Price discovery on common and preferred shares across multiple markets

Cristina Scherrer
School of Economics and Finance, Queen Mary, University of London
E-mail: c.scherrer@qmul.ac.uk

Marcelo Fernandes
School of Economics and Finance, Queen Mary, University of London
E-mail: m.fernandes@qmul.ac.uk

This draft: May 21, 2012

Abstract: We extend the standard price discovery model to estimate the information share (IS) accounting for the information content of both common and preferred non US stocks, their American Depositary Receipts (ADRs) counterparts traded on the New York Stock Exchange and ARCA, and the exchange rate. We gauge the significance of price discovery in the home and foreign markets, through common or preferred stocks. One of the main critiques on the IS methodology is that it does not deliver a single measure when there is contemporaneous correlation among markets. We propose an ordering invariant methodology that delivers a single measure of IS. We employ a tick-by-tick database from the leading exchange in Latin America, namely, Bolsa de Valores de São Paulo (BM&FBovespa). We find that foreign market is more important than the home market for the price discovery of Petrobras, the Brazilian stated-owned oil giant, and Vale, one of the largest mining companies in the world. Additionally, the Brazilian market has lost significant importance after the 2008/2009 financial crisis. During this period, common and preferred stocks shared a single common factor, with voting premium being a stationary process.

JEL classification numbers: G15, G01

Keywords: price discovery, information share, common shares, preferred shares, voting premium.

1 Introduction

Price discovery comprises the determination of how efficiently and timely each market that trades a same asset incorporates news. It refers to the impounding of new information into the security's price (Hasbrouck, 1995). This type of analysis has become a hot topic of study in the past few years for a number of reasons. First, there is an increasing availability of high frequency data sets together with the growth of high frequency trades. Second, most algo trading strategies rely on price discovery analysis.

The Brazilian market has grown importance recently, attracting an increasing number of investors to its well established companies. Home to one of the world biggest oil and mining companies, it turns exciting to study and understand how their stocks behave at home and internationally, as the assets of this companies are also traded in other markets, such as New York Stock Exchange and ARCA. We use a tick-by-tick data set from the leading exchange in Latin America and one of the largest exchanges in the world (market capitalization of USD 1.54 trillion), BM&FBovespa. The reasons for using Brazilian data are twofold: first, it represents a significant large market. Second, there is data availability with many overlapping trading hours with United States. We try to gather all potential information from all possible markets by adding common and preferred shares. Common shares contain a voting right attached to it whereas standard preferred shares do not, or have some restriction. The preferred shares usually present some preference on dividend payments and/or priority on repayment of capital in case of liquidation of the company. The exact specification of the voting right and preference on dividend payments may vary among companies and markets. Preferred shares have been a typical way of financing for Brazilian companies and represent a larger amount of trading volume when compared to common shares. Nowadays Brazilian companies can issue one preferred share for each common one and in the past this relation was two for one.

We measure price discovery on a seven and five-variable model that accounts for common and preferred stocks, their American Depositary Receipts (ADRs) counterparts traded on the New York Stock Exchange and ARCA, as well as the exchange-rate. We include exchange-rate because it allows us to take into consideration exchange-rate risk and it separates exchange-rate movements from price discovery. We test whether price discovery takes place more in the home market, through common and/or preferred stocks, or in the foreign market. As a byproduct, we are able to say something about the voting premium (difference between common and preferred share prices).

An issue that arises with the Information Share measure is the lack of uniqueness and order dependence when there is contemporaneous correlation among markets. As Cholesky decomposition is implemented on the covariance matrix, the order of markets will impact the measure. The proposed way in the literature to deal with this is to take an average of the different measures across all possible orderings. In the case of two markets this is simple and reasonable. However, when the analysis regards a larger number of markets, this approach can be quite extensive. In our study, for instance, we would have to take an average among 5,040 measures in order to have the final Information Share. As a way to tackle this lack of uniqueness on the Information Share measure, we implement an ordering invariant methodology to decompose the covariance matrix. This outcome is particularly important when markets present contemporaneous correlation.

Price discovery framework assumes that all prices of a given security will have the same component in its price structure within all different markets this security is traded. This single component is seen as the common unobserved efficient price. There are two most widely used methodologies to measure price discovery. Information Share (IS) from Hasbrouck (1995) is defined as each market's proportion on the variance of the efficient price innovation. The second methodology is based on Gonzalo and Granger (1995) way to estimate and test the common factor of a permanent and transitory decomposition. Booth et al (1999), Chu et al (1999) and Harris et al (2002) apply this methodology to the price discovery concept, by introducing the component share as an approach to estimate the parameters of price discovery. The market with the higher weight on the permanent component (larger common factor weight) is the dominant market. Hasbrouck (2002), Harris et al (2002), De Jong (2002) and Lehmann (2002) compare the two approaches, finding similarities and differences. Baillie et al (2002) show the relation between the two models and conclude they present similar results if there is no correlation among markets. Grammig, Melvin and Schlag (2005) extends Hasbrouck (1995) model to account for exchange rate. The interpretation of the parameters change slightly with this inclusion, as we will see later in our simulations and empirical exercises. A few variations based on structural models are proposed by Yan and Zivot (2010) and Grammig and Peter (2010). Yan and Zivot (2010) propose a 2-variable model, where there is one transitory and one permanent shock on the efficient price. In a possible extension to n-variable model, restrictions would have to be imposed and order of variables would matter. Grammig and Peter (2010) aim to identify an unique measure when computing information shares. They impose restrictions

on the model based on the distributional properties of financial data, as a way to identify structural innovations in a multiple time series framework. The present paper does not aim to deal with structural models, as a way to avoid imposing restrictions on the model.

Regarding common and preferred stocks literature, there is some evidence that voting premium is a function of two aspects: a premium one might get in case of a merger or an acquisition and private benefits shareholders could obtain once holding the voting rights. The voting premium can vary a lot among countries. In the US, for instance, it is much smaller than other places, such as Italy. The explanation may come from the strong institutions in US which reduce considerably the possibility of getting private benefits on firm's decision and voting process (See Zingales, 1994, 1995). There are other things that may affect the difference between the price of a common and a preferred share. The dividend policy may vary from company to company, affecting the premium attached to the common share. Additionally, investors price liquidity and in some markets one may find a difference in terms of liquidity of these two types of shares.

Our dataset from BM&FBovespa, NYSE and ARCA ranges from December 2007 to November 2009, which entails us to analyze how price discovery works over the 2008/2009 financial crises and the evolution of the measure across different market cycles. We perform six month and yearly analysis to explore the crises dynamics. We use data from Petrobras traded at Bovespa and ADRs of Petrobras traded at NYSE and at ARCA and from Vale traded in Brazil and ADRs traded at NYSE. Petrobras is the largest company in Brazil and energy publicly-traded company. Vale is the second largest mining company in the world.

Bovespa core trading hours run from 9 am to 4 pm Eastern Time (ET) from mid February to mid November and in the rest of the year, from 8 am to 3 pm ET. Basically, Bovespa changes its trading hours in order to keep the highest possible intersection with the US market. The US market opens at 9:30 am and closes at 4 pm ET. Therefore, we have six hours and thirty minute-intersection on the majority of the year (mid February to mid November) and five hours and a half intersection in the remaining 3 months of the year. We believe this comes as a huge advantage compared to other studies in price discovery, which use data from Europe and US, ending up with only 2 to 3 hours of intersection.

We find that the foreign market is more important than the home market for the price discovery of Petrobras. We argue that this is related to the fact that Petrobras is a commodity exporter company. The common shares play a more significant role than the preferred shares in the US market and the opposite is true for the home market. Liquidity seems

to be responsible for this different behavior in Brazil. The Brazilian market has lost importance after the recent financial crises. The voting premium seems to be an integrated of order one process, however during the crises period it became a stationary process. It seems that investors stopped to price the voting rights as a financial asset during the crisis period and therefore, both common and preferred stocks share a single common factor. This is a behavior characteristic of crisis periods, where there is an increase in financial assets' correlation, making company fundamentals (or, as in our study, voting rights and stocks design) to matter less.

For Vale, the results regarding the foreign market are assured. The common and preferred shares have a similar role in the price discovery process before and after the crisis. This distinct result from Petrobras might well be explained by the different design of Vale preferred shares, which have some voting rights. However, the common shares have a much greater importance during the crisis. As in Petrobras, we find three common factors previous the crisis and subsequent to it, along with two common factors throughout the slump.

The remaining of the paper proceeds as follows: Section 2 presents institutional background and the main data features. Section 3 discusses how to deal with high frequency data, including cleansing and aggregation issues. Section 4 describes the estimation procedure. Section 5 and 6 document our empirical results for Petrobras and Vale, whereas Section 7 offers some concluding remarks. The appendix has our Monte Carlo exercises.

2 Institutional Background

The BM&FBovespa is the only stock exchange in Brazil and the leading exchange in Latin America in terms of number of contracts traded. In 2002, Bovespa bought equity membership of the Rio de Janeiro stock exchange (BVRJ) and in 2008, Bovespa merged with the Brazilian Mercantile & Futures Exchange (BM&F), forming the BM&FBovespa. It is a fully electronic exchange (end of open outcry transactions at Bovespa was in 2005 and derivatives transactions was in 2009) and operates under supervision of the CVM (Brazilian Securities Commission). BM&FBovespa markets include equity, commodities and futures, foreign exchange, securities and ETF's (exchange traded funds). The traded volume on the Bovespa segment in 2010 was USD 868 billion. It presented a market capitalization of USD 1.5 trillion in 2010, higher than many European stock exchanges, such as Deutsche Borse (USD 1.4 trillion), BME Spanish (USD 1.2 trillion) and SIX Swiss (USD 1.2 trillion) being one of largest stock exchanges in the world (top 10 in market capitalization). The London

Stock Exchange presents a market capitalization of USD 3.6 trillion

Brazil achieved the investment grade rating from Standard & Poor's in April 2008. Fitch and Moody's increased Brazilian rating in May 2008 and September 2009, respectively. IBOVESPA is the most important index at BM&FBovespa, and it has reflected the improvement in the Brazilian financial market regarding the ratings achievement. IBOVESPA growth rate in 2009 reached 82.7% in BRL terms and 145% in USD terms (partly due to the recovery from 2008 crisis). In 2010 the index was stable, with a growth rate of 1%. The IBOVESPA index is composed by the most traded stocks, being the main indicator of the Brazilian stock market's average performance. The 62 companies composing the IBOVESPA present a combined market value of USD 1.242 trillion, as in the end of December 2010. Both Petrobras and Vale are part of the IBOVESPA.

