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

A non-Normal framework for price discovery: The independent component based information shares measure

Sebastiano Michele Zema [a]

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

2023/03 January 2023
ISSN(ONLINE) 2284-0400

A non-Normal framework for price discovery: The independent component based information shares measure

Sebastiano Michele Zema *

Institute of Economics, Scuola Superiore Sant'Anna

Piazza Martiri della Libertà 56127, Pisa

sebastianomichele.zema@santannapisa.it

January 9, 2023

Abstract

I propose a new measure of price discovery, which I will refer to as the *Independent Component based Information Share* (IC-IS). This measure constitutes a variant of the widespread Information Share, with the main difference being it does not suffer the same identification issues. Under the assumptions of non-normality and independence of the shocks, a rather general theoretical framework leading to the estimation of the IC-IS is illustrated. After testing the robustness of the proposed measures to different non-Normal distributions in a simulated environment, an empirical exercise encompassing different price discovery applications will follow.

Keywords: vector error correction models (VECMs); information shares; market microstructure; independent component analysis; pseudo maximum likelihood; price discovery

JEL classification: C32, C58, G14.

*The author acknowledges the support granted by the PRO3 program "Network Analysis of Economic and Financial Resilience".

The quantification of the contribution of agents and exchanges to the price formation process acquired increasing importance in the literature. Processes of market fragmentation, carried out together with the proliferation of algorithmic trading strategies and the introduction of complex financial products, dramatically increased the complexity of financial markets and made the possibility to measure their informativeness a concrete challenge in the financial environment. In this respect, the information share (IS) measure of Hasbrouck (1995) represents a milestone in the literature, being one of the most widely adopted measures for price discovery as documented by its large 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).

From a market microstructure modeling perspective, the IS build its fundamentals upon the modeling of price changes through vector error correction models (VECM). The main shortfall of the IS measure is it can be uniquely determined only when the VECM residuals are not contemporaneously correlated. Hasbrouck's suggested solution was, in absence of a sound financial theory suggesting appropriate causal relationships, to identify the model by performing the Choleski decomposition on the residuals for all the possible permutations of the variables, which leads to upper and lower bounds for the IS. In empirical applications upper and lower bounds are often very wide because of substantial cross-correlations in the model residuals, raising interpretative ambiguities about the real allocation of information between the analyzed variables.

In this paper I propose a solution by defining a variant of the widespread IS measure, which I will refer to as the *Independent Component based Information Shares* (IC-IS), and for which the identification issues related to the well-established IS are tremendously alleviated. The proposed measure build its fundamentals on the exploitation of non-normality for the estimation and identification of the contemporaneous coefficient matrix through which the shocks reveal in the market. The estimation procedure is implemented adopting the pseudo maximum likelihood (PML) approach of Gouriéroux et al. (2017). Being the newly introduced IC-IS a precise point estimate for the contribution of each variable to price discovery, a simple testing framework for these contributions is also illustrated exploiting the asymptotic properties of the PML estimates.

The article is organized as follows. Section 1 briefly reviews the state of the art, recalling the most recent updates on the topic and the need for further progresses. Section 2 set up the general market microstructure framework on which the proposed methodology will be based on. Section 3 introduces the new price discovery measure, with theory, estimation, and simulations details. Section 4 shows an empirical application on IBM trading data, implementing a variety of price discovery applications. Section 5 concludes.

1 State of the art

The present work is not the first one dealing with such a long-standing issue. Even if several attempts have been made to solve the identification problem associated to the IS measure, a general strategy is not available yet. The idea of estimating unique IS measures by exploiting the distributional properties of the variables was firstly introduced by Grammig and Peter (2013). The authors, inspired by Rigobon (2003), introduced different volatility regimes to identify the IS. The intuition was that the occurrence of extreme price changes causes differences between tail and center correlations, information which can be exploited to reach full identification of the model. Subject to the condition of observing different volatility regimes in the market, which might not always be the case, the above mentioned solution is effective.

A solution to the problem of obtaining unique information share measures can be found also in Lien and Shrestha (2009) and Fernandes and Scherrer (2018). Both authors, handled the problem by computing the IS on the spectral decomposition of either the correlation or the covariance matrix of the innovations. Even if these approaches are effective in getting unique measures, the problem at the origin of the impossibility to obtain a precise quantification of the IS is the lack of an identification procedure commonly accepted. Computing the IS on the factor structure associated to the spectral decomposition of the covariance(correlation) matrix does not provide a solution to the identification issue which constitutes the real problem.

From a more recent data-driven perspective Hasbrouck (2021) proposed to exploit the high-frequency at which quotes and trades occur, modeling in natural time to drastically reduce the range obtained by permuting the variables. The idea is that sampling prices at incredibly short time scales, even at micro or nanoseconds precision, inevitably and drastically reduce the presence of contemporaneous cross-correlations (see also Dias et al., 2021), which consequently leads to narrower IS bounds and discards any possible interpretative ambiguity. Still, modeling in this natural time framework requires to estimate an enormous amount of coefficients. The author handled the problem adopting the heterogeneous autoregressive approach (HAR) proposed by Corsi (2009). Nevertheless, this modeling approach 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, 2021; Buccheri et al., 2021; de Jong, 2021; Ghysels, 2021).

The most recent contribution which tried to provide a solution to the identification problem of the IS, by fixing the permutation indeterminacy of the variables in the model, can be found in Zema (2022). The author proposed the adoption of a causal discovery model, well-established in the machine learning literature, which exploits the non-Normal distributions of the variables to recover the directed acyclic graph (DAG) structure which is more likely to be true given the data. Given the obtained DAG, the associated causal chain was finally used to pick the corresponding permutation of the variable and compute the associated and unique IS measure. However, this approach works if and only if the assumption of the existence of a recursive causal structure in the system holds true.

In this respect, the present work tries to make a step forward in the literature by providing a rather generalized and practical framework for the estimation of unique market information shares when the shocks are non-normally distributed. This will lead to the introduction of the previously mentioned IC-IS, for which is not necessary to assume the presence of either different volatility regimes or causal chains in the system. Moreover, the measure has been found to provide consistent results, even if in a limited sample, with no need to increase the model and computational complexities introduced when working at incredibly high resolutions in natural time.

While the scope of this work is to provide a solution to a long-standing issue in the context of price discovery through the IS measures of Hasbrouck (1995), it should be noted that a variety of other measures and approaches have been proposed in the literature for price discovery (Harris et al., 1995; Booth et al., 1999; De Jong and Schotman, 2010; Putniņš, 2013, see for instance). For the readers interested in having a general overview, comprehensive reviews of different price discovery measures and how they relate with each other can be found in Baillie et al. (2002), Lehmann (2002), and Yan and Zivot (2010).

