Journal of Banking & Finance 59 (2015) 457–468

Contents lists available at ScienceDirect

Journal of Banking & Finance

journal homepage: www.elsevier.com/locate/jbf

The determinants of price discovery: Evidence from US-Canadian cross-listed shares

Bart Frijns *, Aaron Gilbert, Alireza Tourani-Rad

Department of Finance, Auckland University of Technology, Auckland, New Zealand

ARTICLE INFO

Article history:
Received 23 October 2014
Accepted 18 July 2015
Available online 30 July 2015

JEL classification:
C23
G10

Keywords:
Price discovery
Determinants
Market microstructure
United States
Canada

ABSTRACT

We examine the determinants of price discovery for Canadian firms cross-listed on the main US stock exchanges over the period 1996–2011. Sampling at a one-minute frequency, we compute Gonzalo and Granger Component Shares (CS) and employ a system GMM approach to control for persistence in price discovery and endogeneity between CS and its determinants. We find that price discovery is highly persistent and that there is strong evidence of simultaneity between CS and its determinants. We conclude that lower relative spreads and higher relative trading activity increase an exchange's contribution to price discovery. We also document that it is small trades that drive price discovery, particularly since the introduction of decimalization.

1. Introduction

"*Markets have two important functions – liquidity and price discovery*..." (O'Hara, 2003, p. 1335).

Price discovery studies the process by which new information is incorporated into asset prices, and, as the above excerpt from Maureen O'Hara's presidential address highlights, is a central function of a market. In the case of cross-listed assets, price discovery plays a specific role as information can be incorporated into asset prices in different markets. In such a case, price discovery reflects relative informational efficiency and indicates in which market investors prefer to trade.

It has been argued that price discovery should predominantly occur in the home market, as this is the market in which most information about the company is generated (Bacidore and Sofianos, 2002). However, such an argument assumes some degree of market segmentation, where investors cannot easily exploit their information in any other market but their own. In contrast, investors in integrated markets have the ability to choose the cheapest and most liquid trading venue. By trading in the cheaper market, information would first be impounded into that market, leading to an increase in the relative contribution to price discovery of that market. As a consequence, it is likely that the cheaper and more liquid market (i.e. the market that provides better quality) will become more important from a price discovery point of view and suggests that price discovery could be driven by aspects of market quality.[1]

Concurrently, the degree of price discovery could also affect measures of market quality, i.e. an improvement in price discovery may improve aspects of market quality for firms, as price discovery leadership may attract liquidity traders to a market. This, in turn, may increase the liquidity and trading volume of the asset, and lower the spread. Hence, the relation between measures of market quality and price discovery is potentially endogenous.

While the level of price discovery of assets cross-listed in multiple markets has been studied extensively, there are few studies that investigate the drivers and determinants of price discovery.

* Corresponding author at: Department of Finance, Auckland University of Technology, Private Bag 92006, 1142 Auckland, New Zealand. Tel.: +64 9 921 9999x5706; fax: +64 9 921 9940.

E-mail address: bfrijns@aut.ac.nz (B. Frijns).

[1] The location of price discovery is also a concern for exchanges themselves. This importance is, for example, emphasized by the Toronto Stock Exchange Board of Governors "*The TSE cannot afford to have the U.S. markets become the price discovery mechanism for Canadian inter-listed stocks*" (Eun and Sabherwal, 2003 p. 550). It appears that exchanges view price discovery as a measure of the relative efficiency and competitiveness between markets, and as such affects trading activity and liquidity.

http://dx.doi.org/10.1016/j.jbankfin.2015.07.011
0378-4266/

Related to the present study are Eun and Sabherwal (2003) and Frijns et al. (2010). Eun and Sabherwal (2003) use a sample of Canadian/US cross-listed firms to analyze the effects of various explanatory variables, such as, relative trading share of a market, bid-ask spreads on levels of price discovery. Frijns et al. (2010) study Australian-New Zealand cross-listed firms, and use panel regressions to examine the relation between price discovery and various explanatory variables. While both studies provide important information on the relation between price discovery and various market characteristics, they do not address the potential endogeneity issue.

In this paper, we contribute to the literature by addressing the endogeneity issue in price discovery. For a sample of Canadian firms cross-listed on the US exchanges between 1996 and 2011, we use intraday data sampled at a one minute frequency, and compute annual measures of price discovery. We then construct a panel data set and employ a dynamic panel data procedure (system GMM) originally developed by Arellano and Bond (1991), Arellano and Bover (1995), and Blundell and Bond (1998), and recently implemented in the finance literature by Hoechle et al. (2012), and Wintoki et al. (2012), to infer the causal effects of various measures of market quality on price discovery. No other study, to our knowledge, has employed such a long period in measuring price discovery, or addressed the endogeneity issue, allowing us to infer the causal effect of market quality on price discovery.[2]

Our study makes several important contributions. First, our dynamic panel model demonstrates that there is strong persistence in price discovery; a property that has not been documented before. Second, we confirm that there is indeed an endogenous relation between the various measures of market liquidity and price discovery. Our results suggest that lagged price discovery affects various measures of market quality, such as quoting activity and spreads. Furthermore, we document that measures of trading activity (relative volume and relative number of trades between the US and Canada) have a positive causal effect on price discovery. In addition, the relative difference in spreads between markets has a negative causal effect – i.e. a decrease in spread in one market relative to the other market leads to an increase in price discovery in that market. We further find that it is the relative proportion of small trades, fewer than 2500 shares, which lead to an increase in price discovery in a market. Our findings have important implications for exchanges, as these findings indicate ways in which exchanges can improve their price discovery and market quality.

This paper proceeds as follows. In Section 2, we discuss the literature regarding price discovery for assets traded across multiple venues. In Section 3, we briefly discuss the model used to assess price discovery. In Section 4, we provide information on the data used in this study. Section 5 reports the results of our analysis on the determinants of price discovery. Finally, Section 6 concludes.

2. Literature review

There has been a growing literature examining price discovery of stocks listed on multiple exchanges. Originally, papers looked at the relative importance of exchanges within the US, particularly between the NYSE and regional exchanges (Harris et al., 1995; Hasbrouck, 1995). More recently, studies have considered the location of price discovery for cross-listed securities. Despite the assertions of Bacidore and Sofianos (2002) that the home market should dominate in terms of price discovery, the empirical evidence is mixed. Studies looking at Israeli (Lieberman et al., 1999), Mexican (von Furstenberg and Tatora, 2004), Hong Kong (Su and Chong, 2007) and Chinese (Chen et al., 2010) firms cross-listed on the NYSE, Malaysian firms listed in Singapore (Ding et al., 1999), Hong Kong firms in London (Agarwal et al., 2007) and New Zealand and Australian bi-directional cross-listings (Frijns et al., 2010) all find that the foreign market to some extent has an informational role. In contrast, Pascual et al. (2006), and Grammig et al. (2005), who study Spanish and German firms cross-listed on the NYSE, and Alhaj-Yaseen et al. (2014), who study Israeli firms first listed on the NYSE and then subsequently cross-listing back on to the Tel Aviv stock market, find no informational role for the foreign market. Kadapakkam et al. (2003) find price discovery equally split between the London and the Bombay Stock Exchange for cross-listed Indian companies. Hupperets and Menkveld (2002), examining Dutch firms cross-listed on the NYSE, and Eun and Sabherwal (2003), considering Canadian firms cross-listed on the NYSE, find that the location of price discovery varies considerably across firms.

While the majority of extant studies have considered where price discovery occurs for cross-listed stocks, very few have looked at what determines the location of price discovery or how it evolves over time. Harris et al. (2002) relate changes in price discovery to changes in the relative transaction costs between the NYSE and regional exchanges in the US at three discrete points in time. They conclude that higher NYSE spreads reduce the NYSE share of price discovery. Frijns et al. (2010) examine price discovery over time, from 2002 to 2005, using bi-directional cross-listings between the New Zealand Stock Exchange and Australian Stock Exchange. They find clear evidence of an upward trend in the contribution to price discovery of the Australian Stock Exchange, not only for Australian firms but also for New Zealand companies. Using panel regressions they further show that increases in the importance of the Australian market is positively related to the growth in firm size and negatively related to the size of the percentage spread in the Australian market, implying that as firms grow larger and their cost of trading in Australia declines, the Australian market becomes more informative.

Eun and Sabherwal (2003) investigate the contribution of the US stock exchanges (NYSE, AMEX and NASDAQ) to the price discovery of 62 Canadian firms cross-listed on US exchanges, for a six-month period in 1998. They find that while both the US exchange and TSX contribute to price discovery, for most stocks, the US prices adjust more to TSX prices than the other way around. Using cross-sectional regressions to analyze the determinants of price discovery, they find a negative relation with relative spreads, but also find positive relations with the US share of trading and the US share of informative trades.

In a more recent study, Chen and Choi (2012), by incorporating relative measures of informed trades (PIN), observe that the TSX leads in price discovery and also shows a higher PIN than the NYSE. In other words, the trading venue with the heavier intensity of informed trades contributes more to the price discovery of cross-listed pairs.

Frijns et al. (2015) examine the impact of macroeconomic news announcement on price discovery for a sample of Canadian firms cross-listed on the NYSE over the period 2004 to 2011. They demonstrate that there are significant shifts in price discovery on days with macroeconomic news announcement, where for most stocks and most announcements price discovery on the NYSE increases on days with macroeconomic news announcements. They attribute this increase in the NYSE share of price discovery to better information processing capacity of the US market.