Petrobras is a Brazilian publicly-traded company. The majority stockholder is the Government of Brazil, holding 56% of the common shares. It performs as an energy company in the following sectors: exploration and production, refining, oil and natural gas trade and transportation, petrochemicals, and derivatives, electric energy, biofuel among others. It is a leader in the Brazilian oil industry, and aims to be among the top five integrated energy companies in the world by 2020.

Vale was founded as a public company by the Brazilian Government, being privatized in 1997. It is the second largest metals and mining company in the world and the biggest private sector company in Latin America. Vale has a market capitalization of around USD 160 billion. Vale is the world's largest producer of iron ore and iron ore pellets, where the majority of their revenues come from. It is also the world's second largest producer of nickel, beyond their production of manganese, ferroalloys, thermal and coking coal, copper, cobalt, platinum group metals and fertilizer nutrients. They also have operations on energy, logistics and steel, but not as being core business.

Foreign companies usually are traded and listed in the US market through American Depositary Receipts (ADR). An ADR is a physical certificate evidencing ownership of a US dollar denominated form of equity in a foreign company. It represents the shares of the company held on deposit by a custodian bank in the company's home country and carries the corporate and economic rights of the foreign shares, subject to the terms specified on the ADR certificate.

Petrobras and Vale common and preferred shares are listed at NYSE through ADR's on the highest level of an ADR programme a foreign company may sponsor. Our dataset

includes common and preferred shares of Petrobras and Vale traded at Bovespa, ADR's on common and preferred traded at NYSE and NYSE ARCA (this last one only for Petrobras). Common shares traded as ADR represent around 26% of total Petrobras common stocks while common shares traded in Bovespa are 8.5% during the period of our analysis. Looking at preferred shares, ADR's represent around 34% and shares traded in Bovespa are 36% of total Petrobras preferred stocks. For Vale, the ADR's on common shares represent 25%, whereas Bovespa is approximately 22%. For preferred shares, ADR's are 40%. Figure 1 shows the time intersection between Brazilian and US market during the year in Brazilian time. We have six hours and thirty minute-intersection on the majority of the year, delivering a significant advantage for the price discovery analysis compared to other studies in the literature.

3 Data Description

When measuring price discovery, one is interested in looking at which market incorporates news firstly. Intraday data turns to be of extreme relevance in this analysis. Low frequency data (daily data) will not provide the timely incorporation of new information in each different market. Differently, high frequency data provides this information. The use of high frequency data requires a few steps on handling the data before actually estimating the information share of each market, given that it may present microstructure effects, such as bid ask bounces, discreetness of price changes, etc.

First, we use the algorithm proposed by Brownlees and Gallo (2006). The idea of this filter is to remove outliers. High frequency data usually comes from stock exchanges as raw data, which might present entries containing errors, time ordering problems, etc. The algorithm is built in a way to discard observations which do not seem plausible compared to the usual market activity. The algorithm is constructed as follows:

$$\left(\left|p_i - \bar{p}_i\left(k,\delta\right)\right| < 3s_i\left(k\right) + \gamma\right) = \begin{cases} \text{true, observation } i \text{ is kept} \\ \text{false, observation } i \text{ is removed} \end{cases} \tag{1}$$

where $\bar{p}_i(k,\delta)$ is the δ-trimmed sample mean, $s_i\left(k\right)$ is the standard deviation of a neighborhood k around i and γ is a granularity parameter to avoid zero variances from a sequence of k equal prices. The higher the frequency of outliers, the higher should be the chosen δ. The choice of k depends on trading intensity and the granulity parameter is set equal to a multiple of the minimum price variation. The number of outliers are more sensible

Figure 1: Time Intersection (in Brazilian Time)

to changes in γ than in the other two parameters. γ is kept is kept fixed at the minimum price variation of 0.01. We use δ equal to 10% for all series. We specify k according to the number of trades for each stock, it ranges from 20 in the one with lowest number of trades to 60 in the one with the highest. We use different values for k, γ and δ as a robustness check and in order to find the best choice. Tables 1 and 2 give the initial number of entries for each of the price series, number of outliers and final observations after the filtering, for Petrobras and Vale, respectively.

The second step relies on dealing with the non synchronism of the dateset. As seen

Table 1: Cleaning Process - Petrobras

		Asset	Raw obs (M)	Outliers	Final obs (M)
		BRLUSD	4.09	600	4.09
Brazil	Pref	Petr4	9.07	7,353	9.06
	Com	Petr3	2.11	4,812	2.10
NYSE	Com	PBR.N	7.91	3,318	7.91
	Pref	PBRa.N	5.02	4,485	5.02
ARCA	Com	PBR.P	11.82	3,460	11.82
	Pref	PBRa.P	4.87	2,501	4.87
		Total	**44.90**	**26,529**	**44.87**

Raw accounts for initial number of observations (in million) for Exchange rate, preferred and common shares in the Brazilian market, common and preferred ADR's traded at NYSE and common and preferred ADR's traded at ARCA for the period from December 2007 to November 2009. Number of outliers using the filter in (1).

Table 2: Cleaning Process - Vale

		Asset	Raw obs (M)	Outliers	Final obs (M)
		BRLUSD	4.09	600	4.09
Brazil	Pref	Vale5	6.39	5,236	6.38
	Com	Vale3	2.07	8,139	2.06
NYSE	Com	RIO.N	6.93	1,159	6.93
	Pref	RIOp.N	3.58	1,823	3.58
		Total	**18.96**	**16,357**	**18.94**

Raw accounts for initial number of observations (in million) for preferred and common shares in the Brazilian market and common and preferred ADR's traded at NYSE for the period from December 2007 to November 2009. Number of outliers using the filter in (1).

Table 3: Aggregation - Petrobras

		Asset	Initial obs (M)	Missing obs (M)	Agg by 30s (M)	% Zero Returns
		BRLUSD	2.88	0.03	0.35	10%
Brazil	Pref	Petr4	7.79	0.00	0.35	1%
	Com	Petr3	1.94	0.10	0.35	28%
NYSE	Com	PBR.N	7.42	0.00	0.35	1%
	Pref	PBRa.N	4.80	0.02	0.35	5%
ARCA	Com	PBR.P	11.19	0.01	0.35	3%
	Pref	PBRa.P	4.63	0.04	0.35	13%
		Total	**40.66**	**0.21**	**2.47**	

Initial number of observations (in million) for each asset in the trading overlap period. After aggregation in intervals of 30 seconds, we present the missing observations and its percentage against the final number of observations.

in Tables 1 and 2, there are some assets with a much larger amount of trades in a given period, compared to others. We aggregate these series, based on a defined time interval of 30s and 60s (the last one as a robustness check). As the literature has reported there is a trade-off between market microstructure noise when working with tick by tick data (or very short intervals) and contemporaneous correlation (which increases as the frequency of dataset decreases). Grammig, Melvin and Schlag (2005) argue that 10 seconds is the optimal interval. As our dataset do not present such intensity of trades in some assets and our IS estimator is robust to contemporaneous correlation, we work with 30 second-interval. To aggregate the series into 30 second-interval we use the methodology proposed by Harris et al. (1995). For each interval, we identify the last exchange to have the first trade, and acquire the most recent trade from the other exchanges, forming the the first tuple, and so on (called in their paper as the 'replace all' method).

Tables 3 and 4 show the number of observations before and after the aggregation procedure for Petrobras and Vale, respectively. A concern arises regarding the common stocks traded at Bovespa (Petr3 and Vale3), since it presents many missing observations (intervals where there was no trade) that could lead to serial correlation. We use Newey-West to estimate the covariance matrix, which it showed to be not very sensitive to lags (even zero). We also perform a robustness check with 60 second-interval, in order to reduce the number of zero returns, and it does not yield any qualitative difference in the results.

Table 4: Aggregation - Vale

		Asset	Initial obs (M)	Missing obs (M)	Agg by 30s (M)	% Zero Returns
		BRLUSD	2.82	0.03	0.35	10%
Brazil	Pref	Vale5	6.38	0.01	0.35	2%
	Com	Vale3	2.06	0.09	0.35	25%
Nyse	Com	Rio.N	6.92	0.01	0.35	2%
	Pref	Riop.N	3.57	0.04	0.35	10%
		Total	**21.75**	**0.17**	**1.76**	

Initial number of observations (in million) for each asset in the trading overlap period. After aggregation in intervals of 30 seconds, we present the missing observations and its percentage against the final number of observations.

4 Methodology

To estimate a seven-variable model, allowing for common and preferred shares in the home and in the foreign markets, we extend the three variable model including exchange rate proposed by Grammig, Melvin and Schlag (2005). Hasbrouck (1995)'s information share methodology is implemented. Firstly, we estimate a vector error correction (VEC) model using full-information maximum likelihood. The VEC model is parameterized as in (2) and lag length is selected through Schwarz criterion.

$$\Delta y_t = \xi_1 \Delta y_{t-1} + \xi_2 \Delta y_{t-2} + ... + \xi_p \Delta y_{t-p} + \zeta + \xi_0 y_{t-1} + \epsilon_t, \tag{2}$$

where $\xi_0 = \alpha\beta'$, β is the cointegrating vector, α is the error correction term, y_t is an integrated of order one process and ϵ_t is a zero mean white noise process with covariance matrix equal to Ω. We impose restrictions on the constant term for absence of deterministic time trends. Once we have the parameters for the VECM, we back out the vector moving average (VMA) coefficients through dynamic simulation (see Hamilton, 1994). The VMA equation is given by

$$\Delta y_t = \epsilon_t + \psi_1 \epsilon_{t-1} + \psi_2 \epsilon_{t-2} + ... = \Psi(L)\epsilon_t, . \tag{3}$$

Applying Beveridge Nelson decomposition into (3) yields permanent component which is the relevant term to compute the information share.

$$y_t = y_0 + \psi \left(\sum_{s=1}^{t} \epsilon_s \right) \iota + \Psi^* (L) \epsilon_t \tag{4}$$

where ψ = common row vector of $\Psi(1) = I + \psi_1 + \psi_2 + ...$ and ι is a vector of ones. Considering the Hasbrouck (1995) set up, the term $\psi \left(\sum_{s=1}^{t} \epsilon_s \right)$ is common to all prices and

is seen as the efficient price. The variance of the innovation on the efficient price term $\psi\epsilon_t$ is equal to $\psi\Omega\psi'$.

Hasbrouck (1995) calls the market *j's* information share the ratio of the variance of innovation from market j ($\psi_j^2\Omega_{jj}$) to the total innovation variance ($\psi\Omega\psi'$).