2 Measuring price discovery: The general framework

The general market microstructure setting is the one of Hasbrouck (1995). Let $p_t = \{p_{1t}, p_{2t}, \dots , p_{nt}\}$ be a vector of time series log-prices observed in n different exchanges but pertaining the same security. Being the time-series arbitrage linked, their dynamic can be modeled by the vector error correction model (VECM) of Engle and Granger (1987), specified as

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

with $\beta \in \mathbb{R}^{n\times n-1}$ containing the $n-1$ cointegrating vectors $p_1 - p_2$, $p_1 - p_3$, $p_1 - p_n$ and $\alpha \in \mathbb{R}^{n\times n-1}$ being a matrix of loadings. The system in equation 1 is covariance stationary, with $\mathrm{Cov}(u_t) = \Omega$, and admits the common trend representation

$$p_t = p_0 + \Psi(1)\sum_{i=1}^{t} \epsilon_i + \Psi^*(L)u_t \tag{2}$$

where $\Psi(L) = \Psi(1) + (1-L)\Psi^*(L)$ holds and the matrix $\Psi(1)$ can be computed as (Johansen, 1991):

$$\Psi(1) = \beta_\perp \left[\alpha'_\perp \left(I - \sum_{i=1}^{k} \Phi_i\right)\beta_\perp\right]^{-1} \alpha'_\perp. \tag{3}$$

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 $IS_j = \psi_j^2\Omega_{jj}/\psi\Omega\psi'$, with ψ being the common row of $\Psi(1)$ and ψ_j denoting the *j-th* element of ψ corresponding to market j. In many empirical applications Ω is non-diagonal and the information shares are not identified. A practical solution widely adopted in the literature is to consider the Choleski decomposition $\Omega = FF'$ and compute

$$IS_j = \frac{\left([\psi F]_j\right)^2}{\psi\Omega\psi'} \tag{4}$$

for each possible permutation of the variables in the model so to get upper and lower bounds for each IS. Zema (2022) proposed to identify the IS measure by means of a causal search algorithm which exploits the non-Normal distribution of the variables to pick a specific order and performing Choleski accordingly. Still, the proposed solution works only when the assumption of a causal recursive structure among the variables is not violated. In the next section a generalized framework to identify and test the IS measure in a non-Normal setting will be introduced, without imposing any recursive causal structure in the system (i.e., lower triangular matrix of the contemporaneous coefficients).

3 Non-Normal identification: The independent component based information shares

Let consider the n-dimensional vector of price innovations $u_t = [u_{1t}, u_{2t}, ..., u_{nt}]$, with non-diagonal covariance matrix Ω, to be a linear combination of n unobserved shocks $\epsilon_t = [\epsilon_{1t}, \epsilon_{2t}, ..., \epsilon_{nt}]$:

$$u_t = C\epsilon_t \tag{5}$$

Where $C \in \mathbb{R}^{n\times n}$ is an invertible mixing matrix through which the unobserved shocks ϵ_t are revealed in each market. If ϵ is normally distributed, the knowledge of u makes CC' identifiable but C itself cannot be identified. For any non-singular matrix Q, the matrix C and the shocks ϵ_t could be replaced respectively by $C^* = CQ$ and $\epsilon_t^* = Q^{-1}\epsilon_t$ leading to an observationally equivalent model. However, when ϵ_t is not Normal the identification problem almost disappears and C can be identified under few conditions. This follows from well established results (see Comon, 1994; Eriksson and Koivunen, 2004) which lead to the following theorem

Theorem 3.1. *Let $u_t = C\epsilon_t$ and the following conditions hold true:*

i The latent shocks $\epsilon_1, \epsilon_2, ..., \epsilon_n$ are mutually independent: $p(\epsilon_1, \epsilon_2, ..., \epsilon_n) = \prod_i^n p(\epsilon_i)$.

ii The sequence $\epsilon_1, \epsilon_2, ..., \epsilon_n$ contains at most one Normal distribution.

then, C can be identified up to column permutation, sign and scaling.

This brings important implications in our price discovery framework. If the conditions in Theorem 3.1 are satisfied it is possible to introduce the following new information share measure, which I will refer to as the *Independent Component based Information Shares* (IC-IS), for which the identification problem is strongly alleviated:

$$\text{IC-IS}_j = \frac{\left([\psi C^{(p)}]_j\right)^2}{\psi(C^{(p)}C^{(p)\prime})\psi'}. \tag{6}$$

Where $C^{(p)}$ is matrix C after a specific permutation of its columns has been picked, that is $C^{(p)} = CP$. The scaling indeterminacy (i.e., local lack of identification) is easily removed by imposing C to be an orthogonal matrix and pre-whitening the price innovations u_t (Hyvärinen and Oja, 1998, 2000; Moneta et al., 2013; Gouriéroux et al., 2017). Permutation and change in signs of the columns in C (i.e., global lack of identification) imply that, once C is estimated, the order in which the shocks are returned and the signs of their impact are unknown. However, the sign indeterminacy in the columns of C is totally irrelevant in the context of price discovery through the IS measure. Being the newly defined IC-IS still a ratio between two quadratic forms, the sign of the columns of C does not affect the result of the variance allocation mechanism for the efficient price process.

The only remaining cause of lack of identification is the column permutation indeterminacy which leads to the following proposition

Proposition 3.1. *Let P be a permutation matrix such that the matrix $C^{(p)} = CP$ satisfies $|c_{ii}| \geq |c_{ij}| \ \forall \ i \neq j$ and assume the conditions stated in Theorem 3.1 are met. Then, under the following conditions:*

i $E(\epsilon_t) = 0$ *and* $V(\epsilon_t) = Id$

ii C *is orthogonal,*

the IC-IS measures defined in equation 6 are uniquely identified and invariant to arbitrary permutations of the variable in the model.

Proof. See Appendix A. □

Proposition 3.1 ensures the uniqueness of the IC-IS measure by fixing, under a set of conditions, a specific permutation for the columns of C. Assuming that permutation to be true is not such a strong economic imposition: it implies each price series reacts to its own shock more than what other price series do. Then what remains is to obtain an estimate of C and related properties of the IC-IS measure, aspects which will be covered in the remainder of this section[1].

