University of São Paulo
“Luiz de Queiroz” College of Agriculture

Essays on the microstructure of emerging commodities futures markets

Geraldo Costa Junior

Thesis presents to obtain the degree of Doctor in
Science. Area: Applied Economics

Piracicaba
2017

Geraldo Costa Junior
Bachelor of Economics

Essays on the microstructure of emerging commodities futures markets
versão revisada de acordo com a resolução CoPGr 6018 de 2011.

Advisor:
Prof. Dr. **JOÃO GOMES MARTINES FILHO**

Thesis presents to obtain the degree of Doctor in
Science. Area: Applied Economics

Piracicaba
2017

**Dados Internacionais de Catalogação na
Publicação DIVISÃO DE BIBLIOTECA –
DIBD/ESALQ/USP**

Costa Júnior, Geraldo

Essays on the microstructure of emerging commodities futures markets

/ Geraldo Costa Júnior. - - versão revisada de acordo com a resolução CoPGr 6018 de 2011. - -
Piracicaba, 2017.

65 p.

Tese (Doutorado) - - USP / Escola Superior de Agricultura “Luiz
de Queiroz”.

1. Dados de alta frequência 2. Commodities 3. Microestrutura 4.
Dealers I. Título

DEDICATION

To the Lord of the vineyard.

AKNOLEDGEMENTS

I would like to thank my advisor, Prof. João Gomes Martines-Filho, who has supported, encouraged and pushed me throughout this path. It's a privilege to work with Prof. Martines and every meeting was like a new breath of fresh air.

I would like to thank Prof. Andres Trujillo-Barrera, from Wageningen University and Research Centre, for being present since the development of the first ideas of my research project, for his excellent comments, and guidance. I have learned immensely. Thanks to Prof. Joost Pennings for being always supportive.

I would like to thank Prof. Marcelo Fernandes, from Queen Mary University of London, for giving excellent insights to develop the ideas for the last part of this dissertation.

For all of you, I don't have enough words to express my gratitude and admiration.

I would like to thank the Department of Economics, Sociology and Administration at Esalq/USP and to my colleagues: Natalyia, Angel, Mari, and Mirian. Special thanks to the Applied Microeconomics Team (EMA), with whom I had the pleasure to contribute and work with along the PhD years.

I would like to thank the Marketing and Consumer Behavior Group/WUR and the School of Economics and Finance/QMUL for being excellent hosts. Many thanks to the London crew: Andrea, Bruno, Bruna, Gabriele, Carina, and Lorenz B. You guys made London winter warm!

I would like also to thank my friends at Comunidade Corpus Christi and Comunidade Shalom (Netherlands) for the support and friendship during the good and bad times.

To my family: dad, mom, Aline. Thanks for being always there for me and for being the base upon which everything was built.

Thanks to CAPES and CNPq for the scholarship, without which this research would not have been viable.

EPIGRAPH

“Now the kingdom of Heaven is like a landowner going out at daybreak to hire workers for his vineyard. He made an agreement with the workers for one denarius a day and sent them to his vineyard. Going out at about the third hour he saw others standing idle in the market place and said to them, “You go to my vineyard too and I will give you a fair wage.” So they went. At about the sixth hour and again at about the ninth hour, he went out and did the same. Then at about the eleventh hour he went out and found more men standing around, and he said to them, “Why have you been standing here idle all day?” “Because no one has hired us,” they answered. He said to them, “You go into my vineyard too. In the evening, the owner of the vineyard said to his bailiff, “Call the workers and pay them their wages, starting with the last arrivals and ending with the first.” So those who were hired at about the eleventh hour came forward and received one denarius each. When the first came, they expected to get more, but they too received one denarius each.”

Matthew, 20.

CONTENT

RESUMO.....	8
ABSTRACT.....	9
1. INTRODUCTION.....	11
REFERENCES.....	12
2. MODELLING AND FORECASTING VOLATILITY IN EMERGING AGRICULTURAL FUTURES MARKETS.....	15
ABSTRACT.....	15
2.1 INTRODUCTION	15
2.2 REALIZED VOLATILITY: THEORETICAL APPROACH.....	16
2.2.1 CONTINUOUS PROCESS.....	16
2.2.2 INTRODUCTION OF JUMP COMPONENTS	17
2.3 REALIZED VOLATILITY MODELS	17
2.3.1 THE DIEBOLD-MARIANO TEST.....	19
2.4 DATABASE.....	19
2.5 DESCRIPTIVE STATISTICS	20
2.6 MODEL ESTIMATIONS: IN-SAMPLE ANALYSIS.....	21
2.7 FORECASTS: OUT-OF-SAMPLE ANALYSIS.....	23
2.8 CONCLUSION.....	26
REFERENCES.....	26
3. CONCENTRATION AND LIQUIDITY COSTS IN EMERGING COMMODITIES EXCHANGES.....	31
ABSTRACT.....	31
3.1 INTRODUCTION	31
3.2 LITERATURE REVIEW	33
3.3 STRUCTURE OF BM&F-BOVESPA DATA	35
3.4 VARIABLES OF INTEREST	36
3.5 METHODOLOGY	37
3.6 PRELIMINARY ANALYSIS.....	38
3.7 REGRESSION RESULTS.....	41
3.7.1 SEASONALITY ANALYSIS	44
3.8 CONCLUSION	45
REFERENCES.....	46
4. INTER-DEALERS MARKETS DYNAMICS AND COMPETITION: EVIDENCE FROM THE LIVE CATTLE MARKETS... ..	49
ABSTRACT.....	49
4.1 INTRODUCTION	49
4.2 RELATED LITERATURE.....	50

	7
4.3 INSTITUTIONAL DETAILS AND DATA	52
4.4 DATA DESCRIPTION... ..	52
4.5 CHARACTERISTICS OF THE LIVE CATTLE INTER-DEALER MARKET AT BM&F BOVESPA.....	53
4.6 MODELLING DEALERS'S ACTIVITY	55
4.7 DEALERS' DYNAMICS, COMPETITION AND BID-ASK SPREAD.....	59
4.8 CONCLUSION	61
<i>REFERENCES</i>	<i>61</i>
<i>5. CONCLUSION</i>	<i>64</i>

RESUMO

Ensaio sobre a microestrutura de mercados futuros agrícolas emergentes

Negociações nos mercados futuros de commodities passaram por transformações estruturais significativas durante a primeira década dos anos 2000, resultando em uma elevação dos níveis de volume e *open interest*, e também em uma maior facilidade de acesso a esses mercados e inclusão de novos participantes. Beneficiando-se da divulgação de dados de alta frequência possibilitada por estas transformações, esta tese, composta por três artigos, tem por objetivo investigar diferentes aspectos da microestrutura dos mercados de commodities da BM&F-Bovespa. O primeiro artigo analisa a modelagem e previsão de volatilidade realizada nos mercados futuros de milho e boi gordo. Para este fim, utilizou-se o modelo heterogêneo auto regressivo proposto por Corsi (2009), bem como suas extensões adaptadas para a inclusão dos componentes de saltos (*jumps*) (Andersen et al., 2007) e alavancagem (Corsi e Reno, 2012). Utilizando diferentes métricas de comparação, os resultados encontrados mostram que os modelos que incluem os componentes de saltos e os de alavancagem tem melhor desempenho que os demais em análises *in-sample* (modelagem). Por outro lado, a análise das previsões *out-of-sample* mostra que, para o mercado de boi gordo, não há diferença entre os modelos empregados, enquanto que para o mercado de milho verificou-se uma diferenciação preditiva no horizonte diário, porém para os horizontes semanal e mensal, os quatro modelos tiveram performance indistinta. O segundo artigo explora a relação entre volatilidade, volume e *bid-ask spread* nos mercados de milho e boi gordo. Levando em conta que se trata de mercados emergentes, métricas de concentração de mercado foram incluídas na análise. Para capturar a relação entre volatilidade, volume e *bid-ask spread*, um modelo estrutural de três equações simultâneas foi utilizado e a estimação foi feita através do modelo GMM com variáveis instrumentais. Os resultados indicam que os níveis de *bid-ask spread* encontrados para o mercado de boi gordo são maiores que os encontrados para o mercado de milho. Além disso, o *bid-ask spread* é negativamente relacionado ao volume e positivamente relacionado à volatilidade. Entretanto, a intensidade e magnitude da relação entre as variáveis depende dos níveis de liquidez dos mercados analisados. A concentração impacta o mercado de milho e boi gordo de forma diferente. O terceiro artigo investiga a dinâmica da relação entre a atividade dos *dealers* e estrutura do mercado de boi gordo na BM&F-Bovespa. Primeiramente, faz-se uma análise descritiva deste mercado e posteriormente estuda-se o comportamento dos *dealers* e seus determinantes. Os resultados indicam que os *dealers* no mercado de boi gordo não operam em uma estrutura competitiva e que a atividade destes é positivamente relacionada à concentração de mercado, ao *bid-ask spread*, ao número de *dealers* ativos e à quantidade de contratos transacionada pelos *dealers*.

Palavras-chave: Dados de alta frequência; Commodities; Microestrutura; Dealers

ABSTRACT

Essays on the microstructure of emerging commodities futures markets

Commodities futures trading went through unparalleled structural transformation during the first decade of the 2000s, which ultimately resulted in long lasting impacts on the volume and open interest levels as well as on the access to these markets and inclusion of new participants. Benefiting from the new sets of high frequency data made available due to these transformations, this dissertation is composed of three papers that investigate different market microstructure aspects of the commodities futures markets at BM&F-Bovespa. The first paper analyzes the modelling and forecasting of realized volatility in the corn and live cattle markets. For this purpose, the heterogeneous autoregressive model (HAR-RV) proposed by Corsi (2009) was used, as well as its extensions adapted to include jump components (Andersen et al., 2007) and leverage components (Corsi and Reno, 2012). Using measurements to compare both analysis, results show that modelling in-sample realized volatility is best performed if jumps and leverage components are included in the model. Out-of-sample forecasts results for the live cattle market show that there was no statistically significant difference between the models. For the corn markets, difference in the models' forecast performance was found at the daily horizon only. The second paper delves into the relationship between volatility, volume and bid-ask spread in the corn and live cattle markets. Considering that these are emerging agricultural markets, concentration measures were also included. A three-equation structural model was used to capture the relationship between volatility, volume and bid-ask spread and the estimation was performed using the IV-GMM approach. Findings show that bid-ask spread levels are higher for live cattle markets than it is for corn markets. In addition, bid-ask spread is negatively related to volume and positively related to volatility. The significance and magnitude of the responses depend on the level of liquidity in each market. Further, results point out that concentration impacts corn and live cattle differently. The third paper examines the dynamic relationship between dealers activity and market structure in the live cattle inter-dealer market at BM&F-Bovespa. First, a descriptive analysis of the live cattle inter-dealer market structure is carried out and then follows an investigation of the dynamic of dealers' activity and its determinants. Results indicate that the live cattle inter-dealer market is not competitive and that dealers' activity is positively related to market concentration, quoted bid-ask spread, number of active dealers and the dealer's traded quantity.

Keywords: High frequency data; Commodities; Microstructure; Dealers

1. INTRODUCTION

Commodities futures trading in Brazil started with the foundation of the *Bolsa de Mercadorias de São Paulo* in 1917, which would become BM&F in 1991 and BM&F-Bovespa in the beginning of the years 2000s. Like other commodities futures markets around the world, the evolution of commodities futures trading has followed a stable path along the following decades, resulting in a significant increase in the size of these markets.

This increase in size, however, did not translate into the way trade was carried out by its main participants, which remained roughly unchanged since its inception. This configuration was largely transformed during the first decade of the years 2000s. Irwin and Sanders (2012) argue that the structural transformations carried out in commodities futures markets during this period were unparalleled in its history, which has impacted trading in many ways. Among the structural transformations are the introduction of electronic trading platforms and the possibility of financial liquidation for commodities contracts, which have ultimately contributed to make trading easier and, consequently attract new participants in the market and bring down transaction costs. By the end of this process, commodities futures markets became more integrated with financial markets than they have ever been (Domanski and Heath, 2007; Irwin and Sanders, 2012).

The current situation calls for a reexamination of a wide range of issues related to the functioning of commodities futures markets in this new context. As a byproduct of the advances in trading technology, different sets of high frequency data have been made available by many exchanges around the world. Among emerging markets, the BM&F-Bovespa is a reference for having good quality and easy accessible data. A number of studies used high frequency data to better understand the microstructure of commodities markets (Eaves and Williams, 2010; Martinez et al., 2011; Wang, Garcia and Irwin, 2013; Lehecka et al., 2014). However, the microstructure of emerging commodities markets, which must account for its inherent characteristics such as the low levels of liquidity and relatively high degree of market concentration, remains largely unexplored.

This dissertation is composed of three papers that delve into the functioning of emerging commodities markets using high frequency data and the market microstructure approach. Our results show aspects of these markets that were not evident in analysis using lower frequency data.

The second chapter studies the behavior of price variability in the corn and live cattle futures markets at BM&F-Bovespa using high frequency data at a 30-minute interval in the period March-2014/April 2017. I analyzed the data through the lenses of the heterogeneous autoregressive model (HAR-RV), proposed by Corsi (2009). In other words, I gauged how much of the current realized volatility observed in these markets can be explained by its long memory process. For the latter, I controlled for daily, weekly and monthly components. I also used extensions of the HAR-RV to account for jump components (Andersen et al. 2007) and leverage components (Corsi and Reno, 2012) and compared their performances both on modelling and forecasting. Results for the in-sample modelling show that the best way to harness all volatility features in the corn and live cattle future markets is to use models that account for both jump components and leverage effects. However, the out-of-sample forecast analysis show that, for the live cattle market, there is no difference between the models used. On the other hand, for the corn market, difference in predictability rises only at the daily forecast horizon.

Understanding the behavior of volatility in agricultural markets is crucial to the decision-making process of all market participants. Besides, it is a key element to the development of more accurate risk-

mitigation instruments, which are widely used by those seeking to minimize loss resulted from market uncertainties. Nonetheless, it is equally important to understand the relationship between volatility and other market aspects, such as volume and the bid-ask spread.

In this context, chapter 3 describes the relationship between volatility, volume and bid-ask spread in the corn and live cattle futures markets in the March 2004/February 2016 period. I also accounted for market concentration, which is relevant variable when analyzing emerging markets. To investigate the relationship between these variables, I followed Martinez et al. (2011) and Wang, Garcia, and Irwin (2013) and used a three-equation model in which volatility, volume and bid-ask spread are jointly determined. For the estimation, I used a generalized method of moments model with instrumental variables (IV-GMM). Results demonstrate that the average bid-ask spread is lower for the corn market than it is for the live cattle market. Consistent with the literature, I found that bid-ask spread responds negatively to changes in volume and positively to changes in volatility. However, the significance and magnitude of the responses depend on the level of liquidity in each market. In this sense, concentration also impacts these markets differently.

Another relevant aspect of the market microstructure analysis is how market participants (dealers) interact and organize themselves. This topic is of particular interest for emerging markets, where liquidity levels and the number of participants in futures markets are usually low if compared to more mature ones. Studies dedicated to the organization of dealers' markets and competition are well documented for stock and equities markets (Christie and Schultz, 1994; Ellis, Michaely and O'Hara, 2002; Aspris et al. 2012). However, much less is known about the nature, evolution and behavior of dealers in commodities markets.

Hence, in chapter four I analyze the structure and behavior of dealers in the live cattle market in the March 2014/February 2016 period. Following Ellis, Michaely and O'Hara (2002), I model the dealer's decision to be active on any given day of the analyzed period. For this purpose, I used the Probit regression with instrumental variables (IV-Probit). In this case, the dependent variable is 1 when the dealer is active in the market and zero otherwise. Explanatory variables were the ones identified as being potential determinants of the dealer's decision to be active in the market, such as market number of active dealers, volatility, market concentration, quoted bid-ask spread, among others. Subsequently, I estimated the impact of dealers activity on quoted bid-ask spread using a panel fixed effects regression. Findings suggest that live cattle markets at BM&F- Bovespa do not operate in a competitive structure. Further, dealers' probability to be active increase with market concentration, quoted bid-ask spread, number of active dealers and the dealer's traded quantity. As expected, I found that quoted bid-ask spreads in the live cattle market increases with market concentration and decreases with the number of active dealers.

References

- Andersen, T. G.; Bollerslev, T.; Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*. Vol. 89, pp.701-720.
- Aspris, A.; Frino, A.; Lapone, A. (2013). The impact of market maker competition on market quality: evidence from an options exchange. *Australasian Accounting, Business and Finance Journal*, Vol. 6, n. 5, p. 24-36.
- Christie, W. Schultz, P. (1994). Why do Nasdaq market makers avoid odd-eighth quotes? *The Journal of Finance*. Vol. 49, n.5, pp. 1813-1840.

Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*. Vol. 7, pp. 174-196.

Corsi, F.; Reno R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*.

Domanski, D.; A. Heath. (2007). Financial Investors and Commodity Markets. *Bank for International Settlements Quarterly Review*. March 2007, pp.53–67.

Eaves, J.; Williams, J. (2010). Are intraday volume and volatility U-shaped after accounting for public information? *American Journal of Agricultural Economics*. Vol 92, n.1, pp. 212-227.

Ellis, K.; Michaely, R.; O'Hara, M. (2002). The making of a dealer market: from entry to equilibrium in the trading of Nasdaq stocks. *The Journal of Finance*. Vol. LVII, n.5, pp. 2289-2316.

Irwin, S.; Sanders, D. (2012). Financialization and structural change in commodity futures markets. *Journal of Agricultural and Applied Economics*. Vol. 44, n.3, pp. 371-396.

Lehecka, G.; Wang, X.; Garcia, P. (2014). Gone in ten minutes: intraday evidence of announcement effects in the electronic corn futures market. *Applied Economic Perspectives and Policy*. Vol. 36, n.3, pp. 504-526.

Martinez, V.; Gupta, P.; Tse, Y.; Kittiakarasakun, J. (2011). Electronic versus open outcry trading in agricultural commodities futures markets. *Review of Financial Economics*. Vol. 20, pp. 28-36.

Wang, X.; Garcia, P.; Irwin, S.H. (2013). The behavior of bid-ask spreads in the electronically traded corn futures markets. *American Journal of Agricultural Economics*. Vol. 96, n.2, pp. 557-577.

2. MODELLING AND FORECASTING REALIZED VOLATILITY IN EMERGING AGRICULTURAL FUTURES MARKETS

Using intraday data from the corn and live cattle contracts traded at BM&F-Bovespa, this paper focuses on modelling and forecasting realized volatility in the context of emerging commodities markets. For this purpose, four different models are considered in the analysis: the heterogeneous autoregressive model (HAR-RV) developed by Corsi (2009) and its extensions adapted to include jumps (Andersen et al., 2007) and leverage components (Corsi and Reno, 2012). Results for the in-sample modelling show that the best way to harness all volatility features in the corn and live cattle future markets is to use models that account for both jump components and leverage effects. However, the out-of-sample forecast analysis show that, for the live cattle market, there is no difference between the models used. On the other hand, for the corn market, difference in predictability rises only at the daily forecast horizon.

Keywords: realized volatility, jumps, commodities

2.1. Introduction

Modelling and forecasting volatility of agricultural commodities prices has been a central issue for stakeholders in the agricultural industry. Considerable effort has been made over the last decades to develop models capable of predicting the behavior of agricultural prices. The GARCH-type models have been the most commonly used. These models stemmed from the univariate GARCH approach proposed by Engle (1982) and Bollerslev (1986) and consider volatility as a latent factor that can be estimated using daily or higher frequency data (Calonaci 2015; Tian et al. 2017).