Considering a model that does not have a common row on matrix $\Psi(1)$, the Information Share of each market is given by the equation (5), ending up with a square matrix on IS estimates, instead of a row:

$$IS_{ij} = \left(\left[\psi\Omega^{1/2}\right]_{ij}\right)^2 / \left(\psi\Omega\psi'\right)_{ii} \tag{5}$$

where ψ is equal to $\Psi(1) = I + \psi_1 + \psi_2 + ...$

These measures consider Ω being diagonal, which may not be true for empirical applications. In situations when there is correlation among markets, Hasbrouck (1995) proposes two solutions. The first one is to increase the frequency of the data, as with more and shorter intervals, correlation may be lower. The problem arises with the fact that this may reduce, but it does not eliminate contemporaneous correlation. Hasbrouck (1995) proposes a second solution, the use of Cholesky to decompose the covariance matrix. With this solution comes one of the main critiques on the IS estimates: the order of variables matters. Hence, the measure is not unique and the different ordering of variables returns upper and lower bounds. In one hand, these bounds can be very tight, not being an issue, as Grammig, Melvin and Schlag (2005) and Hasbrouck (2003) find. On the other hand, they can diverge, as Grammig and Peter (2010), Kim (2010a,b) and Huang (2002) observe. In this matter, Yan and Zivot (2010) point out that, when contemporaneous correlation is large, IS interpretation is difficult and Grammig and Peter (2010) say that when upper and lower bounds diverge, information share analysis is blurred.

In contrast, we employ the Spectral Theorem to factorize the matrix into a canonical form as a solution to tackle the non diagonal covariance matrix problem. Our proposed IS measure is given by

$$IS_{ij}^s = \left([\psi S]_{ij}\right)^2 / \left(\psi\Omega\psi'\right)_{ii} \tag{6}$$

where S is the squared root of Ω obtained from the eigenvalue decomposition of matrix Ω:

$$\Omega = V\Lambda V^{-1} \;\Rightarrow\; \Omega^{1/2} = V\Lambda^{1/2}V^{-1} \tag{7}$$

where the columns of V are the eigenvectors of Ω and Λ is a diagonal matrix with the respective eigenvalues.

We present a comparison of these two methodologies for the Information Share (IS) using Monte Carlo exercises in the appendix. Lien and Shrestha (2009) present an alternative methodology that delivers an order invariant measure for a two variable model. Our approach is simpler to compute, since in theirs, one needs to compute the correlation matrix and adjustments are necessary to compute the final IS measure.

5 Price Discovery for Petrobras

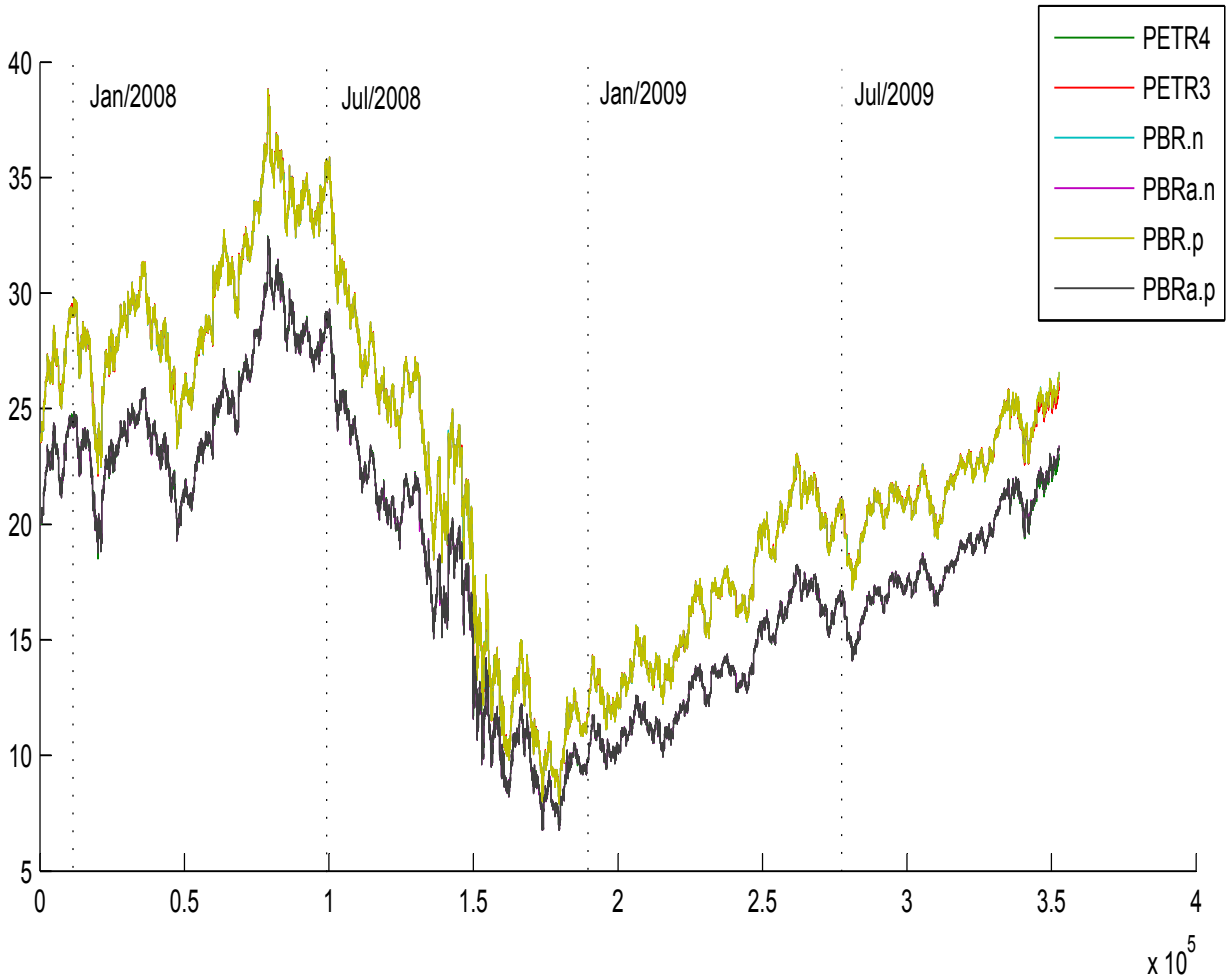
Our study covers six markets where Petrobras shares are traded: common (Petr3) and preferred (Petr4) at Bovespa, ADR of common and preferred shares at NYSE (PBR.n and PBRa.n, respectively) and common and preferred at ARCA (PBR.p and PBRa.p, respectively). In the Brazilian market, we see a large difference in volume of trades between the two stocks. The preferred shares are much more liquid than the common shares (see Table 3). We expect this to have an impact on how the common shares incorporate information. Once a new information arises, if no trades happen at a given market, this market will not be able to reflect this change in prices as fast as the ones that have transactions being executed.

ARCA is a trading platform (former ECN - Electronic Communication Network) that belongs to NYSE. It has a smarter router system. This router is able to check among other exchanges if there is a better quote than the one at ARCA. If this is the case, it executes the order at the venue where the best quote is available. We expect this to result in a more important role for ARCA than NYSE in the price discovery process. We also expect an impact on the IS parameters regarding the relatively high amount of common shares traded in the US market when compared to the Brazilian market and to the total tradeable shares.

We try to understand how the price discovery measure works throughout the crisis period. Therefore, we have four sets of IS estimates: first and second half of 2008 and of 2009 [1]. Figures 2 and 3 show graphically these four periods with all the stocks in their original currencies and in US dollar. The financial crisis reached Brazil in the second half of 2008, where there was a significant drop in prices. Our goal here is to understand what consequences this change has to the price discovery process. All the IS estimates we report rely on the order invariant spectral based decomposition.

[1] Second half of 2009 does not include December.

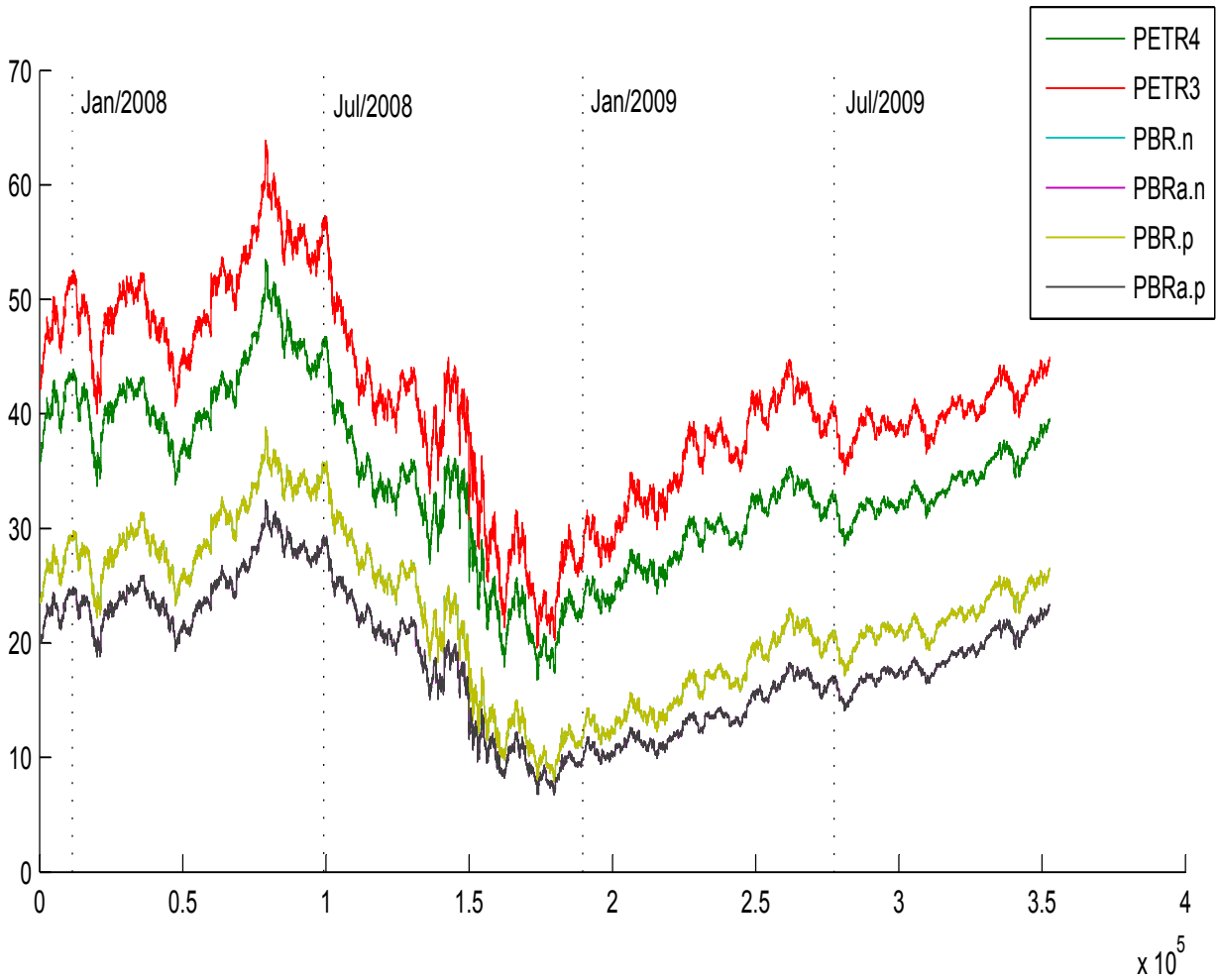
Figure 2: Price Evolution Petrobras - US dollars



All assets priced in US dollars and ADR's converted to one equivalent share.