It is worth noticing the flexibility of such approach. When the permutation just mentioned is not plausible from an economic standpoint, alternative permutation strategies for the columns of C can be implemented accordingly. This could be the case, for instance, of price discovery trough derivative instruments (Blanco et al., 2005; Guidolin et al., 2021; Ahn et al., 2019). Imagine we want to quantify the contribution to the price formation process of leveraged exchange traded funds (ETFs). Leveraged ETFs are synthetic products which do not hold the underlying assets taken as benchmarks for their investment strategy. Still, these products replicate the benchmark's dynamic amplifying its return by opening derivative positions on the underlying (Leung et al., 2017; Shum et al., 2016). The spillover effect on these leveraged ETFs, originating from a shock on the underlying assets, would have an higher magnitude than the shock itself. As a consequence, considering for simplicity a system of two variables only, the permutation matrix P might be such that $C^{(p)} = CP$ satisfies the condition $c_{11}^{(p)} < c_{21}^{(p)}$, where 1 = underlying and 2 = leveraged ETF.

[1]The interested readers might refer to Moneta et al. (2013) and Zema (2022) for more intuitive explanations, and graphical representations, about the advantages of exploiting the non-normality assumption for identification purposes. Comprehensive and rigorous explanations given also by Gouriéroux et al. (2020).

3.1 Estimation and testing

The framework is the pseudo maximum likelihood approach of Gouriéroux et al. (2017). Let consider a set of unknown p.d.f. $g_i(\epsilon_i)$, where $i = 1, \dots, n$, and consider the pseudo log-likelihood function

$$\ln \mathcal{L}_T(C) = \sum_{t=1}^{T} \sum_{i=1}^{n} \ln g_i(c_i^{'} Y_t) \tag{7}$$

where $c_i^{'}$ is the i-th row of the orthogonal matrix C, and $c_i^{'} Y_t = \epsilon_i$. Then, the problem to be solved is

$$\begin{aligned} \hat{C} &= \underset{C}{\operatorname{argmax}} \ \ln \mathcal{L}_T(C) \\ &\text{s.t. } C'C = Id. \end{aligned} \tag{8}$$

The problem (8) is a typical constrained optimization where the constraints consist in the orthogonality condition for C, that is $c_i^{'} c_j = 0$ for $i < j$, and $c_i^{'} c_i = 1 \ \forall \ i$. The first order conditions (FOCs) then can be written as

$$\begin{cases} \sum_{t}^{T} Y_t \dfrac{\partial \ln g_i(\hat{c}_i^{'} Y_t)}{\partial \epsilon_i} - \sum_{j=1}^{n} \hat{\lambda}_{ij} \hat{c}_j = 0, \quad \forall \ i, \\ \hat{c}_i^{'} \hat{c}_i = 1, \quad \forall i \\ \hat{c}_i^{'} \hat{c}_j = 0, \quad i < j. \end{cases} \tag{9}$$

The asymptotic properties of the estimates resulting from the system of equations (9) have been proven under a set of necessary assumptions by Gouriéroux et al. (2017) and summarized in Lemma 3.1[2]

Lemma 3.1. *Suppose the conditions in Theorem 3.1 and Proposition 3.1 hold true. Then, under the following assumptions:*

- *i The functions $\ln g_i$ are twice continuously differentiable.*
- *ii Uniform integrability: $\sup_C |\sum_i^n \ln g_i(c_i^{'} Y_i^{'}) \leq m(Y_i^{'})|$ where $E[m(Y)] < \infty$.*

the PML estimator $\hat{C}_T$ of the true parameter C_0 is asymptotically normal with speed of convergence $1/\sqrt{T}$. That is, being the covariance matrix $\Sigma_C = 1/(E[(\nabla \ln \mathcal{L}(\mathcal{C})|_{C=C_0})^2]$ the reciprocal of the Fisher information matrix, $vec\sqrt{T}(\hat{C}_T - C_0) \sim N(0, \Sigma_C)$.

The assumptions in Proposition 3.1(i)-(ii) consists of regularity conditions needed to guarantee the convergence of the finite-sample estimates to the asymptotic ones as $T \to \infty$. It should be remarked that other estimation strategies, different from the PML approach, could be implemented to get an estimate of the mixing matrix C, computing the IC-IS consequently. In principle, one could opt for a relatively simpler but less general framework as in Lanne et al. (2017), where the non-Normal distributions are assumed to be known (i.e.,

[2]For the sake of clarity and readability, I refer the interested readers directly to Gouriéroux et al. (2017) for the proof of the asymptotic properties of the PML estimator stated in the Lemma.

Student in practice) in a standard maximum likelihood estimation (MLE) approach under the same regularity conditions in 3.1(i)-(ii). Another very popular estimator of C, which is worth mentioning, is the FastICA estimator (Hyvärinen and Oja, 2000) for which the asymptotic properties have been proven by Reyhani et al. (2012) under similar regularity conditions.

The IC-IS measures previously defined do not require the knowledge of a specific non-Normal distribution. For this reason, the PML strategy is appealing since it allows to get an estimate for C when the true probability density functions (p.d.f.) of the unobserved shocks ϵ_t are unknown. A comparison between different estimation strategies is not the objective of this work[3]. Still, when appropriate non-Normal distributions are chosen, given the data of interest, the IC-IS should be robust to the miss-specifications in the pseudo log-likelihood in equation (7). This will be tested both in a simulated environment and empirically.

Given the asymptotic distribution for C, a precise point-estimate of the IC-IS can be obtained and typical testing procedures for the contribution of each market/variable to the variance of the common trend can be performed. This lead to the next proposition

Proposition 3.2. *Let $(\hat{C} - C_0) \sim N(0, \Sigma_c)$ for $T \to \infty$ as shown in Lemma 3.1, and consider the contribution of market j $(\psi\hat{c}_j)$ to the variance of the common trend, being $\hat{c}_j$ the j-th column of $\hat{C}$ and ψ the common row of $\Psi(1)$. Then, the second central moment of $(\psi\hat{c}_j)$ is distributed according to a Gamma distribution with shape parameter $\lambda = 1/2$ and scale parameter $k = 2\psi'\Sigma^c_{jj}\psi$. That is, $(\psi\hat{c}_j - \psi c_j)^2 \sim \Gamma(1/2, 2\psi'\Sigma^c_{jj}\psi)$.*

Proof. See Appendix B. □

Remark 1. The Gamma distribution simply arise as a scaled-χ^2 with scaling parameter equal to $\psi'\Sigma^c_{jj}\psi$, that is $(\psi\hat{c}_j - \psi c_j)^2/\psi'\Sigma^c_{jj}\psi \sim \chi^2_1$. Then $(\psi\hat{c}_j)^2 \sim \sigma\chi'^2(m)$, where $\sigma\chi'^2(m)$ is a non-central $\chi'^2(m)$ with non-centrality parameter $m = \psi c_j$ multiplied by a scaling factor $\sigma = \psi'\Sigma^c_{jj}\psi$.

