithmic trading activity within a LOB in terms of order-to-trade-ratio, that is, the number of messages divided by the number of trades per trading day.[8] The variable RV is used to describe the (daily) volatility in terms of range volatility according to Parkinson (1980), scaled by $\sqrt{252}$ for annualization (in percentage points). And, lastly, the variable QS is used to describe liquidity in terms of relative quoted spread, that is, the difference between best bid and best ask price divided by the mid-price (in basis points). All measures are calculated on a daily basis over our sample period that is January to March 2019, used in the cross-section regressions is the average of the daily values across all three months. We winsorize the variables at the 1st and 99th percentiles. We then take the natural logarithm of all independent variables to control for the skewness of various features (see Appendix A.2) and to improve the interpretability of regression coefficients irrespective of the underlying units. Lastly, we introduce additional dummy variables for the trading venues and stocks in our sample to cover venue- and stock-specific factors beyond market share, algorithmic trading activity, volatility, and liquidity. Summary statistics of the independent variables are provided in the Appendix A.2.

Experiment 1

In experiment 1, we examine the relevance of multi-venue LOB information when predicting the market-wide mid-price. More precisely, we start by comparing the out-of-sample prediction accuracy of the multi-venue model (model 1.1) against the out-of-sample prediction accuracy of the five single-venue models (model 1.2). The scatter plot in figure 1 compares for each stock traded on each venue the prediction accuracy for both multi-venue model and all single-venue models. On the one hand, the multi-venue model (model 1.1) yields values between 65.3% and 70.6%, with a mean of 67.3% and a standard deviation of 1.2%. On the other hand, the five single-venue models (model 1.2) yield values between 54.0% and 67.1%, with a mean of 60.9% and a standard deviation of 2.4%. Among the single-venue models, the model trained on LOB information from the main trading venue performs the best, with a mean of 62.5%. In contrast, the model trained on LOB information from the smallest trading venue Aquis performs the worst, with a mean of 58.2%. The mean values of the remaining single-venue models range between 60.5% and to 61.7%.

Consequently, the multi-venue model yields more consistent results and also outperforms the average prediction accuracy of the single-venue models by 6.3 percentage points. As highlighted by the diagonal, the multi-venue model yields the highest prediction accuracy for each stock in the sample, indicating that multi-venue LOB information is highly relevant when predicting the market-wide mid-price. Based on the results, it further appears that LOB information from a trading venue with a higher market share has also more predictive power with regard to the market-wide mid-price. However, a few instances exist where a lower market share does not go along with a lower prediction accuracy, some single-venue models yield average values above 66%. Consequently,

[8] The order-to-trade ratio is a widely used proxy for algorithmic and high-frequency trading activity. A higher order-trade-ratio indicates higher algorithmic trading intensity. For instance, see Brogaard et al. (2014) or Kemme et al. (2022).



The figure shows the out-of-sample prediction accuracy of the single-venue models on the x-axis and the multi-venue model on the y-axis when predicting the market-wide mid-price. While the multi-venue model is trained on the consolidated LOB of all five trading venues, the single-venue model is trained on the LOB of one trading venue only. The color of the dots indicates the trading venue that serves as input for the single-venue model. The dot size increases with the trading venue's market share, in terms of lit trading volume for the corresponding stock.

Figure 1: Out-of-sample prediction accuracy regarding the market-wide mid-price for multi-venue and single-venue models

these observations indicate that the market share does not completely explain the difference in prediction accuracy.

In the regressions, we investigate the previously introduced factors (independent variables) and their correlation to the difference between multi-venue and single-venue prediction accuracy (dependent variable). The regression results presented in table 2 show that a trading venue's market share is significantly negatively correlated to the difference between multi-venue and single-venue prediction accuracy during continuous trading. This confirms that the additional predictive power that multi-venue information has over single-venue information with regard to the market-wide mid-price decreases with an increase in the trading venue's market share, the trading venue being the same as that used as single-venue input. Depending on the respective model, a 1% increase in market share implies an increase in the difference in prediction accuracy of 0.013 to 0.020 percentage points. Although this magnitude seems rather small, a doubling in market share thereby implies a decrease in the difference in prediction accuracy of up to 1.42 percentage points (ceteris paribus). In terms of short-term price prediction, this is to be considered economically meaningful. There exist also other independent variables that are significantly correlated with the difference in prediction accuracy. For example, the algorithmic trading activity is significantly positively correlated with the dependent variable in all regressions when controlling for stock-specific effects. An increase in OTR of one percent implies an increase in the difference in prediction accuracy of up to 0.022 percentage points. We therefore conclude that LOB with higher algorithmic trading activity have less predictive power with regard to the market-wide mid-price. In terms of volatility, the regression results are rather ambigu-

Table 2: Explaining the difference between multi-venue and single-venue models' prediction accuracy regarding the market-wide mid-price

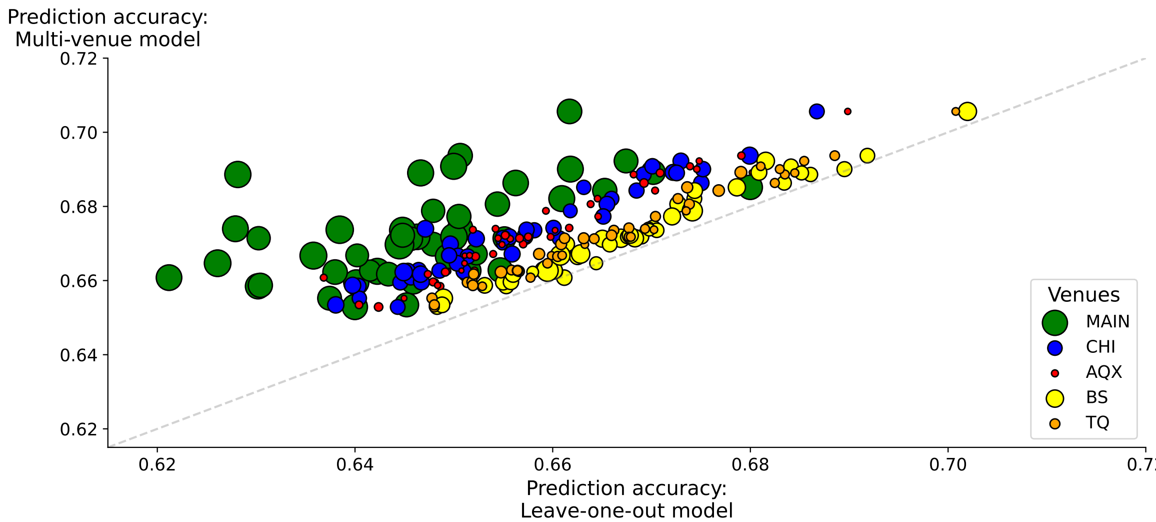
This table shows the results of the cross-section regressions that explain the difference between multi-venue and single-venue model prediction accuracy regarding the market-wide mid-price (in percentage points). Venue dummies (D_v) are four dummy variables for each of the five trading venues in our sample, except for the main trading venue. Stock dummies (D_s) are 45 dummy variables for each of stocks in our sample, except for one. MS refers to the logarithm of a trading venue's market share in terms of continuous trading volume across all five trading venues. OTR refers to the logarithm of the order-to-trade ratio of the respective trading venue. RV is the logarithm of the range volatility of the stock (in percentage points). QS is the logarithm of the quoted spread (in basis points). N is the number of observations. Newey-West HAC standard errors are provided in brackets below the coefficients.