[2] There are several other studies that use intraday data to estimate microstructure measures over long periods of time. Specifically, Easley et al. (2002) use high-frequency data of all common stocks listed on the NYSE over the period 1983–1998 to estimate the probability of informed trading (PIN) for each stock and for each year. They continue to use these yearly estimates in subsequent asset pricing tests. These measures have subsequently been used by Easley et al. (2010) who estimate yearly PINs from 1983–2001, and Aslan et al. (2011) who use yearly PIN estimates for the period 1983–1999.

The literature review demonstrates that there is a vast literature on price discovery, but that this literature predominantly focuses on where price discovery takes place. To date, there have been limited studies on determinants of price discovery. Those studies that have focused on the determinants of price discovery have not considered the potential endogenous relation between these determinants and price discovery. In this paper, we address this issue.

3. Measuring price discovery

To study the informational role of the US and Canadian markets for US-Canadian cross-listed stocks, we investigate the contribution of both markets to price discovery. We follow the standard approach of estimating a vector error correction model (VECM), and computing our price discovery measure from the estimates of the VECM.

Consider a single security that lists in two different markets (in this case, the US and Canada). Let p_{jt}^{US} be the log US dollar price of security j traded in the US market, and let p_{jt}^{CAN} be the log US dollar price of the asset traded on the Toronto Stock Exchange (TSX). If the two assets are identical and completely fungible, then arbitrage implies that the price difference $(p_{jt}^{US} - p_{jt}^{CAN})$ is bounded with probability 1. Stated differently, if the prices, $p_{jt} = (p_{jt}^{US} \quad p_{jt}^{CAN})'$, in the US and Canadian market are for the same asset, then prices should be cointegrated with cointegrating vector $\beta' = (1 \ -1)$. Cointegration in the prices of these assets implies that price changes can be expressed as an error correction model of the form,

$$\Delta p_{jt} = c_j + \alpha_j \beta' p_{jt} + \sum_{i=1}^{l} \Gamma_{ij} \Delta p_{jt-i} + \varepsilon_{jt}, \tag{1}$$

where α_j is the (2×1) vector containing the speed of adjustment coefficients for US prices and Canadian prices and Γ_{ij} are (2×2) matrices containing coefficients on lagged prices. Note that the specification of the cointegrating vector β, implies that we expect the first element of α, $\alpha^{US} \leqslant 0$ and the second element of α, $\alpha^{CAN} \geqslant 0$.

We obtain our price discovery measure from the VECM stated in Eq. (1) by following the permanent-transitory decomposition of Gonzalo and Granger (1995). This measure, called the Component Share (CS), is used by Eun and Sabherwal (2003) in their study on price discovery among US-Canadian cross-listed stocks and compares the speed of adjustment coefficients of the two markets. The lower the speed of adjustment coefficient, the more informative that market is. For example, if the US is completely dominant and the TSX is a pure satellite market, then we expect $\alpha^{US} = 0$ and $\alpha^{CAN} > 0$. Vice versa, if the Canadian market is completely dominant and the US market is a pure satellite, then we expect $|\alpha^{US}| > 0$ and $\alpha^{CAN} = 0$. If neither market is completely dominant $|\alpha^{US}|$ and α^{CAN} will both be positive, but their relative magnitudes will give us an indication of the degree of dominance over the other market. Following Eun and Sabherwal (2003), we compute the Component Share as

$$CS_j^{CAN} = \frac{|\alpha_j^{US}|}{|\alpha_j^{US}| + \alpha_j^{CAN}}, \tag{2}$$

where CS_j^{CAN} is the Component Share for a security on the TSX. Likewise,

$$CS_j^{US} = 1 - CS_j^{CAN}. \tag{3}$$

As an alternative, we could use the Information Shares (IS) proposed by Hasbrouck (1995), which performs a decomposition of the variance of the underlying price process and measures the contribution of each market to this variance. However, using the IS as a measure of price discovery is problematic, when liquidity changes over time. As prices are typically contemporaneously correlated the IS is not unique, and, as shown by Baillie et al. (2002), the contemporaneous correlation increases when the sampling frequency decreases. Booth et al. (2002) suggest that the midpoint of the IS range provides a good measure for price discovery. In general, we observe that over our sample period, liquidity has increased substantially. This increases the contemporaneous correlation and widens the range of the IS. Since the IS is bounded between 0 and 1, an increase in the contemporaneous correlation biases the midpoint of the IS towards 0.5 for both markets (see for example Putnins, 2013, for a numerical example on the impact of sampling on the IS measure). This problem also affects the Information Leadership Share developed by Yan and Zivot (2010) and Putnins (2013), and motivates why we do not use their measure. As shown by both Yan and Zivot (2010) and Putnins (2013), the CS is least affected by time aggregation.

4. Data and descriptive statistics

4.1. Data

We obtain intraday data for Canadian TSX-listed firms that are (or have been) cross-listed on one of the three main US markets: NYSE, NASDAQ and NYSE Amex Equities[3] (hereafter AMEX).[4] These data are obtained from the Thompson Reuter Tick History database (TRTH) maintained by SIRCA[5] for the period 1996–2011, a sample period of 16 years. The data contain all trades and quotes (plus associated volumes) time stamped to the nearest millisecond. Following Eun and Sabherwal (2003) and Grammig et al. (2005), we use the quote midpoints to study price discovery. These midpoints are free from the bid-ask bounce that is typically observed in transaction prices.[6] Subsequently, we sample our data at a one minute frequency and convert all prices to US dollars, using the intraday exchange rate obtained from TRTH.[7]

Table 1 provides an overview of the number of cross-listings per year and per exchange. In 1996, we observe just 48 cross-listings. However, by 2011 the number of cross-listings has increased to 142. There is a period of relatively rapid growth in the number of cross-listings between 1996 and 2007, but from 2007 onwards the number of cross-listings remains relatively constant. In percentage terms, the growth in the number of cross-listings is also notable. In 1996, only 3.2% of TSX firms were cross-listed in the US. By 2003, this is up to 8.2%, and remains around 8% to 9% for the remainder of the sample period.

The pattern of increases in cross-listings has not, however, been uniform across the three different exchanges. The NYSE observes more than a doubling in the number of cross-listings between 1996 and 2002. After 2002, the growth in the number of cross-listings is modest. Firms cross-listing on the NASDAQ observe a rapid growth from 1996 to 2001 up to 43 firms, after

[3] Formerly the AMEX exchange.

[4] Examining Canadian stocks cross-listed in the US offer several advantages. For instance, the trading hours for the TSX and the US markets under investigation fully overlap, with regular trading hours in both markets from 9:30 AM to 4:00 PM (EST). Also, Canadian shares are listed as ordinary shares in the US markets, making them fully fungible.

[5] Securities Industry Research Centre of Asia–Pacific.

[6] In some cases we observe quotes that have a price of zero, when there is no liquidity on one side of the market. We remove those quotes from our sample.

[7] Eun and Sabherwal (2003) sample at a frequency of 10 min. We sample at a higher frequency as liquidity (in terms of number of trades and quotes) has increased substantially over our sample period. Sampling at a lower 10-min frequency would possibly obscure any informational asymmetries that can be observed at higher frequencies.

Table 1
Number of cross-listings per year.

Year	Listed companies TSX group	Total	NYSE listings	NASDAQ listings	AMEX listings
1996	1495	48	20	17	11
1997	1429	64	30	21	13
1998	1433	70	33	25	12
1999	1456	80	37	31	12
2000	1394	93	40	40	13
2001	1299	93	41	43	9
2002	1287	97	46	38	13
2003	1340	107	47	43	17
2004	1410	116	48	40	28
2005	1537	125	50	43	32
2006	1598	129	51	39	39
2007	1613	138	52	35	51
2008	1570	137	49	36	52
2009	1468	138	51	35	52
2010	1517	123	50	26	47
2011	1588	142	57	34	51

Note: This table reports the number of cross-listings per year from the TSX to the various markets in the US. We report the total number of cross-listings and cross-listings on the NYSE, NASDAQ and AMEX, respectively.

which it declines to 26 firms in 2010. Cross-listings on the AMEX remain stable until 2002, but they increase rapidly from 13 to 51 cross-listings starting in 2007.

4.2. Price discovery

In this section, we compute the Gonzalo and Granger (1995) Component Shares. Before estimating these Component Shares, we take several steps to filter the data. First, we clean the data of any obvious recording errors. Second, we exclude those cross-listings that have less than 40 days of quoting activity in both markets in a given year. This filter is applied to ensure that price discovery measures are representative for that given year. Third, we perform augmented Dickey-Fuller (ADF) unit root tests for each stock and for each year. The concept of cointegration only becomes relevant when the time series of stock prices in both markets are non-stationary. Non-stationarity of the time series is therefore a requirement for our cointegration analysis and we exclude any stocks for which the ADF test rejects the presence of a unit root at the 5% level (only in about 6% of the firm-year observations do we reject the presence of a unit root). Panel A of Table 2 presents summary statistics of the ADF tests for the US and Canadian price series considering the full sample and the exchanges the stocks are cross-listed in. Overall, we observe that unit roots are present in most of the price series. Fourth, we perform Johansen's test for

Table 3
Canadian Component Shares by exchange and country.