Over the last decade, however, commodities futures, along with other financial markets, have seen significant technological and structural transformations, which impacted the way trade is carried out in these markets. The introduction of electronic trading platforms has facilitated access of traders to markets, bringing along a variety of new issues, such as new participants entering the markets, an increase in volume and the number of trades, unprecedented volatility behavior, among others. Besides, the 2008 subprime crisis triggered a move of investors from stocks, securities and equities markets seeking refuge in commodities funds, regarded by them safe havens in face of the current turmoil (Irwin and Sanders, 2012).

One byproduct of the technological transformations undergone by agricultural commodities futures markets is the easy access to high frequency data, which contains considerably more information compared to lower frequency data such as daily, weekly or monthly data. Besides, financial data at lower frequencies may pose problems such as excess of noise and imprecision (Andersen and Bollerslev 1998; Wink Junior and Pereira, 2011).

The availability of high frequency data allowed the development of models in which volatility is observable and no longer treated as a latent variable, and therefore, can be better described and understood (Calonaci, 2015). Realized Volatility models were conceived to fill this gap and were first developed by Andersen and Bollerslev (1998), Andersen et al. (2001), Barndoff-Nielsen and Shephard (2002).

A number of models have been developed to describe observable volatility more accurately. Corsi (2009) proposed the Heterogeneous Autoregressive Realized Volatility model (HAR-RV). This model is based on the hypothesis that heterogeneous markets are composed of different agents, who may choose to make transactions at different times, what may ultimately impact future volatility in different ways (Müller et al., 1997). Heterogeneity, in this case, comes from different sources: differences in endowments, institutional constraints, risk profiles, information, geographic location, etc. (Corsi, Audrino and Reno, 2012). Therefore, Corsi (2009) proposed to model volatility persistence using past volatility at different time horizons.

Further developments related to the HAR-RV model were carried out by Andersen et al (2007) and Corsi and Reno (2012). The first decomposed the realized volatility in what can be described by a continuous path and what can be captured

by jump component variables. In the latter work, the authors incorporated leverage and jump components in order to enhance the model's predictability power, also considering different time horizons. There are other models that could be included in the Realized Volatility approach such as the MIDAS model, developed by Ghysels et al. (2004). The MIDAS model is more parsimonious than the previous models, which leaves fewer parameters to be estimated. Besides, its regressors tend to be more consistent and less biased at higher frequencies (Wink Junior and Pereira, 2011). An attempt to give realized volatility a proper treatment within the GARCH-type models were made Baillie et al. (1996). The authors introduced the Fractionally Integrated GARCH models (FIGARCH), which captured intermediate range of persistence in the series and described the slow decaying process of the autocorrelation functions (Calonaci, 2015). However, fractionally integrated models lack clear economic interpretation (Corsi, 2009). More recently, Tian et al. (2017), using data from a commodities futures markets in China, proposed a HAR model with time-varying sparsity.

A number of studies have explored the behavior of volatility in emerging stock markets using high frequency data and the realized volatility approach (Sá-Mota and Fernandes, 2004; Carvalho et al., 2006; Chung et al., 2008; Wink Junior and Pereira, 2011). This paper contributes to the literature by shedding light on the behavior of volatility in emerging agricultural commodities markets, more specifically corn and live cattle, which are the two most liquid commodities markets at BM&F-Bovespa. I compare four models for the estimation of realized volatility which comprehends the HAR-RV proposed by Corsi (2009) and its extensions, the LHAR-RV, the HAR-RV-CJ and the LHAR-RV-CJ models. Following Wink Junior and Pereira (2011), I compare the fit and predictability of the models using in-sample and out-of-sample techniques.

I found that the behavior of realized volatility in corn and live cattle markets are rather different. In the corn futures market, the long-memory components account for a bigger share of the realized volatility than the jump components. For the live cattle futures markets, on the other hand, the jump components explain a bigger part of the realized volatility than the long memory components. In both markets, the leverage components are relevant variables for predicting volatility and should always be considered in the models. And finally, the LHAR-RV-CJ model outperformed all other models in both markets.

2.1.1. Realized Volatility: Theoretical Approach

2.1.1.1. Continuous Process

The first developments within the realized volatility approach start with the price process in continuous time (Andersen and Bollerslev, 1998; Andersen et al. 2003; Barndorff-Nielsen et al., 2002; McAleer and Medeiros, 2008):

$$dp_t = \alpha_t + \sigma_t dW_t, \quad t = 1, 2, \dots, n \quad (1)$$

Where dp_t is the logarithm of price increment, α_t is the drift term, σ_t is the strictly positive volatility process and W_t is the Brownian motion. Assuming the length of the day equal 1, the daily return is:

$$r_t = p(t) - p(t-1) = \int_{t-1}^t \alpha(s) ds + \int_{t-1}^t \sigma(s) W(s) \quad (2)$$

The second term of equation (2) is called Integrated Volatility Variation

The intraday return in period M and on day t is defined as:

$$r_{t,m} = p_{t,m} - p_{t,m-1} \text{ for } m = 1, \dots, M \text{ and } t = 1, \dots, n. \quad (3)$$

The realized variance is defined as the sum of all the intraday squared returns, as follows:

$$RV_t = \sum_m^M r_{t,m}^2 \quad (4)$$

Andersen et al. (2003) demonstrated that in the absence of autocorrelation of returns, the realized variance as shown in equation (4) is a consistent estimator of the Integrated Variance.

$$RV_t \xrightarrow{p} IV_t \quad (5)$$

1.1.1. Introduction of Jump Components

The introduction of jump components, as in addition to the continuous process described in the previous section, came as an effort to incorporate noise and jumps inherent to intraday series into the realized volatility models. Looking back at equation (1), I add the $\gamma(t)dq(t)$ term:

$$dp_t = \alpha_t + \sigma_t dW_t + \gamma(t)dq(t) \quad , t = 1, 2, \dots, n \quad (6)$$

Where $\gamma(t)$ is the size of the jump and $dq(t)$ is a continuous process that equals 1 if there is a jump and zero otherwise. Hence, with the inclusion of the jump component, the daily returns comprehend both the continuous process and the discontinuities, as follows:

$$r_t = \int_{t-1}^t \alpha(t + \delta - 1)d\delta + \int_{t-1}^t \sigma^2(t + \delta - 1)dW(\delta) + \sum_{j=1}^{N(t)} \gamma_j(t) \quad (7)$$

Where $N(t)$ accounts for the number of jumps in the sample. Similarly to the continuous process part, it was demonstrated by Andersen et al. (2007) that the RV_t converges in probability to:

$$plim_{M \rightarrow \infty} RV_t = IV_t + \sum \gamma_t \quad (8)$$

2.2. Realized Volatility Models

I considered four different models, the HAR-RV proposed by Corsi (2009) and its extensions, the LHAR-RV, HAR-RV-CJ, and LHAR-RV-CJ models. Following Calonaci (2015), I decided to use these models in this order because they gradually incorporate the jump and leverage components.

The heterogeneous autoregressive model of realized volatility (HAR-RV) is the most simple model of its kind. As defined by Corsi (2009), it is an additive cascade model of different volatility components each of which generated by different types of market participants. The idea of considering different market participants stems from the heterogeneous market hypothesis (Müller et al., 1997). The HAR-RV model in its logarithmic version is present in Equation (9):

$$\log RV_t^d = \beta_0 + \beta_1 \log RV_{t-1,1}^d + \beta_2 \log RV_{t-5,5}^w + \beta_3 \log RV_{t-22,22}^m + \mu_t \quad (9)$$

Where $RV_{t,h} = \frac{1}{h} \sum_{i=1}^h RV_{t+i-1}$ is the h-step ahead average RV, and $h=1, 5$, and 22 are the daily, weekly and monthly volatility components. The error term μ_t is an i.i.d. random variable with zero mean and unit variance. I considered the logarithmic version in all realized volatility models in this paper for two reasons. Having in mind that financial data may bring about challenges to econometric models (Corsi, 2009), the first reason is to get roughly normal distributions and to avoid negativity issues. The second reason is to compare the models with different independent variables, which is the case of the models analyzed in this paper (Forsberg and Ghysels 2006; Chung et al. 2008; Wink Junior and Pereira 2011).

The LHAR-RV is an extension of the heterogeneous autoregressive model of realized volatility adapted to include the leverage component. It is empirically observed that negative returns are correlated with volatility. The inclusion of the latter was first proposed by Corsi and Reno (2012) and captures the asymmetric return-volatility relation at different time levels. Therefore, besides accounting for volatility persistence and long-memory components (HAR-RV), this extension intends to enhance the model's predictive power by including the leverage components. The LHAR-RV model is exposed in Equation 10:

$$\log RV_t^d = \beta_0 + \beta_1 \log RV_{t-1,1}^d + \beta_2 \log RV_{t-5,5}^w + \beta_3 \log RV_{t-22,22}^m + \beta_4 \rho_t^d + \beta_5 \rho_t^w + \beta_6 \rho_t^m + \mu_t \quad (10)$$

The variables of the right side of Equation (10) are exactly the same of Equation (9), except for the daily, weekly and monthly leverage components are represented by $\rho_t^d, \rho_t^w, \rho_t^m$, respectively.

Another extension of the HAR-RV model considered that fact that financial data often exhibit discontinuities.

According to Corsi and Reno (2012), financial prices are subject to abrupt variations. Jumps are not very frequent and are unpredictable, but are usually positively correlated with volatility. In fact, it was widely perceived in the literature that many log price processes could be better described using a combination between continuous processes (very slowly mean reverting

processes) and a much less persistent jump component (Andersen et al. 2003; Andersen et al. 2007). In this sense, the

Andersen et al. (2007) proposed the HAR-RV-CJ model. The model is described in Equation (11):

$$\log RV_t^d = \beta_0 + \beta_1 \log C_{t-1}^d + \beta_2 \log C_{t-1}^w + \beta_3 \log C_{t-1}^m + \beta_4 \log (J_{t-1}^d + 1) + \beta_5 \log (J_{t-1}^w + 1) + \beta_6 \log (J_{t-1}^m + 1) + \mu_t \quad (11)$$

Where in the continuous process part, $\log C_{t-1}^w = \frac{1}{5} \sum_{i=1}^5 \log C_{t-i}$ and $\log C_{t-1}^m = \frac{1}{22} \sum_{i=1}^{22} \log C_{t-i}$ and in the jump component part, $\log (J_{t-1}^w + 1) = \frac{1}{5} \sum_{i=1}^5 \log J_{t-i}$ and $\log (J_{t-1}^m + 1) = \frac{1}{22} \sum_{i=1}^{22} \log J_{t-i}$.

Finally, the LHAR-RV-CJ model combines the innovations proposed both by Andersen et al. (2007) and Corsi et al. (2012). This is the most comprehensive of all four models analyzed in this paper, as it incorporates both the jump and leverage components. The LHAR-RV-CJ model is shown in Equation (12):

$$\log RV_t^d = \beta_0 + \beta_1 \log C_{t-1}^d + \beta_2 \log C_{t-1}^w + \beta_3 \log C_{t-1}^m + \beta_4 \log (J_{t-1}^d + 1) + \beta_5 \log (J_{t-1}^w + 1) + \beta_6 \log (J_{t-1}^m + 1) + \beta_7 \rho_t^d + \beta_8 \rho_t^w + \beta_9 \rho_t^m + \mu_t \quad (12)$$

2.2.1. The Diebold-Mariano Test

Following Chung et al. (2008) and Wink and Pereira (2011), after estimating the four realized volatility models exposed in this section using in-sample data, I compare the out-of-sample forecasts of these models using the Diebold-Mariano test, as proposed by Harvey et al. (1997). This test considers the out-of-sample forecast errors of a pair of models, e_1 and e_2 , under the null hypothesis that the forecasts provided by the pair of models tested do not differ from each other:

$$H_0 = E[g(e_1) - g(e_2)] = 0 \quad (13)$$

The test statistic is given below:

$$D_1 = \left[\frac{n+1-2a+n^{-1}a(a-1)}{n} \right]^{1/2} [\hat{V}(\bar{d})]^{1/2\bar{d}} \quad (14)$$

Where n is the number of observations and a is the number of horizons used in predictions. The term $\hat{V}(\bar{d})$ is given by:

$$\hat{V}(\bar{d}) = n^{-1}[\widehat{\vartheta}_0 + 2 \sum_{k=1}^{a-1} \widehat{\vartheta}_k] \quad (15)$$

2.3. Database

I use two commodity contracts, corn (CCM) and live cattle (BGI). I chose these contracts firstly because they are actively traded at BM&F Bovespa, and secondly because they have a good level of liquidity and are good representatives of the general behavior of an emerging commodities market. More specifically, I proceeded an analysis of the most traded contracts within each month and I picked the three most traded ones, for every month in my sample.

The data used in this study comes from BM&F Bovespa FTP system. Our database comprises intraday returns related to the two contracts. All contracts are traded on a daily basis. The corn contract is traded electronically from 9:00am-3:30pm, and the live cattle contract from 9:00am-4:00pm. In respect to contract specifications, the contract size for corn is 450 bags of 60 net kilograms and R\$0.01 tick size; and 4407 net kilograms and R\$0.01 tick size for live cattle.

Figure 1– Daily realized volatility for corn and live cattle markets, based on 30-minute intraday return from March 2014 to April 2017.

2.4. Descriptive Statistics

The selection of the intraday frequency is a point of debate in the literature. McAleer and Medeiros (2008) argue that as the frequency increases, so does the precision and microstructure noise. A number of studies identified different optimal frequencies (Andersen et al., 2001; Oomen, 2002; Giot and Laurent, 2004; Wink Junior and Pereira, 2011; Tian et al., 2017). Following Wink Junior and Pereira (2011) and considering the thinness of Brazilian agricultural futures markets, I analyzed the data at three different frequencies: 5-minute, 15-minute, and 30-minute intervals. I used the integrated variance estimator as developed by Barndoff-Nielsen and Sheppard (2002) to decide which frequency would give the most accurate estimator of realized volatility. Table 1 shows that the 95% interval confidence analysis suggests that the 30-minute frequency interval gives the smallest confidence interval both for the live cattle and corn futures markets. Therefore, I will use the intraday data at the 30-minute frequency.

Table 1 - Average size of confidence interval of daily realized volatility.

Market/Interval	5-minute	15-minute	30-minute
Corn	123.2474	58.4390	27.1043
Live Cattle	13.9876	7.1795	3.8834

Descriptive statistics are presented in Table 2. As expected, mean and median values for the two contracts are approximately zero. In addition, Jarque Bera statistics show that none of the returns series follow a normal distribution. It has a high kurtosis value, implying that both returns series are leptokurtic with heavy tail. The live cattle series is slightly less noisy than the corn one, what can be standard deviation values.

Table 2 – Descriptive statistics of returns series.

	Observations	Mean	Median	SD	Kurtosis	Skewness	Jarque Bera
Corn	9843	0.0000	0.0000	0.05	12.452	-0.017785	0.0000
Live Cattle	9514	0.0000	0.0000	0.02	6.8645	0.011439	0.0000

Note: the Jarque Bera values correspond to the test p-values.

The autocorrelation functions for the corn and live cattle realized volatilities are depicted in Figure 2. The slow hyperbolic decay exhibited in both functions suggests the presence of long memory processes.

Figure 2 – Autocorrelation functions for live cattle and corn realized volatilities.

2.5. Model Estimations: *In-Sample Analysis*

I present the estimation results for the corn and live cattle markets on Appendix 1.A and 1.B. Regarding the long-term persistence of volatility, measured by the coefficients $\beta_i, i = (1,2,3)$, I found different results across the two contracts. In general, the coefficients of lagged weekly and monthly volatility have the strongest impact on current volatility. Considering the corn market, the magnitude of the weekly coefficient is significantly bigger than the daily and monthly coefficients, throughout the four models and in time horizons $h = 1$ and $h = 5$. The magnitude of monthly coefficients is bigger than the weekly coefficients in the monthly horizon $h = 22$.

On the other hand, regarding the live cattle market, the long-term persistence variables are significant at the 5% level only in the HAR-RV and LHAR-RV models at time horizon $h = 1$. As forecast horizons increase, long-term persistence variables become more significant. In this case, as was also observed in the corn market, the magnitude of weekly coefficients are bigger than the magnitude of monthly coefficients, except for time horizon $h = 22$. This same pattern was observed by Andersen et al. (2007) and Wink Junior and Pereira (2011) when analyzing long-term persistence of volatility in foreign exchange and stocks markets. However, in all cases in the live cattle market, if I include the jump component terms, long-term persistence variables become less or no longer significant, that is, short memory processes such as jumps are more relevant in explaining current volatility in live cattle markets than long term processes. Interestingly, Carvalho et al. (2006) found no evidence of long memory when analyzing the behavior of realized volatility of five actively traded stocks at BM&F Bovespa. In that case, realized volatility could be modelled and forecasted based on short memory processes.

The estimates of the models HAR-RV-CJ and LHAR-RV-CJ with jump components variables $\beta_i, i = (4,5,6)$ show that the importance of these variables in explaining the behavior of current volatility also differs from market to market. For live cattle contracts, the jump component variables explain most of the current volatility. In this case, the monthly jump component α_6 has the biggest coefficient throughout all forecast horizons, except for the HAR-RV-CJ model in time horizon $h = 1$. In this sense, realized volatility in live cattle markets rely more on short memory than on long memory processes, what brings it closer to stock market realized volatility estimated by Carvalho et al. (2006). Analyzing the jump component variables for the corn market, however, I observed a different pattern. In this case, daily and weekly jump components harness most of the short-term variation and they are more significant in time horizons $h = 1$ and $h = 5$. In addition, I observed that jump components contribute only marginally in explaining volatility in this market, as most of the explanatory power dwells in the long-term persistence variables, even after adding jump components.

It is well known in the literature that volatility tends to increase more after a negative shock than after a positive shock (Bollerslev et al., 2006; Bollerslev et al., 2009; Corsi and Reno, 2012). In this sense, the leverage component, measured by coefficients $\beta_i, i = (4,5,6)$ in LHAR-RV and $\beta_i, i = (7,8,9)$ in LHAR-RV-CJ, incorporates the asymmetric impact of past

returns into the heterogeneous autoregressive structure. I observed negative and strongly significant coefficients for the weekly leverage component in the corn market, at all forecast horizons. The negative coefficient is broadly in accordance with the literature and means that past weekly negative shocks affects daily, weekly and monthly volatility. However, I also observed several positive and statistically significant coefficients associated to the daily and monthly leverage effects across all forecast horizons. Similar results were found for the live cattle market, in which all leverage component coefficients, when significant (mostly daily and monthly), were positive. Therefore, I found a persistent leverage effect after both positive and negative shocks in emerging agricultural futures markets.

To compare the results of the HAR-RV, LHAR-RV, HAR-RV-CJ, and LHAR-RV-CJ, I used the adjusted R^2 and a set of loss functions, namely the mean squared error (MSE), the root mean square error (RMSE) and the mean absolute error (MAE) (Forsberg and Ghysels 2006; Chung et al., 2008; Wink Jr. and Pereira, 2011; Calonaci, 2015). I observed that after gradually incorporating jump and leverage component variables, the models' accuracy improved, what can be verified by the adjusted R^2 and the loss functions values. This improvement, however, may happen through different ways, depending on each market. Adding leverage component variables is more impactful on improving the accuracy of models related to corn markets than it is for live cattle markets. On the other hand, adding jump components enhances the accuracy of models related to live cattle markets more than it does for models related to corn markets. As mentioned before, accounting for jump components radically changes the significance and magnitude of the variables related to the long-memory process $\beta_i, i = (1,2,3)$ in the live cattle analysis. A rather smoother change was observed in the corn market after accounting for jumps. Adding leverage components also had a much smaller impact on the significance and magnitude of long-memory variables coefficients, in both markets. The loss function analysis confirms the previous results. All the three functions, MSE, RMSE and MAE, had their lowest values related to the LHAR-RV-CJ model, at all forecast horizons. Therefore, from the in-sample analysis, I conclude that comparing the same forecast horizons, the LHAR-RV-CJ outperforms all other three models, and this result is valid for both the corn and live cattle markets. The loss function values for the in-sample analysis are disposed in Table 3.