	D_v	D_s	MS	OTR	RV	QS	N	R^2	Adj. R^2
(1)	Yes	No	−1.439**				230	0.366	0.352
			(0.638)						
(2)	No	Yes	−1.552***				230	0.718	0.647
			(0.136)						
(3)	Yes	Yes	−1.533***				230	0.755	0.687
			(0.494)						
(4)	Yes	No		0.176			230	0.352	0.337
				(0.434)					
(5)	No	Yes		2.156***			230	0.548	0.435
				(0.281)					
(6)	Yes	Yes		−0.030			230	0.741	0.668
				(0.664)					
(7)	Yes	No	−1.313*	0.439	−1.507***	2.560***	230	0.449	0.429
			(0.670)	(0.529)	(0.580)	(0.507)			
(8)	No	Yes	−1.761***	0.584**	13.500***	−0.589	230	0.770	0.708
			(0.184)	(0.245)	(4.396)	(0.532)			
(9)	Yes	Yes	−2.046***	−0.450	12.225***	−1.176	230	0.789	0.726
			(0.632)	(0.721)	(4.111)	(1.420)			

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

ous. According to regression (7), an increase in volatility leads to a decrease in the difference between multi-venue and single-venue prediction accuracy. However, when controlling for stock-specific factors, our results indicate that multi-venue LOB information becomes more important in light of higher volatility. Finally, regression setup (7) suggests that information from less liquid LOB has less predictive power with respect to the market-wide mid-price. However, when controlling for stock-specific effects, the coefficient becomes insignificant.

Subsequently, we continue by comparing the out-of-sample prediction accuracy of the multi-venue model (model 1.1) against the out-of-sample prediction accuracy of all five multi-venue leave-one-out models (model 1.3). This is important to evaluate the relevance of single-venue LOB information from the left-out trading venue when predicting the market-wide mid-price. The scatter plot in figure 2 compares for each stock traded on each venue the prediction accuracy for both multi-venue model and all multi-venue leave-one-out models. As already mentioned, the multi-venue model (model 1.1) yields values between 65.3% and 70.6%, with a mean of 67.3% and a standard deviation of 1.2%. On the other hand, the multi-venue leave-one-out models (model 1.3) yields value between 62.1% and 70.2%, with a mean of 65.9% and a standard deviation of 1.4%. Among the multi-venue leave-one-out models, the model that excludes LOB



The figure shows the out-of-sample prediction accuracy of the multi-venue leave-one-out models on the x-axis and the multi-venue model on the y-axis when predicting the market-wide mid-price. While the multi-venue model is trained on the consolidated LOB of all five trading venues, the multi-venue leave-one-out models are trained on a reduced consolidated LOB that covers only four out of the five trading venues. The color of the dots indicates the trading venue that does not serve as input to the consolidated LOB. The dot size increases with the market share of the excluded trading venue, in terms of lit trading volume for the corresponding stock.

Figure 2: Out-of-sample prediction accuracy regarding the market-wide mid-price for multi-venue and multi-venue leave-one-out models

information from the main trading venue performs the worst, with a mean of 64.7%. In contrast, the model trained on multi-venue LOB information from all venues except BATS performs the best, with a mean of 66.8%. The mean of each leave-one-out model that excludes LOB information from Aquis, Chi-X, and Turquoise is 65.8%, 65.7%, and 66.6%, respectively.

Based on these results, it generally appears that LOB information from all trading venues has predictive power with regard to the market-wide mid-price, each stock is more accurately predicted using the multi-venue model than using the corresponding multi-venue leave-one-out model. Again, the contribution of a given trading venue with regard to overall predictive power also appears to be correlated with its market share, given that the multi-venue leave-one-out models that exclude LOB information from either the main trading venue or the second largest trading venue Chi-X show the most substantial decrease in prediction accuracy. However, a given trading venue's market share appears not to explain completely the variation in its contribution, given that the smallest trading venue Aquis turns out to be the third most relevant trading venue (out of five) when predicting the market-wide mid-price. The multi-venue leave-one-out model that excludes LOB information from Aquis yields a decrease in prediction accuracy by 1.4 percentage points, whereas excluding the other trading venues BATS or Turquoise yields a decrease by 0.5 and 0.7 percentage points, respectively. It is for this reason that there need to exist additional factors that contribute to the predictive power of LOB information from a particular trading venue.

In the regressions, we investigate the previously introduced factors (independent

Table 3: Explaining the difference between multi-venue and multi-venue leave-one-out models' prediction accuracy regarding the market-wide mid-price

This table shows the results of the cross-section regressions that explain the difference between multi-venue and multi-venue leave-one-out model prediction accuracy regarding the market-wide mid-price (in percentage points). Venue dummies (D_v) are four dummy variables for each of the five trading venues in our sample, except for the main trading venue. Stock dummies (D_s) are 45 dummy variables for each of stocks in our sample, except for one. MS refers to the logarithm of a trading venue's market share in terms of continuous trading volume across all five trading venues. OTR refers to the logarithm of the order-to-trade ratio of the respective trading venue. RV is the logarithm of the range volatility of the stock (in percentage points). QS is the logarithm of the quoted spread (in basis points). N is the number of observations. Newey-West HAC standard errors are provided in brackets below the coefficients.

	D_v	D_s	MS	OTR	RV	QS	N	R^2	Adj. R^2
(1)	Yes	No	0.480***				230	0.665	0.657
			(0.107)						
(2)	No	Yes	0.428***				230	0.264	0.079
			(0.083)						
(3)	Yes	Yes	0.771***				230	0.770	0.706
			(0.183)						
(4)	Yes	No		−0.116			230	0.655	0.647
				(0.139)					
(5)	No	Yes		0.151*			230	0.097	-0.130
				(0.086)					
(6)	Yes	Yes		−0.732***			230	0.772	0.708
				(0.261)					
(7)	Yes	No	0.438**	−0.100	0.237	−0.445***	230	0.681	0.670
			(0.179)	(0.198)	(0.177)	(0.165)			
(8)	No	Yes	0.220**	0.164	0.677	−3.509***	230	0.719	0.643
			(0.087)	(0.109)	(1.559)	(0.293)			
(9)	Yes	Yes	0.324	−0.602**	−1.981	−0.933	230	0.793	0.731
			(0.213)	(0.301)	(1.878)	(0.624)			

Note: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

variables) and their correlation to the difference between multi-venue and multi-venue leave-one-out prediction accuracy (dependent variable). The regression results presented in table 3 show that a trading venue's market share is significantly positively correlated to the difference between multi-venue and multi-venue leave-one-out prediction accuracy. This indeed indicates that a trading venue with a higher (lower) market share has also more (less) predictive power with regard to the market-wide mid-price. However, as mentioned previously, the market share is not sufficient to explain all the variation in the difference in prediction accuracy. The algorithmic trading activity poses another significant factor that determines the predictive power of LOB information from a particular trading venue. Regression setup (5) indicates that a higher level of algorithmic trading activity in the left-out LOB results in a larger difference between multi-venue and multi-venue leave-one-out prediction accuracy. However, the negative adjusted R^2 shows that the regression model does not adequately fit the data and needs to be interpreted with caution. The models that consider both stock- and venue-specific factors indicate a significant negative relationship between predictive power and algorithmic trading activity. The relationship is statistically significant when controlling for both stock-specific and venue-specific factors. Consistent with the previous results in table 2, these results suggest that LOB information from a trading venue with lower algorithmic trading activity yields more predictive power.

