January 24, 2018

The impact of arbitrage on market liquidity

Dominik Rösch

State University of New York, Buffalo

I thank Lamont Black, Ekkehart Boehmer, Dion Bongaerts, Howard Chan, Tarun Chordia, Ruben Cox, Thierry Foucault, Louis Gagnon, Nicolae Gârleanu, Michael Goldstein, Amit Goyal, Allaudeen Hameed, Terry Hendershott, Shing-yang Hu, Jonathan Kalodimos, Andrew Karolyi, Albert Kyle, Su Li, Albert Menkveld, Pamela Moulton, Maureen O'Hara, Louis Piccotti, Gideon Saar, Piet Sercu, René Stulz, Avanidhar Subrahmanyam, Raman Uppal, Dimitrios Vagias, Mathijs van Dijk, Manuel Vasconcelos, Kumar Venkataraman, Axel Vischer, Avi Wohl, and participants at the 2016 Research in Behavioral Finance Conference (Amsterdam), 2014 Financial Management Conference and Doctoral Consortium (Nashville), the 2014 Northern Finance Conference (Ottawa), the 2014 Asian Finance Conference (Bali), 2014 Eastern Finance Conference (Pittsburgh), the 2013 Erasmus Liquidity Conference (Rotterdam), the 2013 Conference on the Theories and Practices of Securities and Financial Markets (Kaohsiung), the 2013 World Finance and Banking Symposium (Beijing), and at seminars at Babson College, Cass Business School, City University of Hong Kong, Cornell University, Erasmus University, Frankfurt School Of Management, and the University at Buffalo for valuable comments. I am grateful for the hospitality of the Department of Finance at the Johnson Graduate School of Management (Cornell University), where some of the work on this paper was carried out, especially from my hosts, Andrew Karolyi and Pamela Moulton. I also gratefully acknowledge financial support from the Vereniging Trustfonds Erasmus Universiteit Rotterdam and from the Netherlands Organisation for Scientific Research through a "Vidi" grant. This work was carried out on the National e-infrastructure with the support of SURF Foundation. This work is supported in part by NSF ACI-1541215. I thank OneMarket Data for the use of their OneTick software.

Abstract

The impact of arbitrage on market liquidity

I study deviations from the law of one price in Depositary Receipts using tick-by-tick data from the United States and 22 different home markets from 2001 to 2016. Deviations persist, on average, 12 minutes, and mainly arise because of demand pressure. Exploiting institutional details that create exogenous variation in the impediments to arbitrage within and across days, I show that absolute price deviations predict illiquidity, contemporaneously and in the future. Price deviations mainly predict the inventory costs component of the bid-ask spread. Thus, consistent with recent theory, these findings suggest that arbitrageurs tend to trade against demand pressure and thus enhance market integration and liquidity. (*JEL* G14, G15)

Arbitrage enforces the law of one price and thereby improves the informational efficiency of the market. But how arbitrage affects other measures of market quality, in particular market liquidity, is less well understood.

Foucault, Kozhan, and Tham (2017) model arbitrage opportunities to arise exogenously either due to demand or information shocks. If price deviations arise due to information, liquidity providers face higher adverse selection risk. In this case arbitrage can be “toxic”. Traders picking up stale quotes in the Nasdaq Small Order Execution System (SOES)—the so-called “SOES bandits” (Harris and Schultz, 1998)—are example of such toxic arbitrage. Foucault et al. (2017) provide theoretical and empirical evidence that liquidity is lower if toxic arbitrage is likely.

But if arbitrage opportunities arise as a result of demand pressure—such as fire sales by mutual funds or country-specific sentiment (Chan, Hameed, and Lau, 2003)—arbitrageurs trade against market demand and thereby decrease inventory holding costs for liquidity providers. This observation is made succinctly in a survey of the theoretical limits-of-arbitrage literature by Gromb and Vayanos (2010), who state that “arbitrageurs provide liquidity” (p. 258).

In other words, whether arbitrage improves or worsens liquidity depends on the reason why arbitrage opportunities arise. If opportunities arise as a result of demand shocks, arbitrage improves liquidity (Holden, 1995; Gromb and Vayanos, 2010), but if as a result of differences in information, arbitrage worsens liquidity (Domowitz et al., 1998; Foucault et al., 2017; Kumar and Seppi, 1994).

Inspired by this observation, I investigate why price deviations arise and estimate the impact that impediments to arbitrage have on market liquidity. As price deviations arise sometimes as a result of demand shocks and other times as a result of information differences, so arbitrage sometimes improve and other times worsen liquidity. In the extreme case, if both effects are equally strong and cancel each other out, arbitrage will not have a visible effect on liquidity. This is my null hypothesis. Alternatively, if the effect of increased adverse selection dominates, arbitrage worsens liquidity, and if the effect of lower inventory holding costs dominates, arbitrage improves liquidity.

To date the empirical literature finds evidence mainly of illiquidity as an impediment to arbitrage; meanwhile, the evidence of how impediments to arbitrage affect liquidity is scarce.[1] This is not surprising: the natural reverse causality between impediments to arbitrage and liquidity—higher liquidity decreases impediments to arbitrage, and lower impediments to arbitrage encourage arbitrageurs to trade, affecting liquidity—is challenging to address.

I address the challenge by exploiting institutional details of the American Depositary Receipt (ADR) market that create exogenous variation in the impediments to arbitrage within and across days. As Gagnon and Karolyi (2010) lay out, the ADR market is particularly suitable to studying arbitrage because the ADR and the home market share offer identical cash-flows. In addition, in the ADR market arbitrage is almost risk-free because institutions exist that make it possible to convert the ADR into the home market share, and vice versa.

In other markets the arbitrage risk might be even lower. For example, arbitrage between the same stock trading at different exchanges in the U.S. or in Europe would be even less risky. But smart order routing—enforced by Regulation NMS in the U.S. and MiFID in Europe—already ensures that price deviations can not persist. Because no such regulation exists across ADRs and their home market share, these markets are partially segmented and price deviations can persist creating potentially lasting arbitrage opportunities.[2]

Impediments to arbitrage vary exogenously within each day for many ADRs because the ADR and the home market share are trading simultaneously (allowing arbitrage) for only a few hours each day. For many ADRs impediments to arbitrage also vary exogenously across days because corporate actions, such as dividend payments or stock splits, on ADRs and their home market share often do not occur on the same day. During these days when one stock is ex- and the other is cum-dividend, institutions that normally facilitate arbitrage shut down, and it is no longer possible to convert the ADR into a home market share or vice versa (cf. Citibank (2007); Galligan et al. (2014), p. 51).[3] For this and for other reasons that I explain later, on these days impediments

[1] Foucault et al. (2017) only cite Kumar and Seppi (1994), Roll, Schwartz, and Subrahmanyam (2007), and the current paper when discussing papers studying the effect of arbitrage on liquidity (see p. 1058 and footnote 11).

[2] Bodurtha et al. (1995) and Chan et al. (2003) provide evidence that international markets are partially segmented.

[3] I use the terms ex-dividend and cum-dividend to refer to any corporate action which has a direct

to arbitrage are greater.

In my main analysis I examine tick-by-tick bid and ask quotes and trade prices for 194 ADRs and currency-adjusted prices for the associated home market share over the period from February 2001—after decimalization in the U.S.—to December 2016.[4]

With these data I construct two intraday price deviation measures. The first measure is the tick-by-tick difference between the highest bid and the lowest ask price across the ADR and the currency adjusted home market share, denoted price deviations from quotes. If the difference is negative I set price deviations to zero. The second measure is the absolute difference in trade prices of simultaneous trades on both the ADR and the home market share.

From these two intraday price deviations, I construct several daily proxies for the impediments to arbitrage. I assume that the market is reasonably efficient so that "prices reflect information to the point where the marginal benefits of acting ... do not exceed the marginal costs" [Fama (1991), p. 1575]. In other words, the price deviations I observe reflect underlying frictions that impede arbitrage, such as short-selling restrictions, risk, or capital constraints. Such an interpretation is common in the literature, and several studies provide empirical justification for it. For example, Gagnon and Karolyi (2010) show that price deviations in the ADR market positively correlate with holding costs and Hu et al. (2013) interpret large price deviations as "a symptom of a market in severe shortage of arbitrage capital" (p. 2342).

To proxy for the impediments to arbitrage I use the daily maximum and average price deviations from quote prices. I also use the average duration price deviations persist and the average absolute price deviations from simultaneous trade prices.[5] It is common to assume that price deviations follow a mean-reverting process and to interpret high mean-reversion as a sign of high arbitrage activity (as, for example, in Roll et al. (2007)). As elaborated in the main text, above proxies are, in general, negatively related to the speed of mean reversion, justifying the interpretation as inverse

effect on the stock price, not only dividend payments.

[4] The earliest data available in the database is from 1996, but the valid observations in the early years are sparse, so I drop all data before February 2001.

[5] On days when one stock is cum-dividend but the other is not, I adjust prices by the corporate action. As a robustness check, I also consider unadjusted prices.

proxies of arbitrage activity.

Using the intraday data, I first investigate why price deviations arise and whether they do so as a result of nonfundamental demand shocks or of differences in information. Inspired by Schultz and Shive (2010), I identify nonfundamental demand shocks as situations in which price deviations arise as a result of transitory price movements—that is, when one share moves to create the price deviation and later moves back to eliminate it. This identification is based on the understanding that demand shocks are associated with price reversals while new information is associated with a permanent price effect (e.g., Gagnon and Karolyi (2009)). Following Foucault et al. (2017), I consider a price deviation as toxic for one market if the share price of this market moved to eliminate a price deviation that was created with a price movement in the other market. My analysis reveals that in the ADR market most arbitrage opportunities arise due to a nonfundamental demand shock and less than 10% are toxic (compared to around 50% in the foreign exchange market studied by Foucault et al. (2017)).

Next, I use daily measures—as in Foucault et al. (2017) and Roll et al. (2007)—to investigate the impact impediments to arbitrage have on market liquidity. I find that price deviations persist, on average, 12 minutes. Average price deviations computed from quote prices are around 0.80% (as a percentage of the home market share price), similar to the cost-adjusted, absolute end-of-day price deviations reported by Gagnon and Karolyi (2010) of 1.12%.

Inspired by Foucault et al. (2017) I investigate how the arbitrage mix (the percentage of toxic price deviations), the speed of arbitrageurs (the percentage of toxic price deviations ending with a trade), and price deviations affect illiquidity. I extend their analysis by also investigating how nontoxic price deviations are related to illiquidity.

In contrast to Foucault et al. (2017), I find that the fraction of price deviations ending with a trade is negatively related to illiquidity, regardless of whether the price deviation is toxic. This relation potentially occurs because price deviations last much longer and because toxic arbitrage is less common in the ADR market than in the foreign exchange market that Foucault et al. (2017) focus on.

One concern the previous analysis presents is that variables are endogenously determined. The main identification in Foucault et al. (2017) comes from the insight that liquidity providers face higher adverse selection risk only if arbitrageurs can trade faster than local liquidity providers can update their quotes. They address endogeneity concerns using an exogenous shock to the relative speed of arbitrageurs. However, speed seems of secondary importance in the ADR market. Therefore, I focus on exogenous variation in arbitrage activity within (during and outside overlapping trading times) and across days (days between corporate actions).

Using a panel regression, I first investigate liquidity during days when the ADR or the home market share is ex-dividend but the other is still cum-dividend. I find that on these days quoted spreads and effective spreads are around 2 to 3 basis points higher than on other days. These differences are statistically and economically significant. Of course, corporate actions alone might affect liquidity provisions, but they do not seem to explain the above findings. First, on days when both the ADR and the home market go ex-dividend together, spreads are about the same as on other days. Second, on days between corporate actions, quoted spreads are especially high during times when both the ADR and the home market stock are trading compared to when only one is trading. If the corporate action affected liquidity, one would expect the corporate action to affect liquidity throughout the day, not just when the other stock is trading. Therefore, the increase in illiquidity during these days is best explained by cross-market effects and arbitrage activity.

To estimate the effect that impediments to arbitrage have on market liquidity, I start with an instrumental variable panel regression. In the first-stage, I regress price deviations on a dummy set to one if on the specific stock-day the ADR or the home market stock is ex-dividend and the other is cum-dividend. All price-deviation measures are elevated during these days whether or not I adjust prices for the corporate action.

In the second-stage, I regress market liquidity on the fitted value of the first-stage regression and all control variables. In all cases I find that higher price deviations are associated with higher illiquidity, measured by quoted spreads (in particular, the inventory holding costs component of the bid-ask spread), effective spreads, and the difference in quoted spreads during and outside overlapping trading times. These results are always statistically significant at least at the 10% level.

There are several reasons to believe that impediments to arbitrage affect market liquidity not only contemporaneously, but also in the future. First, arbitrageurs might trade against market demand and thereby decrease overnight inventories, thereby improving future liquidity (Comerton-Forde et al., 2010; O'Hara and Oldfield, 1986). Second, large price deviations decrease how informative prices can be and thus lower future liquidity (Cespa and Foucault, 2014). Therefore, I estimate impulse response functions from a panel vector autoregression model (with time and stock fixed effects) using as endogenous variables price deviations, the volatility computed from 5-minute returns, absolute order imbalances, and illiquidity

Price deviations Granger-cause illiquidity, and impulse-response functions indicate that a one-standard-deviation shock to price deviations predicts a contemporaneous increase in illiquidity of three to ten basis points after 15 days.

All in all, I find that higher impediments to arbitrage are associated with higher market illiquidity and predict future illiquidity.

These results are consistent with theory and the findings in the first part of my paper. If price deviations arise as a result of demand shocks (and the first part of my paper indicates that most price deviations do), arbitrage improves liquidity. These results are also consistent with the idea that less informative prices lead to lower liquidity provision (Cespa and Foucault, 2014).

These findings shed light on how policy changes that increase impediments to arbitrage (such as short-sell bans or transaction taxes) could negatively impact the liquidity of financial markets and ultimately increase firms' cost of capital (Amihud and Mendelson, 1986).

Broadly, my paper is related to the literature that investigates how changes to the trading environment affect market quality (e.g. Brogaard et al. (2014); Chaboud et al. (2014a); Chordia et al. (2005, 2008); Hendershott et al. (2011); Menkveld (2013)). More specifically, my paper relates to the empirical limits-of-arbitrage literature (among many significant contributions: Mitchell et al. (2002); Lamont and Thaler (2003); De Jong et al. (2009); Gagnon and Karolyi (2010)). But instead of investigating why price deviations persist and how illiquidity impacts price deviations, I focus on (i) why price deviations arise (Foucault et al., 2017; Schultz and Shive, 2010) and (ii) how price deviations impact liquidity (Ben-David et al., 2014; Choi et al., 2009; Foucault et al., 2017; Lou and Polk, 2013; Roll et al., 2007; Tomio, 2017).

I add to these important contributions in two ways. First, in contrast to prior research (e.g., Foucault et al. (2017) or Roll et al. (2007)) I find evidence that arbitrageurs improve market liquidity. Foucault et al. (2017) test whether toxic arbitrage deteriorates liquidity and how arbitrageurs' trading speed affects the duration price deviations persist. They do not test for the overall effect of arbitrage activity on market liquidity. Roll et al. (2007) show that an increase in the absolute futures-cash basis predicts future illiquidity. In particular, they construct the futures-cash basis from end-of-day prices and suggest a large basis today might attract arbitrageurs the next day, causing abnormally large order imbalances, which lower liquidity. But Roll et al. (2007) do not consider the possibility that price deviations might arise because of order imbalances, and therefore that arbitrageurs might provide liquidity by trading against market demand.

Second, I build upon previous work in the ADR literature, especially Gagnon and Karolyi (2010), who study price deviations in the ADR market, and Moulton and Wei (2009) and Werner and Kleidon (1996), who investigate differences in liquidity during and outside overlapping trading times. To this work I add the following contributions. In contrast to most previous studies I have access to tick-by-tick data for the home market, which allows me to study the impact of price deviations on the difference in liquidity during and outside overlapping trading times. To the best of my knowledge, this paper is the first such application of these data. I provide empirical evidence that a decrease in the impediments to arbitrage decreases the gap between liquidity during and outside overlapping trading times. These results can also provide an explanation for time-variation in liquidity differences during and outside overlapping trading times. Where Werner and Kleidon (1996) find that quoted spreads of ADRs in 1991 are higher during overlapping trading times than they are outside these times, Moulton and Wei (2009), using data from 2003, find the opposite. The decrease in the impediments to arbitrage provides one explanation for the difference in these findings.