Table 3 – In-sample loss functions: MSE, RMSE, and MAE.

	Corn				Live Cattle			
h=1	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
MSE	0.0833	0.0774	0.0823	0.0757	0.1653	0.1543	0.1497	0.1392
RMS	0.2886	0.2783	0.2869	0.2751	0.4065	0.3928	0.3869	0.3732
E								
MAE	0.1335	0.1273	0.1330	0.1262	0.2212	0.2186	0.2099	0.2060
h=5								
MSE	0.0218	0.0204	0.0214	0.0196	0.0799	0.0786	0.0658	0.0646
RMS	0.1476	0.1428	0.1463	0.1400	0.2827	0.2804	0.2565	0.2542
E								
MAE	0.1085	0.1082	0.1069	0.1050	0.1812	0.1813	0.1679	0.1667
h=22								
MSE	0.0193	0.0160	0.0190	0.0157	0.0578	0.0565	0.0397	0.0381
RMS	0.1389	0.1267	0.1379	0.1252	0.2403	0.2377	0.1992	0.1952
E								
MAE	0.1079	0.0996	0.1070	0.0985	0.1759	0.1745	0.1595	0.1563

Table 5 – In-sample Adjusted R^2

	Live Cattle				Corn			
h = 1	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
Adj.R	0.332	0.3723	0.3909	0.4300	0.2456	0.2941	0.2500	0.3058
2								
h = 5								
Adj.R	0.5622	0.5644	0.6372	0.6415	0.5144	0.5420	0.5198	0.5573
2								

h								
= 22								
Adj.R	0.6314	0.6372	0.7451	0.7536	0.3660	0.4691	0.3710	0.4780
2								

The adjusted R^2 values related to the corn and live cattle estimates (Appendix 1.A and 1.B) can be found in Table 5. As mentioned before, I verified through all forecast horizons that the biggest adjusted R^2 values were the ones related to the LHAR-RV-CJ model. In this sense, I also observed that the adjusted R^2 increases with the forecast horizon, as pointed out by Wink Jr and Pereira (2011). Therefore, the best forecast horizons were $h = 5$ for corn markets and $h = 22$ for live cattle markets. It means that the LHAR-RV-CJ models performs the best at the weekly horizon for the live cattle market and at the monthly horizon for the corn market.

2.6. Forecasts: *Out of Sample Analysis*

For the out-of-sample analysis, I divided the total sample of T trading days into I in-sample observations from the first trading day of March 2014 to the last trading day of March 2016 and U out-of-sample observations from the first trading day of April 2016 to the last trading day of March 2017. Hence, $T = I + U$, and a rolling window of I observations was used to reestimate the models and produce 44 (two months) out-of-sample day-ahead forecasts.

To evaluate the forecast performance of the four models, I use both a set of loss functions and the out-of-sample modified Diebold-Mariano test. I selected three of the most common loss functions, namely the mean squared error (MSE), the mean absolute error (MAE) and the root-square mean error (RMSE). The results of the loss functions can be found in Table 4. These results only indicate a general sense of the most accurate model forecast, hence I cannot infer whether the differences among them are significant or not.

Table 4 – Out-of-sample loss functions: MSE, RMSE, and MAE

	Corn				Live Cattle			
h=1	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
MSE	0.0810	0.0965	0.0807	0.0814	0.2841	0.2891	0.2793	0.2841
RMSE	0.2846	0.3106	0.2840	0.2853	0.5330	0.5377	0.5284	0.5330
MAE	0.1031	0.1385	0.1032	0.1061	0.3147	0.3190	0.3024	0.3091
h=5								
MSE	0.0128	0.0141	0.0127	0.0138	0.0992	0.0909	0.0914	0.0915
RMSE	0.1133	0.1188	0.1127	0.1176	0.3148	0.3014	0.3023	0.3024
MAE	0.0815	0.0838	0.0814	0.0839	0.2564	0.2373	0.2209	0.2208
h=22								
MSE	0.0110	0.0139	0.0117	0.0138	0.0896	0.0900	0.0956	0.1204
RMSE	0.1050	0.1180	0.1081	0.1176	0.2994	0.3000	0.3091	0.3469
MAE	0.0875	0.0983	0.0898	0.0967	0.2243	0.2214	0.2404	0.2805

Regarding the MSE, MAE and RMSE values for the corn market shown on Table 4 at forecast horizons $h = 1$ and $h = 5$, it is quite clear that there is very little difference between them. Analyzing the $h = 22$ monthly horizon, however, I observed that the HAR-RV-CJ model has the lowest values for all functions. On the other hand, the MSE, MAE and RMSE values related to the live cattle market show that the HAR-RV and the HAR-RV-CJ have the lowest values at forecast horizons $h = 1$ and $h = 5$. At the $h = 22$ monthly horizon, the loss function values related to the HAR-RV model were the lowest ones.

To analyze the models' forecast accuracy more precisely, I applied the modified Diebold-Mariano test. This test compares the out-of-sample forecast errors of the HAR-RV, LHAR-RV, HAR-RV-CJ and LHAR-RV-CJ models. The test is

carried out in pairs and the null hypothesis of the test is that the two competing forecast models have the same predictive accuracy, and the alternative hypothesis is that one model (the one on the row) forecasts better than the other. Table 6 presents the p-values of the Diebold-Mariano test.

Table 6 - Out-of-Sample Diebold-Mariano Test (p-value)

Corn					Live Cattle				
$h = 1$	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	$h = 1$	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
HAR-RV	-				HAR-RV	-			
LHAR-RV	0.0000	-			LHAR-RV	0.1897	-		
HAR-RV-CJ	0.2585	0.0000	-		HAR-RV-CJ	0.54	0.2769	-	
LHAR-RV-CJ	0.5749	0.0000	0.3319	-	LHAR-RV-CJ	0.9951	0.5303	0.1343	-
$h = 5$	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	$h = 5$	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
HAR-RV	-				HAR-RV	-			
LHAR-RV	0.3115	-			LHAR-RV	0.4754	-		
HAR-RV-CJ	0.7246	0.2477	-		HAR-RV-CJ	0.5625	0.0784	-	
LHAR-RV-CJ	0.456	0.4883	0.3442	-	LHAR-RV-CJ	0.5597	0.6511	0.1015	-
$h = 22$	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	$h = 22$	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
HAR-RV	-				HAR-RV	-			
LHAR-RV	0.1771				LHAR-RV	0.8221	-		
HAR-RV-CJ	0.5919	0.3058			HAR-RV-CJ	0.6402	0.645	-	
LHAR-RV-CJ	0.2911	0.9543	0.2169		LHAR-RV-CJ	0.2032	0.1882	0.3406	-

The results found in the Table 6 partly confirm the loss function analysis. Considering the $h = 1$ forecast horizon, all the p-values are statistically non-significant in the live cattle market analysis, meaning that there is no difference in performance between the four models. On the other hand, the LHAR-RV-CJ model outperformed the LHAR-RV model in this same forecast horizon in the corn market, but I could not reject the null the forecasts perform equally between the LHAR-RV-CJ and HAR-RV models and between LHAR-RV-CJ and the HAR-RV-CJ models. Regarding the $h = 5$ forecast horizon, there is no difference between the four models both in the corn and live cattle markets, considering the 5% significance level and the same result is found for the $h = 22$ forecast horizon. Therefore, while predicting either daily or monthly volatility for corn market at the daily horizon $h = 1$, the inclusion of jump and leverage components improve the model's performance and accuracy. For the other situations, I found no difference in performance between the four models. Wink Junior and Pereira (2011) also found no difference between the HAR-RV and the MIDAS models in forecasting realized volatility for a group of stock at BM&F-Bovespa.

2.7. Conclusion

The goal of this paper was to compare four different models of realized volatility in the context of an emerging agricultural futures market, namely the HAR-RV, LHAR-RV, HAR-RV-CJ and LHAR-RV-CJ models. For this purpose, I used

intraday data with a 30-minute frequency related to the two of the most liquid commodity contracts traded at BM&F-Bovespa: corn (CCM) and live cattle (BGI).

I considered the HAR-RV models proposed by Corsi (2009) and its extensions to account for long memory, leverage and jump components. I observed that although corn and live cattle markets both belong to the category of agricultural commodities, modelling and forecasting realized volatility for each one of these markets may render different results. Considering the mean squared error (MSE), the root mean square error (RMSE), the mean absolute error (MAE) and the adjusted R^2 , I verified that the LHAR-RV-CJ outperforms the other three models in the in-sample analysis.

For the out-of-sample forecasts, however, analyzing the out-of-sample loss function and the Diebold-Mariano test results, I observed that a distinctive predictive power between the models in the corn market rises only at the $h = 1$ monthly horizon. In the live cattle market, on the other hand, no difference between the four models at all forecast horizon analyzed.

This study contributed to extend the realized volatility analysis for Brazilian markets once restricted to the stock market (Sá-Mota and Fernandes, 2004; Carvalho et al., 2006; Wink Junior and Pereira, 2011) to commodities markets. Our findings confirm the applicability of realized volatility methods also on commodities futures markets.

Suggestions for next studies include widen the number of commodities analyzed and possibly investigate the connection between the inherent low liquidity of most commodity markets and the behavior of realized volatility. Another point of interest that remains unexplored is the connection between realized volatility and the structure of the related futures market, i.e., how much of the realized volatility can be explained by the degree of concentration in these markets.

REFERENCES

- Andersen, T.G.; Bollerslev, T. (1998). Answering the skeptics: Yes, standard volatility models do provide accurate forecasts. *International Economic Review*. Vol. 39, pp. 885-905.
- Andersen, T.G.; Bollerslev, T.; Diebold, F.X.; Ebens, H. (2001). The distribution of realized stock return volatility. *Journal of Financial Economics*. Vol. 61, pp. 43-76.
- Andersen, T. G., Bollerslev, T; Diebold, F. X.; Labys, P. (2003). Modelling and forecasting realized volatility. *Econometrica*. Vol. 71(2), pp. 579-625.
- Andersen, T. G., Bollerslev, T., & Diebold, F. X. (2007). Roughing it up: Including jump components in the measurement, modeling, and forecasting of return volatility. *The Review of Economics and Statistics*. Vol. 89, pp.701-720.
- Baillie, R. T., Bollerslev, T., Mikkelsen, H. O. (1996). Fractionally integrated generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. Vol. 74, n.1, pp. 3-30.
- Barndorff-Nielsen, O. E.; Shephard, N. (2002). Econometric analysis of realized volatility and its use in estimating stochastic volatility models. *Journal of the Royal Statistical Society: Series B*. Vol. 64, pp. 253-280.
- Bollerslev, T. (1986). Generalized autoregressive conditional heteroskedasticity. *Journal of Econometrics*. Vol. 31, pp. 307-327.
- Calonaci, F. (2015). Jumps and leverage components, key variables in the description of stock market volatility. Working Paper.
- Carvalho, M. R. C., Freire, M. A. S., Medeiros, M. C.; Souza, L. R. (2006). Modeling and forecasting the volatility of Brazilian asset returns: A realized variance approach. *Revista Brasileira de Finanças*. Vol.4, pp.321-343.
- Chung, H. M., Huang, C. S.; Tseng, T. C. (2008). Modeling and forecasting of realized volatility based on high-frequency data: Evidence from Taiwan. *International Research Journal of Finance and Economics*. Vol. 22, pp.178-191.
- Corsi, F. (2009). A simple approximate long-memory model of realized volatility. *Journal of Financial Econometrics*. Vol. 7, pp. 174-196.
- Corsi, F.; Reno R. (2012). Discrete-time volatility forecasting with persistent leverage effect and the link with continuous-time volatility modeling. *Journal of Business and Economic Statistics*.
- Corsi, F., Audrino, F. & Reno, R. (2012). HAR Modeling for Realized Volatility Forecasting. In: *Handbook of Volatility Models and Their Applications*. (pp. 363-382). New Jersey, USA: John Wiley & Sons, Inc. ISBN 9780470872512
- Engle, R.F. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation. *Econometrica: Journal of the Econometric Society*, pp. 987-1007.
- Forsberg, L. & Ghysels, E. (2006). Why do absolute returns predict volatility so well? *Journal of Financial Econometrics*. Vol. 5, pp. 31-67.
- Ghysels, E.; Santa-Clara, P.; Valkanov, R. (2004). The Midas touch: Mixed data sampling regression models. *Minemo, CIRANO*.
- Irwin, S.; Sanders, D. (2012). Financialization and structural change in commodity futures markets. *Journal of Agricultural and Applied Economics*. Vol. 44, n.3, pp. 371-396

McAleer, M.; Medeiros, M. C. (2008). Realized volatility: A review. *Econometric Reviews*. Vol. 27, pp.10-45.

Müller, U.; Dacorogna, M.; Dav, R.; Olsen, R.; Pictet, O.; Von Weizsacker, J. (1997). Volatilities of different time resolutions – Analysing the dynamics of market components. *Journal of Empirical Finance*. Vol. 4, pp. 213–239.

Sá-Mota, B. & Fernandes, M. (2004). Desempenho de estimadores de volatilidade na Bolsa de Valores de São Paulo. *Revista Brasileira de Finanças*. Vol. 58; pp. 429-448.

Tian, F.; Yang, K.; Chen, L. (2017). Realized volatility forecasting of agricultural commodities futures using the HAR model with time-varying sparsity. *International Journal of Forecasting*. Vol. 33, pp. 132-152.

Wink Junior, M.; Pereira, P. (2011). Modelling and forecasting realized volatility: evidence from Brazil. *Brazilian Review of Econometrics*. Vol. 31, n.2, pp. 315-337.

APPENDIX

Appendix 1.A Table – HAR-RV, LHAR-RV, HAR-RV-CJ, and LHAR-RV-CJ estimates for the corn market at $h = 1, 5, 22$ horizons

	$h = 1$				$h = 5$				$h = 22$			
	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
β_0	0.3431*	0.3216**	0.6536*** 0.4652**	0.6275***		0.4607***	0.6729*** 0.7197**	0.6564***		0.7177***	0.9687***	0.9919***
*	*				*				*			
	(0.1336)	(0.1293)	(0.1583)	(0.1528)	(0.0684)	(0.0665)	(0.0808)	(0.0779)	(0.0654)	(0.0599)	(0.0778)	(0.0710)
β_1	0.0966*	0.1177**	0.0513	-0.0028	0.0264	0.0362	-0.0332	-0.0647**	0.0346	0.0346	-0.0036	-0.0130
	(0.0520)	(0.0505)	(0.0568)	(0.0556)	(0.0266)	(0.0259)	(0.0290)	(0.02838)	(0.0253)	(0.0233)	(0.0276)	(0.0256)
β_2	0.6655*	0.6775***	0.2980** 0.6804**	0.3523***		0.6920***	0.4002*** 0.2539**	0.4376***		0.2329***	0.1046*	0.1004*
*	*				*				*			
	(0.1136)	(0.1100)	(0.1204)	(0.1166)	(0.0583)	(0.0567)	(0.0614)	(0.0594)	(0.0598)	(0.0550)	(0.0594)	(0.0546)
β_3	0.0256 0.2983**	0.0096	0.1231	0.1266	0.0215	0.0053	-0.0077	-0.0197		0.3354***	0.2574***	0.2963***
									*			
	(0.1189)	(0.1155)	(0.1429)	(0.1382)	(0.0609)	(0.0594)	(0.0729)	(0.0704)	(0.0632)	(0.0584)	(0.0704)	(0.0646)
β_4		0.9638***	0.1644	0.3397***		0.3076***	0.1708***	0.2701***		0.0897	0.0866	
			0.1096** (0.1642)			(0.1176)	(0.1181)	(0.0843)		(0.0599)	(0.0601)	
			(0.0748)	(0.0584)		(0.0557)						
β_5		-0.8640**	0.4003*	0.2785*		-	0.2815***	0.2115**		-	0.1676	0.1359
						1.0046***				0.7907***		
		(0.3700)	(0.2142)	(0.2075)		(0.1899)	(0.1093)	(0.1054)		(0.1686)	(0.1062)	(0.0974)
β_6		0.4395	-0.1390	-0.1710		0.9767***	0.0553	0.0448		2.3863***	0.0441	0.0364
		(0.5512)	(0.2256)	(0.2172)		(0.2831)	(0.1151)	(0.1106)		(0.2522)	(0.1125)	(0.1026)
β_7				1.0594*** (0.1679)				0.3951*** (0.0854)				0.1217 (0.0767)

28	β_8	-1.0457*** (0.3766)	-1.1826*** (0.1916)	-0.8445*** (0.1720)
	β_9	0.4871 (0.5473)	0.9979*** (0.2787)	2.4157*** (0.2503)

Appendix 1.B:

Table – HAR-RV, LHAR-RV, HAR-RV-CJ, and LHAR-RV-CJ estimates for the live cattle market at $h = 1, 5, 22$ horizons.

	$h = 1$				$h = 5$				$h = 22$			
	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ	HAR-RV	LHAR-RV	HAR-RV-CJ	LHAR-RV-CJ
β_0	0.4800** *	0.5936** *	-0.2202	-0.1493	0.3635** *	0.4244** *	-0.2767**	-0.2345**	0.335***	0.3528** *	-0.4168***	-0.4355***
	(0.0989)	(0.1035)	(0.1744)	(0.1727)	(0.0689)	(0.0740)	(0.1156)	(0.1176)	(0.0618)	(0.0659)	(0.0906)	(0.0910)
β_1	0.0577	0.0602	0.0393	0.0017	0.1528** *	0.1492** *	0.0174	0.0055	0.0665**	0.0618*	-0.0534*	-0.0496*
	(0.0519)	(0.0508)	(0.0539)	(0.0531)	(0.0361)	(0.0363)	(0.0357)	(0.0362)	(0.0327)	(0.0328)	(0.0286)	(0.0286)
β_2	0.4164** *	0.3777** *	-0.0741	-0.0457	0.3807** *	0.3653** *	0.0512	0.0626	0.2434** *	0.2182** *	-0.1342***	-0.1418***
	(0.0861)	(0.0841)	(0.0907)	(0.0881)	(0.0608)	(0.0609)	(0.0613)	(0.0612)	(0.0608)	(0.0611)	(0.0524)	(0.0516)
β_3	0.2590** *	0.2547** *	-0.1581	-0.1801*	0.2871** *	0.2836** *	-0.2117***	- 0.2225** *	0.5472** *	0.5723** *	-0.0393	-0.0498
	(0.0826)	(0.0805)	(0.1112)	(0.1085)	(0.0588)	(0.0589)	(0.0754)	(0.0756)	(0.0609)	(0.0610)	(0.0619)	(0.0612)
β_4		1.0082** *	-0.0915	-0.0097		0.2343	0.1618**	0.1897**		0.0939	0.1418**	0.1321**
		(0.2137)	(0.1162)	(0.1168)		(0.1526)	(0.0771)	(0.0796)		(0.1294)	(0.0602)	(0.0616)
β_5		-0.7031	0.8360***	0.6721***		-0.2986	0.4169***	0.3517**		0.6880**	0.4934***	0.4567***
		(0.5020)	(0.2380)	(0.2333)		(0.3585)	(0.1587)	(0.1598)		(0.3050)	(0.1238)	(0.1233)
β_6		1.6970**	0.8062***	0.8668***		1.0472**	1.0046***	1.032***		-0.3372	1.1320***	1.2058***
		(0.7395)	(0.2773)	(0.2728)		(0.5280)	(0.1848)	(0.1867)		(0.4494)	(0.1445)	(0.1447)
β_7				1.0352*** (0.2096)				0.2887** (0.1428)				0.1763 (0.1097)
β_8				-0.5883 (0.4942)				-0.2362 (0.3367)				0.6879*** (0.2592)
β_9				1.3116* (0.7094)				0.7776 (0.4833)				-0.5870 (0.3722)

Respectively, *, **, and *** indicate statistical significance at 10, 5 and 1% levels.