The above results imply that typical Wald testing procedures can be easily implemented to test whether the contribution of each market to the price discovery process is significant or not. These results can be summarized as follows. First, when the transitory shocks generating market microstructure noise are not normally distributed, the historical identification problem of the IS measure is tremendously alleviated since it is not necessary to compute all the possible permutations in the model to get an heuristic range of solutions for the IS measure. A unique IC-IS measure, consisting in a precise point estimate for the contribution of each market/variable to the price formation process, can be defined and implemented. Second, this measure can be statistically tested starting from the asymptotic distribution of the estimated matrix C of contemporaneous coefficients.

What remains is to assess the robustness of the IC-IS measure to misspecifications of the pseudo log-likelihood used to get the PML estimator, which will be investigated in a Monte Carlo exercise.

[3] An interesting evaluation study encompassing some of these different approaches can be found, among others, in Moneta and Pallante (2022).

3.2 Montecarlo simulation exercise

In this subsection the robustness of the IC-IS measure to different density functions $g_i(\epsilon_i)$ is investigated. The IC-IS measure, computed starting from the PML estimates, should not be too sensitive to the specification of the pseudo log-likelihood function. In other words, the allocation mechanism for the variance of the common trend should be consistent across different suitable non-Normal density functions chosen. The results will be compared with the standard and well established IS computed with upper and lower bounds associated to different Choleski decompositions.

I simulate N = 500 samples each of length T = 5000. The shocks ϵ_t are drawn from Student distributions with time-varying degree of freedoms v_t to let the variance change over time with a U-shape pattern (Andersen et al., 2012; Bollerslev et al., 2016; Hasbrouck, 2002)

$$\sigma_\epsilon(t) = M + De^{-dt} + We^{-w(1-t)} \tag{10}$$

where parameters are set as $M = 1$, $D = 0.75$, $W = 0.25$, $d = 10$, and $w = 10$ (see Appendix C). Time-varying variance is introduced for two main reasons. First, the U-shape patterns allow to simulate the data more realistically. Intraday financial returns typically display higher levels of volatility both at the beginning and at the end of the trading day, with lower levels of volatility in the middle. Second, being the variance of a Student equal to $v/(v-2)$, the time-varying variance is obtained by letting the degrees of freedoms $v(t)$ of the Student distributions vary accordingly. This introduce an additional source of misspecification with respect to which the robustness of the IC-IS measure is evaluated, being the true distributions not known to the econometrician.

After having specified the 4-dimensional orthogonal mixing matrix C to be estimated, and a matrix S which is used to cross-correlate the shocks ϵ_t to get the correlated price innovations $u_t = SC\epsilon_t$, I simulate 4-dimensional VECMs using the price innovations u_t. For each simulated sample, the VECM is estimated equation by equation given the known cointegration vectors. The price innovations u_t are recovered as residuals and jointly pre-whitened using a Choleski decomposition to both remove the scaling indeterminacy and be compliant with the orthogonality conditions. Then, the PML procedure is performed on the whitened innovations, obtaining the estimates $\hat{C}$ needed to compute the IC-IS as illustrated in section 3.

Being $\mathcal{D}_i$ the true Students with time-varying $v_i(t)$, I set as pseudo distributions $g_i = t(v_i)$, meaning I use Student distributions with different but fixed and predetermined degrees of freedom v_i. Different combinations of degrees of freedom v_i will be considered. The obtained IC-IS are compared with the true IS implied by the simulated model parameters [4], but also with the well-known upper and lower bounds of the IS we would obtain by performing Choleski decompositions over all the possible variable permutations in the model. As an additional robustness check the IC-IS will be computed under an additional source of misspecification: I do not limit the misspecification to the degrees of freedom of the Student but I estimate the IC-IS using other non-Normal distributions as well, namely the Laplace and the Hyperbolic secant distributions.

[4]Parameters are shown in Appendix C, all codes for both the empirical and simulation analysis moreover will be made available publicly.

Table 1: Montecarlo simulation results.

True IS $= [0.0025, 0.1042, 0.5500, 0.3433]$				
	(1) $\epsilon_{1,3} \sim t(3), \epsilon_{2,4} \sim t(4)$			
IC-IS All permutations (Mean)	0.0103 [0.0011, 0.4563] (0.2287)	0.1194 [0.0000, 0.3018] (0.1509)	0.5186 [0.1242, 0.9168] (0.5205)	0.3515 [0.0151, 0.4194] (0.2173)
	(2) $\epsilon_{1,3} \sim t(5), \epsilon_{2,4} \sim t(6)$			
IC-IS All permutations (Mean)	0.0084 [0.0046, 0.5027] (0.2537)	0.1133 [0.0035, 0.3125] (0.158)	0.5313 [0.1191, 0.8695] (0.4943)	0.3469 [0.0163, 0.3716] (0.1940)
	(3) $\epsilon_1 \sim t(4), \epsilon_2 \sim t(5), \epsilon_3 \sim t(7), \epsilon_4 \sim t(12)$			
IC-IS All permutations (Mean range)	0.0072 [0.0000, 0.4280] (0.2140)	0.1042 [0.0016, 0.2800] (0.1408)	0.5371 [0.1316, 0.9381] (0.5349)	0.3515 [0.0180, 0.4387] (0.2284)
	(4) $\epsilon_{1,2,3,4} \sim$ Laplace$(\mu = 0, p = 1)$			
IC-IS All permutations (Mean range)	0.0358 [0.0035, 0.5276] (0.2656)	0.1473 [0.0009, 0.2597] (0.1304)	0.4624 [0.1199, 0.8755] (0.4976)	0.3545 [0.0211, 0.3604] (0.1908)
	(5) $\epsilon_{1,2,3,4} \sim$ Hyperbolic secant			
IC-IS All permutations (Mean range)	0.0102 [0.0009, 0.5170] (0.2590)	0.1066 [0.0000, 0.2557] (0.1278)	0.5405 [0.1265, 0.8969] (0.5117)	0.3427 [0.022, 0.3776] (0.1998)

Notes: Average IC-IS obtained in each Montecarlo simulation. Five different simulation settings from (1) to (5) have been implemented, each of them corresponding to a different specification of the pseudo log-likelihood used to estimate the matrix C and needed to compute the IC-IS. For each specification, the results are compared with the IS obtained performing the standard Choleski procedure with upper and lower bounds and their average. The true IS implied by the data generating process are shown in the first row of the table.

Table 1 reports the average values of the IC-IS obtained from the 500 simulated samples. It is common practice in price discovery applications to get a 'mid-point' of the IS range obtained implementing all the possible Choleski permutations in the model. For this reason I also display that mid-point under the field '(Mean range)', which simply is the average of the upper and lower bounds. Results can be summarized as follows. Independently from the non-Normal distributions chosen in the specification of the pseudo log-likelihood, the proposed IC-IS always allocates the variances across the variables consistently (i.e., the ranking in terms of informativeness is respected and the magnitude of the different shares are close to the true ones). Figure 1 graphically shows some of the scenarios illustrated in table 1, comparing the estimated IC-IS with the true values and showing also the distribution of the IC-IS measure which belongs to the closed interval [0,1].