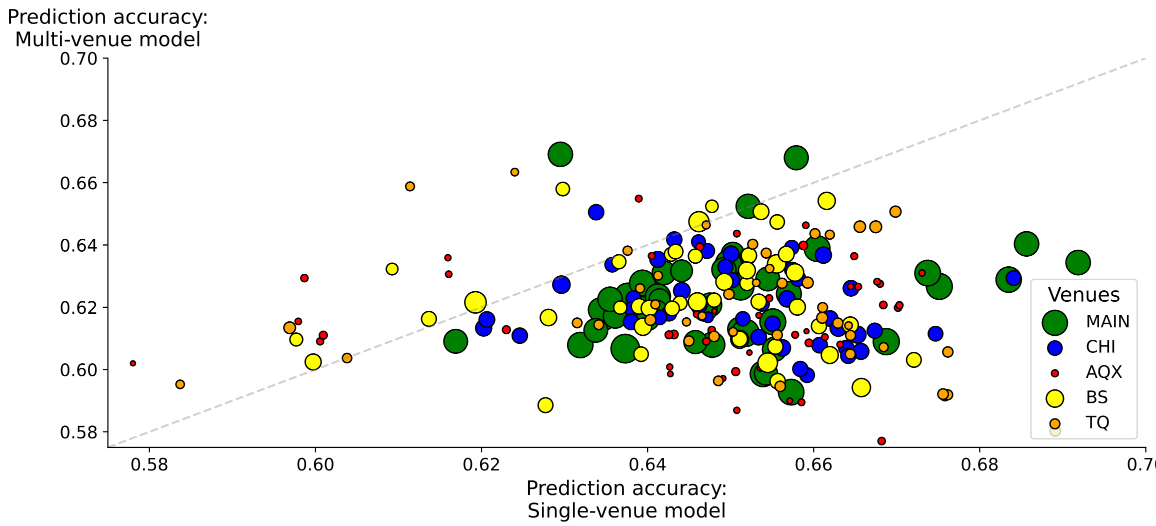
While we do not find evidence that volatility poses a significant factor that determines the predictive power of LOB information with regard to the market-wide mid-price, the estimated coefficients associated with the quoted spread show that left-out LOB information from a trading venue with higher liquidity leads to a higher difference between multi-venue and multi-venue leave-one-out prediction accuracy. The latter finding indicates that a higher level of liquidity in a given LOB results in more predictive power with respect to the market-wide mid-price.

Experiment 2

In experiment 2, we examine the relevance of multi-venue LOB information when predicting the mid-price on an individual trading venue. More precisely, we compare the out-of-sample prediction accuracy of the multi-venue model (model 2.1) against the out-of-sample prediction accuracy of the five single-venue models (model 2.2). Note that each single-venue model is used to predict its own mid-price across all stocks, whereas the multi-venue model is evaluated across all trading venues and all stocks. The scatter plot in figure 3 compares for each stock traded on each venue the prediction accuracy for both multi-venue model and the single-venue models. The multi-venue model (model 2.1), on the one hand, yields values between 57.1% and 66.9%, with a mean of 62.0% and a standard deviation of 1.7%. The average of all single-venue models (model 2.2), on the other hand, yields values between 57.8% and 69.2%, with a mean of 64.8% and a standard deviation of 1.8%. In restriction to the largest trading venue in terms of market share (the main trading venue), the multi-venue model yields the highest mean value of 62.2%. In restriction to the smallest trading venue Aquis, however, the multi-venue model yields the lowest mean value of 61.7%. With regard to the single-venue models, the models yield the highest mean value for Turquoise (65.1%) and the lowest mean value for BATS (64.5%).

In contrast to experiment 1, only in a few instances is the multi-venue model able outperform the single-venue models with regard to the mid-price on the respective trading venue, the set of single-venue models generally performing better. More precisely, the single-venue models outperform the multi-venue model 89.6% of the time. This indicates that multi-venue LOB information, compared to single-venue LOB information, may introduce additional noise that negatively affects prediction accuracy for a particular trading venue. The multi-venue model yields similar performance for each of the trading venues, and most of the small variation in prediction accuracy is explained by the trading venues' market share. This is evident from the multi-venue model performing better for the main trading venue and worse for the smallest trading venue Aquis. Overall, all five single-venue models yield values that are low in variation across trading venues, the existing variation not being explained by market share.

In the regressions, we investigate the previously introduced factors (independent variables) and their correlation to the difference between multi-venue and single-venue prediction accuracy (dependent variable). The (adjusted) R^2 value in table 4 shows that the regression models are less capable of explaining the difference between multi-venue and single-venue prediction accuracy compared to experiment 1. Regression



The figure shows the out-of-sample prediction accuracy of the single-venue models on the x-axis and the multi-venue model on the y-axis when predicting the mid-price on an individual trading venue. While the multi-venue model is trained on the consolidated LOB of all five trading venues, the single-venue model is trained on the LOB of one trading venue only. The color of the dots indicates the trading venue that serves as input for the single-venue model. The dot size increases with the respective trading venue's market share, in terms of lit trading volume for the corresponding stock.

Figure 3: Out-of-sample prediction accuracy regarding the individual trading venue mid-price for multi-venue and single-venue models

setup (1) indicates that a higher market share in a given trading venue results in the single-venue model outperforming the multi-venue model, whereas regression setup (8) points to the opposite direction. Moreover, the signs of the insignificant coefficients of MS vary across the different regressions setups. Consequently and contrary to the previous analyses on market-wide mid-price prediction, our analyses do not provide definitive evidence for a significant (positive) relationship between market share and predictive power. On the other hand, the coefficients for algorithmic trading activity are as expected and indicate a positive correlation with the difference between multi-venue and single-venue prediction accuracy. These results imply that multi-venue information becomes more important relative to single-venue information if the single-venue LOB is subject to higher algorithmic trading activity. Moreover, liquidity significantly affects the difference in prediction accuracy. Again, the results are consistent with the previous regression setups and demonstrate that more liquid LOB have higher predictive power with respect to the mid-price on the individual trading venues. Finally, the results also suggest that, for volatile stocks, multi-venue information becomes more relevant.

Summary

On the one hand, we find that multi-venue LOB information may generally provide additional predictive power, especially with regard to the market-wide mid-price. In this setting, the multi-venue model is generally able to outperform both the respective single-venue and multi-venue leave-one-out model. Consequently, we provide evidence that the multi-venue perspective is relevant for the market-wide price formation process. In general, we are able to show that LOB information from all trading

Table 4: Explaining the difference between multi-venue and single-venue models' prediction accuracy regarding the mid-price on individual trading venues

This table shows the results of the cross-section regressions that explain the difference between multi-venue and single-venue model prediction accuracy regarding the mid-price on individual trading venues (in percentage points). Venue dummies (D_v) are four dummy variables for each of the five trading venues in our sample, except for the main trading venue. Stock dummies (D_s) are 45 dummy variables for each of stocks in our sample, except for one. MS refers to the logarithm of a trading venue's market share in terms of continuous trading volume across all five trading venues. OTR refers to the logarithm of the order-to-trade ratio of the respective trading venue. RV is the logarithm of the range volatility of the stock (in percentage points). QS is the logarithm of the quoted spread (in basis points). N is the number of observations. Newey-West HAC standard errors are provided in brackets below the coefficients.

