# Price Discovery Share: An Order Invariant Measure of Price Discovery<sup>\*</sup>

Shulin Shen<sup>†</sup> Syed Galib Sultan<sup>‡</sup> Eric Zivot<sup>§</sup>

February 14, 2024

#### Abstract

To address the order-dependence of Hasbrouck's (1995) information share (IS) measure of a market's contribution to price discovery, we propose a closely related measure, which we call *price discovery share* (PDS), that is simple to compute, easy to interpret, order invariant, and unique. Our PDS measure is motivated by a widely used method in portfolio risk management to additively decompose portfolio volatility into asset specific contributions. We show analytically and through simulations that PDS provides advantages over IS and the modified IS by Lien and Shrestha (2009).

**Keywords**: price discovery, information share, order-dependence **JEL classification**: C32; G10

<sup>\*</sup>This note is a shortened version of the unpublished working paper Sultan and Zivot (2015) "Price Discovery Share-An Order Invariant Measure of Price Discovery with Application to Exchange-Traded Funds". We thank Shamma Adeeb Alam, Jonathan Brogaard, Yu-Chin Chen, Samuel Henly, Fahad Khalil, Jennifer Koski, Ji Hyung Lee, Xuyang Ma, Ananth Madhavan, Nisha Noor, Syed Sadaf Sultan, Model Validation Group at State Street Corporation and participants at the Annual Conference of The Canadian Economic Association (2014) and The 2015 R in Finance Conference for helpful comments. Financial support from the Robert Richards Endowed Chair and the National Natural Science Foundation of China [Grant No. 71903060] is greatly appreciated. We also thank Yixuan Zhang at Huazhong University of Science and Technology for offering excellent research assistance.

<sup>&</sup>lt;sup>†</sup>School of Economics, Huazhong University of Science and Technology, Wuhan, Hubei, China 430074. E-mail: shen\_shulin@hust.edu.cn

<sup>&</sup>lt;sup>‡</sup>Model Validation Group, State Street Corporation, Boston, MA 02110, USA. E-mail: galibsultan@gmail.com. The opinions and findings presented in this research paper are solely those of the author and do not necessarily reflect the views or positions of State Street Corporation. The author is an employee of State Street Corporation, but the contents of this paper are independent of the company's official stance on the subject matter. This research is conducted in the author's personal capacity and does not constitute an official statement or endorsement by State Street Corporation.

<sup>&</sup>lt;sup>§</sup>Department of Economics, Uninversity of Washington, Seattle, WA 98195, USA. E-mail: ezivot@uw.edu

# Price Discovery Share: An Order Invariant Measure of Price Discovery

#### Abstract

To address the order-dependence of Hasbrouck's (1995) information share (IS) measure of a market's contribution to price discovery, we propose a closely related measure, which we call *price discovery share* (PDS), that is simple to compute, easy to interpret, order invariant, and unique. Our PDS measure is motivated by a widely used method in portfolio risk management to additively decompose portfolio volatility into asset specific contributions. We show analytically and through simulations that PDS provides advantages over IS and the modified IS by Lien and Shrestha (2009).

**Keywords**: price discovery, information share, order-dependence **JEL classification**: C32; G10

## 1 Introduction

Price discovery, the process by which new information is impounded into asset prices through trading activity, is an important research agenda in the financial economics literature and Hasbrouck's (1995) information share (IS) is the most widely used empirical measure to identify and quantify the process of price discovery. However, it is well-documented that IS has a serious identification problem. When idiosyncratic innovations to different market prices in Hasbrouck's (1995) cointegration model are contemporaneously correlated, the IS measure, which is typically reported as a range, can become very wide and does not clearly identify the price leader or the follower. This limitation has been referred to in previous literature as the *order-dependence problem of IS*, because the upper and the lower bound of the range that IS reports depends on the order in which the prices enter into the price vector in the cointegration model.

Numerous studies have proposed different solutions to the order-dependence problem of IS. However, no consensus has emerged so far because all of the proposed alternatives have been found to be either only effective in a particular context or to have their own identification issues. Hasbrouck (1995) suggests sampling trade and quote prices at a high enough frequency such that contemporaneous correlation among the innovations becomes negligible. However, numerous studies find that even with the use of data sampled at a 1-second interval there is still enough residual correlation to produce a wide range for IS. Baillie et al. (2002) argue in support of using the mean or mid-point of the upper and lower bound of the range of IS as a unique measure of IS (IS-mean hereafter). This approach, while intuitively appealing, is ad hoc. Lien and Shrestha (2009) propose a modified information share (MIS) measure that is derived from the spectral decomposition of the innovation correlation matrix.<sup>1</sup> This approach imposes an unintuitive factor structure on the cointegration model, and the calculation of MIS can be numerically unstable when innovations are very highly correlated. De Jong and Schotman (2010) define an order-invariant IS-type measure based on a structural unobserved component model. While the parameters of this unobserved components model have clear interpretations identification of these parameters requires restrictive

<sup>&</sup>lt;sup>1</sup>Lien and Shrestha (2014) propose the generalized information share (GIS), which is basically the same as MIS though under a more generalized cointegration settings.

