The Profitability of Lead-Lag Arbitrage at High-Frequency

Cédric Poutré[1], Georges Dionne[2,†], Gabriel Yergeau[2]

[1]Université de Montréal
[2]HEC Montréal

26 September 2022

Abstract

Any lead-lag effect in an asset pair implies the future returns on the lagging asset have the potential to be predicted from past and present prices of the leader, thus creating statistical arbitrage opportunities. We utilize robust lead-lag indicators to uncover the origin of price discovery and we propose an econometric model exploiting that effect with level 1 data of limit order books (LOB). We also develop a high-frequency trading strategy based on the model predictions to capture arbitrage opportunities. The framework is then evaluated on six months of DAX 30 cross-listed stocks' LOB data obtained from three European exchanges in 2013: Xetra, Chi-X, and BATS. We show that a high-frequency trader can profit from lead-lag relationships because of predictability, even when trading costs, latency, and execution-related risks are considered.

Keywords: Lead-lag relationship; High-frequency trading; Statistical arbitrage; Limit order book; Cross-listed stocks; Econometric models.

JEL codes: C25; C53; C58; G10; G14; G15; G17.

[†] Corresponding author: Canada Research Chair in Risk Management, HEC Montréal, 3000, Chemin de la Côte-Sainte-Catherine, room 4.454, Montreal (Qc) Canada, H3T 2A7. Phone: 514-340-6596. Email: georges.dionne@hec.ca

1 Introduction

Lead-lag relationships have long been a subject of interest in finance. The following are just a few areas that have been explored: stock index futures (Frino & West (2003), Dimpfl & Jung (2012)), cash market and stock index futures (Chan (1992)), stock and stock index futures (Brooks, Garrett, et al. (1999)), stock index and stock index futures (Kawaller et al. (1987), Jong & Nijman (1997), Yang et al. (2012)), stocks (Hou (2007)), spot stock index and stock index futures markets (Herbst et al. (1987), Tse (1995), Judge & Reancharoen (2014)), foreign exchange spot and futures markets (Y.L. Chen & Gau (2010)), and VIX markets (Bollen et al. (2017)). But, the hypothesis that these relationships can potentially be a source of profitable statistical arbitrage is fairly recent. For example, after finding significant lead-lag relationships in NYSE stocks, Curme et al. (2015) discussed the idea that lagged correlations might be exploited by a prediction model. They also believed that the resulting arbitrage opportunities may not be easily exploitable in the presence of market frictions. The same questions were also raised in Basnarkov et al. (2020) in the context of foreign exchange markets. In this paper, we revisit the existence, predictability, and profitability of lead-lag relationships in detail. Our main questions are the following:

1. Can lead-lag relationships be identified in the high-frequency prices of arbitrage-linked assets?

2. If the answer to question 1 is conclusive, can returns in lagging assets be predicted?

3. If the answers to questions 1 and 2 are both affirmative, can the predictability of lagging assets be exploited by high-frequency traders (HFTers), even when important market frictions are considered?

Up to now, the profitability of statistical arbitrage from lead-lag relationships with realistic trading behavior has not been well established. Our goal is to demonstrate its economic viability by proposing a new approach based on robust lead-lag indicators, the direction probability estimation of the lagging asset's return, and the use of LOB information in an high-frequency trading (HFT) arbitrage strategy. We also consider important potential market frictions between multiple exchanges with an application to DAX 30 stocks, all of which are cross-listed in three markets: Xetra in Frankfurt, and Chi-X and BATS, both in London.

Using recent advancements in the estimation of lead-lag, stemming from Hayashi & Yoshida (2005) and Hoffman et al. (2013), we demonstrate that Chi-X led the high-frequency prices of most DAX 30 stocks by mere milliseconds in 2013. This surprising result is in fact in line with other studies empirically demonstrating that the most liquid, actively traded, and least expensive exchange should be the origin of price discovery. This is true in our case, since Chi-X received more quotes and trades for DAX 30 stocks on a daily basis than either Xetra or BATS. Chi-X is also the exchange with the most generous trading rebates and is thus the most competitive option for high-frequency traders, which ultimately establishes Chi-X as the price leader for the cross-listed stocks under study. We also show that all DAX 30 stocks listed at these exchanges are extremely well integrated, because their lags are limited by the speed at which information can travel. This level of precision in the estimation of cross-listed stocks' lead-lag relationships has never been attained before.

Knowing that there is a definitive leader in the prices of cross-listed stocks, we then demonstrate how lagging assets' returns can be predicted accurately using current and past prices observed at two exchanges. A new econometric model, the autoregressive distributed lag multinomial logistic regression, is able to utilize the existing lead-lag relationship between two price processes to predict whether the lagging asset's next return will be positive, null, or negative, with an overall accuracy exceeding 80% out-of-sample. This degree of performance is well maintained throughout our data period, further indicating the robustness of the lead-lag relationship detected in DAX 30 stocks. On our data, the proposed model's accuracy compares favorably with those of models previously suggested in the lead-lag literature, e.g., Huth & Abergel (2014) and Alsayed & McGroarty (2014). It is also a significant departure from ordinary least square models, because it predicts the probabilities of the lagging asset's next return direction instead of predicting the next return itself. We show that this easier task makes it possible to build a more profitable HFT strategy by detecting more potential arbitrage opportunities with superior accuracy. Moreover, as opposed to popular frameworks based on error correction or vector autoregression models, we do not require a uniform sampling scheme of the price processes, which distinguishes our work from prior studies even further.

Fragmented markets make arbitrage opportunities more abundant for HFTers (Foucault & Biais (2014) and O'Hara (2015)). In this case of cross-listed stocks, whenever a lead-lag movement in a lagging asset takes longer than the usual lag to occur (which is measured in milliseconds), an arbitrage opportunity is revealed. Earlier work on high-frequency lead-lag arbitrage failed to generate a profit due to trading costs created by market orders. This occurred with few exceptions, which we will address later. We empirically demonstrate the impossibility of profiting from the usual mid-quote signal coupled with market orders in the context of high-frequency lead-lag arbitrage. Thus, we propose a different strategy, one that makes use of limit orders, thereby reducing the exchange trading costs while also not having to cover the bid-ask spread at every arbitrage opportunity. Furthermore, the trading signal is based on level 1 prices rather than mid-quotes, leading to better-informed decisions compared to earlier studies. In a scenario where latency, trading costs, and execution-related risks are all taken into consideration, we determine that a high-frequency trader colocated at Chi-X is able to generate a net profit surpassing €1.9 million by arbitraging DAX 30 stocks in 2013 at only two exchanges: Xetra and BATS. The presence of market frictions dramatically impedes the trader's capacity to profit more from the detected lead-lag arbitrage opportunities, and risk management procedures are necessary to obtain a satisfying profitability.

The methodology and results in this paper are important from both the academic and practitioner standpoint. First, we contribute to the ongoing discussion about HFTers' arbitrage activities,[1] since the understanding of which is still limited in the empirical research (Y. Chen et al. (2019)). Indeed, our paper demonstrates how HFTs are realistically able to profit from a specific form of statistical arbitrage. Second, we quantify the interconnectedness of international markets in the case of cross-listed stocks by explicitly measuring the time needed between exchanges to incorporate new price information. Third, we further advance the lead-lag literature by providing the first truly profitable high-frequency lead-lag arbitrage strategy and a new econometric model that is able to predict future returns of lagging assets with an accuracy that surpasses earlier models. Furthermore, our framework is applicable to any pair of assets, making it useful for future studies on lead-lag relationships.

Our work falls under the lead-lag arbitrage literature, in which scarcely any studies have attempted to quantify the financial importance of lead-lag relationships. Brooks, Rew, et al. (2001); Huth & Abergel (2014); and Alsayed & McGroarty (2014) are closely related to our paper, especially the last one. However, our study differs from Alsayed & McGroarty (2014) on many points. Firstly, we do not work on a mid-quote basis because, as we show, this leads to suboptimal trading decisions. Each of the three papers above use that setting. We alternatively directly model the best bid and ask price processes, which allows for more precise predictions and better-informed trading decisions. Secondly, we propose an econometric model utilizing all relevant past prices observed in both the lagging and leading assets, instead of a subset of that information. Thirdly, rather than relying on liquidity-taking orders, as in the three above-mentioned papers, we employ liquidity-providing limit orders to avoid important trading costs that render all of their strategies non-viable in practice. It also allows for a more passive trading strategy, which we show to be profitable on our data. Finally, our application covers a new area for lead-lag arbitrage: cross-listed stocks.

The remainder of the paper is organized as follows. Section 2 introduces the literature

[1] Refer to the recent Staff Report on Algorithmic Trading in U.S. Capital Markets of the SEC: `https://www.sec.gov/tm/reports-and-publications/special-studies/algo_trading_report_2020` and the MiFID II Review Report on Algorithmic Trading of the ESMA: `https://www.esma.europa.eu/press-news/esma-news/esma-publishes-mifid-ii-review-report-algorithmic-trading` (both accessed August 12, 2022).

on lead-lag relationships, where an emphasis is put on cross-listed stocks, different high-frequency arbitrage strategies, and lead-lag estimation methods in past studies. Section 3 presents the methodology used to locate and quantify lead-lag relationships. It also details the proposed econometric model in conjunction with the new HFT strategy built around it. The section ends with a description of market frictions and how we include them into our estimations. Section 4 is dedicated to the data from Xetra, Chi-X, and BATS, and also presents the latencies and costs we utilize. Section 5 analyzes the empirical results of our methodology and discusses their implications. Section 6 concludes the paper.

2 Literature Review

As discussed in the introduction, lead-lag relationhips have been observed in most financial assets and instruments. The particular case of cross-listed stocks has been studied at an intraday frequency in Grammig et al. (2005); Pascual et al. (2006); Frijns, Gilbert, et al. (2010); Frijns, Gilbert, et al. (2015); Ghadhab & Hellara (2016); and Frijns, Indriawan, et al. (2018). They all analyze cross-listed stock price discovery based on variations of Hasbrouck's information shares (Hasbrouck (1995)) and/or the component shares of Gonzalo & Granger (1995). Grammig et al. (2005) sample 10-second intervals of mid-quote prices of three German firms cross-listed in New York (NYSE) and Frankfurt (Xetra) from August to October 1999, and find that price discovery mostly originated from the home exchange. Pascual et al. (2006) arrive at the same conclusion in the case of five Spanish ADRs listed on the NYSE and SSE at a one-minute resolution in 2000, as do Frijns, Gilbert, et al. (2010) on four Australian and five New Zealand firms from 2002 to 2007 at a minute level. Ghadhab & Hellara (2016) also corrobarate the idea that local markets are dominant for cross-listed stocks, but find that foreign markets contribute more to price discovery for multiple-listed firms, even more so when their trading costs are lower. Other factors affect the origin of price discovery for cross-listed stocks. Indeed, Frijns, Gilbert, et al. (2015) suggest that a reduced bid-ask spread and a higher trade activity, small trades in particular, have a positive and causal impact on price discovery, from a sample of cross-listed Canadian stocks in the US from 1996 to 2011, at a minute frequency. These recur in Frijns, Indriawan, et al. (2018), which finds a bilateral causality between liquidity in an exchange and its contribution to price discovery. These authors also obtain that algorithmic activity is negatively related to price discovery for Canadian cross-listed stocks in the US from 2004 to 2017. None of the papers mention the possibility of an arbitrageur exploiting these lead-lag relationships, nor do they measure how predictable the lagging assets returns are. We aim to answer these questions by proposing a novel HFT strategy and a new econometric model for cross-listed stocks. Our methodology also considers important limiting factors of arbitrage, mainly, trading costs, latency, and execution-related risks. The proposed model is also computationally simple enough to be used by HFTers in practice.

Very few papers have tried to develop arbitrage strategies or predictive models based on the concept of lead-lag in finance, and none in the context of cross-listed stocks: Judge & Reancharoen (2014) and Li et al. (2022) use daily data; Brooks, Rew, et al. (2001) and Stübinger (2019) focus on uniformly sampled intraday data; and Huth & Abergel (2014) and Alsayed & McGroarty (2014), the closest studies to our paper, also use LOB data. Brooks, Rew, et al. (2001) investigate the lead-lag relationship between the spot index and futures contract of the FTSE 100 at a 10-minute frequency. They are able to predict, one step ahead, the direction of the return in the lagging spot price, with an out-of-sample accuracy approaching 70%, based on a version of the error correction model (ECM) of Engle & Granger (1987). Nonetheless, because of trading costs, their round-trip trade strategy is unable to outperform a passive buy-and-hold strategy. In the same vein, Huth & Abergel (2014) are also not able to profit from the lead-lag relationship they detect in a futures-stock pair, since paying the bid-ask spread at every opportunity is too expensive. Even though their linear regression model predict the next mid-quote return at the next trade of the lagging stock with an accuracy of 60%, the opportunities detected do not cover the market orders costs.

On the other hand, Stübinger (2019) and Alsayed & McGroarty (2014) find economically significant profit-generating strategies by exploiting lead-lag relationships. Stübinger (2019) proposes the "optimal causal path algorithm" to uncover the lead-lag structure between two time series, and then applies it to S&P 500 constituents at a minute level, to identify promising stocks for a pair trading–type strategy. The strategy limits excessive trading by only selecting statistically high returns of the leading stock that also cover the trading costs of market orders. Positions are closed after ℓ minutes, where ℓ is the lag estimated from the optimal causal path algorithm. This trading signal allows the author to significantly outperform a buy-and-hold strategy of the S&P 500 index after transaction costs. But, in a high-frequency setting where lag is measured in milliseconds, as in our study, the trading signal of Stübinger (2019) would result in an insignificant number of trades, since returns at that scale seldom cover the bid-ask spread. Alsayed & McGroarty (2014) profit from lead-lag arbitrage across international futures with a new forecasting framework yielding over 85% accuracy in lagging contracts' mid-quote changes. Their framework is based on the concept of clusters, which are uninterrupted, contiguous observations of prices that allow them to predict mid-quote movements and trade at a high frequency. But, we question the strategy's practical profitability because their profit calculations use mid-quote returns and not actual execution prices. We are proposing a novel high-frequency strategy relying on limit orders to circumvent the profitability issues of earlier studies. Our practical methodology also gets as close as possible to real-life HFT, thus making our results more concrete and accurate. In both Huth & Abergel (2014) and Alsayed & McGroarty (2014), the leading asset leads by mere fractions of a second: around 300 milliseconds in the former and down to 25 milliseconds for a particular pair in the latter. This highlights the importance of newer methodologies enabling sub-second lead-lag estimation.

Considering that today's integrated markets rely heavily on advanced information technology to connect traders and exchanges around the globe, aggregated data at the minute level is not suitable to uncover lead-lag relationships between cross-listed stocks. This is especially true when exchanges are geographically close. As shown in Budish et al. (2015), the correlation of related instruments only breaks down at a millisecond resolution in well-integrated markets, even though their correlation seem nearly perfect at a minute level. But, using sub-second data, i.e., trades and quotes (TAQ) from LOB data, to quantify lead-lag relationships has its challenges: it is neither synchronously nor regularly observed. As noted in Hayashi & Yoshida (2005) and Zhang (2011), among others, earlier estimators based on previous-tick interpolation are severely biased whenever the processes are not synchronously observed. This is true for Granger's causality (Granger (1969)) and for Hasbrouck's information share (Hasbrouck (1995)) models when working with HFT data, because correlation estimates decrease when the processes are synchronously sampled at high frequencies. This downward correlation bias effect was first studied in Epps (1979). Furthermore, if the two processes differ in noise, microstructure frictions, or liquidity, these methods will not be consistent (Putniņš (2013)). Since 2010, some consistent estimators of lead-lag at a high frequency have been proposed (e.g., Hoffman et al. (2013), Hayashi & Koike (2018)), making it possible to depart from previous-tick interpolation and regular sampling of LOB data. It is now possible to use the LOB as is. We are the first to investigate lead-lag relationships of cross-listed stocks at that level of precision, since past causality methods would not have been robust at that time scale. Being able to work at the sub-second horizon is absolutely necessary in our case, because the geographical proximity of the exchanges allows information to flow between them nearly instantly.

3 Methodology and Framework

We introduce the ideas behind the results presented in Section 5. Even though our application covers cross-listed stocks, the general methodology and framework in this section are applicable to any financial market where a high-frequency trader suspects that a lead-lag relationship exists between any pair of assets.

Subsection 3.1 details how we find lead-lag relationships between processes and how to quantify their strength. Subsection 3.2 proposes an econometric model able to exploit an existing lead-lag relationship by predicting the lagging process' future directional movements from past information on the leading process. Subsection 3.3 presents an HFT strategy created from the econometric model predictions. Finally, subsection 3.4 is dedicated to the market frictions we consider when computing our trading profits.

3.1 Lead-Lag Relationships

There are two main schools of thought as regards the ways of mathematically defining and detecting lead-lag relationships: causality methods (e.g., Granger (1969)) or correlation methods (e.g., Herbst et al. (1987)). The latter approach makes it possible to explicitly measure the timing relationship between time series, which provides valuable information in a trading context. Following that literature, there exists a lead-lag relationship in a pair of stochastic processes $(\{X_t\}, \{Y_t\})$ with observations $(\{x_t\}, \{y_t\})$ whenever their cross-correlation with lag ℓ, $\mathrm{Corr}(X_t, Y_{t+\ell})$, is statistically different from 0 for any $\ell \neq 0$. The optimal lag ℓ^* is defined as

$$\ell^* \equiv \underset{\ell \in \mathbb{R}}{\arg\max}\, |\operatorname{Corr}(X_t, Y_{t+\ell})| = \underset{\ell \in \mathbb{R}}{\arg\max}\, |\rho_{X,Y}(\ell)|,$$

where $\rho_{X,Y}(\ell)$ is the lagged Pearson correlation coefficient $\rho_{X,Y}(\ell) \equiv \dfrac{\mathrm{Cov}(X_t, Y_{t+\ell})}{\sqrt{\mathrm{Var}(X_t)\,\mathrm{Var}(Y_t)}}$, $\mathrm{Cov}(X_t, Y_{t+\ell})$ is the lagged cross-covariance of processes $(\{X_t\}, \{Y_t\})$, and $\mathrm{Var}(\cdot)$ is their variance. Whenever $\ell^* \neq 0$, the relationship between $\{X_t\}$ and $\{Y_t\}$ is not contemporaneous and it establishes that there is lead-lag between the processes. When $\ell^* > 0$, $\{X_t\}$ leads $\{Y_t\}$ and vice versa for $\ell^* < 0$. Knowledge of the leader at t can potentially be exploited to forecast the lagging process at $t + \ell^*$.