	Full (%)	NYSE (%)	NASDAQ (%)	AMEX (%)
Panel A: Full sample Component Shares by exchange				
Average	57.93	64.16	44.88	64.01
Std. Dev.	22.21	21.97	21.04	16.44
5th Percentile	18.19	22.99	10.99	35.67
25th Percentile	42.35	47.61	28.95	53.97
Median	59.46	66.88	43.52	64.66
75th Percentile	75.40	82.45	60.91	76.42
95th Percentile	91.23	94.26	79.93	88.88

Note: This Table reports summary statistics for the Canadian Component Shares (CS^{CAN}). We report the average and percentile values for the Component Share for the full sample and per exchange.

Table 2
Error correction summary statistics.

	Full sample		NYSE firms		NASDAQ firms		AMEX firm	
Panel A: Augmented Dicky-Fuller values								
	Canada	US	Canada	US	Canada	US	Canada	US
5th Percentile	−2.7907	−2.8016	−2.5947	−2.5940	−3.0061	−2.8984	−2.8427	−2.8824
25th Percentile	−2.0078	−2.0007	−1.8390	−1.8510	−2.0362	−2.0098	−2.1801	−2.2371
Median	−1.4369	−1.4300	−1.3206	−1.3389	−1.4576	−1.4386	−1.5709	−1.5795
75th Percentile	−0.8163	−0.8194	−0.7268	−0.7247	−0.8995	−0.8698	−0.8432	−0.9039
95th Percentile	0.1824	0.1942	0.1858	0.2413	0.1196	0.1583	0.1183	0.1864
Panel B: Likelihood ratio statistics								
	LR 1	LR 2	LR 1	LR 2	LR 1	LR 2	LR 1	LR 2
5th Percentile	198.12	0.01	247.75	0.01	126.69	0.01	204.77	0.03
25th Percentile	461.82	0.33	631.10	0.27	424.75	0.29	422.48	0.53
Median	1164.05	1.34	1994.60	1.05	853.28	1.47	922.90	1.69
75th Percentile	2363.40	3.32	2933.20	2.75	1673.80	3.34	1750.70	3.98
95th Percentile	6380.80	8.13	8532.80	6.93	5613.10	9.31	3224.00	8.74
Panel C: Speed of adjustment coefficients								
	α_{CAN}	α_{US}	α_{CAN}	α_{US}	α_{CAN}	α_{US}	α_{CAN}	α_{US}
5th Percentile	−0.1332	0.0013	−0.1448	0.0008	−0.1692	0.0018	−0.0529	0.0027
25th Percentile	−0.0459	0.0083	−0.0702	0.0165	−0.0438	0.0065	−0.0238	0.0081
Median	−0.0183	0.0270	−0.0286	0.0634	−0.0177	0.0153	−0.0121	0.0214
75th Percentile	−0.0080	0.0745	−0.0091	0.1088	−0.0086	0.0331	−0.0067	0.0603
95th Percentile	−0.0031	0.1781	−0.0042	0.2459	−0.0029	0.0808	−0.0026	0.1303
Panel D: Cointegrating vector								
	β_{CAN}	β_{US}	β_{CAN}	β_{US}	β_{CAN}	β_{US}	β_{CAN}	β_{US}
5th Percentile	1	−1.0059	1	−1.0023	1	−1.0064	1	−1.0115
25th Percentile	1	−1.0006	1	−1.0004	1	−1.0005	1	−1.0016
Median	1	−0.9999	1	−1.0000	1	−0.9997	1	−1.0001
75th Percentile	1	−0.9990	1	−0.9995	1	−0.9979	1	−0.9985
95th Percentile	1	−0.9936	1	−0.9973	1	−0.9871	1	−0.9923

Note: This table provides summary statistics on Unit root and cointegration tests for the cross-listed stocks in out sample. Panel A reports summary statistics (various percentiles) on the Augmented Dickey-Fuller test; Panel B reports Likelihood ratio statistics testing for cointegration (Johansen's LR statistic) for one cointegrating vector (LR1) and two cointegrating vector (LR2) – i.e. both series are stationary; Panel C reports summary statistics for the speed of adjustment coefficients for the listing in Canada (α_{CAN}) and the US (α_{US}); Finally, Panel D shows summary statistics for the cointegrating vector. Note that all statistics are computed per stock per annum.

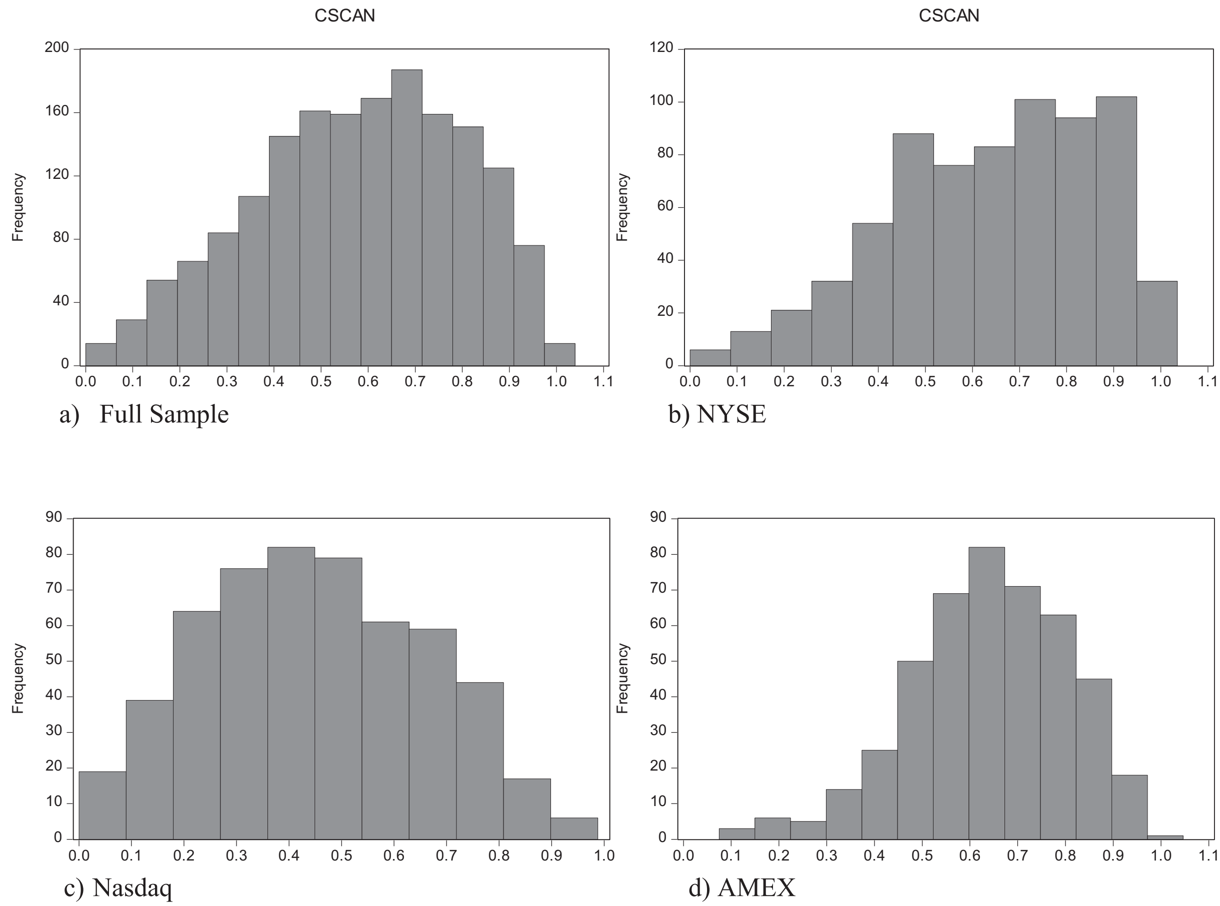

Fig. 1. Histograms of Canadian Component Share.

cointegration, where we test for the presence of no cointegrating relationship versus one cointegrating relationship and remove those firms in the years where the Johansen test does not reject the null of no cointegrating relationship. Panel B of Table 2 presents the Likelihood Ratio test statistics for the test of no cointegrating relationships (LR1) and more than one cointegrating relationships (LR2). Again, we observe that for most of the stocks, we find evidence of the presence of one cointegrating relationship (In approximately 1% of the cases we do not find evidence of cointegration). After excluding any firm-years that fail the above tests, we are left with 1612 firm-year observations.

To assess the level of price discovery for the cross-listed companies in the US and Canadian markets, we estimate the VECM stated in Eq. (1) on a yearly basis following the Johansen procedure, i.e. by estimating the auxiliary regressions and computing the canonical correlations. This provides us with Full Information Maximum Likelihood estimates for the speed of adjustment coefficients and the cointegrating vector, which are summarized in Panels C and D of Table 2.

