INSTITUTE
OF ECONOMICS

Scuola Superiore
Sant'Anna

LEM | Laboratory of Economics and Management

Institute of Economics
Scuola Superiore Sant'Anna

Piazza Martiri della Libertà, 33 - 56127 Pisa, Italy
ph. +39 050 88.33.43
institute.economics@sssup.it

LEM
WORKING PAPER SERIES

Non-Normal Identification for Price Discovery in High-Frequency Financial Markets

Sebastiano Michele Zema [a]

[a] Institute of Economics, Scuola Superiore Sant'Anna, Pisa, Italy.

2020/28 October 2020
ISSN(ONLINE) 2284-0400

Non-Normal Identification for Price Discovery in High-Frequency Financial Markets

Sebastiano Michele Zema*
Institute of Economics, Scuola Superiore Sant'Anna
Piazza Martiri della Libertà 56127, Pisa
sebastianomichele.zema@santannapisa.it

October 23, 2020

Abstract

The possibility to measure the relative contribution of agents and exchanges to the price formation process in high-frequency financial markets acquired increasingly importance in the financial econometric literature. In this paper I propose to adopt fully data-driven approaches to identify structural vector error correction models (SVECM) typically used for price discovery. Exploiting the non-Normal distributions of the variables under consideration, I propose two novel variants of the widespread Information Share (IS) measure which are able to identify the leaders and the followers in the price formation process. The approaches will be illustrated both from a semi-parametric and parametric standpoints, solving the identification problem with no need of increasing the computational complexity which usually arises when working at incredibly short time scales. Finally, an empirical application on IBM intraday data will be provided.

Keywords: Information Shares; Structural VECM; Microstructure noise; Independent Component Analysis; Directed acyclic graphs.

JEL classification: C32, C58, G14.

*I am grateful to Alessio Moneta, Giacomo Bormetti, Fulvio Corsi, and Mattia Guerini for their precious suggestions. I gratefully acknowledge Joel Hasbrouck for sharing his data and useful comments. All errors are my own.

1 Introduction

The past decades have been characterized by dramatic changes in financial markets, where the proliferation of algorithmic trading strategies put aside the intervention of human agents in the price formation process. These algorithms execute orders at incredibly short time scales and there is no doubt anymore they account for most of the trading volumes in developed markets. In addition, processes of market fragmentation have been carried out jointly with the rising of high-frequency trading. This doubly increased the complexity of financial markets, since quotes and trades might be dispersed across different listing venues and at heterogeneous time scales which mix the slower dynamic of humans with the faster dynamic of machines. The possible benefits of fragmented versus consolidated markets have been object of debates for both economists and regulators also in recent times (O'Hara and Ye, 2011; Kwan et al., 2015; Hatheway et al., 2017). As a consequence, the possibility to measure the relative contribution of each exchange in which the asset is listed, to the price formation process, acquired increasing importance in the research environment. Going back in the literature two popular and still widely used measures have been introduced, which are the information share (IS) of Hasbrouck (1995) and the component share (CS) directly based on the permanent-transitory (PT) decomposition introduced in Gonzalo and Granger (1995) (Harris et al., 1995; Booth et al., 1999; Hansen and Lunde, 2006). These two measures build their fundamentals upon the modeling of price changes through VECMs, with the substantial difference that while the CS is defined only in terms of speeds of adjustment toward the common trend (i.e. markets with lower cointegration loadings rapidly adjust and are thus more informative), the IS measure is more concerned with variations in the prices and seeks to measure the amount of information generated by each market. Both approaches have their merits and limits which have been documented by comprehensive discussions in the literature (Baillie et al., 2002; De Jong, 2002; Harris et al., 2002a,b; Hasbrouck, 2002b; Lehmann, 2002). The IS approach, compared to the CS one, has a richer specification since it considers the speed of adjustment together with the relative share of variance of the efficient price process accounted by each market. Still, from a microstructural modeling point of view, the IS can be uniquely determined only when the VECM residuals are uncorrelated given that the presence of substantial contemporaneous correlations hampers the correct identification of the structural shocks occurred in each market. Hasbrouck's suggested solution was, in absence of an established theory providing the causal chain to correctly order the variables in the model, to identify the SVECM using the Choleski decomposition and going through all the possible permutations of the variables so to get upper and lower bounds for the IS of each market. In empirical applications upper and lower bounds are often very wide giving rise to interpretative ambiguities about the real allocation of information between the analyzed variables, making impossible to distinguish between the exchanges which lead the price formation process and exchanges that follow it. From a recent data-driven perspective instead, Hasbrouck (2019) proposed to exploit the high frequency at which quotes and trades occur, modeling thus in natural time to drastically reduce the upper and lower bounds obtained by permuting the variables. Sampling prices at very short time scales, even from microseconds to nanoseconds precision, heavily reduces contemporaneous cross correlations between the listing venues indeed, which by construction leads to narrower IS bounds and allow to discard any interpretive ambiguity. To deal with the enormous amount of coefficients to be estimated in such a natural time framework, the author handled the problem by adopting the heterogeneous autoregressive approach (HAR) proposed by Corsi (2009). Nevertheless, Hasbrouck's modeling approach in natural time raised interesting and useful comments and discussions in the literature, in some cases controversial, directly related to the econometric model specification, treatment of the high level of data sparsity in natural time, and subsequent identification of where price discovery occurs (Brugler and Comerton-Forde, 2019; Buccheri et al., 2019; De Jong, 2019; Ghysels, 2020). Despite the identification issue above mentioned and even if other measures of price discovery have been proposed in the literature (see Lien and Shrestha, 2009; De Jong and Schotman, 2010; Yan and Zivot, 2010; Putniņš, 2013), the IS is still one of the most widely used measures for price discovery as documented by its adoption in recent works as well (Chen and Tsai, 2017; Kryzanowski et al., 2017; Lin et al., 2018; Ahn et al., 2019; Baur and Dimpfl, 2019; Brogaard et al., 2019; Hagströmer and Menkveld, 2019; Entrop et al., 2020). In this article I propose to adopt a completely data driven strategy based on *Independent Component Analysis* (ICA) to identify the SVECM widely adopted in the price discovery context, in order to solve a fundamental problem in the IS measure.

The proposed methodology wisely exploits the non-Normal distributions of stock returns to correctly identify the independent sources driving the shocks, and the associated mixing matrix according to which observed model residuals correlate across markets. The intuition directly arises from the possibility of introducing a data-driven technique in a research area in which is very hard to provide general and robust theory-driven identification strategies, exploiting the empirical properties of financial returns that perfectly fit the non-Normal requirement. Depending on the validity of the underlying assumptions two different and effective estimation procedures will be illustrated, leading to the introduction of two measures which are the *Independent Component based-Information Shares*(IC-IS) and the *Directed Acyclic Graph based-Information Shares*(DAG-IS). The idea of identifying the IS by means of the distributional properties of the variables was firstly introduced by Grammig and Peter (2013). The authors introduced the concept of tail dependence through the adoption in the modeling procedure of different variance regimes, inspired by Rigobon (2003), to identify the contribution of each market to the price discovery process. The intuition was that differences between tail and center correlations, caused by the occurrence of extreme price changes, could be exploited to reach full identification. In particular, following Lanne and Lütkepohl (2010), they assume price innovations to emerge as a mixture of two serially uncorrelated Normal random vectors with different covariance matrices, where one is the identity and the other is a diagonal matrix associated to different variance regimes. Still providing a solution based on the exploitation of the statistical properties of the variables of interest, the methodology proposed in this article differs under many aspects. First, the methodology which I am going to propose can work under any non-Normal distribution and with no need of introducing different volatility regimes to identify the model. Second, the problem will be addressed through both semiparametric and parametric procedures depending also on the validity of the assumptions adopted in the modeling strategy. Third, keeping Hasbrouck (2019) as a clear benchmark, the strategy proposed in this article is found to provide coherent empirical results under different time specifications when identifying the leaders and the followers in the price formation process. For all of these reasons the solution proposed in this article is general, at the cost of introducing the assumption of independent structural shocks in place of uncorrelated ones. Together with the assumption of the presence of an acyclic contemporaneous causal structure (Shimizu et al., 2006; Hyvärinen, 2013), I show we can consistently identify the causal chain in the system and thus the correct permutation of the variable in the VECM with subsequent unique identification of the IS measures. Furthermore, for the cases in which the recursive structure will not be confirmed to exist by the data, I show that the same identification problem can be addressed also relaxing the just mentioned acyclic causal assumption. Modeling through independent components found several successful applications ranging from signal processing and source separation to noise reduction and time series analysis, having the roots in the computer science research area (Comon, 1994; Hyvärinen and Oja, 2000) but gaining increasing attention in economics and finance as well as in the statistical and econometrics literature. Recent developments about the ICA approach can be found particularly in macroeconometrics where the identification issue of structural VAR (SVAR) models is pervasive (Moneta et al., 2013; Gouriéroux et al., 2017; Lanne et al., 2017) but applications can be found also in financial econometric and forecasting studies (Audrino et al., 2005; García-Ferrer et al., 2012; Fabozzi et al., 2016; Hafner et al., ress). The article is organized as follows. In section 2 the general setting is provided, showing the baseline model with its identification issues for price discovery. In section 3, the models and assumptions are illustrated explaining the identification scheme and a simulation exercise is provided to clarify the methodology. In section 4, the possibility of relaxing the recursive structure assumption is addressed through the adoption of an alternative estimation and identification procedure. Section 5 provides an empirical application on IBM 3 October 2016 intraday data, in order to have the results of Hasbrouck (2019) as a clear benchmark to compare with. Conclusions and discussions are provided in section 6.

2 General setting

In this section I briefly go back to the microstructure setting introduced in Hasbrouck (1995), which exploits the vector error correction representation of Engle and Granger (1987), and repeated in Hasbrouck (2019). This will be preparatory for the methodology proposed and illustrated in the next section. The starting point is to consider a set of time series log-prices $\{p_1, p_2, \dots, p_n\}$ observed in n different exchanges but pertaining the same security, thus all arbitrage linked and whose dynamic are modeled by VECM:

$$\Delta p_t = \boldsymbol{\alpha}\boldsymbol{\beta'} p_{t-1} + \sum_{i=1}^{k} \Phi_i \Delta p_{t-k} + \epsilon_t \tag{1}$$

where the matrix $\boldsymbol{\beta}$ contains the $n-1$ cointegrating vectors specified as $p_1 - p_2$, $p_1 - p_3$, $p_1 - p_n$ since all price series naturally cointegrate each other, and $\boldsymbol{\alpha} \in \mathbb{R}^{n \times n}$ is a loading matrix. The system in equation 1 is covariance stationary, with $\text{Cov}(\epsilon_t) = \Omega$, and thus admits a VMA($\infty$) representation

$$\Delta p_t = \Psi(L)\epsilon_t \tag{2}$$

with

$$\Psi(L) = \sum_{i=0}^{\infty} \Psi_i L^i, \tag{3}$$

and also possess, as implied by the Granger representation theorem, a common trend representation given by

$$p_t = p_0 + \Psi(1) \sum_{i=1}^{t} \epsilon_i + \Psi^*(L)\epsilon_t \tag{4}$$

where holds the decomposition $\Psi(L) = \Psi(1) + (1-L)\Psi^*(L)$, which can be seen as the multivariate generalization of the decomposition introduced in Beveridge and Nelson (1981). The second term on the right hand side of equation 4 is the random walk component driving all prices in the system, and thus can be identified as the latent efficient price process, while the last term is the transitory component admitting the VMA(∞). The matrix $\Psi(1)$ can be computed as (Johansen, 1991):

$$\Psi(1) = \beta_\perp \left[\alpha_\perp^{'} \left(I - \sum_{i=1}^{k} \Phi_i \right) \beta_\perp \right]^{-1} \alpha_\perp^{'} \tag{5}$$

and has rank equal to one which is the dimension of the efficient price process behind the observed series, thus all rows of $\Psi(1)$ are identical. The information share measure for market j is the share of variance of the common component which is induced by the *jth* market, which means

$$IS_j = \frac{\psi_j^2 \Omega_{jj}}{\psi \Omega \psi'} \tag{6}$$

with ψ being the common row of $\Psi(1)$ and ψ_j denoting the *j-th* element of ψ corresponding to market j. The above definition uniquely allocate the total variance across markets only if the covariance matrix of the innovations Ω is diagonal, while an identification issue arises when price innovations are correlated. To deal with a non-diagonal Ω two practical solutions have been proposed. The first is to rewrite ϵ_t in terms of orthogonal innovations u_t as

$$\epsilon_t = C u_t \tag{7}$$

where C is the Choleski decomposition of Ω. The IS thus can be computed in terms of the new orthogonal innovations u_t, that is

$$IS_j = \frac{\left([\psi C]_j\right)^2}{\psi \Omega \psi'}. \tag{8}$$

This allocation mechanism defined through the causal chain implied by the lower triangular structure of C depends on the particular order in which the variables are inserted in the VECM, thus the heuristic solution was to consider upper and lower bounds for the IS of each market by considering all the possible variable permutations. The second practical solution consists in drastically reducing the gap between upper and lower bounds, in order to eliminate interpretative ambiguities, estimating the model in natural time at very high resolutions. Non zero cross correlations in Ω naturally arise as the sampling interval increases indeed (Hasbrouck, 2019; Dias et al., 2020), thus they can be minimized by sampling at higher frequencies. This clearly comes at costs, including both the computational aspect of dealing with such a number of observations characterized by high level of sparsity and a suitable model specification to estimate the coefficients still considering a sufficiently long lag-structure in the data. As explained also in Hasbrouck (2003), the upper and lower bounds of the IS measures cannot be interpreted as confidence interval but rather as an identification problem. In the next section I will propose a methodology to uniquely identify, under few assumptions, the correct permutation of the variables in the system in order to recover the exchanges which lead the price discovery and the following ones. Moreover, for the sake of completeness, a further identification strategy still based on the non-normality of the variables will be addressed without the imposition of any recursive causal structure.