In this paper, we rely on high-frequency data, which is notable for being non-synchronous and irregularly observed. "Non-synchronous" means that the two processes are observed at different times, and "irregularly observed" refers to irregular intervals between observation times of the processes. These features drive us to depart from older lead-lag estimation methods used in the literature, as mentioned earlier in Section 2. Hayashi & Yoshida (2005) propose a covariance estimator for non-synchronous and irregularly observed diffusion processes, resulting in the following consistent cross-correlation estimator:

$$\hat{\rho}^{HY}_{X,Y} = \frac{\sum_i \sum_j \Delta X(I^X_i)\Delta Y(I^Y_j)\mathbb{1}_{\{I^X_i \cap I^Y_j \neq \emptyset\}}}{\sqrt{\sum_i [\Delta X(I^X_i)]^2 \sum_j [\Delta Y(I^Y_j)]^2}},$$

where

$$\mathbb{1}_{\{A\}} = \begin{cases} 1, & \text{if A is true,} \\ 0, & \text{if A is false} \end{cases}$$

is the indicator function. The processes $(\{X_t\}, \{Y_t\})$ have discrete observation times $0 = t^X_1 < t^X_2 < \cdots < t^X_n = T^X$ and $0 = t^Y_1 < t^Y_2 < \cdots < t^Y_m = T^Y$ with intervals $I^X_i = (t^X_{i-1}, t^X_i]$, $I^Y_j = (t^Y_{j-1}, t^Y_j]$ and $\Delta X(I^X_i) = x_{t^X_i} - x_{t^X_{i-1}}$, $\Delta Y(I^Y_j) = y_{t^Y_j} - y_{t^Y_{j-1}}$. Hoffman et al. (2013) extended this estimator to include the lag ℓ:

$$\hat{\rho}^{HY}_{X,Y}(\ell) = \frac{\sum_i \sum_j \Delta X(I^X_i)\Delta Y(I^Y_j)\mathbb{1}_{\{I^X_i \cap (I^Y_j)_\ell \neq \emptyset\}}}{\sqrt{\sum_i [\Delta X(I^X_i)]^2 \sum_j [\Delta Y(I^Y_i)]^2}}$$

where $(I^Y_j)_\ell = (t^Y_{j-1} + \ell, t^Y_j + \ell]$. This makes it possible to obtain a practical and unbiased estimation of ℓ^* on HFT data:

$$\widehat{\ell^*} = \underset{\ell \in \mathbb{R}}{\arg\max}\, |\hat{\rho}^{HY}_{X,Y}(\ell)|,$$

which is the estimator used in this paper. In order to quantify the overall side and strength of the lead-lag relationship, Huth & Abergel (2014) introduce the Lead-Lag Ratio (LLR) measuring the asymmetry of the cross-correlation function:

$$LLR_{X,Y} \equiv \frac{\sum_{g \in \mathcal{G}} \hat{\rho}^{HY}_{X,Y}(\ell_g)^2}{\sum_{g \in \mathcal{G}} \hat{\rho}^{HY}_{X,Y}(-\ell_g)^2}$$

for $\mathcal{G}$, a discrete time grid of positive lags. Whenever $LLR_{X,Y} > 1$, $\{X_t\}$ leads $\{Y_t\}$ and the higher $LLR_{X,Y}$ is, the more $\{X_t\}$ leads $\{Y_t\}$. This statistic is also applied to detect lead-lag relationships in our data.

3.2 Econometric Model

We concentrate on the models of Huth & Abergel (2014) and Alsayed & McGroarty (2014) since they are the only studies whose methodologies are directly developed on unsampled LOB data. Huth & Abergel (2014) are predicting the direction of the mid-quote move (up or down) at the next trade of the lagging mid-quote process $\{Y_t\}$ by taking the sign of a linear regression that uses the leader's past mid-quote moves as the only exogenous variables, like so:

$$\widehat{R}^Y_j \equiv \mathrm{sign}(\widehat{\Delta Y}(I^Y_j)) = \mathrm{sign}\left(\sum_{k=1}^{p} \beta_k \sum_{i:t^X_i < t^Y_{j-1}} \Delta X(I^X_i) \mathbb{1}_{\{I^X_i \cap (I^Y_j)_{\ell_k} \neq \emptyset\}} \right),$$

where p is the last statistically significant lag. They set $\beta_k = \hat{\rho}^{HY}_{X,Y}(\ell_k)$ and achieve around 60% directional accuracy on test days. The model's core idea is a binary classification, when in fact, a logistic regression would be more appropriate than taking the sign of a model that is designed for a harder prediction problem. Predictions that fall close to 0 can also be problematic since they lie around the model's decision boundary, where predictions are most uncertain (Nguyen et al. (2022)). Adding a null prediction seems necessary for HFT whenever that occurs. Null predictions have been considered in the next contribution.

Alsayed & McGroarty (2014) define *clusters* as sets of contiguous process variations uninterrupted by variations of a second process observed in parallel. They define $\Big\{ C^X_{i,n} \;\Big|\; i, n \in$

$\mathbb{N}^+\Big\}$ as the set of clusters of process $\{X_t\}$, where the subscript i refers to the cluster index and n the variation index within each cluster. The same definition holds for process $\{Y_t\}$. Figure 1 illustrates the concept of clusters.

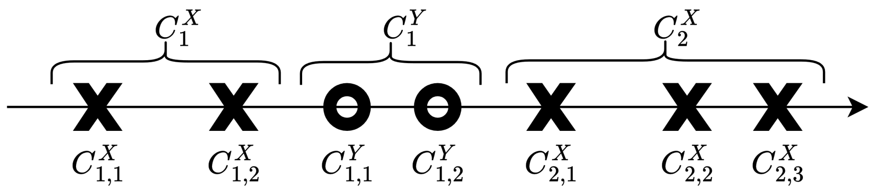

Figure 1: Time-line illustration of dual process clusters. Observations of process $\{X_t\}$ are marked by an "X" and those of $\{Y_t\}$ are marked by an "O." Taken from Alsayed & McGroarty (2014).

Suppose that $\{X_t\}$ leads $\{Y_t\}$, and define $\overline{C}^{\cdot}_{i,n}$ as the mid-quote returns of both processes, Alsayed & McGroarty (2014) predict the next cluster's direction of the lagging asset, $R_{\overline{C}_i^Y} \equiv \operatorname{sign}\left(\sum_n \overline{C}^Y_{i,n}\right)$, with the following rule:

$$\widehat{R}_{\overline{C}_i^Y} = \begin{cases} +1, & \text{if} \max\limits_n \left(\overline{C}^X_{i,n}\right) \geq K^{AM} \\ -1, & \text{if} \min\limits_n \left(\overline{C}^X_{i,n}\right) \leq -K^{AM} \\ 0, & \text{otherwise}, \end{cases}$$

where $K^{AM} \in \mathbb{R}_0^+$ is a preset threshold. They achieve a directional accuracy in excess of 85% on pairings of S&P 500, FTSE 100, and DAX futures contracts in 2012. This high level of accuracy can be explained by the high $LLR_{X,Y}$ in the three asset pairs studied. Only relying on the leader's latest cluster might be hazardous for asset pairs with a weaker lead-lag relationship.

Huth & Abergel (2014) and Alsayed & McGroarty (2014) both offer interesting predictive models that are able to exploit HFT lead-lag relationships in their respective financial contexts. The use in Huth & Abergel (2014) of the leading process' past relevant information, the simplicity of the Alsayed & McGroarty (2014) model, and the trader's ability to set a confidence threshold are all important qualities in HFT econometric models. We extend their contributions by proposing a model that takes into account the aforementioned overlooked aspects. Following Alsayed & McGroarty (2014), we set clusters of the leading price process as $C_i^X = \left\{ C_{i,j}^X \;\middle|\; j = 1, \ldots, n_i^X \in \mathbb{N}^+ \right\}$ and the lagging price process' as $C_i^Y = \left\{ C_{i,j}^Y \;\middle|\; j = 1, \ldots, n_i^Y \in \mathbb{N}^+ \right\}$, where $i = 1, 2, \ldots, N$ for N the number of clusters, and $C_{i,j}^X = \Delta X \left(I^X_{\sum_{k<i} n_k^X + j} \right)$, $C_{i,j}^Y = \Delta Y \left(I^Y_{\sum_{k<i} n_k^Y + j} \right)$ the absolute variations of the two price processes (any price process, not necessarily mid-quote). We define $r_{C_i^X} = \sum_{j=1}^{n_i^X} C_{i,j}^X$ as the total price process variation within cluster C_i^X and the same definition applies for $\{Y_t\}$. Without loss of generality, we assume that the first cluster we observe is from $\{X_t\}$, and the last one is from $\{Y_t\}$. We are interested in predicting the direction of $r_{C_i^Y}$ based on past observations of $(\{X_t\}, \{Y_t\})$, i.e.,

$$R_{C_i^Y} \equiv \operatorname{sign}\left(r_{C_i^Y} \right) = \begin{cases} +1, & \text{if } r_{C_i^Y} > 0 \\ 0, & \text{if } r_{C_i^Y} = 0 \\ -1, & \text{if } r_{C_i^Y} < 0. \end{cases}$$

To do so, we propose the autoregressive distributed lag multinomial logistic regression (ADLMLR) to model $R_{C_i^Y}$. It generalizes the logistic models for autoregressive binary variables introduced in Bonney (1987) in two ways. Firstly, it departs from a binary dependent variable to a multicategorical one, allowing for the modeling of a larger spectrum of systems. Secondly, $\{Y_t\}$ is not only autoregressive, it is autoregressive with a distributed lag for $\{X_t\}$, thus incorporating past values of both processes. This model is an important departure from conventional approaches based on error correction models (ECM) (for example, Engle & Granger (1987), Hasbrouck (1995), Brooks, Rew, et al. (2001), Pascual et al. (2006), Frijns, Gilbert, et al. (2010), Yang et al. (2012), Judge & Reancharoen (2014)) or vector autoregressive models (VAR) (see Hou (2007), Dimpfl & Jung (2012)) since it does not require the processes to be synchronously and regularly observed in time, thanks to the use of clusters. We also depart from an ordinary least squares (OLS) framework to a probabilistic one, where we are interested in predicting the probability of the class of the next return's direction (positive, neutral, or negative) instead of quantifying the return itself. This probabilistic task is easier to accomplish, hence the model predictions are more robust. As we will show, this leads to a greater profitability potential when incorporated into an HFT strategy.

The proposed ADLMLR model for $R_{C_i^Y}$ is as follows. Let

$$\Big(R_{C_i^Y} \mid \mathbf{p}_i = [p_{i,-1}\ p_{i,0}\ p_{i,+1}]\Big) \sim \text{Multinouilli}(p_{i,-1}, p_{i,0}, p_{i,+1})$$

where $p_{i,\cdot} \in [0,1],\ \sum p_{i,\cdot} = 1\ \forall i$, are the conditional probabilities of their respective return direction based on the past observations of $(\{X_t\},\{Y_t\})$. Supposing a (auto)dependence lag of order $D \in \mathbb{N}^+$ for $\{Y_t\}$, we have

$$\begin{aligned}
p_{i,-1} &= P\Big(R_{C_i^Y} = -1 \mid r_{C^Y_{i-D:i-1}}, r_{C^X_{i-D+1:i}}\Big),\\
p_{i,0} &= P\Big(R_{C_i^Y} = 0 \mid r_{C^Y_{i-D:i-1}}, r_{C^X_{i-D+1:i}}\Big),\\
p_{i,+1} &= P\Big(R_{C_i^Y} = +1 \mid r_{C^Y_{i-D:i-1}}, r_{C^X_{i-D+1:i}}\Big),
\end{aligned}$$

where $r_{C_{i-D:i}} = \{r_{C_{i-D}}, r_{C_{i-D+1}}, \ldots, r_{C_{i-1}}, r_{C_i}\}$. We define the conditional probabilities from the *logit* function with an autoregressive distributed lag-like model:

$$\begin{aligned}
\ln\left(\frac{p_{i,-1}}{p_{i,+1}}\right) &= \alpha_{-1} + \sum_{j=0}^{D-1} \beta_{j,-1} r_{C^X_{i-j}} + \sum_{j=1}^{D} \gamma_{j,-1} r_{C^Y_{i-j}},\\
\ln\left(\frac{p_{i,0}}{p_{i,+1}}\right) &= \alpha_0 + \sum_{j=0}^{D-1} \beta_{j,0} r_{C^X_{i-j}} + \sum_{j=1}^{D} \gamma_{j,0} r_{C^Y_{i-j}}.
\end{aligned}$$

Since we also have $\sum p_{i,\cdot} = 1$, we can find the conditional probabilities:

$$
\begin{aligned}
p_{i,-1} &= \frac{e^{\theta_{i,-1}}}{1 + e^{\theta_{i,-1}} + e^{\theta_{i,0}}}, \\
p_{i,0} &= \frac{e^{\theta_{i,0}}}{1 + e^{\theta_{i,-1}} + e^{\theta_{i,0}}}, \\
p_{i,+1} &= \frac{1}{1 + e^{\theta_{i,-1}} + e^{\theta_{i,0}}},
\end{aligned}
$$

where

$$
\begin{aligned}
\theta_{i,-1} &= \alpha_{-1} + \sum_{j=0}^{D-1} \beta_{j,-1} r_{C^X_{i-j}} + \sum_{j=1}^{D} \gamma_{j,-1} r_{C^Y_{i-j}}, \\
\theta_{i,0} &= \alpha_0 + \sum_{j=0}^{D-1} \beta_{j,0} r_{C^X_{i-j}} + \sum_{j=1}^{D} \gamma_{j,0} r_{C^Y_{i-j}}.
\end{aligned}
$$

The parameters of the model $\Theta = \{\alpha_{-1}, \alpha_0, \beta_{0,-1}, \ldots, \beta_{D-1,-1}, \beta_{0,0}, \ldots, \beta_{D-1,0}, \gamma_{1,-1}, \ldots,$ $\gamma_{D,-1}, \gamma_{1,0}, \ldots, \gamma_{D,0}\}$ are found by maximum likelihood estimation of

$$
\mathcal{L}(\Theta) = \prod_{i=D}^{N} (p_{i,-1})^{\mathbb{1}\left\{R_{C^Y_i}=-1\right\}} (p_{i,0})^{\mathbb{1}\left\{R_{C^Y_i}=0\right\}} (p_{i,+1})^{\mathbb{1}\left\{R_{C^Y_i}=+1\right\}},
$$

so that

$$
\widehat{\Theta} = \underset{\Theta \in \mathbb{R}^{4D+2}}{\arg\max}\, \mathcal{L}(\Theta).
$$

We use the BFGS algorithm of Broyden (1970), Fletcher (1970), Goldfarb (1970), and Shanno (1970) to solve for $\widehat{\Theta}$. The largest predicted probability in vector $\widehat{\mathbf{p}}_i = [\widehat{p}_{i,-1}\ \widehat{p}_{i,0}\ \widehat{p}_{i,+1}]$ determines the direction of the total variation in cluster C^Y_i:

$$
\widehat{R}_{C^Y_i} = \begin{cases} +1, & \text{if } (\max(\widehat{\mathbf{p}}_i) = \widehat{p}_{i,+1}) \wedge (\widehat{p}_{i,+1} \geq K) \\ 0, & \text{if } \max(\widehat{\mathbf{p}}_i) = \widehat{p}_{i,0} \\ -1, & \text{if } (\max(\widehat{\mathbf{p}}_i) = \widehat{p}_{i,-1}) \wedge (\widehat{p}_{i,-1} \geq K), \end{cases}
$$

where $K \in [0,\ 1]$ is a preset decision threshold controlling the minimum confidence needed to make a prediction.

3.3 High-Frequency Arbitrage Strategy

With market orders, Brooks, Rew, et al. (2001) and Huth & Abergel (2014) are not able to profit from their predictions, as paying the bid-ask spread at every opportunity is prohibitive for a HFTer, even more so considering exchange trading costs. Predicting thedirection of mid-quote movement is also not the most practical way of building an HFT strategy since orders cannot be executed at that price — another problem discussed in Huth & Abergel (2014). To circumvent these issues, we are predicting the direction of variations in the best bid and best ask prices based on the econometric model introduced in the previous subsection. In other words, a first model instance is used for the best bid price process and a second one is dedicated to the best ask. We are also relying on limit orders to reduce trading costs.

We assume an existing lead-lag relationship between a leader $\{X_t^{Bid/Ask}\}$ and a lagging process $\{Y_t^{Bid/Ask}\}$, which are the best bid/ask price processes. We also assume that our econometric model is able to utilize that relationship to generate adequate predictions. Based on these assumptions, we are interested in profiting from the predicted directions in clusters of $\{Y_t^{Bid/Ask}\}$: $\widehat{R}_{C_i^{Y^{Bid/Ask}}}$. For a tick size of δ, the novel HFT strategy is as follows:

- Bid price process:
  - When $\widehat{R}_{C_i^{Y^{Bid}}} = -1$, do all actions at the same time:
    1. Send a marketable sell limit order of volume V_i^{Bid} at the current value of $\{Y_t^{Bid}\}$;
    2. Send a buy limit order of volume V_i^{Bid} at the current value of $\{Y_t^{Bid}\}$ minus δ;

3. Send a stop buy limit order of volume V_i^{Bid} with stop and limit prices equal to the current value of $\{Y_t^{Bid}\}$ plus 2δ.