This paper is organized as follows. In section 1 I discuss data and variable construction and provide summary statistics. Section 2 investigates why arbitrage opportunities arise. Section 3 estimates how impediments to arbitrage affect contemporaneous liquidity using instrumental variable regressions, and Section 4 investigates dynamic relations using impulse response functions. Section 5 concludes.

1. Data and variable construction

1.1. Data and sample

To investigate the impact of arbitrage on market liquidity, I focus on the American Depositary Receipts market (ADR). I refer to, among others, Baruch et al. (2007), Gagnon and Karolyi (2009, 2010, 2013), and Karolyi (1998) for a detailed explanation and a comprehensive introduction to the ADR market. An ADR represents a tradeable certificate backed by the home market share. The ADR market has many endemic features. For example, the feature of convertibility—both ADR and home market share can be converted to each other—allows the interpretation of price deviations between bid and ask prices at the time an arbitrageur opens the arbitrage position as (almost) risk-free profits.

If the currency-adjusted bid price of the home market share is higher than the ask price of the ADR in the host-market (similarly, if the bid price of the host-market ADR is higher than the ask price of the home market share), an arbitrage opportunity exists to simultaneously short sell the home market share at the bid price, convert the proceeds from the short-sale into USD, and buy the ADR in the host-market at the ask price.[6] Afterward, the ADR can be converted (within one business day and for less than five cents a share (Gagnon and Karolyi, 2010)) into the home market share either through a broker (e.g. Interactive Brokers), a crossing platform (e.g. ADR Max, or ADR Navigator), or the actual depositary bank.[7] After the conversion, the home market share can be delivered to close down the short position, resulting in a risk-free USD profit equal to the difference between the bid of the home market and the ask of the host-market ADR *at the time the arbitrage position was opened*.

Most importantly, I focus on the ADR market because it provides an ideal setting to address endogeneity concerns: Arbitrage activity in the ADR market varies exogenously within the day (whether both or only one of the ADR and the home market stock is

[6] This example is for illustrative purposes only. In real markets short-selling is capital intensive, and an initial margin requirement of the initial value of the share plus 50% is required (in the US, Regulation T), which then also creates exchange rate risk.

[7] This conversion is not available during days when one stock is cum-dividend and the other is ex-dividend, which increases the impediments to arbitrage (cf. Citibank (2007) or Galligan et al. (2014)).

trading) and across days (whether both or only one is ex-dividend).

To construct my sample of ADRs and their respective home market shares, I use standard sources in the DR literature: Datastream, Bank of New York Complete Depositary Receipt Directory (www.adrbnymellon.com), and Deutsche Bank Depositary Receipts Services (adr.db.com). Details about the sample construction can be found in Appendix A.

Initially, I identify 325 ADR/home market pairs for which the ADR is trading at the NYSE or Nasdaq. But because the analysis requires comparing contemporaneous prices across ADRs and home market shares, I drop all countries without an overlap in trading times with the U.S. For all matched pairs, I obtain tick-by-tick data on quotes and trades (time-stamped with at least millisecond precision) as well as their respective sizes from the Thomson Reuters Tick History (TRTH) database from January 1996 (the earliest date available in TRTH) through December 2016. Similarly, I obtain tick-by-tick quotes for all currency pairs required to convert local prices into USD, the currency in which the ADR is quoted in, from TRTH.[8]

Quote and trade data is filtered as described in Appendix B. After filtering, the data contains almost 6 billion updates to the best bid and ask quotes and approximately 4 billion trades, with nearly half on ADRs. I ignore stock-days on which prices of the ADR and the home market share could not be aligned, as described in Appendix B. Because of data availability and because of the above filters I have only sparse observations in the early years of the sample. Therefore, I drop all data before February 2001 (the month the U.S. adopted decimalization). My final sample consists of 194 pairs across 22 exchanges.

[8] The TRTH database is managed by the Securities Industry Research Center of Asia-Pacific (SIRCA) and is used in several recent studies (e.g. Fong et al. (2017); Kahraman and Tookes (2017); Lau et al. (2012); Lai et al. (2014); Marshall et al. (2011)). TRTH is a record of Thomson Reuter's worldwide real-time Integrated Data Network (IDN), and prices are time-stamped as observed by traders relying on feeds from IDN, mitigating concerns that price deviations might reflect inconsistent time stamps across exchanges.

1.2. Measures of price deviations

I construct two price-deviation measures. The first is based on trade prices and the second on quote prices.

From trade prices I calculate the absolute difference between the trade prices of the ADR and the home market stock relative to the midquote price of the home market. In other words, ΔTRD is calculated as:

$$\Delta TRD_{i,t} = \left| \frac{trade.home_{i,t} - trade.adr_{i,t_1}}{mid.home_{i,t}} \right| \tag{1}$$

where $trade.home_{i,t}$ is the currency adjusted trade price for trade t of the home market stock, and $trade.adr_{i,t_1}$ is the bundling adjusted trade price for trade t_1 of the ADR, such that t_1 minimizes the distance to t and both trades occur within one second—that is, $|t - t_1| < 1 seconds$.

From quote prices I calculate the difference between the highest bid and the lowest ask price across the home- and host-market relative to the midquote price of the home market, denoted ΔQTE. If this difference is not positive, I set ΔQTE to zero. ΔQTE is calculated as:

$$\Delta QTE_{i,t} = \max \left(\frac{bid.home_{i,t} - ask.adr_{i,t}}{mid.home_{i,t}}, \frac{bid.adr_{i,t} - ask.home_{i,t}}{mid.home_{i,t}}, 0 \right) \tag{2}$$

where $mid.home_{i,t}$ is the mid-quote price of stock i at time t, and $bid.home_{i,t}$ ($ask.home_{i,t}$) is the bid (ask) of stock i at time t converted to USD using the prevailing bid (ask) of the respective currency pair (for example, DEM and EUR for Germany before and after January 1, 1999, respectively). Further, $bid.adr_{i,t}$ ($ask.adr_{i,t}$) is the bid (ask) at time t of the ADR trading in the U.S. associated with stock i, adjusted for the respective bundling ratio as described in Appendix B. To avoid using stale quotes, I only consider quotes that are at maximum 300 seconds old.[9]

[9] This measure accounts for transaction costs due to bid and ask spreads. As a robustness test, I only consider price deviations above one basis point or above one dollar cent to cover additional transaction costs. In both cases the results are robust. See the Online Appendix Tables A9 and A10.

1.3. Daily price deviations, market illiquidity, and order imbalance

From the two intraday price-deviation measures introduced in the previous section, I construct the following stock-day measures: I calculate the average duration price deviations persist; the average price deviation from simultaneous trade prices (from Eq. 1), denoted $avg(\Delta TRD)$; and the average and maximum price deviation from quote prices (from Eq. 2), denoted $avg(\Delta QTE)$ and $max(\Delta QTE)$, respectively.

As the main illiquidity measures I use the proportional quoted spread ($PQSPR$) and effective spread ($PESPR$). $PQSPR$ is defined as the daily time-weighted average of the difference in the ask and the bid price, scaled by the midquote price. $PESPR$ is defined as the daily average of the absolute difference between the logarithm of the trade price and the logarithm of the midquote price of the prevailing quote. Both measures have been widely used as illiquidity measures (e.g., Roll et al. (2007); Boehmer and Kelley (2009); Moulton and Wei (2009); Schultz and Shive (2010)).[10] While other measures of illiquidity are available (e.g., Amihud (2002) or Pastor and Stambaugh (2003)) these are often too noisy to be used at the stock-day level. To ensure that results are not driven by outliers, I cross-sectionally winsorize price deviation and illiquidity measures at the 99% level on each day.

I further construct a measure of buying or selling pressure. First, I sign every trade in both the home market and the ADR using the Lee and Ready (1991) algorithm.[11] Second, I calculate order imbalance for each stock-day as the absolute difference between the number of buyer- and seller-initiated trades (OIB), as in, Chordia et al. (2008).

1.4. Summary statistics

Table 1 presents cross-sectional summary statistics of time-series averages. In Panel A I compute price deviations across all days for which both the ADR and the home market stock are either cum- or ex-dividend. Price deviations across simultaneous trades

[10] In later tests I also decompose the quoted spread into its adverse selection and its transitory component following Glosten and Harris (1988).

[11] A trade is classified as buyer- (seller-)initiated if it is closer to the ask (bid) of the prevailing quote. A trade at the midpoint of the quote is classified as buyer- (seller-)initiated if the previous price change is positive (negative). Ellis et al. (2000), Lee and Radhakrishna (2000), Odders-White (2000), and Theissen (2001) provide evidence that this algorithm signs around 80% of all trades correctly for Nasdaq, NYSE, and German stocks.

are around 2.74% for the average stock, the time-weighted average price deviations from quotes is around 0.80%, and the average daily maximum price deviation within each day is around 2.12%. For comparison, Gagnon and Karolyi (2010) use end-of-day data from 1993 to 2004 and find (cost-adjusted) average price deviations of 1.12%.

Panel A also reports the average duration of price deviations. After excluding the 93,963 stock-days in which the day started with a price deviation that did not revert over the course of the day, the average duration is 12.41 minutes. In the ADR market, price deviations are thus more persistent than in the Forex market (Chaboud et al., 2014b; Foucault et al., 2017).

In Panel B of Table 1, I compute price deviations on days between corporate actions—that is, when either the host or the home market is cum-dividend but the other is ex-dividend. These days are frequent and common across exchanges, stocks, and time. For the average stock, I observe 21 days between corporate actions, and of the 194 stock-pairs, 137 have at least one day between corporate actions. It is striking that price deviations during these days are much higher than on other days, even when adjusting prices by the corporate action. For example, on days between corporate actions, the average daily maximum price deviation adjusted by the corporate action is about 4.79%; on other days, it is 2.12%. To the best of my knowledge, this is the first time that this important detail when estimating price deviations between home- and host-market stocks has been documented.

In Panel C of Table 1, I report summary statistics for illiquidity and control variables and differences in quoted spreads during and outside overlapping trading times, denoted $\delta PQSPR$. As in Werner and Kleidon (1996) and Moulton and Wei (2009), the overlapping trading time is defined as the time in which both the ADR and the home market are in their continuous trading session. For the home market I examine differences in proportional quoted spread during the overlapping trading time and from 11 UTC (to avoid the general effects of the opening period) until the ADR starts trading. For the ADR, I look at differences in proportional quoted spread during and after the overlapping time (until 17 UTC, to avoid the general effects of the closing period). Therefore, I do not estimate $\delta PQSPR$ for American countries because their trading sessions are practically the same as the trading sessions for their ADRs.

The average home market stock and ADR has a $\delta PQSPR$ of -9 and -4 basis points, respectively. A negative $\delta PQSPR$ indicates that quoted spreads during the overlap are on average lower than outside . Moulton and Wei (2009) similarly document that spreads during the overlap are lower for ADRs.

2. Do price deviations arise as a result of demand shocks or differences in information?

Theory predicts that the impact impediments to arbitrage have on liquidity depends on why arbitrage opportunities arise. If arbitrage opportunities arise as a result of non-fundamental demand shocks, arbitrageurs should act as "cross-sectional market makers" (Holden, 1995) and improve liquidity. But if arbitrage opportunities arise as a result of differences in information, arbitrageurs should increase adverse selection and deteriorate liquidity (Foucault et al., 2017).

I follow Foucault et al. (2017) and Schultz and Shive (2010) to investigate why price deviations arise. If for one particular stock i at time $t-1$ price deviations are zero, but at time t price deviations are positive, at least one bid or ask quote of at least one asset changed from time $t-1$ to time t (this asset—either the ADR, the home market share, or the respective currency pair—is denoted the *First mover*). Similarly, if price deviations are positive until time $\tau - 1 > t$, but zero at time τ, then at least one bid or ask quote of at least one asset changed (this asset is denoted the *Last mover*). In this case I say that the *First mover* creates a price deviation for stock i at time t, and the *Last mover* eliminates the price deviation at time τ.[12]

Table 2 reports the number of price deviations by the *First* and *Last mover*.[13] For each *First mover* I separately report the percentage of all price deviations that are *toxic*. Following Foucault et al. (2017) I consider a price deviation as "toxic" for market m if m was the *Last mover* and the other market was the *First mover*.

Table 2 reports 13,913,834 price deviations in my sample. Price movements in

[12] In cases when the day opens with a price deviation, I consider the asset whose market opened last as the *First mover*. On the other hand, if a price deviation exists and either of the markets closes, I drop this price deviation from the analysis, as I do not know which asset closes down the price deviation. Both cases are infrequent and do not affect the main results in this section.

[13] In the case that the currency pair moves simultaneously with the ADR or home market share, the *First mover* is considered to be the ADR or home market share.

the home market create 4,092,945 of these price deviations, of which 45% are later eliminated because the price of the home market moves back. In only 24% does a price movement of the ADR eliminate the price deviation. The percentage of all toxic price deviations for the home market is 6.50%. Similarly, the percentage of all toxic price deviations for the ADR is 7.03%. In the Online Appendix I show that the results reported in Table 2 are robust to only using the largest price deviation from each stock-day (Table A2).

Table 2 indicates that toxic arbitrage should be rare in the ADR market. This provides initial evidence that arbitrageurs in the ADR market trade against net market demand and act as "cross-sectional market makers" (Holden, 1995) most of the time. Of course, the overall impact of arbitrage on liquidity might still be negative.

To study the overall effect, I now turn to investigate the joint dynamics between the impediments to arbitrage and market liquidity.

3. The impact of impediments to arbitrage on market liquidity: contemporaneous analysis

3.1. Correlations between daily impediments to arbitrage and illiquidity

To understand the joint dynamics between price deviations and illiquidity, a natural first step is to study pairwise correlations.

Table 3 reports pairwise pooled Spearman rank correlations between price-deviation measures and quoted and effective spreads.[14] All correlations are positive and statistically significant at the 1% level. All four price-deviation measures are highly correlated with each other, with the lowest correlation of 61% falling between the maximum price deviation from quotes and the average duration that price deviations persist. Both quoted spread and effective spread are highly correlated with each other across both home and host markets. Correlations are also strong between price deviations and spreads, with correlations ranging from 29% (between average price deviations from quotes and quoted spreads, both for the ADR and the home market share) to 71%

[14] Results are robust (but weaker in magnitude) to using Pearson correlations; see Online Appendix Table A3.

(between average price deviations from trades and effective spreads of the home market share).

The strong positive correlation between price deviations in quotes and quoted spreads is surprising. Mechanically, an increase in quoted spreads in either the home market share or the ADR would lower price deviations. But the finding supports the argument that these price deviations measure impediments to arbitrage: higher illiquidity should be associated with higher impediments to arbitrage. Previous research shows that price deviations correlate with impediments to arbitrage such as imperfect information, short-selling constrains, or funding illiquidity (De Jong et al., 2009; Gagnon and Karolyi, 2010; Lamont and Thaler, 2003; Mitchell et al., 2002; Roll et al., 2007).

3.2. Arbitrageurs' relative speed, arbitrage mix and liquidity in the ADR market

To better understand the relation between illiquidity and impediments to arbitrage, I first adopt the analysis of Foucault et al. (2017) for the ADR market.