3 Model and assumptions

Consider the n-dimensional vector of price innovations $\epsilon_t = [\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{nt}]$ characterized by the non-diagonal covariance matrix Ω. Assume these observed signals to be a linear mixture of hidden components η_t, which can be modeled as

$$\epsilon_t = A_0 \eta_t, \tag{9}$$

where A_0 is a $n \times n$ mixing matrix through which the latent structural shocks η_t are revealed in each market. The equation in 9, which closely resembles equation 7, can be estimated up to permutation, sign, and scaling under some assumptions (Comon, 1994).

Assumption 3.1. The sequence of hidden sources, with finite and non-zero variance, of market microstructure noise η_t possess at most one Normal marginal distribution;

Assumption 3.2. Independence of the latent shocks: $p(\eta_1, \eta_2, ..., \eta_n) = \prod_i^n p(\eta_i)$.

The independence assumption regarding the hidden structural components driving price innovations, could be justified, from a microstructural perspectives, by personal behavioral strategies, institutional rules and technological endowments which possibly differ across listing exchanges and trading platforms. Still, observed price innovations are allowed to correlate each other by means of the mixing matrix A_0 (for example as a consequence of the time aggregation in the sampling process previously mentioned). Since we directly observe only the mixtures, the independence of the hidden sources cannot be tested and has to be assumed. Concerning the non-normality assumption it is more a stylized fact rather than assumption, given the well established leptokurtic shape of the distribution of financial returns. The independence assumption of the non-Normal structural shocks η_t is a stronger concept than uncorrelatedness which is not sufficient alone to get rid of all the dependence information in the data. This additional information is what allows to reach full identification of the model if there exists a contemporaneous causal chain between the variables in the system, thus leading to the third and last assumption.

Assumption 3.3. The observed price innovations ϵ_t can be arranged in a causal chain, meaning that their data generating process possesses a *directed acyclic graph structure* (DAG) (Spirtes et al., 2000).

Under assumption 3.3 we can model the system in equation 9 as the following structural model,

$$\epsilon_t = B_0 \epsilon_t + \eta_t \tag{10}$$

where $A_0 = (I - B_0)^{-1}$ and the assumption of acyclical contemporaneous causal structure implies there exists an appropriate ordering of the variables according to which B_0 is lower triangular. We refer to this model as the *Linear Non-Gaussian Acyclic Model* (LiNGAM) introduced by Shimizu et al. (2006) in the research field of non-Normal Bayesian networks. To easily understand why non-normality is fundamental in the above specified model, consider for simplicity the two-dimensional case with ϵ_1 and ϵ_2. The structural model should identify one variable as the exogenous and the other as the dependent one. In the gaussian case independence and uncorrelatedness coincide and there would be no way to understand whether $\epsilon_1 \to \epsilon_2$, or vice versa, relying on the covariance matrix of η_t (Hyvärinen, 2013). This because the spherical symmetry of the joint normal distribution of the random vector η_t makes impossible to uniquely identify the matrix A_0, which could be estimated only up to an arbitrary orthogonal transformation as usually done with the PCA approach in the Gaussian case (Moneta et al., 2013). In what follows, the estimation steps necessary to achieve the structural identification of the VECM with associated information shares are illustrated.

3.1 Quantifying non-normality and recovering the independent components

The first step necessary for the identification process is to recover the non-Normal and statistically independent sources η_t from the observed price innovations ϵ_t. This requires the adoption of suitable measures which quantify the non-normality of a random variable. The estimation can thus proceed by estimating the mixing matrix A_0 such that the non-normality of η_t is maximized. There are many approaches to estimate the model in 9 based for instance on the maximization of the kurtosis, negentropy, or minimization of the mutual information between the random variables. All methodologies are closely related and exploit the central limit theorem. The additive mixture ϵ_t of independent and non-normal components η_t, is always closer to a Normal distribution than the latter. Thus, maximizing the non-normality of η_t directly relates to finding a direction of the space through the inverse of A_0 such that their mutual dependence is minimized. Going to the optimization schemes implemented so far in the literature, in this work the FastICA algorithm of Hyvärinen and Oja (2000) is firstly adopted being one of the most popular algorithm whose performances have been assessed theoretically and empirically, both from a computational and statistical perspective, and for which efficient variants of the algorithm have been also provided (Koldovsky et al., 2006; Miettinen et al., 2017; Moneta and Pallante, 2020). The algorithm solves the optimization problem quantifying the non-normality in terms of *approximated negentropy*. The entropy (amount of information) for a continuous random variables x is defined as

$$H(x) = -\int f(x)\, log\, f(x)\, dx. \tag{11}$$

Given that a normal variable has the largest entropy among random variables of equal variance (Cover and Thomas, 1991), one could optimally quantify, at least theoretically, the non-normality of a random variable by looking at the difference between its entropy and the one of a normal variable with the same variance. The so called *negentropy* is thus defined as

$$J(x) = H(\mathcal{N}) - H(x). \tag{12}$$

However, this would require in practice the knowledge of the probability density function from which the data are generated. For this reason the algorithm deals with an useful approximation of the negentropy of a random variable which takes the form

$$J(x) \approx \left[E\left(g(x)\right) - E\left(g(\mathcal{Z})\right)\right]^2, \tag{13}$$

where $\mathcal{Z}$ is a standardized normal and $g(\cdot)$ is any suitable non-quadratic function used to approximate the negentropy given the data (Hyvärinen and Oja, 1998), here $g(x) = -e^{-x^2/2}$. First, the mixtures are centered to be zero mean and whitened (i.e. uncorrelated and with their variances equal to one) which means I work with the quantities $\boldsymbol{z} = PD^{-1/2}\boldsymbol{\epsilon}$, where PDP^t is the spectral decomposition of the covariance matrix of the mixtures Ω. The algorithm searches for a vector $\boldsymbol{w}$ which maximizes the non-normality of $\boldsymbol{w}^t\boldsymbol{z}$ measured as shown by equation 13, that is

$$\hat{\boldsymbol{w}} = argmax_{\boldsymbol{w}} E(J(\boldsymbol{w^t z})). \tag{14}$$

The asymptotic properties of the resulting estimates are studied in detail by Reyhani et al. (2012). In particular, under the assumptions of Lipschitz continuity for both the first and second order derivatives of $g(\cdot)$, denoted by $g'(\cdot)$ and $g''(\cdot)$, boundedness of $g''(\cdot)$, that $E(\boldsymbol{z}) = 0$ and all the moments of $\boldsymbol{z}$ up to the fourth exist, then the FastICA estimator

$$\hat{\boldsymbol{w}} = \{\boldsymbol{w} : E(\boldsymbol{z}g(\boldsymbol{w}^t\boldsymbol{z}) = 0))\}, \tag{15}$$

with $E(\boldsymbol{z}g(\boldsymbol{w}^t\boldsymbol{z}) = 0$ being the first order condition for the maximization problem in 14, is consistent and asymptotically normal, that is

$$\sqrt{n}(\hat{\boldsymbol{w}} - \boldsymbol{w}) \xrightarrow{d} \mathcal{N}(0, \Omega). \tag{16}$$

Other studies proposed to use non-Normal distributions to identify structural shocks in SVAR models (Lanne and Lütkepohl, 2010; Lanne et al., 2017; Gouriéroux et al., 2017) by assuming specific density functions for the structural shocks and deriving the asymptotic properties of their estimates. Here we start following a flexible semi-parametric approach which allows for any form of non-normality in the data, while a parametric approach will be implemented when relaxing the recursive structure assumption.

3.2 Identifying the acyclical causal structure

Until now I made use only of assumptions 3.1 and 3.2 to estimate η_t and the mixing matrix A_0 up to permutation, sign, and scaling. The permutation indeterminacy in particular prevent the possibility to determine an appropriate order for the variables. I thus introduce at this point assumption 3.3 to identify the structural model by adapting to our context a causal search algorithm, well established in the machine learning research area (Shimizu et al., 2006; Hyvärinen et al., 2010), in which the acyclicity assumption makes possible to exploit statistical dependencies to recover a unique causal chain between price innovations in the SVECM. As a consequence I will be able to impose a specific order of the variables in the Choleski decomposition. In algorithm 1, the whole procedure to finally get the IS measure for each market without permutation indeterminacy is illustrated. While step 3 deals with the scaling indeterminacy of the ICA estimation, steps 2 and 4 deal with the sign and permutation indeterminacy which is the crucial problem we have when we want to identify the IS measures for each market, leading to proposition 3.1.

Algorithm 1 VECM-LiNGAM algorithm for IS measures

1: Estimate the VECM equation by equation given the known cointegrating relationships, and perform the ICA estimation on the model residuals (any suitable ICA estimator) to recover A_0 and η_t.

2: Given the unmixing matrix $W = A_0^{-1}$, find the permutation of the rows of W such that the permuted version W^* minimize $\sum_i^n 1/|W_{ii}^*|$.

3: Divide each row of W^* by its diagonal element so to get a matrix $\tilde{W}$ with ones in the main diagonal.

4: Let $\tilde{B}_0 = I - \tilde{W}$ be the estimate of B_0. Find a permutation matrix Z such that $Z\tilde{B}_0Z'$ as close as possible to be strictly lower triangular. Set the upper triangular elements to zero and permute back to get the matrix $\widehat{B}_0$ containing the directed acyclical graphical structure (DAG). A non zero element b_{ij} in matrix $\widehat{B}_0$ indicates the variable in position j to cause the variable in position i.

5: Thus, order the variable in the VECM according to the DAG structure obtained and perform Choleski on the estimated price innovations. Compute the IS measures.

It is useful to note that a test of statistical significance for the non zero elements of $\tilde{B}_0$ can be performed with no difficulties by boostrapping if a sufficiently long time series is available, which is the case for high-frequency data.

Proposition 3.1. Suppose that assumptions 3.1, 3.2 and 3.3 hold true. Then the Information Shares computed by following algorithm 1 are uniquely identified.

Proof. Let $\boldsymbol{\sigma} = \{\sigma_1, ..., \sigma_{n!}\}$, with

$$\sigma_i = \begin{pmatrix} 1 & 2 & \dots & n \\ \sigma_i(1) & \sigma_i(2) & \dots & \sigma_i(n) \end{pmatrix}$$

and $\sigma_i(\cdot) : \{1, ..., n\} \to \{1, ..., n\}$, be the set of all possible permutations of the n variables in the model. Consider the set of the Cholesky factors, of the covariance matrices, associated to each permutation of the variables $\boldsymbol{C_{(\sigma)}} = \{C_{(\sigma_1)}, ..., C_{(\sigma_{n!})}\}$. The uniqueness of the Information Share follows directly from the fact that given the estimates of the independent components, there is only one permutation, among the possible ones, yielding a strictly lower triangular matrix $\hat{B}_0$ representing the DAG structure of the variables in the model (for the proof see Shimizu et al., 2006). Then, being σ_i^* and $C_{(\sigma_i^*)}$ unique solutions, the identified Information Shares given the estimated DAG structure and computed as

$$DAG - IS_j = \frac{\left(\left[\psi C_{(\sigma_i^*)}\right]_j\right)^2}{\psi \Omega \psi'} \tag{17}$$

are unique. □

Thus, the identification scheme proposed ensures the uniqueness of the permutation according to which the price innovations in ϵ_t are mapped in a one-to-one correspondence with the structural shocks η_t. In the next section, a simulation exercise is provided to clarify the methodology.

3.2.1 An illustrative simulation exercise

Here I present the proposed identification mechanism on simulated data. In light of assumptions 3.1 and 3.2 I generate samples of T=5000 observations of independent sources η_t from an *Exponential Power Distribution* (EPD) whose density function is defined as

$$f(\eta \mid p,\ \mu,\ \sigma_p) = \frac{p}{2\sigma_p p^{1/p}\Gamma(1+1/p)} exp\left(-\frac{1}{p}\left|\frac{\eta-\mu}{\sigma_p}\right|^p\right) \tag{18}$$

where

$$\begin{aligned}\Gamma(1+1/p) &= \int_0^\infty \eta^{1/p} e^{-\eta} dx \\ &= (1/p)!\end{aligned} \tag{19}$$

is the gamma function. The variances are governed through the scale parameter σ_p according to

$$\sigma^2 = \frac{\sigma_p^2 \Gamma(3/p)}{\Gamma(1/p)} \tag{20}$$

Since we need η_t to be non-Normal, I choose to simulate from the EPD density (see Nardon and Pianca, 2009; Kalke and Richter, 2013, for extensive discussions about simulation methodologies) to have flexibility in modeling through the additional shape parameter p. The EPD become a normal when $p = 2$ and allows for fat tails by setting $p < 2$ (DiCiccio and Monti, 2004; Nadarajah, 2005), which is useful in the present setting to simulate data displaying excess kurtosis as financial price changes do. When $p = 1$ the distribution converges to a Laplace, I start with a shape parameter $p = 1.2$ which implies an excess kurtosis of 1.8 according to

$$k = \frac{\Gamma(1/p)\Gamma(5/p)}{\Gamma(3/p)^2} - 3. \tag{21}$$

Typically, intraday financial returns display higher levels of volatility both at the beginning and at the end of the trading day, and lower levels of volatility in the middle. For this reason I let the variance of the distributions from which I simulate η_t vary over time, modelling it through the diurnal U-shape pattern (Hasbrouck, 2002a; Andersen et al., 2012; Bollerslev et al., 2016).