- When $\widehat{R}_{C_i^{Y^{Bid}}} \in \{0,\ 1\}$: do nothing.
- When a position has been open for M minutes, send a market buy order to close.

- Ask price process:
  - When $\widehat{R}_{C_i^{Y^{Ask}}} = 1$, do all actions at the same time:
    1. Send a marketable buy limit order of volume V_i^{Ask} at the current value of $\{Y_t^{Ask}\}$;
    2. Send a sell limit order of volume V_i^{Ask} at the current value of $\{Y_t^{Ask}\}$ plus δ;
    3. Send a stop sell limit order of volume V_i^{Ask} with stop and limit prices equal to the current value of $\{Y_t^{Bid}\}$ minus 2δ.
  - When $\widehat{R}_{C_i^{Y^{Ask}}} \in \{-1,\ 0\}$: do nothing.
  - When a position has been open for M minutes, send a market sell order to close.

A short (long) position is open when the marketable sell (buy) limit order hits the market and the buy (sell) limit order tries to close it whenever the lagging process $\{Y_t^{Bid}\}$ ($\{Y_t^{Ask}\}$) moves in the predicted direction. This allows us to capture a potential profit of δ when our econometric model makes a good prediction. No new position is open until the previous one has been closed. The stop limit orders are employed for risk management in the case of a wrong prediction; the same goes for closing market orders. Additional details of the strategy are presented in Appendix A.

3.4 Market Frictions: Latency, Risks, and Costs

In order to be as practical as possible, we use the Deltix QuantOffice trading software suite. This software only manages back-office operations and replays the LOB messages for backtesting purposes, letting us get closer to real-life high-frequency trading. It is possible to bypass the software and implement an equivalent testing program. We utilize the professional suite to ensure the quality of the results.

Latency is of paramount importance in HFT, as shown in Poutré et al. (2021). So, we use a simplified version of their methodology to account for latency in our empirical results. By latency, we mean the total time it takes for a trader to interact with the market when new information arrives. Hasbrouck & Saar (2013) measure latency on three components: the time it takes for a trader to learn about an event, to generate a response, and for the exchange to act on that response. Considering an HFT colocated at the leading exchange, the first two components of latency are the amount of time required for information generated at a lagging exchange to arrive and its treatment by the HFTer's server and trading algorithm. This is due to the finite speed of light causing a delay in the observed LOB between the source of information (lagging exchange) and its point of observation (leading exchange). To replicate that relativistic effect for a HFTer, we wait for an amount of time equal to the true one-way information transportation time plus its treatment time before entering the lagging exchange's data into the HFT strategy, thus delaying it. So, for a HFTer colocated at the leading exchange, it is as if its trading algorithm only observes past LOB states of the geographically distant lagging exchange, as it would in practice. Moving forward, this will be referred to as the first half of latency.

The last component of latency, which we will refer to as the second half of latency, is treated similarly. When the HFT strategy of Section 3.3 generates a trade signal, the orders are only sent to the execution engine after a time delay that corresponds to the same one-way information transportation time between exchanges, plus the receiving exchange's matching engine delay. So, a HFTer cannot interact infinitely rapidly with a geographically distant lagging exchange, as is the case in practice. For convenience, we assume that the HFT server is able to process a stream of level 1 data with the same efficiency as an exchange server. This allows us to use the same total latency value for the first and second halves of latency. In the next section, Table 3 presents the latency values employed.

The high-frequency strategy is exposed to both execution and non-execution risks since it utilizes market and limit orders. Those risks are taken into account using a set of professional rules determining if, when, and at what price the orders sent would have been executed in practice. The details are presented in Poutré et al. (2021). We also compute exchange trading costs after an order's execution, which are shown in Table 4 of the next section. Liquidity removal costs for marketable limit and market orders, and liquidity-providing costs for limit orders are taken into account.

4 Data

DAX 30 (which was extended to DAX 40 on November 24, 2020) is a German stock index containing 30 of the country's largest blue chip companies. Table 1 lists its constituents in 2013, and Table 2 details some of their stylized facts. Xetra, operated by Deutsche Börse AG at the Frankfurt Stock Exchange, is the reference order-driven trading venue for German stocks and has normal trading hours of 9:00 a.m. to 5:30 p.m. CET.[2] Chi-X Europe, also an order-driven exchange, is a cost-effective pan-European alternative to the largest European exchanges, with continuous trading hours between 8:00 a.m. and 4:30 p.m. GMT, located in London. Finally, BATS Europe (Better Alternative Trading System) is another London-based pan-European stock exchange, founded in 2008. BATS Europe was a direct competitor to Chi-X Europe, with the same normal trading hours, but it ultimately acquired the latter in 2011.

Our data covers DAX 30 stocks in the three European exchanges listed above: Xetra, Chi-X, and BATS, and spans six months in 2013, from February to July, inclusively, thus covering 125 trading days. Xetra's raw data contains every market event sent by the exchange, and

[2]Xetra offers the "continuous trading with auctions" service for its more liquid securities. Call auctions occur three times in a regular trading day for DAX 30 stocks: from 8.50 am to 9.00 am at the earliest (opening auction), from 1:00 p.m. to 1:02 p.m. at the earliest (intraday auction), and from 5:30 p.m. to 5:35 p.m. at the earliest (closing auction), with random end times. Continuous trading occurs in between auctions and only these periods are used in our study. See `https://www.xetra.com/xetra-en/trading/trading-models/continuous-trading-with-auctions` for the detailed trading models of Xetra.

we use the Xetra Parser software of Bilodeau (2013) to rebuild the first level of the LOB at microsecond precision for each update. The timestamps are then rounded to the nearest greater millisecond, for use in conjunction with the following data sets. The data of Chi-X and BATS was acquired from BEDOFIH (Base Européenne de Données Financières à Haute Fréquence) and it contains the trades and quotes at a millisecond precision of the first 20 LOB levels, but only the first level is used in this study. The London-based exchanges lag one hour behind Xetra because of different time zones, but all their normal trading hours overlap completely, from opening to closing.

Table 1: DAX 30 constituents from February to July 2013.

Ticker	**Company**	**Prime Standard Sector**
ADS	Adidas	Consumer
ALV	Allianz	Insurance
BAS	BASF	Chemicals
BAYN	Bayer	Chemicals
BEI	Beiersdorf	Consumer
BMW	BMW	Automobile
CBK	Commerzbank	Banks
CON	Continental	Automotive
DAI	Daimler AG	Automobile
DB1	Deutsche Börse	Financial Services
DBK	Deutsche Bank	Banks
DPW	Deutsche Post	Transportation & Logistics
DTE	Deutsche Telekom	Telecommunication
EOAN	E.ON	Utilities
FME	Fresenius Medical Care	Pharma & Healthcare
FRE	Fresenius	Pharma & Healthcare
HEI	HeidelbergCement	Construction
HEN3	Henkel	Consumer
IFX	Infineon Technologies	Technology
LHA	Deutsche Lufthansa	Transportation & Logistics
LIN	Linde	Chemicals
LXS	Lanxess	Chemicals
MRK	Merck	Pharma & Healthcare
MUV2	Munich Re	Insurance
RWE	RWE	Utilities
SAP	SAP	Software
SDF	K+S	Chemicals
SIE	Siemens	Industrial
TKA	Thyssenkrupp	Industrial
VOW3	Volkswagen AG	Automobile

Table 2: Stylized facts of the DAX 30 stocks from February to July 2013.

		Xetra			Chi-X			BATS		
Ticker	**Market Cap ($B)**	**Daily Trades**	**Daily Quotes**	**Daily Volume**	**Daily Trades**	**Daily Quotes**	**Daily Volume**	**Daily Trades**	**Daily Quotes**	**Daily Volume**
ADS	18.62	4 065.37	45 892.78	717 000.01	4 754.04	90 547.10	351 458.26	926.17	29 146.06	58 485.46
ALV	62.80	4 750.37	63 093.81	1 568 537.26	4 812.42	83 739.54	498 074.92	2 079.18	45 718.41	166 271.06
BAS	86.42	7 924.10	95 070.57	2 481 711.35	9 282.54	170 845.96	1 038 437.31	2 434.26	77 506.34	196 013.78
BAYN	78.62	6 687.76	76 045.05	1 661 952.61	10 481.14	157 124.54	935 971.42	1 764.42	58 777.43	127 529.10
BEI	15.28	2 196.01	35 603.93	379 020.22	2 721.87	63 721.97	202 624.60	432.15	22 821.12	25 154.20
BMW	50.68	5 919.62	73 191.91	1 483 157.74	7 250.10	122 096.66	651 476.35	1 484.31	52 993.49	133 672.82
CBK	9.10	7 638.83	64 760.41	30 018 813.68	4 709.13	93 422.61	4 263 311.16	1 221.26	37 149.70	951 537.95
CON	19.05	3 224.14	44 399.19	429 689.75	3 124.57	76 207.83	167 750.59	806.37	25 400.49	42 794.27
DAI	48.04	9 351.49	92 221.54	3 627 361.09	10 005.53	168 268.54	1 376 246.22	1 750.92	63 342.74	157 744.60
DB1	11.24	3 100.38	38 989.17	650 748.07	2 236.83	66 416.24	192 806.29	433.70	21 276.93	28 260.58
DBK	40.49	11 003.10	119 713.20	5 723 773.42	12 085.97	211 213.81	2 336 862.62	2 819.78	105 425.97	443 193.01
DPW	21.85	3 039.34	35 751.58	3 120 529.90	4 083.18	63 482.83	1 564 405.99	1 229.61	29 143.98	373 672.02
DTE	40.43	6 725.98	62 365.61	12 449 292.22	8 727.14	161 058.58	4 918 937.39	1 458.01	71 097.87	834 711.04
EOAN	29.26	5 587.51	64 101.83	9 228 846.25	5 407.94	85 763.46	2 751 795.26	1 610.38	45 003.34	622 245.14
FME	21.21	2 807.30	40 390.46	700 928.89	3 695.55	124 150.12	334 206.74	1 133.18	52 963.98	97 659.41
FRE	16.90	2 711.75	33 894.09	340 475.18	3 680.93	69 445.98	208 713.06	422.45	17 585.19	23 680.45
HEI	9.35	3 317.76	39 276.53	701 370.92	3 524.97	73 321.34	318 912.25	611.42	19 791.14	46 476.07
HEN3	32.19	2 676.42	41 266.30	465 748.64	3 205.38	70 077.32	243 584.49	457.59	24 517.36	36 561.23
IFX	7.17	4 376.42	44 864.96	6 605 088.00	3 946.65	97 041.22	2 250 187.47	984.75	41 728.13	582 747.08
LHA	7.13	2 953.59	36 487.69	2 585 342.06	2 811.82	51 522.81	803 350.74	761.57	19 633.18	209 178.17
LIN	32.42	2 296.01	43 861.00	381 474.56	2 803.41	62 408.79	172 591.52	1 293.37	33 669.54	68 399.42
LXS	6.00	4 184.40	43 147.40	823 462.50	3 252.21	74 075.39	231 579.72	362.42	18 749.78	21 017.78
MRK	23.63	1 392.85	24 840.28	192 377.91	1 906.78	29 967.90	95 725.38	578.02	16 777.25	19 133.98
MUV2	24.40	2 929.64	50 244.27	563 913.20	3 408.33	69 053.92	226 159.97	1 258.56	36 379.85	66 005.40
RWE	20.81	5 389.74	63 971.63	2 909 747.86	5 284.01	102 906.34	928 932.98	920.52	39 324.37	121 170.85
SAP	95.68	6 538.16	67 550.11	2 476 079.29	8 518.55	150 156.54	1 256 262.95	3 144.94	94 518.34	365 707.80
SIE	91.61	7 861.28	72 557.95	2 069 163.66	12 098.28	190 943.03	1 072 561.42	3 015.94	102 190.10	196 583.27
TKA	9.95	3 556.34	41 221.83	3 014 566.10	3 291.10	54 980.92	918 368.38	622.98	22 309.68	142 966.09
VOW3	84.29	5 059.38	61 500.74	936 956.81	5 180.32	89 913.40	321 970.15	2 306.03	57 458.02	115 518.84

Table 3 details the latency to generate our results,[3] and Table 4 shows the trading costs of the three exchanges in 2013.[4] Table 5 documents the rules used by Xetra to determine stocks' tick sizes.[5] Chi-X and BATS subsequently use the same tick sizes for cross-listed stocks also traded at Xetra.

Table 3: Latency for the two exchanges links used in the strategy.

Link	**One-Way Transportation Latency (ms)**	**Exchange Latency (ms)**	**Total Latency (ms)**	**Total Latency Used (ms)**
Chi-X / Xetra	4.15	1.10	5.25	**5**
Chi-X / BATS	~ 0	0.165	0.165	**1**

Table 4: Trading costs associated with sending liquidity-removing and liquidity-providing orders at Xetra, Chi-X, and BATS in 2013.

Exchange	**Liquidity Removal (bps)**	**Liquidity Providing (bps)**
Xetra	0.36	0.36
Chi-X	0.30	(0.15)
BATS	0.15	0.00

Table 5: Tick size δ rules at Xetra

Price Range (€)	$\boldsymbol{\delta}$ **(€)**
$[0, 10)$	0.001
$[10, 50)$	0.005
$[50, 100)$	0.01
$[100, \infty)$	0.05

[3]Table 3 presents latencies found from multiple sources. Note that Chi-X and BATS servers are located in Equinix Slough (LD4), 32km west of Central London, and Xetra servers are in Frankfurt (FR2). Also note that one-way transportation latency is half of a round trip. Sources used are: `https://www.marketsmedia.com/extent-of-adoption-of-microwave-technology-in-europe-revealed` (Chi-X/Xetra one-way on fiber optics to be conservative), Deutsche Börse Group (2013) (Xetra exchange latency), and `https://cdn.cboe.com/resources/press_releases/BATS_Europe_Latency_Update_FINAL.pdf` (BATS exchange latency). Total latencies are rounded to the nearest non-zero integer.

[4]Deutsche Börse Group (2012) contains the trading costs of DAX stocks at Xetra, and `https://www.cboe.com/europe/equities/notices/41029/fee_schedule/` the trading costs of Chi-X and BATS. All trading costs are effective January 2, 2013.

[5]`https://www.xetra.com/xetra-en/trading/trading-models/trading-parameter-tick-size`. All websites referenced in this section were accessed on September 7, 2022.

5 Results and Analysis

5.1 Empirical Lead-Lag Relationships

Table 6 presents the mid-quote lead-lag estimation of Chi-X/Xetra and Chi-X/BATS cross-listed stocks on our data with the discrete time grid $\mathcal{G} = \{0, 1, \ldots, 50, 55, \ldots, 100, 200, \ldots, 1000, 2000, \ldots, 15000\}$ ms.

Table 6: Mid-quote lead-lag estimation using the Hoffman et al. (2013) estimator and Huth & Abergel (2014) $LLR_{X,Y}$ for the links Chi-X/Xetra and Chi-X/BATS on our data.

	Chi-X / Xetra				**Chi-X / BATS**			
Ticker	**Leader**	$\mathbf{LLR_{X,Y}}$	$\widehat{\ell}^*$ **(ms)**	$\widehat{\rho}^{HY}_{X,Y}(\widehat{\ell}^*)$	**Leader**	$\mathbf{LLR_{X,Y}}$	$\widehat{\ell}^*$ **(ms)**	$\widehat{\rho}^{HY}_{X,Y}(\widehat{\ell}^*)$
ADS	Chi-X	1.15	10	0.025	Chi-X	2.94	4	0.034
ALV	Chi-X	2.12	8	0.046	Chi-X	4.00	2	0.157
BAS	Chi-X	1.81	8	0.034	Chi-X	3.32	1	0.039
BAYN	Chi-X	1.93	9	0.065	Chi-X	1.36	2	0.065
BEI	Chi-X	1.07	6	0.059	Chi-X	1.64	2	0.154
BMW	Chi-X	1.21	6	0.094	Chi-X	2.83	4	0.098
CBK	Chi-X	2.21	10	0.077	Chi-X	3.36	1	0.034
CON	Chi-X	1.37	7	0.039	Chi-X	1.89	10	0.033
DAI	Chi-X	1.35	7	0.052	Chi-X	1.07	1	0.051
DB1	Chi-X	1.25	5	0.031	Chi-X	3.58	2	0.120
DBK	Chi-X	1.73	5	0.100	Chi-X	2.81	4	0.105
DPW	Chi-X	1.85	8	0.060	Chi-X	2.77	1	0.060
DTE	Chi-X	2.34	9	0.039	Chi-X	2.12	1	0.206
EOAN	Chi-X	3.98	7	0.030	Chi-X	1.31	0	0.038
FME	Chi-X	1.19	7	0.035	Chi-X	2.89	2	0.032
FRE	Chi-X	1.01	9	0.025	Chi-X	2.16	1	0.085
HEI	Chi-X	1.53	6	0.033	Chi-X	1.07	1	0.306
HEN3	-	-	-	-	Chi-X	7.26	1	0.047
IFX	Chi-X	1.26	7	0.034	Chi-X	2.38	3	0.045
LHA	Chi-X	1.29	6	0.072	Chi-X	7.76	1	0.138
LIN	Chi-X	2.20	8	0.063	Chi-X	1.93	1	0.087
LXS	Chi-X	1.12	10	0.035	Chi-X	2.49	10	0.026
MRK	Chi-X	1.48	7	0.088	Chi-X	1.80	1	0.094
MUV2	Chi-X	1.90	8	0.019	Chi-X	2.89	2	0.061
RWE	Chi-X	1.27	8	0.032	-	-	-	-
SAP	Chi-X	1.56	8	0.062	Chi-X	1.30	0	0.100
SIE	Chi-X	1.92	7	0.064	Chi-X	1.55	0	0.144
TKA	Chi-X	1.59	7	0.047	Chi-X	1.55	1	0.100
VOW3	Chi-X	1.69	8	0.021	Chi-X	1.69	3	0.044

Chi-X leads almost every DAX 30 cross-listed stock also quoted at Xetra and BATS. Exceptions are HEN3 and RWE, where no definitive lead-lag relationship exists between Chi-X/Xetra and Chi-X/BATS, respectively. These stocks are excluded from the rest of the section. An important observation is that $\widehat{\ell}^*$ (measured in milliseconds) is lower-bounded by the actual latency observed between the markets in 2013, i.e., around 4–5 milliseconds for Chi-X/Xetra and around 0–1 millisecond for Chi-X/BATS (see latencies in Section 4). This demonstrates the reliability of the Hoffman et al. (2013) lag estimation. Any lead-lag movement in the lagging exchange that takes longer than latency is theoretically exploitable by a HFTer. The number of potential arbitrage opportunities is presented in the next subsection.