Table 5 gives the results for the first half of 2008. We find four cointegrating vectors, which lead to three common factors. These common factors are assigned to be the efficient exchange rate, efficient unobserved price and voting premium. The voting premium seems not to be a stationary process. The third cointegrating vector shows that only getting the preferred share minus the common share in the Brazilian market does not yield into a stationary process. It is also necessary to add the voting premium from the US market in order to get a stationary process. This may come as a surprise, since one might expect to find a constant voting premium. However, considering some literature on the topic (Zingales, 1994, 1995), the voting premium should be a function of the controlling premium one may get in case of a merger or acquisition. Also, it should be related to the private benefits and asymmetric information one may have when controlling the company or voting for

Figure 3: Price Evolution Petrobras - Original Currency



All assets priced in the original currency and ADR's converted to one equivalent share.

influential decisions. The premium could be seen as a proportion of the stock price itself, and hence, it would be an integrated of order 1 process. Moreover, the private benefits could be priced as an integrated of order 1 process. Other points that may affect the voting premium are credit risk, dividend policy, voting restrictions and liquidity of the two shares.

Still in Table 5, the information share estimates for Petr4 are much higher than the ones for Petr3. This feature arises from the fact that the preferred share is traded with a much greater intensity than the common, affecting the way and the speed in which Petr3 incorporates new information. This leads us to the conclusion that liquidity matters on determining price discovery.

As expected, ARCA is impounding more information into the securities prices than NYSE. This comes from the fact that ARCA has a smart order router which identifies the

markets with the best prices continuously and also transfers orders to those markets. That is, if an order is sent to ARCA, a check may be made of where it would best be executed (ARCA, Nasdaq, other local exchanges or electronic communication networks). Once the venue with the best price is identified (it might take only a fraction of a second), the order is sent to this venue. A second issue regarding the US market comes from the common stock, which is much more important than the preferred in the price discovery process. The common shares traded through ADR represent 20% of the total shares, which is a considerably amount, specially when considering that 54% already belong to the Brazilian Government.

At a first glance, a surprising result arises from the event that exchange rate is not exogenous. This is in contrast to Grammig, Melvin and Schlag (2005), who find exogenous exchange rate. We claim that Petrobras is a commodity exporter company, which may explain this result. We see a strong correlation between the exchange rate and international oil returns (higher than 0.42 for this period), which corroborates our finding of non exogeneity of the exchange rate. Regarding exchange rate shocks, the US market is the one to incorporate that.

During the second half of 2008, the financial crisis affects Brazil. The home currency devaluates 58% from beginning of August to middle of November against the US dollar, recovering value during the year of 2009. Investments decisions start to be postponed and canceled by companies. By the end of the year, Brazilian most important financial index losses almost half of its value, when compared to the highest pick during the year.

Table 6 shows the IS estimates for this period. There is a significant decrease in the stock's price in all markets. There are now five cointegrating vectors and only two common factors, namely efficient exchange rate and efficient price. Voting premium seems to be a stationary process now. The fifth cointegrating vector also entails this conclusion, since we are able to have a stationary process by getting preferred minus common share. This is a considerable difference from the results in the previous six months. The voting premium seems not to have a pricing process anymore. Therefore, preferred and common shares respond in the same way to innovation in other markets, differently than in the first half of 2008. Additionally, the Brazilian market loses a lot of importance in this period, specially the preferred stock, which had a considerable role on the price discovery process. At the same time, ARCA gains even more importance. This could be the result of a change in the pool of investors, with an increase of foreign investors. Exchange rate shocks are now

incorporated by the Brazilian market and exchange rate is even less exogenous, which might be an effect of the Brazilian currency considerably devaluation during this crisis period.

Table 7 displays the results for the first half of 2009, which are quite similar from the second half of 2008: only two common factors and Brazilian market loosing importance.

Finally, Table 8 shows how the information shares estimates behave after this turbulent period. We are back to the situation prior the crisis, where we have three common factors and the voting premium does not seem to be a stationary process. The Brazilian market looses even more importance when compared to both first and second half of 2008. We see a shift on information coming from the Brazilian market to information originated in the US. Additionally, NYSE gains a more significant role on the price discovery process as well as the preferred shares in the US market. NYSE is the exchange where block trades happen, hence we see this gain in importance as an effect of the increase of pension funds that started to invest in the Brazilian market given the investment grade rating given by Moodys in September 2009. Exchange rate is still endogenous, but less so, similar to the results in the first half of 2008.

As a robustness check, we estimate the IS measures with the data set sampled at 60 seconds. The results we get are very similar from the ones with 30 seconds. We also estimate in years and the whole sample period (2008 - 2009). The IS estimates do not change significantly and confirm the results found on the six month sample: 1.Foreign market being more important to the price discovery process, in particular ARCA. 2.Common share in the foreign markets more responsible for impounding information into securities prices, and the opposite for the Brazilian market. 3.Overall importance for common shares in US market.

Table 5: First half 2008

	Information Share							Cointegrating Vector			
	BRLUSD	Petr4	Petr3	PBR.N	PBRa.N	PBR.P	PBRa.P				
BRLUSD	0.53 (0.027)	0.00 (0.002)	0.04 (0.007)	0.05 (0.008)	0.07 (0.009)	0.16 (0.017)	0.15 (0.014)	-0.02	0.00	0.01	1.00
Petr4	0.00 (0.001)	0.38 (0.025)	0.03 (0.007)	0.12 (0.015)	0.10 (0.012)	0.25 (0.023)	0.11 (0.015)	0.02	0.00	1.00	-0.01
Petr3	0.00 (0.002)	0.23 (0.019)	0.06 (0.01)	0.20 (0.023)	0.06 (0.008)	0.39 (0.026)	0.06 (0.01)	0.00	0.00	-1.00	-1.00
PBR.N	0.02 (0.004)	0.16 (0.017)	0.03 (0.006)	0.20 (0.023)	0.07 (0.01)	0.42 (0.027)	0.09 (0.014)	1.00	0.00	1.01	1.00
PBRa.N	0.04 (0.006)	0.26 (0.021)	0.01 (0.004)	0.13 (0.016)	0.12 (0.014)	0.29 (0.024)	0.15 (0.017)	0.00	1.00	-1.00	0.00
PBR.P	0.02 (0.004)	0.16 (0.017)	0.03 (0.006)	0.20 (0.023)	0.07 (0.01)	0.42 (0.027)	0.09 (0.014)	-1.00	0.00	0.00	-0.04
PBRa.P	0.04 (0.006)	0.26 (0.021)	0.01 (0.004)	0.13 (0.016)	0.12 (0.014)	0.29 (0.024)	0.15 (0.017)	-0.02	-1.00	0.00	0.04

'4'= Preferred, 'a'= Preferred, 'N'= NYSE and 'P'= ARCA. T = 87,781 obs. The bootstrap standard errors are in parenthesis.

Table 6: Second half 2008

	Information Share							Cointegrating Vector				
	BRLUSD	Petr4	Petr3	PBR.N	PBRa.N	PBR.P	PBRa.P					
BRLUSD	0.38 (0.071)	0.04 (0.027)	0.00 (0.006)	0.09 (0.045)	0.04 (0.029)	0.29 (0.105)	0.14 (0.082)	0.00	0.00	1.00	1.00	-0.04
Petr4	0.04 (0.016)	0.14 (0.056)	0.05 (0.017)	0.17 (0.053)	0.07 (0.037)	0.36 (0.105)	0.17 (0.094)	0.00	0.00	-0.01	-1.00	0.00
Petr3	0.04 (0.015)	0.14 (0.057)	0.05 (0.017)	0.17 (0.053)	0.07 (0.037)	0.36 (0.105)	0.17 (0.093)	0.00	0.00	-1.00	-0.01	0.00
PBR.N	0.00 (0.007)	0.05 (0.032)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.107)	1.00	0.00	0.00	0.00	1.00
PBRa.N	0.01 (0.007)	0.05 (0.031)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.108)	0.00	1.00	0.00	0.00	-1.00
PBR.P	0.00 (0.007)	0.05 (0.032)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.107)	-1.00	0.00	1.00	0.00	0.00
PBRa.P	0.01 (0.007)	0.05 (0.031)	0.02 (0.013)	0.19 (0.057)	0.08 (0.041)	0.45 (0.114)	0.21 (0.108)	0.00	-1.00	0.00	1.00	-0.04

'4'= Preferred, 'a'= Preferred, 'N'= NYSE and 'P'= ARCA. T = 90,431 obs. The bootstrap standard errors are in parenthesis.

Table 7: First half 2009

	Information Share							Cointegrating Vector				
	BRLUSD	Petr4	Petr3	PBR.N	PBRa.N	PBR.P	PBRa.P					
BRLUSD	0.39 (0.042)	0.00 (0.017)	0.02 (0.004)	0.10 (0.022)	0.10 (0.031)	0.14 (0.044)	0.25 (0.056)	0.00	0.00	1.00	1.00	0.00
Petr4	0.00 (0.002)	0.08 (0.053)	0.05 (0.008)	0.26 (0.037)	0.06 (0.024)	0.41 (0.078)	0.14 (0.042)	0.00	0.00	-0.03	-1.00	1.00
Petr3	0.00 (0.004)	0.09 (0.056)	0.06 (0.01)	0.26 (0.037)	0.06 (0.022)	0.41 (0.078)	0.12 (0.038)	0.00	0.00	-1.00	-0.01	-1.00
PBR.N	0.03 (0.01)	0.06 (0.048)	0.02 (0.005)	0.24 (0.037)	0.08 (0.027)	0.38 (0.075)	0.18 (0.047)	1.00	0.00	0.00	0.00	0.00
PBRa.N	0.06 (0.016)	0.05 (0.044)	0.01 (0.004)	0.23 (0.035)	0.09 (0.029)	0.35 (0.072)	0.21 (0.05)	0.00	1.00	0.00	0.00	0.00
PBR.P	0.03 (0.01)	0.06 (0.048)	0.02 (0.005)	0.24 (0.037)	0.08 (0.027)	0.38 (0.075)	0.18 (0.047)	-1.00	0.00	1.00	0.00	0.97
PBRa.P	0.06 (0.016)	0.05 (0.044)	0.01 (0.004)	0.23 (0.035)	0.09 (0.029)	0.35 (0.072)	0.21 (0.05)	0.00	-1.00	0.02	1.00	-0.97

'4'= Preferred, 'a'= Preferred, 'N'= NYSE and 'P'= ARCA. T = 87,797 obs. The bootstrap standard errors are in parenthesis.