	D_v	D_s	MS	OTR	RV	QS	N	R^2	Adj. R^2
(1)	Yes	No	−1.832**				230	0.043	0.022
			(0.887)						
(2)	No	Yes	0.161				230	0.401	0.251
			(0.207)						
(3)	Yes	Yes	−0.920				230	0.422	0.261
			(0.898)						
(4)	Yes	No		0.397			230	0.022	-0.000
				(0.591)					
(5)	No	Yes		0.064			230	0.398	0.247
				(0.395)					
(6)	Yes	Yes		0.965			230	0.424	0.263
				(0.972)					
(7)	Yes	No	−0.918	0.029	4.292***	0.075	230	0.204	0.175
			(0.896)	(0.578)	(0.690)	(0.594)			
(8)	No	Yes	0.548**	0.897**	2.061	2.962***	230	0.443	0.291
			(0.273)	(0.402)	(4.415)	(0.821)			
(9)	Yes	Yes	0.325	0.741	0.091	6.611***	230	0.467	0.306
			(0.873)	(0.963)	(4.617)	(1.915)			

Note: $^{*} p < 0.1$; $^{**} p < 0.05$; $^{***} p < 0.01$

venues in our sample provides additional predictive power with regard to the market-wide mid-price. Cross-sectional regressions further demonstrate that the importance of multi-venue LOB information is significantly linked to the degree of fragmentation, algorithmic trading activity, liquidity, and volatility. The relevance of LOB information from a particular trading venue is significantly linked to its market share, as well as to the degree of algorithmic trading intensity and liquidity. On the other hand, we find that multi-venue LOB information may in some settings also introduce additional noise that worsens the performance compared to single-venue LOB information, especially when predicting the mid-price on that particular trading venue. In this setting, the single-venue model is generally able to outperform the multi-venue model.

Postrequisites

After we report the results, we describe several ways in which we ensure adequate robustness of our results. Due to large-scale training data and the corresponding computational cost, our robustness checks are limited to only a few but nonetheless meaningful experiments. As mentioned before, we conduct each experiment three times and report for each model the average prediction accuracy across all three runs. This way, we ensure that our findings are robust in spite of non-deterministic outcomes observed in terms of model training. Of particular interest is the standard deviation of the prediction accuracy within a given stock but across multiple runs, because it shows whether our results vary across runs in the cross-section of stocks. Table A.4 in the Appendix reports for each stock the standard deviation across multiple runs, ultimately aggregated across all stocks in the form of a mean value. The values vary between and 0.15 and 0.84 percentage points, the majority of values lying below 0.3 percentage points. These values are rather low and therefore highlight that all models in and of themselves yield consistent values across both multiple runs, not only in aggregate, but also when accounting for the cross-section of stocks. In regard to the number of LOB states per unit of time, the multi-venue input has a higher information density than the single-venue input, and the single venue model is therefore effectively trained on a larger window in terms of clock time. To verify that this is not the reason for the single-venue model (model 2.2) outperforming the multi-venue model (model 2.1), we repeat experiment 2 and set the input length for the multi-venue model to 500 observations, keeping the input length for the single-venue model at 100 observations. This does not change our results.

6 Discussion

In experiment 1, we investigate the extent to which multi-venue LOB information is relevant when predicting the market-wide mid-price. With respect to our research question (see chapter 2), we finally conclude that LOB information from multiple trading venues provides additional predictive power with regard to short-term changes in the market-wide mid-price. The predictive power of LOB information from an individual trading venue is significantly positively correlated with market share and liquidity. Furthermore, the algorithmic trading activity appears to negatively affect predictive power. Given that we are the first research group to evaluate the difference in predictive power between single-venue and multi-venue LOB information, our results have several implications for market microstructure research.

We find that the market share of a trading venue (in a given stock) is an important criterion in selecting LOB information to explain the market-wide mid-price, which is in line with previous studies on price discovery across multiple markets (e.g., Hasbrouck (1995) or Ozturk et al. (2017)). It is important to note, however, that these studies investigate a different market environment (United States) and rely on a lower timestamp resolution (up to one-minute intervals). Moreover, we also find that the degree of algorithmic trading activity observed in LOB information is negatively correlated with its predictive power. This result is in line with previous studies such as Weller (2018) which find that algorithmic trading activity one the one hand increases price efficiency with respect to acquired information, but on the other hand decreases the amount of information in the market to which prices respond. In addition, we find that the higher the liquidity in the LOB, the less informative is multi-venue information with regard to short-term changes in the market-wide mid-price. While, in the short-term, predictive power is generally lower for more liquid stocks (Chordia et al., 2008), we provide evidence that a liquid LOB yields more predictive power than an illiquid LOB given the same prediction task.

Our results indicate that it is highly important to consider information from multiple trading venues when modeling the price formation process in a fragmented market environment. That is, LOB information from each additional trading venue is able to improve the approximation of the price formation process as reflected in the multi-venue output. To achieve the maximum in predictive power when observing only a subset of trading venues, however, the selection criteria should exceed a trading venue's market share and also include its liquidity, volatility, and the level of algorithmic trading activity.

From a practitioner perspective, our results indicate that the subscription of data feeds from multiple trading venues may provide an edge for more sophisticated market participants that are able to trade on low-latency signals. For market participants that engage in multi-market strategies, LOB information from each additional trading venue may provide additional predictive power with regard to the market-wide mid-price and is therefore worth considering. However, as our research currently ignores latency across different trading venues, this predictive power may not actually be exploitable in practice. In light of the plans of the European Commission (European Commission, 2020) to introduce a consolidated tape in the European Union market environment, multi-venue LOB information may in the future be available to everyone, and it may thereby affect the predictive power that this information currently holds.

In experiment 2, we investigate the extent to which multi-venue LOB information is relevant when predicting the mid-price on a particular trading venue. With respect to our research question (see chapter 2), we finally conclude that LOB information from multiple trading venues generally does not provide additional predictive power with regard to short-term changes in the mid-price on a given trading venue. More precisely, this holds for 89.6% of the instances in our sample, each instance being a given stock traded on a given venue. The variation in the difference in predictive power between multi-venue and single-venue prediction accuracy cannot be adequately explained by market share, algorithmic trading activity, liquidity, and volatility.