Interestingly, the fact that Chi-X is the leader of DAX 30 stocks is a direct counterexample of some earlier papers where the home market was the main source of price discovery (Grammig et al. (2005), Pascual et al. (2006), Menkveld et al. (2007) and Frijns, Gilbert, et al. (2010)), but it aligns with other contributions demonstrating that the most liquid and most actively traded market leads price discovery (Poshakwale & Theobald (2004), Frijns, Gilbert, et al. (2015), Frijns, Indriawan, et al. (2018)) (see Table 2 in Section 4 for stylized facts). It is also in line with the hypothesis that the market with lower transaction costs will be the source of price discovery (Abhyankar (1995), Brooks, Rew, et al. (2001)) in the case of Chi-X/Xetra relationships (see trading costs in Section 4). This is also known as the "trading cost hypothesis" introduced in Fleming et al. (1996). In the case of the Chi-X/BATS relationships, even though the liquidity-removal cost is higher at Chi-X, HFTs seem to be more active at that exchange than at BATS probably because of the higher liquidity-providing rebates given at Chi-X. Thus, by being colocated at Chi-X, a HFTer should have the best chance of exploiting these lead-lag relationships in DAX 30 stocks, even if Xetra is their home exchange.

From Table 3, we can answer our first question. Indeed, the exchange that is most liquid, most actively traded, and has the highest liquidity-providing rebates will lead the high-frequency prices in the case of cross-listed stocks, even if it is not the home exchange. In our application, Chi-X is the definitive leader of DAX 30 stocks, over Xetra and BATS, for the aforementioned reasons.

5.2 Econometric Model Performance

We choose a lag order of $D = 10$, given that trials on the first two weeks of data show that $r_{C^X_{i-D}}$ and $r_{C^Y_{i-D}}$ are always statistically insignificant in the model for $D > 12$. The model is also losing some predictive power with $D < 10$, so setting $D = 10$ is a good middle ground. The same D is used during the entire six months and for every stock. The models are recurrently trained every five days with past data and are used out-of-sample through the next five-day period, as shown in Figure 2. "Test" sections are out-of-sample periods

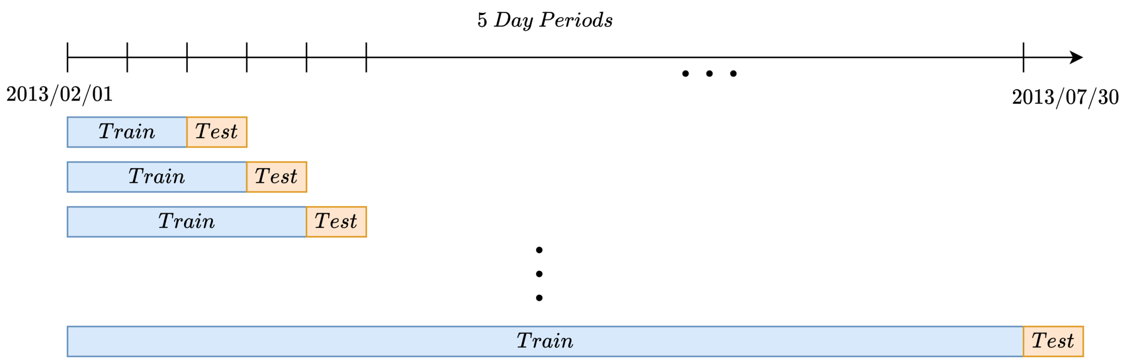

Figure 2: Schema depicting the recurrent training and out-of-sample testing of our model every five days from February 1 to July 31, 2013

where live trading decisions are generated based on the predictions of the models estimated on "train" periods consisting of past days. The first two five-day periods are reserved for the first training iteration, and the first out-of-sample period is the following five days. Other training frequencies were tested, but the model's performance did not significantly change. The decision threshold $K \in [0,\ 1]$ plays an important role in selecting the right opportunities to trade on. Figure 3 exemplifies its effect on the quality of predictions and the number of potential opportunities generated by the model. Increasing K generally results in a higher accuracy in the training sample, but only up to a certain point, at which it tends to decrease. It also drastically reduces the number of potential opportunities, since less and less predicted probability $\max(\widehat{\mathbf{p}}_i) \geq K$ when $K \to 1$. The peak is found every time a model is trained and it is used to select the trading opportunities out-of-sample. This is done independently

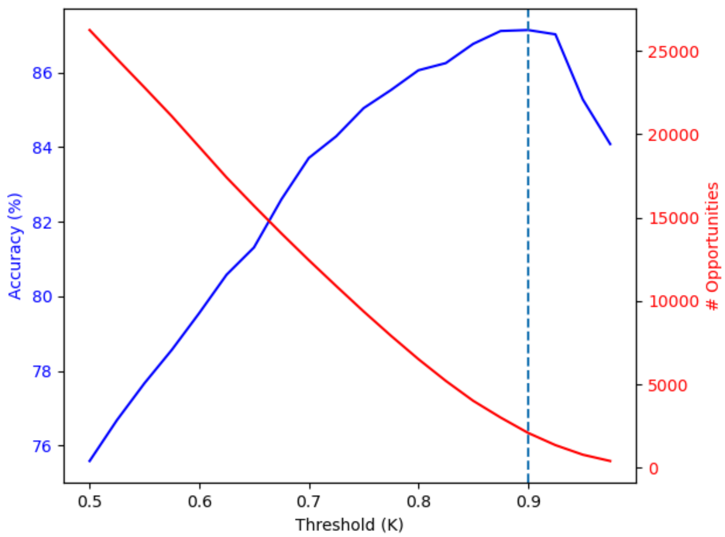

Figure 3: Example of threshold K's effect on model performance, fitted on the bid price processes of Chi-X:DBKd and Xetra:DBK during the first training iteration. The blue line depicts the accuracy and the red one represents the number of potential opportunities, both as a function of K. The dotted vertical line is the peak of the accuracy function on the training sample.

for every stock at each exchange.

We use the model of Alsayed & McGroarty (2014) as a benchmark because a clear comparison can be made between their model and ours. Moreover, the data in both studies come from similar periods. Their predictive framework currently has the best accuracy in the lead-lag arbitrage literature, so it is a suitable point of comparison. The number of potential lead-lag arbitrage opportunities on processes $(\{X_t\}, \{Y_t\})$ is defined as

$$\text{Potential Opportunities}_{\{X,Y\}} = PO_{\{X,Y\}} = \sum_{i=1}^{N} \mathbb{1}_{\left\{\widehat{R}_{C_i^Y} \neq 0\right\}},$$

which represents the number of non-null movement predictions made by a model for the next cluster of the lagging process $\{Y_t\}$. The model accuracy is then defined as

$$\text{Accuracy}_{\{X,Y\}} = \frac{1}{PO_{\{X,Y\}}} \sum_{i=1}^{N} \mathbb{1}_{\left\{\left(\widehat{R}_{C_i^Y} = R_{C_i^Y}\right) \wedge \left(\widehat{R}_{C_i^Y} \neq 0\right)\right\}},$$

the ratio of correct non-null predictions to the total number of potential opportunities. We exclude the null predictions in the accuracy measurement because they do not generate trades. We want to focus on the model's accuracy on actual opportunities. Table 7 summarizes the performance of the Alsayed & McGroarty (2014) predictive model on the mid-quote from our data (see Section 3.2 for details) where δ is the tick size. For the complete per-ticker performance, see Tables 17 and 18 in Appendix B.

Table 7: Alsayed & McGroarty (2014) mid-quote direction performance summary on the six months of data for multiple K^{AM}s.

Threshold (K^{AM})	**Xetra Accuracy**	**Xetra Potential Opportunities**	**BATS Accuracy**	**BATS Potential Opportunities**	**Total Accuracy**	**Total Potential Opportunities**
δ	71.72%	5 187 749	70.37%	4 833 712	71.07%	10 021 461
2δ	70.69%	1 037 573	70.88%	908 307	70.78%	1 945 880
3δ	66.76%	351 333	68.82%	285 449	67.68%	636 782
4δ	64.35%	192 933	67.46%	148 695	65.71%	341 628
5δ	63.20%	126 730	66.72%	95 555	64.71%	222 285
6δ	62.60%	85 101	66.38%	63 081	64.21%	148 182
7δ	62.39%	57 869	66.25%	42 805	64.03%	100 674
8δ	62.54%	38 922	65.88%	28 837	63.96%	67 759
9δ	62.38%	26 356	66.11%	19 655	63.97%	46 011
10δ	62.93%	18 599	65.90%	14 116	64.21%	32 715

Table 8 presents the out-of-sample performance summary of our econometric model on the best bid and ask price processes obtained on the Chi-X/Xetra and Chi-X/BATS lead-lag relationships. Multiple dynamic thresholds are tested to study the importance of K. We begin at the peak, i.e., the values of K on the training sets that generate the highest accuracy from the set $K \in \{0.35, 0.375, 0.40, \ldots, 1\}$, and then decrease K from that starting point by increments of 0.025. For the complete per-ticker performance of our model for both best bid and ask prices processes at Xetra and BATS, see Tables 19 to 22 in Appendix B.

From Tables 7 and 8, we can see that we compare favorably in terms of accuracy. As mentioned earlier, depending only on the latest cluster observation of the leading asset can be hazardous whenever the lead-lag relationship is not as strong as the ones observed in Alsayed & McGroarty (2014), as defined by the $LLR_{X,Y}$. In our cross-listed stock case, fully utilizing the leading and lagging assets' past prices resulted in an average absolute

Table 8: ADLMLR out-of-sample performance summary on the six months of data for multiple Ks.

Threshold (K)	**Xetra Accuracy**	**Xetra Potential Opportunities**	**BATS Accuracy**	**BATS Potential Opportunities**	**Total Accuracy**	**Total Potential Opportunities**
Peak	84.22%	915 666	78.30%	708 580	81.64%	1 624 246
Peak - 0.025	84.25%	1 093 229	78.54%	868 951	81.72%	1 962 180
Peak - 0.050	83.94%	1 262 096	78.42%	1 042 930	81.44%	2 305 026
Peak - 0.075	83.50%	1 428 723	78.09%	1 231 914	81.00%	2 660 637
Peak - 0.100	82.84%	1 614 729	77.55%	1 401 910	80.38%	3 016 639
Peak - 0.125	82.11%	1 817 528	77.00%	1 568 967	79.74%	3 386 495
Peak - 0.150	81.32%	2 028 380	76.29%	1 709 488	79.02%	3 737 868
Peak - 0.175	80.64%	2 162 587	75.37%	1 836 598	78.22%	3 999 185
Peak - 0.200	79.84%	2 264 035	74.60%	1 878 981	77.46%	4 143 016

increase of 10% in total accuracy. As expected, by decreasing the threshold K, we are able to increase the number of potential opportunities at the expense of a lower model accuracy. The financial effect of K is presented in the next subsection.

We also compare the performance of the ADLMLR model to a standard autoregressive distributed lag (ADL) model, where ADLMLR is a classification model trained with maximum likelihood and ADL is a closely related regression model fitted using the OLS method. In Section 3.2, we made the case that ADLMLR has a greater profitability potential than its regression counterpart, which we show here. First, we define the ADL model closest to ADLMLR:

$$r_{C_i^Y} = \alpha + \sum_{j=0}^{D-1} \beta_j r_{C_{i-j}^X} + \sum_{j=1}^{D} \gamma_j r_{C_{i-j}^Y} + \varepsilon_j$$

where $\varepsilon_j \overset{iid}{\sim} N(0, \sigma^2)$ and $D \in \mathbb{N}^+$. In order for that model's performance to be compared to ADLMLR's, the predicted directions of the total variation in cluster C_i^Y are computed as follows:

$$\widehat{R}_{C_i^Y}^{ADL} = \begin{cases} +1, & \text{if } \widehat{r}_{C_i^Y} \geq K^{ADL} \\ 0, & \text{if } -K^{ADL} < \widehat{r}_{C_i^Y} < K^{ADL} \\ -1, & \text{if } \widehat{r}_{C_i^Y} \leq -K^{ADL}. \end{cases}$$

Again, $K^{ADL} \in \mathbb{R}_0^+$ is a preset threshold found dynamically, as described at the beginning of this subsection. Notice that, when we set $D = 1$, $\widehat{\alpha} = 0$, $\widehat{\beta}_0 = 1$, $\widehat{\gamma}_1 = 0$, the model is almost equivalent to Alsayed & McGroarty (2014) (they use the minimum and maximum returns within the leader's cluster, not its total return). Also, when $D = p$, $K^{ADL} = 0$, $\widehat{\alpha} = 0$, and $\widehat{\gamma}_j = 0 \ \forall j$, we get a model similar to Huth & Abergel (2014), but on a quote basis instead of a trade basis. Hence, the ADL model in conjunction with the direction prediction method is a generalization of the predictive framework employed in both studies. Table 9 presents the out-of-sample performance summary of that framework on the best bid and ask price processes selected from a grid of $K^{ADL} \in \{0, \delta, 2\delta, \ldots, 10\delta\}$ with $D = 10$. At its peak, the comparable ADL model achieves an accuracy of 79.64% on a total

Table 9: ADL out-of-sample performance summary on the six months of data

Threshold (K^{ADL})	**Xetra Accuracy**	**Xetra Potential Opportunities**	**BATS Accuracy**	**BATS Potential Opportunities**	**Total Accuracy**	**Total Potential Opportunities**
Peak	84.43%	634 435	75.73%	777 847	79.64%	1 412 282

of 1.4 million potential arbitrage opportunities. On the other hand, as seen in Table 8, the ADLMLR model can reach the same level of accuracy, but on 3.4 million arbitrage opportunities, which is over 140% more than what ADL generates. At its peak, ADLMLR's accuracy outperforms ADL's by an absolute 2% while creating over 200,000 more potential opportunities. This demonstrates that the classification framework of ADLMLR indeed produces a greater profitability potential, as compared to an equivalent regression framework.

To understand how the leading exchange affects the predictive model's performance, we set $\beta_{j,-1} = \beta_{j,0} = 0, \ \forall j = 0, \ldots, D-1$ in the ADLMLR model so that only past cluster returns in the lagging exchange are used to generate predictions for the cross-listed stock at that same exchange. Table 10 shows the results when $K \in \{0.35, 0.375, 0.40, \ldots, 1\}$ is dynamically set at the peak. Not utilizing the lead-lag relationship between Chi-X and the lagging exchanges Xetra and BATS dramatically lowers the model's accuracy compared to Table 8. In fact, it does not significantly outperform a naive forecasting model randomly

Table 10: ADLMLR out-of-sample performance summary on the six months of data without the leading exchange observations ($\beta_{j,-1} = \beta_{j,0} = 0,\ \forall j = 0, \dots, D-1$)

Threshold (K)	**Xetra Accuracy**	**Xetra Potential Opportunities**	**BATS Accuracy**	**BATS Potential Opportunities**	**Total Accuracy**	**Total Potential Opportunities**
Peak	43.62%	1 690 915	42.57%	1 014 306	43.23%	2 705 221

predicting positive or negative returns in the lagging exchange. This random model is able to get an accuracy of 40.10% at Xetra and 40.48% at BATS. Hence, relying only on Xetra and BATS to predict their own future returns is hardly possible because of the poor accuracy. This is in line with the efficient market hypothesis (Fama (1970)). But, using prices observed at Chi-X enables accurate return predictions at lagging exchanges. This is a direct violation of the hypothesis. This is another proof of an existing lead-lag relationship for DAX 30 stocks at these three European exchanges. Additionally, when we set $\gamma_{j,-1} = \gamma_{j,0} = 0,\ \forall j = 1, \dots, D$ without constraining the βs, ADLMLR's accuracy decreases slightly compared to Table 8. This means that the best model employs both the leading and lagging exchange prices to generate its predictions; this is the one used through the remainder of the paper. Huth & Abergel (2014) and Alsayed & McGroarty (2014) only incorporate a subset of that information, but we are able to utilize it all.

We are interested in ADLMLR's performance through time in order to make sure that it is long-lasting and well founded. Figure 4 illustrates the out-of-sample aggregated accuracy of our econometric models when K is set at peak training accuracy and $D = 10$ for every stock and every trading period. The models' out-of-sample accuracies are fairly stationary in time, varying by about 3%, and centered at the temporal mean during the entirety of our data sample. Therefore, ADLMLR is able to generate a robust predictive function based on the lead-lag effect observed between Chi-X/Xetra and Chi-X/BATS. The model performs on average 6% better at Xetra and it constantly outperforms the one fitted at BATS.