Panel C presents the speed of adjustment coefficient for Canada and the US (α^{CAN} and α^{US}, respectively). As mentioned before, if the Canadian market is informationally dominant then we expect α^{CAN} to be close to zero with α^{US} to be high (in absolute terms) relative to α^{CAN}. The results in Panel C show a wide variation. The 25th percentile values for the full sample show higher absolute values for α^{CAN} than for α^{US}, indicating that the Canadian market is dominated by the US for at least 25% of the sample. However, the median values indicate that the Canadian market is generally the more informationally important market. This is consistent for the NYSE and AMEX sample, although the 5th percentile α^{CAN} for the AMEX is considerably lower than for the NYSE. For the NASDAQ, the median α^{CAN} is large than α^{US} suggesting the US dominates for most Canadian firms on the NASDAQ. Panel D of Table 2 presents the estimates for the cointegrating vector, which, by definition, is 1 for β^{CAN} and, by theory, should be -1 for β^{US}. Overall, we observe a very tight range of values for β^{US} close to -1, and the median value for β^{US} is very close to -1.

In Table 3, we report summary statistics on the $CS^{CA\,N}$, i.e. the proportion of price discovery occurring on the TSX, over the full sample period as computed in Eq. (2). On average, we observe that nearly 58% of price discovery occurs in the Canadian market. However, the Canadian market does not dominate for all shares, as the standard deviation of CS^{CAN} is about 22%. For more than 25% of our sample CS^{CAN} is less than 50%, indicating that in these instances the US exchanges dominate price discovery. Consistent with Eun and Sabherwal (2003), we observe a wide range in CS^{CAN}, between 18.19% and 91.23% at the 5th and 95th percentiles, respectively. The 5th percentile value suggests that for a subgroup of cross-listed firms, the TSX provides very little of the price discovery, while for the firms at the 95th percentile the US exchanges play almost no role in producing information.[8] The relatively wide distribution and higher average Component Share for the Canadian market is confirmed by the histogram in Fig. 1.

We observe similar patterns for both the NYSE and the AMEX. On average, around 64% of the price discovery occurs in the home

[8] Gonzalo and Granger (1995) provide a formal test for whether the contribution to the permanent component (i.e. the Component Share) is significant or not. This test shows that for 1.9% of the firm year observation CS^{CAN} is not significantly different from zero, while for 7.9% of the firms the CS^{US} is not significantly different from zero.

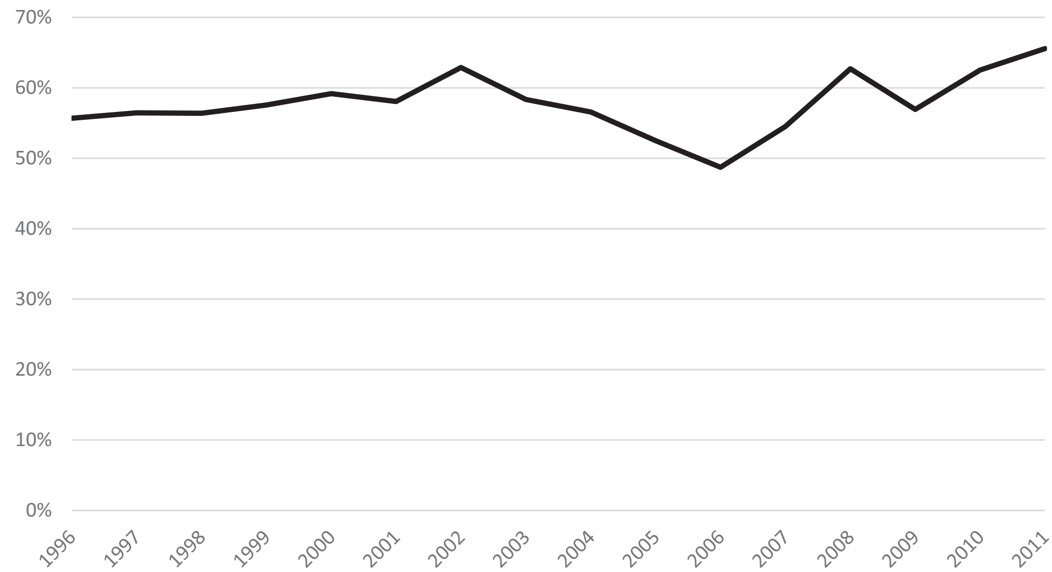

Fig. 2. Full Sample Canadian Component Shares over Time.

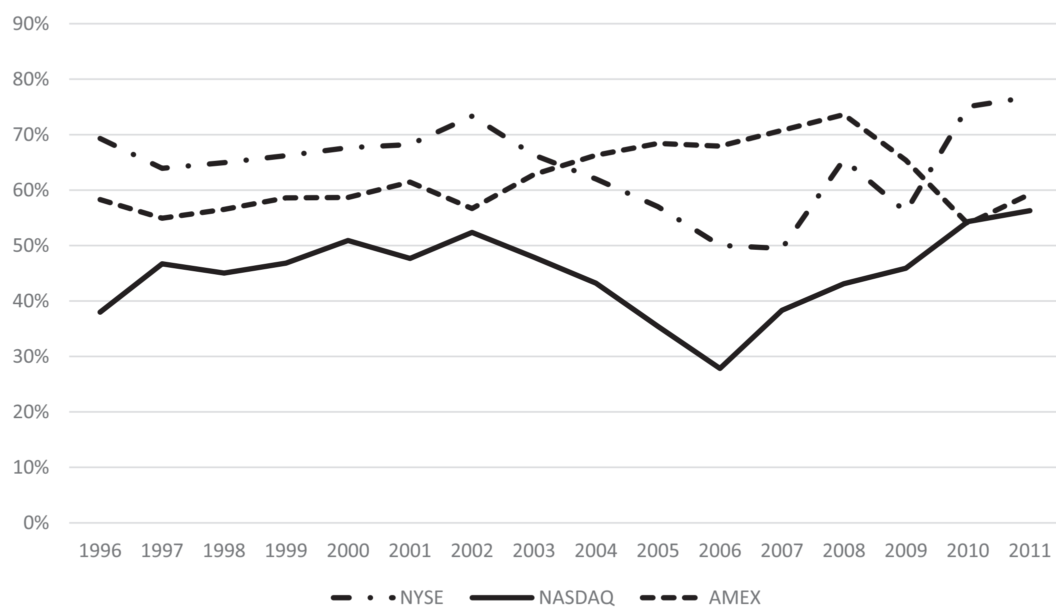

Fig. 3. Sample Canadian Component Shares over Time by Exchange.

market, with a standard deviation of about 22% and values ranging from 22.99% to 94.26% for the NYSE, and a standard deviation of 16.44% and values ranging from 35.67% to 88.88% for the AMEX, for the 5th and 95th percentiles, respectively. However, for Canadian firms cross-listed on the NASDAQ, it is the US that is more important, on average. Specifically, the average and median CS^{CAN} is below 50%, indicating that for more than half these companies the NASDAQ is informationally more important. The standard deviation is about 21% and the 5th and 95th percentile values are 10.99% and 79.93%, respectively. This finding seems to suggest that the NASDAQ has an informational advantage for the firms listing on its exchange, although again there is wide variation in the values. The difference between the three exchanges is better seen in the histograms in Fig. 1, which show that for most of the firms cross-listing on the NYSE the Canadian market is dominant. Similarly, for firms cross-listing on AMEX, the Canadian market is dominant, but there seems to be less dispersion in the distribution of the Component Shares. Finally, for the NASDAQ we observe a histogram that has more weight in the left tail rather than the right tail.

In Fig. 2, we plot the average CS^{CAN} per year for both the full sample. We observe a gradual increase from 56% to 63% between 1996 and 2002, which declines to 49% over the next 4 years, before increasing to about 66% by 2011. Most of the yearly changes are relatively small, although we do observe a sharp increase between 2006 and 2008. We observe similar patterns and trends in both the NYSE and NASDAQ samples, presented in Fig. 3, with only relatively small changes in most years. The AMEX differs in that we observe a generally increasing trend from the early part of the sample until 2008, after we which we see a relatively rapid reduction in price discovery. On the whole, price discovery appears to be persistent and changes gradually over time.

4.3. Measures of market quality and control variables

To identify the determinants of price discovery we consider a number of variables that measure various aspects of the relative liquidity or competitiveness of the US and Canadian exchanges. The measures of market quality that we include are Rel_Quotes_{it}, which is the number of quotes on the TSX divided by the number of quotes on the US exchange for company i in year t; Rel_Vol_{it}, which is the number of shares traded on the TSX divided by the number of shares traded on the US exchange for company i in year t; Rel_Trades_{it}, which is the number of trades on the TSX divided by the number of trades on the US exchange for company i in year t; and $Rel\_Spreads_{it}$, which is the average percentage spread on the TSX divided by the average percentage spread on the US exchange for company i in year t. In addition, we decompose the relative trades into $Rel\_Small\_Trades_{it}$, $Rel\_Mid\_Trades_{it}$, and $Rel\_Large\_Trades_{it}$. We define small, medium and large trades in the same manner as Eun and Sabherwal (2003), where small trades are those of less than 2500 shares, medium are 2501–10,000 and large are greater than 10,000. All these data are collected from TRTH.

In addition, we collect data on several control variables. We obtain $LogMV_{it}$, which is the natural log of the market capitalization for company i in year t; $LogYrsListed_{it}$, the natural log of the number of years that company i has been listed on the US exchange as at year t. These data are obtained from DataStream. We also include two dummies, $NASDAQ_i$ and $AMEX_i$ which equal 1 if company i is listed on the NASDAQ or AMEX, respectively.