The idea of Foucault et al.'s model and empirical analysis is that if toxic arbitrage opportunities are common and arbitrageurs have a speed advantage to liquidity providers, arbitrageurs create adverse selection and lower liquidity. As defined in Section 2, a price deviation is toxic for one market if the share price of this market moved to eliminate a price deviation that was created by a price movement in the other market. Foucault et al. (2017) explain illiquidity of asset i—in Foucault, either one of three currency pairs; in this paper, either the ADR or a home market stock—on day d using the following regression model:

$$\begin{aligned} Illiq_{i,d} = \omega_i \times \chi_m + a_0 t + a_1 \pi_{i,d} + a_2 \phi_{i,d} + a_3 \alpha_{i,d} + a_4 \sigma_{i,d} \\ + a_5 Vola_{i,d} + a_6 Trsize_{i,d} + a_7 Quotes_{i,d} + a_8 Ted_{i,d} + \epsilon_{i,d} \end{aligned}$$

where on day d and for asset i, $\pi_{i,d}$ is the number of toxic price deviations that end with a trade divided by the number of toxic price deviations; $\phi_{i,d}$ is the number of toxic price deviations divided by the number of all price deviations; $\alpha_{i,d}$ is the number of price deviations divided by the number of trades; $\sigma_{i,d}$ is the average deviation in midquote prices; $Vola_{i,d}$ is the five-minute midreturn volatility; $Quotes$ is the number of updates to the quote; $Trsize$ is the average trade size; and Ted is the Ted spread, or the difference between three-month USD LIBOR and Treasury Bills. Further, t represents a time trend, and $\omega_i \times \chi_m$ are two-dimensional stock-month fixed effects.[15]

Table 4 shows the results. Setting endogeneity issues aside for now, I find that π_{Toxic} is negative and statistically significant. If the percentage of toxic price deviations ending with a trade increases by 1%, quoted spreads decrease an economically insignificant amount of 0.00008%, or 0.008 basis points. Similarly, the same ratio calculated using nontoxic price deviations ($\pi_{notToxic}$) is also negatively related to illiquidity, and statistically but not economically significant. Because these tests reject the hypothesis that π_{Toxic} is positively related to illiquidity (the finding in Foucault et al. (2017)) arbitrage might have a different effect on liquidity in the ADR market than it has in the foreign exchange market.

Foucault et al. (2017) focus on the importance of the relative speed of arbitrageurs and market-makers, proxied by π_{Toxic}, the percentage of toxic price deviations ending with a trade. After all, arbitrageurs only increase adverse selection risk for local market-makers if they can trade faster than market-makers can change their quotes. Speed seems to be particularly important in the foreign exchange market where, on average, price deviations only last for 1.5 seconds.

But in a market where arbitrage positions are more complex, speed might be of secondary importance. Indeed, the ability to determine the optimal time to trade after observing prices deviate might be more important: arbitrageurs neither want to trade too early, because of noise trader risk and margin requirements, nor too late, because of competition with other arbitrageurs bringing prices back inline (cf. Jarrow (2010) or Liu and Longstaff (2004)). The secondary importance of speed in the ADR market can be seen in the average duration of a price deviation of around 12 minutes.

In a market where the relative speed of arbitrageurs is of secondary importance and toxic price deviations are rare, the percentage of toxic price deviations ending with a trade might proxy for overall arbitrage activity. That is, the emphasis is on trading against price deviations rather than on that the price deviation happen to be identified as toxic. Supporting this interpretation is the finding that both π_{Toxic} and $\pi_{notToxic}$ are

[15] Results are robust to estimating panel regressions separately for home market stocks and their respective ADRs, see Table A4 in the Online Appendix.

negatively related to illiquidity.

The second important observation is that price deviations (σ) are positively related to illiquidity, and this relation is statistically and economically significant. In the three-period model of Foucault et al. (2017) σ is exogenously given, and the market-maker reacts by increasing the spread given σ. In multiperiod models, though, it seems necessary to endogenize σ because arbitrageurs are not just price takers (Jarrow, 2010). Previous research shows that price deviations can range from nearly half the share price over weeks for dual listed stocks (De Jong et al., 2009) to a few basis points over just a few seconds in the foreign exchange market (Foucault et al., 2017). Clearly, arbitrage (or the lack thereof) partly explains these huge variations in observable price deviations.

In continuous time it is common to model price deviations as a zero-mean, mean-reverting process such as a Brownian Bridge (Brennan and Schwartz, 1990; Roll et al., 2007). In such models it is common to interpret the speed of mean reversion as a measure of arbitrage activity.

As shown in Appendix C, σ is in general negatively correlated to the speed of mean reversion and hence an inverse measure of arbitrage activity. This is important, because following this interpretation, Table 4 indicates that arbitrage activity and illiquidity are negatively correlated.

But results in Table 4 need to be interpreted with caution. As previously mentioned, arbitrage activity and illiquidity are jointly determined, meaning that estimates are biased. The rest of this paper attempts to address these endogeneity concerns.[16]

3.3. Liquidity on days between corporate actions

In the previous two sections, correlations and regression results indicate a strong positive relation between illiquidity and price deviations. The goal of this paper is to study the effect impediments to arbitrage have on market liquidity. Empirically, it

[16] Foucault et al. (2017) are aware of endogeneity concerns and use AutoQuote on Reuters D-3000—that allowed traders to automate order submission—as an instrument related to the relative speed of arbitrageurs. In the current ADR content a similar structural break occurred in the NYSE with the introduction of the "Hybrid Market" (Hendershott et al., 2011). The main results in Table 4 are robust to using the introduction of the "Hybrid Market" as an instrument, except that the percentage of toxic price deviations ending with a trade is not statistical significantly related to illiquidity (see the Online Appendix, Table A5).

is challenging to study the effects of variables that are jointly determined, for example, because of reverse causality. This is likely the case between the impediments to arbitrage and liquidity: higher liquidity lowers impediments to arbitrage, and lower impediments to arbitrage encourage arbitrageurs to trade, which might improve liquidity either because prices become more informative (Cespa and Foucault, 2014) or, if arbitrage opportunities arise because of demand shocks, because arbitrageurs trade against market demand.

One way to address this challenge is to find a variable that is correlated with impediments to arbitrage but not directly correlated with illiquidity—that is, an instrument. While it is challenging to motivate and statistically impossible to verify that both assumptions hold, as a suitable candidate I propose a dummy variable that is one on days between corporate actions, that is, when either the host or the home market is cum-dividend but the other is ex-dividend.

For example, the Royal Bank of Scotland (RBS) proposed a distribution of rights that was approved during the annual shareholder meeting on May 14, 2008. This resulted in a stock dividend for the RBS stock in London with ex-date of May 15, 2008. But because the rights were not registered under the United States Securities Act of 1933, the Depository Bank sold these rights (from owning the home market stock underlying the ADR) in the home market and passed on the proceeds to the ADR holders as a special dividend. ADR holders received a special cash-dividend of USD 0.674089 with ex-date of May 29, 2008.

Accordingly, price deviations between May 15 and May 28 spiked with an average of USD 0.92 of the daily maximum difference between the bid of the ADR and the currency-adjusted ask of the home market. While these large price deviations (of almost 20%) do not reflect possible arbitrage profits, these days are likely characterized by higher impediments to arbitrage because of additional risk. Consider the simplest case in which holders of the home market share receive a cash dividend. Even in this case the final dividend payment for the ADR holder is unknown. After receiving the dividend, in general weeks $after$ the ex-date, the Depository Bank needs to exchange the home market currency into U.S. dollars and then pay the ADR holders. The holding period of the arbitrage position also increases significantly as prices will not converge until both stocks are ex-dividend and the Depository bank does not convert the home market share to its ADR (or vice versa) during these days (cf. Citibank (2007)).

In summary, arbitrageurs introduce uncertainty when adjusting prices for the corporate action to compute their profits and the expected holding period of the arbitrage position is much longer than on other days, making arbitrage more risky (or costly if this additional risk would be hedged away).

As such, it is not surprising that during these days price deviations are especially high even after adjusting prices by corporate actions (with the exact adjustment factor only known ex-post). As reported in Table 1, price deviations adjusted by the corporate action are more than twice as high on days between corporate actions as they are on other days. After adjusting the quotes by the dividend payment in the example before (i.e. subtracting USD 0.674089 from all bid and ask quotes of the ADR), price deviations are USD 0.24, almost 5% of the share price and more than three standard deviations higher than the average price deviation for RBS in the first quarter of 2008.

To motivate using days between corporate actions as an instrument, I first investigate whether liquidity of a stock is affected on these days. To have a valid instrument, liquidity should only be affected on these days because of changes in the impediments to arbitrage. So far relatively little evidence of how liquidity is affected by dividend and stock-split decisions exists. This is not surprising; theoretically, it is not always easy to argue why these decisions should matter at all (e.g., compare Merton H. Miller (1961)). Recently, both theoretical and empirical evidence indicates that these decisions might matter. For example, Muscarella and Vetsuypens (1996) show that liquidity improves after stock splits, and Banerjee et al. (2007) show that illiquid firms pay higher dividends. But it is even more difficult to argue why liquidity should be affected on days when either the host or the home market stock is cum-dividend but the other is ex-dividend. I believe liquidity is affected because the increased price deviation segments both markets and increases the impediments to arbitrage.

Table 5 shows the results of panel regressions explaining illiquidity by a dummy variable that is set to one on days between corporate actions. To avoid endogeneity issues, I estimate regressions without any control variables. To control for unobserved heterogeneity, I use two dimensional stock-months fixed effects (following the advice of Gormley and Matsa (2014)). In additional tests, I control for other, probably endogenously determined, variables as in Table 4. To rule out that corporate actions by themselves affect liquidity, I also control for days in which both the ADR and home market stock go ex-dividend together. Finally, I control for the number of trades and order imbalance, defined as the absolute difference in the number of buyer versus seller initiated trades.

In Panel A of Table 5 I investigate how days between corporate actions affect quoted spreads. The results indicate that on days between corporate actions quoted spreads are statistically significantly higher by around two basis points.

Similarly, in Panel B of Table 5 I proxy illiquidity by effective spread, and find that on days between corporate actions effective spreads are higher by almost three basis points.

To further rule out that corporate actions by itself affect liquidity, I investigate how days between corporate actions affect the difference in liquidity during and outside overlapping trading times.

Panel C of Table 5 reports the results where illquidity is proxied by the difference in quoted spread during and outside overlapping trading times (i.e., the difference in spreads when both the home- and the host-market stock are trading and when only one stock is trading) ($\delta PQSPR$).[17] Because of the different dependent variables in Panel C compared to Panel A and B, I use different control variables. Moulton and Wei (2009) examine two explanations for differences in liquidity during and outside overlapping trading times: (i) concentrated trading, and (ii) increased competition. Like Moulton and Wei (2009) I proxy the former by the difference between the number of trades during and outside the overlapping trading times ($\delta Trades$). I proxy the competition from the other exchange by the percentage of trades (during the overlapping trading times) that occur on the home market versus in the U.S. ($\Delta Trades$). I also control for the difference during and outside overlapping trading times of all other variables used in Panels A and B.

Using only fixed effects, the coefficient for the dummy variable in Panel C is estimated at 0.018 and is statistically significantly different from zero. With additional

[17] Because of the focus on the difference in variables during and outside overlapping trading times, I exclude countries with similar trading hours as the U.S. In particular, I drop all stock-pairs if the home market stock is from Argentina, Brazil, Chile, Mexico, and Peru.

controls the coefficient is 0.014 and is also statistically significant. These results indicate that on days between corporate actions quoted spreads are one to two basis points higher during overlapping trading periods than they are outside these periods. This is consistent with the idea that it is not the corporate action itself that affects liquidity, because this should affect liquidity throughout the whole trading day, but rather spillover effects from the other market. For example, the increased price deviation that resulted from the corporate action could lower traders' ability to learn from the other market, which lowers liquidity (Cespa and Foucault, 2014). Alternatively, the increase in the impediments to arbitrage during these days could stop arbitrageurs from trading on these days. Assuming arbitrageurs provide liquidity this would lead to a decrease in liquidity during these days.

3.4. Days between corporate actions as an instrument

In this section I estimate a two-stage panel regression with two-dimensional stock-month fixed effects as given below:

$$\Delta Price_{i,d} = FE + \beta_0 \times D^{EX}_{i,d} + \boldsymbol{\zeta_0} \times \boldsymbol{Controls}_{i,d} + \eta_{i,d} \tag{3a}$$

$$Illiq_{i,d} = FE + \beta_1 \times \widehat{\Delta Price}_{i,d} + \boldsymbol{\zeta_1} \times \boldsymbol{Controls}_{i,d} + \epsilon_{i,d} \tag{3b}$$

where $\Delta Price_{i,d}$ is a price deviation measure for stock i on day d, and $D^{EX}_{i,d}$ is a dummy variable set to one on days between corporate actions (the instrument, and motivated in the previous section). In the second equation $Illiq_{i,d}$ is a measure for illiquidity, $\widehat{\Delta Price}_{i,d}$ is the fitted value from the first equation (the first stage), $Controls_{i,d}$ are control variables as in Table 5, and FE are two-dimensional stock-month fixed effects.

The advantage of a panel regression, compared to individual stock time-series regressions, is that a panel regression can address omitted variables. For example, if time-varying funding liquidity influences the impediments to arbitrage and simultaneously market liquidity (as in Brunnermeier and Pedersen (2008)) this could cause an omitted variable bias.[18] If, however, funding liquidity affects stocks equally, adding time-fixed effects will control for the stock-invariant differences in time. Similarly, I can

[18] I control for the Ted spread, a measure for funding illiquidity, but the Ted spread is based on U.S. interest rates and might not fully capture changes in funding illiquidity of the home market.

control for time-invariant heterogeneity by using individual-fixed effects.

Table 6 shows the results of the first stage from estimating Eq. 3. I proxy impediments to arbitrage by the average duration price deviations persist, the average price deviation from simultaneous trade prices, and the average and maximum price deviation from quote prices (adjusted and not adjusted by corporate actions). I find that the dummy variable $D_{i,d}^{EX}$ is positively and statistically significantly related to all five proxies. For example, keeping all other variables constant, on days between corporate actions maximum price deviation from quotes adjusted by corporate actions are higher by 1.775%, which is statistically significant at the 1% level (computed from standard errors clustered by stock).

Table 7 shows the results of the second stage from estimating Eq. 3. Again I use all five proxies for the impediments to arbitrage, as in Table 6. Panel A of Table 7 report results where illiquidity is proxied by quoted spread ($PQSPR$). Results indicate that all five proxies are positively associated with illiquidity and the relation is statistically significant at least at the 10% level. For example, results indicate that a 1% increase in the maximum price deviation from quotes adjusted by corporate actions increases quoted spreads by 1.3 basis points.

Overall, control variables have the expected sign consistent with previous literature. Spreads decrease by the share price and by the number of trades. Spreads increase with volatility and the TED spread.

Panel B of Table 7 report results where illquidity is proxied by effective spreads ($PESPR$). Again all five different price-deviation measures are positively and statistically significant associated with illiquidity.

To further address the concern that the corporate action by itself affects liquidity, I exploit exogenous variation in the impediments to arbitrage within the day. Panel C of Table 7 report results where illiquidity is proxied by the difference in quoted spread during and outside overlapping trading times ($\delta PQSPR$).

In all ten regressions the estimated slope coefficient of all five price-deviation measures is positive and statistically significant at least at the 10% level. For example, the results indicate that a 1% increase in the maximum price deviation from quotes adjusted by corporate actions increases quoted spreads during the overlapping trading times compared to outside by 0.9 basis points. These results indicate that if the impediments to arbitrage increase, illiquidity increases during the time arbitrageurs are active (during overlapping trading times quoted spreads increase) relative to when they are not active.

Throughout the paper I estimate panel regressions with both home market and ADR stocks simultaneously. A valid concern might be that the results so far are driven by one subsection of the data. Similarly, it is of interest to find out whether the results are opposite in any particular subsection. After all, the sample contains a long time-series of 16 years and a large cross-section in ADRs and home market stocks from 22 different countries. Therefore, in the Online Appendix Table A6, I estimate regressions separately for home market stocks and their respective ADRs and also across three different regions and time-periods. In total I estimate 18 panel regressions for each of the illiquidity and price deviation proxies. While results are more often statistical significant for ADRs than for home market stocks, there is no indication that results are driven by any particular subset or that results might be of opposite sign in any particular subset. In only one case I estimate a statistically negative coefficient between illiquidity and price deviation, compared to 31 estimates that are statistically positive. Therefore, to maximize the power of the statistical tests it seems suitable to pool all data into one regression.

Another concern might be that results are spurious as both the dependent (illiquidity) and dependent variable (price deviations) are scaled by the same variable, home market price. Therefore, in the Online Appendix Table A8, I estimate regressions using the maximum price deviation within each stock-day measured in USD. In the Online Appendix I also estimate contemporaneous effects of impediments to arbitrage on liquidity excluding price deviations below one basis point and below one dollar cent per share to cover additional transaction costs (see Tables A9 and A10). Results are robust to these different specifications.