$$\sigma_{\eta t} = C + Ae^{-at} + Be^{-b(1-t)} \tag{22}$$

where parameters are set as in Andersen et al. (2012), that is $C = 0.88929198$, $A = 0.75$, $B = 0.25$, $a = 10$, and $b = 10$. Modeling stochastic volatility is clearly not the objective of this section and this is the reason why the variance of η_t vary only in a deterministic manner, showing the robustness of the proposed approach in recovering the IS even in presence of time-varying parameters. Moreover, the stochastic behavior is still present given the random extraction from distributions with time-varying variances. The simulated non-Normal and independent shocks η_t are then mixed through matrix A to obtain the observed correlated signals ϵ_t. We thus simulate the data starting with a 2-dimensional VECM, where the signals ϵ_t are obtained as a mixture of the simulated structural shocks η_t through the mixing matrix

$$A_0 = \begin{pmatrix} 0.9 & 0 \\ 0.4 & 0.3 \end{pmatrix}. \tag{23}$$

Details about the simulation setting and parameters are provided in the supplemental appendix. Note that the lower triangular structure of the mixing matrix implies a causal chain in the SVECM which goes from the first market to the second one. The shocks in η_t are set to be independent and such that $Cov(\eta_t) = \Sigma_t$ is diagonal with equal variances, the information shares of the two markets are affected by the speed of adjustments in $\boldsymbol{\alpha}$ as well. In particular, the first market is much faster in adjusting towards the simulated efficient price process than the second one. With the specified parameters, the true IS measures (i.e. the ones given by Choleski if the variables are correctly ordered with respect to the simulated causal chain) are $IS_1 = 0.9151$ and $IS_2 = 0.0849$, which means the second market is dominated by the first one in the price discovery process. I thus recover η_t and an estimate of A_0 up to sign, permutation, and scaling as illustrated in section 3, and I proceed with algorithm 1 which yields the DAG structure contained in

$$\widehat{B_0} = \begin{pmatrix} 0 & 0 \\ 0.46 & 0 \end{pmatrix}. \tag{24}$$

The non zero element in b_{21} indicates that the structural shocks propagate from the first market to the second one, which is consistent with the causal chain imposed in the simulation. The identification of the causal chain is necessary to decide the correct permutation of the variables when computing the IS measure. Given the above detected acyclical structure, the estimated IS measures become $\widehat{IS_1} = 0.9151$ and $\widehat{IS_2} = 0.0849$ which coincide with the true one with no surprise since the correct permutation of variables has been detected. Note that applying all the possible permutations as usually done with the IS measure, the bounds would have been very wide with $IS_1 = [0.6583, 0.9198]$ and $IS_2 = [0.08, 0.3417]$. Even if in both cases we are able to attribute most of the price discovery to the first market, the result sensibly changes. In the correct case the second market would not contribute at all to the price discovery, while with the wrong permutation we would attribute more than a 30 percent share which become a significant contribution to the price discovery process. In the 2-dimensional case the application aimed at recovering the correct causal chain was straightforward and with few room for errors. It is then interesting to repeat the simulation with a higher number of variables involved, to see whether the proposed methodology is still able to consistently reconstruct the simulated process. In the light of the empirical application provided in the next section, in which no more than 4-variables will be contemporaneously considered, I simulate a 4-dimensional VECM process still driven by only one common stochastic trend. Again, I refer the readers to the supplemental appendix which contains the model's parameters adopted in the simulations. The signals ϵ_t are obtained by mixing η_t through the matrix

$$A_0 = \begin{pmatrix} 0.9 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0.5 & 0.2 & 0.7 & 0 \\ 0.3 & 0.5 & 0.3 & 0.1 \end{pmatrix} \tag{25}$$

whose lower triangular structure implies a causal chain from the first to the forth variables passing through the second and the third ones. With the specified parameters, the true IS measures are $IS_1 = 0.58$, $IS_2 = 0.01$, $IS_3 = 0.39$, and $IS_4 = 0.02$. The identification procedure yields the following acyclic structure

$$\widehat{B_0} = \begin{pmatrix} 0 & 0 & 0 & 0 \\ 0.44 & 0 & 0 & 0 \\ 0.42 & 0.43 & 0 & 0 \\ 0.2 & 0.68 & 0.43 & 0 \end{pmatrix}, \tag{26}$$

which means the estimated DAG structure consistently recover the causal chain from the first variable to the fourth, passing before through the second and third variables. Figure 1 shows the scatter plots for the reduced form residuals ϵ_t, clearly correlated as imposed in the data generating process (DGP), and the recovered independent structural sources η_t. Note that the estimated mixing matrix, upon which the causal search algorithm 1 is performed, closely resemble the true A_0 up to sign indeterminacy as shown below

$$\widehat{A_0} = \begin{pmatrix} -1 & 0.01 & 0.03 & 0.004 \\ -0.43 & 0.69 & 0.04 & 0.01 \\ -0.59 & 0.26 & -0.75 & 0.005 \\ -0.34 & 0.58 & -0.3 & 0.1 \end{pmatrix}. \tag{27}$$

The computation of the ISs going through all the possible permutations would provide us with $IS_1 = [0.1, 0.58], IS_2 = [0.01, 0.32], IS_3 = [0.1, 0.6]$, and $IS_4 = [0.01, 0.31]$, which make impossible to correctly disentangle the contribution of each market to the variance of the efficient price process. However, recovering the correct causal chain by means of the proposed identification strategy we are able to correctly permute the variables to get the true IS measures. The above simulations have been performed to illustrate the possibility of recovering, with a data-driven identification strategy, the permutation of variables in the VECM which is consistent with the real causal chain implied by the simulated DGP. While in the present setting the acyclical structure has been imposed and the procedure has been able to detect it, when working with real data there are no guarantees that the assumed recursive structure is respected. For this reason, an alternative identification approach for price discovery will be now illustrated.

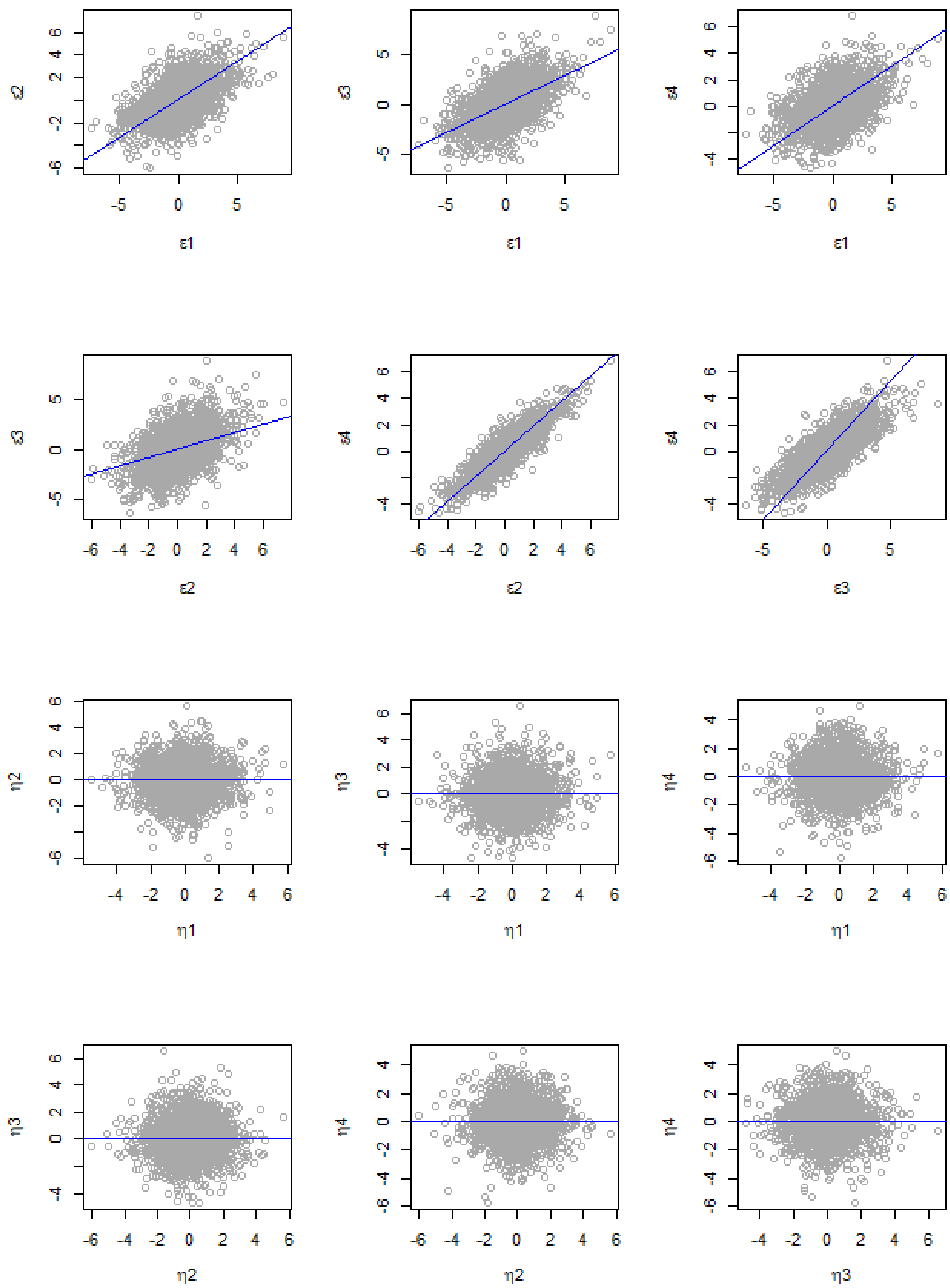

Figure 1: Scatter plots for the simulated reduced form residuals (top half) and estimated latent structural shocks (bottom half).

4 Relaxing the recursive structure assumption

4.1 Motivations

The LiNGAM approach for identifying the SVECM model in price discovery analyses always return an acyclical contemporaneous structure which is imposed *a-priori* in the identification scheme. This also happens with the Choleski decomposition, with the main difference that the variables' order is decided relying on economic theory when possible. Thus, it might be necessary to test the significance of the coefficient matrix B_0 of instantaneous effects, to understand whether the typically imposed assumption of acyclical contemporaneous structure is supported by the data or not. This can be done with a bootstrap approach if sufficiently long time series are available, in order to test both the stability of the causal chain recovered by the algorithm and the statistical significance of the coefficients in B_0. Also parametric approaches, however, have been recently introduced in the literature to estimate and identify SVAR models through ICA, giving the possibility of conducing statistical inference. These approaches do not impose a recursive structure which is considered only as a special case, but they do impose specific non-normal distributions for the error terms in order to compute the likelihood function to maximize in the estimation process. Two recent data-driven identifications of non-Normal SVAR models have been introduced in the econometric literature which are the maximum likelihood (ML) estimator of Lanne et al. (2017) which exploits the identification scheme of Ilmonen et al. (2011), and the pseudo-ML (PML) estimator of Gouriéroux et al. (2017). Both approaches give the possibility to statistically test usual identifying restrictions applied to SVAR models in macroeconomics, leading to a unique instantaneous coefficient matrix without imposing a recursive structure. Thus, they can be taken as alternative parametric identification schemes to compare with the semi-parametric approach previously illustrated when assumption 3.3 is not supported by the data. In what follows, given the typical leptokurtic behavior of the variables under analysis, the ML estimator under t-Student distributions will be implemented. Moreover, the identification scheme from Ilmonen et al. (2011) can be suitably adapted to our context in order to solve the lack of identification of the IS measure.

4.2 Identification and ML estimation

Consider again the SVECM model

$$\Delta p_t = \boldsymbol{\alpha\beta'} p_{t-1} + \sum_{i=1}^{k} \Phi_i \Delta p_{t-k} + A_0 \eta_t \tag{28}$$

where $A_0\eta_t = \epsilon_t$. The problem is again the identification of the matrix of instantaneous effects A_0, since for any nonsingular $n \times n$ matrix Q the matrix A_0 and the shocks in η_t can be replaced by $A_0 Q$ and $Q^{-1}\eta_t$ leading to an observationally equivalent model and thus to an identification problem. In typical SVAR analyses the covariance matrix of the structural shocks is restricted to a diagonal matrix so that $Q = DC$ where C is orthogonal and D is the diagonal matrix. This is what happens when identifying VAR models with Choleski indeed. As shown at the beginning of this section, under the assumptions 3.1 and 3.2 we can restrict the orthogonal matrix C to a permutation matrix P in order to allow only for permutation and scale indeterminacy in the columns of A_0. I now get rid of the assumption 3.3 and I follow a different identification procedure. Given the set $\mathcal{M}_n$ of nonsingular $n \times n$ matrices, and given the equivalence relation $\sim$, then $A_0^i \sim A_0^j \iff A_0^j = A_0^i DP$ for some diagonal matrix D and some permutation matrix P, with $A_0^i, A_0^j \in \mathcal{M}_n$, for $\forall i, j \in \mathcal{N}$ with $i \neq j$, defining a set of observationally equivalent SVECM processes. In the following, the steps necessary to get complete identification, thus selecting a particular SVECM model from its equivalence class, are illustrated.