Table 4 presents summary statistics of the explanatory variables we employ. We express our explanatory variables as relative ratios where we divide the Canadian value by the US value. Panel A considers the relative ratio of trades on the Canadian exchange to the US exchange. For the full sample, we observe considerable variation for the firm-year observations. At the 5th percentile, we observe that the US exchange has over seven times the number of trades of the Canadian exchange (0.1369), while at the 95th percentile the Canadian exchange has 14 times the trades of the US market. The median value of 1.673, however, suggests that the Canadian market retains the majority of the trading for more than half the sample. Interestingly, when we split the sample by exchange, over three-quarters of the firms-years for the NASDAQ companies have the greater portion of their trades occurring in the US. By contrast, less than 25% of the NYSE and AMEX firms-years have more of their trading in the US.

Panel B, which presents the relative number of quotes, shows a much tighter range, 0.47 to 4.36 for the full sample. We also observe less difference between the NASDAQ and the other exchanges, a median of 0.99 compared to 1.16 for the NYSE with similar values at the 5th and 95th percentile.

Panel C presents the results for relative spreads. A lower relative spread denotes a lower spread in the Canadian market. For the full sample, we observe considerable variation, ranging from spreads in Canada that are 2.5 times smaller than those of the US, to spreads that are twice as large. Consistent with results reported in Panel A, more than 25% of firm-years have smaller spreads on the US for the NYSE and AMEX samples, and more than 50% for NASDAQ. The percentages with lower spreads in the US are similar to the percentages with higher US shares of price discovery, providing some support for Eun and Sabherwal's (2003) finding that spreads and price discovery are related.

Panel D presents the relative volume. As with relative trades and quotes, a higher number denotes greater activity on the Canadian market. The results are mostly consistent with the other panels, with just over 25% of firms having the majority of their trading volume in the US while the median indicates clear dominance by the TSX. Again we observe that for most NASDAQ

Table 4
Regression variable summary statistics.

	Full sample	NYSE	NASDAQ	AMEX
Panel A: Relative number of trades				
5th Percentile	0.1369	0.5817	0.0983	0.4354
25th Percentile	0.6144	1.3225	0.1959	1.1961
Median	1.6730	2.5084	0.4167	2.6684
75th Percentile	4.0033	6.4318	0.9750	4.6383
95th Percentile	14.5099	24.8311	3.3224	11.2263
Panel B: Relative number of quotes				
5th Percentile	0.4707	0.5623	0.4091	0.4457
25th Percentile	0.8421	0.9076	0.7337	0.9055
Median	1.1429	1.1580	0.9945	1.5583
75th Percentile	1.6162	1.4943	1.2734	2.8784
95th Percentile	4.3576	2.3575	2.8388	6.1611
Panel C: Relative spreads				
5th Percentile	0.3751	0.3784	0.4280	0.3464
25th Percentile	0.6299	0.5982	0.8596	0.5604
Median	0.9185	0.8329	1.2732	0.7645
75th Percentile	1.3169	1.1662	1.7168	1.0326
95th Percentile	2.0679	1.7733	2.3782	1.7326
Panel D: Relative volume				
5th Percentile	0.1279	0.4667	0.0781	0.3412
25th Percentile	0.6000	1.3108	0.1909	1.2222
Median	1.8257	3.0000	0.4225	3.0000
75th Percentile	5.5590	10.4000	1.1905	7.1214
95th Percentile	30.6905	42.2222	6.2963	21.3115
Panel E: Relative number of small trades				
5th Percentile	0.1330	0.5416	0.0832	0.4390
25th Percentile	0.6116	1.3846	0.1893	1.4222
Median	1.7695	2.7306	0.4208	2.8524
75th Percentile	4.2499	6.9764	1.0396	4.8017
95th Percentile	16.4676	27.2885	3.8773	12.4828
Panel F: Relative number of medium trades				
5th Percentile	0.0534	0.1929	0.0207	0.2467
25th Percentile	0.4423	1.0429	0.1367	1.1269
Median	1.6901	3.0702	0.3360	3.0267
75th Percentile	6.3446	13.6054	1.1496	7.4266
95th Percentile	42.4545	80.0171	9.7850	27.1651
Panel G: Relative number of large trades				
5th Percentile	0.0559	0.2585	0.0076	0.1354
25th Percentile	0.4345	1.1687	0.1099	0.6552
Median	1.9319	4.3927	0.3489	3.0138
75th Percentile	9.0000	27.1622	1.7778	9.2778
95th Percentile	109.9167	196.0378	22.0000	45.0000
Panel H: Market capitalization ($US mill)				
5th Percentile	17.22	180.95	7.21	17.22
25th Percentile	148.52	1346.74	85.22	87.39
Median	610.38	3797.19	210.75	219.24
75th Percentile	3362.27	12883.90	588.05	507.39
95th Percentile	26493.40	36542.90	3127.42	3211.47

Note: This Table shows summary statistics for the various measures of market quality and for the control variables we include in Eq. (4). We report averages and values at different percentiles for Relative number of trades (Panel A); Relative number of quotes (Panel B); Relative spreads (Panel C); Relative volume (Panel D); Relative number of small trades (Panel E); Relative Number of Medium Trades (Panel F); Relative number of large trades (Panel G); and Market capitalization (Panel H). Small trades are defined as trades for less than 2500 shares; Medium-sized trades are volumes between 2501 and 10,000; Large trades are for more than 10,000 shares.

cross-listed companies the majority of the volume is traded in the US. We also split volume into three additional categories, relative percentage of small, medium and large trades. The results for the three panels, E, F, G, are again broadly consistent with the other panels. The only exception is for large trades for AMEX firms, where the US has a greater share at the 25th percentile.

Panel H presents summary statistics on the market capitalization of the sample. We observe a wide range from small companies, the 5th percentile for the full sample is just US $17.5 million, to reasonably large firms, the 95th percentile is US $26.5 billion. When we look at the exchange sub-samples, we observe

Table 5
Dynamic lag structure and endogeneity.

Panel A: Dynamic lag structure of price discovery

	$Logit\_CS_{it-1}$	$Logit\_CS_{it-2}$	$LogMV_{it}$	$LogYrsListed_{it}$	Year Dummies	R^2
$Logit\_CS_{it}$	0.2246***	0.1048	−0.0717	−0.6236***	YES	0.210
	(3.64)	(1.39)	(−1.26)	(−2.65)		
$Logit\_CS_{it}$		0.1800***	−0.0441	−0.6784***	YES	0.170
		(3.01)	(−0.78)	(−1.98)		

Panel B: Relation between market quality measures and lagged price discovery

	Rel_Quotes_{it}	Rel_Vol_{it}	$Rel\_Spreads_{it}$
$Logit\_CS_{it-1}$	0.0809**	0.4414	−0.0505**
	(2.25)	(1.31)	(−2.15)
Log_MV_{it-1}	0.0510	0.0859	−0.0393**
	(1.53)	(0.30)	(−2.35)
$LogYrsListed_{it-1}$	0.7576***	−0.9691	0.0329
	(3.77)	(−0.89)	(0.73)
Rel_Quotes_{it-1}	0.1972	0.1004	−0.0011
	(2.89)	(0.54)	(−0.18)
Rel_Vol_{it-1}	0.0024	0.3992***	−0.0018**
	(0.72)	(6.19)	(−2.52)
$Rel\_Spreads_{it-1}$	0.0112	−0.0782	0.3200***
	(0.13)	(−0.21)	(3.00)
Year Dummy	YES	YES	YES
R^2	0.2048	0.4120	0.3052

Panel C: Strict exogeneity tests

	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$
Log_MV_{it}	−0.0187	−0.0187	−0.0277	−0.0290	−0.0193
	(−0.50)	(−0.50)	(−0.77)	(−0.80)	(−0.41)
$LogYrsListed_{it}$	−0.2821***	−0.2534***	−0.2480***	−0.2589***	−0.5168***
	(−2.70)	(−2.51)	(−2.46)	(−2.57)	(−2.85)
Rel_Quotes_{it}	−0.0302**	−0.0263**	−0.0249**	−0.0307**	−0.0305**
	(−2.34)	(−2.11)	(−2.01)	(−2.50)	(−2.50)
Rel_Vol_{it}	0.0123***	0.0084**	0.0118***	0.0084***	0.0083**
	(3.38)	(2.08)	(3.30)	(2.12)	(2.17)
$Rel\_Spreads_{it}$	−0.9105***	−0.9067***	−0.8195***	−0.8206***	−0.8091***
	(−7.74)	(−7.75)	(−7.57)	(−7.46)	(−7.52)
Rel_Quotes_{it+1}	0.0264**			0.0239**	0.0245**
	(2.22)			(2.18)	(2.27)
Rel_Vol_{it+1}		0.0103**		0.0085**	0.0081**
		(2.53)		(2.09)	(2.04)
$Rel\_Spreads_{it+1}$			−0.2620***	−0.2423***	−0.2447***
			(−3.89)	(−3.55)	(−3.65)
Log_MV_{it+1}					−0.0107
					(−0.25)
$LogYrsListed_{it+1}$					0.4181*
					(1.72)
Year Dummy	YES	YES	YES	YES	YES

Note: This Table reports various tests for the dynamic lag structure of price discovery and assesses the endogenous relation between price discovery and various measures of market quality. Panel A assesses the dynamic nature of the Component Share. Panel B examines the relation between current measures of market quality and lagged price discovery. Finally, Panel C presents a test for strict exogeneity by regression current values of price discovery on future values of market quality. All models are estimated by OLS and include year dummies. All standard errors are robust, controlling for clustering at the firm level (see Petersen, 2009). *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

a marked difference between the NYSE and the other exchanges. The NYSE, not surprisingly, is dominated by considerably larger firms, in fact the 5th percentile is around the same level as the median firm size for the NASDAQ and AMEX exchanges. Equally, the 95th percentile for the NASDAQ and the AMEX is smaller than the median for the NYSE. Firms on the NASDAQ and AMEX by contrast are of fairly similar sizes.