From Table 8 and Figure 4, we demonstrate that if there is a lead-lag relationship between any two assets, an adequate econometric model fully utilizing current and past observations of both assets is able to predict the lagging returns with respectable accuracy. In our case, a

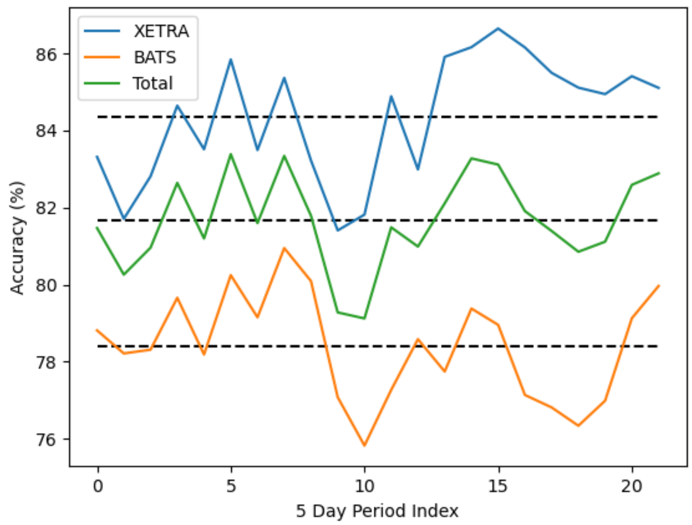

Figure 4: Out-of-sample accuracy in time, weighted on Table 6 selected DAX 30 stocks of our econometric models for Xetra and BATS at each 5-day period from February 1 to July 31, 2013.

generalized form of autoregressive logistic regression can predict the next cluster movement of Xetra's and BATS' best bid and ask prices out-of-sample with an average accuracy exceeding 80%. This is possible because Chi-X led the DAX 30 cross-listed stocks prices.

5.3 Statistical Arbitrage Performance

We compute the performance of the HFT arbitrage strategy of Section 3.3 in two scenarios to determine the lead-lag relationships' financial significance. In the first scenario, we only consider the first half of latency. We observe the LOBs of Xetra and BATS at a delay because the physical distance between these exchanges and Chi-X causes the information to arrive late at that location. In the second scenario, the first half of latency is still considered, but now orders sent to Xetra and BATS also arrive at a delay to account for the second half of latency. Both scenarios consider trading costs and assume the colocation of a server at Chi-X. This allows us to empirically study the effect of latency on the arbitrage strategy's performance.

Table 11 details the performance of the HFT strategy when latency is considered in the case of information arrival, but not when sending orders (scenario 1). By being colocated at Chi-X, we receive Xetra's TAQ data five milliseconds after it is sent by the exchange, and BATS' data is received after one millisecond. But, orders are immediately integrated into Xetra's and BATS's LOBs whenever they are sent by the strategy. As in Table 8, we begin at the peak, i.e., the values of K on the training sets that generate the highest accuracy from the set $K \in \{0.35, 0.375, 0.40, \ldots, 1\}$, and then decrease K from that starting point by increments of 0.025.

Table 11: Performance summary of the HFT arbitrage strategy on six months of 2013 data for the first scenario and multiple Ks.

Threshold (K)	**Xetra Profits (before costs)**	**Xetra Net Profits**	**BATS Profits (before costs)**	**BATS Net Profits**	**Total Profits (before costs)**	**Total Net Profits**
Peak	€ 607 012.83	€ 121 976.42	€ 405 892.76	€ 327 639.91	€ 1 012 905.59	€ 449 616.33
Peak - 0.025	€ 739 347.13	€ 137 268.97	€ 521 868.46	€ 422 886.51	€ 1 261 215.59	€ 560 155.48
Peak - 0.050	€ 880 286.64	€ 151 088.22	€ 634 577.10	€ 514 435.27	€ 1 514 863.74	€ 665 523.48
Peak - 0.075	€ 1 043 393.11	€ 173 071.48	€ 730 251.28	€ 589 171.97	€ 1 773 644.39	€ 762 243.45
Peak - 0.100	€ 1 236 157.68	€ 198 565.95	€ 800 106.40	€ 642 167.31	€ 2 036 264.08	€ 840 733.26
Peak - 0.125	€ 1 443 020.65	€ 214 671.18	€ 847 086.00	€ 674 313.30	€ 2 290 106.65	€ 888 984.48
Peak - 0.150	€ 1 667 881.72	€ 246 440.08	€ 874 899.96	€ 691 162.67	€ 2 542 781.68	€ 937 602.75
Peak - 0.175	€ 1 878 796.74	€ 290 017.36	€ 874 616.61	€ 681 707.30	€ 2 753 413.35	€ 971 724.66
Peak - 0.200	€ 2 058 341.88	€ 318 595.84	€ 849 585.18	€ 652 700.51	€ 2 907 927.06	€ 971 296.35

We stop at $K = \text{Peak} - 0.200$ because it is the point at which the strategy's profitability starts to diminish and continues to do so past that threshold. Table 12 presents the performance of the HFT strategy when latency is also included when sending orders to the market, while still considering information arrival latency (scenario 2), meaning that orders sent to Xetra take five milliseconds to arrive in the LOB, and orders sent to BATS arrive after one millisecond from a colocated server at Chi-X. Full latency is thus considered, being the most realistic scenario, in accounting for important market frictions.

Comparing Table 11 with Table 12, we notice that adding latency to the orders sent by the HFT strategy plays an important role in its net profitability, especially at Xetra. Indeed, net profits at that exchange are reduced by 15%—20%, but the strategy still remains profitable. On the other hand, net profits at BATS do not change dramatically (around 5% change). The geographical proximity of BATS to Chi-X and its lower trading activity and

Table 12: Performance summary of the HFT arbitrage strategy on six months of 2013 data for the second scenario and multiple Ks.

Threshold (K)	**Xetra Profits (before costs)**	**Xetra Net Profits**	**BATS Profits (before costs)**	**BATS Net Profits**	**Total Profits (before costs)**	**Total Net Profits**
Peak	€ 555 628.57	€ 99 890.54	€ 423 989.76	€ 346 240.12	€ 979 618.33	€ 446 130.66
Peak - 0.025	€ 678 371.27	€ 111 282.59	€ 542 331.65	€ 443 909.93	€ 1 220 702.92	€ 555 192.52
Peak - 0.050	€ 809 847.42	€ 120 902.05	€ 657 113.35	€ 537 572.71	€ 1 466 960.77	€ 658 474.76
Peak - 0.075	€ 962 084.37	€ 136 661.09	€ 752 227.96	€ 614 798.29	€ 1 714 312.33	€ 751 459.38
Peak - 0.100	€ 1 146 414.12	€ 158 359.51	€ 828 672.00	€ 671 389.06	€ 1 975 086.12	€ 829 748.57
Peak - 0.125	€ 1 349 241.14	€ 174 914.23	€ 879 732.91	€ 707 631.92	€ 2 228 974.05	€ 882 546.15
Peak - 0.150	€ 1 566 425.26	€ 203 051.37	€ 908 244.56	€ 725 160.98	€ 2 474 669.82	€ 928 212.35
Peak - 0.175	€ 1 773 586.32	€ 244 850.48	€ 910 667.44	€ 718 406.20	€ 2 684 253.76	€ 963 256.68
Peak - 0.200	€ 1 945 885.20	€ 268 122.62	€ 885 587.51	€ 689 348.67	€ 2 831 472.71	€ 957 471.29

liquidity compared to Xetra makes it so that latency does not play an important role on the net profitability. Because of its higher trading costs, its geographical distance to the leading exchange, and its higher level of trading and quoting activity, as compared to BATS, generating net profits from lead-lag arbitrage at Xetra is more challenging. From these results, we show that a HFTer is able to exploit the lead-lag relationship that exists for most DAX 30 stocks cross-listed at Xetra, Chi-X, and BATS even when full latency, non-execution risk, and trade costs are considered. From Table 12, we see that a HFTer can realistically generate an annual net profit of over €1.9 million on DAX 30 stocks alone from the three exchanges, or more than €33,000 on average per cross-listed stock, per exchange. Table 13 presents the detailed performance of the Alsayed & McGroarty (2014) strategy with the most accurate K^{AM}.

The most accurate version of the Alsayed & McGroarty (2014) mid-quote strategy is not able to cover the bid-ask spread and the transaction costs. Almost 100% of the trades in this strategy are not profitable, because there needs to be a variation in the best bid (when closing a long position) and in the best ask (when closing a short position) greater than the bid-ask spread, plus the transaction costs, within a single cluster, which lasts on average around two seconds at both exchanges. This profitable situation occurs 0.65% of the time at Xetra and 0.05% at BATS. Larger values of K^{AM} do not generate better results in terms of net profit per trade, and no K^{AM}s generate a net profitable strategy.

Table 13: Detailed performance of the Alsayed & McGroarty (2014) strategy on six months of 2013 data in the second scenario with the most accurate threshold $K^{AM} = \delta$.

	Xetra	**BATS**
Gross Profit	€ 29 646.60	€ 2 414.60
Loss	-€ 11 530 611.00	-€ - 23 819 383.60
Trading Costs	-€ 1 597 281.21	-€ 317 594.47
Total Net Profit	-€ 13 098 245.61	-€ 24 134 563.47
Median Net Daily Profit	-€ 110 406.68	-€ 201 944.76
Mean Net Daily Profit	- € 115 913.68	-€ 213 580.21
Most Profitable Date (Net Profit)	5/20/2013 (-€ 59 006.03)	7/23/2013 (-€ 97 640.02)
Fifth Most Profitable Date (Net Profit)	7/22/2013 (-€ 63 408.32)	7/22/2013 (-€ 112 638.24)
Least Profitable Date (Net Profit)	2/26/2013 (-€ 290 537.19)	5/2/2013 (-€ 448 107.81)
Fifth Least Profitable Date (Net Profit)	2/21/2013 (-€ 136 761.52)	2/21/2013 (-€ 239 111.62)
Median Trade Time	0.050s	0.021s
Mean Trade Time	2.17s	2.17s
# Net Profitable Trades	27 350	1 917
# Net Unprofitable Trades	4 196 171	4 124 243
# Trades	4 223 521	4 126 160
% Net Profitable Trades	0.65%	0.05%
Mean Volume per Trade	100	100
Mean Net Profit per Profitable Trade	€ 0.68	€ 1.17
Mean Net Profit per Unprofitable Trade	-€ 3.12	-€ 5.85

We also demonstrate that a mid-quote-based market order HFT strategy, like the one in Huth & Abergel (2014) and Alsayed & McGroarty (2014) is not viable in practice. To do so, we assume a perfect model that is always able to predict the exact mid-quote return of the lagging asset's next cluster. If that return is above (under) a threshold $K^{Perfect}$ ($-K^{Perfect}$), the strategy opens a long (short) position with a buy (sell) market order at the best ask (bid) right before the lagging asset's next cluster begins. The position is then closed with an opposite market order precisely when the lagging asset's cluster ends. This is the buy-and-hold HFT strategy of Alsayed & McGroarty (2014). Huth & Abergel (2014) employ the same type of strategy, but on a trade basis with a threshold of 0. Table 14 presents this best case mid-quote-based market order HFT strategy on our data in the second scenario.

Even though the predictive model is perfectly accurate on the next mid-quote return of the lagging asset, gross profits never cover the bid-ask spread cost of market orders. This is the only source of losses in Table 14. Thus, it is impossible to profit from high-frequency lead-lag arbitrage from mid-quote return predictions and a market order—based HFT strategy. It also shows that at the millisecond scale, asset returns rarely cover market order trading costs. This means that the trading signal of Stübinger (2019) would also generate inconsiderable

Table 14: Performance of a best case mid-quote-based market order HFT strategy on six months of 2013 data in the second scenario for multiple $K^{Perfect}$.

Threshold ($K^{Perfect}$)	**# Trades**	**% Net Profitable Trades**	**Gross Profit (€)**	**Loss (€)**	**Trading Costs (€)**	**Total Net Profits (€)**
δ	11 383 116	0.50%	69 458.80	-44 766 290.20	-2 677 746.80	-47 374 578.20
2δ	2 881 086	1.36%	49 935.80	-17 000 431.90	-596 335.46	-17 546 831.56
3δ	1 226 077	1.67%	30 536.10	-10 368 211.80	-197 800.70	-10 535 476.40
4δ	723 858	1.40%	20 857.80	-7 427 306.80	-94 245.93	-7 500 694.93
5δ	427 531	1.30%	15 601.00	-5 414 911.90	-52 933.31	-5 454 244.21
6δ	303 438	1.27%	13 100.10	-4 097 537.00	-34 348.47	-4 118 785.37
7δ	180 751	1.50%	11 385.10	-2 990 559.20	-21 215.59	-3 000 389.69
8δ	113 714	1.82%	10 195.90	-2 332 527.70	-14 061.15	-2 336 392.95
9δ	71 894	2.15%	9 057.30	-1 873 618.20	-9 477.26	-1 874 038.16
10δ	47 844	2.54%	8 251.40	-1 537 993.50	-6 420.21	-1 535 892.31

profits in this setting. Switching from market orders to limit orders eliminates the necessity of covering the bid-ask spread and facilitates access to profitability. It is also important to know what side(s) of the LOB will generate the non-zero mid-quote return to capture arbitrage opportunities and mid-quote returns do not provide that information. Predicting the best bid and best ask returns allows better-informed trading decisions. Table 15 presents the detailed performance of our limit order-based strategy with the most profitable K in the second scenario.

Table 15: Detailed performance of the HFT strategy on six months of 2013 data in the second scenario with the most profitable threshold $K = \text{Peak} - 0.175$.

	Xetra	**BATS**
Gross Profit	€ 3 365 103.46	€ 2 108 945.01
Loss	-€ 1 591 517.13	-€ 1 198 277.57
Trading Costs	-€ 1 528 735.84	-€ 192 261.24
Total Net Profit	€ 244 850.48	€ 718 406.20
Median Net Daily Profit	€ 1 942.99	€ 6 071.93
Mean Net Daily Profit	€ 2 166.82	€ 6 357.58
Most Profitable Date (Net Profit)	6/11/2013 (€ 9 987.13)	2/26/2013 (€ 16 118.00)
Fifth Most Profitable Date (Net Profit)	6/24/2013 (€ 5 721.97)	2/25/2013 (€ 11 053.02)
Least Profitable Date (Net Profit)	5/2/2013 (-€ 2 289.61)	5/9/2013 (€ 2 811.62)
Fifth Least Profitable Date (Net Profit)	2/21/2013 (€ 1 237.24)	7/26/2013 (€ 4 173.32)
Median Trade Time	1.02s	1.44s
Mean Trade Time	27.82s	28.45s
# Net Profitable Trades	1 158 049	1 002 859
# Net Unprofitable Trades	223 452	223 998
# Trades	1 381 501	1 226 857
% Net Profitable Trades	83.83%	81.74%
Mean Volume per Trade	503.64	352.29
Mean Net Profit per Profitable Trade	€ 1.79	€ 1.95
Mean Net Profit per Unprofitable Trade	-€ 8.20	-€ 5.51

Gross profits are considerable in both exchanges. But, losses incurred from execution-related risks are also sizeable, drastically decreasing the net profitability of the strategy, by approximately 50%. Losses occur whenever the model predictions are wrong, which directly results in limit orders not being executed because the lagging assets' level 1 prices have drifted away from the specified limit price. At that point, losses are cut by stop limit orders. When these limit orders are also not executed, market orders are sent to finally close the position after M minutes (15 minutes for Xetra and 20 for BATS; see Appendix A for details). Losses can also occur even when the model is right, but limit orders remain in the queue without ever being executed.

Exchange trading costs are also significant, especially at Xetra, given its prohibitive fee structure relative to BATS. This was expected, given the large number of trades and their limited profitability because of the brief holding period typical of HFT strategies. Overall, considering losses and trading costs, a net profit margin of 7% was obtained at Xetra and 34% at BATS, where the significant difference stems from that difference in their fee structure and from the longer latency to trade on Xetra from Chi-X. All order types are expensive at Xetra, whereas liquidity-providing orders are free at BATS and liquidity-taking fees are less than half of Xetra's (see Table 4 for all fees).

Median trading times are quick at both exchanges, though slightly longer at BATS because of its lower level of trading activity compared to Xetra (see Table 2). Mean trading times are greater than the median, given the non-execution risk of limit orders, which can stay for up to M minutes in the LOB without being executed. The proportions of net profitable trades are in line with model accuracy for both exchanges. Once again, we notice the importance of execution-related risks from the difference between the performance of profitable and non-profitable trades. In fact, the mean loss incurred is over 4.58 times greater than the mean profit per trade at Xetra, and the same ratio is over 2.82 at BATS. Without risk management procedures, these ratios are even greater. Table 16 presents the detailed performance of the HFT strategy excluding stop-limit orders, maximum level 1 price variation, and the no-microstructure-change rule (see Appendix A for details). We leave the time breaker of M minutes before closing positions; otherwise they can stay open for days and no trade occurs in that time because the strategy waits for the previous position to close before opening the next. This is a consequence of the non-execution risk of limit orders.

Table 16: Detailed performance of the HFT strategy on six months of 2013 data in the second scenario with the most profitable threshold $K = \text{Peak} - 0.175$ without risk management.