Our initial characterization of price formation is a global process that depends on a set of local processes (one per trading venue), indirectly interconnected through trading activity (see chapter 2). In our results, however, we find that LOB information from other trading venues generally does not have predictive power with regard to a given trading venue. Consequently, these results suggest that the interconnection between trading venues is rather weak, meaning that the multi-market trading activity of market participants does not introduce significant patterns, at least not in the timeframe that we consider (in the realm of milliseconds). In this short timeframe, price formation is rather only a set of local processes that in themselves explain most of their own future price development. In contrast to this result, the studies of Clapham and Zimmermann (2016) and Ozturk et al. (2017) find that each trading venue in their sample contributes at least marginally to the price determination at one of the other trading venues. Note, however, that these studies focus on price discovery rather than on price prediction. This discrepancy may be due to the fact that price discovery does not necessarily imply price predictability (Narayan and Smyth, 2015), and that their results on low-resolution market data may not be transferable to high-resolution market data (Hasbrouck, 2021). Surprisingly, we do not find consistent evidence that the difference in prediction accuracy between single-venue and multi-venue model increases with a trading venue's market share, which is a common finding in the literature on price discovery (Ozturk et al., 2017).

Moreover, we provide evidence that multi-venue information is more important if the considered single-venue is subject to a high level of algorithmic trading activity. A possible explanation for this finding is a phenomenon referred to as *ghost liquidity* that can occur in a fragmented market environment with a high level of algorithmic trading activity (Degryse et al., 2021).[9] In essence, this phenomenon implies that, provided a high level of algorithmic trading activity, the liquidity observed on a given trading venue is more likely to be mirrored on other trading venues and to be cancelled if an order is filled at one of the other trading venues. Consequently, the LOB of a given trading venue may not adequately represent the actual level of supply and demand for a security and lead to overly optimistic or pessimistic price predictions. In addition, we find that multi-venue information becomes more relevant in light of higher volatility. Note, however, that in the existing literature, there are mixed results on whether the contribution that multiple trading venues have with regard to price discovery depends on volatility (e.g., Comerton-Forde and Putniņš, 2015; Ghadhab and Hellara, 2016). Finally, and consistent with our previous results on market-wide mid-price prediction, we find that multi-venue LOB information becomes less relevant when predicting the mid-price change on a given trading venue with high liquidity.

This result suggests that the price in the most liquid LOB is primarily driven by its own level of supply and demand, which is plausible since the liquid LOB covers most of the market-wide level of demand and supply, and thus provides the most information about future trading intentions.

Our results have implications for practitioners and regulators. For high-frequency trading firms that engage in market making strategies at only a single trading venue, our results suggest that the subscription of other trading venues' market data feeds is likely not going to provide additional informational value, and it may even introduce additional noise. However, multi-venue LOB information may become more relevant in the future, should the European Commission (European Commission, 2020) go through with their plans on a consolidated tape. A pre-trade tape in combination with a trade-through rule would mechanically interconnect the individual trading venues and thereby unify the decentralized nature of price formation in the European Union market environment. However, experts raise concerns about the feasibility of a pre-trade tape given the latency issues that result from the geographical distance between European trading venues (Peterhoff et al., 2021).

In conclusion, our results are consistent with previous studies such as Tsantekidis

[9] According to Degryse et al. (2021), ghost liquidity means that identical limit orders are sitting simultaneously at multiple trading venues as they are submitted by market participants in the context of a multi-market strategy. If and order is filled at one of the trading venues, its duplicates on the remaining trading venues are cancelled and, consequently, multi-venue LOB information may consistently overestimate the market-wide liquidity that is actually available.

et al. (2017b) or Sirignano and Cont (2019) that use deep learning models for short-term price prediction, demonstrating that LOB information has significant predictive power. However, our setting cannot be perfectly compared to other studies, given that it differs in terms of learning task (ternary classification instead of binary classification), sampling frequency of the model input (subsampled instead of event-based), and forecasting horizon. Especially in terms of forecasting horizon, most studies use a fixed amount of time, either event-time (number of events) or clock-time (e.g., number of seconds), instead of the time of the next mid-price change. Our study extends the literature in that it demonstrates that multi-venue and single-venue output are best predicted based on multi-venue and single-venue input, respectively. In other words, this means that multi-venue input is beneficial only if the objective is to predict also multi-venue output. As mentioned before, this does not necessarily contradict market efficiency since market frictions can make it impossible to exploit the prediction results in terms of a profitable trading strategy. Although our study does not control for market frictions such as latency or trading costs, our results suggest that a high-frequency trading firm engaging in multi-market trading strategies would likely be successful in building profitable trading strategies, using an accurate short-term price prediction model as a trading signal.

Finally, we discuss the overarching limitations of our study and point out potential directions for future research. First, our data engineering approach both consolidates and aggregates level 2 market data from multiple trading venues, thereby resulting in an overall information loss. On the one hand, the consolidation leads to the exclusion of all liquidity that does not contribute to the overall ten best price levels per side. This challenge cannot really be overcome as the main purpose of consolidation is to achieve this effect. On the other hand, aggregation abstracts from the distribution of liquidity across trading venues. This challenge can be overcome by keeping the individual liquidity of each trading venue, that is, not aggregating quantity and the number of underlying orders per consolidated price level.

Second, our model architecture is based on an LSTM-block that cannot adequately model temporal and spatial structure that are reflected in our data. On the one hand, we cannot use our architecture to capture the spatial structure of the market environment through interconnections between the individual trading venues (and between stocks). This challenge can be overcome by using a graph neural network architecture to learn the spatial embedding from our data. On the other hand, we cannot use our architecture to capture our time series in event-time as it presumes an equidistant temporal structure. This challenge can be overcome by expanding the graph neural network architecture to work on a continuous-time dynamic graph, similar to the work of Rossi et al. (2020), that would learn based on the actual timestamp of each observation.

Last, our general approach assumes zero latency between trading venues. Consequently, our models potentially cannot capture patterns reflective of multi-market trading activity, for the simple reason that market participants are not fast enough. This challenge can only be overcome by introducing additional data engineering and making assumptions about latency based on the geographical locations of both trading venues and market participants.

7 Conclusion

The price formation process in financial markets is non-trivial since demand and supply for a given security are split across multiple trading venues in a fragmented market environment, with changes in liquidity occurring on a nanosecond-basis. In consequence, we suggest that both market participants and researchers need to take into consideration the entirety of demand and supply, consolidated across all trading venues, in order to explain both global and also any local mid-price dynamics.

In this study, we investigate price formation in the European Union market environment by evaluating the predictive power of LOB information with regard to the future mid-price development, using level 2 market data from the five most relevant trading venues for EURO STOXX 50 constituents. More precisely, we use LOB consolidation to generate an adequate data representation, and then train a set of deep learning models that differ only with respect to input and output, each based on either a single trading venue or multiple trading venues. In terms of model output, we predict the next mid-price change either locally, at each individual trading venue (single-venue), or market-wide, across multiple trading venues (multi-venue). In terms of model input, we use regular LOB information from a single trading venue (single-venue), consolidated LOB information from all trading venues (multi-venue), and consolidated LOB information from all except one trading venues (multi-venue leave-one-out).

By analyzing our results, we find that training models on LOB information from additional trading venues substantially increases the out-of-sample prediction accuracy with regard to the market-wide mid-price. In this setting, the predictive power of LOB information from an additional trading venue is significantly positively correlated with the trading venue's market share and its liquidity. The reflected level of algorithmic trading activity has a significant negative impact on the predictive power of LOB information. With regard to the mid-price on an individual trading venue, however, LOB information from other trading venues introduces additional noise for most stocks in our sample and therefore generally does not improve prediction accuracy. This result indicates that price formation at an individual trading venue is rather isolated from other trading venues in the short-term. However, our experiments do not consider latency between different trading venues as this would require us to introduce additional assumptions. Overall, our results are helpful for both researchers and practitioners that seek to model or predict price formation in a fragmented market environment, and that need to decide which trading venues are worth considering.