5. Results

5.1. Determinants of price discovery

To examine the causal relation between price discovery and various measures of market quality, we follow Wintoki et al. (2012) by employing a dynamic panel GMM estimator. The dynamic panel GMM estimator can provide consistent and unbiased estimates when there is endogeneity and a dynamic relation between the dependent and independent variables. In our case, we can expect endogeneity induced by potential reverse causality. On the one hand, we may expect that various aspects of market quality have a causal effect on price discovery i.e. relative improvements in market quality may positively affect the contribution to price discovery of a market. Concurrently, the degree of price discovery may affect measures of market quality i.e. an improvement in price discovery may improve aspects of market quality for that firm. At the same time, we also may expect that there is persistence in the measures of market quality and price discovery. As demonstrated by Wintoki et al. (2012), the presence of simultaneity and persistence means that OLS, or a fixed effects estimator, would produce biased estimates of the causal relation between market quality and price discovery. By applying the dynamic panel GMM estimator, we can capture the causal relation between market quality and price discovery, by estimating the following equation:

$$Logit\_CS_{it} = \alpha + \sum_j \beta_j Logit\_CS_{it-j} + \boldsymbol{\delta}\mathbf{MQ_{it}} + \boldsymbol{\gamma}\mathbf{Controls_{it}} + \eta_i + \varepsilon_{it}, \tag{4}$$

where $Logit\_CS_{it}$ is the logit transformation of the CS^{CAN}. We apply this transformation as the Component Share is bounded between zero and one. $\mathbf{MQ_{it}}$ is the vector of market quality measures discussed in Section 4.3, $\mathbf{Controls_{it}}$ is the vector of control variables, and η_i captures a firm-level fixed effect. In Eq. (4), we treat all market quality measures as potentially endogenous, whereas the controls are assumed to be exogenous.

Although Eq. (4) can be estimated directly using the dynamic panel GMM estimator, there are several steps prior to this estimation that are worth documenting to confirm the accuracy of the model specification. First, we confirm the dynamic completeness of the model, by obtaining the correct lag structure for the dependent variable. To obtain this, we estimate the following model by OLS for different values of j,

$$Logit\_CS_{it} = \alpha + \sum_j \beta_j Logit\_CS_{it-j} + \boldsymbol{\gamma}\mathbf{Controls_{it}} + \eta_i + \varepsilon_{it}, \tag{5}$$

and report the results of this regression in Table 5, Panel A, where we control for clustering at the firm level in the calculation of the standard errors (see Petersen, 2009). In the first row of Panel A, we add two lags of the dependent variable. We observe that the first lag is significant at the 1% level, while the second lag is insignificant. This suggests an AR(1) structure, where there is a dynamic relation between the current and previous year's values of CS. Of the control variables, we note that only the log of years listed is significant and negative (i.e. the longer the firm is cross-listed in the US, the lower its CS becomes). In the second row of Panel A, we remove the first lag, and note that the second lag now becomes significant at the 1% level. Hence, the persistence captured by the first lag is taken over by the second lag. This is desirable if lagged values of *Logit_CS* are to be a useful instrument in the dynamic panel GMM.

We next assess whether the measures of relative market quality, which we assume will have a causal effect on price discovery, are endogenous. We do this by assessing whether the variables of interest (*Rel_Quotes*, *Rel_Vol* and *Rel_Spread*) are affected by lagged values of price discovery. Finding such a relationship would indicate a possible endogenous relation between these variables and price discovery. To assess this, we perform the following regression,

$$MQ^k_{it} = \alpha + \beta Logit\_CS_{it-1} + \boldsymbol{\delta}\mathbf{MQ_{it-1}} + \boldsymbol{\gamma}\mathbf{Controls_{it-1}} + \eta_i + \varepsilon_{it}, \tag{6}$$

where we relate the level of the measure for market quality to lagged values of price discovery, market quality and controls.

In Panel B of Table 5, we report the results for these regressions. We observe that the relative number of quotes between the US and Canada are positively related to the lagged values of price discovery. This finding implies that if price discovery is high in Canada in the previous year, this increases the number of quotes in the Canadian market compared with the US market in the subsequent year, suggesting that relative quoting activity is potentially endogenous. We also note that relative quote activity is persistent over time, as lagged relative quotes are significant and positive determinants of the current relative quote activity. For relative volume, we observe that the relation with lagged values of price discovery is insignificant, but there is persistence in relative volume as documented by the significant coefficient on lagged relative volume. For relative spreads, we observe that there is a significant negative relation with lagged values of price discovery, suggesting that if lagged price discovery in Canada was higher in the past, then the spreads will be relatively lower in Canada in the subsequent year. This result again suggests that this variable is potentially endogenous. We also find that relative spreads are persistent over time. These findings generally support the importance of price discovery for exchanges. Higher price discovery gives firms an advantage in terms of both quote activity and spreads.

Wintoki et al. (2012) suggest an additional test for strict exogeneity, by regressing current values of the dependent variable on future values of potentially endogenous variables. This regression is of the form:

$$Logit\_CS_{it} = \alpha + \boldsymbol{\delta}\mathbf{MQ_{it}} + \boldsymbol{\gamma}\mathbf{Controls_{it}} + \boldsymbol{\lambda}\mathbf{MQ_{it+1}} + \eta_i + \varepsilon_{it}, \tag{7}$$

where the test for strict exogeneity centers on λ. If λ is significantly different from zero, we can reject the null hypothesis of strict exogeneity.

In Panel C of Table 5, we report the results for the regression in Eq. (7), where we add the measures of market quality, one-by-one and add them all together. All variables of interest yield significant coefficients on their future values, suggesting that price discovery has an effect on these variables. Hence the notion of strict exogeneity with respect to these variables can be rejected. These tests demonstrate why attempting to determine the drivers of price discovery using traditional OLS analysis is inadequate. There is clear evidence of a simultaneity problem which biases the results based on OLS regressions.

In Table 6, we report the results for Eq. (4) where we estimate the causal relation between the measures of market quality and price discovery. We estimate this model by system GMM, where we use a two-step procedure in the estimation. As it has been shown that this procedure typically produces downward biased

Table 6
System GMM estimates of the determinants of price discovery.

	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$
$Logit\_CS_{it-1}$	0.4100***	0.3811***	0.3649***
	(6.37)	(4.71)	(3.80)
Rel_Quotes_{it}	−0.0294	−0.0712	−0.1216
	(−0.32)	(−0.83)	(−1.32)
Rel_Vol_{it}	0.0288**		
	(2.09)		
Rel_Trades_{it}		0.0723**	
		(2.37)	
$Rel\_Small\ Trades_{it}$			0.0524*
			(1.86)
$Rel\_Med\ Trades_{it}$			−0.0016
			(−0.13)
$Rel\_Large\ Trades_{it}$			−0.0005
			(−0.33)
$Rel\_Spreads_{it}$	−0.7078**	−0.7090*	−0.9005**
	(−2.60)	(−1.88)	(−2.32)
Log_MV_{it}	0.0030	−0.0193	0.0135
	(0.12)	(−0.57)	(0.58)
$LogYrsListed_{it}$	0.0049	0.0220	−0.0696
	(0.05)	(0.20)	(−0.67)
Year Dummies	YES	YES	YES
Observations	1328	1328	1290
AR(1) in first differences	(0.000)	(0.000)	(0.000)
AR(2) in first differences	(0.131)	(0.217)	(0.198)
Hansen test for overidentifying restrictions	(0.376)	(0.312)	(0.178)
Difference-in-Hansen test	(0.911)	(0.940)	(0.443)

Note: This Table reports results for the model that assesses the causal relation between various measures of market quality and price discovery (Logit transformation of the Component Share). The model is estimated by system GMM, where the measures for market quality are treated as endogenous and the control variables as exogenous. We use lags two and three as the internal instruments and use the two-step GMM estimation procedure. We employ the Windmeijer (2005) correction in the calculation of standard errors and report t-statistics in parentheses. *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

Table 7
Results by exchange.