Table 8: Second half 2009 (without December)

	Information Share							Cointegrating Vector			
	BRLUSD	Petr4	Petr3	PBR.N	PBRa.N	PBR.P	PBRa.P				
BRLUSD	0.47 (0.072)	0.00 (0.01)	0.01 (0.011)	0.13 (0.015)	0.09 (0.017)	0.15 (0.019)	0.15 (0.021)	0.01	0.00	0.00	1.00
Petr4	0.02 (0.014)	0.09 (0.046)	0.01 (0.005)	0.21 (0.014)	0.18 (0.017)	0.24 (0.017)	0.25 (0.018)	-0.01	0.00	1.00	-0.17
Petr3	0.03 (0.016)	0.03 (0.025)	0.02 (0.013)	0.33 (0.017)	0.09 (0.014)	0.37 (0.018)	0.13 (0.015)	0.00	0.00	-1.00	-1.00
PBR.N	0.01 (0.008)	0.02 (0.019)	0.01 (0.007)	0.32 (0.017)	0.12 (0.015)	0.36 (0.018)	0.18 (0.016)	1.00	0.00	1.01	1.00
PBRa.N	0.02 (0.012)	0.05 (0.033)	0.00 (0.005)	0.22 (0.015)	0.19 (0.018)	0.25 (0.017)	0.27 (0.019)	0.00	1.00	-1.00	0.00
PBR.P	0.01 (0.008)	0.02 (0.019)	0.01 (0.007)	0.32 (0.017)	0.12 (0.015)	0.36 (0.018)	0.18 (0.016)	-1.00	0.00	0.00	0.03
PBRa.P	0.02 (0.012)	0.05 (0.033)	0.00 (0.005)	0.22 (0.015)	0.19 (0.018)	0.25 (0.017)	0.27 (0.019)	0.01	-1.00	0.00	0.06

'4'= Preferred, 'a'= Preferred, 'N'= NYSE and 'P'= ARCA. T = 75,010 obs. The bootstrap standard errors are in parenthesis.

6 Price Discovery for Vale

We estimate a 5-variable model for Vale, including common and preferred shares traded in the Brazilian market (Vale3 and Vale5, respectively) and in the US market at NYSE (RIO.N and RIOp.N, respectively), plus exchange rate. We expect to have a similar behavior for Vale and Petrobras, since both companies trade commodities and are greatly important players worldwide in their core business. Preferred shares at Vale are designed in a slightly different way than the ones at Petrobras. They are denominated as "class A" and each preferred share gives the right of one vote in deliberations at the General Assemblies, exactly as the common shares. Preferred shares have also the same political rights as the common shares, apart from the right to vote for the election of members in the Administration Board. This special characteristics may cause an effect on our IS estimates for the preferred shares, such as an increase in its importance.

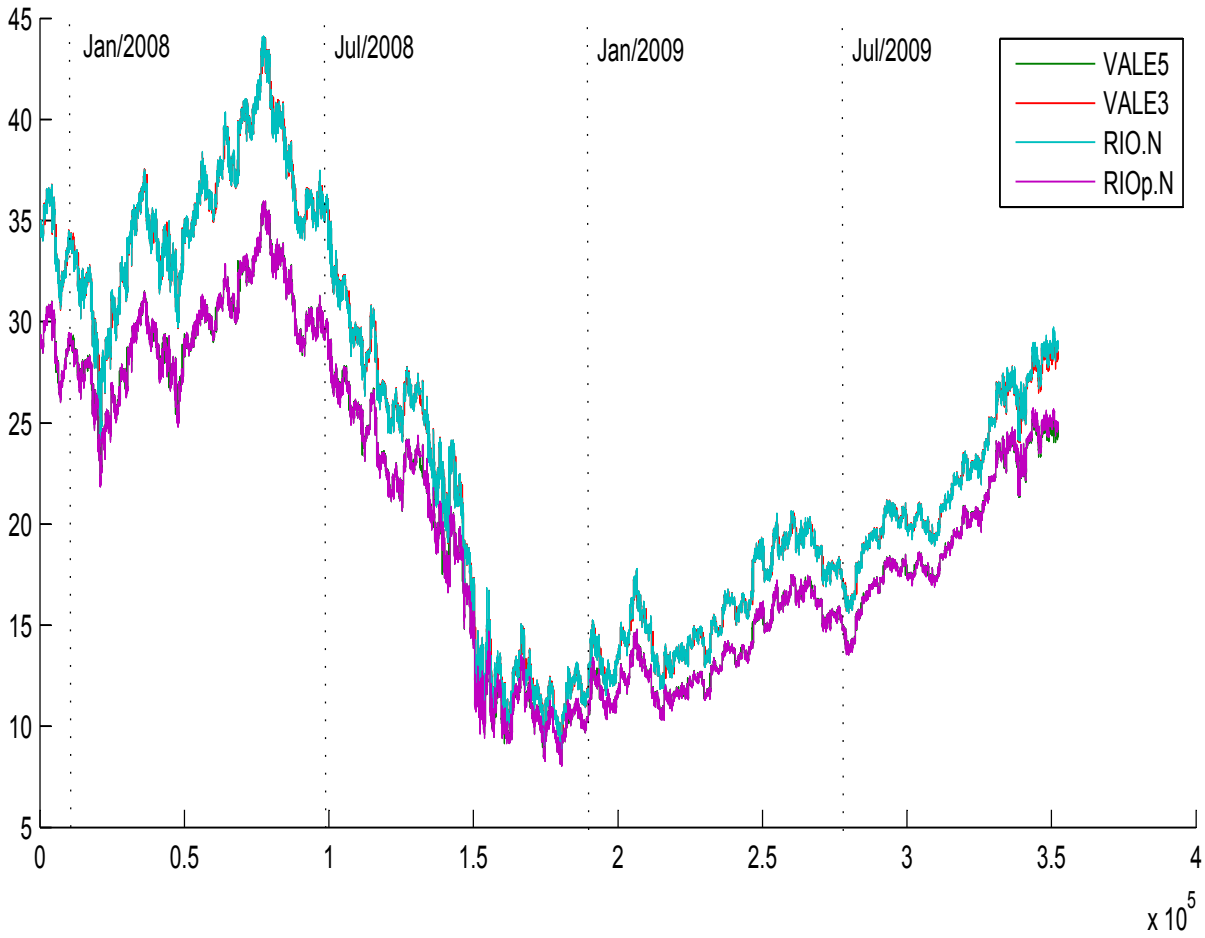
Our sample period time is the same as for Petrobras and we present the estimates in a six-month basis. Tables 9 to 12 state the results. As in Petrobras, we find three common factors in the first half of 2008, which are seen as the efficient price, efficient exchange rate and voting premium. The US market is the one with a principal role in the price discovery process, presenting a similar pattern for common and preferred shares. Shocks on the US traded common shares are highly incorporated by the other stocks, specially the common ones and the same behavior is true for shocks on the US preferred shares: high parameters for all stocks, but even higher for the preferred ones. Additionally, the preferred shares at the US market have a slightly grater importance than the common, which goes in contrast from what we see at Petrobras. This feature might be explained by the different design of the preferred stocks of Vale and Petrobras.

Common and preferred shares' roles are quite similar for the price discovery process in the Brazilian market and their importance in impounding information is much lower than their counterparts in the US market. The exchange rate is not exogenous and the US market is the one to be impacted by exchange rate shocks, while its parameters for the Brazilian market are close to zero. Again, this is not a surprise if we take into account the high correlation among exchange rate and metal returns (0.53 for this period, considering the S&P Industrial Metals Spot index).

In the second half of 2008, we find 3 cointegrating vectors, and only two common factors. This change is the same we find for Petrobras, showing a similar behavior for the two stocks during the crisis period. An interesting change comes from the much higher IS estimates

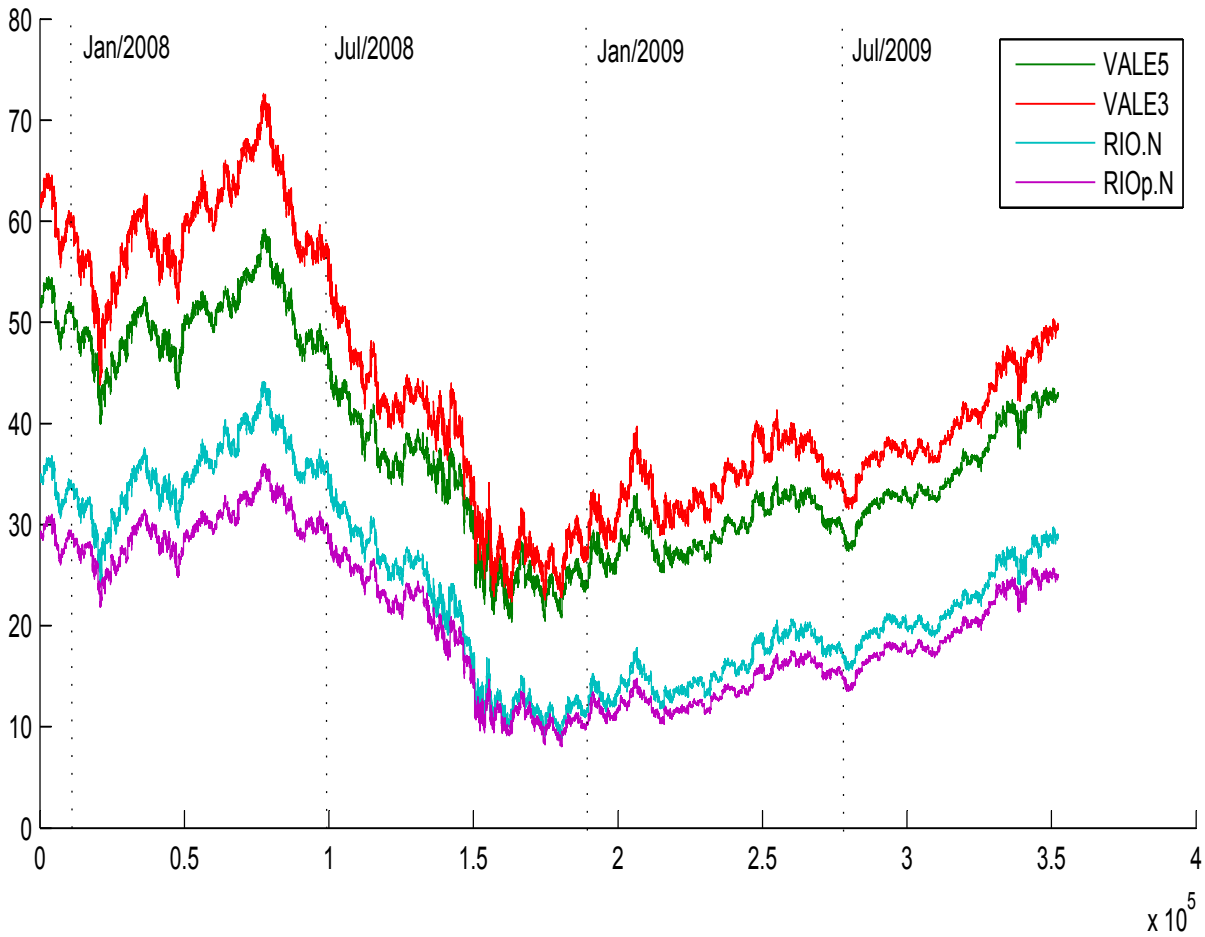
assigned for the common share in the US market, accompanied by lower estimates for the preferred shares in the same market. In Brazil, we also observe this pattern, but in a much smaller dimension. At a period of instability, the common shares are more important and faster in incorporating news than the preferred shares. Figure 4 displays the price evolution for Vale shares in USD dollars and figure 5 is in their original currency. From both graphs, the period of instability can be seen with a huge drop in prices, followed by a slow recovery afterwards.