Remark 2. Given $(\psi\hat{c}_j - \psi c_j)^2 \sim \Gamma(1/2, 2\psi'\Sigma^c_{jj}\psi)$, the quantity X= $(\psi\hat{c}_j$-$\psi c_j)^2/\sum_{j=1}^{N}(\psi\hat{c}_j$ - $\psi c_j)^2$ follows a Beta distribution being a ratio of independent Gamma distributions of the form $\Gamma_j/\sum_{j=1}^{N}\Gamma_j$. Then, the non-central version of this quantity is exactly the IC-IS measure which arise as the ratio of non-central chi-square distributions, which follows the non-central Beta distribution (see Johnson et al., 1995).

Overall, the methodology performs well and it is robust to the different sources of misspecifications stemming both from the choice of specific non-normal distributions and time-varying variances. Most importantly, from this simulation exercise is possible to appreciate how the Choleski permutation procedure would yield very large ranges from which is almost impossible to disentangle the real informativeness of the variables in the model. Taking the mean of these ranges alleviates the problem only partially. Only two variables out of four get shares which are comparable, to some extent, to the true ones by taking the mid-range of the Choleski, while the other two variables are largely either over or underestimated. After having assessed the IC-IS in a simulated environment, an empirical application of the IC-IS on real high-frequency data will follow in the next section.

4 Empirical Application

To evaluate the goodness of the proposed measures, I perform the empirical application on the same IBM data adopted by Hasbrouck (2021), for the day 3 October 2016, which have been shared under the authorization of the NYSE. This allows to keep detailed analyses already established in the literature as a benchmark to compare with, making clearer the interpretation of the obtained results. The recent results of Hasbrouck (2021) have been already reproduced also by Zema (2022), for this reason they will be simply reported here with no need to recompute them.

The econometric analysis is performed on IBM's trades and quotes recorded the day 3 October 2016, with each record reporting both participants and Securities Information Processor (SIP) timestamps, with a sample for the day consisting of around 30.000 observations for each variable. The objective is to evaluate empirically the robustness of the IC-IS measure under different time settings. For this reason, the IC-IS will be implemented both

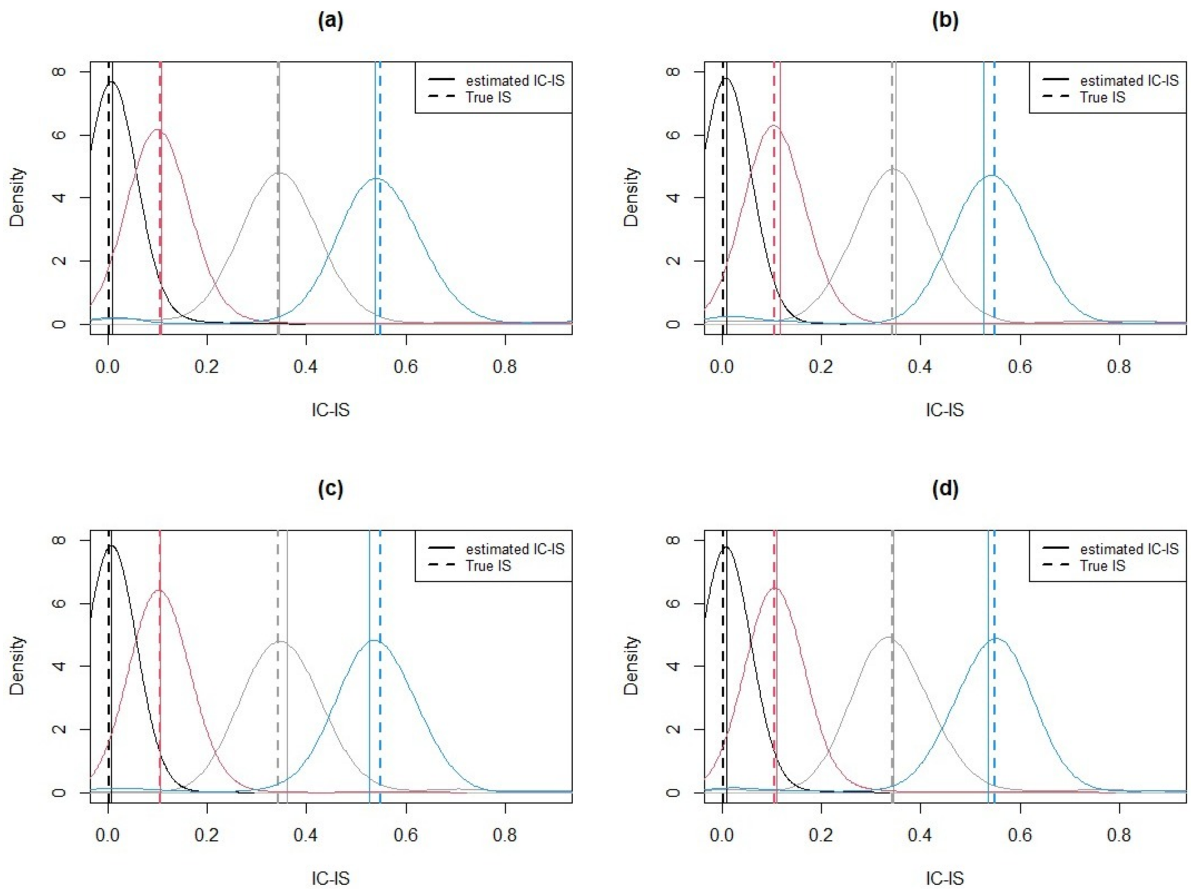

Figure 1: Comparison between the estimated IC-IS (continuous vertical lines) and true IS (dashed lines) measures for each simulated variable, together with the underline distributions of the IC-IS obtained from the N=500 Montecarlo samples. Each color corresponds to a different variable.

in natural and event time, setting a relatively low level of resolution (i.e., second precision) for the data in the natural time specification. This will allow to check the robustness of the obtained results without increasing both the computational complexity and data sparsity introduced when working at very high-frequencies.

The empirical analysis follows three main lines of investigation. The first study focuses on the analysis of participants and SIP timestamps, quantifying the impact of time reporting differentials on the measurement of price discovery. SIP data are needed to establish a consolidated and transparent way to disseminate market data to the public audience. Starting from participants trades and quotes, the SIP compute and publicly disseminate national best bids (NBBs) and offers (NBOs) at which brokers are required to trade, by the regulation, when acting on behalf of their customers. Since SIP data are by construction delayed signals of the participants ones, one expects to attribute leadership in the price discovery process to the participants-based data almost entirely. To perform the analysis, a 4-variables VECM will be estimated including both SIP national NBBs and NBOs plus participants bid and ask prices.