The numbers in parenthesis are standard errors.

3. CONCENTRATION AND LIQUIDITY COSTS IN EMERGING COMMODITIES EXCHANGES

This paper analyzes the relationship between liquidity costs (proxied by the bid-ask spread), volume and volatility in the context of an emerging agricultural futures markets. Considering the intrinsic characteristics of emerging markets, we control for market liquidity and market concentration. To analyze this relationship, we used intraday data from corn and live cattle contracts traded at BM&F-Bovespa from March 2014 to February 2016. The methods used consisted of a structural three-equation and the IV-GMM model. Results reveal a liquid corn market and a not so liquid live cattle market. The bid-ask spread responds negatively to volume and positively to volatility. The significance and magnitude of the responses depend on the level of liquidity in each market. Further, results point out that concentration impacts corn and live cattle differently. While in the live cattle market an increase in concentration contributes to reduce volume, in the corn market, it also contributes to reduce bid-ask spread levels.

Keywords: liquidity costs, concentration, commodities

3.1. Introduction

Agriculture in emerging countries has undergone significant transformations in the past decades. Countries such as Brazil, China and India went from net importers of primary goods to global powers in production, exports and imports of a wide range of agricultural products. Brazil currently ranks among the world's largest corn producers, result of a progressive growth process that took place in the last four decades, having seen its production increase four times since 1980. In 2000, Brazil was ranked among the world top-five producers of 31 commodities, and this number rose to 36 in 2008 (Rada and Buccola, 2012). From 2005 to 2015, Brazil's annual corn export growth averaged 21%, making it the second largest corn exporter. Also, the country has historically been one of the largest live cattle and meat producers. As of 2015, Brazil was responsible for roughly 16.3% of all beef produced in the world and was also the second largest exporter of that commodity.

Alongside structural changes in production, commodities futures markets have seen dramatic transformations in recent years, including an increase in trading, consolidation of exchanges, and a shift from pit floor to electronically trading platforms (Irwin and Sanders 2012). Investments in commodities went from \$15 billion in 2003 to \$250 billion in 2009 (Irwin and Sanders 2011). The new investment influx has brought along new types of traders, mainly those seeking in commodity indexes an alternative to diversify their portfolio (Cheng et al. 2014; Adams and Glück 2015).

Both the increase in production and the modernization of trading platforms had positive impacts on the volume and number of transactions of the corn and live cattle markets at BM&F Bovespa. Among commodities markets, these two figure among those with high levels of liquidity, however the corn market has roughly twice as much volume as the live cattle market.

In spite of the transformations, Brazilian commodities futures markets still have some characteristics intrinsic to emerging markets. These usually have a lower number of transactions and trading volume if compared to more developed ones. The most worrisome consequences of this deficiency are increased liquidity

costs and reduced returns (Lesmond, 2005). Besides, low liquidity levels may hinder price discovery, increase the chances of high volatility and price manipulation. As a result, markets may become dysfunctional and have its hedging and trading activities undermined. Therefore, well-functioning futures markets are required to keep liquidity at an acceptable level. (Wang, Garcia and Irwin 2013). In this sense, liquidity is a key factor to the development of emerging commodities markets and understanding its determinants is urgent and crucial.

Another relevant factor in the microstructure of commodities futures markets is the degree of concentration among its traders and dealers. This topic is of particular interest for markets with lower levels of liquidity, since it is argued that increased levels of concentration may potentially worsen market quality, resulting in higher bid-ask spreads and volatility and lower volume (Branch and Freed 1977; Hamilton 1979; Cohen and Conroy 1990; McNish and Wood 1996).

As a result of advances in technology, different sets of high frequency data have been made available, and as a consequence, underlying market characteristics have been exposed. These two facts combined fueled a rise in the number of studies dedicated to understand the microstructure of commodities futures markets (Eaves and Williams 2010; Martinez et al. 2011; Shah and Brorsen 2011; Kauffman 2013; Wang, Garcia and Irwing 2013; Lehecka, Wang and Garcia 2014). However, little is known nowadays about the microstructure of commodities futures markets in emerging countries, where markets are thin or simply do not exist, and high frequency data are not always available. Brazil is possibly an exception to this rule, being an emerging market with good quality data.

This paper's contribution is to understand one key element for market efficiency, which is the behavior of liquidity costs (proxied with bid-ask spread) in emerging financial and commodities futures markets and its relation with volume and volatility. Using the observed bid-ask spread from the BM&F Bovespa for March 2014-February 2016, this is the first paper to measure liquidity costs in the intraday level for commodities markets with varying levels of liquidity, which is this case the Brazilian corn and live cattle markets, and to analyze its joint structural determination with volume and volatility. The fact that these markets exhibit different levels of liquidity allows us to proceed a comparative analysis. To the best of our knowledge, this paper is also the first to gauge the impact of market concentration on liquidity costs, volume and volatility in emerging commodities futures markets. The period of analysis includes the 2014/2015 record corn harvest of 84.7 million tons and the subsequent smaller and troubled 2015/2016 harvest. The latter was of roughly 76.2 million tons and was affected mainly by severe weather conditions and drops in productivity, according to CONAB. The live cattle production also exhibited a downward trajectory during this period, going from 49.6 in 2014 to 48.2 million heads in 2015. USDA points that the decrease in production is due to high retention of cows and low volume of rains in the most important producing areas. Further, during this period BM&F Bovespa had its electronic platforms fully operating, with most of the commodities contracts ending in financial liquidation. These market structure changes made access to these markets easier to traders and dealers and paved the way to enhanced trading and hedging performances.

I estimated a three-equation structural model using the IV-GMM approach to investigate the relationship of BAS with volume and volatility (Wang and Yau 2000; Martinez et al. 2011; Wang, Garcia and Irwin 2013). All equations include the concentration variable, measured by the Herfindahl index as a way to account for market concentration effects on BAS, volume and volatility. The analysis identifies patterns in liquidity costs varying according to the level of liquidity in those markets. Because of its higher volume, the corn

futures market exhibited lower liquidity costs than the live cattle market. The relation between BAS, volume and volatility tends to mimic that of more developed futures markets as liquidity levels increase. By more developed futures markets I mean those with high levels of liquidity and number of transactions. I also found concentration to have different impacts of the corn and live cattle markets. In the former, an increase in concentration leads lower levels of volume and BAS, and in the latter, the negative impact of concentration on volume was significantly more than in the corn market.

3.2. Literature Review

The first studies of liquidity costs in financial and commodities markets measured the bid-ask spread (BAS). The widespread conclusion of these studies is that the BAS is affected negatively by volume and positively by volatility. This pattern has also been observed in agricultural markets, however with some exceptions.

The negative association between trading volume and BAS is supported by a wide range of studies. Thompson, Eales and Seibold (1993) estimated BAS and compared liquidity costs in the wheat futures markets both in Chicago and Kansas City futures exchanges. They found higher liquidity costs in the latter and attributed it to its lower trading volume. Wang, Yau and Baptiste (1997) analyzed agricultural, metals and financial futures contracts, finding BAS to be negatively correlated with trading volume and positively correlated to price variability. Besides, trading volume associated to agricultural futures was found to be more inelastic to variations in liquidity costs than trading volume associated to non-agricultural contracts.

More recently, Frank and Garcia (2011) used Bayesian methods to measure liquidity costs in the live cattle and lean hog futures market at CME. Their findings show that live cattle contracts have lower liquidity costs, due to its higher levels of trading volume and volume per transaction. The authors also found a positive correlation between price volatility and nearly all bid-ask spread estimations. Shah, Brorsen and Andersen (2012) reached a similar conclusion after comparing wheat futures and options market in Kansas City. The wheat futures markets exhibited a lower BAS due its higher volume level. In addition, it was found that liquidity costs in options markets tended to increase as options premia increased. Wang, Garcia and Irwin (2013) analyzed the behavior of BAS in CME's electronically traded corn futures market. They verified high levels of liquidity in the new platform and a relatively low BAS. The authors also confirmed the negative relationship between volume and BAS and the positive relationship between BAS and volatility, pointing to the fact the nature of the relation between BAS, volume and volatility remains unchanged even if subject to increased liquidity.

In studies on stocks, equities, and financial derivatives markets, trading volume and price volatility have been identified as the two main determinants of the bid-ask spread. The relationship between price volatility and trading volume is well documented in the literature and revolves around the Simultaneous Information Arrival Hypothesis (Copeland 1976; Jennings, Starks, and Fellingham 1981) and the Mixed Distribution Hypothesis (Clark 1973; Tauchen and Pitts 1983). The joint behavior of BAS, trading volume and price volatility can be explained by the inventory cost model (Stoll, 1978), the information asymmetry model (Copeland and Galai 1988), and the order processing model (Demsetz 1968; Tinic 1972). The common conclusion is the positive relation between BAS and price volatility, while a negative relation between BAS and volume (Wang and Yau, 2000). This pattern has been observed in commodities markets, with some exceptions, however. Regarding

emerging commodities markets, the lower levels of liquidity could potentially influence the behavior of volume and volatility and make it deviate from the pattern observed in other markets. In fact, Atilgan et al. (2016) argue that it has been observed that a positive relationship between volume and volatility in several emerging markets.

In the past few years commodities futures markets have undergone significant structural changes which resulted in a progressive increase in volume and open interest as well as a diversification of trades and participants (Irwin and Sanders, 2012). Much of the increase in volume followed the move to electronic trading pits, which made markets more accessible and trading easier. However, no consensus has been achieved in the literature about the effect of electronic trading on the relationship between BAS, volume and volatility (Frank and Garcia, 2010).

A number of studies compared electronic and outcry trading in commodities futures. (Martinez et al. 2011; Shah and Brorsen 2011; Wang, Garcia and Irwin 2013). All studies concluded that the introduction of electronic trading and the new volume influx and lower price volatility had a significant impact on lowering BAS. In addition, Martinez et al. (2011) found that volume migrated from the outcry system to the electronic system and they also found the price discovery occurs firstly in the latter; Shah and Brorsen (2011) detected more volatility clusters in the outcry system than in the electronic system.

Bryant and Haigh (2004) found different results analyzing the relationship between BAS and price volatility in the early 2000s, when the transition to electronic platforms had just started. They argued that the positive relation between trading volume and BAS is explained by the substantial increase in price volatility and adverse selection suffered by market makers. In this context, Wang and Yau (2000) identified endogeneity in the relationship between BAS, trading volume, and price volatility. Analyzing two financial and two metals futures contracts traded at CME and COMEX, they found BAS, volume, and volatility to be dynamically and simultaneously determined. A system of dynamic structural equations was also used by Martinez et al. (2011) and Wang, Garcia and Irwin (2013).

Another aspect of the market microstructure analysis that can influence the relationship between BAS, volume and volatility is the level of concentration in the market, measured based on the number of dealers, dealers and market makers. Unlike equities and securities, there are no market makers operating in BM&F Bovespa FX, interest rate, live cattle and corn futures markets. Therefore, all the trade is carried out within an inter-dealership market. Literature has focused the behavior of the market maker acting as a monopolist (Tinic 1972; Ho and Stoll 1981). Ho and Stoll (1983) extended the Ho and Stoll (1981) model to allow for the existence of multiple market makers under competition. Among their conclusions is that the impact of volatility on BAS will be stronger as more concentrated the market is. Hagerty and McDonald (1996) studied the behavior of dealers in a market microstructure frame and found that monopolistic dealers charge traders market BAS while competitive dealers may charge traders a lower spread based on how informed they are. Studies regarding the impact of market concentration on other aspects of market microstructure are more recent. Klock and McCormick (2002) found that the relation between quote aggressiveness and order flow is stronger if the market is competitive. Van Ness et al. (2005) investigated the impact of market concentration on adverse selection costs using data from Nasdaq. They concluded that more competitive markets have lower adverse selection costs because of market makers' knowledge of order flows. King et al. (2013) analyze the evolution of FX markets and argue that concentration increased after the move to electronic platforms, since the investment to trade on

these platforms are high and small traders cannot afford it. Besides, unlike other financial markets, where market makers are usually present, dealers and dealers have a prominent role in FX markets.

There are not many empirical studies on the impact of concentration on market quality at the microstructure level probably due to the lack of available data containing information on the firms trading in the market (Branch and Freed 1977; McNish and Wood 1996). These authors were mainly concerned with the increased competition between exchanges and fragmentation of order flow. Except for Mendelson (1987), who concluded that concentration increases traded quantity and reduces volatility, all authors concluded that competition has a significant effect on reducing the bid-ask spread. Further, Branch and Freed (1977) argue that less concentration has a bigger impact on reducing spreads than volume. More recently, Kröber et al. (2013), analyzing equity markets in the London Stock Exchange, found that most market quality measures improve with less concentration. More specifically, volatility was significantly reduced with more fragmentation and competition. Trading volume was reduced under less concentration, but global volume increased.

Studies regarding the behavior of liquidity costs in emerging commodities markets have concluded that it does not exhibit gross deviations from what has been observed in more developed commodities markets, despite the lower liquidity levels. Liu et al. (2016) analyzed the behavior of liquidity costs in the copper, aluminum, gold and rubber futures markets at the Shanghai Futures Exchange and found the estimated BAS to be negatively related to trading volume and skewness, but positively related to volatility and kurtosis. In Brazil, Marquezin (2013) analyzed liquidity costs for the soybeans futures market at BM&F-Bovespa and found them to be relatively high. The author also found that these costs have a negative relationship with trading volume, open interest, the number of days to the expiration; and a positive relation with the mean volume per contract. Marquezin and De Mattos (2014) estimated the liquidity costs of the live cattle futures market at BM&F- Bovespa using four different estimators. They found relatively low liquidity costs and in addition, a negative relation between volume and liquidity costs was found, as expected.

3.3. Structure of BM&F Bovespa Data

Data by BM&F-Bovespa at a tick-by-tick frequency is organized in three files. The first file contains the information related to the buy offers (OFER_CPA), and the second file the sell offers (OFER_VDA). The variables in these two files are the date and time of the event, to the milliseconds, price of the (buy and sell) offers, volume, order number, state of the offer (new, update, cancel, traded, etc.), and the contract identifier¹, and the buy or sell sequence number. The last file (NEG) records all transactions and in addition to all variables cited previously, it contains the sequence numbers relative to the buy and sell orders that compose each transaction. Therefore, all three files are connected through the sequence numbers.

Since the main purpose of this paper is to identify the determinants of liquidity costs across emerging futures markets, an accurate definition of the liquidity costs is necessary. In the literature, the most used proxy for liquidity costs is the bid-ask spread. I used data from the transaction, buy offers, and sell offers files, I measured the bid-ask spread directly, by reconstructing the order book and the top of the book. This is a distinct

¹ i.e., BGIH11, where BGI is the code for live cattle futures, H is the maturity month, in this case March, and 11 is the maturity year

feature, since a significant number of studies use estimators based on transaction data to calculate the bid-ask spread (Roll 1984; Thompson and Waller 1987; Shah and Brorsen 2011; Frank and Garcia 2011; Martinez et al. 2011; Marquezin 2013; Marquezin and De Mattos 2014).

For the analysis, I picked two different contracts traded at BM&F-Bovespa, namely the corn (CCM) and live cattle (BGI) contracts. More specifically, I picked the two most traded contracts in every month, regardless of their expiration date. The first one is a highly traded contract and the second one relates to the mildly traded contracts. Doing this, I accounted for the level of liquidity in each market and capture potential heterogeneity that may exist across emerging markets. The ultimate outcome comprehends a thorough analysis of the BAS and its determinants in this context.

The live cattle contract is traded electronically from 9:00am-4:00pm, and the corn contract from 9:00am-3.30pm- Brasilia time. In respect to contract specifications, the contract size for corn is 450 bags of 60 net kilograms and R\$0.01 tick size and 4407 net kilograms and R\$0.01 tick size for live cattle. Corn contracts maturity months are January, March, May, July, August, September, and November. Live cattle contracts expire every month.

The period of analysis starts in March 3rd 2014 and ends in February 29th 2016, totaling 24 months. In order to obtain the bid-ask spread, I created a database gathering information from all the three files. For every transaction in the transactions file, we have the sequential numbers of the corresponding bid and ask offers. I defined the bid-ask spread as the difference between the bid and ask offers immediately before the liquidation of the transaction, that is, when the bid offer meets the ask offer and trade occurs.

3.4. Variables of Interest

Our database is composed of the following variables:

1. Sequence number of the bid offer;
2. Sequence number of the ask offer;
3. Date and time;
4. Bid Price;
5. Ask Price;
6. Volume;
7. Contract identifier;
8. Order number;
9. Transaction status.
10. Spread

For each transaction in the transactions file I identified the bid and ask offers associated to it in the buying offers file and in the ask offers file. Using variables date and time and transaction status, it was possible to identify the transaction's life cycle, from the moment it was created to the moment a deal was achieved. The bid price, ask price and the related volume were taken from the buy offers and sell offers files and inform, respectively, the bid and ask prices immediately before the transaction ends, that is, a trade occurs. The spread variable is the bid-ask spread, that is, the difference between the bid and ask prices.

3.5. Methodology

I investigated the relationship between BAS, volume and volatility in emerging futures markets considering a potential endogeneity bias. I used a three-equation model in which these three variables are jointly determined (Hausman 1978; Wang and Yau 2000; Martinez et. al 2011; Wang, Garcia, and Irwin 2013). For this purpose, I followed Wang and Yau (2000) and Wang, Garcia and Irwin (2013) and use a GMM model three - equation model:

$$(1) \text{BAS}_{i,t} = \beta_0 + \beta_1 \text{Volume}_{i,t} + \beta_2 \text{Volatility}_{i,t} + \beta_3 \text{BAS}_{i,t-1} + \beta_4 \text{Months}_{i,t} + \beta_5 \text{Days of the Week}_{i,t} + \beta_6 \text{Concentration}_{i,t} + \varepsilon$$

$$(2) \text{Volume}_{i,t} = \beta_0 + \beta_1 \text{BAS}_{i,t} + \beta_2 \text{Volatility}_{i,t} + \beta_3 \text{Volume}_{i,t-1} + \beta_4 \text{Months}_{i,t} + \beta_5 \text{Days of the Week}_{i,t} + \beta_6 \text{Concentration}_{i,t} + \varepsilon$$

$$(3) \text{Volatility}_{i,t} = \beta_0 + \beta_1 \text{BAS}_{i,t} + \beta_2 \text{Volume}_{i,t} + \beta_3 \text{Volatility}_{i,t-1} + \beta_4 \text{Months}_{i,t} + \beta_5 \text{Days of the Week}_{i,t} + \beta_6 \text{Concentration}_{i,t} + \varepsilon$$

In the model described above, *BAS* is the daily average of bid-ask spreads. I calculated the bid-ask spread updated by every incoming order. The bid prices were obtained from the buy offers file, the ask prices from the sell offers file. A similar, however different, approach was carried out by Barros and Fernandes (2014) when analyzing market depth for stocks traded at BM&F Bovespa. *Volume* corresponds to the daily sum of the volumes associated to each traded transaction, and *Volatility* is the daily standard deviation of traded prices. I controlled for seasonality including dummies for each month of the year. Following Garcia and Frank (2010) and Wang, Garcia and Irwin (2013) I also included day of the week dummies. As in Wang and Yau (2000), all variables were put in the log form, except for the *Month* and *Day of the Week* variables. This transformation allowed us to readily interpret the coefficients as elasticities and also contributes to stabilize the variance of the error terms.