3.5. Spread decomposition

So far the results are consistent with two explanations. First, arbitrageurs might improve market liquidity by trading against market demand thereby lowering inventory holding costs as envisioned by, for example, Holden (1995). Second, arbitrageurs might make prices more informative, which improves liquidity (Cespa and Foucault, 2014).

One way to distinguish between both effects is to investigate which component of the bid-ask spread is affected by arbitrage. Since at least Stoll (1978) and Glosten and Milgrom (1985) research shows that illiquidity arises because of both inventory holding and adverse selection risk. If arbitrageurs trade against market demand, one would expect that they affect the component due to inventory holding risk. On the other hand, if arbitrageurs make prices more informative, one would expect that they affect the component due to adverse selection risk.

To test which component is affected, in the following, I decompose the bid-ask spread following Glosten and Harris (1988). The main idea is similar to the idea behind classifying price deviations as toxic or due to price pressure in Section 2: price effects due to inventory holding risk should be transitory and incorporating new information (which creates adverse selection risk) should have a permanent price effect.

As Glosten and Harris (1988) I focus on estimating a restricted version of their general model, given as Eq. (5) of their paper and as Eq. (4) below. Eq. (4) is estimated by stock-day (with at least ten trades) using ordinary least square (OLS).[19]

$$P_{i,d,t} - P_{i,d,t-1} = c_{0,i,d}(Q_{i,d,t} - Q_{i,d,t-1}) + z_{1,i,d}Q_{i,d,t}V_{i,d,t} + \epsilon_{i,d,t} \tag{4}$$

where $P_{i,d,t}$ is the t-th traded price (converted to USD) of stock i on day d, $Q_{i,d,t}$ is the sign of the trade (estimated using Lee and Ready (1991)) and $V_{i,d,t}$ is its dollar volume. To increase the power of the regression, I estimate Eq. (4) over the whole trading period and not just during overlapping trading times.

Table 8 shows the results of the second stage from estimating Eq. 3. In particular, Table 8 shows the results from panel regressions with two-dimensional month-ADR fixed effects. But now illiquidity is measured as the transitory component (Panel A: $c_{0,i,d}$ in Eq. 4) or the adverse selection component (Panel B: $z_{1,i,d}$ in Eq. 4) of the bid-ask spread. Both components are scaled to reflect the cost of a round trip in dollar cents,

[19] As mentioned in footnote 1 of Glosten and Harris (1988) estimation using OLS would be inefficient because of the round-off errors but considering that in my sample (at least ADR) prices are quoted in cents rounding errors are likely not that problematic.

i.e., both $c_{0,i,d}$ and $z_{1,i,d}$ are multiplied by 200.[20] Note, that regressions are estimated using ADRs only. Compared to before, results are sensitive to whether I pool both ADR and home market stocks in one panel or estimate the effect separately.

Panel A of Table 8 shows that all five price deviation measures are positively and statistically significantly related to the transitory component of the bid-ask spread. Results in Panel A of Table 8 show that, for example, if the duration price deviations persist increases by one hour (i.e., $INARB_{i,d}$ increases by 60) the transitory component of the bid-ask spread increases by 2.4 dollar cents. An economically large effect considering that the average transitory component of the bid-ask spread is just 1.2 cents with a standard deviation of 3.1 cents.

Panel B shows that four out of five price deviation measures are negatively related to the adverse selection component of the bid-ask spread, but none of the estimates are statistically significant.

Panel C and D of Table 8 report above analysis using home market stocks only. Results indicate that for home market stocks price deviations have a positive effect on the transitory component of the bid-ask spread, which is economically very large, but statistically insignificant. This can have several explanations, first, of course, it might indicate that in the home market arbitrage activity does not affect the transitory component (nor the adverse selection component) of the bid-ask spread. But results for the home market might also be less reliable. Compared to previous variables, such as quoted or effective spreads, estimating Eq. 4 is more challenging, for example, differences in reporting standards or wider tick-sizes in the home market would lead to inefficient estimates.

In short, results from Table 8 suggest that arbitrageurs mainly affect the transitory component of the bid-ask spread. These results support the idea that arbitrageurs lower inventory holding costs by trading against order imbalances.

Instrumental variable regressions allow to address endogeneity between two variables and estimate the contemporaneous effect of one variable on the other. But the impact of impediments to arbitrage on liquidity does not need to be contemporaneous

[20] As explained by Glosten and Harris (1988) the average dollar spread for a round-trip of V dollars is given by $2(c_0 + z_1 V)$.

alone. O'Hara and Oldfield (1986) and Comerton-Forde et al. (2010) provide theoretical and empirical evidence that overnight inventories affect future liquidity. If, for example, arbitrageurs trade against net market demand, an increase in the impediments to arbitrage might lead to larger order imbalances, which could predict a decrease in liquidity. To understand the longer term relation between price deviations and liquidity, I estimate vector autoregressions in the following section.

4. The impact of impediments to arbitrage on market liquidity: predictive analysis

Vector autoregressions (VARs) regress each variable on lagged versions of itself and on lagged versions of all other variables in the system. Especially, impulse response functions (IRF) constructed from a VAR are commonly used to yield important information about the dynamics of jointly determined variables (cf. Roll et al. (2007)). An IRF estimated from a VAR tracks the response on one variable from an impulse to another variable and hence allows investigating longer term effects. Using the Cholesky decomposition to calculate orthogonalized impulse responses an IRF also allows estimating contemporaneous effects. Because in the Cholesky decomposition a variable only has a contemporaneous effect on other variables if it enters the system of equations before the other variables, theory needs to guide the ordering of the variables (Doan, 2010).

In the following I fix the order to price deviation, market order imbalance, volatility, and illiquidity. The order is motivated by: First, Table 2 indicates that most price deviations arise because of a demand shock, and hence arbitrageurs would trade against market demand, contemporaneously affecting order imbalance. This motivates using price deviations as the first variable. Second, previous literature indicates that order imbalance contemporaneously affects volatility and liquidity (Chordia et al., 2002). This motivates the order between measures of order imbalance and measures of volatility and liquidity.

To avoid spurious results, I detrend all variables using a linear time-trend. For all detrended variables the Im–Pesaran–Shin test for unbalanced panels rejects the existence of a unit root with p-values less than 0.01.

In the following I estimate a panel VAR with day and stock fixed effects and 15 lags (chosen by the Schwarz information criteria) of the following vector of endogenous variables ($\boldsymbol{V}$): maximum price deviations from quotes adjusted by corporate actions, order imbalance, volatility, and quoted spreads.

$$\boldsymbol{V_{i,d}} = \boldsymbol{FE} + \boldsymbol{\rho} \times \boldsymbol{BetweenCorpAct_{i,d}} + \sum_{l=1}^{5} \boldsymbol{\beta} \times \boldsymbol{V_{i,d-l}} + \boldsymbol{\epsilon_{i,d}} \tag{5}$$

Equation 5 is estimated using generalized method of moments (GMM) as recommended by Arellano and Bond (1991) using all stocks with at least one year of data and days with at least ten stocks.

Table 9 reports Granger causality tests: the sum of all lagged coefficients for each variable and the associated p-values whether this sum is statistically significantly different from zero and whether the lagged coefficients are jointly different from zero. As before Panel A, Panel B, and Panel C differ in the way illiquidity is measured.

The results in Panel A indicate Granger causality between price deviations and illiquidity and vice versa. The results in Panel B and Panel C are largely consistent with results in Panel A. Because Granger causality tests are based on only one equation of the VAR they cannot provide a complete picture of the joined dynamics of all variables. A much better approach is constructing impulse response functions.

Figure 1, Figure 2, and Figure 3 show cumulative impulse response functions constructed from the panel VAR in Table 9 Panel A, B, and C, respectively. Figure 1 shows how shocks to price deviations, order imbalance, volatility, and quoted spreads affect each other contemporaneously and up to 15 days in the future. The first row presents the effect of a one standard deviation shock to price deviations on itself in the first column, and on order imbalance, volatility, and quoted spreads in the second, third, and fourth columns,respectively.

The focus of this paper is how price deviations effect illiquidity (shown in the upper, right corner of Figure 1). A shock to price deviations predicts a strong increase in quoted spreads. A one standard deviation shock to price deviations predicts an increase in quoted spread of around three basis points in the next 15 days. Of these three basis points around one-third of a basis point is contemporaneous, which is lower than previous results from the instrumental variable regression in Table 7. A shock to price deviations also predicts an increase in order imbalance contemporaneously and in the future.

Figure 2 provide impulse response functions estimated with effective spreads, instead of quoted spreads as in Figure 1. Results are much stronger, indicating that a shock to price deviations predict an increase in effective spreads of almost 5 basis points contemporaneously and over 10 basis points after 15 days.

So far, results indicate that if price deviations increase, illiquidity increases both contemporaneously and over the next days. One explanation for this finding is that if price deviations increase, traders can learn less from prices in the other market, which should lower liquidity (Cespa and Foucault, 2014). Another explanation is that arbitrageurs normally trade against market demand and thereby improve liquidity.

But while Figure 1 indicates that price deviations predict illiquidity, it does not need to be causal. For example, arbitrageurs might be able to predict general changes in liquidity and in anticipation of increasing market or funding illiquidity step out of the market (Shleifer and Vishny, 1997; Bernardo and Welch, 2004), potentially causing higher price deviations. In this case liquidity would deteriorate, regardless of whether arbitrageurs step out of the market or decide to continue to be active. In other words, one concern might be that an omitted variable (that has a stock-specific effect) could drive the predictive power of price deviations on market liquidity. To investigate this question, I again exploit exogenous variation in the impediments to arbitrage within the day.

Figure 3 provide impulse response functions estimated with the intraday difference in quoted spreads during and outside overlapping trading times. Consistent with previous results a one standard deviation increase in price deviations predicts a statistically significant contemporaneous and future increase in quoted spreads during the overlap compared to outside. In other words, the results in Figure 3 indicate that price deviations do not predict a general increase in illiquidity, but rather an increase in illiquidity during the overlapping trading times compared to outside.

In the Online Appendix I report several robustness tests. I estimate IRFs from a VAR using the other three proxies for arbitrage activity (Figures A1, A2, and A3).

And I estimate IRFs from a VAR using price deviations calculated from mid-quote prices (A4) and from U.S. dollars (A5). To rule out that results are driven by the ordering of the variables, I estimate IRFs using the reverse order of the input variables (Figure A6). I estimate IRFs separately for home market stocks and ADRs (Figure A7 and Figure A8). I estimate IRFs from a VAR using weekly data by taking averages across all variables within the week (Figures A9, A10, and A11).

The main results reported in this paper are robust to these changes, except when using average price deviations from quotes (Figure A3). Further, results are only robust in the long-run when using the average time a price deviation persists as an alternative proxy for the impediments to arbitrage (Figure A1).

5. Conclusion

Arbitrageurs enforce the law of one price by trading against mispricings, but whether by doing so arbitrageurs provide liquidity depends on the reason for the arbitrage opportunity to arise. In this paper I study price deviations in the American Depositary Receipts market, because here arbitrage is almost risk-free and institutional details provide exogenous variation in arbitrage activity within and across days.

Results show that large price deviations are associated with contemporaneous and future illiquidity, an increase in quoted spreads, effective spreads, and the transitory component of the bid-ask spreads. Illiquidity is particularly affected during overlapping trading times, i.e., when arbitrageurs are active. These results are consistent with the idea that arbitrageurs provide liquidity by trading against net market demand, or as Foucault et al. (2013) put it, arbitrageurs are "leaning against the wind" (p. 336). Arbitrageurs in the ADR market are indeed "cross-sectional market-makers" (Holden, 1995).

One way to encourage arbitrage activity is to introduce portfolio margins, where the offsetting position between the home market stock and the associated ADR are incorporated in the margin requirements. Within the U.S., a similar concept is already approved by the SEC for example for index options.

Appendix A: Sample construction

This appendix describes details of the sample construction. I first retrieve all dead and alive American Depositary Receipts (ADRs) from Datastream which are traded at the New York Stock Exchange (NYSE) or Nasdaq. I focus on the NYSE and Nasdaq as host-markets because together they capture almost 90% of the worldwide total trading in the DR market of USD 3.5 trillion in 2010 (Cole-Fontayn, 2011).

I identify the home market share, associated to any of the ADRs, using Datastream and verify the match using data from adrbnymellon.com and adr.db.com. Both websites offer a list of DRs and an ISIN code for the home market share.

As the analysis requires intraday data for which I use the Thomson Reuters Tick History (TRTH) database, I filter out any DR for which I could not establish the RIC (the primary identifier in TRTH) for either the DR or the home market stock. Upon request Datastream provides a RIC field, however this field is empty for around 50% of all DRs. In the case of a missing RIC field for the ADR or for the home market shares I use the TRTH API to search for a RIC code by ISIN.

For every ISIN the RIC from the major exchange of the home market country is chosen. This way I identify 351 ADRs and their respective home market share. A similar setup (i.e. using intraday data from TRTH, albeit for an event study) is considered by Berkman and Nguyen (2010), who identify 277 ADR and home market pairs. Further Gagnon and Karolyi (2010) identifies 506 U.S. cross-listed stocks, but they include other forms of cross-listing than ADRs, such as Canadian ordinaries (in total, 127 pairs) and New York registered shares.

To avoid non-synchronous prices, I only focus on countries with overlapping trading times to the U.S. This excludes all Asian countries (China, Hong Kong, India, Indonesia, Japan, Philippines, South Korea, and Taiwan) as well as Australia and New Zealand. I am left with 238 pairs from the following 27 countries: Argentina (12 pairs), Austria (1), Belgium (2), Brazil (27), Chile (13), Colombia (1), Denmark (1), Finland (1), France (13), Germany (12), Greece (2), Ireland (3), Israel (9), Italy (4), Luxembourg (2), Mexico (23), Netherlands (12), Norway (3), Peru (3), Portugal (1), Spain (8), Sweden (6), Switzerland (5), Turkey (1), United Kingdom (58), Russian Federation (3), and South Africa (12).

Of these 238 stocks I filter out 30 (and with it five countries drop), because I could not align prices of the home market with prices of the ADR, as described in more detail on page 32.

Of the remaining 208 stocks, I have 194 stocks with at least one year of data from 2001 to 2016 (the sample period for the main analysis). In the internet appendix I report summary statistics for the main variables used in this paper separately for each of the 22 countries (Table A1).

Appendix B: Data filters

This appendix describes the quote and trade data filters, based on all available data from 1996 to 2016. I discard non-positive bid and ask quotes (in total 196,061 quotes), quotes where the ask is lower or equal to the bid quote (193,936,018 quotes), and quotes outside the continuous trading session (814,278,837 quotes).[21] Further, I remove outliers (144,258 quotes). An outlier is defined as a bid (ask) quote that differs by more than 10% of the average of the ten surrounding bid (ask) quotes. In a similar way trade prices are filtered. Of the in total 4.45 billion trades, I discard trades that fall outside the continuous trading session in the U.S. (in total 13,631,796 trades) and on the respective home market (50,938,498). I also discard trades with a non-positive price (3,190,264 trades) or a price that is more than 10% different from the trade price of the ten surrounding trades for ADRs (1,529 trades) and home market shares (13,542,154). Further, I discard trades of more than 100,000 shares for ADRs (170,142 trades) and in the home market (7,060,211), because large trades are often negotiated before they get reported.

To make prices comparable between the home market stock and the ADR, bid and ask quotes of the ADR are converted according to the bundling ratio which I got from either Datastream, adrbnymellon.com, or adr.db.com. Bundling ratios can be time-varying but all sources only report the latest bundling ratio. To adjust the bundling ratio over time, I get all corporate actions from TRTH for all ADRs and home market

[21] Most quotes outside continuous trading times are from the ADR. ADR prices seem to update when the home market share is traded, regardless of whether the ADR is in its continuous trading session.

shares. Changes in the bundling ratio can occur, for example because of solo stock-splits. To verify the accuracy of the bundling ratio I plot the daily average, currency adjusted mid-quote ratios for each stock in the sample (unreported). If the ratio does not vary around one and does not resemble a step function, the stock is dropped from the sample. As such 30 stocks are dropped from the sample because prices of the home market could not be aligned to prices of the ADR.

For 82 ADRs the bundling ratio changed over the sample, with a maximum of three changes for four ADRs. For 59 ADRs the bundling ratio changed once over the sample.