The second study will quantify price discovery across different exchanges instead. Financial instruments are often traded on multiple markets. In particular, public companies can have their stocks traded contemporaneously both in the primary listing exchanges (i.e., where the initial public offering occurred) and other exchanges indeed (same examples include cross-listing, dark pools, OTC markets among others). The VECM here will include IBM bids and offers placed on the primary listing exchange, plus best bids and offers taken from all the other exchanges in which IBM were traded except the primary one.

Finally, the third study analyzes the contribution to price discovery of trades and quotes. Here, the model will include trades occurred on lit and dark pools plus NBBs and NBOs quotes from participant timestamps. Differently from lit pools characterized by stricter regulatory requirements (such as NYSE, NASDAQ, or LSE among others), dark pools are alternative private trading venues with no regulatory transparency requirements. The rationale behind the existence of these dark pools is to let institutional investors trade large volumes of securities without making their hand visible. Trading with an order book not visible to the public avoids potential adverse price effects generated by large movements in the market, still with detrimental effects in terms of transparency. While the analysis of the benefits of lit versus dark pools is not the objective of the analysis from a regulatory perspective, it is interesting to investigate how trades occurred on these two different venues contribute to the price discovery process with respect to quotes.

The three VECM models implemented to perform the three empirical analyses mentioned are the following:

1. $\mathrm{p}_{\mathrm{t}}^{\mathrm{Model1}} = \left[\mathrm{NBB}_{\mathrm{t}}^{\mathrm{Participants}}, \mathrm{NBO}_{\mathrm{t}}^{\mathrm{Participants}}, \mathrm{NBB}_{\mathrm{t}}^{\mathrm{SIP}}, \mathrm{NBO}_{\mathrm{t}}^{\mathrm{SIP}}\right]$;
2. $\mathrm{p}_{\mathrm{t}}^{\mathrm{Model2}} = \left[\mathrm{NBB}_{\mathrm{t}}^{\mathrm{OtherExchanges}}, \mathrm{NBO}_{\mathrm{t}}^{\mathrm{OtherExchanges}}, \mathrm{Bid}_{\mathrm{t}}^{\mathrm{Primary}}, \mathrm{Ask}_{\mathrm{t}}^{\mathrm{Primary}}\right]$;
3. $\mathrm{p}_{\mathrm{t}}^{\mathrm{Model3}} = \left[\mathrm{NBB}_{\mathrm{t}}^{\mathrm{Participants}}, \mathrm{NBO}_{\mathrm{t}}^{\mathrm{Participants}}, \mathrm{Trade}_{\mathrm{t}}^{\mathrm{LitPools}}, \mathrm{Trade}_{\mathrm{t}}^{\mathrm{DarkPools}}\right]$.

The results of the analysis are then displayed and commented in the next subsection.

4.1 Results

For each model, related to a given price discovery analysis, the IC-IS measure is computed and compared with the standard Choleski based IS in which upper and lower bounds are computed by going through all the possible permutations of the variables in the model. The results are shown in table 2 and can be summarized as follows.

In the participant versus SIP timestamps analysis, the IC-IS attributed importance to the participants in the price discovery process almost entirely, as expected, with a 99 percent share of price discovery in the event-time specification. Noticeably, the IC-IS consistently attributed to the participants 86 percent of price discovery even in the 1-second resolution in natural time. This is not the case for the standard approach where very large upper and lower bounds for the IS are obtained, making impossible to disentangle the real contribution of price discovery in natural time.

In the primary listing versus other exchanges analysis the same hold. The IC-IS attribute most of the price discovery to the primary listing exchange consistently across the two time specifications. Interestingly, in the event-time framework where the identification problem for the IS is relaxed (cross-correlations in event-time are lower compared to the natural-time setting) the IC-IS gives a 78 percent share to the primary exchange, thus attributing much more importance than the one attributed by the Choleski permutation procedure (with a range for the primary listing being 46-56 percent). In the natural time the IC-IS still manage to give leadership in price discovery to the primary market with a 55 percent share, while the permutation procedure would still yield very large bounds.

Finally, in the quotes versus trade analysis results suggest that quote are more informative than trades. Since trades in dark pools had a totally negligible contribution in terms of price discovery, their shares have been added to the ones obtained for the lit pools. The IC-IS consistently attribute the majority of the information to quotes both in natural and event-time. Still, as in the previous cases, it is possible to appreciate discrepancies across the two different time settings in terms of shares magnitude.

The IC-IS does not solve the cross-correlation problem which arise aggregating information inside each second interval, and the event-time analysis should be considered in this respect more reliable being the time counter updated any time new information arrives. It's still valuable to notice how the measure still provide consistent and reasonable results even with a relatively low level of resolution in natural time, something which hardly happen with a Choleski permutation procedure. In the event-time, where the identification issue is relaxed, results are consistent with the upper and lower bound except for the primary versus other exchange analysis, with the IC-IS giving much more importance to the primary listing compared to the classical IS.

As illustrated in the previous section the contribution of each variable to the price discovery process, quantified by the IC-IS measure, can be easily tested defining the Wald statistics $\hat{W}_{j,T} = (\hat{\psi}\hat{c}_j - \psi c_j)^2 / \hat{\psi}' \hat{\Sigma}^c_{jj} \hat{\psi}$ whose asymptotic distribution is $\chi^2(1)$. The null hypothesis is $H_0 : \psi c_j = 0$ (i.e., the contribution of the generic variable j to the price discovery process is not significant), versus $H_1 : \psi c_j \neq 0$. The tables 3 and 4 display the Wald tests for each variable, in each price discovery analysis, in natural and event-time respectively.

Table 2: IC-IS and IS comparison: Summary results.

	IC-IS		All permutations			
	participants	SIP	participants		SIP	
			Min	Max	Min	Max
1-sec	0.86	0.14	0.002	0.999	0.001	0.998
Event time	0.997	0.002	0.943	0.999	0.001	0.057
	primary	non-primary	primary		non-primary	
			Min	Max	Min	Max
1-sec	0.55	0.45	0.12	0.994	0.006	0.88
Event time	0.78	0.22	0.46	0.56	0.44	0.54
	Quotes	Trades	Quotes		Trades	
			Min	Max	Min	Max
1-sec	0.90	0.10	0.39	0.979	0.021	0.61
Event time	0.67	0.33	0.61	0.67	0.33	0.39

Notes: Information shares measures for each price discovery analysis, comparing the IC-IS with the classical IS based on all permutations. In natural-time(1-sec), 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. Trades include both lit and dark trades, given that the contribution of the latter is negligible. The all permutations approach yielded results consistent with Hasbrouck (2021). For the specification of the pseudo log-likelihood pairs of Student distributions with 3 and 4 d.o.f. have been used, but results have been found to be consistent with the adoption of other heavy tail distributions such as the Laplace or the Hyperbolic secant.