	NYSE			NASDAQ			AMEX		
	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$	$Logit\_CS_{it}$
$Logit\_CS_{it-1}$	0.3326***	0.2789***	0.1103	0.2842***	0.3484***	0.265	0.3290*	0.3394**	0.5725***
	(4.52)	(3.12)	(0.68)	(2.77)	(2.99)	(1.59)	(1.86)	(2.25)	(3.24)
Rel_Quotes_{it}	−0.1441	0.1704	−0.6984	−0.1153*	−0.0388	−0.0209	0.1179	0.1222	0.0321
	(−0.33)	(0.43)	(−0.87)	(−1.86)	(−1.35)	(−0.20)	(1.34)	(1.51)	(0.13)
Rel_Vol_{it}	0.0302**			0.1466*			0.0173		
	(2.49)			(1.75)			(0.68)		
Rel_Trades_{it}		0.0485**			0.0435			0.0812**	
		(2.59)			(0.43)			(2.19)	
$Rel\_Small\ Trades_{it}$			0.0527**			0.0173			0.0380
			(2.07)			(0.54)			(0.13)
$Rel\_Med\ Trades_{it}$			−0.0180			0.0024			0.0153
			(−0.74)			(0.07)			(0.26)
$Rel\_Large\ Trades_{it}$			0.0011			−0.0097			0.0005
			(0.47)			(−0.99)			(0.09)
$Rel\_Spreads_{it}$	−0.4774*	−0.3723	−1.298**	−0.7021**	−0.8403*	−1.105***	−1.210*	−1.126*	0.5586
	(−1.64)	(−0.93)	(−2.01)	(−2.76)	(−3.58)	(−3.81)	(−1.62)	(−1.79)	(0.57)
Log_MV_{it}	0.0889**	0.0622	0.1128	−0.0384	−0.0413	−0.0432	−0.0041	−0.0569	0.0746
	(2.08)	(1.34)	(1.52)	(−1.38)	(−1.42)	(−1.15)	(−0.04)	(−0.74)	(0.64)
$LogYrsListed_{it}$	−0.2080*	−0.1782	−0.2611*	0.1515	0.1690*	0.1284	−0.0420	0.0387	−0.0589
	(−1.73)	(−1.42)	(−1.97)	(1.41)	(1.73)	(1.12)	(−0.32)	(0.33)	(−0.25)
Year Dummies	YES	YES	YES	YES	YES	YES	YES	YES	YES
Observations	569	569	554	433	433	423	72	72	70
AR(1) in first differences	(0.000)	(0.000)	(0.022)	(0.004)	(0.003)	(0.004)	(0.000)	(0.000)	(0.002)
AR(2) in first differences	(0.456)	(0.616)	(0.923)	(0.818)	(0.730)	(0.753)	(0.310)	(0.509)	(0.218)
Hansen test for overidentifying restrictions	(0.160)	(0.191)	(0.060)	(0.914)	(0.887)	(0.677)	(0.275)	(0.799)	(0.049)
Difference-in-Hansen test	(0.094)	(0.203)	(0.071)	(0.778)	(0.954)	(0.561)	(0.546)	(0.655)	(0.007)

Note: This Table reports results for the model that assesses the causal relation between various measures of market quality and price discovery (Logit transformation of the Component Share) for Canadian firms cross-listed on different US exchanges. The model is estimated by system GMM, where the measures for market quality are treated as endogenous and the control variables as exogenous. We use lags two and three as the internal instruments and use the two-step GMM estimation procedure. We employ the Windmeijer (2005) correction in the calculation of standard errors and report *t*-statistics in parentheses. *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

estimates of the standard errors, we use the Windmeijer (2005) correction in the computation of the standard errors. First, the results show that the coefficient on the lagged price discovery is 0.41, which is significant at the 1% level. This suggests that there is persistence in price discovery over time. Hence, once a market becomes dominant, this is not likely to change rapidly over time. Second, the results show that the relative number of quotes is insignificant in all specifications, suggesting that there is no causal effect of quote activity on price discovery. For relative volume, we observe a positive and significant coefficient. Hence, there is a positive causal effect of relative volume on price discovery, i.e. if the trading volume in Canada increases relative to the trading volume in the US, price discovery in Canada tends to increase. For relative spreads, we observe a negative and significant coefficient, suggesting that a decrease in the spread in one market relative to the other market leads to an increase in price discovery in that market. These results are consistent with Eun and Sabherwal (2003). The control variables in this specification are all insignificant.

We further report various diagnostic statistics on the model that we estimate in Table 6. First, we report the *p*-values for the AR(1) and AR(2) test in the first-differenced residuals. Since the first-differencing of the data introduces first-order autocorrelation in the residuals, the highly significant *p*-value for the AR(1) test is expected. However, for the AR(2) test, we find a *p*-value of 0.131. Hence, there is no statistical evidence for second order autocorrelation, suggesting that the dynamic lag structure of the model is sufficient (i.e. one lag for the price discovery variable). The *p*-value for the Hansen test for overidentifying restrictions is insignificant at 0.376, suggesting that our instruments are valid. Finally, we report the results for the Difference-in-Hansen test. This test suggests that we cannot reject the null of exogeneity for a subset of the instruments.

In the second column, we report the results when we replace the relative volume, with the relative number of trades occurring in each market. On the whole, our earlier findings do not change by introducing this variable, relative quotes remains insignificant, while relative spread remains significant and relative trades is highly significant.

Finally, the last column of Table 6 reports the results for relative trades of different sizes, small, medium and large trades. The literature on stealth trading (Barclay and Warner, 1993; and Chakravarty, 2001) suggests that an informed trader will try to hide their private information by splitting large orders into smaller ones. The findings of prior literature suggest that it is the medium-sized trades that carry most price information, confirming the idea of order splitting. The argument for the medium-sized trades being most informative is that an optimal point is sought between the price impact of trades (i.e. trades revealing information) and the cost of trades. Small trades may have a small price impact, but are expensive as they lead to high trading costs whereas large trades have a high price impact but low costs. An optimal point is found somewhere in the middle (see e.g. Chakravarty, 2001). In line with this, Eun and Sabherwal (2003) document, in their cross-sectional regression, that price discovery is mainly related to the relative volume in medium-sized trades. Hence, the literature mostly suggests that it is medium-sized trades that are most informative.

However, it is important to note that these prior studies were conducted in a pre-decimalization period, where prices were quoted in fractions, resulting in a relatively wide bid-ask spreads and therefore a relatively high cost of trading. One can expect that as the spread reduces, the cost of trading reduces, and the size of the orders will decrease as well, shifting the optimal trade size towards smaller trades. Indeed, Goldstein and Kavajecz (2000)

Table 8
Results for sub-period analysis.

	Pre 2001			Post 2002		
	Logit_CS$_{it}$	*Logit_CS$_{it}$*	*Logit_CS$_{it}$*	*Logit_CS$_{it}$*	*Logit_CS$_{it}$*	*Logit_CS$_{it}$*
Logit_CS$_{it-1}$	0.3254**	0.2874**	0.2946**	0.4044***	0.3820***	0.3348**
	(2.58)	(2.29)	(2.05)	(4.64)	(4.02)	(3.58)
Rel_Quotes$_{it}$	0.0141	0.0247	0.0560	−0.0297	−0.0565	−0.0337
	(0.22)	(0.41)	(0.76)	(−0.25)	(−0.51)	(−0.30)
Rel_Vol$_{it}$	0.0073			0.0360		
	(0.50)			(1.63)		
Rel_Trades$_{it}$		0.0174**			0.0788	
		(0.94)			(1.56)	
Rel_Small Trades$_{it}$			−0.0006			0.1113**
			(0.76)			(2.30)
Rel_Med Trades$_{it}$			0.0404			−0.0217
			(1.16)			(−1.50)
Rel_Large Trades$_{it}$			0.0008			−0.0002
			(0.11)			(−0.15)
Rel_Spreads$_{it}$	−0.8702***	−0.8777***	−0.6469***	−0.9519*	−0.8876*	−1.059**
	(−3.91)	(−3.70)	(−2.70)	(−1.86)	(−1.95)	(−2.39)
Log_MV$_{it}$	0.0909***	0.0809	0.0568*	−0.0301	−0.0329	0.0072
	(3.94)	(2.78)	(1.68)	(−0.78)	(−0.86)	(0.28)
LogYrsListed$_{it}$	−0.0790	−0.1085	−0.1221	0.1128	0.0816	0.0295
	(−0.74)	(−1.05)	(−1.31)	(0.72)	(0.67)	(0.28)
Year Dummies	YES	YES	YES	YES	YES	YES
Observations	287	287	272	1041	1041	1018
AR(1) in first differences	(0.030)	(0.016)	(0.142)	(0.000)	(0.000)	(0.000)
AR(2) in first differences	(0.898)	(0.887)	(0.853)	(0.129)	(0.086)	(0.144)
Hansen test for overidentifying restrictions	(0.542)	(0.439)	(0.643)	(0.330)	(0.094)	(0.090)
Difference-in-Hansen test	(0.770)	(0.561)	(0.302)	(0.732)	(0.122)	(0.068)

Note: This Table reports results for the model that assesses the causal relation between various measures of market quality and price discovery (Logit transformation of the Component Share) for the period up to and including 2001 (pre-decimalization) and the period from 2002 onwards. The model is estimated by system GMM, where the measures for market quality are treated as endogenous and the control variables as exogenous. We use lags two and three as the internal instruments and use the two-step GMM estimation procedure. We employ the Windmeijer (2005) correction in the calculation of standard errors and report *t*-statistics in parentheses. *, **, and *** indicate significance at the 10%, 5% and 1% level, respectively.

show that order size decreases significantly after the NYSE reduced tick size from eights to sixteenths. Similarly, Chakravarty et al. (2005) note that after the decimalization in 2001, order size decreased, significantly so for the largest 50% of stocks traded on the NYSE. These findings suggest that small trades may have become more informative than large trades, due to the reduction in trading costs.

The results presented in Table 6, suggest that it is indeed the relative small trades that cause a shift in price discovery, i.e. the more small trades that occur in one market relative to the other, the higher the price discovery will be in that market.[9]

5.2. Robustness tests

The previous section evaluated the dynamic relation of price discovery and assessed the determinants of price discovery. In this section, we perform several robustness tests by examining whether the results are consistent for cross-listing on the different US markets, and over different subsamples.