To further ensure that stocks are mapped properly and prices are adjusted correctly, I drop any stock-day if (adjusted) price deviations ΔQTE are higher than 100% (of the mid-quote of the home market, in total 273 stock-days) or higher than USD 5 (1,285 stock-days). These very large price deviations are driven by two stocks from Argentina with 718 and 536 days with price deviations above USD 5.[22]

Appendix C: Large price deviations indicate less arbitrage activity

In this appendix I provide motivation for interpreting large price deviations as an indication of low arbitrage activity. If price deviations follow a mean-reverting Ornstein–Uhlenbeck process, then the duration price deviations persist and the magnitude of a price deviation are, in general, negatively correlated to the speed of mean reversion.

Assume that price deviations X follow an Ornstein–Uhlenbeck process:

$$dX(t) = -\theta X(t)dt + \sigma dW(t) \tag{6}$$

with $X(0) = 0$ and where $\theta > 0$ is the speed of mean-reversion, $\sigma > 0$ is the instantaneous volatility, and $W(t)$ is a standard Brownian motion. Ignoring the stochastic fluctuations of $X(t)$, $X(t)$ reverts to its mean exponentially at the rate of θ.

The (unrealized) profits of an arbitrage at time t and entered at time $\mathcal{T}_{0,p} = inf\{t \geq 0 \mid X(t) = p\}$ are given by $Y_t = X_{\mathcal{T}_{0,p}} - X_t$, where the assumption is that transaction costs are zero and arbitrageurs only trade on positive deviations, i.e., $p > 0$ (the case of

[22] The main results of this paper are robust when dropping these filters, see the Internet Appendix Table A11.

$p < 0$ is equivalent). For simplicity, I ignore that arbitrageurs could convert one asset to the other and thereby close down the position, instead I assume that arbitrageurs close down the position once the price deviation reverts back to zero at time $\mathcal{T}_{0,p,0} = inf\{t \geq \mathcal{T}_{0,p} \mid X(t) = 0\}$ with a profit of $X_{\mathcal{T}_{0,p}} - X_{\mathcal{T}_{0,p,0}} = X_{\mathcal{T}_{0,p}} = p$

The interpretation is that $\mathcal{T}_{0,p,0}$ is one trading cycle (also called one excursion, the first passage time of 0 after hitting p), where the arbitrageur sets up the position if price deviations are reasonably large (given by p) and closes down the position when price deviations disappear (i.e., $X(t) = 0$).

The following Lemma provides insights how trading cycles are related to mean-reversion.

Lemma 1. *The expected duration of one trading cycle is given by:*

$$d(p, \theta, \sigma) = E[\mathcal{T}_{0,p}] + E[\mathcal{T}_{p,0}] = \frac{2\sqrt{\pi}}{\theta} \sum_{k=0}^{\infty} \frac{(p\sqrt{\theta}/\sigma)^{2k+1}}{k!(2k+1)} \tag{7}$$

Proof. See Hollstein (2010), Corollary 4.1.2. Ricciardi and Sato (1988) and Finch (2004) provide the solution for the special case of $\theta = 1$ and $\sigma^2 = 2$, from which the general case also follows (see p. 46 of Ricciardi and Sato (1988)). $\square$

Theorem 1. *The hypothesis is that the following empirically observable variables are negatively correlated to the speed of mean reversion:*

1. *the maximum and average price deviations computed from quote or simultaneous trade prices [see Eq. (2) and Eq. (1)].*
2. *the average time price deviations persist, i.e., the duration of one trading cycle $E[\mathcal{T}_{0,p,0}]$.*

Proof. The first hypothesis follows directly from the variance of $X(t)$, which is $\frac{\sigma^2}{2\theta}$. Observing large price deviations indicate a high variance for X, hence indicating a high instantaneous variance σ^2 or low mean-reversion θ. In other words, keeping σ constant, observing large price deviations hence indicates low mean-reversion.

The second hypothesis follows from above Lemma. The first derivative of $d(p, \theta, \sigma)$ in Equation 7 with respect to θ is given by:

$$\frac{\partial d(p,\theta,\sigma)}{\partial\theta} = 2\sqrt{\pi}\frac{p}{\sigma}\left(-\frac{1}{\theta\sqrt{\theta}} + \sum_{k=1}^{\infty}(2k-1)\theta^{k-1/2}\frac{(p/\sigma)^{2k}}{k!(2k+1)}\right) \tag{8}$$

This derivative is negative if

$$\frac{1}{\theta\sqrt{\theta}} > \sum_{k=1}^{\infty}(2k-1)\theta^{k-1/2}\frac{(p/\sigma)^{2k}}{k!(2k+1)} \tag{9}$$

It follows that for reasonable parameter choices ($p/\sigma < 1$ and $\theta < 0.8$) the first derivative of $d(p,\theta,\sigma)$ in Equation 7 with respect to θ is negative, because:

$$\begin{aligned}
&\sum_{k=1}^{\infty}(2k-1)\theta^{k+1}\frac{(p/\sigma)^{2k}}{k!(2k+1)} \\
<&\sum_{k=1}^{\infty}\frac{\theta^{k+1}}{k!} \quad \text{if} \quad p/\sigma < 1 \\
=&-\theta + \theta \times exp(\theta) \\
<&1 \quad \text{if} \quad \theta < 0.8
\end{aligned}$$

And in this case the second hypothesis follows.

□

Amihud, Y. (2002). Illiquidity and stock returns: cross-section and time-series effects. *Journal of Financial Markets 5*, 31–56.

Amihud, Y. and H. Mendelson (1986). Asset pricing and the bid-ask spread. *Journal of Financial Economics 17*, 223–249.

Arellano, M. and S. Bond (1991). Some tests of specification for panel data: Monte carlo evidence and an application to employment equations. *Review of Economic Studies 58*(2), 277–297.

Banerjee, S., V. A. Gatchev, and P. A. Spindt (2007). Stock market liquidity and firm dividend policy. *Journal of Financial and Quantitative Analysis 42*, 369–397.

Baruch, S., G. A. Karolyi, and M. L. Lemmon (2007). Multimarket trading and liquidity: Theory and evidence. *The Journal of Finance 62*(5), 2169–2200.

Ben-David, I., F. Franzoni, and R. Moussawi (2014). Do ETFs Increase Volatility? *working paper, Ohio State University*.

Berkman, H. and N. H. Nguyen (2010). Domestic liquidity and cross-listing in the United States. *Journal of Banking & Finance 34*, 1139–1151.

Bernardo, A. E. and I. Welch (2004). Liquidity and financial market runs. *The Quarterly Journal of Economics 119*, 135–158.

Bodurtha, Jr, J. N., D.-S. Kim, and C. M. Lee (1995). Closed-end country funds and u.s. market sentiment. *Review of Financial Studies 8*, 879–918.

Boehmer, E. and E. K. Kelley (2009). Institutional investors and the informational efficiency of prices. *Review of Financial Studies 22*, 3563–3594.

Brennan, M. and E. Schwartz (1990). Arbitrage in stock index futures. *The Journal of Business 63*, 7–31.

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

Brunnermeier, M. K. and L. H. Pedersen (2008). Market Liquidity and Funding Liquidity. *Review of Financial Studies 22*, 2201–2238.

Cespa, G. and T. Foucault (2014). Illiquidity contagion and liquidity crashes. *Review of Financial Studies 27*, 1615–1660.

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Vega (2014a). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance 69*, 2045–2084.

Chaboud, A., B. Chiquoine, E. Hjalmarsson, and C. Vega (2014b). Rise of the machines: Algorithmic trading in the foreign exchange market. *The Journal of Finance 69*, 2045–2084.

Chan, K., A. Hameed, and S. T. Lau (2003). What if trading location is different from business location? evidence from the jardine group. *The Journal of Finance 58*(3), 1221–1246.

Choi, D., M. Getmansky, and H. Tookes (2009). Convertible bond arbitrage, liquidity externalities, and stock prices. *Journal of Financial Economics 91*, 227–251.

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

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

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

Citibank (2007). Manufacturas de papel – ratio change, par value change and reverse split. `https://depositaryreceipts.citi.com/adr/notices/previewPdf.aspx?caId=18` [Online; accessed 18-April-2017].

Cole-Fontayn, M. (2011). The Depositary Receipt Market 2010 Yearbook. *BNY Mellon*.

Comerton-Forde, C., T. Hendershott, C. M. Jones, P. C. Moulton, and M. S. Seasholes (2010). Time Variation in Liquidity: The Role of Market-Maker Inventories and Revenues. *Journal of Finance 65*, 295–331.

De Jong, A., L. Rosenthal, and M. van Dijk (2009). The Risk and Return of Arbitrage in Dual-Listed Companies. *Review of Finance 13*, 495–520.

Doan, T. A. (2010). RATS Handbook for Vector Autoregressions. *working paper, Estima*.

Domowitz, I., J. Glen, and A. Madhavan (1998). International Cross-Listing and Order Flow Migration : Evidence from an Emerging Market. *Journal of Finance 53*, 2001–2027.

Ellis, K., R. Michaely, and M. O'Hara (2000). The accuracy of trade classification rules: Evidence from nasdaq. *The Journal of Financial and Quantitative Analysis 35*, 529–551.

Fama, E. F. (1991). Efficient capital markets: II. *Journal of Finance 46*, 1575–1617.

Finch, S. (2004). Ornstein-uhlenbeck process. https://oeis.org/A249417/a249417.pdf [Online; accessed 17-November-2017].

Fong, K., W. C. Holden, and C. Trzcinka (2017). What Are the Best Liquidity Proxies for Global Research? *Review of Finance 21*(4), 1355–1401.

Foucault, T., R. Kozhan, and W. W. Tham (2017). Toxic Arbitrage. *Review of Financial Studies 30*, 1053–1094.

Foucault, T., M. Pagano, and A. Röell (2013). *Market Liquidity Theory, Evidence, and Policy*. New York: Oxford University Press.

Gagnon, L. and A. G. Karolyi (2010). Multi-market trading and arbitrage. *Journal of Financial Economics 97*, 53–80.

Gagnon, L. and A. G. Karolyi (2013). *Do International Cross-listings Still Matter? in Evidence on Financial Globalization and Crises*. London: Elsevier.

Gagnon, L. and G. A. Karolyi (2009). Information, Trading Volume, and International Stock Return Comovements: Evidence from Cross-Listed Stocks. *Journal of Financial and Quantitative Analysis 44*, 953–986.

Galligan, C., M. Vlasevich, and P. Cheng (2014). Traders' guide to global equity markets. http://www.convergex.com/uploads/Traders_Guide_2014_Q4.pdf [Online; accessed 18-April-2017].

Glosten, L. and P. Milgrom (1985). Bid, ask and transaction prices in a specialist market with heterogeneously informed traders. *Journal of financial economics 14*, 71–100.

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

Gormley, T. A. and D. A. Matsa (2014). Common errors: How to (and not to) control for unobserved heterogeneity. *The Review of Financial Studies 27*, 617–661.

Gromb, D. and D. Vayanos (2010). Limits of arbitrage: The state of the theory. *Annual Review of Financial Economics 2*, 251–275.

Harris, J. H. and P. H. Schultz (1998). The trading profits of {SOES} bandits. *Journal of Financial Economics 50*, 39 – 62.

Hendershott, T., C. M. Jones, and A. J. Menkveld (2011). Does algorithmic trading improve liquidity? *Journal of Finance 66*, 1–33.

Holden, C. W. (1995). Index arbitrage as cross-sectional market making. *Journal of Futures Markets 15*, 423–455.

Hollstein, M. (2010). Option Valuation, Optimization and Excursions of Commodity Indices. *PhD dissertation, Technische University Kaiserslautern.*

Hu, G. X., J. Pan, and J. Wang (2013). Noise as Information for Illiquidity. *Journal of Finance 68*, 2341–2382.

Jarrow, R. A. (2010). Optimal Trading of Arbitrage Opportunities with Market Impac. *International Review of Applied Financial Issues and Economics 2.3*, 552–575.

Kahraman, B. and H. E. Tookes (2017). Trader leverage and liquidity. *The Journal of Finance 72*, 1567–1610.

Karolyi, G. A. (1998). Why do companies list shares abroad?: A survey of the evidence and its managerial implications. *Financial Markets, Institutions & Instruments*, 1–60.

Kumar, P. and D. J. Seppi (1994). Information and index arbitrage. *Journal of Business 67*, 481–509.

Lai, S., L. Ng, and B. Zhang (2014). Does PIN Affect Equity Prices Around the World? *Forthcoming Journal of Financial Economics*.

Lamont, O. a. and R. H. Thaler (2003). Can the Market Add and Subtract? Mispricing in Tech Stock Carve-outs. *Journal of Political Economy 111*, 227–268.

Lau, S., L. Ng, and B. Zhang (2012). Information environment and equity risk premium volatility around the world. *Management Science 58*, 1322–1340.

Lee, C. M. and B. Radhakrishna (2000). Inferring investor behavior: Evidence from TORQ data. *Journal of Financial Markets 3*, 83–111.

Lee, C. M. and M. J. Ready (1991). Inferring trade direction from intraday data. *Journal of Finance 46*, 733–746.

Liu, J. and F. A. Longstaff (2004). Losing money on arbitrage: Optimal dynamic portfolio choice in markets with arbitrage opportunities. *The Review of Financial Studies 17*, 611–641.

Lou, D. and C. Polk (2013). Comomentum: Inferring arbitrage activity from return correlations. *working paper, London School of Economics*.

Marshall, B. R., N. H. Nguyen, and N. Visaltanachoti (2011). Commodity Liquidity Measurement and Transaction Costs. *Review of Financial Studies 25*, 599–638.

Menkveld, A. (2013). High frequency trading and the new-market makers. *Journal of Financial Markets 16*, 712–740.

Merton H. Miller, F. M. (1961). Dividend policy, growth, and the valuation of shares. *The Journal of Business 34*, 411–433.

Mitchell, M., T. Pulvino, and E. Stafford (2002). Limited Arbitrage in Equity Markets. *Journal of Finance 57*, 551–584.

Moulton, P. C. and L. Wei (2009). A tale of two time zones: The impact of substitutes on cross-listed stock liquidity. *Journal of Financial Markets 12*, 570–591.

Muscarella, C. and M. Vetsuypens (1996). Stock splits: Signaling or liquidity? The case of ADR 'solo-splits'. *Journal of Financial Economics 42*, 3–26.

Odders-White, E. R. (2000). On the occurrence and consequences of inaccurate trade classification. *Journal of Financial Markets 3*, 259–286.

O'Hara, M. and G. S. Oldfield (1986). The microeconomics of market making. *Journal of Financial and Quantitative Analysis 21*, 361–376.

Pastor, L. and R. Stambaugh (2003). Liquidity risk and expected stock returns. *Journal of Political Economy 111*, 642–685.

Ricciardi, L. M. and S. Sato (1988). First-passage-time density and moments of the ornstein-uhlenbeck process. *Journal of Applied Probability 25*(1), 43–57.

Roll, R. W., E. S. Schwartz, and A. Subrahmanyam (2007). Liquidity and The Law of One Price: The Case of the Cash/Futures Basis. *Journal of Finance 62*, 2201–2234.

Schultz, P. and S. Shive (2010). Mispricing of dual-class shares: Profit opportunities, arbitrage, and trading. *Journal of Financial Economics 98*, 524–549.

Shleifer, A. and R. W. Vishny (1997). The Limits of Arbitrage. *Journal of Finance 52*, 35–55.

Stoll, H. (1978). The supply of dealer services in securities markets. *The Journal of Finance 33*, 1133–1151.

Theissen, E. (2001). A test of the accuracy of the lee/ready trade classification algorithm. *Journal of International Financial Markets, Institutions and Money 11*, 147 – 165.

Tomio, D. (2017). Arbitraging Liquidity. *working paper, Darden School of Business*.

Werner, I. M. and A. W. Kleidon (1996). UK and US trading of British cross-listed stocks: An intraday analysis of market integration. *Review of Financial Studies 9*, 619–664.