Identification scheme. *For a given $A_0 \in \mathcal{M}_n$, take the sequence of transformations $A_0 \to A_0 D_1^+ \to A_0 D_1^+ P \to A_0 D_1^+ P D_2$ with: (i) D_1^+ a positive definite diagonal matrix such that the columns of $A_0 D_1^+$ have Euclidean norm one; (ii) P is the permutation matrix such that the generic element of $A_0 D_1 P$ in position ij satisfies $|c_{ii}| > |c_{ij}|$, $\forall\ i < j$; (iii) D_2 is the diagonal matrix such that all diagonal elements of $A_0 D_1^+ P D_2$ equals one. Let $\mathcal{A} \subseteq \mathcal{M}_n$ be the set containing all those A_0 for which the matrices D_1^+, P, and D_2 exist. Then, given the linear application $G(\cdot) : \mathcal{A} \to \mathcal{A}$ such that $G(A_0) = A_0 D_1^+ P D_2$, we define the set $\mathcal{A}_0$ consisting of unique, separate representative matrices from each equivalence class in $\mathcal{A}$ as $\mathcal{A}_0 = G(\mathcal{A}) = \{\tilde{A}_0 \in \mathcal{M}_n : \tilde{A}_0 = G(A_0)$ for $A_0 \in \mathcal{A}\}$.*

The identification scheme above makes possible to implement a particular scaling and permutation to identify a unique matrix A_0 from the equivalence class containing the observationally equivalent SVECM models. The identification strategy imposed in the estimation procedure is quite appealing since the set ε of matrices which have been excluded from the identification scheme has Lebesgue measure zero in $\mathbb{R}^{n\times n}$ (for the proof I refer the readers to Lanne et al., 2017), meaning that the exclusion of those models is not relevant in practice. The SVECM is estimated via ML as follows. The parameters of the model to be estimated are $\theta = [\boldsymbol{\gamma}, A_0, \lambda]$ and will be estimated in separate steps, where $\boldsymbol{\gamma} = (\boldsymbol{\alpha}, \Phi_1, ..., \Phi_k)$ and λ contains any additional parameter characterizing the specific density function chosen. Consider the generic standardized log-likelihood function of the model

$$L_t(\theta) = \frac{1}{T}\sum_{i=1}^{T} l_t(\theta) \tag{29}$$

where $l_t(\theta)$ is the sample log-likelihood at time t which depend on the specific non-Normal error density function $f_x(\boldsymbol{x};\theta)$ chosen. As previously anticipated, given the leptokurtic behavior of financial returns, the Student's t-distribution will be employed in the next section's empirical analysis and thus λ contains the degrees of freedom. First the reduced form VECM is estimated equation by equation via the least squares (LS) estimator, given the known cointegrating relationships, obtaining the parameters' estimate $\tilde{\boldsymbol{\gamma}}$. In the second step, given $\tilde{\boldsymbol{\gamma}}$, the reduce form residuals are recovered and the parameters related to $A_0\eta_t$ are estimated maximizing the log-likelihood function of the errors with respect to A_0 and λ (see Lange et al., ress, for the code implementation of this step). Full identification is reached by imposing the set of restrictions previously provided through the identification scheme when maximizing the likelihood. Finally, the estimates are further replaced in the log-likelihood of the model and a second maximization takes place in order to recover the ML estimates $\hat{\boldsymbol{\gamma}}$ as well. Under the usual regularity assumptions on the differentiability and integrability of respectively the log-likelihood function and its partial derivatives, ensuring the score function to be zero mean and with finite and positive definite covariance matrix evaluated at the true parameters $\Sigma_{\theta^T} = E[(l_t'(\theta)|_{\theta=\theta^T})(l_t'(\theta)|_{\theta=\theta^T})']$, then the ML estimator possesses the usual properties of consistency and asymptotic normality. That is

$$\sqrt{T}(\hat{\theta} - \theta^T) \xrightarrow{d} \mathcal{N}(0, \Sigma_{\theta^T}^{-1}). \tag{30}$$

This allow us to provide an estimate of the Information Share in terms of the independent structural shocks as

$$IC-IS_j := \frac{\left([\psi A_0]_j\right)^2}{\psi(A_0 \Sigma_\eta A_0')\psi'}, \tag{31}$$

where $\Sigma_\eta = \boldsymbol{I}_n$ and ψ is computed in the same manner with the SVECM parameters $\hat{\gamma}$ and the cointegration matrix $\boldsymbol{\beta}$. I refer to the modified measure as the *Independent Component based-Information Shares* (IC-IS) for which the identification problem is solved.

Proposition 4.1. The IC-IS measure is invariant to arbitrary permutations of the variables in the model.

Proof. Let $\tilde{A}_0 = A_0 D_1^+$ be the estimated mixing matrix of the independent shocks, with columns having Euclidean norm one. As a consequence of assumption 3.1 and assumption 3.2, column permutations create a lack of identification for $\tilde{A}_0$. Consider the mixing matrix identified through the transformation $\tilde{A}_0 \to \tilde{A}_0 P$, with P being the permutation matrix of the identification scheme such that the generic element of $\tilde{A}_0 P$ in position ij satisfies $|c_{ii}| > |c_{ij}|$, $\forall\ i < j$, and let be $\tilde{A}_0 P^{(r)}$ the estimated mixing matrix with randomly permuted columns through the random permutation matrix $P^{(r)}$. Then, there always exist a permutation matrix $Q = {P^{(r)}}^T P$ yielding $\tilde{A}_0 P^{(r)} Q = \tilde{A}_0 P$. □

Exploiting basic properties of the permutation matrices, the proof above simply shows that independently from the arbitrary column ordering of the mixing matrix we will always be able to find an appropriate permutation matrix, since the product of two permutation matrices is a permutation matrix itself, such that the identification scheme imposed in the estimation procedure is respected. In the next section an empirical application is provided, implementing the proposed methodologies on real high-frequency data.

5 Empirical application

5.1 Benchmarking the models

Bringing the procedure on high-frequency data exposes to several caveats, mostly related to the sparsity of the data and to model specification issues. To have a benchmark to compare with, I empirically test the proposed methodologies on the same IBM data adopted by Hasbrouck (2019), for the day 3 October 2016, which have been shared under the authorization of the NYSE making this analysis possible. I thus try to disentangle the relative contribution to the price discovery process of listing and non-listing exchanges, participant-based and SIP-based quotes, trades and quotes. As previously illustrated, the main power of the approaches relies in the exploitation of the non-Normal distributions to separate the sources of noise in each variable. In this respect it becomes interesting to test the model stability both in natural and event time, adopting a relatively low level of resolution (i.e. second precision) in the data for the natural time specification. This to eventually check the consistency of the obtained results in both time specifications without increasing the computational complexity and data sparsity introduced when working at very high frequencies.

5.2 IBM, 3 October 2016

As anticipated above, the empirical application focuses on some detailed analyses already conducted in the literature in order to have a direct comparison which makes clearer the interpretation of the obtained results. The econometric analysis is performed on IBM's quotes and trades for the day 3 October 2016, with each record reporting both participant-based and SIP-based timestamps. VECM models are thus estimated both in natural and event-time with a maximum lag $k = 10$, and then the data-driven identification strategies for the IS measures are implemented. The first study disentangles the impact of time reporting differentials on the quantification of price discovery measures, through the estimation of a 4-variables VECM including national best bids (NBBs) and offers (NBOs) constructed from both participant and SIP timestamps. Given that the SIP timestamps are by construction delayed signals of the participant ones, one expects to attribute the price discovery to the participant-based data. I then proceed with the second analysis which consists in quantifying the price discovery in both the primary listing and other exchanges. The VECM will include bids and offers placed on the primary listing, plus best bids and offers taken from all the exchanges except the primary one. Finally, the third study is aimed at determining the relative contributions of trades and quotes. I thus insert in the model trades occurred on lit and dark pools separately, plus NBBs and NBOs quotes from participant timestamps. In figure 2, the quantile-quantile plots for the VECM reduced form residuals are displayed. It can be noticed that the residuals are visibly leptokurtic as expected (the normality hypothesis was soundly rejected at the 1% by different tests usually adopted as the Jarque-Bera and the Shapiro-Wilk tests). This to justify the suitability of an identification approach based on the requirement of non-Normality.

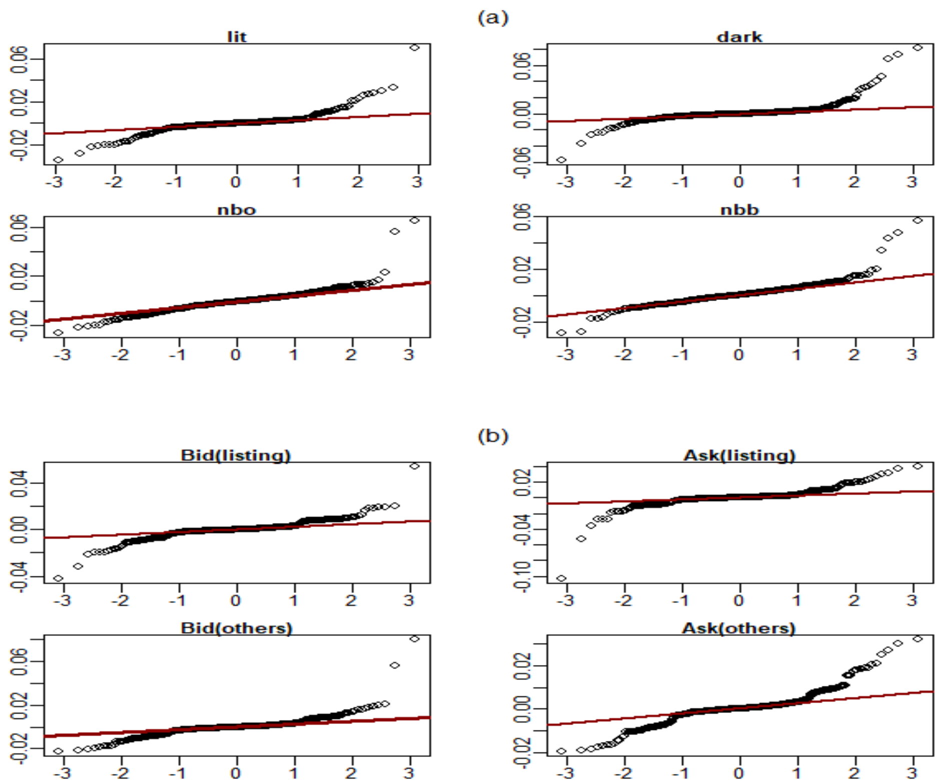

Figure 2: Quantile-quantile plots of the VECM reduced form residuals. In Panel (a) are displayed the model residuals related to the price discovery analysis across trades and quotes, while in panel (b) the one across exchanges using quotes.

The residuals of the models estimated for the participant versus SIP timestamps are not reported in the quantile-quantile plots to avoid useless redundancies, given that the variables would be again NBBs and NBOs with just the time-delays differentials in reporting them. For each model related to a given price discovery analysis, the two identification procedure proposed in this work, leading respectively to the DAG-IS and IC-IS measures, are performed and compared with the approach in which upper and lower bounds are computed by going through all the possible permutations in Choleski. While table 1 shows the estimated coefficients of the structural matrix A_0 for each analysis and identification procedure, table 2 summarizes the information shares estimated for each variable. The autoregressive and loading coefficients, for each estimated VECM, are not reported here for the sake of brevity and can be found in the supplemental online appendix. However, as also reported in Hasbrouck (2019), estimates are mostly insignificant at the 1-second resolution while they are very significant in the event-time specification. The DAG-IS measure, defined by adopting the LiNGAM approach, is able to identify the participant timestamps as the dominating ones thus suggesting the correct variable's order in the system even in the low resolution case (1-second precision) in which a permutation approach does not solve the identification issue given the very wide upper and lower bounds. There is no doubt in the event time specification instead, where also the approach based on all the possible permutations identify the participant timestamps as the variables leading the price formation process. Noticeably the result differs when identifying the model with the ML approach leading to the IC-IS measure, given that it attributes higher importance to the SIP timestamps especially in the natural time framework, where up to the 30 percent of information share is attributed. Still, also the IC-IS measure is able to uniquely identify the participants as the the ones who lead the price discovery and the SIP timestamps as the following ones. Also in the price discovery across exchange analysis, both approaches yield identification results which consistently identify the primary listing exchange as the leader both in natural and event-time. This would not be possible using the heuristic solution with upper and lower bounds. It has to be noticed, however, that the DAG-IS works by finding a permutation respecting the most the statistical dependencies of the data but does not solve the temporal aggregation issue we have when using low levels of resolution. This means that if we discard price variations in each market by aggregating over seconds, the measurement will be obviously overestimated or vice versa. This is not the case for the parametric approach which yielded a consistent result across the two different time specifications, despite it attributes to the listing exchange around 13 percent more (0.56 vs 0.691) than what obtained by assuming a recursive structure as done either in LiNGAM or all Choleski permutations.

Table 1: Estimated instantaneous effect matrices A_0.

NATURAL-TIME (1-SEC)									
participant VS SIP timestamps									
LiNGAM	1	0	0	0	MLE t-student	1	**-0.62**	**1.003**	-0.74
	0.34	1	**-0.36**	0		**0.67**	1	-0.062	**0.9**
	-0.99	0	1	0		**1.002**	**-0.61**	1	**-1.09**
	0.016	**-1.001**	-0.016	1		**0.67**	**1.23**	-0.06	1
primary VS non-primary									
LiNGAM	1	0.026	**-0.45**	**-0.22**	MLE t-student	1	0.007	-0.014	**0.06**
	0	1	**-0.23**	**-0.45**		**0.35**	1	**0.053**	**0.15**
	0	0	1	0		**0.74**	**0.04**	1	**0.07**
	0	0	-0.35	1		**0.1**	**0.06**	**0.04**	1
Quotes VS Trades									
LiNGAM	1	0	-0.0013	0	MLE t-student	1	-0.0017	-0.008	**0.057**
	0.012	1	0	**0.039**		-0.0027	1	-0.0053	**0.043**
	-0.062	0	1	0		**0.032**	**0.096**	1	0.0002
	-0.051	0	**0.071**	1		**0.067**	0.0048	**0.18**	1
EVENT-TIME									
participant VS SIP timestamps									
LiNGAM	1	-0.05	-0.038	-0.046	MLE t-student	1	-0.008	0.0006	**-0.025**
	0	1	0.063	0		**0.052**	1	**-0.14**	**0.15**
	0	0	1	0		0.002	**0.004**	1	**0.03**
	0	**-0.12**	**0.13**	1		**-0.052**	**0.023**	**0.011**	1
primary VS non-primary									
LiNGAM	1	0	**-0.33**	-0.012	MLE t-student	1	-0.00047	**0.032**	-0.004
	0.08	1	**-0.015**	**-0.034**		**-0.033**	1	**0.016**	**0.044**
	0	0	1	0		**0.036**	-0.0054	1	**0.021**
	0	0	-002	1		**0.01**	0.0027	**0.019**	1
Quotes VS Trades									
LiNGAM	1	0	0	0	MLE t-student	1	**-0.014**	**-0.014**	0.002
	-0.011	1	-0.0083	0.019		**-0.006**	1	**-0.035**	-0.004
	-0.032	0	1	0		**0.05**	**0.01**	1	**-0.03**
	-0.033	0	**-0.028**	1		**0.03**	0.009	**0.1**	1

Notes: Coefficients in bold are significants at the 1% level in both LiNGAM and MLE t-student approaches. For the LiNGAM approach, statistical significance has been tested using standard errors from 1000 bootstrap samples.