In Table 7, we report the results for cross-listing on the different exchanges. For cross-listings on the NYSE, we observe persistence in price discovery in two of the three model specifications (only in the case where we break down trades into different sizes does the persistence in price discovery become insignificant). Consistent with the full sample, we find that both relative volume (positive) and relative spreads (negative) are significant determinants of price discovery. We also note that when we replace relative volume with relative trades, and when we split trades into size categories, trades and especially small trades significantly determine price discovery. In addition, we note that, in the specification with relative volume, market capitalization is positively related to price discovery, i.e. the larger firms have higher Component Shares in the US, and years listed has a negative sign.

For firms that cross-list on NASDAQ, we again note that the persistence in price discovery largely remains. Overall, the most important determinant for firms cross-listed on NASDAQ is the relative spread, but we also find evidence for relative quotes and relative volume. Interestingly, in this specification, we find no evidence for relative trades, or trades of different sizes.

In the last columns, we report the results for stocks listed on AMEX. For this exchange, we observe that the persistence in price discovery is present in all specifications. Relative spreads again seem to be an important determinant of price discovery, and we further find evidence for relative trades. We find no evidence that trades of different sizes have an impact on price discovery, but we do note that the sample size for stocks cross-listed on AMEX is relatively small.

As a second robustness test, we split the sample into two subperiods, 1996–2001, and 2002-onwards, where the first period broadly covers the period prior to US decimalization and the second period cover the post-decimalization period. We report the results for this subsample analysis in Table 8.

As we note, persistence in price discovery is present in both sub-periods, although it is stronger in the later period. We also note that relative spreads are significant determinants of price discovery in both subperiods. We find that small trades are only significant in the post-decimalization period. These results may be due to decimalization (smaller tick size) and perhaps the increased

[9] This result contrasts the findings of Eun and Sabherwal (2003), who document that it is the relative medium sized trades that are related to price discovery. However, their sample covers 6 months in 1998, reflecting the situation in a pre-decimalization period. We have performed a similar cross-sectional regression to Eun and Sabherwal (2003) using only the data for 1998 and find that indeed during that period medium-sized trades are positively related to price discovery. These results are available on request.

speed of execution and an increase in smaller trades that has occurred after decimalization.

6. Conclusion

In this paper, we examine the determinants of price discovery for a large sample of Canadian stocks that cross-list in the US over the sample period from 1996–2011. An important issue in this analysis is the presence of endogeneity, where causality may run from either price discovery to measures of liquidity and market quality, or the other way around. We resolve the endogeneity problem by making use of a system GMM approach, where we model price discovery in a dynamic panel data framework (see also Wintoki et al., 2012).

We show that there is strong positive persistence in price discovery, suggesting that once price discovery has been gained by a market it tends to remain there. We also observe that the relation between price discovery and various measures of market quality is indeed endogenous. We document that lagged measures of price discovery affect various aspects of market quality. We further demonstrate that the relative trading activity (either measured by volume or number of trades) has a significant positive causal effect on price discovery, while relative spreads have a significant negative causal effect on price discovery. Finally, we note that it is predominantly the relative number of small trades that affects price discovery.

Overall, our results suggest that improvements in market quality, such as a reduction in spreads and an increase in trade activity, have a positive and causal effect on price discovery, i.e. making the market more informationally efficient. Further results, based on our strict exogeneity test, show that price discovery has a reinforcing effect on market quality, i.e. improvement in price discovery leads to improvement in market quality. In addition, the finding that price discovery is persistent suggests that once price discovery is gained by a particular market, it is sticky and will persist. Hence, efforts to improve market quality and thus price discovery offer exchanges the opportunity to gain a long-term advantage over its competitors.

References

Agarwal, S., Liu, C., Rhee, S., 2007. Where does price discovery occur for stocks traded in multiple markets? Evidence from Hong Kong and London. Journal of International Money and Finance 26, 46–63.

Alhaj-Yaseen, Y., Lam, E., Barkoulas, J., 2014. Price discovery for cross-listed firms with foreign IPOs. International Review of Financial Analysis 31, 80–87.

Aslan, H., Easley, D., Hvidkjaer, S., O'Hara, M., 2011. Firm characteristics and informed trading: implications for asset pricing. Journal of Empirical Finance 18, 782–801.

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

Arellano, M., Bover, O., 1995. Another look at the instrumental variable estimation of error component models. Journal of Econometrics 68, 29–51.

Bacidore, J., Sofianos, G., 2002. Liquidity provision and specialist trading in NYSE-listed non-U.S. stocks. Journal of Financial Economics 63, 133–158.

Baillie, R., Booth, G., Tse, Y., Zabotina, T., 2002. Price discovery and common factor models. Journal of Financial Markets 5, 309–321.

Barclay, M.J., Warner, J.B., 1993. Stealth trading and volatility: which trades move prices? Journal of Financial Economics 34, 281–305.

Blundell, R., Bond, S., 1998. Initial conditions and moment restrictions in dynamic panel data models. Journal of Econometrics 87, 115–143.

Booth, G., Baillie, R., Tse, Y., Zabotina, T., 2002. Price discovery and common factor models. Journal of Financial Markets 5, 309–321.

Chakravarty, S., 2001. Stealth trading: which traders' trades move stock prices? Journal of Financial Economics 61, 289–307.

Chakravarty, S., Panchapagesan, V., Wood, R., 2005. Did decimalization hurt institutional investors? Journal of Financial Markets 8, 400–420.

Chen, H., Choi, P., 2012. Does information vault Niagara falls? Cross-listed trading in New York and Toronto. Journal of Empirical Finance 19, 175–199.

Chen, K., Guangzhong, L., Wu, L., 2010. Price discovery for segmented US-listed Chinese stocks: location or market quality? Journal of Business Finance and Accounting 38, 242–269.

Ding, D., Harris, F., Lau, S., McInish, T., 1999. An investigation of price discovery in informationally-linked markets: equity trading in Malaysia and Singapore. Journal of Multinational Financial Management 9, 317–329.

Easley, D., Hvidkjaer, S., O'Hara, M., 2002. Is information risk a determinant of asset returns? Journal of Finance 5, 2185–2221.

Easley, D., Hvidkjaer, S., O'Hara, M., 2010. Factoring information into returns. Journal of Financial and Quantitative Analysis 45, 293–309.

Eun, C.S., Sabherwal, S., 2003. Cross-border listing and price discovery: evidence from US-listed Canadian stocks. Journal of Finance 58, 549–575.

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

Frijns, B., Idriawan, I., Tourani-Rad, A., 2015. Macroeconomic news announcements and price discovery: evidence from Canadian-U.S. cross-listed firms. Journal of Empirical Finance 32, 35–48.

Goldstein, M., Kavajecz, K., 2000. Eights, sixteenths, and market depth: changes in tick size and liquidity provision on the NYSE. Journal of Financial Economics 56, 125–149.

Gonzalo, J., Granger, C., 1995. Estimation of common long-memory components in cointegrated systems. Journal of Business and Economic Statistics 13, 27–36.

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

Harris, F., McInish, T., Shoesmith, G., Wood, R., 1995. Cointegration, error correction, and price discovery on three informationally-linked security markets. Journal of Financial and Quantitative Analysis 30, 563–579.

Harris, F., McInish, T., Wood, R., 2002. Security price adjustment across exchanges: an investigation of common factor components for Dow stocks. Journal of Financial Markets 5, 341–348.

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

Hoechle, D., Schmid, M., Walter, I., Yermack, D., 2012. How much of the diversification discount can be explained by poor corporate governance? Journal of Financial Economics 102, 41–60.

Hupperets, E., Menkveld, A., 2002. Intraday analysis of market integration: Dutch blue chips traded in Amsterdam and New York. Journal of Financial Markets 5, 57–82.

Kadapakkam, P., Misra, L., Tse, Y., 2003. International price discovery for emerging stock markets: evidence from Indian GDRs. Review of Quantitative Finance and Accounting 21, 179–199.

Lieberman, O., Ben-Zion, U., Hauser, S., 1999. A characterization of the price behavior of international dual stocks: an error correction approach. Journal of International Money and Finance 18, 289–304.

O'Hara, M., 2003. Presidential address: liquidity and price discovery. Journal of Finance 58, 1335–1354.

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

Petersen, M.A., 2009. Estimating standard errors in finance panel data sets: comparing approaches. Review of Financial Studies 22, 435–480.

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

Su, Q., Chong, T., 2007. Determining the contributions to price discovery for Chinese cross-listed stocks. Pacific-Basin Finance Journal 15, 140–153.

von Furstenberg, G., Tatora, C., 2004. Bolsa or NYSE: price discovery for Mexican shares. Journal of International Financial Markets Institutions and Money 14, 295–311.

Windmeijer, F., 2005. A finite sample correction for the variance of linear efficient two-step GMM estimators. Journal of Econometrics 126, 25–51.

Wintoki, M., Linck, J., Netter, J., 2012. Endogeneity and the dynamics of internal corporate governance. Journal of Financial Economics 105, 581–606.

Yan, B., Zivot, E., 2010. A structural analysis of price discovery measures. Journal of Financial Markets 134, 1–19.