Table 3: Wald tests in natural time at 1-second resolution.

Natural time (1-sec) analysis				
Model 1: Participant VS SIP				
	NBBpart	NBOpart	NBBsip	NBOsip
W	78.083	52.106	21.672	9.069
p-value	9.87e-19	5.25e-13	3.23e-06	2.59e-03
Model 2: Primary listing VS Other Exchanges				
	NBB(others)	NBO(others)	Bid(primary)	Ask(primary)
W	40.507	4.867	13.385	34.439
p-value	1.95e-10	2.737e-02	2.54e-04	4.39e-09
Model 3: Trades VS Quotes				
	Trades(Lit)	Trades(Dark)	NBBpart	NBOpart
W	2.758	0.015	14.431	10.063
p-value	0.096	0.903	1.4e-4	1.5e-3

Notes: Wald tests for the contribution of each variable j to the price discovery process in natural time, W is the Wald statistic. The test is performed using the $\chi^2(1)$ under the null hypothesis $H_0 : \psi c_j = 0$. The contribution of both lit and dark trades to price discovery is statistically equal to zero ($\alpha = 0.05$) and only quotes are informative at 1-second resolution.

Table 4: Wald tests in event-time.

Event-time analysis				
Model 1: Participant VS SIP				
	NBBpart	NBOpart	NBBsip	NBOsip
W	83.988	50.594	0.026	0.261
p-value	4.97e-20	1.14e-12	0.871	0.609
Model 2: Primary listing VS Other Exchanges				
	NBB(others)	NBO(others)	Bid(primary)	Ask(primary)
W	7.634	0.002	14.797	14.604
p-value	5e-03	0.962	1.19e-04	1.33e-04
Model 3: Trades VS Quotes				
	Trades(Lit)	Trades(Dark)	NBBpart	NBOpart
W	12.753	0.009	18.34	11.105
p-value	3.55e-04	0.923	1.85e-05	8.61e-04

Notes: Wald tests for the contribution of each variable j to the price discovery process in event-time, W is the Wald statistic. The test is performed using the $\chi^2(1)$ under the null hypothesis $H_0 : \psi c_j = 0$. SIP timestamps are both statistically equal to zero ($\alpha = 0.05$) as well as NBO from exchanges that are not the primary listing one. Dark trades are not informative also in event-time.

In a natural time framework at 1-second resolution, all variables appear to be informative except trades. The contribution of both lit and dark trades is statistically equal to zero and only quotes are informative. Lit trades, despite having an IC-IS of a 10 percent magnitude, have high level of uncertainty at 1 second resolution indeed.

In the event-time analysis SIP timestamps completely lose statistical significance as expected, also quotes placed outside the primary listing exchange partially lose significance being the contribution of NBO(others) statistically equal to zero. Finally, in event-time lit trades are informative while dark trades remain uninformative.

Overall, the empirical results obtained with the proposed IC-IS measure are consistent across different time-specifications in choosing the leaders in the price formation process. Results are also coherent with Hasbrouck (2021) except they attribute a greater importance to the primary listing exchange. These results have been achieved without either increasing the modeling and computational complexity which arise when working at incredibly short time scales, or introducing the rather restrictive directed acyclic graph structure assumption of Zema (2022). This leads to the conclusion.

5 Conclusion

Measuring the contributions to the price formation process through the estimation of unique information share measures represents a long-standing issue in empirical finance and market microstructure modeling. While several attempts have been made in the literature, a general and relatively easy to implement procedure is not available yet. To this end, a new measure of price discovery, namely the IC-IS, has been introduced. The measure, does not suffer from the identification issues inherited by the historical IS measure.

Differently from both the historical IS and the DAG-IS defined by Zema (2022), for which a recursive structure is imposed in the system through the adoption of Choleski decompositions, the IC-IS provides a less restrictive framework in which no triangular structure assumptions is needed to resolve the identification issues. Moreover, this new measure neither require the adoption of different volatility regimes nor require to model in natural-time at incredibly short time-scales. For these reasons, the IC-IS could bring new insights about the way in which the contributions to price discovery, through the variance of the efficient price process as historically proposed by Hasbrouck (1995), are both identified and quantified.

Even if the IC-IS can be adopted as a standalone measure, being a testing framework available for it, the greatest benefits might come with an adoption which is complementary to other established measures, especially when no sound prescription in favor of a specific approach is available. The empirical application on IBM data, performed keeping the results of Hasbrouck (2021) as a sound benchmark in the literature, provided consistent and reasonable results indeed, raising the possibility for future applications in the field to benefit from the new measure.

Appendix A. Proof of proposition 3.1

Proof. Consider the estimated orthogonal mixing matrix of the independent shocks $\hat{C}$. As a consequence of Theorem 3.1 columns permutations in $\hat{C}$ create a lack of identification.

Being $\hat{C}$ orthogonal, $c_i^{'}c_j = 0$ for $i < j$, and $c_i^{'}c_i = 1 \; \forall \; i$. Let then P be the permutation matrix such that $\hat{C}^p = \hat{C}P$ satisfies the identification condition $|c_{ii}| \geq |c_{ij}| \; \forall \; i \neq j$ stated in Proposition 3.1.

Consider now the estimated mixing matrix $\hat{C}P^{(r)}$ with columns randomly permuted by a random permutation matrix $P^{(r)}$. Then, there always exists a permutation matrix $P^* = P^{(r)^{'}}P$ such that $\hat{C}P^{(r)}P^* = \hat{C}P$.

□

Exploiting basic properties of permutation matrices, the proof above simply shows that independently from the arbitrary column ordering obtained after estimating the mixing matrix C, it is always possible to find an appropriate permutation matrix, being the product of two permutation matrices a permutation matrix itself, such that the identification conditions in Proposition 3.1 are met.

Appendix B. Proof of proposition 3.2

The proof for the asymptotic distribution of $(\psi\hat{c}_j - \psi c_j)^2$, recalling c_j is the j-th column of C, is a trivial application of the delta method in the multivariate case knowing that asymptotically $vec\sqrt{T}(\hat{C}_T - C) \sim N(0, \Sigma_C)$.