In addition, considering the intrinsic characteristics of emerging markets, we include the *Concentration* variable, which accounts for the intraday Herfindahl-Hirschman index. The HHI index is commonly employed in the economics literature to assess the concentration of industries. In the market microstructure literature, it was first used by Ticnic (1972). In this paper, it is calculated using the volume traded by each market maker (firm) in the market for every trading day:

$$(4) \text{MS}_i = \frac{x_i}{\sum_{j=1}^n x_j}$$

$$(5) \text{HHI}_i = \sum_{i=1}^n \text{MS}_i^2$$

Where x_i is the participation of each of the n market maker. Through equation (4) I obtained the market share of each market maker for each trading day. In equation (5) the market share of each firm is squared and then summed up.

Regarding expected signs of the coefficients in the three-equation model, it has been observed in the literature that increases in volume tends to decrease the BAS, whereas increases in volatility tends to increase it. Then, in equation (1), the volume coefficient is expected to have a negative sign and the Volatility coefficient a positive sign. (Frank and Garcia 2010; Shah, Brorsen and Andersen 2012). The Concentration coefficient is expected to have a negative sign, as more competitive markets bring about lower levels of BAS (Branch and Freed 1977; Hamilton 1979; Cohen and Conroy 1990; McInish and Wood 1996, Van Ness et al. 2013).

In equation (2), BAS is expected to have a negative coefficient because higher transaction costs lead to lower profitability and consequentially lower volume (Wang, Garcia and Irwin, 2013). On the other hand, it is well known in the literature that volatility has a positive impact on volume, either contemporaneously, giving support to the Mixed Distribution Hypothesis (Clark 1973; Tauchen and Pitts 1983) or through lagged values, as states the Sequential Information Hypothesis (Copeland 1976; Jennings, Starks, and Fellingham 1981; Smirlock and Starks 1984). The impact of Concentration on volume can be dubious. Mendelson(1987) and Kröber et al. (2013) found Concentration to positively impact volume, whereas McInish and Wood (1996) found the opposite.

In equation (3), the BAS coefficient is expected to be positive, since a wider BAS naturally leads to a higher volatility in prices. As stated previously, a positive relationship is expected between volume and volatility, however, a negative signed coefficient is expected to the lagged volume variable, due to a potential overreaction to new information (Wang, Garcia and Irwin, 2013). Concentration is expected to positively impact Volatility (Branch and Freed 1977; Hamilton 1979; Cohen and Conroy 1990; McInish and Wood 1996).

For each of the markets analyzed I performed the unit root augmented Dickey-Fuller and the Durbin- Wu-Hausman endogeneity test among BAS, volume, volatility, and Concentration.

The outcome of the test will define what methodology will be used in each case. In the case of presence of endogeneity among the three variables, I used GMM-Instrumental Variables model to estimate the three- equation system described at the beginning of this section (Wang and Yau 2000; Martinez et al. 2011; and Wang, Garcia and Irwin 2013). If no endogeneity is found, then I estimated the three-equation model using either a Seemingly Unrelated Regression (SUR) if the error terms of the equations are correlated or Ordinary Least Squares (OLS) if the error terms are uncorrelated. For the Hausman's test I adopted the same procedure as in Martinez et al (2011). For the BAS equation (1), I tested whether the variables *Volume* or *Volatility* are endogenous. For this purpose, I regressed *Volume* on its set of explanatory variables as shown in equation (2), except for *BAS* and *Volatility*, and generate the regression residual ε_1 . The same procedure is carried out for *Volatility*, and then, leaving out *BAS* and *Volume* in equation (3), the regression residual ε_2 is kept. The next step is to run the BAS equation (1) including the two residuals. Based on the F-test, if they are jointly significantly, then *Volume* and *Volatility* are endogenous in the BAS equation.

3.6. Preliminary Analysis

Table 1 shows summary statistics of the two contracts in our analysis. Corn, and live cattle samples have 474, and 467 daily observations. The variables presented below are continuous in time and correspond to

the most traded contracts in each month. The corn contract is more traded than the live cattle contract, with an average of 2192.3 contracts traded by day. It is roughly 16% more than the average daily volume of live cattle markets. The average BAS and volatility follow an opposite direction to that of volume throughout all markets. It is expected that average volatility in thinly traded markets is higher than in more liquid markets (Adjemian, 2016). In this case, the average daily volatility observed in the live cattle market exceeds the volatility observed in more traded corn market.

Table 1 – Descriptive Statistics

	Live Cattle				Corn			
	Mean	Std. Dev.	Min.	Max.	Mean	Std. Dev.	Min.	Max.
BAS	0.1723	0.0839	0.0529	1.0855	0.0724	0.0225	0.0338	0.1512
Volume	1889.401	1428.317	33	13847	2192.365	1312.06	110	8803
Volatility	1.9995	8.2957	0.0203	74.29	0.2547	0.4739	0.0331	4.3915
Concentration	0.2459	0.0978	0.1133	0.8256	0.2741	0.0868	0.1143	0.6709

Figure 1 reports the level of market concentration in all three futures markets. I identified in our sample 41 dealers trading in the corn market, and 42 in the live cattle market in the same period. The corn futures markets is slightly more concentrated than the live cattle market, with an average Herfindahl index of 0.27 against 0.24 for the latter. Figure 1 shows the HHI index along the trading days for each market. The red line in Figure 1 indicates what the level of concentration would be if the markets were competitive.

Figure1 – Daily Herfindahl index for live cattle and corn markets, conversely².

The results for the unit root tests are reported in Table 2. The null hypothesis of the Augmented Dickey-Fuller test is rejected for the BAS, volume, volatility, and concentration variables in all contracts at the 1% level. Therefore, all series are stationary and no differencing is needed.

² This figure shows the daily Herfindahl (HH) index for the live cattle (right) and corn (left) markets. The index was calculated based on the quantities bought and sold by all brokers along the trading day. Each point on the graph represents one trading day and its HH index.

Table 2 - Augmented Dickey-Fuller test.

	BAS	Volume	Volatility	Concentration
Live Cattle	-11.703 ***	-11.238 ***	-12.728 ***	-13.748 ***
Corn	-9.667 ***	-10.333 ***	-12.293 ***	-17.364 ***

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

Following Wang and Yau (2000), Martinez et al. (2011) and Wang, Garcia and Irwin (2013), I used GMM-Instrumental Variables model to estimate the three-equation system described at the beginning of this section. In equation (1), I used lagged volume and the first difference of volatility as instruments (Wang, Garcia and Irwin 2013; Frank and Garcia 2011); in equation (2) I used lagged BAS and the first difference of volatility as instruments (Wang, Garcia and Irwin 2013); and in equation (3), the instruments were lagged BAS (Martinez et al. 2011; Wang, Garcia and Irwin 2013) and lagged concentration. Since I have found no consensus in the literature for an instrumental variable for concentration, I followed the same procedure adopted for the other variables and used lagged concentration. Prior to estimation, I applied the Dickey-Fuller test for stationarity and the augmented Durbin-Wu-Hausman test for endogeneity bias among BAS, volume and volatility. Post estimation tests include the Cumby-Huizing modified Breusch-Godfrey test for autocorrelation and the Pagan- Hall test for heteroskedasticity. The results of the Breusch-Godfrey test for autocorrelation and the Breusch- Pagan for heteroskedasticity are disposed in Table 3. I found mixed results in both tests. Autocorrelation was detected in the volatility equation for live cattle and in the volatility and BAS equations for corn. The null of homoscedasticity was rejected mostly in volume and volatility equations, for both markets.

Table 3 – Heteroskedasticity, Autocorrelation, and Weak Identification tests.

	Live Cattle			Corn		
Test/Equation	BAS	Volume	Volatility	BAS	Volume	Volatility
	t value	t value	t value	t value	t value	t value
Heteroskedasticity	37.241***	55.419***	49.131***	23.586	35.642**	50.181***
Autocorelation	0.717	2.551	8.210**	4.666**	0.785	3.707*
Weak Identification	23.773	244.304	9.212	11.916	265.171	120.042

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

I also tested for endogeneity among BAS, volume, volatility, and Concentration. For this purpose, I applied the the Durbin-Wu-Hausman test for endogeneity and the outcome is presented in Table 4. Overall, I found more endogeneity issues in the less liquid live cattle market than in the more liquid corn market. For the former, volume and volatility are endogenous in the BAS equation, volatility was found endogenous in the volume equation and BAS, volume, and concentration were found endogenous in the volatility equation. In the corn system of equations, concentration is endogenous in the BAS equation, volatility is endogenous in the volume equation, and volume is endogenous in the volatility equation. I applied the Stock-Yogo test for weak instruments and could reject the null of large bias, meaning that all instruments are well identified. Results for the weak identification can be found in Table 3.

Table 4 – Durbin-Wu-Hausman endogeneity test

Variable/Equation	Live Cattle			Corn		
	BAS	Volume	Volatility	BAS	Volume	Volatility
	t value	t value	t value	t value	t value	t value
BAS		0.022	0.085		0.362	0.306
Volume	6.792***		17.261***	0.116		6.664***
Volatility	10.873***	12.821***		2.227	3.037*	
Concentration	0.127	0.032	2.742*	3.545*	1.407	0.054

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

3.7. Regression Results: *BAS, Volume, Volatility and Concentration Estimates*

The results obtained differ according to the level of liquidity in each market. In general, I found coefficients to be more statistically significant in the corn market, which is the most traded one. Lagged variables followed this same trend. Lagged BAS, lagged volume, and lagged volatility were all significant in the corn and live cattle markets. In particular, lagged BAS had the highest coefficient among all lagged variables, meaning that it is the most persistent variable in both corn and live cattle markets. Martinez et al. (2011) and Wang, Garcia, and Irwin (2013) also found significant persistence of BAS and volatility in the CME corn futures market, and lagged BAS showed to be the most persistent variable in that case.

Table 5 – IV-GMM Output

<i>BAS Equation</i>	Live Cattle		Corn	
	<i>Coefficient</i>	<i>z</i>	<i>Coefficient</i>	<i>z</i>
Constant	1.1086*	1.7300	0.0194	0.0400
Volume	0.1006	1.3300	-0.1387***	-5.2200
Volatility	-0.0703**	-2.3100	0.0700***	4.4700
Lagged BAS	0.4973***	10.1000	0.4201***	7.9500
Concentration	0.0455	0.7500	-0.4463**	-1.9800
January	-0.2985***	-3.9700	-0.0123	-0.2400
February	-0.1986***	-2.5600	0.0030	0.0500
March	-0.1828**	-1.9800	-0.1150**	-2.0900
April	-0.2821***	-2.9400	-0.1292**	-2.4700
May	-0.3320***	-4.1000	-0.1748***	-2.9800
June	-0.2427***	-2.6600	-0.0991*	-1.8000
July	-0.2264***	-2.5700	-0.1266**	-2.2900
August	-0.1388*	-1.6800	-0.1430**	-2.5500
September	-0.3302***	-3.5000	-0.0586	-1.1300
October	-0.2313***	-2.9200	-0.0086	-0.1700
November	-0.0837	-1.0700	-0.0220	-0.4000
Monday	0.0956*	1.9000	-0.0667**	-1.9400
Tuesday	-0.1225**	-2.5000	-0.1019***	-3.0800
Wednesday	-0.0379	-0.8000	-0.0183	-0.5400
Thursday	-0.0616	-1.2700	0.0165	0.4900
<i>Volume Equation</i>				

Constant	1.0723	0.9700	1.2971	1.4100
BAS	-0.3790***	-4.8500	-0.6423***	-7.1000
Volatility	0.2585***	7.4200	0.3493***	7.2900
Lagged Volume	0.3207***	7.9100	0.4039***	10.4000
Concentration	-0.3606***	-4.2600	-0.1330*	-1.7900
January	0.0080	0.0600	0.0587	0.5700
February	-0.1242	-0.9300	0.1638	1.5600
March	0.4600***	3.4100	0.1072	0.9700
April	0.4855***	3.6400	0.0157	0.1500
May	0.1824	1.3700	0.2994***	2.6900
June	0.4122***	3.0800	0.3942***	3.6300
July	0.3261**	2.4100	0.2215**	2.0100
August	-0.0892	-0.6100	-0.1025	-0.8500
September	0.4772***	3.6200	0.2869***	2.7700
October	0.2532**	2.0000	0.2311**	2.2700
November	0.2574**	2.0300	0.0598	0.5700
Monday	-0.1440*	-1.7200	-0.0827	-1.2000
Tuesday	0.3093***	3.7700	0.1254*	1.8500
Wednesday	0.2116***	2.6500	0.1631**	2.4600
Thursday	0.1505*	1.8500	0.1089	1.6200

<i>Volatility Equation</i>				
Constant	0.1461	0.0600	-2.6154**	-2.0300
BAS	-0.1854	-0.8700	0.5015***	3.3300
Volume	-0.0595	-0.1800	0.2533**	2.1800
Lagged Volatility	0.3892***	8.3200	0.3117***	6.8400
Concentration	0.5661	0.6400	0.0066	0.0600
January	-0.5066*	-1.8100	0.1431	0.9800
February	-0.6707**	-2.1900	-0.0547	-0.3600
March	-0.0025	-0.0100	-0.0457	-0.2900
April	-0.0086	-0.0300	-0.0379	-0.2500
May	0.0234	0.0700	-0.2420	-1.4900
June	0.1150	0.3400	-0.1666	-0.9900
July	0.2556	0.6800	0.0483	0.2900
August	1.1884***	3.6300	0.8548***	5.3000
September	-0.2546	-0.7800	-0.0531	-0.3400
October	0.1872	0.5500	0.0464	0.3000
November	-0.2122	-0.7300	0.2262	1.5200
Monday	-0.3710*	-1.9300	0.0572	0.5800
Tuesday	-0.1472	-0.8200	-0.1063	-1.1200
Wednesday	-0.0463	-0.2600	-0.1061	-1.1200
Thursday	-0.2296	-1.3000	-0.0226	-0.2300

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

As expected, I found a positive relationship between volume and volatility in the corn and live cattle markets. Positive changes in volume bring new information to the markets. In face of the new information, both

traders and liquidity providers (dealers) adjust their positions, which contributes to increase price volatility (Wang, Garcia and Irwin 2013). The fact that volume and volatility are contemporaneously correlated gives support to the Mixed Distribution Hypothesis (Clark 1973; Tauchen and Pitts 1983). In the corn volume equation, a 10% increase in volatility causes a 3,49% increase in volume. In the case of live cattle futures, a 10% increase in volatility triggers a 2,58% increase in volume. On the other hand, the impact of a change in volume on volatility is smaller than the other way around. In the corn market, a 10% increase in volume causes a 2,53% increase in volatility and no impact of volume on volatility was found in the live cattle market. The volume and volatility created by the new information arrival will affect BAS differently, depending on the nature of the information shock and on the depth of the market. The more liquid the market, the smaller tends to be the impact of volume and volatility on BAS. (Wang, Garcia and Irwin, 2013).

The relationship between BAS and volume was different in each market, but overall, I found that BAS has a bigger impact on volume than the contrary. Particularly, I found that volume has no impact on BAS in the live cattle market, but its impact on the corn market is significant. The existence of a stronger bond between volume and BAS in the latter may be justified by its higher liquidity levels. Further, coefficient signs were consistent with expectations, meaning that there is a negative correlation between the two variables: increases in volume tends to reduce BAS levels. More specifically, in the corn market, a 10% increase in volume decreases BAS in 1.38%. On the other hand, in the same market, BAS has a much bigger impact on volume as a 10% increase in BAS reduces volume in roughly 6.42%.

Interestingly, this pattern is rather different in the live cattle market. The volume coefficient in the live cattle BAS equation was not statistically significant, but had a positive sign. A positive relationship between BAS and volume was found by Bryant and Haigh (2004) in coffee and cocoa futures traded at LIFFE. A possible explanation is that an increase in volume triggers an increase in price volatility and following adverse selection problems. The final effect is an increase in the BAS. The bigger impact of BAS on volume than the other way around suggests that liquidity providers (market makers or dealers) do not simply react to changes in volume, but rather use the BAS to manage their order inventory. Posting more aggressive bid and ask prices, liquidity providers attract order flow, increase volume and reduce transaction costs. On the other hand, despite of having observed no impact of volume on BAS in the live cattle market, the contrary is true. Increases in BAS tends to reduce volume, however in a smaller magnitude as observed in the corn market. Indeed, a 10% increase in BAS tends to reduce volume in 3.79%.

The literature has broadly identified a positive relationship between BAS and volatility. Such relationship was also observed in the corn market, having BAS more impact on volatility than the contrary. It is so that wider BAS allows prices to vary more, thus leading to a heightened volatility. A 10% increase in BAS increases volatility in roughly 5%. Conversely, a 10% increase in volatility causes an increase in BAS of about 0.7%. The higher sensibility of volume and volatility with respect to changes in BAS was also found by Wang and Yau (2000) and Wang, Garcia and Irwin (2013). I found volatility to have a negative impact on BAS in the live cattle market. A 10% increase in volatility reduces BAS in roughly 0.7%. However, BAS has no impact on volatility, since its coefficient is not statistically significant. Therefore, in the corn market, changes in volatility tends to increase BAS while in the live cattle market, changes in volatility tend to reduce BAS. However, changes in BAS brings about spikes in volatility only in the corn market.

Regarding the impact of concentration on BAS, volume and volatility, our results are roughly in line with the literature. Regarding the impact of concentration in the corn market, I found that a 10% increase in concentration reduces volume in about 1.33% and BAS in about 4.46%, and apparently has no effect on volatility. The negative relationship between concentration and volume can be largely found in the literature (Branch and Freed 1977; Hamilton 1979; Cohen and Conroy 1990; and McInish and Wood, 1996). The negative relationship between concentration and BAS, however, is quite unique. It means that as market concentration increases, bid-ask spread levels tend to decrease. Most studies in the literature points out that increases in concentration drives up BAS levels, which is not the case in the corn markets at BM&F-Bovespa. In the live cattle market, a 10% increase in concentration tends to decrease volume in 3.59%, roughly two and a half times more in comparison to the corn market. Therefore, an increase in concentration has a bigger negative impact on the volume traded in live cattle markets than it has in corn markets. No significant relationship was found between concentration and BAS and concentration and volatility in the live cattle market.

In conclusion, concentration among dealers seems to have different effects in the corn and live cattle futures markets. In the first, more concentration means less volume and smaller levels of BAS, in the latter it means less volume. The implication of this finding is pertinent to market regulators since dealers competition may not necessarily contribute to enhancing market quality. In the latter, the impact of concentration on volume had a bigger magnitude than the impact found in the corn market.

3.7.1. Seasonality Analysis

I verified the existence of seasonality patterns including dummy variables for months and days of the week. As expected, the results are significant and differ from market to market. In the corn market, BAS is lowest in the months of May and August. In the volume equation, positive and significant coefficients were found for May, June, July, September and October. Volume increases between May and June, when it reaches its highest point. The second highest volume coefficient is found in September. The monthly effects were less pronounced in the volatility equation. Volatility is lowest during the months of May and June, what confirms the negative relation between volume and volatility. The highest volatility is found in August. Day-of-the-week effects suggest that BAS is lowest on Tuesdays and volume is highest on Wednesdays. No significant effects were found in the volatility equation. These findings are in line with the Brazilian corn production cycle. The country currently has two corn harvests in the year. The first is the summer crop, in which harvest occurs in the first half of the year between the months of January and April. The winter crop is mostly harvested in the second half of the year, between May and August (Silveira and Mattos 2015). Therefore, the two most significant points of low BAS, high volume and low volatility occur at the end of the summer and winter harvest cycles.