Figure 4: Price Evolution Vale - US Dollars



All assets priced in US dollars.

Figure 5: Price Evolution Vale - Original Currency



All assets priced in the original currency. Brazilian shares are priced in Reais and ADR's in USD.

Analyzing the first half of 2009, there are still only two common factors and the common share in the US market is still the most important one for the price discovery process, although the preferred shares present higher IS estimates than in the previous six months.

In the second half of 2009, we are back to two cointegrating vectors, three common factors and the preferred shares recover all its importance. The Brazilian market does not recover the importance it had before the turbulent period and looses it even more, just as found for Petrobras.

We perform some robustness check in yearly basis and for the whole sample period (2008 - 2009). The IS estimates do not change significantly and confirm the results found on six month basis: 1.Foreign market being more important to the price discovery process. 2.Common and preferred shares with a similar role in impounding information into securities

prices. 3. Non exogeinity of the exchange rate. 4.Three common factors, seen as unobserved efficient price, efficient exchange rate and voting premium.

Table 9: First half 2008

	Information Share					Cointegrating Vector	
	BRLUSD	Vale5	Vale3	RIO.N	Riop.N		
BRLUSD	0.42 (0.039)	0.01 (0.006)	0.00 (0.002)	0.21 (0.021)	0.36 (0.029)	1.00	0.00
Vale5	0.00 (0.003)	0.24 (0.028)	0.14 (0.019)	0.22 (0.02)	0.41 (0.029)	0.04	1.00
Vale3	0.00 (0.003)	0.10 (0.018)	0.25 (0.026)	0.42 (0.026)	0.23 (0.023)	-1.05	-0.99
RIO.N	0.02 (0.007)	0.06 (0.014)	0.18 (0.022)	0.45 (0.026)	0.30 (0.026)	1.02	0.99
Riop.N	0.02 (0.008)	0.14 (0.022)	0.09 (0.017)	0.27 (0.021)	0.48 (0.031)	0.00	-0.99

'5'= Preferred, 'p'= Preferred, 'N'= NYSE. T = 87,966 obs. The bootstrap standard errors are in parenthesis.

Table 10: Second half 2008

	Information Share					Cointegrating Vector		
	BRLUSD	Vale5	Vale3	RIO.N	Riop.N			
BRLUSD	0.54 (0.05)	0.04 (0.032)	0.01 (0.006)	0.32 (0.053)	0.10 (0.07)	1.00	-1.02	0.14
Vale5	0.00 (0.002)	0.03 (0.032)	0.23 (0.026)	0.71 (0.064)	0.03 (0.034)	0.00	1.00	1.00
Vale3	0.00 (0.005)	0.04 (0.036)	0.24 (0.027)	0.69 (0.065)	0.02 (0.03)	-0.99	0.00	-0.84
RIO.N	0.05 (0.016)	0.01 (0.018)	0.17 (0.022)	0.72 (0.059)	0.05 (0.047)	0.99	0.00	0.00
Riop.N	0.09 (0.022)	0.00 (0.014)	0.14 (0.021)	0.71 (0.058)	0.06 (0.051)	0.00	-1.01	0.00

'5'= Preferred, 'p'= Preferred, 'N'= NYSE. T = 90,143 obs. The bootstrap standard errors are in parenthesis.

Table 11: First half 2009

	Information Share					Cointegrating Vector		
	BRLUSD	Vale5	Vale3	RIO.N	Riop.N			
BRLUSD	0.40 (0.043)	0.04 (0.044)	0.01 (0.012)	0.26 (0.051)	0.29 (0.083)	1.00	-1.00	0.16
Vale5	0.01 (0.006)	0.00 (0.02)	0.27 (0.035)	0.48 (0.064)	0.23 (0.076)	0.00	1.00	1.00
Vale3	0.02 (0.009)	0.01 (0.022)	0.29 (0.036)	0.47 (0.064)	0.22 (0.072)	-1.02	0.00	-0.78
RIO.N	0.01 (0.004)	0.00 (0.016)	0.22 (0.032)	0.49 (0.063)	0.28 (0.085)	1.01	0.00	0.00
Riop.N	0.02 (0.009)	0.00 (0.017)	0.19 (0.031)	0.49 (0.062)	0.30 (0.089)	0.00	-1.00	0.00

'5'= Preferred, 'p'= Preferred, 'N'= NYSE. T = 86,256 obs. The bootstrap standard errors are in parenthesis.

Table 12: Second half 2009 (without December)

	Information Share					Cointegrating Vector	
	BRLUSD	Vale5	Vale3	RIO.N	Riop.N		
BRLUSD	0.49 (0.046)	0.00 (0.006)	0.01 (0.012)	0.19 (0.02)	0.31 (0.045)	1.00	0.00
Vale5	0.00 (0.003)	0.19 (0.041)	0.08 (0.029)	0.28 (0.022)	0.44 (0.052)	0.15	1.00
Vale3	0.00 (0.005)	0.08 (0.026)	0.11 (0.035)	0.49 (0.027)	0.31 (0.042)	-1.36	-1.02
RIO.N	0.02 (0.009)	0.05 (0.021)	0.09 (0.032)	0.48 (0.026)	0.36 (0.045)	1.13	0.99
Riop.N	0.04 (0.011)	0.11 (0.031)	0.07 (0.027)	0.33 (0.022)	0.46 (0.052)	0.00	-0.97

'5'= Preferred, 'p'= Preferred, 'N'= NYSE. T = 76,311 obs. The bootstrap standard errors are in parenthesis.

7 Price Discovery and Corporate Ownership

There is a vast literature on measuring voting rights or on determining what affects voting premium. There is also an extensive literature on measuring price discovery on different markets. However, there is no literature, as far as our knowledge goes, that relates these two topics. The Brazilian market enhances this type of study, as companies over there face the possibility of issuing a large number of preferred shares (two preferred shares for each one common before 2001, and one preferred for each one common afterwards), allowing the preferred shares to have enough liquidity and hence, be target of study. The Brazilian markets has a peculiar situation where 1% percent of voting rights may correspond to a much lower percentage of cash flow rights for many companies, once it is usual for them to issue preferred shares. For Petrobras, for instance, the relation between number of common shares over number of preferred shares is 0.75. For Vale, this relation is 0.65. We find that common shares are more important to the price discovery process of Petrobras in the US market, and the reason why this is not true for the Brazilian market is due to liquidity. However, in terms of the overall process, the common shares in US are more responsible in impounding information when compared to any other share. This happens even though the common shares carry a lower percentage of cash flow rights compared to voting rights. Hence, one might conclude that voting rights carry more information than cash flow rights. We corroborate this finding looking at Vale results. The fact that preferred shares have similar voting rights as the common ones makes these two types of shares to contribute in an alike way to the price discovery process. This conclusion is strengthened by the shift in importance between common and preferred shares during the crises period, where common shares gain even higher weight in the price discovery process.

8 Conclusion

We expand IS framework to estimate a 7-variable model that accounts for the information content of both Common and Preferred stocks (Brazilian market), their ADRs counterparts traded on NYSE and ARCA, and Exchange rate for Petrobras and for Vale (5-variable model), the two Brazilian largest companies. Our goal is to gather all possible information for the price discovery process. We propose a new framework which yields an unique measure for the IS. In particular, this is most suitable when the covariance matrix of innovations is not diagonal. This single measure becomes a huge advantage, specially when working with a higher number of variables. In our case, analyzing all possible permutations to get upper

and lower bounds would result in an average of 5,040 different measures.

We find that foreign market is more important than the home market for the price discovery of two commodity exporter companies. This comes as a different result than the ones previously found in the literature. For Petrobras, the preferred shares play a more significant role than the common shares in the Brazilian market and the opposite is true for the US market. Liquidity seems to be responsible for this behavior in the Brazilian market. For Vale, the specific characteristics of its preferred shares give a different result: they are slightly more important than the common shares before and after the crisis, but not during the instable period. Regarding the voting premium, we find similar conclusions for both Petrobras and Vale. During the first half of 2008 and second half of 2009, the voting premium is an integrated of order one process, meaning that investors price voting rights as a financial asset. During the crisis, however, we find that voting premium resumes to be an integrated of order zero process, and both common and preferred stocks share a single common factor. This behavior is characteristic of crisis periods, where financial assets' correlation increases significantly, making investors not to price as much company fundamentals (in our case, stocks design) as during non crisis period. Finally, the Brazilian market has lost significant importance during the 2008/2009 financial crisis and it does not recover its importance afterwards, movement corroborated by Petrobras and Vale results.

References

[1] Baillie, Richard T., Booth, G. Geoffrey, Tse, Yiuman, Zabotina, Tatyana, 2002, Price Discovery and common factor models, *Journal of Financial Markets* 5, 309-321.