 $assumptions.^2$ 

In this paper, we propose a new price discovery measure that is closely related to IS, simple to compute, easy to interpret, order-invariant, and unique. Our measure of price discovery is motivated by a widely used method in portfolio risk management (see, e.g., Bruder and Roncalli, 2012) to additively decompose portfolio volatility into asset specific contributions. A notable feature of Hasbrouck's (1995) cointegration model for capturing price discovery is that the volatility of the efficient price innovation (permanent shock) is linearly homogeneous in the common factor weights of each market's innovation just as portfolio volatility is linearly homogeneous to its portfolio weights. We use this property and apply Euler's theorem to additively decompose the permanent shock volatility into components attributed to each market. These components are defined as the contributions of each market to the price discovery process. We convert these market contributions to market shares by normalizing by the permanent shock volatility. Our new measure of price discovery for each market is this contribution share which we call *price discovery share* (PDS).

PDS is applicable to the general *n*-asset cointegration model. In the bivariate case, we provide analytical comparisons between IS, IS-mean, MIS, De-Jong and Schotman's IS, and PDS. We show analytically that the difference between PDS across two markets is as least as large as the corresponding difference between MIS and so is better able to reveal a dominance-satellite relationship between markets than MIS. Using simulations from a standard market microstructure model, we compare estimates of IS-mean, MIS, and PDS under different sampling frequencies and show that PDS is more informative about a market's price leadership as the sampling frequency lowers from 1 second to 5 minutes.

The remainder of the paper is organized as follows. In Section 2, we describe the reducedform cointegration framework and define our new measure of price discovery, PDS. In Section 3, we compare PDS to existing price discovery measures. In Section 4, we examine the performances of PDS, IS-mean, and MIS under different sampling frequencies from simulated market data from a stylized market microstructure model of asset prices. A brief summary

 $<sup>^{2}</sup>$ Grammig and Peter (2014) also propose a unique measure for price discovery which is derived by exploiting two properties of price changes: fat tails and tail-dependence. However, this measure is not volatility based like the other IS-type measures.

of the paper's findings is provided in Section 5.

## 2 Model description

Let  $\mathbf{P}_t = (p_{1,t}, \cdots, p_{n,t})'$  denote an  $n \times 1$  vector of I(1) log prices with a single common stochastic component, or a unique fundamental value, that drives all arbitrage-linked prices. As a result, there are n - 1 cointegrating vectors  $\boldsymbol{\theta}_i$  such that,  $\boldsymbol{\theta}'_i \mathbf{P}_t \sim I(0)$ . We use the following  $(n - 1) \times n$  matrix to denote a basis for the cointegrating space:

$$\boldsymbol{\Theta}' = \begin{bmatrix} \boldsymbol{\theta}'_1 \\ \vdots \\ \vdots \\ \boldsymbol{\theta}'_{n-1} \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} = \left( \mathbf{1}_{n-1} \quad \vdots \quad -\mathbf{I}_{n-1} \right), \tag{1}$$

where  $\mathbf{1}_{n-1}$  is an  $(n-1) \times 1$  vector of ones and  $\mathbf{I}_{n-1}$  is the identity matrix of dimension n-1. Since  $\Delta \mathbf{P}_t$  is I(0), it has a Wold representation:

$$\Delta \mathbf{P}_t = \mathbf{\Psi}(L)\mathbf{e}_t = \mathbf{e}_t + \mathbf{\Psi}_1\mathbf{e}_{t-1} + \mathbf{\Psi}_2\mathbf{e}_{t-2} + \dots, \qquad (2)$$

where  $\Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k, \Psi_0 = \mathbf{I}_n, \mathbf{e}_t = (e_{1t}, \dots, e_{nt})'$  and  $\mathbf{e}_t \sim iid(0, \Sigma)$  where  $\Sigma$  is an  $n \times n$  matrix with elements  $\sigma_{ij}$ . Using the Beveridge-Nelson decomposition (Beveridge and Nelson, 1981), we can write:

$$\mathbf{P}_t = \mathbf{P}_0 + \boldsymbol{\Psi}(1) \sum_{j=0}^t \mathbf{e}_j + \boldsymbol{\Psi}^*(L) \mathbf{e}_t, \qquad (3)$$

where  $\Psi(1) = \sum_{k=0}^{\infty} \Psi_k, \Psi^*(L) = \sum_{k=0}^{\infty} \Psi_k^* L^k, \Psi_k^* = -\sum_{j=k+1}^{\infty} \Psi_j$ , and  $\Psi^*(L)\mathbf{e}_t \sim I(0)$ .