The seasonality found in live cattle futures markets was quite similar to the one found in the corn futures market. I found BAS to be lowest in May and second-lowest in September. In the volume equation, the highest coefficients were found for the months of April and September and in the volatility equation the highest coefficient was found in the month of August. Therefore, April and September are the months when BAS and is lowest and volume is highest. Regarding day-of-the-week effects, like in corn markets, Tuesday is when the

BAS is lowest. It is also the day when volume is highest and no significant day-of-the-week coefficient was found for the volatility equation.

3.8. Conclusion

Trading commodities futures has changed significantly over the past few years. The introduction of electronic platforms made access to these markets easier and improved transparency. These transformations have also been incorporated in emerging commodities futures markets, even though the latter differ in some aspects from more mature markets, notably in the levels of liquidity. I investigated the behavior of the bid-ask spread (BAS) and its relationship with volume and volatility in the live cattle and corn futures markets at BM&F Bovespa. I proceeded a comparative analysis considering different levels of liquidity across markets and controlled for the impact of market concentration on the three variables using a structural equation framework.

Our findings demonstrate that the average BAS is lower for the corn market relative to the live cattle market. This pattern is justified by the lower level of liquidity in the latter if compared to the corn market. It is an evidence that maintaining lower levels BAS is crucial to the well-functioning of futures markets. Consistent with the literature, our analysis reveals that BAS responds negatively to changes in volume and positively to changes in volatility. However, the responses also differ in magnitude according to the level of liquidity in each market. In general, the more liquid corn market responded more to such changes, as its coefficients were in general more statistically significant than the ones in the live cattle market. Market concentration also had different impacts on each market.

Our results are generally in line with the findings in the literature. (Thompson, Eales and Seibold 1993; Martinez et al. 2011; Shah and Brorsen 2011; Wang, Garcia and Irwin 2013). However, the similarity depends on the level of liquidity of the markets analyzed. The more liquid the market at BM&F Bovespa, the closer the relation between BAS, volume and volatility will be to the already consolidated results. In this sense, the results found for the corn market are closer to the results found by Martinez et al. (2011) and Wang, Garcia and Irwin (2013) both in terms of magnitude and signals than the ones found for the live cattle market. This study also highlights the degree of market concentration as a relevant variable when analyzing the microstructure of emerging futures markets. The impact concentration can have on market quality depends on the market being analyzed and on its level of liquidity. Our findings point to the fact that an increase in concentration contributes to decrease volume in live cattle and corn markets. The negative impact of concentration on volume is bigger for the former. Besides, concentration also leads to lower levels of BAS in corn markets.

In addition to the higher level of liquidity observed in the corn futures market, another factor that may explain the lower BAS levels in this market is the presence of foreign investors trading in this market. Between March 2014 and February 2016, the share of contracts traded by foreign investors rose progressively in the corn futures markets. However, this same trend was not observed as intensively in the live cattle market. Foreign investors contribute to improving general market quality as they help generate lower BAS levels and increase competition among traders (Lee and Chung, 2016). The impact of foreign investment in Bovespa on concentration and the link with world market prices form an avenue for future research

Directions for further research include verify if the same pattern of relation between BAS, volume, and volatility also occur for other futures markets at BM&F Bovespa and the impact of market concentration on this structure. We also know little about market microstructure of commodities options markets at emerging exchanges and how similar it could possibly be to the commodities futures markets structure. Further, the recent surge in high frequency and algorithm trading could possibly be present also in emerging exchanges. Since these markets have

intrinsic characteristics, the impact of high frequency trading can potentially deviate from that observed in mature exchanges and are still largely unknown.

References

- Adams, Z.; Glück, T. (2015). Financialization of commodities markets: a passing trend or the new normal? *Journal of Banking and Finance*. Vol. 60, pp. 93-111.
- Atilgan, Y.; Demirtas, K.; Simsek, K. (2016). Derivatives markets in emerging economies: a survey. *International Review of Economics and Finance*. Vol. 42, pp. 88-102.
- Barros, C.; Fernandes, M. (2014). Market Depth at the BM&F Bovespa. *Brazilian Review of Econometrics*. Vol. 34, n.1, pp. 25-44.
- Branch, B.; Freed, W. (1977). Bid-ask spreads on the AMEX and the big board. *The Journal of Finance*. Vol. 32, n.1, pp. 159-163.
- Bryant, H. L., & Haigh, M. S. (2002). Bid-ask spreads in commodity futures markets. [Paper]. *Department of Agricultural and Resource Economics*, College Park, MD. Retrieved from <http://ageconsearch.umn.edu/bitstream/28587/1/wp02-07.pdf>.
- Clark, P. (1973). A subordinated process model with finite variance for speculative prices. *Econometrica*. Vol. 41, n.1, pp. 135-155.
- Comerton-Forde, C.; Rydger, J. (2004). A review of stock market microstructure. *Equity Market Microstructure Review*. Available at SSRN: <https://ssrn.com/abstract=710801> or <http://dx.doi.org/10.2139/ssrn.710801>
- Copeland, T. (1976). A model of asset trading under the assumption of sequential information arrival. *The Journal of Finance*. Vol. 31, pp. 1149-1168.
- Copeland, T.; Galai, G. (1983). Information effects on the bid-ask spread. *Journal of Finance*. Vol. 38, n.5, pp. 1457-1469.
- Demsetz, H. (1968). The cost of transacting. *The Quarterly Journal of Economics*. Vol. 82, n.1, pp. 33-53.
- Eaves, J.; Williams, J. (2010). Are intraday volume and volatility U-shaped after accounting for public information? *American Journal of Agricultural Economics*. Vol 92, n.1, pp. 212-227.
- Frank, J.; Garcia, P. (2011). Bid-ask spreads, volume, and volatility: evidence from livestock markets. *American Journal of Agricultural Economics*. Vol. 93, n. 1, pp. 209-225.
- Hagerty, K.; McDonald, R.L. 1996. Brokerage, market fragmentation, and securities market regulation. In: *The Industrial Organization and Regulation of the Securities Industry*. The University of Chicago Press.
- Hamilton, J. (1979). Marketplace fragmentation, competition, and the efficiency of the stock exchange. *The Journal of Finance*. Vol 34, n.1, pp. 171-187.
- Ho, T.; Stoll, H. (1981). Optimal dealer pricing under transactions and return uncertainty. *Journal of Financial Economics*. Vol. 9, pp. 47-73.
- Ho, T.; Stoll, H. (1983). The dynamics of dealer markets under competition. *The Journal of Finance*. Vol. 38, n.4, pp. 1053-1074.
- Irwin, S. H.; Sanders, D. R. (2012). Financialization and structural change in commodities futures markets. *Journal of Agricultural and Applied Economics*. Vol.44, n.3, pp.371-396.
- Jennings, R.; Starks, L.; Fellingham, J. (1981). An equilibrium model of asset trading with sequential information arrival. *The Journal of Finance*. Vol. 36, n.1, pp. 143-161.
- Kauffman, N. (2013). Have extended trading hours made agricultural commodity markets more risky? *Economic Review*, Q III, pp. 67-94.

- King, M.; Osler, C.; Rime, D. (2013). The market microstructure approach to foreign exchange: looking back and looking forward. *Journal of International Money and Finance*. Vol. 38, pp. 95-119.
- Klock, M.; McCormick, D.T. 1999. The impact of market maker competition on Nasdaq spreads. *The Financial Review*. Vol. 34, p. 55-74.
- Lehecka, G.; Wang, X.; Garcia, P. (2014). Gone in ten minutes: intraday evidence of announcement effects in the electronic corn futures market. *Applied Economic Perspectives and Policy*. Vol. 36, n.3, pp. 504-526.
- Lesmond, D. (2005). Liquidity of emerging markets. *Journal of Financial Economics*. Vol. 77, n.2, pp. 411-452.
- Liu, Q.; Hua, R.; An, Y. (2016). Determinants and information content of intraday bid-ask spreads: evidence from Chinese commodity futures markets. *Pacific-Basin Finance Journal*. Vol. 38, pp.135-148.
- Marquezin, C. (2013). Custo de liquidez do contrato future de soja na BM&F Bovespa, no período de 2010 a 2013 (Masters thesis). Retrieved from : <http://www.locus.ufv.br/handle/123456789/81?show=full>
- Marquezin, C.; De Mattos, L. (2014). Custo de liquidez do contrato futuro de boi gordo da BM&F Bovespa. *Revista de Administração Mackenzie*. Vol. 15, n.4, pp164-192.
- Martinez, V.; Gupta, P.; Tse, Y.; Kittiakarasakun, J. (2011). Electronic versus open outcry trading in agricultural commodities futures markets. *Review of Financial Economics*. Vol. 20, pp. 28-36.
- Mattos, F.; Silveira, R. (2015). Future price response to crop reports in grain markets. *The Journal of Futures Markets*. Vol. 36, n.10, pp. 923-942.
- Mendelson, H. (1987). Consolidation, fragmentation, and market performance. *The Journal of Financial and Quantitative Analysis*. Vol. 22, n.2, pp. 189-207.
- Mihaljek, D.; Packer, F. (2010). Derivatives in emerging markets. *BIS Quarterly Review*. December, pp. 43-58.
- Rada, N.; Buccola, S. (2012). Agricultural policy and productivity: evidence from Brazilian censuses. *Agricultural Economics*. Vol. 43, pp. 355-367.
- Roll, R. (1984). A simple implicit measure of the effective bid-ask spread in an efficient market. *The Journal of Finance*. Vol. XXXIX, n.4, pp. 1127-1139.
- Shah, S., Brorsen, B. W. (2011). Electronic vs . Open Outcry : Side-by-Side Trading of KCBT Wheat Futures, 36(1), 48-62.
- Stoll, H. (1978). The supply of dealer services in securities markets. *Journal of Finance*. Vol. 33, pp. 1133-51.
- Tauchen, G.; Pitts, M. (1983). The price variability-volume relationship on speculative markets. *Econometrica*. Vol. 51, pp. 485-505.
- Tinic, S. (1972). The economics of liquidity services. *The Quarterly Journal of Economics*. Vol. 86, n.1, pp. 79-93.
- Thompson, S.; Waller, M. (1987). The execution cost of trading in commodity futures markets. *Food Research Institute Studies*. Vol XX, n.2, pp. 141-163.
- Thompson, S.; Eales, J.; Seibold, D. (1993). Comparison of liquidity costs between the Kansas City and Chicago wheat futures contracts. *Journal of Agricultural and Resource Economics*. Vol. 18, n.2, pp. 185-197.
- Upper, C.; Valli, M. (2016). Emerging derivatives markets? *BIS Quarterly Review*, December, pp. 67-80.
- Van Ness, B.; Van Ness, R.; Warr, R. (2005). The impact of market maker concentration on adverse-selection costs for Nasdaq stocks. *The Journal of Financial Research*. Vol. XXVIII, n.3, pp. 461-485.

Wang, G.; Yau, J.; Baptiste, T. (1997). Trading volume and transaction costs in futures markets. *The Journal of Futures Markets*. Vol. 17, n.7, pp. 757-780.

Wang, G.; Yau, J. (2000). Trading volume, bid-ask spread, and price volatility in futures markets. *The Journal of Futures Markets*, Vol. 20, n.10, pp. 943-970.

Wang, X.; Garcia, P.; Irwin, S.H. (2013). The behavior of bid-ask spreads in the electronically traded corn futures markets. *American Journal of Agricultural Economics*. Vol. 96, n.2, pp. 557-577.

Wood, R.; McInish, T. (1996). Competition, Fragmentation and Market Quality. In: *The Industrial Organization and Regulation of the Securities Industry*. University of Chicago Press, p 63-92.

4. INTER-DEALERS MARKETS DYNAMICS AND COMPETITION: EVIDENCE FROM LIVE CATTLE MARKETS

This paper examines the dynamic relationship between dealers' activity and market microstructure in the live cattle inter-dealer market at BM&F-Bovespa. Firstly, a descriptive analysis of the live cattle inter-dealer market structure is carried out in order to answer questions such as: Is the live cattle market competitive in the sense of many dealers active in the price discovery process, or is there a dominant dealer? How important is the dealers' activity? Does market concentration affect dealers' profits and activity? The next step is an investigation about the dynamic of dealers' activity and its determinants. The behavior of dealers' activity is modelled using an instrumental variable probit model, giving value 1 when the dealer is active on any given day and 0 otherwise. Results indicate that the live cattle inter-dealer market is not competitive and that dealers' activity is positively related to market concentration, quoted bid-ask spread, number of active dealers and the dealer's traded quantity. Besides, results also point that the impact of the concentration and number of dealers on the bid-ask spread observed in the live cattle market.

Keywords: dealers; liquidity; competition

4.1. Introduction

Derivative contracts have always been an essential tool in commodities trading as they provide a way for producers and consumers to hedge the price risk of their spot positions and overcome markets frictions such as informational asymmetry and transportation and storage costs (Etula, 2013). In recent years, however, commodity derivatives have seen dramatic changes, including a strong increase in trading, consolidation of exchanges, and a shift from pit floor to electronic trading platforms (Irwin and Sanders 2012). Such structural changes have made the access to these markets easier to hedgers and speculators, thus leading to a substantial increase in volume and number of transactions.

Despite the advances, there are still significant differences among commodities futures exchanges around the world. On the one hand, stands the mature markets, where liquidity levels are high, there is a great variety of participants and good degree of transparency. On the other hand, emerging trading venues struggle with lower liquidity, what ultimately leads to higher bid-ask spread levels and increased volatility. Further, it may hinder price discovery and increase the probability of price manipulation, complicating hedging and trading activities (Lesmond, 2005).

The lower liquidity levels found in emerging commodities futures markets may find its roots on the underlying inter-dealer market structure. It is argued that increased concentration in the inter-dealer market may have negative consequences on the overall market quality. In addition to lower volume, it also contributes to increase bid-ask spreads and volatility (Branch & Freed, 1977; McInish & Wood, 1996, Van Ness et al., 2005).

There is no clear consensus about the impact of these transformations on the market microstructure of commodities futures (Frank and Garcia, 2011). In fact, the microstructure of commodities markets is not well known. This is particularly the case for emerging markets for which high frequency data are not always available. Brazil is possibly an exception to this rule, boasting good quality data for its large spot and futures markets. Moreover, despite the many studies looking at the organization of dealers' equity markets (Christie and Schultz 1994; Ellis, Michaely and O'Hara 2002; Aspris, Frino and Lepone 2012), there are very few about commodities markets.

The goals of this paper are twofold. Following Ellis, Michaely and O'Hara (2002), I first verify how competitive the inter-dealer market structure is in the live cattle markets at BM&FBovespa. Secondly, I aim to better understand the behavior and dynamics of the dealers in this market and their impact on liquidity costs (proxied by the bid-ask spread). I aim to answer whether there are many active dealers in the price discovery at the live cattle market and whether market concentration affects

dealers' profits and activity. Addressing these issues provides a way to understand not only some specific properties of dealers' markets, but also how these markets evolve over time.

I use tick-by-tick data from the BM&F Bovespa between March 2014 and February 2016 in order to examine the microstructures of the live cattle inter-dealer market. I focus on the live cattle contract since it is one of the most liquid commodities markets at the BM&F Bovespa, therefore providing a reliable picture of the mechanics of an emerging commodities market.

To the best of our knowledge, I am the first to investigate the structure of an inter-dealer market in the context of an emerging commodities exchange and to assess its impact on the bid-ask spread. Market structure is a pressing issue in emerging markets where liquidity levels are usually lower if compared to more mature commodities markets and the probability of existing dominant players is higher. More concentrated markets entail higher transaction costs, smaller participation and ultimately less efficiency.

In order to account for the determinants of dealers' activity in the live cattle market, I estimate by instrumental variables (IV) a probit model for whether the dealer is active or not. I subsequently estimate the impact of dealers' activity on quoted bid-ask spread (BAS) using a panel fixed-effects regression. I find that live cattle markets at BM&F Bovespa are not very competitive in the sense that a few dealers dominate the price discovery. In addition, I also document that the dealers' probability to be active increases with market concentration, quoted BAS, number of other active dealers, and the dealer's traded quantity. Lastly, I find that quoted BAS increases with market concentration and decreases with the number of active dealers.

Our findings go roughly in the same direction as those of Ellis et al. (2002). Similarly to what was found by the authors in the equities inter-dealer market at NASDAQ, the live cattle inter-dealer market at BM&F Bovespa does not operate on a competitive setting. However, the inequalities found in the latter were starker. While in the former market the top third market makers trade around 77% and 82.3% of all contracts, in the live cattle market the top third brokers trade around 95% of all contracts. In addition, I also found that the degree of concentration is a determinant of the bid-ask spread in the live cattle market, but it is more impactful in the latter market than it is in the equities market at NASDAQ.

The rest of this paper is as follows. Section 2 discusses the literature related to the inter-dealer market organization. Section 3 describes the institutional framework at the BM&F Bovespa futures market, whereas Section 4 explains the data details. Section 5 discusses the results of the IV-Probit model and fixed-effects regressions. Section 6 then performs a descriptive analysis of the live cattle inter-dealer market at BM&F Bovespa. Sections 7 and 8 report the main empirical findings as what regards respectively dealers' activity and the impact of dealers' activity in the quoted bid-ask spread. The final section provides a concluding summary.

4.2 Related Literature

Derivatives trading in Brazil occur either through exchange-traded futures or over the counter (OTC) markets. Like most exchanges, BM&F Bovespa now has electronic trading platforms that provide easier access to traders. Derivatives are traded in an inter-dealer system, where entry and exit are approximately free. The interaction of dealers' quotes with orders provides liquidity to traders and firms, lowers transactions costs, enhances price discovery and, ultimately, contributes to improve market efficiency. In principle, this price-setting structure of the dealer market closely approximates the competitive price formation process depicted by standard economic theory (Ellis, Michaely and O'Hara 2002; Dodd and Griffith-Jones 2007).

Concerns about the possibility of a noncompetitive structure in dealers-brokers markets motivated a number of studies. Christie and Schultz (1994) argue that the absence of odd-eighth quotes in nearly 70% of all stocks traded at NASDAQ could be an evidence of collusive behavior among dealers and market makers. The absence of odd-eighth quotes drove up bid-ask spreads and consequentially increased liquidity costs. Christie, Harris, and Schultz (1994) further detect a sharp drop on bid-ask spread after the release of Christie and Harris' paper, implying an abandonment of the collusive behavior by dealers and market

makers. In subsequent analysis, Barclay (1997) rule out the hypothesis that the higher bid-ask spread values observed at NASDAQ could be attributed to higher market making costs, thus contributing to the mounting evidence of collusive behavior. Using trade data for 30 stocks, Schultz (2000) confirm the results in Christie, Harris and Schultz (1994) and Barclay (1997), and raise the point that differences between dealers in aspects such as capitalization and industry specialization may hinder the achievement of a competitive equilibrium.