Table 1 – Cross-sectional summary statistics of time-series averages, 2001 - 2016

This table reports the cross-sectional average, standard deviation, and 5, 50, and 95 percentile of the time-series average by stock of price deviation and liquidity measures. Panel A reports summary statistics computed as averages over all days on which both the host (the ADR) and the home market stock are either cum- or ex-dividend. Panel A reports the average number of minutes for a price deviation to persist, conditional on price deviations that arise and vanish within the same day ($INARB$); The daily average in trade prices, computed from simultaneous trades, i.e. trades within one second on both the host and the home-market share ($avg(\Delta TRD)$) (as in Eq. 1); and the daily time-weighted average and maximum price deviation computed from quotes ($avg(\Delta QTE)$ and $max(\Delta QTE)$), computed for every stock and in every second as the difference in the highest bid and the lowest ask price across the home- and host-market relative to the mid price of the home market (as in Eq. 2). Panel B provide summary statistics of price deviations on days between corporate actions, i.e., when either the host or the home-market is cum-dividend but the other is ex-dividend. *# days*, provides cross-sectional summary statistics of the total number of days between corporate actions. During these days I also compute price deviations from quotes, which are not adjusted for the corporate action ($UQTE$). Panel C reports cross-sectional summary statistics of liquidity and control variables, the daily time-weighted average proportional quoted spread for the home- and host-market share ($PQSPR$), the daily proportional effective spread for the home- and host-market share ($PESPR$), the difference in quoted spread for the home-market share between the overlapping trading times and from 11 UTC until the host-market opens ($\delta PQSPR$ *Home*), the difference in quoted spread for the host-market share between the overlapping trading times and from the time the home-market closes until 17 UTC ($\delta PQSPR$ *Host*), volatility for the home- and host-market stock, defined as the annualized volatility calculated from 5-minute mid-quote returns (*Volatility*), average trade size in the number of shares for the home- and host-market stock (*Trsize*), number of trades for the home- and host-market stock (*Trades*), number of quote updates to the best bid and ask for the home- and host-market stock (*Quotes*), the absolute order imbalance (the number of buyer minus seller initiated trades) for the home- and host-market (OIB), *size* calculated as the market cap for each stock at the beginning of each year, and *price* calculated as the end-of-day price from Datastream. Both *size* and *price* are in USD. All variables (except, $\delta PQSPR$, *price*, and *size*) are measured during the overlapping trading times only, i.e. when both the home- and host-market share are trading. All variables are scaled as denoted in the table, in particular all illiquidity and price deviations are in percent. All price deviation and spread measures are cross-sectionally winsorized each day at the 99% level. Data to compute *size* and *price* is from Datastream all other data underlying the computations is from TRTH.

	avg	stddev	p5	median (p50)	p95
Panel A: Price deviations outside days between corporate actions:					
$INARB$ [minutes]	12.41	18.54	0.51	6.41	39.27
$avg(\Delta TRD)$ [%]	2.74	5.54	0.12	0.77	10.10
$avg(\Delta QTE)$ [%]	0.80	2.13	0.00	0.21	5.76
$max(\Delta QTE)$ [%]	2.12	4.50	0.14	0.77	10.33
Panel B: Price deviations during days between corporate actions:					
# days	21	28	1	11	74
$INARB$ [minutes]	12.70	17.43	0.29	5.81	47.22
$avg(\Delta TRD)$ [%]	4.65	8.97	0.15	0.99	23.47
$avg(\Delta QTE)$ [%]	2.47	4.87	0.02	0.54	13.80
$max(\Delta QTE)$ [%]	4.79	8.15	0.36	1.38	25.25
$max(\Delta UQTE)$ [%]	10.75	22.47	0.72	2.33	80.14

Table 1 continued						
		avg	stddev	p5	median (p50)	p95
Panel C: Liquidity and control variables						
PQSPR [%]	Home	0.64	1.20	0.06	0.21	2.78
	Host	0.72	1.11	0.06	0.27	2.76
PESPR [%]	Home	0.64	1.05	0.05	0.30	2.14
	Host	0.53	0.85	0.05	0.20	2.10
$\delta PQSPR$ [%]	Home	-0.04	0.22	-0.70	-0.00	0.03
	Host	-0.09	0.34	-0.78	-0.01	0.23
Volatility	Home	0.23	0.08	0.13	0.19	0.38
	Host	0.22	0.08	0.12	0.19	0.35
Trsize ($*10,000$)	Home	0.32	0.49	0.02	0.17	1.12
	Host	0.04	0.02	0.02	0.04	0.08
Trades ($*10,000$)	Home	0.16	0.21	0.002	0.09	0.51
	Host	0.23	0.58	0.001	0.06	0.82
Quotes ($*10,000$)	Home	0.18	0.18	0.004	0.15	0.48
	Host	0.28	0.31	0.01	0.19	0.88
$\|OIB\|$ ($*1,000$)	Home	0.20	0.29	0.01	0.12	0.83
	Host	0.17	0.29	0.003	0.08	0.65
Size [billion USD]		41.30	234.96	0.29	7.91	101.05
Price [USD]		42.34	258.63	0.55	8.81	67.63

Table 2 – Average number of daily price deviations and reasons for why they arise, 2001 - 2016 This table presents the total number of price deviations (*# Price deviations*) by the asset that moves to create the deviation (*First mover*) and by the asset that moves to eliminate it (*Last mover*). The first column indicates the asset that moves to create the price deviation: either the home-market share (*Home*), the host-market share (*Host*), both the home- and the host-market share (*Both*), or the respective currency pair (*Forex*). The second column (*#Price deviations*) indicates the total number of price deviations across all stocks and days in this category. The third column (*%Toxic*) indicates the percentage of all price deviations that are toxic, when one share moves to create the price deviation and later the other moves back to eliminate it (for example, if the *Home*-market share is the first mover *%Toxic* is defined as the number of price deviations starting in the *Host*-market and ending in the *Home* market as a percentage of all price deviations). The rest of the columns *Home*, *Host*, *Both*, and *Forex* indicate the percentage of all price deviations that get eliminated because of a movement in the respective asset. All data underlying the computations are from TRTH.

			Last mover:			
First mover	#Price deviations	%Toxic	*%Home*	*%Host*	*%Both*	*%Forex*
Home	4,092,945	6.50	44.65	23.90	16.39	15.06
Host	4,235,914	7.03	21.33	47.09	16.12	15.47
Both	2,484,622		27.36	29.35	29.74	13.55
Forex	3,100,353		20.62	22.54	13.03	43.81

Table 3 – Pooled Spearman rank correlations of daily price deviation and illiquidity measures, 2001 - 2016

This table reports pooled Spearman rank correlations between the following daily measures: the average duration price deviations persist ($INARB$), the average price deviation from simultaneous trades ($avg(\Delta TRD)$), the average and maximum price deviations from quotes ($avg(\Delta QTE)$ and $max(\Delta QTE)$, respectively), home- and host-market proportional quoted ($PQSPR$) and effective spread ($PESPR$). For a description of these variables I refer to Table 1. All measures are computed during the overlapping trading time, i.e. when both the home- and host-market are in their continuous trading session. All correlations are significant at the 1% level. All data underlying the computations are from TRTH.

					$PQSPR$		$PESPR$	
	$INARB$	$avg(\Delta TRD)$	$avg(\Delta QTE)$	$max(\Delta QTE)$	$Home$	$Host$	$Home$	$Host$
$INARB$	100%							
$avg(\Delta TRD)$	67%	100%						
$avg(\Delta QTE)$	79%	72%	100%					
$max(\Delta QTE)$	61%	71%	78%	100%				
$PQSPR_{Home}$	35%	52%	29%	60%	100%			
$PQSPR_{Host}$	37%	54%	29%	55%	69%	100%		
$PESPR_{Home}$	44%	71%	41%	65%	79%	61%	100%	
$PESPR_{Host}$	40%	59%	34%	59%	69%	96%	65%	100%

Table 4 – Arbitrageurs' relative speed, arbitrage mix and liquidity in the ADR market, 2001 - 2016

This table presents results of regressions explaining illiquidity by arbitrageurs' relative speed (π), arbitrage mix (ϕ), and price deviations (σ). Each column presents the results of the following panel regression with two dimensional stock-month fixed effects:

$$Illiq_{i,d} = FE + a_1\pi_{i,d} + a_2\phi_{i,d} + a_3\alpha_{i,d} + a_4\sigma_{i,d} + \boldsymbol{\zeta} \times \boldsymbol{Controls}_{i,d} + \epsilon_{i,d}$$

where illiquidity ($Illiq_{i,d}$) of home- or host-market stock i on day d is measured by the proportional quoted spread (Panel A: $PQSPR$) and proportional effective spread (Panel B: $PESPR$). $\pi_{i,d}$ is the number of toxic price deviations that end with a trade divided by the number of toxic price deviations; $\phi_{i,d}$ is the number of toxic price deviations divided by the number of all price deviations; $\alpha_{i,d}$ is the number of price deviations divided by the number of trades; $\sigma_{i,d}$ is the average deviation in mid-quote prices. And where $\boldsymbol{Controls}$ is a vector of control variables: $Vola$ is the 5-minute mid-return volatility; $Trsize$ is the average trade size (in 10,000 shares); $Quotes$ is the number of quote updates (in 10,000); Ted is the Ted spread, the difference between 3-Month USD LIBOR and Treasury Bills. All regressions are estimated with a linear time-trend (unreported). All stock specific variables are measured during the overlapping trading time, i.e. when both the home- and host-market are in their continuous trading session. Standard errors are clustered by stock and statistical significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. $p-$values are in parentheses below the coefficients. *Ted* is from the Federal Reserve Bank of St. Louis and all other data underlying the computations are from TRTH.

	Panel A: $PQSPR_{i,d}$		Panel B: $PESPR_{i,d}$	
$\pi_{i,d,Toxic}$	-0.008***	-0.007***	-0.003*	-0.003*
	(0.00)	(0.00)	(0.08)	(0.09)
$\pi_{i,d,notToxic}$		-0.006***		0.006
		(0.01)		(0.14)
$\phi_{i,d}$	0.021	0.025	-0.010	-0.001
	(0.17)	(0.15)	(0.53)	(0.96)
$\alpha_{i,d}$	0.006***	0.006***	0.002	0.002
	(0.00)	(0.00)	(0.18)	(0.17)
$\sigma_{i,d,Toxic}$	0.104***	0.082***	0.103***	0.075***
	(0.00)	(0.01)	(0.00)	(0.00)
$\sigma_{i,d}$		0.048**		0.056***
		(0.01)		(0.00)
$Vola_{i,d}$	0.141**	0.140**	0.036	0.034
	(0.02)	(0.02)	(0.18)	(0.18)
$Trsize_{i,d}$	-0.005	0.000	-0.054	-0.049
	(0.77)	(0.99)	(0.14)	(0.18)
$Quotes_{i,d}$	-0.006	-0.005	-0.001	-0.001
	(0.11)	(0.11)	(0.68)	(0.67)
Ted_d	0.024***	0.025***	0.030***	0.030***
	(0.00)	(0.00)	(0.00)	(0.00)
Adj. R^2 [%]	75.49	74.69	41.21	40.41
Obs.	513,837	511,777	512,611	510,615

Table 5 – Illiquidity during days between corporate actions, 2001 - 2016
This table presents results of how liquidity varies on days between corporate actions, i.e., when either the host or the home-market is cum-dividend but the other is ex-dividend ($Between_{i,d}$). Each column presents the results of the following panel regression with two dimensional stock-month fixed effects:

$$Illiq_{i,d} = FE + \beta \times Between_{i,d} + \boldsymbol{\zeta} \times \boldsymbol{Controls}_{i,d} + \epsilon_{i,d}$$

where illiquidity ($Illiq_{i,d}$) of home- or host-market stock i on day d is measured by the proportional quoted spread (Panel A: $PQSPR$), proportional effective spread (Panel B: $PESPR$), and the difference in $PQSPR$ during and outside overlapping trading times (Panel C: $\delta PQSPR$), as defined in Table 1. And where $\boldsymbol{Controls}$ is a vector of control variables: *CorpAct* is a dummy variable set to one on days on which both the home and the host-market go ex-dividend; and the other variables are as defined in Table 4 and OIB is the order imbalance, measured as the difference between the number of buyer and seller initiated trades; and $Trades$ is the number of trades (in 10,000). All stock specific variables (except *price*) are measured during the overlapping trading time, i.e. when both the home- and host-market are in their continuous trading session. In Panel C, a δ before any variable indicates that this variable is constructed as the difference during and overlapping trading times. In Panel C, I drop all stock-pairs if the home-market stock is from Argentina, Brazil, Chile, Mexico, and Peru. Standard errors are clustered by stock and statistical significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. $p-$values are in parentheses below the coefficients. Data to compute *price* are from Datastream, *Ted* is from the Federal Reserve Bank of St. Louis, and all other data underlying the computations are from TRTH.

	Panel A: $PQSPR_{i,d}$		Panel B: $PESPR_{i,d}$		Panel C: $\delta PQSPR_{i,d}$	
$Between_{i,d}$	0.023***	0.022***	0.026**	0.027***	0.018**	0.014***
	(0.00)	(0.01)	(0.01)	(0.01)	(0.02)	(0.00)
$CorpAct_{i,d}$		0.004		0.001		0.001
		(0.49)		(0.89)		(0.69)
$log(price)_{i,d}$		-0.364***		-0.316***		
		(0.00)		(0.00)		
$Vola_{i,d}$		0.190***		0.046*		
		(0.00)		(0.09)		
$OIB_{i,d}$		-0.000		-0.000		
		(0.79)		(0.86)		
$Trades_{i,d}$		-0.021***		0.003		
		(0.00)		(0.49)		
$Quotes_{i,d}$		-0.006		-0.004		
		(0.11)		(0.12)		
$Trsize_{i,d}$		0.034		-0.009		
		(0.37)		(0.79)		
Ted_d		0.039***		0.038***		
		(0.00)		(0.00)		
$\delta Vola_{i,d}$						0.193***
						(0.00)
$\delta OIB_{i,d}$						0.000
						(0.22)
$\delta Trades_{i,d}$						-0.005***
						(0.00)
$\delta Quotes_{i,d}$						-0.001*
						(0.08)
$\delta Trsize_{i,d}$						-0.001
						(0.83)
$\Delta Trades_{i,d}$						0.033
						(0.24)
Adj. R^2 [%]	64.72	68.13	34.95	35.41	28.71	30.08
Obs.	825,228	798,243	813,584	792,003	527,912	525,087

Table 6 – Price deviations during days between corporate actions, 2001 - 2016
This table presents the first stage regressions of instrumenting price deviations when explaining illiquidity by a dummy variable which is one on days between corporate actions ($Between_{i,d}$), i.e., when either the host or the home-market is cum-dividend but the other is ex-dividend. Each column presents the results of the following panel regression with two dimensional stock-month fixed effects:

$$\Delta Price_{i,d} = FE + \beta_0 \times BetweenCorpAct_{i,d} + \boldsymbol{\zeta_0} \times \boldsymbol{Controls}_{i,d} + \epsilon_{i,d}$$

where price deviations ($\Delta Price_{i,d}$) between home- and host-market stock i on day d are measured by the average duration price deviations persist ($INARB_{i,d}$), the average price deviation from simultaneous trade prices across the host and the home-market ($avg(\Delta TRD_{i,d})$) the average ($avg(\Delta QTE_d)$) and maximum ($max(\Delta QTE_{i,d})$) price deviation between simultaneous bid and ask price across the home- and host-market relative to the mid price of the home market, and the average price deviation from quotes unadjusted by the corporate action ($avg(\Delta UQTE_{i,d})$). And where $\boldsymbol{Controls}$ is a vector of control variables. For a description of these variables I refer to Table 1 and Table 5. Standard errors are clustered by stock and statistical significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. $p-$values are in parentheses below the coefficients. Data to compute *price* are from Datastream, *Ted* is from the Federal Reserve Bank of St. Louis, and all other data underlying the computations are from TRTH.