	Xetra	**BATS**
Gross Profit	€ 1 569 778.02	€ 1 146 940.41
Loss	-€ 767 202.96	-€ 604 726.34
Trading Costs	-€ 624 995.42	-€ 100 639
Total Net Profit	€ 177 579.64	€ 441 575.05
Median Net Daily Profit	€ 1 453.54	€ 3 766.97
Mean Net Daily Profit	€ 1 571.50	€ 3 907.74
Most Profitable Data (Net Profit)	6/13/2013 (€ 4 707.61)	3/6/2013 (€ 15 650.04)
Fifth Most Profitable Date (Net Profit)	5/23/2013 (€ 3 909.88)	2/27/2013 (€ 7 581.48)
Least Profitable Date (Net Profit)	7/5/2013 (-€ 2 870.90)	7/19/2013 (-€ 2 684.23)
Fifth Least Profitable Date (Net Profit)	2/21/2013 (€ 1 320.75)	6/21/2013 (€ 699.45)
Median Trade Time	1.80s	1.76s
Mean Trade Time	75.22s	79.81s
# Net Profitable Trades	456 981	546 489
# Net Unprofitable Trades	55 701	17 172
# Trades	512 682	563 661
% Net Profitable Trades	89.14%	96.95%
Mean Volume per Trade	338.45	213.47
Mean Net Profit per Profitable Trade	€ 2.14	€ 1.92
Mean Net Profit per Unprofitable Trade	-€ 14.33	-€ 35.52

As expected, the ratio of mean loss to mean profit per trade incurred at Xetra climbs from 4.58 to 6.70 and soars from 2.82 to 18.50 at BATS. More importantly, the net profitability decreased by 27% at Xetra and by 39% at BATS. Nonetheless, the strategy remains profitable at both exchanges. The largest difference between Table 15 and Table 16 comes from the absence of stop-limit orders. Without them, the positions stay open as long as the profit-taking limit orders are not executed, up to M minutes. The average trade duration more than doubles, hence reducing the number of arbitrage opportunities captured by about the same quotient. Risk management procedures are thus useful in preventing large losses by mitigating the non-execution risk of limit orders while also closing positions rapidly when prices drift away for them.

Figure 5 presents the cumulative net profit of the most profitable version of the strategy in scenario 2 (see Table 15). The strategy has minimal drawdown and constantly generates a positive net profit on a daily basis. Table 15 answers our final question. If a lead-lag

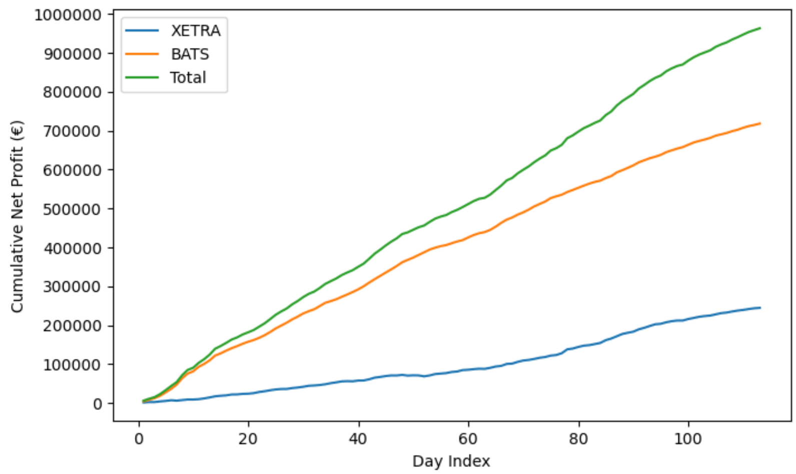

Figure 5: Cumulative net profit of the HFT strategy on a daily basis for Table 6 selected DAX 30 stocks from February 1 to July 31 2013 in the second scenario with the most profitable threshold $K = Peak - 0.175$.

relationship exists between two assets and if a predictive model is able to exploit it, a HFTer can in fact realistically earn a profit. As shown in the same table, the execution-related risks were the main impediment to lead-lag arbitrage, followed by trading costs and latency, based on Table 11 versus Table 12. Nonetheless, an HFT strategy that was colocated at the leading exchange and that relied mainly on limit orders was able to profit from the lead-lag relationship that existed between DAX 30 stocks cross-listed at Xetra, Chi-X, and BATS in 2013.

6 Conclusion

In this paper, we investigate the existence, predictability, and profitability of lead-lag relationships at a high frequency with an application to DAX 30 stocks cross-listed at Xetra in Frankfurt, and Chi-X and BATS, both in London, during six months of 2013. Using the robust lead-lag estimator of Hoffman et al. (2013) and the lead-lag ratio of Huth & Abergel

(2014), we first show that Chi-X leads level 1 prices by mere milliseconds. This is in line with previous studies showing that the most actively traded, liquid, and least expensive exchange will ultimately be the price discovery origin of arbitrage-linked assets. The lead-lag estimation demonstrates the great interconnectedness between the three exchanges by showing that their lag is approaching, or even equating, the physical speed limit at which information could travel between them at that time. This level of precision is the highest in the cross-listed stocks lead-lag literature. It was previously unattainable because of the Epps effect (Epps (1979)), which would have generated biased estimations at high frequencies for previous-tick-based methodologies employed by prior studies (Zhang (2011)).

After establishing the existence of lead-lag relationships in DAX 30 cross-listed stocks, we develop a new predictive modeling framework based on the concept of clusters proposed by Alsayed & McGroarty (2014), in conjunction with a new, generalized version of the autoregressive logistic regression. Clusters allow us to depart from uniformly sampled observations to instead employ the unadulterated LOB updates. Our econometric model employs past and current asset prices to forecast a classification of the next clusters' return: positive, null, or negative. This probabilistic framework generates an out-of-sample return accuracy exceeding 80%, with a solidly maintained performance throughout our data period, thereby comparing advantageously to the other models put forth in the literature. Indeed, the proposed approach is able to detect substantially more potential arbitrage opportunities, with an even greater accuracy than previous regression models.

We then introduce a new high-frequency trading strategy built around our predictive model to profit from the detected lead-lag relationships. Previous studies failed to generate viable high-frequency strategies because of the steep costs associated with market orders (Brooks, Rew, et al. (2001), Huth & Abergel (2014)). In these studies, paying the bid-ask spread and the exchange trading costs was too prohibitive to exploit intraday lead-lag relationships. To go further, we empirically demonstrate the non-viability of mid—quote and market order-based strategies in the context of high-frequency lead-lag arbitrage. The results show the quasi-impossibility of such a strategy to cover even the bid-ask spread when lags exist at the sub-second scale. The strategy we propose relies instead on limit orders and LOB signals to cut on these costs, at the expense of adding a non-execution risk. In a scenario where major market frictions are present, we demonstrate that high-frequency traders could realistically earn a profit with our limit order—based strategy. More precisely, they could generate an annual net profit above €1.9 million from DAX 30 stocks alone and only two exchanges (Xetra and BATS) with a colocated server at Chi-X. We show that execution-related risks, trading costs, and latency (in that order) are important impediments to lead-lag arbitrage, and that risk management measures are necessary to alleviate their impact on profitability.

Our goal was to demonstrate how a high-frequency trader would theoretically be able to profit from lead-lag arbitrage and empirically show that possibility with a pragmatic approach. We intended to develop a complete framework incorporating the detection, prediction, and trading of lead-lag relationships for any pair of assets. The framework empirically achieved that for cross-listed stocks, hence advancing knowledge on lead-lag in high-frequency markets and answering queries about their financial importance (Curme et al. (2015), Basnarkov et al. (2020)). The proposed framework is also general enough to be used on any pair of assets.

Our study covered the application of high-frequency lead-lag relationships in an arbitrage context. Li et al. (2022) demonstrate how the daily lead-lag effect significantly improves the profitability of alpha-factor strategies. In that sense, the statistical relationship, predictive model, and backtesting methodology presented in this paper could also be investigated for other types of strategies, like market making. Being able to predict an asset's level 1 prices from another related and leading asset would probably prove beneficial for market markers. It would also be worthwhile to quantify the financial viability of lead-lag relationships in other asset classes and markets, and during different time periods with the proposed framework, or any other that might come.

References

Abhyankar, A.N. (1995). Return and Volatility Dynamics in the FT-SE 100 Stock Index and Stock Index Futures Markets. Journal of Futures Markets, 15.4, 457–488.

Alsayed, H. & McGroarty, F. (2014). Ultra-High-Frequency Algorithmic Arbitrage Across International Index Futures. Journal of Forecasting, 33, 391–408.

Basnarkov, L., Stojkoski, V., Utkovski, Z., & Kocarev, L. (2020). Lead-lag relationships in foreign exchange markets. Physica A: Statistical Mechanics and its Applications, 539.1.

Bilodeau, Y. (2013). Xetra parser [computer software]. HEC Montréal.

Bollen, N.P.B., O'Neill, M.J., & Whaley, R.E. (2017). Tail Wags Dog: Intraday Price Discovery in VIX Markets. Journal of Financial Markets, 37.5, 431–451.

Bonney, G.E. (1987). Logistic Regression for Dependent Binary Observations. Biometrics, 43.4, 951–973.

Brooks, C., Garrett, I., & Hinich, M.J. (1999). An alternative approach to investigating lead-lag relationships between stock and stock index futures markets. Applied Financial Economics, 9, 605–613.

Brooks, C., Rew, A.G., & Ritson, S. (2001). A trading strategy based on the lead-lag relationship between the spot index and futures contract for the FTSE 100. International Journal of Forecasting, 17, 31–44.

Broyden, C.G. (1970). The convergence of a class of double-rank minimization algorithms. Journal of the Institute of Mathematics and Its Applications, 6, 76–90.

Budish, E., Cramton, P., & Shim, J. (2015). The High-Frequency Trading Arms Race: Frequent Batch Auctions as a Market Design Response. The Quaterly Journal of Economics, 130.4, 1547–1621.

Chan, K. (1992). A Further Analysis of the Lead-Lag Relationship Between the Cash Market and Stock Index Futures Market. The Review of Financial Studies, 5.1, 123–152.

Chen, Y., Da, Z., & Huang, D. (2019). Arbitrage trading: The long and the short of it. The Review of Financial Studies, 32, 1608–1646.

Chen, Y.L. & Gau, Y.F. (2010). News announcements and price discovery in foreign exchange spot and futures markets. Journal of Banking and Finance, 34, 1628–1636.

Curme, C., Tumminello, M., Mantegna, R.N., Stanley, H.E., & Kenett, D.Y. (2015). Emergence of statistically validated financial intraday lead-lag relationships. Quantitative Finance, 15.8, 1375–1386.

Deutsche Börse Group (2012). 124/2012 Amendment to the Price List for the Utilization of the Exchange EDP of FWB Frankfurt Stock Exchange and of the EDP XONTRO. https://www.deutsche-boerse-cash-market.com/dbcm-en/newsroom/circulars/Xetra-circulars-mailings. Accessed on 7 September 2022.

Deutsche Börse Group (2013). Presentation: Investor Day 2013. https://www.deutsche-boerse.com/dbg-en/investor-relations/presentations. Accessed on 7 September 2022.

Dimpfl, T. & Jung, R.C. (2012). Financial market spillovers around the globe. Applied Financial Economics, 22.1, 45–57.

Engle, R.F. & Granger, C.W.J. (1987). Cointegration and error correction: representation, estimation and testing. Econometrica, 55, 251–276.

Epps, T.W. (1979). Comovements in stock prices in the very short-run. Journal of the American Statistical Association, 74.366, 291–298.

Fama, E.F. (1970). Efficient Capital Markets: A Review of Theory and Empirical Work. The Journal of Finance, 25.2, 383–417.

Fleming, J., Ostdiek, B., & Whaley, R.E. (1996). Trading costs and the relative rates of price discovery in stock, futures and options markets. Journal of Futures Markets, 4, 353–387.

Fletcher, R. (1970). A New Approach to Variable Metric Algorithms. Computer Journal, 13.3, 317–322.

Foucault, T. & Biais, B. (2014). HFT and market quality. Bankers, Markets & Investors, 128, 5–19.

Frijns, B., Gilbert, A., & Tourani-Rad, A. (2010). The dynamics of price discovery for cross-listed shares: Evidence from Australia and New Zealand. Journal of Banking and Finance, 34, 498–508.

Frijns, B., Gilbert, A., & Tourani-Rad, A. (2015). The determinants of price discovery: Evidence from US-Canadian cross-listed shares. Journal of Banking and Finance, 59, 457–468.

Frijns, B., Indriawan, I., & Tourani-Rad, A. (2018). The interactions between price discovery, liquidity and algorithmic trading for U.S.-Canadian cross-listed shares. International Review of Financial Analysis, 56, 136–152.

Frino, A. & West, A. (2003). The impact of transaction costs on price discovery: Evidence from cross-listed stock index futures contracts. Pacific-Basin Finance Journal, 11, 139–151.

Ghadhab, I. & Hellara, S. (2016). Price discovery of cross-listed stocks. International Review of Financial Analysis, 44, 177–188.

Goldfarb, D. (1970). A Family of Variable Metric Updates Derived by Variational Means. Mathematics of Computation, 24.109, 23–26.

Gonzalo, J. & Granger, C.W.J. (1995). Estimation of common long-memory components in cointegrated systems. Journal of Business and Economic Statistics, 13, 1–9.

Grammig, J., Melvin, M., & Schlag, C. (2005). Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects. Journal of Empirical Finance, 12, 139–164.

Granger, C.W.J. (1969). Investigating causal relations by econometric models and cross-spectral methods. Econometrica, 37.3, 424–438.

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

Hasbrouck, J. & Saar, G. (2013). Low-Latency Trading. Journal of Financial Markets, 16, 646–679.

Hayashi, T. & Koike, Y. (2018). Wavelet-Based Methods for High-Frequency Lead-Lag Analysis. SIAM Journal of Financial Mathematics, 9.4, 1208–1248.

Hayashi, T. & Yoshida, N. (2005). On covariance estimation of non-synchronously observed diffusion processes. Bernouilli, 11.2, 359–379.

Herbst, A.F., McCormack, J.P., & West, E.N. (1987). Investigation of a Lead-Lag Relationship between Spot Stock Indices and Their Futures Contracts. The Journal of Futures Markets, 7.4, 373–381.

Hoffman, M., Rosenbaum, M., & Yoshida, N. (2013). Estimation of the Lead-lag Parameter from Non-Synchronous Data. Bernouilli, 19.2, 426–461.

Hou, K. (2007). Industry Diffusion and the Lead-Lag Effect in Stock Returns. The Review of Financial Studies, 20.4, 1113–1138.

Huth, A. & Abergel, F. (2014). High frequency lead/lag relationships - Empirical facts. Journal of Empirical Finance, 26, 41–58.

Jong, F. & Nijman, T. (1997). High frequency analysis of lead-lag relationships between financial markets. Journal of Empirical Finance, 4.2–3, 259–277.

Judge, A. & Reancharoen, T. (2014). An empirical examination of the lead-lag relationship between spot and futures markets Evidence from Thailand. Pacific-Basin Finance Journal, 29, 335–358.

Kawaller, I.G., Koch, P.D., & Koch, T.W. (1987). The Temporal Price Relationship between S& P 500 Futures and the S& P 500 Index. The Journal of Finance, 42.5, 1309–1329.

Li, Y., Wang, T., Sun, B., & Liu, C. (2022). Detecting the lead–lag effect in stock markets: definition, patterns, and investment strategies. Financial Innovation, 8.51. DOI: 10.1186/s40854-022-00356-3. URL: https://doi.org/10.1186/s40854-022-00356-3.

Menkveld, A.J., Koopman, S.J., & Lucas, A. (2007). Modeling Around-the-Clock Price Discovery for Cross-Listed Stocks Using State Space Methods. Journal of Business and Economic Statistics, 25.2, 213–225.

Nguyen, V.L., Shaker, M.H., & Hüllermeier, E. (2022). How to measure uncertainty in uncertainty sampling for active learning. Machine Learning, 111, 89–122.

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

Pascual, R., Pascual-Fuster, B., & Climent, F. (2006). Cross-listing, price discovery and the informativeness of the trading process. Journal of Financial Markets, 9, 144–161.

Poshakwale, S. & Theobald, M. (2004). Market capitalisation, cross-correlations, the lead/lag structure and microstructure effets in the Indian stock market. Journal of International Financial Markets, Institutions and Money, 14.4, 385–400.

Poutré, C., Dionne, G., & Yergeau, G. (2021). International High-Frequency Arbitrage for Cross-Listed Stocks. URL: https://ssrn.com/abstract_id=3890433.

Putniņš, T.J. (2013). What do price discovery metrics really measure? Journal of Empirical Finance, 23, 68–83.

Shanno, D.F. (1970). Conditioning of quasi-Newton methods for function minimizations. Mathematics of Computation, 24.111, 647–656.

Stübinger, J. (2019). Statistical arbitrage with optimal causal paths on high-frequency data of the S&P 500. Quantitative Finance, 19.6, 921–935.

Tse, Y.K. (1995). Lead-Lag Relationship Between Spot Index and Futures Price of the Nikkei Stock Average. Journal of Forecasting, 14, 553–563.

Yang, J., Yang, Z., & Zhou, Y. (2012). Intraday Price Discovery and Volatility Transmission in Stock Index and Stock Index Futures Markets: Evidence from China. The Journal of Futures Markets, 32.2, 99–121.

Zhang, L. (2011). Estimating covariation: Epps effect, microstructure noise. Journal of Econometrics, 160.1, 33–47.

Appendices

A High-Frequency Strategy - Additional Details

The strategy has two important variables controlling its performance: the time breaker's delay M, and the order volume $V_t^{Bid/Ask}$. To select M at Xetra and BATS, we tested its financial effect on the first out-of-sample period. We ran the HFT strategy in that timeframe with $M \in \{5, 6, 7, 8, 9, 10, 15, 30, 60, 90, 120, 300, 600, 900, 1200, 1800, 3600\}$ seconds at the two exchanges. $M = 900$ seconds produced the greatest profitability in that first period at Xetra, and $M = 1200$ seconds at BATS. These values where then set for the entirety of our data, since dynamically selecting them (like we did for K) is computationally very expensive. As shown in Figure 5, net profits are fairly constant in time, a sign that the strategy does not suffer from a preset M.