.

Table 2: Information shares: Summary results.

	DAG-IS		IC-IS		All permutations			
	participants	SIP	participants	SIP	participants		SIP	
					Min	Max	Min	Max
1-sec	0.999	0.001	0.7	0.3	0.002	0.999	0.001	0.998
Event time	0.962	0.038	0.89	0.11	0.943	0.999	0.001	0.057
	primary	non-primary	primary	non-primary	primary		non-primary	
					Min	Max	Min	Max
1-sec	0.994	0.006	0.695	0.305	0.12	0.994	0.006	0.88
Event time	0.56	0.44	0.691	0.309	0.46	0.56	0.44	0.54
	Quotes	Trades	Quotes	Trades	Quotes		Trades	
					Min	Max	Min	Max
1-sec	0.67	0.33	0.65	0.35	0.39	0.979	0.021	0.61
Event time	0.64	0.36	0.62	0.38	0.61	0.67	0.33	0.39

Notes: Information shares measures obtained for each identification procedure and for each price discovery analysis across participants and SIP timestamps, trades and quotes, and exchanges. In the natural-time(1-sec) setting the most recent price observed in a given second interval is taken. In the event time specification, the time counter is incremented whenever there is an update to any variable in the system instead. Trades comprises both lit and dark trades, given that the contribution of the latter to the IS measure is negligible.

Finally, no sound difference between methodologies has been detected when measuring the informational content of quotes and trades. Quotes are more informative than trades and the finding is noticeably consistent across methodologies and time specifications. Since the contribution of dark trades turns out to be negligible, their shares have been put together with the ones of lit trades differentiating only between trades and quotes. Discrepancies in the results however can easily arise given the different assumptions and estimation techniques employed in the identification of the structural matrix A_0. Thus, the choice of employing a particular technique should depend on which assumptions behind the methodology are believed to be more realistic for the kind of data under treatment. Coefficients in table 1 show that, with few exceptions, the assumption of a recursive structure seems to be not supported by the data also in event-time, raising concerns about the lower triangular structure imposed both by LiNGAM and Choleski. When the assumption of a recursive structural relation between the variables is hardly justifiable, a more general approach is recommended but still keeping in mind that the resulting IC-IS will depend on the particular density function chosen. On the contrary, the DAG-IS is a good measure when we can recover a plausible causal chain between the variables involved. The fact that the DAG-IS is able to identify the leaders and the followers, even if a causal chain between all the variables is not supported by the data, should not surprise. This because the few significant coefficients in A_0 are not sufficient to describe the entire causal chain, but they are at least sufficient to tell us which variable, associated to some statistically significant causal relations, should occupy the firsts positions in the Choleski decomposition. Consequently, it is still possible to understand who leads the price formation process despite the shares will be quite imprecise. Given that each methodology possesses different strengths and shortcomings, providing both the DAG-IS and IC-IS could be a robust way to identify, with few room for errors in case of coherent results across methodologies, the variables leading the price discovery process. Otherwise, they should be carefully chosen depending on the plausibility of the underlying assumptions. Overall, the results obtained in the empirical application just illustrated are coherent in choosing the leaders in the price formation process, and in line with the results of Hasbrouck (2019) but without increasing the modeling and computational complexity introduced by working at incredibly short time-scales.

6 Conclusions

Measuring the informational content of fragmented financial markets acquired increasing importance over time for both academics and practitioners. This article proposes two novel methodologies, exploiting the typical non-Normal distributions of financial returns, to uniquely identify one of the most widely adopted measures for price discovery and for which no identification solutions had been proposed for almost twenty years until the first approach proposed by Grammig and Peter (2013). Differently from the cited approach, with this article I put forward two procedures in which the Information Shares measures can be always determined with no need of exploiting the possible presence of different volatility regimes caused by extreme price changes, thus providing a general identification framework for price discovery analyses. To this purpose two alternative measures are introduced, namely the DAG-IS and IC-IS measures depending on whether the presence of an acyclical causal structure among the variables is supported by the data or not. The new estimation procedures have been assessed both theoretically and empirically. Keeping the empirical analysis of Hasbrouck (2019) as a direct benchmark to compare with, the proposed procedure is found to yield coherent results even across different time specifications, being able to correctly identify the leaders in the price formation process. Given the flexibility of the modeling strategy which can be assessed from both a semiparametric and parametric perspectives, depending also on the few assumptions needed, future applications in the field might benefit from the revisited Information Shares measures here introduced.

References

Ahn, K., Y. Bi, and S. Sohn (2019). Price discovery among sse 50 index-based spot, futures, and options markets. *Journal of Futures Markets 39*(2), 238–259.

Andersen, T. G., D. Dobrev, and E. Schaumburg (2012). Jump-robust volatility estimation using nearest neighbor truncation. *Journal of Econometrics 169*(1), 75–93.

Audrino, F., G. Barone-Adesi, and A. Mira (2005). The stability of factor models of interest rates. *Journal of Financial Econometrics 3*(3), 422–441.

Baillie, R. T., G. G. Booth, Y. Tse, and T. Zabotina (2002). Price discovery and common factor models. *Journal of financial markets 5*(3), 309–321.

Baur, D. G. and T. Dimpfl (2019). Price discovery in bitcoin spot or futures? *Journal of Futures Markets 39*(7), 803–817.

Beveridge, S. and C. R. Nelson (1981). A new approach to decomposition of economic time series into permanent and transitory components with particular attention to measurement of the 'business cycle'. *Journal of Monetary economics 7*(2), 151–174.

Bollerslev, T., A. J. Patton, and R. Quaedvlieg (2016). Exploiting the errors: A simple approach for improved volatility forecasting. *Journal of Econometrics 192*(1), 1 – 18.

Booth, G. G., R. W. So, and Y. Tse (1999). Price discovery in the german equity index derivatives markets. *Journal of Futures Markets: Futures, Options, and Other Derivative Products 19*(6), 619–643.

Brogaard, J., T. Hendershott, and R. Riordan (2019). Price discovery without trading: Evidence from limit orders. *The Journal of Finance 74*(4), 1621–1658.

Brugler, J. and C. Comerton-Forde (2019). Comment on: Price Discovery in High Resolution*. *Journal of Financial Econometrics*. nbz005.

Buccheri, G., G. Bormetti, F. Corsi, and F. Lillo (2019). Comment on: Price Discovery in High Resolution. *Journal of Financial Econometrics*. nbz008.

Chen, Y.-L. and W.-C. Tsai (2017). Determinants of price discovery in the vix futures market. *Journal of Empirical Finance 43*, 59–73.

Comon, P. (1994). Independent component analysis, a new concept? *Signal processing 36*(3), 287–314.

Corsi, F. (2009). A Simple Approximate Long-Memory Model of Realized Volatility. *Journal of Financial Econometrics 7*(2), 174–196.

Cover, T. M. and J. A. Thomas (1991). Information theory and statistics. *Elements of Information Theory 1*(1), 279–335.

De Jong, F. (2002). Measures of contributions to price discovery: A comparison. *Journal of Financial markets 5*(3), 323–327.

De Jong, F. (2019). Comment on: Price Discovery in High Resolution*. *Journal of Financial Econometrics*. nbz006.

De Jong, F. and P. C. Schotman (2010). Price discovery in fragmented markets. *Journal of Financial Econometrics 8*(1), 1–28.

Dias, G. F., M. Fernandes, and C. M. Scherrer (2020). Price Discovery in a Continuous-Time Setting. *Journal of Financial Econometrics*. nbz030.

DiCiccio, T. J. and A. C. Monti (2004). Inferential aspects of the skew exponential power distribution. *Journal of the American Statistical Association 99*(466), 439–450.

Engle, R. F. and C. W. Granger (1987). Co-integration and error correction: representation, estimation, and testing. *Econometrica: journal of the Econometric Society*, 251–276.

Entrop, O., B. Frijns, and M. Seruset (2020). The determinants of price discovery on bitcoin markets. *Journal of Futures Markets 40*(5), 816–837.

Fabozzi, F. J., R. Giacometti, and N. Tsuchida (2016). Factor decomposition of the eurozone sovereign cds spreads. *Journal of International Money and Finance 65*, 1–23.

García-Ferrer, A., E. González-Prieto, and D. Peña (2012). A conditionally heteroskedastic independent factor model with an application to financial stock returns. *International Journal of Forecasting 28*(1), 70–93.

Ghysels, E. (2020). Comment on: Price Discovery in High Resolution and the Analysis of Mixed Frequency Data*. *Journal of Financial Econometrics*. nbz007.

Gonzalo, J. and C. Granger (1995). Estimation of common long-memory components in cointegrated systems. *Journal of Business & Economic Statistics 13*(1), 27–35.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2017). Statistical inference for independent component analysis: Application to structural var models. *Journal of Econometrics 196*(1), 111–126.

Grammig, J. and F. J. Peter (2013). Telltale tails: A new approach to estimating unique market information shares. *Journal of Financial and Quantitative Analysis*, 459–488.

Hafner, C. M., H. Herwartz, and S. Maxand (in press). Identification of structural multivariate garch models. *Journal of Econometrics*.

Hagströmer, B. and A. J. Menkveld (2019). Information revelation in decentralized markets. *The Journal of Finance 74*(6), 2751–2787.

Hansen, P. R. and A. Lunde (2006). Realized variance and market microstructure noise. *Journal of Business & Economic Statistics 24*(2), 127–161.

Harris, F. H. d., T. H. McInish, G. L. Shoesmith, and R. A. Wood (1995). Cointegration, error correction, and price discovery on informationally linked security markets. *The Journal of Financial and Quantitative Analysis 30*(4), 563–579.

Harris, F. H. d., T. H. McInish, and R. A. Wood (2002a). Common factor components versus information shares: a reply. *Journal of Financial Markets 5*(3), 341–348.

Harris, F. H. d., T. H. McInish, and R. A. Wood (2002b). Security price adjustment across exchanges: an investigation of common factor components for dow stocks. *Journal of financial markets 5*(3), 277–308.

Hasbrouck, J. (1995). One security, many markets: Determining the contributions to price discovery. *The journal of Finance 50*(4), 1175–1199.

Hasbrouck, J. (2002a). The dynamics of discrete bid and ask quotes. *The Journal of Finance 54*(6), 2109–2142.

Hasbrouck, J. (2002b). Stalking the "efficient price" in market microstructure specifications: an overview. *Journal of Financial Markets 5*(3), 329–339.

Hasbrouck, J. (2003). Intraday price formation in us equity index markets. *The Journal of Finance 58*(6), 2375–2400.

Hasbrouck, J. (2019). Price Discovery in High Resolution*. *Journal of Financial Econometrics*. nbz027.

Hatheway, F., A. Kwan, and H. Zheng (2017). An empirical analysis of market segmentation on us equity markets. *Journal of Financial and Quantitative Analysis 52*(6), 2399–2427.

Hyvärinen, A. (2013). Independent component analysis: recent advances. *Philosophical Transactions of the Royal Society A: Mathematical, Physical and Engineering Sciences 371*(1984), 20110534.

Hyvärinen, A. and E. Oja (1998). Independent component analysis by general nonlinear hebbian-like learning rules. *signal processing 64*(3), 301–313.

Hyvärinen, A. and E. Oja (2000). Independent component analysis: algorithms and applications. *Neural networks 13*(4-5), 411–430.

Hyvärinen, A., K. Zhang, S. Shimizu, and P. O. Hoyer (2010). Estimation of a structural vector autoregression model using non-gaussianity. *Journal of Machine Learning Research 11*(5).

Ilmonen, P., D. Paindaveine, et al. (2011). Semiparametrically efficient inference based on signed ranks in symmetric independent component models. *the Annals of Statistics 39*(5), 2448–2476.

Johansen, S. (1991). Estimation and hypothesis testing of cointegration vectors in gaussian vector autoregressive models. *Econometrica: journal of the Econometric Society*, 1551–1580.

Kalke, S. and W.-D. Richter (2013). Simulation of the p-generalized gaussian distribution. *Journal of Statistical Computation and Simulation 83*(4), 641–667.

Koldovsky, Z., P. Tichavsky, and E. Oja (2006). Efficient variant of algorithm fastica for independent component analysis attaining the cramér-rao lower bound. *IEEE Transactions on neural networks 17*(5), 1265–1277.

Kryzanowski, L., S. Perrakis, and R. Zhong (2017). Price discovery in equity and cds markets. *Journal of Financial Markets 35*, 21–46.

Kwan, A., R. Masulis, and T. H. McInish (2015). Trading rules, competition for order flow and market fragmentation. *Journal of Financial Economics 115*(2), 330–348.

Lange, A., B. Dalheimer, H. Herwartz, and S. Maxand (in press). svars: An R package for data-driven identification in multivariate time series analysis. *Journal of Statistical Software*.

Lanne, M. and H. Lütkepohl (2010). Structural vector autoregressions with nonnormal residuals. *Journal of Business & Economic Statistics 28*(1), 159–168.