Evidence on the contrary comes from Wahal (1997), who analyzing quoted instead of effective bid-ask spreads, find that dealer pricing did not deviate from competitive equilibrium at NASDAQ National Market. However, the author could neither deny nor confirm the existence of collusive behavior among market makers. In this sense, Dutta and Madhavan (1997) argue that the free entry and exit system at NASDAQ markets make the existence of noncompetitive behavior almost impossible. Further, they conclude that spreads can be larger than competitive equilibrium values even if dealers do not cooperate. Klock and McCormick (1999) emphasize the role of market makers in reducing the bid-ask spread and suggested that due to the relatively high number of market makers in NASDAQ markets, the occurrence of collusive behavior is unlikely.

Despite its relative scarcity, due to data limitations, recent literature points that attaining a competitive behavior on dealers' markets is quite hard. Ellis, Michaely and O'Hara (2002) analyze the microstructure of dealers markets for a wide range of stocks traded at NASDAQ. They conclude that these markets are not competitive. Indeed, it is very often the case that a dominant dealer exists. Aspris, Frino and Lepone (2012) investigate the dynamic relationship between competition, liquidity provision, and market structure in the Australian options market (ASX). They conclude that free entry/exit of a market does not necessarily lead to a competitive market setting. They also argue that, to better explain why markets are not competitive, one must control for other factors, such as liquidity and degree of market concentration.

There has been no shortage of models in the literature providing a general idea of how dealers, brokers and market makers behave. Most of these models focus on the behavior of the market maker acting as a monopolist (Tinic 1972; Ho and Stoll 1981). By allowing for the existence of multiple market makers under competition, Ho and Stoll (1983) show that the impact of volatility on BAS increases with market concentration. Hagerty and McDonald (1996) find that monopolistic brokers charge traders the quoted BAS, whereas competitive brokers may charge traders a lower spread based on how informed they are. This implies lower quoted BAS as competition increases among brokers.

The behavior of market participants is central in understanding the microstructure of dealers-brokers markets. The common idea about the functioning of a dealers-brokers market structure is that it should approximately achieve a competitive equilibrium, since there is in most cases free entry and exit from the markets (Ellis, Michaely and O'Hara 2002; Aspris et al 2012).

A number of studies focused on analyzing the entry and exit of dealers, brokers and market makers. In general, their results indicate that there is a positive relationship between market entry and trading volume. They find mixed results regarding market entry and volatility, however. Wahal (1997) uncovers a negative association between them, reflecting the fact that high volatility pushes up the cost of carrying inventory, thus hindering market entry. Aspris, Frino and Lepone (2012) argue for a positive association between market entry and volatility because hedging and other risk management techniques are more profitable for dealers and market makers in an environment of high price variability. Although Ellis, Michaely and O'Hara (2002) find a negative association between market entry and concentration, they observe no significant relationship between the probability of entry/exit and the number of active market makers. Aspris, Frino and Lapone (2012) also verify a negative association between the number of active market makers and the probability of entry, meaning that markets with few dealers face higher probability of entry. Another relevant aspect of entry/exit broadly identified in the literature is the dealer's profitability. Ellis, Michaely and O'Hara (2002) unveil a weakly positive relation between dealer's profit and market entry, whereas Wahal (1997) and Aspris, Frino and Lepone (2012) find a positive association between market entry and spreads.

4.3. Institutional Details and Data

Like many futures and options exchanges, BM&F Bovespa operates with an open electronic limit order book and trades are made with the intermediation of a dealer or broker associated with BM&F Bovespa. Brokers only execute buy and sell orders coming from their clients. They do not carry inventories, and therefore, they cannot make profits from trading. On the other hand, dealers also execute buy and sell orders coming from their clients, but they can trade for their own portfolios, and make profits from trading. During the March 2014 – February 2016, brokers made up the majority of firms operating in the live cattle inter-dealer market. To be granted access to operate in the market, all dealers must meet the minimum financial and operational requirements established by BM&FBovespa. There are no market makers operating in the live cattle futures markets and all transactions are carried out either by brokers or dealers. This is the reason why I do not model the behavior of market makers as in other papers (Wahal, 1997; Ellis et al., 2002; Aspris et al. 2012). Therefore, unlike other more liquid and developed futures markets, there is formally no market participant providing liquidity and price references. This gives BM&FBovespa commodities futures markets a quite unique environment. There are 41 different brokers and dealers operating during the sample period.

The BM&FBovespa database organizes transactions data in three files. The first and second files contain information on buy offers (OFER_CPA) and sell offers (OFER_VDA), respectively. These files include date and time of the event to the millisecond, offer price (buy or sell), volume, order number, state of the offer (new, update, cancel, traded, etc), contract identifier, buy or sell sequence number, and broker identifier. The last file (NEG) records all transactions and, in addition to all above variables, contains the sequence numbers relative to the buy and sell orders that compose each transaction. These sequence numbers thus connect the three files.

I restrict attention to the live cattle contract traded at the BM&FBovespa. The live cattle contract trades from 9:00 to 16:00. In respect to contract specifications, contract size is of 4,407 net kilograms, with a tick size of R\$ 0.01 and expiration on the last trading day of the month. The sample period ranges from March 1st, 2014 to February 29th, 2016 (24 months).

4.4. Data description

Our database is composed of the following variables: *Event Dummy*, *Number of Dealers*, *Total Quantity*, *Total Profit*, *Concentration*, *Dealer's Traded Quantity*, *Dealer's Profit*, *Quoted Bid-ask Spread*, *Volatility*, *Time to Expiry* and a dummy variable to identify the first production cycle (*Season*).

In order to obtain the *Quoted Bid-ask Spread*, I reconstruct the order book and the top of the book. More specifically, I used all transactions from the buy and sell files, filtered them according to the order status and trading time, and obtained the lowest ask and highest bid. This is a distinct feature, since a significant number of studies use estimators based on transaction data to calculate the bid-ask spread.

The *Number of Dealers* is the daily average number of different active dealers; *Total Quantity* is the sum of the quantity traded by all dealers per day; *Total Profit* is the sum of the profits made by all dealers per day; *Concentration* is the Herfindahl index calculated per day; *Dealer's Quantity* is the quantity traded by each dealer per day; *Dealer's Profit* is the profit made by each dealer per day; *Quoted Bid-ask Spread* is the average of the bid-ask spread per day; *Volatility* is the difference between the highest and lowest price among traded contracts in the trading day; *Season*, or the first production cycle period, comprises the months from December to May, when the supply of live cattle is highest; *Time to Expiry* is a set of five dummy variables that account for the time to maturity day (last trading day of the month). In this case, I divided the trading month in maturity day, one to five days to maturity, six to ten days to maturity, eleven to fifteen days to maturity and more than sixteen days to maturity; *Event Dummy* is a dummy variable assigned the value of one if the dealer is active and zero otherwise.

Table 1 reports the descriptive statistics for the constructed series of variables over the period March 2014 – February 2016. The variables presented below are daily, continuous in time and correspond to the most traded contracts in each month.

Table 1

Variable	Mean	Std. Dev.	Min.	Max.
Quoted BAS	0.1723	0.0839	0.0529	1.0855
Number of Active Dealers*	19.8666	2.6747	14	25
Dealer's Traded Quantity	95.9148	291.9197	0	5524
Total Traded Quantity	3561.983	2557.451	42	17608
Volatility	2.0736	8.3275	0.0203	74.2996
Concentration	0.2459	0.0978	0.1133	0.8256

*per day

4.5. Characteristics of the live cattle inter-dealer market at the BM&FBovespa

I analyze dealers' participation from March 2014 to February 2016 period. For this purpose, I divide each month into five parts: maturity day (the last trading day of each month), one to five days to maturity day, six to ten days to maturity day, eleven to fifteen days to maturity and more the sixteen days to maturity day. Table 4 exhibits the average number of active dealers per day, the average quantity traded by dealers per day, and the average daily BAS.

Table 4

Average Number of Brokers	Maturity	1 to 5	6 to 10	11 to 15	16 and more	Whole Month
Mean	14.79	15.5	16.07	16.77	17.17	16.25
Median	17	15	16	17	17	16
Minimum	4	5	6	8	8	4
Maximum	24	28	28	25	27	28
Amplitude	20	23	22	17	19	24
Average Daily Trading Quantity						
Mean	3866	3478	3217	3627	3965	3562
Median	3504	2902	2899	3268	3326	3094
Minimum	42	66	202	314	440	42
Maximum	17610	11540	13870	10790	14120	17610
Amplitude	17568	11474	13668	10476	13680	17568
Average Daily Bid-Ask Spread						
Mean	0.18	0.16	0.17	0.18	0.18	0.17
Median	0.13	0.16	0.16	0.16	0.15	0.16
Minimum	0.05	0.07	0.07	0.09	0.09	0.05
Maximum	1.09	0.33	0.49	0.73	0.56	1.09
Amplitude	1.04	0.26	0.42	0.64	0.47	1.04

There is substantial variation in the number of active dealers across the maturity month, ranging from a minimum of four active dealers on maturity days to a maximum of 28 active dealers within one to ten days to maturity. The average number of active dealers is roughly stable across the month, presenting a smooth downward trend as maturity day approaches. Ellis, Michaely and O'Hara (2002) also found the average number of dealers to be relatively stable along the time when analyzing the NASDAQ stock markets. Regarding the average traded quantity, there's an upward trend starting from the tenth day to maturity up to maturity day. On maturity day is also when the amplitude is highest, that is, when the smallest and biggest number of

contracts are traded. This result is expected since it is usually on the last trading day when hedgers and speculators either end or roll their contract positions.

The average daily BAS analysis shows that liquidity costs tend to take the opposite direction from traded volume. From the tenth day to maturity, the median BAS exhibits a downward trend until maturity day, when median BAS is lowest. It is also on maturity day when the BAS amplitude is highest. The minimum BAS value recorded for maturity days is R\$ 0.05 and the maximum is R\$ 1.04. I do not find such an amplitude in any other period along the month. The negative relation between traded quantity/volume and BAS in commodity markets is well established and broadly found in the literature (Wang and Yau 2000; Wang, Garcia and Irwin 2013). Therefore, along closest to maturity months I observe that traded quantity tends to increase and BAS tends to decrease as maturity day approaches. In addition, the average number of active dealers remains roughly stable across the month, with a tendency to decrease as maturity day draws near.

Table 5

Quantity Traded, by Broker	Maturity	1 to 5	6 to 10	11 to 15	16 and more	Whole Month
% Quantity by largest Dealer	26.06	26.58	29.29	28.20	30.69	28.46
% Quantity by smallest Dealer	0	0	3.89E-050	0	0	8.88E-06
% Quantity by top third	89.17	95.38	94.85	94.53	95.67	94.76
% Quantity by middle third	8.74	3.61	4.53	4.88	3.40	4.39
% Quantity by bottom third	2.08	0.99	0.61	0.57	0.91	0.84

Although traded quantity by dealers follows an expected upward trend as maturity day gets close, Table 5 reveals strong inequalities in traded quantity across dealers. For this purpose, I calculate the daily trading quantity by each dealer. I also divide the total number of dealers in a stock into thirds, and I measure the fraction of trading quantity carried out by the top third, middle third, and bottom third each day.

I find that most of the trades is concentrated in the top subset of dealers, which accounts, on average, for 95% of all traded quantity. The bottom two thirds of dealers never respond for more than 5% of overall traded quantity. There is little change on these shares across the month, however a small downward trend is verified on maturity day, meaning that the smaller dealers tend to be more active on maturity day. Therefore, the live cattle market at BM&FBovespa is not competitive in the sense of many dealers active in the price discovery process, being the case that the top three dealers dominate the biggest part of traded contracts.

In order to better analyze how concentrated the live cattle market at BM&F Bovespa is, I calculate the Herfindahl index for every day of the analysis. If the live cattle market is a monopoly then the Herfindahl index is equal to one, while a perfectly competitive market in which dealers equally shared volume would yield an index of $1/n$ (represented by the red line in Figure 1), being n the number of dealers (Ellis, Michaely and O'Hara 2002). Figure 1 shows the Herfindahl index for the whole period March 2014-February 2016 period.

Figure 1

The values given by the index show that the live cattle inter-dealer market at BM&FBovespa is not monopolistic, however it is far from being competitive. Further analysis points to the fact that the concentration does not exhibit strong variations along the month. Similar results were found by Ellis, Michaely and O'Hara (2002) in NASDAQs stock markets. The number of dealers in a market may have little economic relevance for market behavior. A more meaningful statistic is the number of active dealers and understanding the way they are organized and how they behave is crucial to have a deep and comprehensive understanding of the microstructure of live cattle markets.

An interesting feature of this market is that brokers have very different frequency of trading. The top four brokers in terms of quantity traded were active in almost every single day of the period of analysis. The frequency of trading decreases as quantity traded decreases as depicted in Figure 2. The bottom third of brokers was active in 12.5% of days of the March 2014-February 2016 period, whereas the middle third and the top third brokers were active in 46.9% and 82.3% of days in the period of analysis on average, respectively. Therefore, dominant brokers are those that trade the biggest shares of contracts in the market and at the same time the ones that trade most frequently.

5.

Figure 2

4.6 Modeling dealers' activity

I model the dealer's probability of being active within intervals of 5, 10 and 20 days. As for controls, I include the *Number of Dealers*, *Total Quantity*, *Total Profit*, *Concentration*, *Dealer's Traded Quantity*, *Dealer's Profit*, *Quoted Bid-ask Spread*, *Volatility*, and a dummy variable to identify the first production cycle season (*Season*) as regressors.

Regarding expected signs of the coefficients in equation (1), following the discussion in the literature section, it is expected that both *Total Quantity* and the *Dealer's Traded Quantity* have a positive coefficient, meaning that brokers and dealers tend to be active when trading activity increases. A positive coefficient is also expected for *Quoted Bid-ask Spread* (Wahal, 1997; Aspris et al. 2012) and for the profit related variables (Ellis et al. 2002). The *Concentration* coefficient is expected to have a negative sign, as more competitive markets bring about lower levels of quoted BAS (Ellis et al. 2002). A negative coefficient is also expected for the *Number of Dealers* (Aspris et al. 2012). In what concerns the *Volatility* coefficient, either a positive or a negative sign is expected. A positive sign is justified by Aspris et al. (2012) as a demand for hedging and other risk mitigation techniques. On the other hand, Wahal (1997) argues that a negative coefficient would reflect the costs of carrying inventories in a high volatility environment.

The IV-Probit regression is depicted in equation (1):

$$(1) \Pr(\text{Dealer is active for contract } j \text{ in period } t) \\ = \beta_0 + \beta_1 \text{Number of Dealers}_{j,t-1} + \beta_2 \text{Total Quantity}_{j,t-1} + \beta_3 \text{Total Profit}_{j,t-1} \\ + \beta_4 \text{Concentration}_{j,t-1} + \beta_5 \text{Dealer's Quantity}_{j,t-1} + \beta_6 \text{Dealer's Profit}_{j,t-1} \\ + \beta_7 \text{Quoted Bid/Ask Spread}_{j,t-1} + \beta_8 \text{Volatility}_{j,t-1} + \beta_9 \text{Season} + \beta_{dealer}$$

I compute the Herfindahl index using the quantity traded by each broker in the market for every trading day:

$$(4) HHI_i = \sum_{i=1}^n MS_i^2 = \sum_{i=1}^n (x_i / \sum_{j=1}^n x_j)^2$$

where MS_i denotes the market share of dealer i and x_i is the participation of each of the n dealer.

As for the dealer's profit, I proceed as in Ellis, Michaely and O'Hara (2002):

$$(5) Profit_t = \begin{cases} N_S(t)(P_S - P_B) & \text{if } N_B > N_S \\ N_B(t)(P_S - P_B) & \text{if } N_B < N_S \end{cases}$$

where N_S is the number of contracts sold by each dealer in any given trading day; N_B is the number of contracts bought by each dealer in any given trading day; P_S is the selling price (ask price), P_B is the buying price (bid price), and t denotes trading day. When the broker buys more contracts than sells ($N_B > N_S$), then his profit comes from reselling these contracts (N_S). I compute daily buy and sell prices (P_B and P_S , respectively) as weighted averages of the transactions prices:

$$(6) P = \begin{cases} P_B = \frac{\sum_{j=1}^{B_t} N_B(j) \cdot P(j)}{N_B(t)} \\ P_S = \frac{\sum_{j=1}^{S_t} N_S(j) \cdot P(j)}{N_S(t)} \end{cases}$$

In order to model the dealer's probability of being active, I use a probit regression with fixed effects. A number of studies considered the possibility of endogeneity in market microstructure models, most notably among variables such as bid-ask spread, volume, volatility (Wang and Yau 2000; Martinez et al. 2011; Wang, Garcia and Irwin 2013), and number of dealers (Grossman and Miller 1988; Diamond and Verrechia 1991). Accordingly, I perform the Wald test for endogeneity for each one of the independent variables in equations (1).

Table 2 reports the test results. The null hypothesis is that the variable under consideration can be treated as exogenous. In this sense, the number of dealers, concentration, volatility, and total traded quantity were found to be endogenous in the 10-day and 20-day window regressions. The dealer's traded quantity was also considered an endogenous variable in the latter regression. The 5-day window regression exhibited fewer endogenous regressions than the other two. In that case, the number of dealers, concentration, and total traded quantity can be regarded as endogenous, the last two at the 10% significance level.

Table 2

Endogeneity Test	5-day	10-day	20-day
Number of Dealers	0.0004***	0.0000***	0.0000***
Volatility	0.6709	0.0000***	0.0000***
Concentration	0.0559*	0.0002***	0.0000***
Quoted BAS	0.3782	0.33	0.5422
Dealer's Profit	0.4923	0.8138	0.1425
Total Profit	0.7561	0.6461	0.7249
Dealer's Traded Quantity	0.2347	0.4292	0.0009***
Total Traded Quantity	0.0502*	0.0000***	0.0000***

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

For the 5-day window, I use the past values of total traded quantity, dealer's traded quantity, volatility and concentration as well as the first difference of the number of dealers as instruments. For the 10-day and 20-day windows, the list of instruments reduces to the first difference of the number of dealers as well as past values of total volume, concentration and volatility.

In equation (2) I apply the augmented Durbin-Wu-Hausman test for endogeneity bias among volatility, concentration and total traded quantity. As endogeneity was detected, I used lagged volatility as an instrument for volatility, lagged concentration as an instrument for concentration and lagged total traded quantity as an instrument for total traded quantity. Table 3 documents the findings of the Durbin-Wu-Hausman tests. Our results show that volatility, number of dealers, concentration, and total traded quantity are endogenous in equation (2), considering the 5-day and 10-day windows. Regarding equation (2) using the 20-day window, only concentration cannot be considered endogenous.

Table 3

Endogeneity Test	5-day	10-day	20-day
Number of Dealers			
Volatility	0.0000***	0.0000***	0.0000***
Concentration	0.0000***	0.0000***	0.583
Total Traded Quantity	0.0000***	0.0000***	0.0431**

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

Table 6 reports the results for the IV-Probit regression. In general, the coefficient estimates are consistent with expectations and statistically significant across the 5-days, 10-days and 20-days windows of analysis.

Table 6

Dependent Variable: Active Broker	5-day	10-day	20-day
--	--------------	---------------	---------------

Number of Dealers	0.2182*** (0.0393)	0.4447*** (0.0410)	0.7276*** (0.0525)
Volatility	0.1650** (0.0741)	0.1579* (0.0881)	0.1955* (0.1079)
Concentration	6.5604*** (1.6155)	17.7842*** (1.9215)	34.7301*** (2.7846)
Total Traded Quantity	-0.0001*** (0.0000)	-0.0003*** (0.0000)	-0.0006*** (0.0000)
Dealer's Traded Quantity	0.0045*** (0.0005)	0.0051*** (0.0005)	0.0053*** (0.0005)
Season	-0.0426 (0.0411)	0.0233 (0.0367)	-0.1038*** (0.0322)
Quoted BAS	0.2306 (0.4705)	1.9295*** (0.5272)	3.1964*** (0.6428)
Total Profit	-0.0001*** (0.0000)	-0.0005*** (0.0000)	-0.0010*** (0.0000)
Dealer's Profit	0.0005 (0.0009)	-0.0006 (0.0011)	0.0020 (0.0015)

Respectively, *, **, and *** indicate statistically significance at 10, 5, and 1% levels.