[2] Booth, G. Geoffrey, So, Raymond W., 1999, Price Discovery in the German Equity Index Derivatives Markets, *The Journal of Futures Markets* 19, No 6, 619-643.

[3] Brownlees, C.T. , Gallo, G.M., 2006, Financial econometric analysis at ultra-high frequency: Data handling concerns,*Computational Statistics & Data Analysis* 51, 2232-2245.

[4] Chu, Quentin C., Hsieh, Wen-liang G. and Tse, Yiuman, 1999, Price Discovery on the S&P 500 index markets: An analysis of spot index, index futures and SPDRs, *International Review of Financial Analysis* 8, No 1, 21-34.

[5] De Jong, Frank, 2002, Measures of contributions to price discovery: a comparison, *Journal of Financial Markets* 5, 323-327.

[6] Gonzalo, Jesus and Granger, Clive, 1995, Estimation of Common Long-Memory Components in Cointegrated Systems, *Journal of Business & Economic Statistics*, 13 (January 1995), No 1, 27-35.

[7] Grammig, Joachim, Melvin, Michael and Schlag, Christian, 2005, Internationally cross-listed stock prices during overlapping trading hours: price discovery and exchange rate effects, *Journal of Empirical Finance* 12, 139-164.

[8] Grammig, Joachim and Peter, Franziska, J., 2010, Tell-Tale Tails A data driven approach to estimate unique market information shares, working paper.

[9] Hamilton, James D., 1994, Time Series Analysis, *Princeton University Press*, 1994.

[10] Harris, Frederick, H deB, McInish, Thomas H, Shoesmith, Gary L. and Wood, Robert, A., 1995, Cointegration, Error Correction, and Price Discovery on Informationally Linked Security Markets, *Journal of Financial and Quantitative Analysis* 30, No 4, 563-578.

[11] Harris, Frederick, H deB, McInish, Thomas H, and Wood, Robert, A., 2002, Common factor components versus information shares:a reply, *Journal of Financial Markets* , No 5, 341-348.

[12] Hasbrouck, Joel, 1995, One Security, Many Markets: Determining the Contributions to Price Discovery, *The Journal of Finance* 50, No 4 (Sep., 1995), 1175-1198.

[13] Huang, Roger D., 2002, The Quality of ECN and Nasdaq Market Maker Quotes, *The Journal of Finance* 57, No 3 (Jun., 2002), 1285-1319.

[14] Kim, Cherl Hyun, 2010a, Price Discovery in Crude Oil Prices, *Working Paper*

[15] Kim, Cherl Hyun, 2010b, Cross-Border Listing and Price Discovery: Canadian Blue Chips Traded in TSX and NYSE, *Working Paper*

[16] Lehmann, Bruce N., 2002, Some desiderata for the measurement of price discovery across markets, *Journal of Financial Markets* 5, 259-276.

[17] Lien, Donald. and Shrestha, Keshab, 2009, A new information share measure, *The Journal of Futures Markets* 29, No 4, 377-395.

[18] Yan, Bingcheng and Zivot, Eric, 2010, A structural analysis of price discovery measures, *Journal of Financial Markets* 13, 1-19.

[19] Zingales, Luigi, 1994, The value of the voting right: A study of the Milan Stock Exchange experience, *The Review of Financial Studies* 7, No 1, 125-148.

[20] Zingales, Luigi, 1995, What determines the value of corporate votes?, *The Quarterly Journal of Economics* Nov, 1995, 1047-1073.

9 Appendix

9.1 Monte Carlo Simulations

The main purpose of this section is to compare the Cholesky and Spectral-based decompositions of the covariance matrix. We create three different data generation process (DGP). The simpler case (denoted by MC1) includes three variables: the log of the exchange rate (E_t), the log of the stock's price in the home market (P_t^h) and the log of the stock price in the foreign market (P_t^f). In this setup, P_t^f follows the home market, whereas exchange rate is exogenous.

$$
\begin{aligned}
E_t &= E_{t-1} + u_t \\
P_t^h &= P_{t-1}^h + v_t \\
P_t^f &= P_{t-1}^h + E_{t-1} + w_t
\end{aligned} \tag{8}
$$

where (u_t, v_t, w_t) are innovations with zero mean and serially uncorrelated. In following two DGP's (denoted by MC2 and MC3), we include the log of common ($P_{c,t}$) and preferred ($P_{p,t}$) stocks in the home and foreign market. We add a constant voting power (vp) in the common stock at home and international markets. In MC2, the voting power is set to be equal for the home and foreign common stocks, whereas in MC3 they are set to be different. What differentiates the DGP's, besides the number of variables, is the nature of dependency among stocks' prices in the two markets. In MC2, we assume common and preferred prices are generated independently at home, and the preferred and common in the foreign market follow, respectively, the preferred and common in the home market, namely,

$$
\begin{aligned}
E_t &= E_{t-1} + u_t \\
P_{p,t}^h &= P_{p,t-1}^h + v_t \\
P_{c,t}^h &= P_{c,t-1}^h + e_t \\
P_{p,t}^f &= P_{p,t-1}^h + E_{t-1} + w_t \\
P_{c,t}^f &= P_{c,t-1}^h + E_{t-1} + n_t.
\end{aligned} \tag{9}
$$

The third DGP considers the common stock price following the preferred at home mar-

ket, which will lead to the fact that two remaining prices also follow the home preferred.

$$
\begin{aligned}
E_t &= E_{t-1} + u_t \\
P^h_{p,t} &= P^h_{p,t-1} + v_t \\
P^h_{c,t} &= P^h_{p,t-1} + vp + e_t \\
P^f_{p,t} &= P^h_{p,t-1} + E_{t-1} + w_t \\
P^f_{c,t} &= P^h_{p,t-1} + vp + E_{t-1} + n_t
\end{aligned} \tag{10}
$$

We assume that the disturbances in all MC exercises are normally distributed with zero mean and variance equal to one. We perform 1,000 replications and we fix sample size equal to 10,000 observations. We also discard the first 500 observations to alleviate the dependence on initial values.

We consider two set of simulations for all DGP's. In the first one, the covariance matrix is diagonal and hence the use of Cholesky does not impact the IS when we change the variables' order. This is what we could expect to happen in empirical applications using high frequency data, since the correlation among markets reduces as frequency increases (Hasbrouck, 1995). This is in line with the results by Grammig, Melvin and Schlag (2005), whose upper and lower bounds for the IS became very similar using high frequency data sampling (10-s intervals). However, this is not what Grammig and Peter (2010), Kim (2010a,b) and Huang (2002) find. We will deal with this case in our second set of simulations, where the covariance matrix is non diagonal.

Table 13 reports results on the first set of simulations with a diagonal covariance matrix for MC1. The first three columns show the theoretical value of the Information Share computed using the Spectral Theorem. Next, we present the results of the IS estimates (mean and standard deviation among all simulations) using the Spectral Theorem. Finally, there are the IS results using the Cholesky decomposition of the covariance matrix. As expected, since we impose a diagonal covariance matrix, Cholesky and spectral based decomposition do not have any considerable difference.

The two methodologies we use to decompose the matrix perform equally well for all DGP's where innovation terms are uncorrelated. This is exactly what we would expect given that we are working with a diagonal covariance matrix, hence Cholesky decomposition does not impose any actual restriction to the matrix. As a matter of space, we do not report the results for MC2 and MC3 since they are all equal to the theoretical value, as in MC1.

To assess the impact of a non-diagonal covariance matrix in the Information Share when

Table 13: IS - MC1

	theoretical value			spectral			Cholesky		
	E	R^f	R^h	E	R^f	R^h	E	R^f	R^h
E	1.00	0.00	0.00	1.00 (0.002)	0.00 (0.001)	0.00 (0.002)	1.00 (0.002)	0.00 (0.001)	0.00 (0.002)
R^f	0.50	0.00	0.50	0.50 (0.032)	0.00 (0.001)	0.50 (0.032)	0.50 (0.032)	0.00 (0.002)	0.50 (0.032)
R^h	0.00	0.00	1.00	0.00 (0.002)	0.00 (0.002)	1.00 (0.002)	0.00 (0.002)	0.00 (0.002)	1.00 (0.002)

Information Shares estimates (mean) and standard deviation among all simulations using Cholesky decomposition, Spectral decomposition and the Theoretical value calculated using Spectral decomposition. Diagonal covariance matrix.

using Cholesky decomposition, we now consider non zero correlations among markets. The covariance matrix used when generating the dataset for MC1 is given by

$$\Omega_{MC1} = \begin{pmatrix} 1.00 & 0.40 & 0.10 \\ 0.40 & 1.00 & 0.50 \\ 0.10 & 0.50 & 1.00 \end{pmatrix} \tag{11}$$

while the one used for MC2 is given by

$$\Omega_{MC2} = \begin{pmatrix} 1.0 & 0.1 & 0.3 & 0.4 & 0.1 \\ 0.1 & 1.0 & 0.2 & 0.4 & 0.4 \\ 0.3 & 0.2 & 1.0 & 0.4 & 0.4 \\ 0.4 & 0.4 & 0.4 & 1.0 & 0.2 \\ 0.1 & 0.4 & 0.4 & 0.2 & 1.0 \end{pmatrix} \tag{12}$$

and the one for MC3 is

$$\Omega_{MC3} = \begin{pmatrix} 1.0 & 0.5 & 0.5 & 0.2 & 0.5 \\ 0.5 & 1.0 & 0.7 & 0.8 & 0.9 \\ 0.5 & 0.7 & 1.0 & 0.5 & 0.7 \\ 0.2 & 0.8 & 0.5 & 1.0 & 0.7 \\ 0.5 & 0.9 & 0.7 & 0.7 & 1.0 \end{pmatrix}. \tag{13}$$

The covariance matrix used on MC3 is based on our empirical correlation among the series. Tables 14 and 15 report the results for MC1 and MC2, respectively. First, we show the theoretical value and the results using spectral decomposition. Then, we present the IS with Cholesky decomposition for the upper bound of E_t and lower bound of $P^h_{p,t}$ (P^h_t for MC1), and the average between upper and lower bounds for E_t and $P^h_{p,t}$. To this purpose, we compute the IS with exchange rate being the first variable and home price being the last to find the upper bound for exchange rate and the lower bound for home price. The order

of all other variables in the middle are kept unchanged. With exchange rate ordered first, this variable has no zeros imposed by Cholesky, whereas the preferred stock at home market ordered last has all zeros, but one. When we change the order to compute the upper and lower bounds for home price and exchange rate, respectively, the opposite happens. We choose these two variables to be the ones to have the two extremes of upper and lower bounds, because we build our series with them being the major drivers of the other series. We find considerable differences among the upper and lower bounds. Although the average of the two bounds present a better estimate than the single upper or lower bounds, it is still not so close to the theoretical value. The result confirms the problem of using Cholesky decomposition. It imposes a restriction that may not be true (lower triangularization form), causing the order of variables to matter on the estimation process and resulting in dissimilar measures. In MC1, we also tried to get the average among all possible combinations (not only the upper and lower bounds of the two variables determining the third one). There is not a significant difference on the IS estimates for the home price and exchange rate, but the IS estimate for the foreign price is worsen off compared to the theoretical value.