Proof. Given the continuous differentiable function $f(\hat{c}_j) = \psi\hat{c}_j$, consider the first order Taylor series expansion around the true value c_j (higher order terms are exactly zero being $f(\hat{c}_j)$ linear)

$$f(\hat{c}_{jT}) = f(c_j) + \nabla f(c_j)^{'}(\hat{c}_{jT} - c_j)$$

that is

$$\psi\hat{c}_{jT} = \psi c_j + \psi(\hat{c}_{jT} - c_j),$$

then from the Slutsky's Theorem (Gut, 2005) follows that if $\sqrt{T}(\hat{c}_{jT} - c_j) \xrightarrow{d} N(0, \Sigma^{c_j})$, being Σ^{c_j} the covariance matrix of the j-th column of C, then

$$\sqrt{T}(f(\hat{c}_j) - f(c_j)) \xrightarrow{d} N(0, \nabla f(c_j)^{'}\Sigma^{c_j}\nabla f(c_j))$$

that is

$$\sqrt{T}(\psi\hat{c}_j - \psi c_j) \xrightarrow{d} N(0, \psi\Sigma^{c_j}\psi^{'}).$$

It follows that $(\psi\hat{c}_j - \psi c_j)/\sqrt{\psi\Sigma^{c_j}\psi^{'}} \sim N(0,1)$ which implies $(\psi\hat{c}_j - \psi c_j)^2/\psi\Sigma^{c_j}\psi^{'} \sim \chi^2(1)$. Then, $(\psi\hat{c}_j - \psi c_j)^2 \sim \psi\Sigma^{c_j}\psi^{'}\chi^2(1) \rightarrow (\psi\hat{c}_j - \psi c_j)^2 \sim \Gamma(1/2, 2\psi^{'}\Sigma^{c}_{jj}\psi)$ being the χ^2 a special case of a Gamma with parameters $\lambda =$ d.o.f./2 and $k = 2$, where the scale parameter k absorbs the variance $\psi^{'}\Sigma^{c}_{jj}\psi$ and d.o.f $= 1$.

□

Appendix C: Simulation setting and parameters

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

$$\Pi(L)p_t = u_t \tag{B.1}$$

where

$$\Pi(L) \equiv I_n - \sum_i^k \Pi_i L^i \tag{B.2}$$

$$\alpha\beta' = (\sum_i^k \Pi_i - I_n) \tag{B.3}$$

$$\Phi_s = -(\Pi_{s+1} + \Pi_{s+2} + ... + \Pi_k) \tag{B.4}$$

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 cointegration vectors and has rank equal to *n-1*. I simulate the system with 1 lag only for simplicity, so parameters are

$$\alpha = \begin{pmatrix} 0.025 & 0.05 & 0.03 \\ 0.08 & 0.07 & 0.06 \\ 0.1 & 0.01 & 0.04 \\ 0.09 & 0.06 & 0.09 \end{pmatrix}, \quad \beta = \begin{pmatrix} 1 & \\ \vdots & -\mathrm{I_{n-1}} \\ 1 & \end{pmatrix}, \Phi_1 = \begin{pmatrix} 0.4 & -0.9 & -0.25 & 0.3 \\ 0.6 & 0.35 & 0.55 & -0.1 \\ 0.2 & -0.2 & -0.7 & 0.4 \\ 0.1 & 0.35 & 0.6 & 0.1 \end{pmatrix},$$

and $\Phi_1 = -\Pi_2$, $\Pi_1 = \alpha\beta' + I - \Pi_2$. Finally, the matrix S used in $u_t = SC\epsilon_t$ is

$$S = \begin{pmatrix} 0.9 & 0 & 0 & 0 \\ 0.4 & 0.6 & 0 & 0 \\ 0.5 & 0.2 & 0.7 & 0 \\ 0.3 & 0.5 & 0.3 & 0.1 \end{pmatrix}.$$

It must be mentioned that typically the diurnal U-shape pattern is quantified in the literature by setting $M \approx 0.89$. Here $M = 1$ simply to guarantee the existence of the variance of the Students from which shocks are generated, being the degrees of freedom of the Students mapped over time in a one-to-one relation with the time-varying variances. This useful shift still preserves a U-shape patterns and does not hamper in any way the IC-IS measure in the simulation exercise.

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.

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.

Blanco, R., S. Brennan, and I. W. Marsh (2005). An empirical analysis of the dynamic relation between investment-grade bonds and credit default swaps. *The journal of Finance 60*(5), 2255–2281.

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 (2021). Comment on: Price discovery in high resolution. *Journal of Financial Econometrics 19*(3), 431–438.

Buccheri, G., G. Bormetti, F. Corsi, and F. Lillo (2021). Comment on: Price discovery in high resolution. *Journal of Financial Econometrics 19*(3), 439–451.

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.

de Jong, F. (2021). Comment on: Price Discovery in High Resolution*. *Journal of Financial Econometrics 19*(3), 452–458.

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 (2021). Price discovery in a continuous-time setting. *Journal of Financial Econometrics 19*(5), 985–1008.

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.

Eriksson, J. and V. Koivunen (2004). Identifiability, separability, and uniqueness of linear ica models. *IEEE signal processing letters 11*(7), 601–604.

Fernandes, M. and C. M. Scherrer (2018). Price discovery in dual-class shares across multiple markets. *Journal of Futures Markets 38*(1), 129–155.

Ghysels, E. (2021). Comment on: Price discovery in high resolution and the analysis of mixed frequency data. *Journal of Financial Econometrics 19*(3), 459–464.

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.

Gouriéroux, C., A. Monfort, and J.-P. Renne (2020). Identification and estimation in non-fundamental structural varma models. *The Review of Economic Studies 87*(4), 1915–1953.

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.

Guidolin, M., M. Pedio, and A. Tosi (2021). Time-varying price discovery in sovereign credit markets. *Finance Research Letters 38*, 101388.

Gut, A. (2005). *Probability: a graduate course*, Volume 200. Springer.

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

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.

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

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

Hasbrouck, J. (2021). Price discovery in high resolution. *Journal of Financial Econometrics 19*(3), 395–430.

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.

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

Johnson, N. L., S. Kotz, and N. Balakrishnan (1995). *Continuous univariate distributions, volume 2*, Volume 289. John wiley & sons.

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

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.

Leung, T., M. Lorig, and A. Pascucci (2017). Leveraged etf implied volatilities from etf dynamics. *Mathematical Finance 27*(4), 1035–1068.

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.

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 (2022). Identification of structural var models via independent component analysis: a performance evaluation study. *Journal of Economic Dynamics and Control 144*, 104530.

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.

Shum, P., W. Hejazi, E. Haryanto, and A. Rodier (2016). Intraday share price volatility and leveraged etf rebalancing. *Review of Finance 20*(6), 2379–2409.

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

Zema, S. M. (2022). Directed acyclic graph based information shares for price discovery. *Journal of Economic Dynamics and Control 139*, 104434.