Lanne, M., M. Meitz, and P. Saikkonen (2017). Identification and estimation of non-gaussian structural vector autoregressions. *Journal of Econometrics 196*(2), 288–304.

Lehmann, B. N. (2002). Some desiderata for the measurement of price discovery across markets. *Journal of Financial Markets 5*(3), 259 – 276. Price Discovery.

Lien, D. and K. Shrestha (2009). A new information share measure. *Journal of Futures Markets: Futures, Options, and Other Derivative Products 29*(4), 377–395.

Lin, C.-B., R. K. Chou, and G. H. Wang (2018). Investor sentiment and price discovery: Evidence from the pricing dynamics between the futures and spot markets. *Journal of Banking & Finance 90*, 17–31.

Miettinen, J., K. Nordhausen, H. Oja, S. Taskinen, and J. Virta (2017). The squared symmetric fastica estimator. *Signal Processing 131*, 402–411.

Moneta, A., D. Entner, P. O. Hoyer, and A. Coad (2013). Causal inference by independent component analysis: Theory and applications. *Oxford Bulletin of Economics and Statistics 75*(5), 705–730.

Moneta, A. and G. Pallante (2020). Identification of Structural VAR Models via Independent Component Analysis: A Performance Evaluation Study. *LEM Papers Series* (2020/24).

Nadarajah, S. (2005). A generalized normal distribution. *Journal of Applied Statistics 32*(7), 685–694.

Nardon, M. and P. Pianca (2009). Simulation techniques for generalized gaussian densities. *Journal of Statistical Computation and Simulation 79*(11), 1317–1329.

O'Hara, M. and M. Ye (2011). Is market fragmentation harming market quality? *Journal of Financial Economics 100*(3), 459 – 474.

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

Reyhani, N., J. Ylipaavalniemi, R. Vigário, and E. Oja (2012). Consistency and asymptotic normality of fastica and bootstrap fastica. *Signal processing 92*(8), 1767–1778.

Rigobon, R. (2003). Identification through heteroskedasticity. *Review of Economics and Statistics 85*(4), 777–792.

Shimizu, S., P. O. Hoyer, A. Hyvärinen, and A. Kerminen (2006). A linear non-gaussian acyclic model for causal discovery. *Journal of Machine Learning Research 7*(Oct), 2003–2030.

Spirtes, P., C. N. Glymour, R. Scheines, and D. Heckerman (2000). *Causation, prediction, and search*. MIT press.

Yan, B. and E. Zivot (2010). A structural analysis of price discovery measures. *Journal of Financial Markets 13*(1), 1–19.

Appendix

This supplemental appendix contains all the details about the simulation setting. The tables containing the VECM estimates for to the empirical application are also displayed.

Appendix A: Simulation details

Data for the illustrative exercise are simulated from the equivalent VAR representation of the VECM adopted in the paper as follows

$$\Pi(L)p_t = \epsilon_t \tag{32}$$

where

$$\Pi(L) \equiv I_n - \sum_i^k \Pi_i L^i \tag{33}$$

$$\boldsymbol{\alpha\beta'} = (\sum_i^k \Pi_i - I_n) \tag{34}$$

$$\phi_s = -(\Pi_{s+1} + \Pi_{s+2} + ... + \Pi_k) \tag{35}$$

for $s = 1, 2, ..., k-1$, and such that $|I_n - \Pi_1 z - \Pi_2 z^2 - ... - \Pi_k z^k| = 0$ has only one unit root since the system is driven by only one common stochastic trend. Consequently, the matrix $\boldsymbol{\beta}$ contains the known cointegrating vectors and has rank equal to *n-1*. In the two-dimensional case the parameters are

$$\boldsymbol{\alpha} = \begin{pmatrix} 0.1 \\ 0.5 \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.45 \\ 0.45 & 0.32 \end{pmatrix}, \quad \phi_1 = \begin{pmatrix} 0.6 & 0.3 \\ -0.7 & -0.9 \end{pmatrix}$$

$$\boldsymbol{\beta'} = \begin{pmatrix} 1 \\ -1 \end{pmatrix}, \quad \Pi_2 = \begin{pmatrix} -0.6 & -0.3 \\ 0.7 & 0.9 \end{pmatrix} \quad \Pi_1 = \begin{pmatrix} 1.7 & 0.2 \\ -0.2 & -0.4 \end{pmatrix},$$

while in the four-dimensional case are

$$\boldsymbol{\alpha} = \begin{pmatrix} 0.025 & 0.05 & 0.03 \\ 0.08 & 0.07 & 0.06 \\ 0.1 & 0.01 & 0.04 \\ 0.09 & 0.06 & 0.09 \end{pmatrix}, \quad \boldsymbol{\Omega} = \begin{pmatrix} 1 & 0.45 & 0.57 & 0.34 \\ 0.45 & 0.67 & 0.4 & 0.54 \\ 0.57 & 0.4 & 0.98 & 0.58 \\ 0.34 & 0.54 & 0.58 & 0.56 \end{pmatrix},$$

$$\phi_1 = \begin{pmatrix} 0.2 & -0.2 & -0.7 & 0.4 \\ 0.1 & 0.35 & 0.6 & 0.1 \\ 0.6 & 0.35 & 0.55 & -0.1 \\ 0.4 & -0.9 & -0.25 & 0.3 \end{pmatrix}, \quad \Pi_1 = \begin{pmatrix} 1.305 & -0.225 & -0.75 & 0.37 \\ 0.31 & 1.270 & 0.53 & 0.04 \\ 0.75 & 0.25 & 1.54 & -0.14 \\ 0.64 & -0.99 & 0-.31 & 1.21 \end{pmatrix},$$

$$\Pi_2 = \begin{pmatrix} -0.2 & 0.2 & 0.7 & -0.4 \\ -0.1 & -0.35 & -0.6 & -0.1 \\ -0.6 & -0.35 & -0.55 & 0.1 \\ -0.4 & 0.9 & 0.25 & -0.3 \end{pmatrix}, \quad \boldsymbol{\beta'} = \begin{pmatrix} 1 & \\ \vdots & -\boldsymbol{I_{n-1}} \\ 1 & \end{pmatrix}.$$

Appendix B: Additional tables

Below, the tables containing the estimates of the VECM coefficients for each empirical analysis are directly reported from the console output. With a1, a2, a3 I indicate the coefficients in the speed of adjustment matrix $\boldsymbol{\alpha}$. The four dependent variables at time t, which are not displayed with labels on the columns because of graphical issues caused by the large number of lags, have to be read (clearly from the left to the right) following the same order in which they appear as lagged exogenous variables in the first column.

Table 3: Primary VS non-primary exchanges in natural-time (1-sec) using quotes.

	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
NBBotherL1	4.461e-04	1.555e-03	9.082e-05	1.691e-03	-2.042e-03	0.0042	-8.415e-04	0.00524
NBOotherL1	7.019e-04	1.270e-03	1.096e-03	1.381e-03	-1.961e-03	0.00349	-5.120e-03	0.00428
NBBlistL1	-2.571e-03	4.849e-03	1.124e-02*	5.272e-03	4.438e-03	0.01333	7.902e-04	0.01635
NBOlistL1	5.012e-03	4.491e-03	-7.042e-03	4.882e-03	-5.879e-03	0.01235	7.773e-03	0.01514
NBBotherL2	-1.057e-04	1.556e-03	-3.656e-04	1.691e-03	-2.473e-03	0.00427	-1.001e-03	0.00524
NBOotherL2	1.053e-03	1.271e-03	-2.153e-03	1.382e-03	3.606e-04	0.00349	-8.363e-04	0.00428
NBBlistL2	-1.470e-02*	4.909e-03	2.898e-02*	5.337e-03	1.226e-02	0.01350	1.952e-02	0.01655
NBOlistL2	3.365e-02*	4.561e-03	-2.019e-02*	4.958e-03	3.996e-03	0.01254	-4.463e-03	0.01537
NBBotherL3	2.065e-03	1.560e-03	1.055e-03	1.696e-03	-9.143e-04	0.00429	-7.546e-04	0.00526
NBOotherL3	3.589e-04	1.271e-03	5.240e-04	1.382e-03	8.533e-04	0.00349	-1.575e-03	0.00428
NBBlistL3	-3.358e-02*	4.939e-03	2.970e-02*	5.369e-03	4.643e-03	0.01358	6.187e-03	0.01665
NBOlistL3	2.903e-02*	4.580e-03	-2.502e-02*	4.979e-03	6.407e-03	0.01259	3.937e-03	0.01544
NBBotherL4	1.730e-03	1.560e-03	-4.223e-04	1.696e-03	-6.053e-04	0.00429	-1.259e-03	0.00526
NBOotherL4	1.181e-03	1.270e-03	-5.052e-04	1.381e-03	3.267e-02*	0.00349	-8.110e-03	0.00428
NBBlistL4	-1.837e-02*	4.952e-03	3.424e-02*	5.383e-03	-1.615e-03	0.01361	1.173e-02	0.01669
NBOlistL4	2.655e-02*	4.586e-03	-3.612e-02*	4.986e-03	9.590e-04	0.01261	2.027e-03	0.01546
NBBotherL5	7.102e-04	1.560e-03	1.042e-03	1.696e-03	-2.301e-03	0.00429	1.259e-04	0.00526
NBOotherL5	5.380e-04	1.270e-03	6.282e-04	1.381e-03	-5.343e-04	0.00349	-1.790e-03	0.00428
NBBlistL5	-1.967e-02*	4.959e-03	3.529e-02*	5.391e-03	9.511e-03	0.01363	2.611e-02	0.01672
NBOlistL5	3.110e-02*	4.590e-03	-2.031e-02*	4.990e-03	8.762e-03	0.01262	-1.267e-03	0.01547
NBBotherL6	1.583e-03	1.561e-03	8.106e-04	1.697e-03	4.914e-03	0.00429	-1.603e-02*	0.00526
NBOotherL6	-8.537e-04	1.269e-03	5.279e-04	1.379e-03	-3.696e-02*	0.00348	-6.421e-02*	0.00427
NBBlistL6	-4.629e-02*	4.958e-03	4.436e-02*	5.390e-03	1.031e-02	0.01363	3.326e-02*	0.01671
NBOlistL6	3.900e-02*	4.591e-03	-4.446e-02*	4.991e-03	2.362e-02	0.01262	2.696e-02	0.01548
NBBotherL7	1.565e-03	1.561e-03	-2.901e-04	1.697e-03	-1.677e-03	0.00429	-1.652e-03	0.00526
NBOotherL7	8.768e-04	1.269e-03	5.720e-04	1.379e-03	-1.913e-03	0.00348	-1.434e-03	0.00427
NBBlistL7	-3.497e-02*	4.954e-03	3.784e-02*	5.386e-03	1.298e-03	0.01362	2.734e-02	0.01670
NBOlistL7	5.179e-02*	4.583e-03	-2.943e-02*	4.982e-03	3.982e-02	0.01260	8.151e-03	0.01545
NBBotherL8	2.254e-03	1.562e-03	2.395e-03	1.698e-03	-4.961e-03	0.00429	-4.692e-03	0.00526
NBOotherL8	1.464e-03	1.269e-03	1.661e-03	1.379e-03	-1.012e-03	0.00348	-2.093e-03	0.00427
NBBlistL8	-2.187e-02*	4.948e-03	6.024e-02*	5.379e-03	2.623e-02	0.01360	-2.095e-03	0.01668
NBOlistL8	6.906e-02*	4.572e-03	-2.672e-02*	4.970e-03	2.896e-02	0.01257	4.466e-02*	0.01541
NBBotherL9	3.393e-03*	1.561e-03	3.006e-03	1.697e-03	-2.752e-02*	0.00429	4.301e-05	0.00526
NBOotherL9	2.072e-03	1.266e-03	3.137e-03*	1.377e-03	-4.619e-02*	0.00348	-5.244e-03	0.00427
NBBlistL9	-4.350e-02*	4.932e-03	7.171e-02*	5.362e-03	4.337e-02*	0.01356	5.587e-02*	0.01663
NBOlistL9	9.795e-02*	4.541e-03	-3.601e-02*	4.937e-03	7.750e-02*	0.01248	1.484e-02	0.01531
NBBotherL10	6.825e-03*	1.561e-03	4.798e-03*	1.697e-03	-4.029e-03	0.00429	2.257e-03	0.00526
NBOotherL10	3.634e-03*	1.266e-03	8.238e-03*	1.376e-03	3.406e-03	0.00348	-3.890e-03	0.00427
NBBlistL10	-5.183e-02*	4.873e-03	1.447e-01*	5.297e-03	6.344e-02*	0.01340	6.346e-02*	0.01643
NBOlistL10	1.100e-01*	4.470e-03	-9.194e-02*	4.860e-03	6.964e-02*	0.01229	7.372e-02*	0.01507
a1	1.353e-04	1.064e-04	-1.851e-04	1.156e-04	1.335e-04*	0.0002	-2.634e-04	0.00035
a2	9.190e-06	3.904e-05	-4.198e-05	4.244e-05	4.962e-04*	0.00010	1.807e-04	0.00013
a3	-1.623e-05	4.581e-05	-8.578e-06	4.980e-05	5.687e-04*	0.00012	1.540e-04	0.00015

$^{*}p < 0.05$

Table 4: Primary VS non-primary exchanges in event-time using quotes.