As for $V_t^{Bid/Ask}$, it follows the median level 1 volume of the last 500 LOB updates, rounded to the closest lowest 100 to trade on round lots. Using more LOB updates does not significantly affect the volume sent by the strategy and does not greatly impact the strategy's performance. More formally, given LOB update indices $t = 1, 2, \ldots, T$, the order volume that is sent by the HFT strategy is calculated by

$$V_t^{Bid/Ask} = 100 \left\lfloor \frac{\tilde{v}_t^{Bid/Ask}}{100} \right\rfloor, \ \forall t \geq 500$$

where $\tilde{v}_t^{Bid/Ask}$ is the empirical median of the sample $v_{t-499:t}^{Bid/Ask}$, for $v_t^{Bid/Ask} \in \mathbb{N}^+$ the best bid/ask volume at index t. No order is sent to the market before observing 500 LOB updates. This is done independently for every stock at Xetra and BATS. Using a windowed median volume limits the market impact of the strategy and the liquidity risk, because the orders dynamically and conservatively follow the liquidity present in the LOB.

To mitigate risk even more, orders are only sent when three conditions are respected:

1. The last in-cluster return of the leader C^X_{i,n^X_i} is not generated by a trade;

2. The realized local variation of level 1 price at the lagging exchange is under a preset threshold;

3. No microstructure change has occurred.

The first condition is present so that the strategy does not to open a position whenever child orders hit the same ticker at multiple exchanges and at the same time. When that occurs, the LOBs of all exchanges move in the same direction at the same time. The strategy cannot profit from that situation since it depends on delayed movements of the LOB at the lagging exchange.

The second condition limits execution-related risks by not opening a position when the volatility of level 1 prices of the LOB is too great, as measured from the previous 50 prices. Given LOB update indexes $t = 1, 2, \dots T$, the realized local variation is defined as

$$RLV_t^{Bid/Ask} = \sum_{i=0}^{49} \left| p_{t-i}^{Bid/Ask} - p_{t-i-1}^{Bid/Ask} \right|,$$

where $p_t^{Bid/Ask} \in \mathbb{R}^+$ the best bid/ask price at index t. Whenever $RLV_t^{Bid/Ask} > \delta W$ for $\delta \in \mathbb{R}^+$ the tick size and $W \in \mathbb{R}_0^+$ a preset threshold, the strategy does not send orders. W is found from the set $\{5, 10, 25, 50, 75, 100, 150, 200, 250, 500\}$ in the same way as M. $W = 100$ at Xetra and $W = 25$ at BATS.

The third condition relates to changes in the tick size of the stock. Whenever this microstructure shock occurs, the strategy stops trading the given ticker until it returns to its initial tick size. This is for simplicity, because the models would need a more complex fitting method to accommodate such an event. See Table 5 for the tick size rules.

B Econometric Model Performance - Additional Results

Table 17: Alsayed & McGroarty (2014) mid-quote direction predictions computed on six months of data for each ticker at Xetra.

	Accuracy					Potential Opportunities				
Ticker \ K^{AM}	**δ**	**2δ**	**3δ**	**4δ**	**5δ**	**δ**	**2δ**	**3δ**	**4δ**	**5δ**
ADS	71.73%	70.09%	66.10%	64.53%	64.16%	207 093	51 518	22 702	13 452	9 151
ALV	74.75%	64.11%	70.57%	75.58%	76.09%	41 165	1 120	316	172	138
BAS	75.07%	75.22%	66.57%	64.39%	63.36%	227 108	28 800	9 027	5 218	3 428
BAYN	73.85%	75.73%	67.79%	64.09%	63.19%	251 731	41 221	13 036	6 897	4 507
BEI	74.88%	71.65%	66.18%	63.49%	62.70%	110 355	21 086	8 992	5 368	3 697
BMW	74.69%	75.52%	68.71%	66.20%	63.27%	236 652	38 946	11 764	5 911	3 458
CBK	65.76%	67.46%	65.92%	63.64%	62.19%	339 938	99 731	26 519	14 294	10 138
CON	70.63%	69.95%	67.23%	65.26%	64.05%	222 551	71 282	31 413	17 958	11 853
DAI	73.50%	72.81%	67.13%	64.08%	62.26%	440 231	91 651	34 176	18 742	11 843
DB1	69.11%	67.82%	65.97%	63.45%	61.12%	201 577	62 592	24 100	13 997	9 530
DBK	73.05%	73.51%	69.02%	65.00%	63.57%	416 043	61 684	15 043	6 782	3 895
DPW	73.55%	67.99%	65.74%	66.67%	70.11%	41 428	4 792	2 201	1 026	435
DTE	66.62%	69.87%	70.03%	69.12%	68.98%	430 101	68 517	9 369	4 015	2 495
EOAN	69.12%	64.31%	64.30%	64.43%	65.03%	40 405	1 681	479	194	143
FME	73.10%	68.36%	64.72%	62.99%	62.31%	114 853	20 998	10 183	6 552	4 590
FRE	71.30%	71.24%	68.13%	64.80%	63.59%	197 753	64 364	30 590	17 911	12 534
HEI	72.77%	70.98%	66.93%	65.48%	64.74%	174 606	38 452	14 460	8 213	5 385
HEN3	-	-	-	-	-	-	-	-	-	-
IFX	72.55%	69.66%	66.74%	65.31%	64.66%	281 850	78 420	21 257	10 756	7 505
LHA	70.92%	64.87%	61.73%	60.33%	64.07%	54 939	5 346	1 769	663	398
LIN	71.27%	65.19%	66.60%	68.57%	69.71%	24 285	2 215	991	385	241
LXS	69.45%	67.71%	65.46%	63.27%	63.28%	219 530	55 455	18 168	10 097	6 500
MRK	67.86%	65.75%	63.70%	62.84%	61.88%	23 184	3 425	1 193	705	522
MUV2	73.23%	63.18%	58.61%	60.67%	60.24%	33 239	2 010	691	239	166
RWE	72.20%	69.76%	64.77%	61.57%	58.82%	171 798	23 412	7 991	4 267	2 295
SAP	73.57%	71.46%	67.13%	65.98%	63.31%	145 500	19 100	7 140	3 957	2 164
SIE	75.06%	74.51%	66.10%	63.62%	62.94%	243 347	32 407	10 405	6 063	3 999
TKA	68.21%	64.93%	61.17%	59.04%	63.31%	73 814	7 277	2 045	791	387
VOW3	74.12%	67.13%	60.44%	60.52%	64.56%	77 109	6 021	1 691	580	285

Table 18: Alsayed & McGroarty (2014) mid-quote direction predictions computed on six months of data for each ticker at BATS.

	Accuracy					Potential Opportunities				
Ticker \K^{AM}	$\boldsymbol{\delta}$	$\boldsymbol{2\delta}$	$\boldsymbol{3\delta}$	$\boldsymbol{4\delta}$	$\boldsymbol{5\delta}$	$\boldsymbol{\delta}$	$\boldsymbol{2\delta}$	$\boldsymbol{3\delta}$	$\boldsymbol{4\delta}$	$\boldsymbol{5\delta}$
ADS	72.51%	72.42%	69.19%	67.92%	67.77%	186 741	44 983	18 573	10 650	7 115
ALV	65.17%	70.67%	78.06%	81.15%	80.00%	40 214	1 040	319	191	160
BAS	76.58%	77.14%	70.22%	68.74%	68.41%	226 470	26 610	7 384	4 171	2 741
BAYN	75.83%	76.90%	70.43%	67.43%	67.25%	244 096	37 637	10 625	5 164	3 258
BEI	73.23%	74.96%	71.75%	69.47%	69.47%	108 102	19 604	7 770	4 366	2 817
BMW	72.53%	72.88%	68.70%	67.88%	65.19%	211 945	34 597	9 453	4 399	2 439
CBK	60.51%	64.05%	64.79%	64.84%	64.07%	314 401	88 256	23 712	12 619	8 984
CON	71.37%	71.14%	69.29%	67.96%	66.96%	159 198	52 819	22 658	12 242	7 924
DAI	70.45%	70.31%	66.20%	64.21%	63.58%	429 938	86 667	30 589	16 030	9 785
DB1	69.37%	70.42%	70.56%	69.60%	68.62%	161 937	49 272	18 013	9 583	6 132
DBK	76.19%	76.86%	72.58%	69.54%	67.83%	410 400	61 851	14 984	6 539	3 671
DPW	71.83%	70.41%	69.59%	70.44%	69.88%	41 742	4 160	1 812	866	415
DTE	64.38%	68.59%	68.28%	65.85%	64.89%	484 711	71 700	9 915	4 351	2 757
EOAN	70.07%	72.50%	72.79%	72.15%	71.43%	39 904	1 567	463	219	168
FME	70.85%	70.82%	68.51%	67.62%	68.06%	119 564	18 461	8 152	4 981	3 460
FRE	69.27%	69.83%	68.80%	67.65%	67.28%	128 239	43 937	19 725	10 732	7 229
HEI	68.29%	70.38%	70.42%	69.05%	67.84%	132 583	31 014	10 982	6 029	4 005
HEN3	69.07%	68.53%	67.16%	65.85%	65.97%	116 053	27 137	10 051	5 159	3 185
IFX	66.41%	67.40%	66.43%	65.20%	64.79%	248 676	72 051	19 118	9 509	6 624
LHA	73.99%	71.35%	68.56%	65.73%	68.62%	54 420	4 793	1 323	499	290
LIN	69.97%	70.91%	70.19%	68.75%	67.36%	24 303	1 860	805	336	239
LXS	68.33%	69.75%	70.31%	70.33%	70.28%	164 776	42 692	13 558	7 199	4 644
MRK	68.19%	69.52%	66.95%	68.11%	67.67%	24 930	3 094	932	508	365
MUV2	72.97%	75.38%	70.33%	65.49%	64.38%	33 414	1 775	583	226	160
RWE	-	-	-	-	-	-	-	-	-	-
SAP	73.95%	74.05%	71.38%	69.79%	67.24%	145 724	17 172	5 822	3 059	1 688
SIE	71.84%	75.52%	71.00%	69.32%	68.68%	258 849	30 501	8 470	4 527	2 880
TKA	72.05%	71.84%	71.47%	67.47%	65.14%	75 764	6 849	1 714	667	350
VOW3	71.70%	72.69%	69.29%	68.80%	60.70%	80 921	5 482	1 361	468	257

Table 19: ADLMLR bid price process direction predictions computed on six months of data for each ticker at Xetra and multiple Ks.

	Accuracy								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	81.83%	81.55%	80.98%	80.34%	79.42%	78.65%	77.54%	74.43%	71.49%
ALV	77.74%	77.66%	74.09%	71.63%	71.86%	70.07%	68.58%	67.97%	67.12%
BAS	86.12%	86.01%	85.60%	84.83%	83.79%	82.61%	81.36%	80.23%	79.09%
BAYN	87.20%	86.77%	86.02%	85.18%	84.33%	83.39%	82.40%	81.43%	80.47%
BEI	88.09%	87.82%	87.48%	87.02%	86.58%	85.81%	85.15%	84.36%	83.57%
BMW	88.20%	87.91%	87.61%	86.88%	86.10%	85.08%	84.07%	82.99%	81.90%
CBK	82.46%	82.09%	81.25%	79.76%	76.96%	69.22%	69.70%	71.32%	72.40%
CON	79.78%	79.83%	79.65%	79.46%	78.94%	78.29%	77.50%	77.31%	76.11%
DAI	84.54%	84.54%	84.26%	83.78%	83.15%	82.49%	81.58%	80.65%	79.70%
DB1	80.67%	80.45%	80.14%	79.88%	79.22%	78.43%	77.44%	76.60%	71.82%
DBK	88.64%	86.37%	86.06%	85.77%	85.21%	84.63%	83.96%	83.25%	82.35%
DPW	86.44%	86.55%	85.84%	85.16%	83.88%	82.58%	80.95%	79.92%	78.90%
DTE	88.19%	88.12%	87.68%	87.11%	86.30%	85.36%	84.42%	83.44%	82.52%
EOAN	85.86%	84.73%	82.66%	80.28%	76.46%	74.41%	72.80%	71.31%	70.04%
FME	84.86%	85.15%	84.97%	84.68%	84.18%	83.68%	83.03%	82.27%	81.53%
FRE	80.52%	80.44%	79.35%	76.65%	68.62%	68.87%	69.15%	70.06%	70.75%
HEI	85.53%	85.42%	85.02%	84.72%	84.13%	83.47%	82.61%	81.74%	80.88%
HEN3	-	-	-	-	-	-	-	-	-
IFX	85.90%	85.89%	85.73%	85.49%	85.18%	84.93%	84.65%	84.22%	83.70%
LHA	85.25%	85.54%	84.96%	84.61%	83.95%	83.15%	82.41%	81.31%	80.35%
LIN	81.43%	81.61%	80.93%	79.30%	76.46%	75.04%	73.38%	71.78%	70.72%
LXS	81.14%	81.54%	81.78%	81.52%	81.25%	80.64%	80.10%	79.46%	78.70%
MRK	82.44%	82.84%	82.33%	81.48%	80.82%	80.34%	79.81%	78.82%	78.54%
MUV2	82.27%	82.32%	81.08%	79.61%	77.70%	75.40%	72.98%	71.23%	69.91%
RWE	85.24%	85.21%	84.81%	84.27%	83.51%	82.64%	81.74%	80.77%	79.61%
SAP	84.82%	84.63%	83.99%	83.16%	82.21%	81.06%	79.85%	78.63%	77.40%
SIE	87.34%	87.39%	86.73%	86.06%	85.11%	83.95%	82.76%	81.88%	80.77%
TKA	86.45%	86.70%	86.27%	85.68%	84.94%	84.03%	82.98%	81.63%	80.43%
VOW3	85.96%	85.03%	83.96%	82.55%	80.74%	79.49%	78.12%	76.73%	75.81%
	Potential Opportunities								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	30 921	35 286	39 968	45 093	50 575	50 272	28 165	10 629	4 297
ALV	283	394	575	980	1 663	2 743	4 032	5 601	7 281
BAS	18 160	23 046	28 044	33 558	39 700	46 766	54 954	64 071	73 405
BAYN	26 195	31 442	36 851	42 762	49 519	57 000	65 163	73 748	82 099
BEI	15 945	18 433	21 030	23 664	26 402	29 328	32 213	35 120	37 811
BMW	25 793	30 810	35 998	41 453	47 254	53 655	60 483	67 707	74 969
CBK	33 676	40 120	41 088	35 718	31 174	10 751	12 381	11 114	12 799
CON	26 077	29 178	32 502	36 021	39 907	43 920	48 190	38 752	23 615
DAI	54 762	63 187	71 882	80 629	90 099	99 954	110 593	121 646	132 795
DB1	19 627	22 824	26 454	30 525	35 270	40 394	45 854	35 404	7 226
DBK	36 681	34 372	42 140	50 664	59 908	70 333	81 715	94 324	107 699
DPW	3 406	4 140	5 000	6 233	7 871	9 861	12 165	14 388	16 578
DTE	15 898	21 367	27 264	33 409	39 635	46 320	53 142	60 140	67 593
EOAN	1 266	1 755	2 364	3 463	5 560	8 698	13 005	17 242	20 964
FME	12 382	14 521	16 831	19 370	22 244	25 238	28 402	31 444	34 313
FRE	28 444	26 874	24 332	16 284	5 360	6 136	6 845	4 108	4 861
HEI	18 116	20 956	23 910	26 966	30 117	33 293	36 637	40 151	43 507
HEN3	-	-	-	-	-	-	-	-	-
IFX	31 499	36 515	41 742	47 484	54 301	62 839	72 185	80 774	83 824
LHA	6 115	7 515	8 892	10 548	12 648	15 196	17 983	20 867	23 714
LIN	1 422	1 767	2 166	2 739	3 705	5 053	6 810	8 821	10 925
LXS	11 547	13 920	16 593	19 606	23 095	26 695	30 538	34 767	39 377
MRK	1 657	2 028	2 490	3 040	3 691	4 414	5 135	5 893	6 539
MUV2	1 145	1 493	1 866	2 428	3 318	4 573	6 239	8 210	10 290
RWE	17 519	21 529	25 494	29 483	34 003	38 907	44 709	51 344	58 359
SAP	13 213	15 657	18 212	21 188	24 919	29 452	34 764	40 652	46 600
SIE	18 183	22 674	27 466	32 574	38 546	45 283	52 742	60 341	67 885
TKA	5 462	7 323	9 218	11 086	12 822	14 743	16 816	19 029	21 456
VOW3	5 593	6 848	8 296	10 197	12 599	15 432	18 646	22 050	25 340

Table 20: ADLMLR bid price process direction predictions computed on six months of data for each ticker at BATS and multiple Ks.