Our findings are a first step towards a deeper understanding of price formation in a fragmented market environment, and we encourage the research community to transfer our approach to a different equity universe, to other asset classes, and to a different model architecture. In our opinion, of particular interest would be a comparison between the European Union and United States market environment that may have important implications for the introduction of a consolidated tape in the European Union.

References

Beber, A. and Caglio, C. (2005). Order submission strategies and information: Empirical evidence from the NYSE. In *EFA 2003 Annual Conference Paper*.

Brogaard, J., Hendershott, T., and Riordan, R. (2014). High-frequency trading and price discovery. *The Review of Financial Studies*, 27(8):2267–2306.

Brogaard, J., Hendershott, T., and Riordan, R. (2019). Price discovery without trading: Evidence from limit orders. *The Journal of Finance*, 74(4):1621–1658.

Cao, C., Hansch, O., and Wang, X. (2009). The information content of an open limit-order book. *Journal of Futures Markets: Futures, Options, and Other Derivative Products*, 29(1):16–41.

Chordia, T., Roll, R., and Subrahmanyam, A. (2005). Evidence on the speed of convergence to market efficiency. *Journal of Financial Economics*, 76(2):271–292.

Chordia, T., Roll, R., and Subrahmanyam, A. (2008). Liquidity and market efficiency. *Journal of Financial Economics*, 87(2):249–268.

Clapham, B. and Zimmermann, K. (2016). Price discovery and convergence in fragmented securities markets. *International Journal of Managerial Finance*, 12(4):381–407.

Comerton-Forde, C. and Putniņš, T. J. (2015). Dark trading and price discovery. *Journal of Financial Economics*, 118(1):70–92.

Cont, R., Stoikov, S., and Talreja, R. (2010). A stochastic model for order book dynamics. *Operations research*, 58(3):549–563.

De Prado, M. L. (2018). *Advances in financial machine learning*. John Wiley & Sons.

Degryse, H., De Winne, R., Gresse, C., and Payne, R. (2021). Cross-venue liquidity provision: High frequency trading and ghost liquidity. *Working Paper*.

European Commission (2017). Commission Delegated Regulation (EU) 2017/574 of 7 June 2016 supplementing Directive 2014/65/EU of the European Parliament and of the Council with regard to regulatory technical standards for the level of accuracy of business clocks. *Official Journal of the European Union L87*, 60:147–151.

European Commission (2020). Capital markets union 2020 action plan: A capital markets union for people and businesses.

Foucault, T., Moinas, S., and Theissen, E. (2007). Does anonymity matter in electronic limit order markets? *The Review of Financial Studies*, 20(5):1707–1747.

Ghadhab, I. and Hellara, S. (2016). Price discovery of cross-listed firms. *International Review of Financial Analysis*, 44:177–188.

Gomber, P., Sagade, S., Theissen, E., Weber, M. C., and Westheide, C. (2017). Competition between equity markets: A review of the consolidation versus fragmentation debate. *Journal of Economic Surveys*, 31(3):792–814.

Goodfellow, I., Bengio, Y., and Courville, A. (2016). *Deep learning*. MIT press.

Härdle, W. K., Hautsch, N., and Mihoci, A. (2012). Modelling and forecasting liquidity supply using semiparametric factor dynamics. *Journal of Empirical Finance*, 19(4):610–625.

Harris, L. (1991). *Liquidity, Trading Rules, and Electronic Trading System*. Monograph Series in Finance and Economics. New York University-Salomon Center.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The Journal of Finance*, 50(4):1175–1199.

Hasbrouck, J. (2021). Price discovery in high resolution. *Journal of Financial Econometrics*, 19(3):395–430.

Haslag, P. H. and Ringgenberg, M. C. (2022). The demise of the NYSE and NASDAQ: Market quality in the age of market fragmentation. *Journal of Financial and Quantitative Analysis. Forthcoming.*

Hochreiter, S. and Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8):1735–1780.

Hornik, K., Stinchcombe, M., and White, H. (1989). Multilayer feedforward networks are universal approximators. *Neural Networks*, 2(5):359–366.

Huang, R. D. and Stoll, H. R. (1994). Market microstructure and stock return predictions. *The Review of Financial Studies*, 7(1):179–213.

Kemme, D. M., McInish, T. H., and Zhang, J. (2022). Market fairness and efficiency: Evidence from the Tokyo Stock Exchange. *Journal of Banking & Finance*, 134:106309.

Madhavan, A. (2000). Market microstructure: A survey. *Journal of Financial Markets*, 3(3):205–258.

Narayan, S. and Smyth, R. (2015). The financial econometrics of price discovery and predictability. *International Review of Financial Analysis*, 42:380–393.

Ntakaris, A., Mirone, G., Kanniainen, J., Gabbouj, M., and Iosifidis, A. (2019). Feature engineering for mid-price prediction with deep learning. *IEEE Access*, 7:82390–82412.

Ozturk, S. R., van der Wel, M., and van Dijk, D. (2017). Intraday price discovery in fragmented markets. *Journal of Financial Markets*, 32:28–48.

O'Hara, M. (2015). High frequency market microstructure. *Journal of Financial Economics*, 116(2):257–270.

Parkinson, M. (1980). The extreme value method for estimating the variance of the rate of return. *Journal of business*, pages 61–65.

Pascual, R. and Veredas, D. (2009). What pieces of limit order book information matter in explaining order choice by patient and impatient traders? *Quantitative Finance*, 9(5):527–545.

Peterhoff, D., Wuergler, T., and Burkl, M. (2021). Caught on tape - a consolidated tape for europe. Technical report, Oliver Wyman.

Petrella, G. (2010). MiFID, Reg NMS and competition across trading venues in Europe and the USA. *Journal of Financial Regulation and Compliance*, 18(3):257–271.

Rossi, E., Chamberlain, B., Frasca, F., Eynard, D., Monti, F., and Bronstein, M. (2020). Temporal graph networks for deep learning on dynamic graphs. *arXiv preprint arXiv:2006.10637*.

Sirignano, J. and Cont, R. (2019). Universal features of price formation in financial markets: Perspectives from deep learning. *Quantitative Finance*, 19(9):1449–1459.

Tashiro, D., Matsushima, H., Izumi, K., and Sakaji, H. (2019). Encoding of high-frequency order information and prediction of short-term stock price by deep learning. *Quantitative Finance*, 19(9):1499–1506.

Tripathi, A., Null, V., and Dixit, A. (2020). Limit order books: A systematic review of literature. *Qualitative Research in Financial Markets*, 12(4):505–541.

Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., and Iosifidis, A. (2017a). Forecasting stock prices from the limit order book using convolutional neural networks. *2017 IEEE 19th Conference on Business Informatics (CBI)*, 1:7–12.

Tsantekidis, A., Passalis, N., Tefas, A., Kanniainen, J., Gabbouj, M., and Iosifidis, A. (2017b). Using deep learning to detect price change indications in financial markets. *2017 25th European Signal Processing Conference (EUSIPCO)*, pages 2511–2515.