	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
NBBotherL1	-0.240*	0.00489	-0.0405*	0.0107	0.137*	0.00471	0.0813*	0.00494
NBOotherL1	-0.006*	0.00221	-0.1348*	0.0048	0.027*	0.00215	0.0389*	0.00225
NBBlistL1	0.132*	0.00502	0.1418*	0.0109	-0.316*	0.00482	0.114*	0.00504
NBOlistL1	0.070*	0.00474	0.1809*	0.0104	0.114*	0.00463	-0.263*	0.00482
NBBotherL2	-0.017*	0.00509	-0.0358*	0.0112	0.098*	0.00494	0.0772*	0.00513
NBOotherL2	0.0140*	0.00224	0.0729*	0.0049	0.007*	0.00219	0.00948*	0.00227
NBBlistL2	0.0765*	0.00535	0.0946*	0.0119	-0.095*	0.00525	0.0871*	0.00549
NBOlistL2	0.0455*	0.00494	-0.0017	0.0110	0.101*	0.00486	-0.0726*	0.00504
NBBotherL3	-0.026*	0.00509	-0.1957*	0.0113	0.058*	0.00497	0.0558*	0.00522
NBOotherL3	0.0125*	0.00221	0.0389*	0.0049	-0.001	0.00214	0.00274	0.00223
NBBlistL3	0.0523*	0.00535	0.0675*	0.0120	-0.129*	0.00525	0.0707*	0.00545
NBOlistL3	0.0340*	0.00494	-0.03*	0.0111	0.075*	0.00487	-0.111*	0.00516
NBBotherL4	0.0281*	0.00507	-0.0972*	0.0113	0.057*	0.00505	0.0551*	0.00527
NBOotherL4	0.0057*	0.00223	0.0700*	0.0049	-0.002	0.00214	0.00193	0.00223
NBBlistL4	0.0420*	0.00532	0.0866*	0.0121	-0.035*	0.00533	0.0606*	0.00554
NBOlistL4	0.0328*	0.00494	-0.0169	0.0111	0.073*	0.00488	-0.0309*	0.00506
NBBotherL5	-0.0175*	0.00481	-0.1435*	0.0114	0.033*	0.00502	0.0421*	0.00527
NBOotherL5	0.0106*	0.00221	0.0521*	0.0049	-0.004*	0.00218	-0.00236	0.00228
NBBlistL5	0.0289*	0.00501	0.0791*	0.0121	-0.063*	0.00532	0.0409*	0.00558
NBOlistL5	0.0209*	0.00479	-0.0238*	0.0111	0.057*	0.00488	-0.0682*	0.00505
NBBotherL6	0.012*	0.005	-0.1081*	0.0114	0.029*	0.00501	0.0372*	0.00524
NBOotherL6	0.003	0.00227	0.0890*	0.0049	-0.005*	0.00217	-0.00349	0.00223
NBBlistL6	0.01	0.00502	0.0749*	0.0121	-0.046*	0.00532	0.0275*	0.00557
NBOlistL6	0.13*	0.0052	-0.02	0.0110	0.047*	0.00487	-0.0442*	0.00504
NBBotherL7	-0.02*	0.00509	-0.1124*	0.0113	0.014*	0.00501	0.0379*	0.00521
NBOotherL7	0.009*	0.00224	0.0689*	0.0049	-0.006*	0.00217	-0.00525*	0.00222
NBBlistL7	-0.0048*	0.00552	0.0680*	0.0120	-0.037*	0.00530	0.0285*	0.00553
NBOlistL7	0.0125*	0.00507	-0.0081	0.0110	0.039*	0.00486	-0.0371*	0.00508
NBBotherL8	-0.026*	0.00509	-0.1461*	0.0114	0.018*	0.00501	0.0310*	0.00524
NBOotherL8	0.01*	0.00225	0.0695*	0.0049	-0.007*	0.00216	-0.00816*	0.00225
NBBlistL8	0.0069	0.00545	0.0391*	0.0119	-0.022*	0.00523	0.00979	0.00546
NBOlistL8	0.0262*	0.00501	-0.0251*	0.0109	0.036*	0.00481	-0.0464*	0.00502
NBBotherL9	-0.023*	0.00533	-0.0818*	0.0116	0.022*	0.00511	0.0314*	0.00534
NBOotherL9	0.00789*	0.00224	0.0777*	0.0049	-0.005*	0.00215	-0.00899*	0.00225
NBBlistL9	0.0085	0.00536	0.0388*	0.0116	-0.026*	0.00514	0.0236*	0.00537
NBOlistL9	0.01*	0.00497	-0.0218*	0.0108	0.029*	0.00476	-0.0217*	0.00498
NBBotherL10	-0.019*	0.00507	0.0251*	0.0111	0.016*	0.00486	0.0308*	0.00508
NBOotherL10	0.0032	0.00223	0.0789*	0.0048	-0.007*	0.00213	-0.00573*	0.00222
NBBlistL10	0.0047	0.00503	0.0294*	0.0109	-0.004	0.00482	0.0191*	0.00504
NBOlistL10	0.012*	0.00479	-0.0294*	0.0104	0.016*	0.00457	-0.0151*	0.00477
a1	-0.0403*	0.00175	-0.0616*	0.0040	0.032*	0.00179	0.0110*	0.00187
a2	0.0173*	0.00111	0.1560*	0.0029	-0.019*	0.00128	-0.0233*	0.00133
a3	0.0434*	0.00173	0.0518*	0.0040	-0.029*	0.00177	-0.0150*	0.00185

$^{*}p < 0.05$

Table 5: Trades (Lit and Dark) VS Quotes (NBB and NBO) in event-time.

	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
Lit.L1	-0.168*	0.00496	0.0163*	0.00609	0.153*	0.00549	0.0701*	0.00480
DL1	0.0453*	0.00406	-0.187*	0.00499	-0.00434	0.00450	0.0131*	0.00393
NBBL1	0.0709*	0.00449	0.0190*	0.00552	-0.220*	0.00498	0.378*	0.00435
NBOL1	0.0930*	0.00512	0.00953	0.00628	0.124*	0.00566	-0.418*	0.00495
Lit.L2	0.0208*	0.00567	-0.0399*	0.00695	0.0541*	0.00627	0.132*	0.00548
DL2	0.0268*	0.00416	-0.0845*	0.00511	0.0627*	0.00461	0.000749	0.00403
NBBL2	0.121*	0.00501	0.00494	0.00614	-0.0733*	0.00554	0.155*	0.00485
NBOL2	0.0515*	0.00535	0.0263*	0.00656	-0.0868*	0.00592	0.000711	0.00517
Lit.L3	0.0340*	0.00552	-0.0427*	0.00677	0.0802*	0.00611	-0.0970*	0.00534
DL3	0.0151*	0.00418	-0.0553*	0.00513	0.0207*	0.00463	0.00916*	0.00405
NBBL3	0.00387	0.00422	0.00787	0.00518	-0.00762	0.00467	0.0494*	0.00408
NBOL3	-0.0856*	0.00520	0.0805*	0.00638	-0.240*	0.00576	0.0408*	0.00504
Lit.L4	0.0136*	0.00549	-0.055*	0.00674	-0.0225*	0.00608	0.0702*	0.00532
DL4	0.0156*	0.00419	-0.0288*	0.00514	-0.00419	0.00463	0.00278	0.00405
NBBL4	-0.0323*	0.00418	-0.0373*	0.00514	0.0211*	0.00463	-0.113*	0.00405
NBOL4	-0.0974*	0.00522	0.0508*	0.00640	0.128*	0.00577	0.116*	0.00505
Lit.L5	0.0763*	0.00541	-0.000814	0.00663	0.0567*	0.00598	0.103*	0.00523
DL5	0.00791	0.00416	-0.00586	0.00511	0.0114*	0.00461	-0.00212	0.00403
NBBL5	-0.0974*	0.00412	-0.0202*	0.00505	-0.166*	0.00456	-0.0569*	0.00399
NBOL5	-0.0285*	0.00522	0.0312*	0.00637	0.331*	0.00575	-0.0246*	0.00503
Lit.L6	0.0528*	0.00542	-0.0339*	0.00664	0.0905*	0.00599	-0.0271*	0.00524
DL6	0.0150*	0.00415	-0.00909	0.00509	0.0148*	0.00459	-0.0133*	0.00401
NBBL6	-0.0113*	0.00404	-0.0172*	0.00496	-0.058*	0.00447	-0.0307*	0.00391
NBOL6	-0.00398	0.00538	0.0547*	0.00660	-0.0542*	0.00596	-0.0872*	0.00521
Lit.L7	0.0469*	0.00535	-0.0642*	0.00656	0.0101	0.00592	0.0709*	0.00517
DL7	-0.00367	0.00412	-0.00601	0.00506	0.0149*	0.00456	-0.00328	0.00399
NBBL7	-0.0802*	0.00404	-0.0342*	0.00495	-0.0988*	0.00447	-0.121*	0.00391
NBOL7	0.00589	0.00531	0.0377*	0.00651	0.220*	0.00587	0.151*	0.00514
Lit.L8	0.0595*	0.00524	-0.0273*	0.00643	-0.0180*	0.00580	-0.0114*	0.00507
DL8	0.00667	0.00409	-0.00498	0.00501	0.000959	0.00452	0.0249*	0.00396
NBBL8	0.0179*	0.00405	0.00849	0.00497	-0.104*	0.00449	-0.0608*	0.00392
NBOL8	0.103*	0.00538	-0.0165*	0.00660	0.0205*	0.00595	-0.0925*	0.00521
Lit.L9	0.0807*	0.00502	0.0171*	0.00617	0.0273*	0.00556	0.0297*	0.00486
DL9	0.00664	0.00405	0.000682	0.00497	0.0105*	0.00448	-0.0212*	0.00392
NBBL9	-0.0894*	0.00399	0.0258*	0.00489	-0.254*	0.00441	-0.151*	0.00386
NBOL9	0.0336*	0.00523	-0.0304*	0.00641	-0.0367*	0.00579	0.0238*	0.00506
Lit.L10	0.0615*	0.00490	0.138*	0.00601	-0.0168*	0.00542	0.0313*	0.00474
DL10	0.00083	0.00401	0.000297	0.00491	0.0144*	0.00443	0.00374	0.00388
NBBL10	-0.127*	0.00395	0.0307*	0.00485	-0.123*	0.00437	-0.133	* 0.0038280
NBOL10	-0.0152*	0.00488	-0.0543*	0.00599	0.0286*	0.00541	-0.00417	0.00473
a1	-0.0644*	0.00196	0.00561*	0.00241	0.0501*	0.00217	0.0166*	0.00190
a2	-0.0047*	0.00199	0.0889*	0.00244	-0.03*	0.00220	-0.0102*	0.00193
a3	0.139*	0.00373	-0.0797*	0.00458	-0.0717*	0.00413	-0.0606*	0.00361

$^{*}p < 0.05$

Table 6: Trades (Lit and Dark) VS Quotes (NBB and NBO) in natural-time (1-sec).

	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
Lit.L1	1.437e-04	.657e-03	4.042e-05	2.194e-03	1.022e-05	3.938e-03	2.472e-05	2.738e-03
DL1	6.543e-04	1.734e-03	1.236e-03	4.145e-03	2.256e-03	4.275e-03	1.466e-03	1.879e-03
NBBL1	5.341e-03	4.788e-03	4.221e-02*	6.351e-03	6.721e-02*	2.311e-03	3.921e-02	4.314e-03
NBOL1	4.342e-03	5.576e-03	-7.041e-03	7.683e-03	-4.021e-03	2.612e-03	1.127e-03*	5.315e-03
Lit.L2	1.243e-04	1.752e-03	1.254e-04	2.392e-03	1.254e-04	5.363e-03	2.356e-04	4.123e-03
DL2	2.011e-02*	1.411e-03	5.323e-04	1.295e-03	3.321e-04	1.465e-03	3.991e-04	2.368e-03
NBBL2	1.536e-02*	3.256e-03	2.294e-02*	2.432e-03	1.235e-02*	3.231e-03	1.812e-02*	2.251e-03
NBOL2	2.758e-02*	5.735e-03	2.415e-02*	5.128e-03	3.766e-02*	3.122e-03	2.896e-03	2.852e-03
Lit.L3	1.435e-03	2.545e-03	4.135e-03	5.696e-03	4.726e-03	7.468e-03	1.714e-03	4.256e-03
DL3	3.534e-04	5.324e-03	6.124e-04	2.416e-03	5.118e-04	1.444e-03	5.118e-04	1.444e-03
NBBL3	5.342e-02*	5.641e-03	2.970e-02	5.369e-03	2.995e-03	2.311e-03	2.111e-03	3.115e-03
NBOL3	-3.903e-03	5.340e-03	-2.502e-02*	3.979e-03	-1.534e-02*	4.877e-03	4.334e-02*	2.337e-03
Lit.L4	2.634e-03	1.743e-03	-4.223e-04	1.246e-03	4.223e-04	7.252e-03	1.223e-04	2.255e-03
DL4	1.132e-03	1.659e-03	5.031e-04	1.285e-03	2.027e-04	2.145e-03	3.487e-04	3.169e-03
NBBL4	-3.478e-02*	4.342e-03	3.752e-02*	7.943e-03	3.752e-02*	7.248e-03	2.232e-02*	4.253e-03
NBOL4	5.757e-03	5.673e-03	-3.612e-02*	4.986e-03	1.312e-02*	5.216e-03	1.312e-02*	2.216e-03
Lit.L5	8.722e-04	1.395e-03	1.241e-03	1.612e-03	-2.461e-03	7.362e-03	5.128e-03	1.442e-03
DL5	5.380e-04	1.270e-03	6.277e-04	1.343e-03	1.578e-04	1.343e-03	2.836e-04	1.974e-03
NBBL5	3.257e-03	3.369e-03	3.529e-02*	5.391e-03	2.193e-02*	1.121e-03	1.287e-03	2.321e-03
NBOL5	1.410e-03	1.577e-03	-1.231e-02*	2.924e-03	1.283e-02*	2.857e-03	3.532e-03	5.321e-03
Lit.L6	1.641e-03	1.231e-03	3.446e-04	1.245e-03	3.446e-04	2.441e-03	3.499e-03	2.541e-03
DL6	-7.458e-04	2.649e-03	7.288e-04	1.432e-03	5.856e-04	3.352e-03	5.856e-04	3.652e-03
NBBL6	4.458e-02*	3.249e-03	4.425e-02*	3.355e-03	2.448e-02*	2.255e-03	2.448e-02*	2.735e-03
NBOL6	2.131e-02	3.543e-03	4.542e-02*	4.123e-03	4.542e-03	4.123e-03	3.943e-03	3.121e-03
Lit.L7	2.265e-03	2.598e-03	1.348e-04	1.762e-03	-1.877e-04	2.741e-03	3.117e-04	1.241e-03
DL7	-3.791e-04	1.774e-03	-6.131e-04	2.678e-03	3.742e-04	4.621e-03	1.242e-04	2.731e-03
NBBL7	2.548e-03	2.213e-03	3.741e-03	5.316e-03	1.511e-03	2.834e-03	2.238e-03	3.887e-03
NBOL7	5.179e-02*	2.541e-03	1.148e-03	4.982e-03	1.148e-03	2.927e-03	2.147e-03	1.267e-03
Lit.L8	3.331e-03	1.944e-03	4.235e-03	2.548e-03	5.336e-03	5.851e-03	4.341e-03	4.723e-03
DL8	1.293e-03	3.242e-03	-1.753e-03	1.432e-03	-1.633e-03	3.212e-03	4.633e-04	2.743e-03
NBBL8	-2.187e-02	4.948e-03	3.021e-03	5.129e-03	7.017e-03	4.189e-03	1.615e-02*	6.719e-03
NBOL8	2.216e-02*	3.572e-03	-2.672e-03	2.135e-03	-2.654e-03	2.234e-03	1.754e-03	1.633e-03
Lit.L9	6.393e-03	7.361e-03	1.906e-03	1.958e-03	1.346e-03	2.158e-03	2.948e-03	3.278e-03
DL9	3.221e-03	3.346e-03	1.837e-03	1.398e-03	1.292e-03	1.688e-03	1.292e-03	1.688e-03
NBBL9	5.261e-02*	2.288e-03	5.241e-03	6.464e-03	4.245e-03	6.934e-03	4.235e-03	4.738e-03
NBOL9	9.795e-02*	4.541e-03	3.211e-03	3.987e-03	3.216e-03	4.517e-03	2.234e-03	7.313e-03
Lit.L10	5.135e-03	6.231e-03	5.798e-03*	2.498e-03	4.741e-03*	1.192e-03	5.741e-02*	6.298e-03
DL10	1.3534e-03	1.476e-03	2.211e-03	5.126e-03	1.912e-03	1.1367e-03	1.114e-03	2.444e-03
NBBL10	1.199e-02*	5.673e-03	3.645e-05	5.198e-03	3.645e-05	4.931e-03	1.236e-05	3.951e-03
NBOL10	4.175e-04	2.430e-03	-4.131e-02	4.860e-03	-2.131e-02	4.860e-03	3.276e-02	2.410e-03
a1	3.322e-04*	1.025e-04	-2.151e-04	1.156e-04	-2.151e-03*	1.156e-04	2.151e-03*	2.287e-04
a2	9.190e-06*	3.904e-06	-3.283e-05	4.244e-05	1.283e-04*	4.244e-05	2.227e-05	4.211e-05
a3	-3.265e-05	6.534e-06	7.174e-06	6.642e-05	3.572e-06	4.421e-05	5.489e-04*	3.771e-05