	$INARB_{i,d}$		$avg(\Delta TRD_{i,d})$		$avg(\Delta QTE_{i,d})$		$max(\Delta QTE_{i,d})$		$max(\Delta UQTE_{i,d})$	
$Between_{i,d}$	4.100***	4.279***	1.867***	1.921***	1.089***	1.103***	1.775***	1.800***	5.500***	5.445***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
$log(price)_{i,d}$		-1.465		-1.048***		-0.443**		-1.496***		-1.078
		(0.13)		(0.00)		(0.02)		(0.00)		(0.38)
$Vola_{i,d}$		-0.764*		0.477**		0.054*		0.167*		0.171*
		(0.07)		(0.01)		(0.09)		(0.07)		(0.08)
$OIB_{i,d}$		0.000		0.000		0.000		0.000		0.000**
		(0.31)		(0.73)		(0.13)		(0.18)		(0.01)
$Trades_{i,d}$		-1.693***		0.005		0.038**		0.149***		0.104**
		(0.01)		(0.79)		(0.01)		(0.00)		(0.01)
$Quotes_{i,d}$		-0.858*		-0.018		0.010		0.037*		0.054**
		(0.06)		(0.20)		(0.25)		(0.07)		(0.05)
$Trsize_{i,d}$		-1.056***		-0.023		0.084***		0.097***		0.132***
		(0.01)		(0.64)		(0.00)		(0.00)		(0.00)
Ted_d		-1.188***		-0.113***		-0.032		0.079**		0.030
		(0.00)		(0.00)		(0.20)		(0.05)		(0.51)
Adj. R^2 [%]	36.80	37.24	75.95	76.16	88.69	89.26	87.24	88.09	80.90	81.14
F	31.73	10.32	10.47	7.32	36.54	8.52	46.88	14.43	37.40	20.21
Obs.	637,268	617,477	720,436	703,073	826,224	798,235	826,224	798,235	826,224	798,235

Table 7 – Instrumental variable regressions to address contemporaneous effects of impediments to arbitrage on illiquidity, 2001 - 2016

This table presents the second stage regressions of instrumenting price deviations when explaining illiquidity by a dummy variable which is one on days between corporate actions, i.e., when either the host or the home-market is cum-dividend but the other is ex-dividend. Each column presents the results of the following panel regression with two dimensional stock-month fixed effects:

$$Illiq_{i,d} = FE + \beta_1 \times \widehat{\Delta Price}_{i,d} + \boldsymbol{\zeta_1} \times \boldsymbol{Controls}_{i,d} + \epsilon_{i,d}$$

where illiquidity ($Illiq_{i,d}$) of home- or host-market stock i on day d is measured by the proportional quoted spread (Panel A: $PQSPR$), proportional effective spread (Panel B: $PESPR$), and the difference in PQSPR during and outside overlapping trading times (Panel C: $\delta PQSPR$), as defined in Table 1. $\widehat{\Delta Price_{i,d}}$ is the fitted value from first stage regressions reported in Table 6. And $\boldsymbol{Controls}$ is a vector of control variables. For a description of these variables I refer to Table 1 and Table 5. In Panel C, I drop all stock-pairs if the home-market stock is from Argentina, Brazil, Chile, Mexico, and Peru. Standard errors are clustered by stock and statistical significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. R^2 statistics are not reported, because they cannot be properly interpreted in two stage regressions. $p-$values are in parentheses below the coefficients. Data to compute *price* are from Datastream, *Ted* is from the Federal Reserve Bank of St. Louis, and all other data underlying the computations are from TRTH.

	Panel A: $PQSPR_{i,d}$									
$\widehat{INARB_{i,d}}$	0.006**	0.006**								
	(0.03)	(0.03)								
$\widehat{avg(\Delta TRD_{i,d})}$			0.008**	0.005*						
			(0.04)	(0.09)						
$\widehat{avg(\Delta QTE_{i,d})}$					0.021**	0.021**				
					(0.01)	(0.02)				
$\widehat{max(\Delta QTE_{i,d})}$							0.013**	0.013**		
							(0.01)	(0.02)		
$\widehat{max(\Delta UQTE_{i,d})}$									0.004**	0.004**
									(0.01)	(0.01)
$log(price)_{i,d}$		-0.351***		-0.136***		-0.355***		-0.345***		-0.359***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
$Vola_{i,d}$		0.178***		1.502***		0.188***		0.187***		0.189***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
$OIB_{i,d}$		-0.000		0.000		-0.000		-0.000		-0.000
		(0.49)		(0.15)		(0.69)		(0.70)		(0.73)
$Trades_{i,d}$		-0.012*		-0.068***		-0.022***		-0.023***		-0.022***
		(0.08)		(0.00)		(0.00)		(0.00)		(0.00)
$Quotes_{i,d}$		0.000		-0.042**		-0.006		-0.006*		-0.006
		(0.94)		(0.02)		(0.10)		(0.10)		(0.10)
$Trsize_{i,d}$		0.083		0.018		0.032		0.033		0.033
		(0.16)		(0.53)		(0.40)		(0.39)		(0.38)
Ted_d		0.040***		0.004		0.039***		0.038***		0.039***
		(0.00)		(0.46)		(0.00)		(0.00)		(0.00)

Table 7 continued

Obs.	636,806	617,451	719,962	703,069	825,148	798,163	825,148	798,163	825,148	798,163
	Panel B: $PESPR_{i,d}$									
$\widehat{INARB_{i,d}}$	0.007**	0.007**								
	(0.04)	(0.04)								
$\widehat{avg(\Delta TRD_{i,d})}$			0.015*	0.014*						
			(0.05)	(0.06)						
$\widehat{avg(\Delta QTE_{i,d})}$					0.024**	0.025**				
					(0.02)	(0.02)				
$\widehat{max(\Delta QTE_{i,d})}$							0.015**	0.015**		
							(0.02)	(0.01)		
$\widehat{max(\Delta UQTE_{i,d})}$									0.005**	0.005**
									(0.01)	(0.01)
$log(price)_{i,d}$		-0.293***		-0.154***		-0.306***		-0.293***		-0.311***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
$Vola_{i,d}$		0.042*		0.808***		0.044*		0.043*		0.045*
		(0.08)		(0.00)		(0.09)		(0.09)		(0.09)
$OIB_{i,d}$		-0.000		0.000		-0.000		-0.000		-0.000
		(0.44)		(0.99)		(0.84)		(0.84)		(0.85)
$Trades_{i,d}$		0.015*		-0.025***		0.002		0.000		0.002
		(0.07)		(0.00)		(0.66)		(0.93)		(0.58)
$Quotes_{i,d}$		0.004		-0.024**		-0.004		-0.004*		-0.004
		(0.33)		(0.03)		(0.11)		(0.09)		(0.10)
$Trsize_{i,d}$		0.030		-0.009		-0.014		-0.014		-0.013
		(0.59)		(0.89)		(0.68)		(0.69)		(0.71)
Ted_d		0.043***		0.012**		0.039***		0.037***		0.038***
		(0.00)		(0.02)		(0.00)		(0.00)		(0.00)
Obs.	631,068	614,268	719,894	702,970	813,504	791,923	813,504	791,923	813,504	791,923
	Panel C: $\delta PQSPR_{i,d}$									
$\widehat{INARB_{i,d}}$	0.014*	0.010**								
	(0.06)	(0.03)								
$\widehat{avg(\Delta TRD_{i,d})}$			0.008*	0.005**						
			(0.10)	(0.05)						
$\widehat{avg(\Delta QTE_{i,d})}$					0.012**	0.009**				
					(0.04)	(0.01)				
$\widehat{max(\Delta QTE_{i,d})}$							0.009**	0.007**		
							(0.04)	(0.01)		
$\widehat{max(\Delta UQTE_{i,d})}$									0.005**	0.004***
									(0.02)	(0.00)
$\delta Vola_{i,d}$		0.168***		0.185***		0.195***		0.192***		0.193***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)

Table 7 continued

$\delta OIB_{i,d}$		-0.000		0.000		0.000		0.000		0.000
		(0.87)		(0.15)		(0.19)		(0.20)		(0.23)
$\delta Trades_{i,d}$		-0.004**		-0.005***		-0.005***		-0.005***		-0.005***
		(0.03)		(0.00)		(0.00)		(0.00)		(0.00)
$\delta Quotes_{i,d}$		-0.001		-0.001*		-0.001*		-0.001*		-0.001*
		(0.26)		(0.07)		(0.07)		(0.07)		(0.07)
$\delta Trsize_{i,d}$		-0.012		0.002		-0.001		-0.001		-0.001
		(0.21)		(0.68)		(0.82)		(0.80)		(0.81)
$\Delta Trades_{i,d}$		0.033		0.037		0.033		0.033		0.033
		(0.28)		(0.12)		(0.24)		(0.24)		(0.24)
Obs.	406,864	404,829	497,964	497,459	527,844	525,018	527,844	525,018	527,844	525,018

Table 8 – Instrumental variable regressions to address contemporaneous effects of impediments to arbitrage on the components of the bid-ask spread, 2001 - 2016

This table presents the second stage regressions of instrumenting price deviations when explaining the components of the bid-ask spread by a dummy variable which is one on days between corporate actions, i.e., when either the host or the home-market is cum-dividend but the other is ex-dividend. Following Glosten and Harris (1988), the bid-ask spread is decomposed into its component due to adverse selection and its transitory component, due to, for example, inventory holding costs (Equation 4). Each column presents the results of the following panel regression with two dimensional stock-month fixed effects:

$$Illiq_{i,d} = FE + \beta_1 \times \widehat{\Delta Price}_{i,d} + \boldsymbol{\zeta_1} \times \boldsymbol{Controls}_{i,d} + \epsilon_{i,d}$$

where illiquidity ($Illiq_{i,d}$) is the transitory component (Panels A and C) or the component due to adverse selection risk (Panels B and D) of the bid-ask spread of ADR (Panels A and B) or home-market stock (Panels C and D) i on day d. $\widehat{\Delta Price_{i,d}}$ is the fitted value from (unreported) first stage regressions. And ***Controls*** is a vector of control variables. For a description of these variables I refer to Table 1 and Table 5. Standard errors are clustered by stock and statistical significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. R^2 statistics are not reported, because they cannot be properly interpreted in two stage regressions. $p-$values are in parentheses below the coefficients. Data to compute *price* are from Datastream, *Ted* is from the Federal Reserve Bank of St. Louis, and all other data underlying the computations are from TRTH.

Panel A: Transitory component of the bid-ask spread for ADRs

$\widehat{INARB_{i,d}}$	0.044***	0.045***								
	(0.00)	(0.01)								
$\widehat{avg(\Delta TRD_{i,d})}$			0.062*	0.061*						
			(0.07)	(0.07)						
$\widehat{avg(\Delta QTE_{i,d})}$					0.125**	0.132**				
					(0.02)	(0.03)				
$\widehat{max(\Delta QTE_{i,d})}$							0.078**	0.083**		
							(0.02)	(0.02)		
$\widehat{max(\Delta UQTE_{i,d})}$									0.027**	0.028**
									(0.01)	(0.02)
$log(price)_{i,d}$		0.004***		0.004***		0.004***		0.004***		0.003***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
$Vola_{i,d}$		0.000*		0.005***		0.000		0.000		0.000
		(0.09)		(0.00)		(0.19)		(0.20)		(0.19)
$OIB_{i,d}$		-0.000		0.000*		0.000		0.000		0.000
		(0.54)		(0.10)		(0.51)		(0.52)		(0.42)
$Trades_{i,d}$		0.000**		-0.000*		0.000		-0.000		0.000
		(0.03)		(0.08)		(0.80)		(0.45)		(0.58)
$Quotes_{i,d}$		0.000*		0.000		0.000*		0.000*		0.000*
		(0.10)		(0.16)		(0.08)		(0.08)		(0.08)
$Trsize_{i,d}$		-0.000		-0.000		-0.000		-0.000		-0.000
		(0.86)		(0.56)		(0.67)		(0.63)		(0.75)
Ted_d		0.002***		0.002***		0.002***		0.002***		0.002***

Table 8 continued

		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Obs.	269,418	261,853	312,761	305,687	348,018	337,519	348,018	337,519	348,018	337,519

Panel B: Adverse selection component of the bid-ask spread for ADRs

$\widehat{INARB_{i,d}}$	-0.000	-0.000								
	(0.56)	(0.65)								
$\widehat{avg(\Delta TRD_{i,d})}$			0.000	0.000						
			(0.87)	(0.83)						
$\widehat{avg(\Delta QTE_{i,d})}$					-0.000	-0.000				
					(0.54)	(0.69)				
$\widehat{max(\Delta QTE_{i,d})}$							-0.000	-0.000		
							(0.54)	(0.69)		
$\widehat{max(\Delta UQTE_{i,d})}$									-0.000	-0.000
									(0.55)	(0.69)
$log(price)_{i,d}$		-0.000**		-0.000**		-0.000**		-0.000**		-0.000**
		(0.01)		(0.04)		(0.01)		(0.01)		(0.02)
$Vola_{i,d}$		-0.000		0.000***		-0.000		-0.000		-0.000
		(0.79)		(0.00)		(0.97)		(0.99)		(0.97)
$OIB_{i,d}$		-0.000		-0.000***		-0.000***		-0.000***		-0.000***
		(0.28)		(0.00)		(0.00)		(0.00)		(0.00)
$Trades_{i,d}$		-0.000		-0.000***		-0.000		-0.000		-0.000*
		(0.49)		(0.01)		(0.12)		(0.46)		(0.09)
$Quotes_{i,d}$		-0.000		-0.000		0.000		0.000		0.000
		(0.81)		(0.13)		(0.65)		(0.61)		(0.63)
$Trsize_{i,d}$		-0.000***		-0.000***		-0.000***		-0.000***		-0.000***
		(0.00)		(0.00)		(0.00)		(0.00)		(0.00)
Ted_d		0.000**		0.000**		0.000**		0.000**		0.000**
		(0.03)		(0.04)		(0.02)		(0.02)		(0.02)
Obs.	266,682	259,213	311,673	304,621	344,794	334,428	344,794	334,428	344,794	334,428

Panel C: Transitory component of the bid-ask spread for home-market stocks

$\widehat{INARB_{i,d}}$	1.180	1.156								
	(0.24)	(0.25)								
$\widehat{avg(\Delta TRD_{i,d})}$			-0.178	-0.236						
			(0.85)	(0.80)						
$\widehat{avg(\Delta QTE_{i,d})}$					1.874	1.858				
					(0.48)	(0.49)				
$\widehat{max(\Delta QTE_{i,d})}$							1.217	1.208		
							(0.47)	(0.49)		
$\widehat{max(\Delta UQTE_{i,d})}$									0.391	0.403
									(0.47)	(0.49)
$log(price)_{i,d}$		6.964*		3.009		3.478**		4.356*		4.676

Table 8 continued

		(0.07)		(0.11)		(0.02)		(0.08)		(0.11)
$Vola_{i,d}$		27.268		2.489		-2.274		-5.980		-1.519
		(0.21)		(0.23)		(0.70)		(0.59)		(0.75)
$OIB_{i,d}$		0.000		0.000		0.000		0.000		0.000
		(0.79)		(0.22)		(0.22)		(0.24)		(0.22)
$Trades_{i,d}$		6.335*		3.692**		3.817**		3.775**		3.803**
		(0.06)		(0.03)		(0.02)		(0.02)		(0.02)
$Quotes_{i,d}$		-7.974*		-8.112**		-7.861**		-7.093*		-7.807**
		(0.09)		(0.03)		(0.04)		(0.08)		(0.04)
$Trsize_{i,d}$		0.338		-0.171		-0.078		-0.017		0.024
		(0.68)		(0.69)		(0.73)		(0.95)		(0.93)
Ted_d		1.664**		0.859**		0.960**		0.847**		0.869**
		(0.05)		(0.03)		(0.02)		(0.01)		(0.01)
Obs.	309,225	300,590	358,809	349,318	395,910	385,349	395,910	385,349	395,910	385,349

Panel D: Adverse selection component of the bid-ask spread for home-market stocks