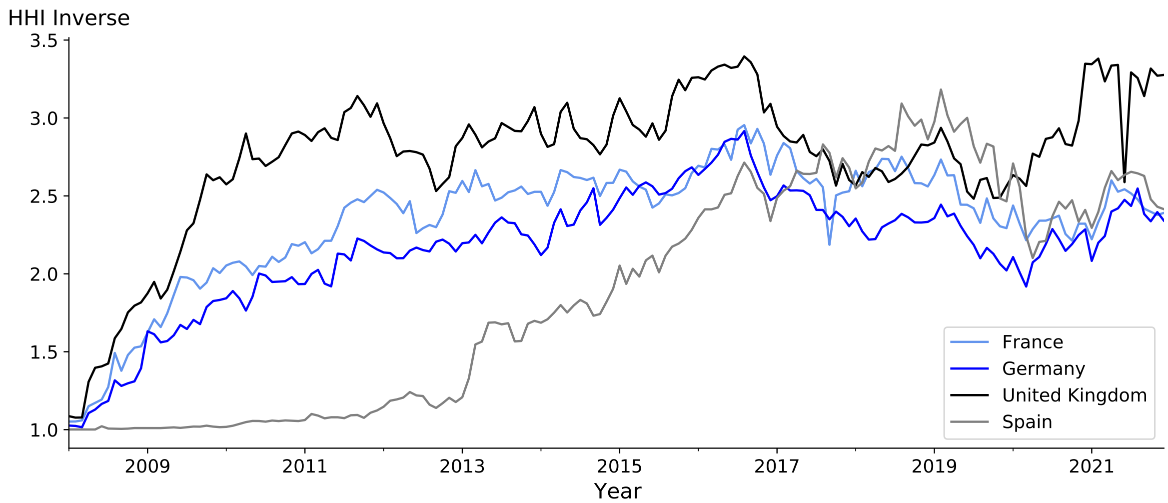
Weller, B. M. (2018). Does algorithmic trading reduce information acquisition? *The Review of Financial Studies*, 31(6):2184–2226.

Wu, L., Xu, K., and Meng, Q. (2021). Information flow and price discovery dynamics. *Review of Quantitative Finance and Accounting*, 56(1):329–367.

Zaznov, I., Kunkel, J., Dufour, A., and Badii, A. (2022). Predicting stock price changes based on the limit order book: a survey. *Mathematics*, 10(8):1234.

Zhang, Z., Zohren, S., and Roberts, S. (2019). DeepLOB: Deep convolutional neural networks for limit order books. *IEEE Transactions on Signal Processing*, 67(11):3001–3012.

A Appendix



This figure shows the development of the inverse of the popular Herfindahl-Index across major European equity markets. A higher value indicates a higher level of fragmentation, and thus a higher level of competition within markets.

Figure A.1: Inverse of the Herfindahl Index in European Equity Markets

A.1 Data Generator

From the data pool, we sample without replacement individual time series and feed them as input to the model. We cannot fit these large and high-dimensional time series into memory, however, and we follow a heuristic approach to maximize heterogeneity in our training data while constantly reloading chunks of data into memory, without which the universal approach to model training would not work. We load into memory a chunk of size 10,000 that is composed of multiple random data regions (a particular stock over a particular period) and, out of this chunk, sample random mini-batches of size 32 that are ultimately fed to the model input. Since each chunk may only be composed of a limited number of data regions, we further increase heterogeneity within each data region by dropping 199 out of 200 consecutive time series so as to minimize the overlap between individual time series. When a given chunk is exhausted, we continue with the next chunk until we have sampled the specified total of mini-batches per epoch.

A.2 Timestamp Precision

In the prediction of mid-price changes across multiple trading venues based on high-resolution market data, timestamp precision is paramount. If the timestamps are not accurate and synchronized across all trading venues, they may result in a wrong sequence of LOB states, and thereby in our prediction model being biased - either by incorporating future states that already account for new information, or by ignoring observations from the recent past that provide valuable information about the next price movement. To ensure the highest possible timestamp precision and to avoid information leakage, we rely on data post MiFID-II only. Since the introduction of MiFID II

Table A.1: Data sample: stocks and trading venue market shares

This table shows all companies in our sample plus the corresponding RIC (Reuters Instrument Code) of their primary listing. The columns MAIN to TQ show each trading venue's respective share of lit trading volume during continuous trading, relative to all five trading venues, during the first quarter of 2019 (in percentage points). The shares are calculated on a daily basis using our observation window between 08:30 and 16:00 UTC. MAIN refers to the primary exchange of the stock. From left to right we cover the primary exchange, Aquis, Bats, Chi-X and Turquoise. In total we cover 46 stocks and five venues.

Company	RIC - Primary Listing	MAIN	AQX	BS	CHI	TQ
ASML Holding	ASML.AS	54.1	3.0	15.4	18.8	8.7
AXA	AXAF.PA	50.0	4.3	18.9	18.8	8.0
Adidas	ADSGn.DE	48.1	4.3	17.1	23.2	7.4
Air Liquide	AIRP.PA	42.2	5.0	27.0	17.1	8.7
Airbus Se	AIR.PA	45.4	4.4	19.1	21.3	9.7
Allianz Se	ALVG.DE	52.9	3.8	19.1	17.5	6.6
Amadeus It Group	AMA.MC	45.6	3.0	24.9	19.9	6.5
Anheuser-Busch	ABI.BR	51.2	4.4	17.8	20.8	5.8
BASF SE	BASFn.DE	55.3	3.6	16.2	18.6	6.3
BBVA	BBVA.MC	46.6	1.7	20.2	20.2	11.3
BMW AG	BMWG.DE	53.4	3.6	14.1	22.3	6.5
BNP Paribas	BNPP.PA	54.5	2.6	16.8	18.2	7.9
Banco Santander Sa	SAN.MC	61.6	1.4	17.6	12.0	7.3
Bayer AG	BAYGn.DE	52.2	4.5	16.1	20.4	6.7
Daimler	DAIGn.DE	54.4	3.9	17.1	18.9	5.8
Deutsche Post	DPWGn.DE	54.4	4.9	15.8	19.6	5.3
Deutsche Telekom	DTEGn.DE	52.1	5.8	16.4	19.1	6.5
Enel	ENEI.MI	48.4	3.1	26.4	17.4	4.7
Engie	ENGIE.PA	47.0	4.9	19.0	20.5	8.6
Eni	ENI.MI	54.0	3.4	17.5	17.2	7.9
Essilor International	ESLX.PA	48.1	4.2	21.8	18.3	7.6
Fresenius Se	FREG.DE	56.7	3.5	14.8	18.7	6.2
Groupe Danone	DANO.PA	44.2	3.4	23.4	21.3	7.7
ING Groep	INGA.AS	58.2	2.2	16.4	17.9	5.3
Iberdrola Sa	IBE.MC	47.4	3.7	21.1	19.6	8.2
Inditex	ITX.MC	46.6	3.4	23.2	18.9	8.0
Intesa Sanpaolo	ISP.MI	70.1	2.6	15.3	8.0	4.0
Kering	PRTP.PA	35.9	2.0	37.3	17.3	7.6
Koninklijke Ahold	AD.AS	51.2	3.6	18.5	18.3	8.4
Koninklijke Philips	PHG.AS	47.7	2.4	19.6	21.4	8.8
L Oreal S.A.	OREP.PA	45.1	4.0	23.5	18.4	9.1
LVMH	LVMH.PA	45.5	2.7	27.0	16.4	8.4
Munich Re	MUVGn.DE	53.7	3.4	17.8	18.8	6.2
Orange	ORAN.PA	51.4	5.8	17.4	17.4	8.0
SAP SE	SAPG.DE	43.4	3.2	24.2	19.8	9.4
Safran	SAF.PA	47.4	5.1	21.2	18.2	8.1
Sanofi-Aventis	SASY.PA	38.1	3.8	27.5	20.5	10.0
Schneider Electric Se	SCHN.PA	46.0	4.6	19.8	21.5	8.1
Siemens	SIEGn.DE	43.3	3.5	32.8	15.2	5.1
Societe Generale	SOGN.PA	57.3	2.5	13.4	19.7	7.1
Telefonica Sa	TEF.MC	47.5	3.8	20.0	18.2	10.5
Total	TOTF.PA	43.0	2.6	30.3	18.2	5.8
Vinci	SGEF.PA	44.8	4.1	21.2	21.6	8.3
Vivendi	VIV.PA	46.1	5.0	20.7	19.4	8.8
Volkswagen	VOWG_p.DE	60.4	3.8	12.5	18.7	4.6
Average		49.8	3.7	20.3	18.8	7.5