The results using the eigenvalue decomposition present a much better estimation of the theoretical value of the Information Share. As a consequence of the eigenvalue decomposition, there is no upper or lower bound for the IS, just a single estimate. The order of variables does not matter anymore, there will be always one single measure. This is a huge advantage against the Cholesky decomposition, which is commonly implemented to measure Information Share. The increase of the sampling frequency (a solution proposed by the literature) may reduce the correlation among markets, not yielding much difference between IS estimates computed with Cholesky or spectral. This may be true for some datasets, but not for all, as we will see further in our empirical application. The increasing number of trading algorithms have made the markets correlated even when observations are sampled with a relatively high frequency.

When we increase the correlation among markets, the Cholesky IS estimates deviate even more from the theoretical value. Although the average entails a better estimate of the IS than upper and lower bounds, the difference among upper and lower bounds increases significantly. The IS interpretation is difficult in such situations. Table 16 reports the results for MC3, where we impose a higher covariance among the markets, stated by (13). IS estimates obtained using eigenvalue decomposition are still very close to the theoretical value. They also show a slightly higher standard deviation among all the simulations when

compared to Cholesky. This may be explained by the restrictions that Cholesky decomposition imposes. As our Monte Carlo exercises show, even when we have a high covariance among markets, Cholesky decomposition imposes a zero when decomposing the covariance matrix for all variables on the upper triangular, which affects the IS measure giving a lower parameter weight for the variables ordered last. The only variable which has no zeros imposed by this decomposition is the one ordered first.

In summary, we propose the use of eigenvalue decomposition on the covariance matrix while computing the IS parameters based on a reduced form VECM approach. The eigenvalue decomposition can be greatly useful when this matrix is not diagonal, once it eliminates the problem of having a non unique measure. This situation may happen very often either when high frequency data is not available, or when correlation among markets still prevails even when high frequency data is used.

Table 14: Theoretical value and spectral measures - MC1

	theoretical value			spectral			Cholesky (E upper)			Cholesky (E lower)			Cholesky (Avg)		
	E	R^f	R^h	E	R^f	R^h	E	R^f	R^h	E	R^f	R^h	E	R^f	R^h
E	0.96	0.04	0.00	0.96 (0.013)	0.04 (0.013)	0.00 (0.002)	1.00 (0.029)	0.00 (0.029)	0.00 (0.007)	0.82 (0.002)	0.16 (0.001)	0.01 (0.001)	0.91 (0.013)	0.08 (0.012)	0.01 (0.003)
R^f	0.46	0.10	0.45	0.46 (0.031)	0.10 (0.019)	0.45 (0.031)	0.55 (0.03)	0.12 (0.021)	0.34 (0.031)	0.37 (0.031)	0.08 (0.017)	0.55 (0.031)	0.46 (0.03)	0.10 (0.018)	0.44 (0.029)
R^h	0.00	0.07	0.93	0.00 (0.002)	0.07 (0.016)	0.93 (0.016)	0.01 (0.001)	0.25 (0.001)	0.74 (0.002)	0.00 (0.007)	0.00 (0.025)	1.00 (0.025)	0.01 (0.004)	0.13 (0.014)	0.87 (0.014)

Theoretical value calculated using spectral decomposition, mean and standard deviation among all simulations of IS estimates using spectral decomposition and Cholesky decomposition. Cholesky (E upper) bound was computed with exchange rate ordered first and home market ordered last and Cholesky (E lower) bound was computed in the opposite way. Cholesky (Avg) is the average between Cholesky (E upper) and Cholesky (E lower).

Table 15: Theoretical value and spectral measures - MC2

	theoretical value					spectral					Cholesky (E upper)					Cholesky (E lower)					Cholesky (Avg)				
	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h
E	0.94	0.00	0.02	0.04	0.00	0.94 (0.015)	0.00 (0.002)	0.02 (0.009)	0.04 (0.012)	0.00 (0.003)	1.00 (0.003)	0.00 (0.001)	0.00 (0.001)	0.00 (0.001)	0.00 (0.002)	0.81 (0.025)	0.01 (0.005)	0.08 (0.018)	0.10 (0.02)	0.01 (0.007)	0.90 (0.013)	0.00 (0.003)	0.04 (0.009)	0.05 (0.01)	0.01 (0.004)
C^h	0.00	0.92	0.00	0.04	0.04	0.00 (0.002)	0.92 (0.018)	0.00 (0.004)	0.04 (0.013)	0.04 (0.012)	0.01 (0.007)	0.99 (0.007)	0.00 (0.001)	0.00 (0.002)	0.00 (0.001)	0.00 (0.002)	0.84 (0.025)	0.00 (0.001)	0.00 (0.002)	0.16 (0.025)	0.01 (0.004)	0.91 (0.013)	0.00 (0.001)	0.00 (0.001)	0.08 (0.013)
R^f	0.46	0.02	0.05	0.03	0.44	0.45 (0.031)	0.02 (0.01)	0.05 (0.014)	0.03 (0.011)	0.44 (0.032)	0.55 (0.032)	0.07 (0.017)	0.05 (0.014)	0.00 (0.004)	0.33 (0.03)	0.37 (0.032)	0.00 (0.003)	0.04 (0.012)	0.04 (0.013)	0.55 (0.032)	0.46 (0.03)	0.04 (0.01)	0.04 (0.013)	0.02 (0.005)	0.44 (0.029)
C^f	0.45	0.44	0.02	0.07	0.02	0.45 (0.032)	0.44 (0.032)	0.02 (0.008)	0.07 (0.017)	0.02 (0.01)	0.55 (0.032)	0.45 (0.032)	0.00 (0.001)	0.00 (0.002)	0.00 (0.002)	0.37 (0.031)	0.44 (0.033)	0.04 (0.012)	0.04 (0.014)	0.11 (0.022)	0.46 (0.03)	0.44 (0.031)	0.02 (0.006)	0.02 (0.007)	0.06 (0.011)
R^h	0.00	0.04	0.04	0.00	0.92	0.00 (0.002)	0.04 (0.013)	0.04 (0.013)	0.00 (0.004)	0.91 (0.019)	0.01 (0.007)	0.15 (0.024)	0.10 (0.021)	0.01 (0.005)	0.72 (0.031)	0.00 (0.002)	0.00 (0.001)	0.00 (0.002)	0.00 (0.002)	1.00 (0.003)	0.01 (0.004)	0.08 (0.012)	0.05 (0.01)	0.00 (0.003)	0.86 (0.016)

Theoretical value calculated using spectral decomposition, mean and standard deviation among all simulations of IS estimates using spectral decomposition and Cholesky decomposition. Cholesky (E upper) bound was computed with exchange rate ordered first and home market ordered last and Cholesky (E lower) bound was computed in the opposite way. Cholesky (Avg) is the average between Cholesky (E upper) and Cholesky (E lower).

Table 16: Theoretical value and spectral measures - MC3

	theoretical value					spectral					Cholesky (E upper)					Cholesky (E lower)					Cholesky (Avg)				
	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h	E	C^h	R^f	C^f	R^h
E	0.87	0.04	0.04	0.00	0.04	0.87 (0.021)	0.04 (0.014)	0.05 (0.014)	0.00 (0.002)	0.04 (0.013)	1.00 (0.003)	0.00 (0.001)	0.00 (0.001)	0.00 (0.002)	0.00 (0.002)	0.60 (0.031)	0.01 (0.008)	0.04 (0.013)	0.09 (0.02)	0.25 (0.028)	0.80 (0.016)	0.01 (0.004)	0.02 (0.006)	0.05 (0.01)	0.13 (0.014)
C^h	0.04	0.19	0.09	0.08	0.59	0.04 (0.013)	0.19 (0.026)	0.09 (0.018)	0.09 (0.018)	0.59 (0.033)	0.25 (0.03)	0.56 (0.033)	0.01 (0.006)	0.00 (0.002)	0.18 (0.027)	0.00 (0.001)	0.00 (0.002)	0.00 (0.002)	0.00 (0.001)	1.00 (0.003)	0.13 (0.015)	0.28 (0.017)	0.00 (0.003)	0.00 (0.001)	0.59 (0.014)
R^f	0.43	0.14	0.08	0.03	0.32	0.43 (0.032)	0.14 (0.023)	0.09 (0.018)	0.03 (0.011)	0.32 (0.031)	0.74 (0.029)	0.19 (0.026)	0.00 (0.004)	0.00 (0.002)	0.06 (0.016)	0.20 (0.025)	0.01 (0.005)	0.01 (0.008)	0.03 (0.012)	0.75 (0.028)	0.47 (0.026)	0.10 (0.014)	0.01 (0.006)	0.02 (0.006)	0.41 (0.019)
C^f	0.43	0.14	0.08	0.03	0.32	0.43 (0.032)	0.14 (0.023)	0.09 (0.018)	0.03 (0.011)	0.32 (0.031)	0.74 (0.029)	0.19 (0.026)	0.00 (0.004)	0.00 (0.002)	0.06 (0.016)	0.20 (0.025)	0.01 (0.005)	0.01 (0.008)	0.03 (0.012)	0.75 (0.028)	0.47 (0.026)	0.10 (0.014)	0.01 (0.006)	0.02 (0.006)	0.41 (0.019)
R^h	0.04	0.19	0.09	0.08	0.59	0.04 (0.013)	0.19 (0.026)	0.09 (0.018)	0.09 (0.018)	0.59 (0.033)	0.25 (0.03)	0.56 (0.033)	0.01 (0.006)	0.00 (0.002)	0.18 (0.027)	0.00 (0.001)	0.00 (0.002)	0.00 (0.002)	0.00 (0.001)	1.00 (0.003)	0.13 (0.015)	0.28 (0.017)	0.00 (0.003)	0.00 (0.001)	0.59 (0.014)

Theoretical value calculated using spectral decomposition, mean and standard deviation among all simulations of IS estimates using spectral decomposition and Cholesky decomposition. Cholesky (E upper) bound was computed with exchange rate ordered first and home market ordered last and Cholesky (E lower) bound was computed in the opposite way. Cholesky (Avg) is the average between Cholesky (E upper) and Cholesky (E lower).