$\widehat{INARB_{i,d}}$	0.000	0.000								
	(0.85)	(0.76)								
$\widehat{avg(\Delta TRD_{i,d})}$			-0.000	-0.000						
			(0.41)	(0.51)						
$\widehat{avg(\Delta QTE_{i,d})}$					-0.000	-0.000				
					(0.38)	(0.51)				
$\widehat{max(\Delta QTE_{i,d})}$							-0.000	-0.000		
							(0.38)	(0.51)		
$\widehat{max(\Delta UQTE_{i,d})}$									-0.000	-0.000
									(0.39)	(0.52)
$log(price)_{i,d}$		-0.000		-0.000		-0.000		-0.000		-0.000
		(0.35)		(0.39)		(0.32)		(0.28)		(0.27)
$Vola_{i,d}$		0.000		0.000*		0.000**		0.000*		0.000**
		(0.42)		(0.06)		(0.02)		(0.08)		(0.02)
$OIB_{i,d}$		0.000		0.000		0.000		0.000		0.000
		(0.34)		(0.29)		(0.29)		(0.28)		(0.29)
$Trades_{i,d}$		0.000		0.000		0.000		0.000		0.000
		(0.45)		(0.44)		(0.38)		(0.38)		(0.38)
$Quotes_{i,d}$		-0.000		-0.000		-0.000		-0.000		-0.000
		(0.21)		(0.13)		(0.14)		(0.14)		(0.14)
$Trsize_{i,d}$		-0.000***		-0.000***		-0.000***		-0.000***		-0.000***
		(0.00)		(0.00)		(0.01)		(0.00)		(0.00)
Ted_d		-0.000		-0.000		-0.000		-0.000		-0.000
		(0.86)		(0.31)		(0.37)		(0.49)		(0.45)
Obs.	308,824	300,205	358,656	349,173	395,417	384,878	395,417	384,878	395,417	384,878

Table 9 – Vector autoregressions and Granger causality to investigate long term relation between impediments to arbitrage and illiquidity, 2001 - 2016

This table reports results from three panel vector autoregressions. Panel vector autoregressions are estimated using day and stock fixed effects and price deviations, order imbalance (OIB), 5-minute mid-return volatility ($Vola$), and illiquidity ($Illiq$) as endogeneous variables with a lag-length of 15-lags (based on the Schwarz information criterion).

$$LHS_{i,d} = FE + \sum_{l=1}^{x} \beta_l \Delta Price_{i,d-l} + \sum_{l=1}^{x} \gamma_l OIB_{i,d-l} + \sum_{l=1}^{x} \delta_l Illiq_{i,d-l} + \epsilon_{i,d}$$

where illiquidity ($Illiq_{i,d}$) is measured by the proportional quoted spread (Panel A: $PQSPR$), proportional effective spread (Panel B: $PESPR$), and the difference in $PQSPR$ during and outside overlapping trading times (Panel C: $\delta PQSPR$). All variables are measured during the overlapping trading time, i.e. when both the home market and the cross-listed market are in their continuous trading session. For a detailed description of these variables I refer to Table 1. In Panel C, I drop all stock-pairs if the home-market stock is from Argentina, Brazil, Chile, Mexico, and Peru. In all cases the table reports the sum of all lagged variables for each endogeneous variable. $p-$values are in parentheses below the coefficients, based on a χ^2 test for testing the null hypothesis that the sum of the lagged coefficients (in each row) is equal to zero when explaining the column variable. Standard errors are clustered by stock and statistical significance at the 1%, 5%, and 10% level is indicated by ***, **, and *, respectively. The next row reports the p-value of testing joined significance of the lagged values. All data underlying the computations are from TRTH.

Panel A: $PQSPR$				
	$max(\Delta QTE)_{i,d}$	$OIB_{i,d}$	$Vola_{i,d}$	$Illiq_{i,d}$
$\sum_{l=1}^{15} max(\Delta QTE)_{i,d-l}$	0.958***	0.518*	0.001***	0.002***
	(0.00)	(0.08)	(0.00)	(0.00)
	(0.00)	(0.06)	(0.00)	(0.00)
$\sum_{l=1}^{15} OIB_{i,d-l}$	0.000	0.555***	0.000***	0.000
	(0.13)	(0.00)	(0.00)	(0.72)
	(0.01)	(0.00)	(0.00)	(0.60)
$\sum_{l=1}^{15} Vola_{i,d-l}$	0.002	229.308***	0.592***	-0.153***
	(0.99)	(0.00)	(0.00)	(0.00)
	(0.00)	(0.00)	(0.00)	(0.00)
$\sum_{l=1}^{15} PQSPR_{i,d-l}$	0.054**	-16.886***	0.001	0.881***
	(0.02)	(0.00)	(0.76)	(0.00)
	(0.00)	(0.10)	(0.00)	(0.00)
$StockFE$	Yes	Yes	Yes	Yes
$DayFE$	Yes	Yes	Yes	Yes
Observations	476,908	476,908	476,908	476,908

Panel B: $PESPR$				
$\sum_{l=1}^{15} max(\Delta QTE)_{i,d-l}$	0.959***	-0.206	0.001***	0.006***
	(0.00)	(0.56)	(0.00)	(0.00)
	(0.00)	(0.18)	(0.00)	(0.00)
$\sum_{l=1}^{15} OIB_{i,d-l}$	0.000	0.556***	0.000**	0.000

Table 9 continued

	$max(\Delta QTE)_{i,d}$	$OIB_{i,d}$	$Vola_{i,d}$	$Illiq_{i,d}$
	(0.17)	(0.00)	(0.01)	(0.49)
	(0.00)	(0.00)	(0.00)	(0.08)
$\sum_{l=1}^{15} Vola_{i,d-l}$	-0.125	269.584***	0.733***	0.111
	(0.39)	(0.00)	(0.00)	(0.23)
	(0.00)	(0.01)	(0.00)	(0.00)
$\sum_{l=1}^{15} PESPR_{i,d-l}$	0.044**	16.598	0.000	0.706***
	(0.02)	(0.27)	(0.62)	(0.00)
	(0.46)	(0.09)	(0.02)	(0.00)
$StockFE$	Yes	Yes	Yes	Yes
$DayFE$	Yes	Yes	Yes	Yes
Observations	461,986	461,986	461,986	461,986
Panel C: $\delta PQSPR$				
$\sum_{l=1}^{15} max(\Delta QTE)_{i,w-l}$	0.901***	1.736	0.000**	0.001**
	(0.00)	(0.10)	(0.03)	(0.03)
	(0.00)	(0.02)	(0.00)	(0.28)
$\sum_{l=1}^{15} \delta OIB_{i,w-l}$	0.000	0.479***	0.000	0.000
	(0.70)	(0.00)	(0.32)	(0.29)
	(0.01)	(0.00)	(0.56)	(0.46)
$\sum_{l=1}^{15} \delta Vola_{i,w-l}$	-0.424*	30.481	0.771***	-0.087***
	(0.07)	(0.72)	(0.00)	(0.00)
	(0.00)	(0.37)	(0.00)	(0.00)
$\sum_{l=1}^{15} \delta PQSPR_{i,w-l}$	-0.006	-13.751	-0.008***	0.694***
	(0.85)	(0.23)	(0.00)	(0.00)
	(0.00)	(0.69)	(0.00)	(0.00)
$StockFE$	Yes	Yes	Yes	Yes
$DayFE$	Yes	Yes	Yes	Yes
Observations	379,213	379,213	379,213	379,213

Figure 1 – Impulse response functions from price deviations and quoted spreads, 2001- 2016
This figure shows impulse response functions (IRF) from panel vector autoregression (VAR) estimated as in Panel A of Table 9. For a description of the VAR and these variables I refer to Table 1 and Table 9. All IRF in the first row show responses to a Cholesky one standard-deviation shock to price deviations on day 0 (the contemporaneous effect) to day 15, with the first column showing responses on itself (price deviations), the second on order imbalance, the third on volatility, and the last column on quoted spreads. Each figure shows bootstrapped 95% confidence bands based on 1000 runs (lower, upper). All data underlying the computations are from TRTH.

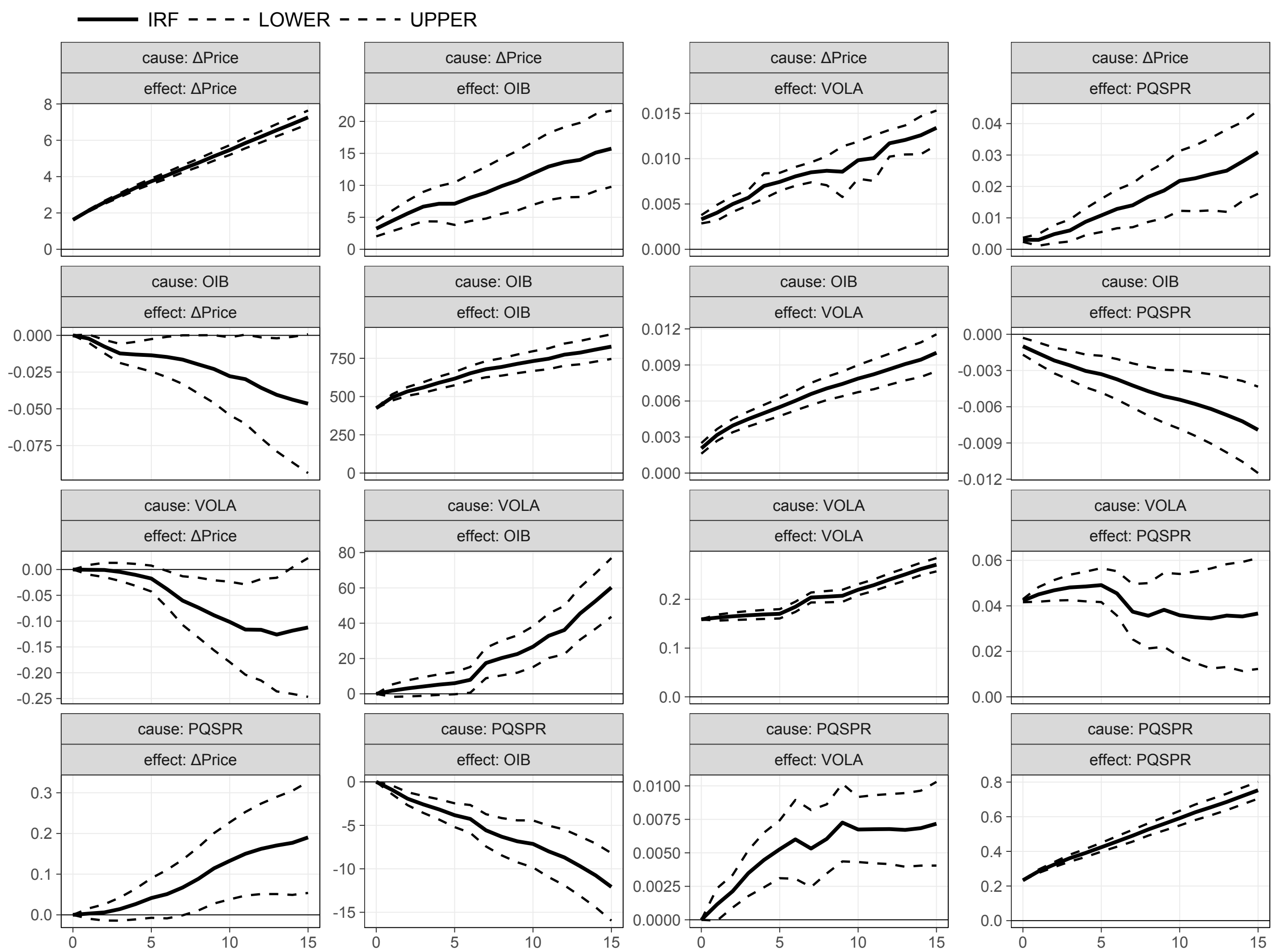

Figure 2 – Impulse response functions from price deviations and effective spreads, 2001- 2016
This figure shows impulse response functions (IRF) from panel vector autoregression (VAR) estimated as in Panel B of Table 9. For a description of the VAR and these variables I refer to Table 1 and Table 9. All IRF in the first row show responses to a Cholesky one standard-deviation shock to price deviations on day 0 (the contemporaneous effect) to day 15, with the first column showing responses on itself (price deviations), the second on order imbalance, the third on volatility, and the last column on effective spreads. Each figure shows bootstrapped 95% confidence bands based on 1000 runs (lower, upper). All data underlying the computations are from TRTH.

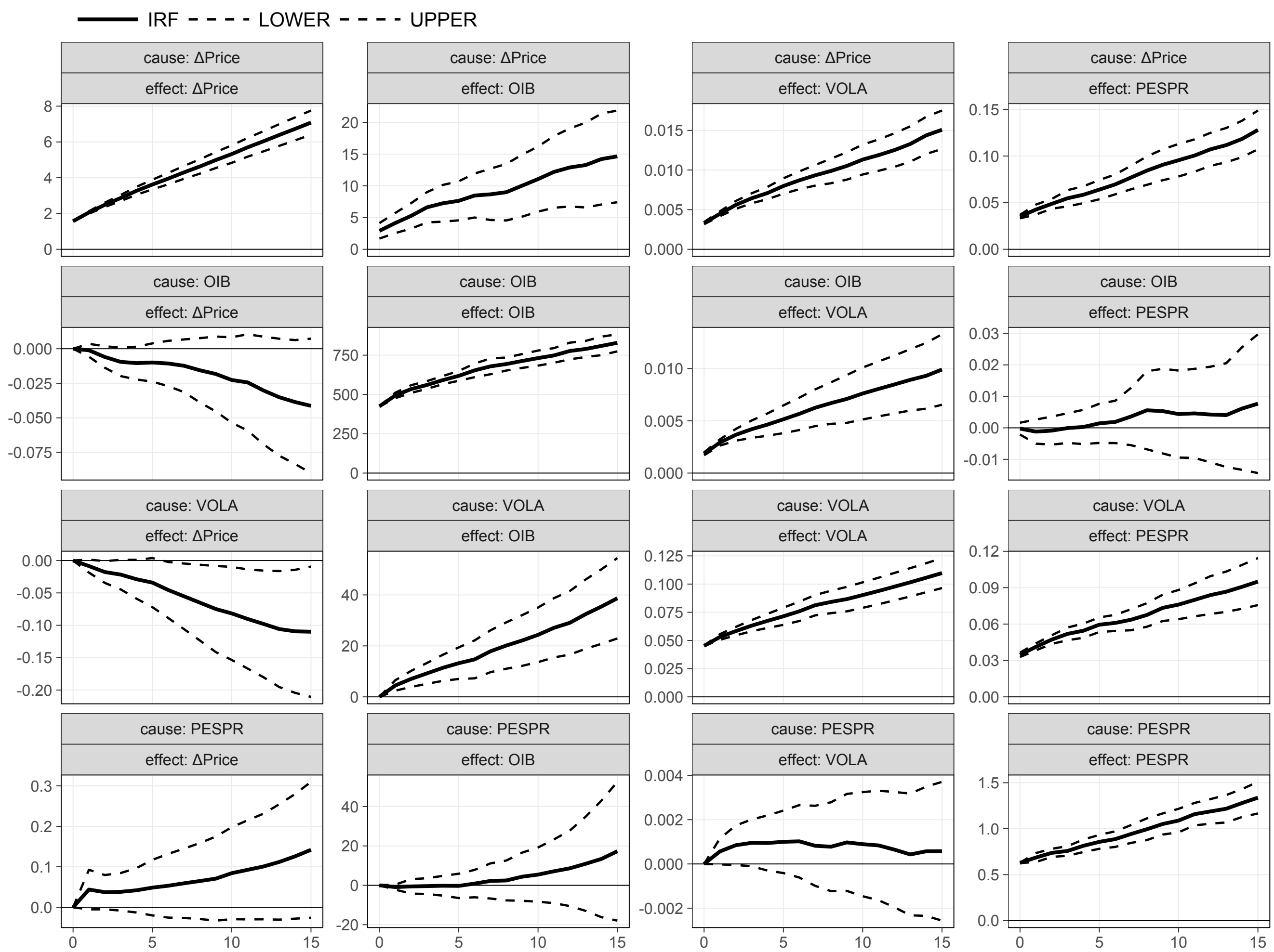

Figure 3 – Impulse response functions from price deviations and differences in quoted spread during and outside overlapping trading times, 2001- 2016

This figure shows impulse response functions (IRF) from panel vector autoregression (VAR) estimated as in Panel C of Table 9. For a description of the VAR and these variables I refer to Table 1 and Table 9. All IRF in the first row show responses to a Cholesky one standard-deviation shock to price deviations on day 0 (the contemporaneous effect) to day 15, with the first column showing responses on itself (price deviations), the second, third, and fourth on differences in order imbalance, volatility, and quoted spread during and outside overlapping trading times, respectively. Each figure shows bootstrapped 95% confidence bands based on 1000 runs (lower, upper). All data underlying the computations are from TRTH.