$^*p < 0.05$

Table 7: Participants VS SIP timestamps in natural-time (1-sec).

	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
NBBpartL1	-0.163	0.161	-0.0236	0.181	0.0651	0.161	-0.0137	0.181
NBOpartL1	-0.123	0.0989	-0.189	0.110	-0.129	0.0986	0.0275	0.110
NBBsipL1	0.171	0.161	0.0369	0.181	-0.0565	0.161	0.0271	0.181
NBOsipL1	0.135	0.0988	0.196	0.110	0.141	0.0988	-0.0213	0.110
NBBpartL2	-0.131	0.177	0.0654	0.198	-0.0754	0.177	0.0827	0.199
NBOpartL2	-0.0835	0.109	0.0468	0.123	-0.0822	0.109	0.0232	0.123
NBBsipL2	0.111	0.177	-0.0577	0.198	0.0554	0.177	-0.0749	0.199
NBOsipL2	0.0534	0.109	-0.0427	0.123	0.0516	0.109	-0.0190	0.123
NBBpartL3	-0.0982	0.176	0.0578	0.197	-0.0390	0.176	0.0542	0.198
NBOpartL3	0.0482	0.108	0.103	0.122	0.0505	0.108	0.0655	0.122
NBBsipL3	0.101	0.176	-0.0556	0.197	0.0420	0.176	-0.0519	0.198
NBOsipL3	-0.0427	0.108	-0.0972	0.121	-0.0446	0.108	-0.0591	0.122
NBBpartL4	0.111	0.175	0.0542	0.196	0.143	0.175	0.0572	0.196
NBOpartL4	0.0462	0.107	0.0590	0.120	0.0558	0.107	0.0317	0.120
NBBsipL4	-0.108	0.175	-0.0511	0.196	-0.140	0.175	-0.0540	0.196
NBOsipL4	-0.0422	0.107	-0.0555	0.120	-0.0517	0.107	-0.0281	0.120
NBBpartL5	-0.107	0.174	-0.189	0.194	-0.0923	0.174	-0.176	0.195
NBOpartL5	-0.0249	0.106	-0.153	0.119	-0.0263	0.106	-0.101	0.119
NBBsipL5	0.115	0.173	0.199	0.194	0.100	0.174	0.187	0.195
NBOsipL5	-0.0155	0.106	0.0793	0.119	-0.0141	0.106	0.0272	0.119
NBBpartL6	-0.0601	0.172	-0.0665	0.193	-0.0337	0.172	-0.0691	0.193
NBOpartL6	-0.0318	0.104	-0.0341	0.116	-0.0368	0.104	-0.0583	0.117
NBBsipL6	0.0616	0.173	0.0700	0.193	0.0353	0.172	0.0727	0.193
NBOsipL6	0.0354	0.104	0.0361	0.116	0.0404	0.104	0.0604	0.116
NBBpartL7	-0.00157	0.170	-0.0555	0.191	0.00923	0.170	-0.0650	0.191
NBOpartL7	-0.0732	0.103	-0.0142	0.115	-0.0769	0.103	-0.0411	0.115
NBBsipL7	-0.000990	0.170	0.0577	0.191	-0.0118	0.170	0.0671	0.191
NBOsipL7	0.114	0.103	0.00996	0.115	0.118	0.103	0.0370	0.115
NBBpartL8	-0.0686	0.168	-0.0441	0.188	-0.0680	0.168	-0.0428	0.188
NBOpartL8	0.0554	0.102	-0.137	0.114	0.0574	0.102	-0.169	0.114
NBBsipL8	0.0688	0.168	0.0463	0.188	0.0682	0.168	0.0450	0.188
NBOsipL8	-0.0526	0.102	0.138	0.114	-0.0545	0.102	0.170	0.114
NBBpartL9	0.0134	0.165	0.0119	0.185	0.0146	0.165	0.0149	0.185
NBOpartL9	-0.0513	0.101	-0.0687	0.113	-0.0533	0.101	-0.0900	0.113
NBBsipL9	-0.0125	0.165	-0.00987	0.185	-0.0136	0.165	-0.0129	0.185
NBOsipL9	0.0542	0.101	0.0680	0.113	0.0563	0.101	0.0893	0.113
NBBpartL10	-0.0704	0.161	-0.0389	0.181	-0.0750	0.161	-0.0440	0.181
NBOpartL10	0.000351	0.0989	0.00211	0.110	-0.00763	0.0989	-0.0203	0.110
NBBsipL10	0.0715	0.161	.0405	0.180	0.0761	0.161	0.0459	0.181
NBOsipL10	-0.00350	0.0988	-0.00721	0.110	0.00455	0.0989	0.0152	0.110
a1	-0.0207	0.0775	-0.1266	0.0868	-0.0153	0.0775	0.0555	0.0869
a2	-0.0621	0.108	-0.00182	0.121	0.0604	0.108	-0.00220	0.122
a3	0.0207	0.0775	0.126	0.0868	0.0153	0.0775	-0.0555	0.0869

$^{*}p < 0.05$

Table 8: Participants VS SIP timestamps in event-time .

	coef.	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**	**coef.**	**s.e.**
NBBpartL1	-0.0664*	0.00495	0.0270*	0.00394	0.898*	0.00192	0.00584	0.00273
NBOpartL1	0.0256*	0.00653	-0.147*	0.00520	-0.0342*	0.00254	0.488*	0.00360
NBBsipL1	-0.136*	0.0129	0.0895*	0.0103	-0.694*	0.00505	0.125*	0.00714
NBOsipL1	0.0213*	0.00946	0.288*	0.00753	0.0682*	0.00368	-0.740*	0.00521
NBBpartL2	-0.290*	0.0220	-0.0218	0.0175	0.0542*	0.00860	0.0668*	0.0121
NBOpartL2	0.121*	0.0148	-0.161*	0.0118	0.0828*	0.00578	-0.0930*	0.00818
NBBsipL2	0.225*	0.0206	0.0995*	0.0164	-0.0100*	0.00803	-0.0283	0.0113
NBOsipL2	-0.0642*	0.0119	0.149*	0.00954	-0.0366*	0.00467	-0.0823*	0.00660
NBBpartL3	-0.265*	0.0208	-0.160*	0.0165	-0.00294	0.00810	0.0275	0.0114
NBOpartL3	0.180*	0.0128	-0.115*	0.0102	0.0505*	0.00500	0.1701*	0.00707
NBBsipL3	0.150*	0.0192	0.106*	0.0153	-0.0404*	0.00749	0.0879*	0.0106
NBOsipL3	0.106*	0.0102	0.0834	0.00817	-0.0276*	0.00399	0.0926*	0.00565
NBBpartL4	-0.159*	0.0194	-0.122*	0.0155	0.0394*	0.00758	0.0185	0.0107
NBOpartL4	0.225*	0.0114	-0.0924*	0.00910	0.0212*	0.00445	-0.194*	0.00630
NBBsipL4	0.119*	0.0172	0.163*	0.0137	-0.0287*	0.00672	0.0345*	0.00951
NBOsipL4	-0.0309*	0.00981	-0.185*	0.00782	0.00825	0.00382	-0.0720*	0.00541
NBBpartL5	-0.181*	0.0177	-0.0764*	0.0141	0.0194*	0.00690	-0.00956	0.00976
NBOpartL5	0.0286*	0.0106	0.138*	0.00848	0.0197*	0.00415	0.0455*	0.00587
NBBsipL5	0.230*	0.0161	0.101*	0.0128	-0.0412*	0.00628	0.121*	0.00889
NBOsipL5	-0.116*	0.00937	0.108*	0.00746	-0.00764	0.00365	0.0209*	0.00516
NBBpartL6	-0.274*	0.0166	-0.193*	0.0132	0.0351*	0.00648	-0.127*	0.00916
NBOpartL6	-0.0054	0.0102	-0.00701	0.00818	0.0139*	0.00400	0.00524	0.00566
NBBsipL6	0.175*	0.0143	0.00975	0.0113	0.00381	0.00557	-0.00535	0.00788
NBOsipL6	-0.00723	0.00747	-0.0801*	0.00595	0.00857*	0.00291	-0.0450*	0.00412
NBBpartL7	-0.196*	0.0150	0.00729	0.0119	-0.009088	0.00584	-0.0291*	0.00827
NBOpartL7	0.0824*	0.00900	-0.0518*	0.00717	0.00439	0.00350	-0.0370*	0.00496
NBBsipL7	0.133*	0.0136	-0.0120	0.0108	0.00218	0.00531	-0.0393*	0.00751
NBOsipL7	-0.0307*	0.00644	0.0718*	0.00513	-0.0101*	0.00251	0.0391*	0.00355
NBBpartL8	-0.409*	0.0143	-0.111*	0.0114	-0.00439	0.00560	-0.01377*	0.00792
NBOpartL8	-0.0261*	0.00780	-0.0255*	0.00621	-0.00534	0.00304	0.0130*	0.00430
NBBsipL8	-0.0284*	0.0114	0.0199	0.00914	0.00631	0.00447	0.0149*	0.00632
NBOsipL8	-0.145*	0.00462	0.0181*	0.00367	-0.000240	0.00180	0.0111*	0.00254
NBBpartL9	-0.0295*	0.0121	-0.00468	0.00966	-0.0411*	0.00472	-0.0150*	0.00668
NBOpartL9	0.133*	0.00662	0.0744*	0.00527	0.000521	0.00258	0.0527*	0.003650
NBBsipL9	-0.194*	0.00824	0.0455*	0.00656	-0.0131*	0.0032	0.0305*	0.00454
NBOsipL9	0.120*	0.00384	0.0166*	0.00306	-0.0199*	0.00149	0.0280*	0.00212
NBBpartL10	-0.0938*	0.00902	-0.165*	0.00718	0.0163*	0.00351	-0.0543*	0.00497
NBOpartL10	0.204*	0.00605	-0.0343*	0.00482	0.00888*	0.00236	-0.0253*	0.00333
NBBsipL10	-0.0328*	0.00432	0.0901*	0.00344	-0.0297*	0.00168	0.0685*	0.00238
NBOsipL10	-0.0357*	0.00321	-0.000353	0.00256	0.0174*	0.00125	-0.00813*	0.00177
a1	0.0277*	0.0170	-0.108*	0.0135	-0.152*	0.00662	0.859*	0.00937
a2	0.211*	0.0229	0.149*	0.0182	0.643*	0.00892	-0.165*	0.0126
a3	-0.0576*	0.0169	0.119*	0.0135	0.147*	0.00661	-0.851*	0.00935

$^{*}p < 0.05$