	Accuracy								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	76.02%	76.27%	76.45%	76.05%	75.55%	74.97%	74.13%	73.87%	73.42%
ALV	79.26%	79.19%	79.01%	78.89%	78.44%	77.96%	77.51%	76.68%	75.78%
BAS	79.29%	79.67%	79.58%	79.28%	78.88%	78.11%	77.20%	76.17%	75.11%
BAYN	78.90%	79.05%	78.97%	78.67%	78.24%	77.70%	77.06%	76.15%	75.24%
BEI	80.07%	79.66%	79.11%	78.32%	77.30%	76.08%	74.28%	72.27%	69.60%
BMW	79.18%	79.14%	78.95%	78.87%	78.53%	78.28%	77.88%	77.30%	76.51%
CBK	69.38%	70.46%	71.52%	71.85%	72.06%	71.94%	71.80%	71.44%	71.17%
CON	75.42%	75.39%	75.48%	75.38%	75.19%	74.91%	75.41%	74.41%	74.14%
DAI	74.64%	74.59%	74.55%	74.25%	73.99%	73.58%	73.07%	72.31%	72.17%
DB1	78.16%	77.98%	78.01%	77.76%	77.31%	76.92%	76.53%	75.63%	74.78%
DBK	86.20%	82.93%	83.27%	83.32%	83.26%	83.05%	82.70%	82.32%	81.77%
DPW	77.03%	76.77%	76.52%	75.68%	74.60%	73.29%	71.85%	70.60%	69.09%
DTE	83.86%	83.44%	82.96%	82.33%	81.55%	80.63%	79.69%	78.63%	77.47%
EOAN	84.17%	83.09%	82.33%	81.35%	80.37%	79.42%	79.20%	78.44%	77.66%
FME	76.99%	76.88%	76.83%	76.27%	75.59%	74.32%	72.88%	71.37%	70.67%
FRE	78.26%	78.03%	77.46%	76.90%	73.93%	75.86%	71.98%	71.87%	72.15%
HEI	79.79%	79.26%	78.85%	77.98%	77.30%	76.56%	75.81%	74.77%	73.68%
HEN3	76.69%	76.60%	76.05%	75.30%	74.46%	73.85%	73.13%	72.11%	72.91%
IFX	78.26%	78.23%	78.03%	77.62%	76.98%	76.11%	74.81%	72.17%	69.50%
LHA	86.66%	87.12%	87.05%	86.20%	85.49%	84.76%	84.09%	83.19%	82.35%
LIN	78.68%	78.31%	78.45%	78.27%	77.95%	77.45%	76.89%	76.19%	75.22%
LXS	80.46%	80.86%	80.83%	81.04%	80.98%	80.67%	79.94%	78.94%	78.07%
MRK	75.78%	75.82%	75.24%	74.69%	73.31%	72.02%	70.73%	69.68%	68.25%
MUV2	79.79%	79.21%	79.27%	78.82%	78.13%	77.57%	76.93%	76.09%	74.94%
RWE	-	-	-	-	-	-	-	-	-
SAP	78.35%	78.11%	77.70%	77.18%	76.54%	75.94%	75.04%	73.63%	71.73%
SIE	78.00%	78.10%	77.83%	77.33%	76.66%	75.80%	74.56%	73.28%	72.07%
TKA	84.76%	84.89%	84.07%	83.90%	83.30%	82.59%	81.60%	80.62%	79.39%
VOW3	76.87%	76.15%	75.41%	74.32%	73.35%	72.44%	71.61%	70.89%	69.98%

	Potential Opportunities								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	16 658	19 906	23 755	28 191	30 847	33 653	39 495	23 346	7 419
ALV	4 870	6 344	7 446	8 206	8 709	9 071	9 351	9 677	10 098
BAS	20 357	24 871	30 271	36 608	43 650	51 218	59 218	68 748	79 569
BAYN	25 715	31 102	37 115	43 786	51 053	59 118	68 008	77 702	87 659
BEI	14 178	16 890	19 827	23 281	27 085	26 959	29 558	31 972	31 720
BMW	16 988	20 273	24 103	28 490	33 255	38 726	44 826	51 328	58 495
CBK	6 026	7 875	9 899	12 208	14 584	17 098	19 777	22 585	25 456
CON	18 493	21 154	24 325	27 894	31 800	36 206	7 939	7 339	7 502
DAI	27 757	32 729	38 415	44 663	51 605	59 466	68 222	78 085	62 966
DB1	7 293	8 680	10 281	12 158	14 211	16 655	19 480	22 834	26 717
DBK	30 084	26 839	33 668	41 185	49 402	58 302	68 056	78 582	89 943
DPW	6 952	8 389	9 972	11 647	13 573	15 775	17 121	18 375	20 361
DTE	19 527	24 681	30 420	36 474	42 941	49 656	56 636	63 938	71 807
EOAN	2 091	4 359	7 070	9 811	12 817	14 278	15 315	16 757	18 021
FME	10 772	13 039	15 681	18 746	22 141	26 056	30 596	35 639	33 514
FRE	16 801	19 567	21 713	25 015	18 485	8 513	4 575	2 574	3 163
HEI	10 130	11 770	13 553	15 539	17 796	20 102	22 509	25 152	28 017
HEN3	12 073	13 663	15 491	17 561	19 849	22 378	20 441	21 923	20 310
IFX	24 511	29 214	34 819	41 175	48 318	56 331	58 735	56 609	37 044
LHA	3 906	5 705	7 584	9 524	11 398	13 261	15 265	17 372	19 775
LIN	5 484	7 013	8 264	9 260	10 026	10 618	11 162	11 695	12 327
LXS	4 135	5 163	6 347	7 658	9 341	11 274	13 570	16 139	19 026
MRK	3 030	3 581	4 132	4 626	5 181	5 722	6 369	6 976	7 707
MUV2	3 617	5 042	6 265	7 312	8 176	8 885	9 537	10 194	10 932
RWE	-	-	-	-	-	-	-	-	-
SAP	15 259	19 120	22 756	26 421	30 150	34 157	38 404	43 569	50 664
SIE	16 256	20 633	25 579	31 271	37 625	44 290	51 478	58 937	66 880
TKA	4 009	5 651	7 628	9 474	11 436	13 528	15 714	17 975	20 742
VOW3	9 527	12 781	16 099	19 535	22 849	25 774	28 312	30 527	32 671

Table 21: ADLMLR ask price process direction predictions computed on six months of data for each ticker at Xetra and multiple Ks.

	Accuracy								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	80.93%	80.70%	80.31%	79.77%	79.07%	78.12%	77.15%	76.30%	71.59%
ALV	76.67%	75.88%	75.05%	73.24%	70.24%	68.91%	67.52%	66.33%	65.63%
BAS	86.35%	86.17%	85.55%	84.68%	83.39%	82.18%	80.90%	79.67%	78.53%
BAYN	84.84%	84.50%	83.97%	83.53%	82.73%	81.97%	80.95%	79.85%	78.75%
BEI	86.76%	86.66%	86.26%	85.87%	85.24%	84.56%	83.80%	82.85%	82.44%
BMW	87.94%	87.72%	87.34%	86.69%	85.86%	84.81%	83.78%	82.68%	81.61%
CBK	82.25%	72.07%	72.24%	72.32%	72.67%	72.49%	72.39%	72.33%	72.30%
CON	80.06%	80.13%	80.08%	79.89%	79.51%	78.80%	78.00%	76.70%	73.93%
DAI	84.60%	84.50%	84.28%	83.81%	83.25%	82.49%	81.67%	80.69%	79.84%
DB1	80.54%	80.27%	79.80%	79.52%	78.94%	78.38%	77.57%	76.78%	71.94%
DBK	86.07%	86.07%	85.72%	85.35%	84.76%	84.16%	83.43%	82.62%	81.84%
DPW	85.17%	85.04%	84.45%	83.83%	82.28%	80.88%	79.68%	78.67%	77.89%
DTE	89.66%	89.47%	89.04%	88.35%	87.56%	86.40%	85.24%	84.20%	83.18%
EOAN	85.81%	85.34%	83.16%	80.45%	76.95%	74.60%	72.32%	70.48%	68.98%
FME	83.81%	84.17%	84.15%	84.19%	83.77%	83.18%	82.77%	82.27%	81.69%
FRE	77.17%	77.40%	76.43%	75.08%	71.96%	71.22%	68.70%	69.45%	69.93%
HEI	85.62%	85.64%	85.40%	85.10%	84.55%	83.88%	83.02%	82.21%	81.23%
HEN3	-	-	-	-	-	-	-	-	-
IFX	86.37%	86.15%	85.93%	85.59%	85.21%	84.93%	84.64%	84.27%	83.75%
LHA	86.46%	86.32%	85.82%	85.38%	84.59%	83.56%	82.65%	81.79%	80.80%
LIN	81.37%	80.72%	79.17%	77.35%	75.54%	73.91%	72.37%	71.00%	69.78%
LXS	80.99%	81.36%	81.38%	81.35%	81.13%	80.64%	80.16%	79.63%	78.83%
MRK	78.51%	78.30%	78.86%	79.60%	78.80%	78.39%	77.71%	77.50%	76.88%
MUV2	82.58%	82.21%	81.91%	80.51%	79.26%	76.54%	74.53%	72.91%	70.84%
RWE	83.42%	83.44%	83.38%	82.92%	82.06%	81.22%	80.30%	79.30%	78.37%
SAP	85.16%	84.90%	84.10%	83.09%	81.79%	80.69%	79.47%	78.25%	77.23%
SIE	86.55%	86.82%	86.51%	85.89%	84.89%	83.93%	82.95%	82.04%	81.08%
TKA	85.78%	85.86%	85.55%	84.95%	84.12%	83.15%	82.12%	81.07%	79.77%
VOW3	85.55%	85.74%	85.15%	83.91%	82.54%	80.97%	79.43%	77.85%	76.76%

	Potential Opportunities								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	28 608	32 698	37 266	42 006	47 189	52 778	58 452	29 678	7 789
ALV	643	937	1 511	2 590	4 295	6 461	9 008	11 716	14 293
BAS	20 883	25 677	30 490	35 730	42 148	49 724	58 185	67 347	76 681
BAYN	21 160	25 974	31 020	36 515	42 747	49 702	57 503	66 176	75 580
BEI	17 641	20 042	22 644	25 499	28 429	31 471	34 480	37 391	35 580
BMW	24 655	29 771	34 987	40 413	46 077	52 271	58 846	65 856	73 228
CBK	13 372	4 758	5 655	6 614	7 666	8 691	9 886	11 214	12 760
CON	25 429	28 756	32 516	36 670	41 169	45 994	40 291	21 207	3 126
DAI	55 247	64 215	73 575	83 238	93 416	104 200	116 037	128 391	126 253
DB1	19 476	22 574	26 268	30 435	35 137	40 248	45 788	32 260	9 796
DBK	27 432	34 604	42 219	50 449	59 533	69 492	80 640	92 856	105 613
DPW	2 934	3 602	4 314	5 311	6 789	8 567	10 623	12 804	14 900
DTE	19 116	24 869	30 456	36 400	42 540	49 112	56 183	63 660	71 108
EOAN	1 283	1 848	2 560	3 549	5 301	7 866	10 960	14 473	18 096
FME	11 992	14 147	16 433	19 057	22 004	25 328	28 525	31 738	34 802
FRE	20 999	22 862	21 715	13 417	5 556	6 314	2 361	2 959	3 495
HEI	18 568	21 667	24 912	28 208	31 675	35 262	38 996	42 720	46 319
HEN3	-	-	-	-	-	-	-	-	-
IFX	31 910	36 935	41 767	47 211	53 824	62 397	71 723	80 246	86 962
LHA	6 181	7 661	9 079	10 863	13 014	15 616	18 424	21 423	24 285
LIN	1 369	1 675	2 108	2 804	3 827	5 208	6 891	8 716	10 637
LXS	10 966	13 216	15 819	18 753	22 050	25 553	29 378	33 623	38 364
MRK	2 634	3 102	3 141	3 549	4 245	5 020	5 765	6 468	7 097
MUV2	1 039	1 321	1 658	2 109	2 714	3 636	4 978	6 625	8 563
RWE	16 029	19 631	23 356	27 363	31 711	36 679	42 362	48 561	55 273
SAP	13 183	15 591	18 104	21 135	25 025	29 603	34 963	40 685	46 425
SIE	18 078	22 682	27 290	32 468	38 319	44 929	52 046	59 466	66 941
TKA	6 469	8 270	10 084	11 851	13 679	15 658	17 771	20 199	22 816
VOW3	3 904	5 013	6 229	7 706	9 501	11 683	14 275	17 308	20 482

Table 22: ADLMLR ask price process direction predictions computed on six months of data for each ticker at BATS and multiple Ks.

	Accuracy								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	75.51%	75.59%	75.38%	74.97%	74.27%	73.58%	73.45%	72.85%	71.07%
ALV	78.12%	78.48%	78.11%	77.81%	77.17%	76.85%	76.43%	75.85%	75.08%
BAS	79.84%	79.86%	79.48%	79.05%	78.30%	77.42%	76.45%	75.19%	73.98%
BAYN	78.91%	78.75%	78.68%	78.39%	77.97%	77.41%	76.68%	75.93%	74.95%
BEI	79.32%	78.99%	78.54%	78.01%	77.30%	76.62%	75.45%	74.03%	71.76%
BMW	79.23%	79.09%	79.13%	78.74%	78.33%	77.63%	76.81%	76.46%	76.81%
CBK	69.27%	70.24%	71.29%	71.16%	71.54%	71.59%	71.41%	71.29%	71.20%
CON	75.31%	75.53%	75.52%	75.57%	75.74%	76.59%	76.68%	76.78%	76.39%
DAI	75.23%	75.21%	75.19%	75.07%	74.86%	74.58%	74.09%	73.30%	72.83%
DB1	79.00%	79.07%	78.53%	77.85%	77.13%	76.37%	75.56%	74.41%	73.18%
DBK	82.45%	82.52%	82.46%	82.49%	83.45%	83.09%	82.75%	82.38%	81.77%
DPW	77.55%	77.27%	77.03%	76.09%	74.84%	73.41%	71.81%	70.26%	68.58%
DTE	82.42%	82.36%	81.96%	81.45%	80.76%	79.87%	78.70%	80.18%	78.93%
EOAN	80.89%	81.06%	80.58%	80.14%	79.22%	78.24%	77.22%	76.10%	74.72%
FME	75.82%	76.08%	76.32%	76.20%	75.62%	74.94%	73.79%	72.26%	70.94%
FRE	77.24%	77.20%	77.10%	76.59%	74.86%	74.45%	74.74%	73.92%	74.01%
HEI	79.33%	78.87%	78.36%	77.77%	77.10%	76.30%	75.31%	74.32%	73.28%
HEN3	77.17%	76.66%	76.22%	75.64%	74.49%	73.61%	72.36%	71.62%	70.63%
IFX	79.22%	79.05%	78.64%	78.14%	77.47%	76.29%	74.75%	72.95%	70.28%
LHA	87.42%	87.73%	87.34%	86.88%	86.27%	85.70%	84.63%	83.68%	82.67%
LIN	77.53%	78.15%	78.42%	78.21%	77.81%	77.44%	77.02%	76.44%	75.60%
LXS	80.49%	80.25%	79.63%	79.01%	78.23%	77.23%	75.72%	74.31%	73.23%
MRK	76.76%	76.38%	75.43%	74.78%	74.47%	73.47%	72.14%	70.55%	68.90%
MUV2	80.26%	81.38%	80.53%	79.96%	79.16%	78.77%	78.40%	77.76%	77.13%
RWE	-	-	-	-	-	-	-	-	-
SAP	78.82%	78.60%	78.19%	77.65%	76.93%	76.12%	75.26%	74.03%	72.65%
SIE	77.48%	77.37%	77.17%	76.66%	75.93%	74.91%	73.81%	72.59%	71.31%
TKA	84.70%	84.83%	84.10%	83.52%	82.76%	81.86%	80.92%	79.67%	78.62%
VOW3	76.27%	75.59%	75.40%	74.62%	73.65%	72.76%	71.79%	70.94%	69.84%
	Potential Opportunities								
Ticker \K	**Peak**	**Peak - 0.025**	**Peak - 0.050**	**Peak - 0.075**	**Peak - 0.100**	**Peak - 0.125**	**Peak - 0.150**	**Peak - 0.175**	**Peak - 0.200**
ADS	20 460	24 126	28 180	32 813	38 273	44 420	24 071	10 687	7 498
ALV	3 747	5 078	6 145	6 867	7 452	7 811	8 105	8 411	8 787
BAS	25 399	30 853	37 300	44 306	51 788	59 716	68 571	78 824	90 271
BAYN	26 801	32 133	38 087	44 553	51 822	59 762	68 834	78 496	88 362
BEI	12 084	14 542	17 175	20 086	23 480	27 283	31 779	33 811	33 689
BMW	21 650	25 444	29 913	34 879	40 243	46 361	52 962	55 564	50 812
CBK	4 992	5 985	7 238	8 611	10 063	11 543	13 101	14 694	16 337
CON	15 444	17 739	20 425	23 540	26 320	23 886	20 533	9 466	9 026
DAI	32 913	38 482	45 002	52 334	60 786	70 333	81 285	93 833	74 003
DB1	10 432	12 211	14 248	16 478	19 200	22 245	25 835	29 944	32 847
DBK	21 476	28 297	35 242	42 448	37 368	44 493	51 933	60 041	68 775
DPW	6 143	7 524	8 941	10 453	12 158	14 026	16 268	17 189	16 961
DTE	14 109	18 768	24 218	30 079	36 409	43 447	52 738	37 771	43 339
EOAN	2 993	5 818	8 898	11 518	13 743	15 553	17 149	18 670	20 339
FME	9 976	12 049	14 631	17 657	21 001	24 838	29 117	33 994	36 397
FRE	17 959	20 639	23 794	27 208	24 996	5 228	6 310	3 903	4 625
HEI	9 777	11 465	13 325	15 292	17 544	19 940	22 604	25 513	28 574
HEN3	13 097	14 941	17 055	19 271	18 646	21 219	21 552	22 979	13 939
IFX	25 927	30 711	36 085	42 126	48 862	54 363	57 507	59 465	39 525
LHA	4 736	6 846	9 041	11 155	13 133	15 062	17 074	19 219	21 531
LIN	4 855	6 385	7 677	8 748	9 590	10 188	10 670	11 131	11 693
LXS	8 759	10 471	12 631	15 159	18 011	21 241	25 009	29 290	21 101
MRK	3 158	3 772	4 388	4 905	5 328	5 817	6 357	6 991	7 726
MUV2	2 001	2 917	3 858	4 655	5 235	5 663	5 971	6 209	6 490
RWE	-	-	-	-	-	-	-	-	-
SAP	19 983	24 210	28 468	32 327	35 884	39 382	42 942	46 889	51 624
SIE	17 497	21 767	26 841	32 650	39 001	46 036	53 309	60 974	69 181
TKA	5 608	7 616	9 624	11 699	13 721	15 863	18 062	20 423	22 864
VOW3	7 942	10 377	13 140	16 190	19 353	22 387	25 298	27 822	30 296