Market concentration, quoted bid-ask spread and volatility entail the largest effects on dealers' activity on any given day. Market concentration has the largest impact, guarding a positive relationship with dealers' activity. This means that dealers are more active when market concentration is high. Similarly, the quoted bid-ask spread exhibits a positive association with the dealer's probability of being active. These results reflect the situation described in the previous section. The live cattle inter-dealer market at BM&F Bovespa is dominated by a few brokers and their presence is ultimately associated with higher bid-ask spread and concentration levels. This result is in line with Wahal's (1997) and Aspris' et al. (2012) findings regarding equities and options markets.

There is also a positive relationship between daily volatility and the probability of being active, mirroring the increase in hedging demand in times of higher uncertainty. This positive association is consistent with Aspris et al.'s (2012) evidence for the Australian Options market.

The number of active dealers in the market has a moderate but significant and positive impact on the broker's decision to be active. That is, dealers and brokers tend to be active when other dealers or brokers are also active. Aspris, Frino and Lepone (2012) find the number of market makers to matter in the decision to entry the market, however within a negative association.

Regarding traded quantity, the probability of being active in the market increases with the dealer's individual traded quantity and decreases with the overall market traded quantity (the sum of the quantity traded by all active dealers) in any given

day. I also find dealers' individual profit to have no significant relationship with the probability of being active in the market. However, I observe that the overall profitability (i.e., the sum of all profits made by all active dealers) negatively impacts the probability of being active in the market.

Following Rosen and Sheinkman (1993) study on live cattle market cycles, I create a variable to capture the effect of seasonality in our data. For this purpose, I divide the period of analysis in two. The first period corresponds to the *Season* or live cattle's first cycle, ranging from December to May. This is when the live cattle supply is highest and the most traded contract is the one maturing in May. The second cycle, also known as *off-season*, goes from June up to November. This is the period when most producers opt to intensify the production and the supply is relatively short. The most traded contract in this period is the one maturing in October. I found a negative and significant coefficient on the 20-day window regression, meaning that dealers tend to be more active in the *off-season* period if compared to the *season* period.

In conclusion, the live cattle inter-dealer market at BM&F Bovespa is largely asymmetrical. Trading activity, which is dominated by a handful of dealers and brokers, tends to be associated with high levels of market concentration and quoted BAS. This relationship is further investigated in the next section. Dealers also tend to be active as volatility increases, which is intuitive since futures markets provide hedging and risk management techniques to cope with increased volatility. Dealers tend to be active when other dealers are active, and they tend to be more active during the *off-season* period. Further, the probability to be active increases with its individual traded quantity. On the other hand, dealers' probability to be active decreases as overall traded quantity and profitability rises.

4.7. Dealers' Dynamics, Competition and Bid-Ask Spread

To examine the association between dealers and brokers' activity and the impact on quoted bid-ask spreads (BAS), I use a panel data model with fixed effects and clustered standard errors. Similarly, I construct 5-day, 10-day and 20-day event windows.

$$(2) \text{BAS}_{i,t} = \beta_0 + \beta_1 \text{Concentration}_{i,t-1} + \beta_2 \text{Number of Dealers}_{i,t-1} + \beta_3 \text{Volatility}_{i,t-1} + \beta_4 \text{Total Quantity}_{i,t-1} + \beta_5 \text{Time to Expiry}_{i,t} + \beta_6 \text{Event Dummy}_{i,t} + \beta_7 \text{Season}_{i,t} + \beta_i$$

Regarding the expected coefficients in equation (2), the *Concentration* coefficient is expected to have a positive sign, as less competitive markets bring about greater levels of BAS (Branch & Freed, 1977; Cohen & Conroy, 1990; Hamilton, 1979; McInish & Wood, 1996; Ellis, Michaely and O'Hara, 2002; Van Ness et al., 2005; Aspris, Frino and Lepone, 2012). Since more competition contributes to drive down quoted BAS levels, then the *Number of Dealers* coefficient tends to be negative. A positive sign is expected for the *Volatility* coefficient. The positive relationship between quoted BAS and volatility is well documented in the literature (Wang and Yau 2000; Wang, Garcia and Irwin, 2013). *Total Quantity* is expected to have a negative sign, since the greater the volume, the smaller quoted BAS levels tend to be (Ellis et al., 2002, Wang, Garcia and Irwin, 2013). *Time to Expiry* and *Event Dummy* are expected to have positive and negative signs, respectively, meaning that quoted BAS tends to decrease as maturity day approaches and when dealers are active in the market (Aspris et al., 2012).

Table 7 reports the results for the fixed effects regression. As expected, a positive relationship was found between quoted BAS and concentration, and between quoted BAS and volatility. These two variables are the main determinants of the quoted BAS in our analysis, since they exhibit the biggest coefficients among all. In addition, a negative association was found between quoted BAS and the number of dealers and between quoted BAS and overall traded quantity. Therefore, these results are consistent with previous studies (Ellis, Michaely and O'Hara 2002; Aspris, Frino and Lepone 2012).

Table 7

Dependent Variable: BAS	5-day	10-day	20-day
Concentration	0.3552***	0.5873***	0.8782***

	(0.0170)	(-0.0195)	(0.0231)
Volatility	0.1078***	0.1167***	0.1023***
	(0.0025)	(0.0027)	(0.0033)
Total Traded Quantity	-0.00001***	-0.0000***	-0.0000***
	(0.0000)	(0.0000)	(0.0000)
Number of Dealers ²	-0.0044***	-0.0202***	-0.0472***
	(0.0007)	(0.0012)	(0.0023)
Active Dealer	-0.0090***	-0.0075***	-0.0072***
	(0.0016)	(0.0016)	(0.0016)
Maturity	-0.0035**	-0.0061	-0.0028
	(0.0031)	(0.0030)	(0.0032)
1 to 5 days to Maturity	-0.019***	-0.0232***	-0.0249***
	(0.0018)	(0.0018)	(0.0018)
6 to 10 days to Maturity	-0.0095***	-0.0104***	-0.0076***
	(0.0018)	(0.0018)	(0.0018)
11 to 15 days to Maturity	-0.0020	-0.0007	0.0026
	(0.0018)	(0.0017)	(0.0018)
Season	-0.0031***	-0.0099**	-0.0174***
	(0.0012)	(0.0013)	(0.0014)

Respectively, *, **, and *** indicate statistical significance at 10, 5, and 1% levels.

² first difference of the number of dealers

Our results also point to the fact that quoted BAS tend to be smaller as maturity day approaches. The negative coefficient of the *Event Dummy* variable means that dealer's activity is on average associated with a significant decline in quoted bid-ask spreads. This result is robust for 5-day, 10-day and 20-day event windows. Our results are in line with Hagerty and McDonald (1996), who argued that BAS tend to be lower when brokers are under competition than in situations when brokers are monopolists. Lower BAS can be translated in easier trading since liquidity costs are lower.

Therefore, traders in the live cattle market at BM&F Bovespa would benefit when either more dealers or brokers enter the market or when the existing ones increase their participation in the total traded volume. Thus, a more competitive inter-dealer market structure contributes to significantly lower liquidity costs levels and improve overall market quality.

Another point of discussion is the impact of foreign investors on the microstructure of emerging futures markets. Lee and Chung (2016) argues that the presence of foreign investors improves general market quality as they contribute to lower BAS levels and increase competition among traders. However, commodities futures markets well connected to foreign investors are quite rare in the context of emerging markets. This is precisely the case of the live cattle market studied in this paper. Unlike the corn or

coffee futures markets at BM&F Bovespa, which have a significant part of their contracts traded by non-resident traders, most of the trading activity observed in the live cattle market is carried out by resident investors. Therefore, the quality observed in the live cattle futures markets is mostly defined by domestic factors and as I demonstrate in this paper, it could be enhanced with a more competitive inter-dealer system.

4.8. Conclusion

This paper sought to analyze the organization of the live cattle inter-dealer market at BM&F Bovespa on the microstructure level. More specifically, I focused on the organization and behavior of dealers and brokers as well as the impact of this structure on the quoted bid-ask spread (BAS), a proxy used for liquidity costs.

I find that despite the relatively high number of dealers and brokers operating in the live cattle market at BM&F Bovespa, the role played by some of these dealers can be different from the role played by others. On any given day of our period of analysis, a single dealer or broker typically dominates trading, executing approximately 28% of the traded quantity. If I consider the top three participants, this share goes up to roughly 95% of all traded contracts. These results clearly show the asymmetric nature of the relationship among dealers in the live cattle market.

An in-depth analysis of the degree of competitiveness concluded that the live cattle inter-dealer market deviates from a competitive setting. In this sense, our findings go in line with what has been found for other inter-dealer markets. Moreover, no significant difference was found between commodities and equities and stock markets (Ellis et al. 2002; Aspris et al. 2012). As a matter of fact, the sheer number of dealers in a market may have little economic relevance for market behavior. A more meaningful statistic is the number of active dealers and how they interact with one another and respond to market signals such as quoted BAS and market concentration.

The probability of finding active dealers in the live cattle market increases with market concentration and quoted BAS. This fact mirrors the asymmetric nature of the inter-dealer market, where a few dealers or brokers dominate the trading activity and are active in almost every single day of the March 2014-February 2016 period. There is also a positive association between the probability of being active on any given day and volatility, the number of other active dealers, and the dealer's traded quantity.

Results pertaining the liquidity costs analysis show that dealers' activity contribute to reduce quoted BAS as well as the number of active dealers and total traded quantity. On the other hand, market concentration and volatility contribute to increase quoted BAS. Therefore, a more competitive inter-dealer structure would lower liquidity costs and enhance the quality of the live cattle market at BM&F Bovespa.

In general, our findings suggest that in emerging commodities exchanges, where liquidity levels are usually low, one possible determinant of the high bid-ask spread levels observed in these markets is the structure of the underlying inter-dealer market. In this sense, increasing the participation of dealers and brokers in trading

The implication of our findings is pertinent both to commodities exchanges in emerging markets, and to market regulators since inter-dealer competition contributes to enhancing liquidity and overall market quality.

The impact of foreign investment in Bovespa with respect to concentration and the linkages with world market prices is a venue for future research.

References

- Aspris, A.; Frino, A.; Lapone, A. (2013). The impact of market maker competition on market quality: evidence from an options exchange. *Australasian Accounting, Business and Finance Journal*, Vol. 6, n. 5, p. 24-36.
<http://ro.uow.edu.au/cgi/viewcontent.cgi?article=1396&context=aabfj>
- Branch, B., & Freed, W. (1977). Bid-ask spreads on the AMEX and the big board. *The Journal of Finance*, 32(1), 159–163.
<https://doi.org/10.1111/j.1540-6261.1977.tb03249.x>

- Barclay, M.J. (1997). Bid-ask spreads and the avoidance of odd-eight quotes on Nasdaq: an examination of exchange listings. *Journal of Financial Economics*. Vol. 45, p.35-60. [https://doi.org/10.1016/S0304-405X\(97\)00010-X](https://doi.org/10.1016/S0304-405X(97)00010-X)
- Christie, W. Schultz, P. (1994). Why do Nasdaq market makers avoid odd-eighth quotes? *The Journal of Finance*. Vol. 49, n.5, p. 1813-1840. <https://doi.org/10.1111/j.1540-6261.1994.tb04782.x>
- Christie, W.; Harris, J.; Schultz, P. (1994). Why did Nasdaq market makers stop avoiding odd-eighth quotes? *The Journal of Finance*. Vol. 49, n.5, p. 1841-1860. <https://doi.org/10.2307/2329273>
- Cohen, K. J., & Conroy, R. M. (1990). An empirical study of the effect of rule 19c-3. *The Journal of Law and Economics*, 33(1), 277–305. <https://doi.org/10.1086/467206>
- Diamond, D.; Verrechia, R. (1991). Disclosure, liquidity and the cost of capital. *The Journal of Finance*. Vol. 46, n.4, p. 1325-1359. <https://doi.org/10.1111/j.1540-6261.1991.tb04620.x>
- Dodd, R.; Griffith-Jones, S. (2007). Brazil's derivative markets: hedging, central bank intervention and regulation. CEPAL: series informes y estudios especiales n. 20. <http://repositorio.cepal.org/handle/11362/8868>
- Dutta, P.K.; Madhavan, A. (1997). Competition and collusion in dealer markets. *Journal of Finance*. Vol. 52, p. 245-276. <https://doi.org/10.2307/2329563>
- Ellis, K.; Michaely, R.; O'Hara, M. (2002). The making of a dealer market: from entry to equilibrium in the trading of Nasdaq stocks. *The Journal of Finance*. Vol. LVII, n.5, p. 2289-2316. <https://doi.org/10.1111/1540-6261.00496>
- Etula, E. (2013). Broker-dealer risk appetite and commodity returns. *Journal of Financial Econometrics*, Vol. 11, n.3, p. 486-521. <https://doi.org/10.1093/jfinec/nbs024>
- Frank, J.; Garcia, P. (2011). Bid-ask spreads, volume and volatility: evidence from live stock markets. *American Journal of Agricultural Economics*. Vol. 93, n. 1, p. 209-225. <http://www.jstor.org/stable/41240269>.
- Grossman, S.; Miller, M. (1988). Liquidity and market structure. *The Journal of Finance*. Vol. 43, n.3, p. 617-633. <https://doi.org/10.1111/j.1540-6261.1988.tb04594.x>
- Hagerty, K.; McDonald, R.L. (1996). Brokerage, market fragmentation, and securities market regulation. In: *The Industrial Organization and Regulation of the Securities Industry*. The University of Chicago Press.
- Hamilton, J. L. (1979). Marketplace fragmentation, competition, and the efficiency of the stock exchange. *The Journal of Finance*, 34(1), 171. <https://doi.org/10.2307/2327151>
- Irwin, S. H.; Sanders, D. R. (2012). Financialization and structural change in commodities futures markets. *Journal of Agricultural and Applied Economics*. Vol.44, n.3, p.371-396. <https://doi.org/10.1017/S1074070800000481>
- Klock, M.; McCormick, D.T. (1999). The impact of market maker competition on Nasdaq spreads. *The Financial Review*. Vol. 34, p. 55-74. <https://doi.org/10.1111/j.1540-6288.1999.tb00469.x>
- Lee, J.; Chung, K. (2016). Foreign Ownership, Legal System, and Stock Market Liquidity. Available at SSRN: <https://ssrn.com/abstract=2563154> or <http://dx.doi.org/10.2139/ssrn.2563154>
- Lesmond, D. (2005). Liquidity of emerging markets. *Journal of Financial Economics*, 77(2), 411–452. <https://doi.org/10.1016/j.jfineco.2004.01.005>
- McInish, T. H., & Wood, R. A. (1996). Competition, fragmentation, and market quality. In *The industrial organization and regulation of the securities industry* (pp. 63–92). University of Chicago Press.
- Schultz, P. (2000). Regulatory and legal pressures and the cost of Nasdaq trading. *The Review of Financial Studies*. Vol. 13, n. 4, p. 917-957. <https://doi.org/10.1093/rfs/13.4.917>
- Van Ness, B. F., Van Ness, R. A., & Warr, R. S. (2005). The impact of market maker concentration on adverse-selection costs for NASDAQ stocks. *Journal of Financial Research*, 28(3), 461–485. <https://doi.org/10.1111/j.1475-6803.2005.00134.x>
- Wahal, S. (1997). Entry, exit, market makers, and the bid-ask spread. *The Review of Financial Studies*. Vol. 10, n.3, p. 871-901. <https://doi.org/10.1093/rfs/10.3.871>

Wang, X.; Garcia, P.; Irwin, S.H. (2013). The behavior of bid-ask spreads in the electronically traded corn futures markets. *American Journal of Agricultural Economics*. Vol. 96, n.2, pp. 557-577.
<http://ajae.oxfordjournals.org/cgi/doi/10.1093/ajae/aat096>

5. CONCLUSION

The structural transformations undergone by commodities futures markets during the first decade of the 2000s have left indelible impacts on the volume, open interest and on the access to these markets. These transformations occurred also in emerging markets, being the BM&F-Bovespa no exception. Nonetheless, emerging markets have intrinsic characteristics such as lower levels of volume and transactions if compared to more developed markets, higher degree of market concentration and higher levels of liquidity costs.

Benefiting from the use of high frequency data made possible by such transformations, this dissertation is composed of three papers that sought to understand certain aspects of the microstructure of commodities futures markets that were not evident through the lenses of higher frequency data, such as daily, weekly or monthly data.

In the first paper (chapter 2) I modelled and forecasted realized volatility for the corn and live cattle markets at BM&F-Bovespa. For this purpose, I used the heterogenous autoregressive model developed by Corsi (2009) and its extensions adapted to include jump and leverage components. The modelling period went from March 2014 to March 2016 and the forecasting period went from April 2016 to March 2016. Results show that realized volatility both in the corn and live cattle markets can be understood as a combination of long-memory processes and short-term processes, being the latter represented by jumps and leverage components. This result contributes to shed new light on the issue regarding price volatility in Brazilian commodities futures markets, since most of the work previously done in this field used models that treated volatility as a latent variable, and such aspects could not be clearly verified.

The second paper (chapter 3) analyzed the relationship between volatility, volume, and liquidity costs, proxied by the bid-ask spread, in the context of two emerging commodity markets – corn and live cattle – during the March 2014-February 2016 period. Having in mind that emerging market usually face problems related to market thinness, I decided to control for liquidity. Another frequent issue in emerging markets relates to its concentration, what led me to control also so market concentration. For the estimation and considering potential endogeneity problems, I used a three-equation structural model and the IV-GMM approach. My findings point that liquidity costs are higher for the less liquid live cattle market than for the more liquid corn market, which was expected. I also found that bid-ask spread is negatively related to volume and positively related to volatility, as well defined in the literature. However, the significance and magnitude of the responses depend on the level of liquidity in each market. The more liquid the market, the clearer and the better defined are the responses. Besides the impact of concentration on each market is different. While in the live cattle market an increase in concentration contributes to decrease volume, in the corn market it contributes to decrease both the levels of bid-ask spread and volume.

Finally, in the third paper (chapter 4), I investigated the live cattle inter-dealers market during the March 2014-February 2016 period. In this paper I was particularly interested in the behavior of the dealers in the commodities markets and how it could potentially affect market variables, such as the bid-ask spread. I firstly carry out a descriptive analysis of the dealers market and reach the conclusion that they do not operate on a competitive structure. Then I model the behavior of dealers' activity using an instrumental variable probit model, giving value 1 when the dealer is active on any given day and 0 otherwise. Results show that dealer's activity is positively related to market concentration, quoted bid-ask spread, number of active dealers and the dealer traded quantity. Another result from the third paper is the impact of the concentration and number of dealers on the bid-ask spread observed in the live cattle market. I found that these markets would benefit from lower bid-ask spread levels, hence, lower liquidity costs, if a more competitive structure is achieved in the live cattle inter-dealers market.

The relatively recent availability of high frequency data gives access to the once not so well understood

mechanisms of the commodities futures markets. This dissertation unveils a tiny fraction of these mechanisms by shedding light on a few aspects of the microstructure of these markets. More research is needed in this field to better understand how commodities futures markets work in its details and in the smallest fractions of time. The ultimate result will be gains in efficiency and better functioning market.