Table A.2: Summary statistics of independent variables in the OLS regressions

This table shows the distribution of the independent variables that are used in the cross-sectional regressions in chapter 5. Each observation represents one LOB in our sample which consists of 46 stocks traded on five venues, i.e., 230 stock-venue combinations in total. Note that these values are not yet winsorized and logarithmized.

	N	mean	std	min	25%	50%	75%	max
Market share	230	20.00	16.69	1.44	6.33	17.39	23.38	70.13
Order-trade-ratio	230	65.76	38.20	21.06	38.59	51.78	83.41	225.52
Range volatility	230	14.90	4.51	9.52	11.74	14.06	16.90	55.20
Quoted spread	230	4.36	1.38	1.88	3.21	4.26	5.37	8.36

in January 2018, European trading venues have been obligated to provide synchronized timestamps of millisecond precision (or higher) that diverge with a maximum of one millisecond from the Coordinated Universal Time (UTC) (European Commission, 2017). To take full advantage of the timestamp precision provided by the trading venue operators, we use the Thomson Reuters Tick History *Raw Legacy Market Depth* format to build our dataset.

Despite the millisecond timestamp precision, the LOB data from all five trading venues comprises observations where multiple updates occur within the same millisecond, especially in the most liquid stocks. While the sequence of updates is correctly recorded within the respective trading venue, the true sequence of LOB updates across multiple trading venues is ambiguous if the same timestamp exists across multiple trading venues.

To address this problem, we employ a simple solution based on timestamp randomization - group all data points based on timestamp and trading venue and, given a group of identical timestamps, add a random and unique timedelta $d \in [0, t_0 - t_1]$ to each data point such that the order inherent within each group of identical trading venues is preserved. On the one hand, we thereby assume a random order of market updates across multiple trading venues and, on the other hand, preserve the original order within each individual trading venue. While, consequently, this approach may not completely prevent data leakage across venues, the information leakage is limited to less than 1 ms, in most cases being substantially smaller. Moreover, it is unreasonable to assume multi-venue strategy induced patterns in a sub-millisecond time horizon as, for reasons of geographical distance, the lower bound on round-trip latency will generally be higher when trading across multiple trading venues. Lastly, we train and evaluate each model multiple times to address the risk of introducing systematic bias.

Table A.3: Overview of models' out-of-sample prediction accuracy

This table shows the out-of-sample prediction accuracy for all models. The prediction accuracy is calculated based on the mean across all three runs and is expressed in percentage points. The table shows that that multi-venue and single-venue output are best predicted based on multi-venue and single-venue input, respectively. Moreover, the prediction accuracy varies in the cross-section of stocks.

Model Input	Model Output	Mean	Std. Dev.
Multi Venue	Multi Venue	67.27	1.23
Single Venue - AQX	Multi Venue	58.21	2.02
Single Venue - BS	Multi Venue	61.15	2.28
Single Venue - CHI	Multi Venue	61.78	1.51
Single Venue - MAIN	Multi Venue	62.54	1.14
Single Venue - TQ	Multi Venue	60.60	2.45
Leave-one-out - AQX	Multi Venue	65.82	1.15
Leave-one-out - BS	Multi Venue	66.81	1.21
Leave-one-out - CHI	Multi Venue	65.67	1.21
Leave-one-out - MAIN	Multi Venue	64.66	1.23
Leave-one-out - TQ	Multi Venue	66.60	1.24
Multi Venue	Single Venue - AQX	61.65	1.67
Single Venue - AQX	Single Venue - AQX	64.67	2.25
Multi Venue	Single Venue - BS	62.20	1.71
Single Venue - BS	Single Venue - BS	64.46	1.64
Multi Venue	Single Venue - CHI	62.19	1.28
Single Venue - CHI	Single Venue - CHI	65.11	1.36
Multi Venue	Single Venue - MAIN	62.25	1.57
Single Venue - MAIN	Single Venue - MAIN	64.92	1.51
Multi Venue	Single Venue - TQ	61.86	2.04
Single Venue - TQ	Single Venue - TQ	65.06	2.11

Table A.4: Robustness results for models' out-of-sample prediction accuracy

This table shows the out-of-sample prediction accuracy for all models across all three runs. While the mean is calculated across runs and stocks (equal to table A.3), the standard deviation in this table is the average of all standard deviations that are calculated for each stock across the three runs (in percentage points). The table highlights that the models yield a consistent out-of-sample prediction accuracy across all runs in the cross-section of stocks, demonstrating that our results are robust with respect to the non-deterministic nature of neural network training.

Model Input	Model Output	Mean	Std. Dev.
Multi Venue	Multi Venue	67.27	0.16
Single Venue - AQX	Multi Venue	58.21	0.84
Single Venue - BS	Multi Venue	61.15	0.21
Single Venue - CHI	Multi Venue	61.78	0.32
Single Venue - MAIN	Multi Venue	62.54	0.25
Single Venue - TQ	Multi Venue	60.60	0.79
Leave-one-out - AQX	Multi Venue	65.82	0.18
Leave-one-out - BS	Multi Venue	66.81	0.19
Leave-one-out - CHI	Multi Venue	65.67	0.33
Leave-one-out - MAIN	Multi Venue	64.66	0.25
Leave-one-out - TQ	Multi Venue	66.60	0.15
Multi Venue	Single Venue - AQX	61.65	0.16
Single Venue - AQX	Single Venue - AQX	64.67	0.41
Multi Venue	Single Venue - BS	62.20	0.37
Single Venue - BS	Single Venue - BS	64.46	0.21
Multi Venue	Single Venue - CHI	62.19	0.39
Single Venue - CHI	Single Venue - CHI	65.11	0.28
Multi Venue	Single Venue - MAIN	62.25	0.28
Single Venue - MAIN	Single Venue - MAIN	64.92	0.19
Multi Venue	Single Venue - TQ	61.86	0.25
Single Venue - TQ	Single Venue - TQ	65.06	0.36