The matrix  $\Psi(1)$  contains the cumulative impacts of the innovation  $\mathbf{e}_t$  on all future price movements, and thus measures the long-run impact of  $\mathbf{e}_t$  on prices. As shown in Hasbrouck (1995), the restriction  $\Theta' \Psi(1) = 0$  implies that  $\Psi(1)$  has rank one and can be expressed as

$$\Psi(1) = \mathbf{1}_n \psi' = \begin{bmatrix} \psi_1 & \cdots & \psi_n \\ \vdots & \ddots & \vdots \\ \psi_1 & \cdots & \psi_n \end{bmatrix},$$
(4)

where  $\boldsymbol{\psi} = (\psi_1, ..., \psi_n)'$ . Since the rows of  $\Psi(1)$  are identical, the long-run impact of  $\mathbf{e}_t$  on each price is identical. Substituting (4) into (3) gives:

$$\mathbf{P}_{t} = \mathbf{P}_{0} + \mathbf{1}_{n} \sum_{j=0}^{t} \eta_{j}^{P} + \tilde{\boldsymbol{\varepsilon}}_{t} = \mathbf{P}_{0} + \mathbf{1}_{n} m_{t} + \tilde{\boldsymbol{\varepsilon}}_{t}, \qquad (5)$$

where  $\eta_t^P = \boldsymbol{\psi}' \mathbf{e}_t$  is the component of price changes that is permanently impounded into prices due to new information, and  $m_t = m_{t-1} + \eta_t^P = \sum_{j=0}^t \eta_j^P$  is the random walk component that is common to all prices. Transient pricing errors such as bid-ask bounces and inventory adjustments are absorbed by the I(0) component  $\tilde{\boldsymbol{\varepsilon}}_t = \boldsymbol{\Psi}^*(L)\mathbf{e}_t$ .

#### 2.1 Information Share

Hasbrouck (1995) defines the information share of *i*-th market as its share of the permanent shock variance,  $\operatorname{var}(\eta_t^p) = \psi' \Sigma \psi$ . There are two cases to consider.

Case 1.  $\Sigma$  is diagonal:

$$IS_i = \frac{(\psi_i \sigma_i)^2}{\psi' \Sigma \psi}, i = 1, ..., n.$$
(6)

Case 2.  $\Sigma$  is non-diagonal:

$$\mathrm{IS}_{i} = \frac{((\boldsymbol{\psi}'\mathbf{F})_{i})^{2}}{\boldsymbol{\psi}'\boldsymbol{\Sigma}\boldsymbol{\psi}}, i = 1, ..., n.$$
(7)

where  $(\psi' \mathbf{F})_i$  is the *i*-th element of  $\psi' \mathbf{F}$  and  $\mathbf{F}$  is a lower triangular matrix (Cholesky factor) such that  $\mathbf{FF}' = \Sigma$ . The value of  $\mathbf{F}$  and hence, also the value of  $IS_i$ , depends on the ordering in which the individual prices enter into the vector of price,  $\mathbf{P}_i$ . Therefore, when  $\Sigma$  is nondiagonal, Hasbroucks's approach can only provide upper and lower bounds for  $IS_i$  based on all possible orderings of prices in the vector. In particular, Baillie et al.(2002) show that largest information share for a given market occurs when its price is placed first in the price vector.

In practice,  $IS_i$  is computed from an empirical vector error correction model (VECM) of the form:

$$\Delta \mathbf{P}_{t} = \mathbf{A}(\Theta' \mathbf{P}_{t-1} - \boldsymbol{\mu}) + \sum_{j=1}^{K-1} \Gamma_{j} \Delta \mathbf{P}_{t-j} + \mathbf{e}_{t},$$
(8)

where **A** is an  $n \times (n - 1)$  matrix of error correction parameters,  $\Gamma_j$  is an  $n \times n$  matrix of short-run coefficients. The lag length, K, is typically chosen by some model selection criterion such as BIC or AIC. Because the cointegrating matrix  $\Theta'$  is known, equation (8) can be estimated by least squares equation-by-equation. The long-run impact matrix,  $\Psi(1)$ can be computed directly using Johansen's factorization:

$$\Psi(1) = \Theta_{\perp} (\mathbf{A}_{\perp}' \Gamma(1) \Theta_{\perp})^{-1} \mathbf{A}_{\perp}', \qquad (9)$$

where  $\Theta_{\perp}$  and  $\mathbf{A}_{\perp}$  are vectors satisfying  $\Theta' \Theta_{\perp} = 0$  and  $\mathbf{A}' \mathbf{A}_{\perp} = 0$ , respectively. Also,  $\Gamma(1) = \mathbf{I}_n - \sum_{j=1}^{K-1} \Gamma_j$ .

### 2.2 Price Discovery Share

Our new measure of price discovery is motivated by the additive decomposition of portfolio volatility that is widely used in portfolio risk management. Recall, the permanent shock is defined as a weighted average of individual market innovations  $\eta_t^P = \psi' \mathbf{e}_t$ . The volatility of the permanent shock is  $\sigma_\eta(\psi) = (\psi' \Sigma \psi)^{1/2}$ . Now  $\sigma_\eta(\psi)$  is linearly homogenous in  $\psi$  since  $\sigma_\eta(c \cdot \psi) = c \cdot \sigma_\eta(\psi)$  for any constant c. As a result we can apply Euler's theorem and derive the following additive decomposition of  $\sigma_\eta(\psi)$ :

$$\sigma_{\eta}(\boldsymbol{\psi}) = \boldsymbol{\psi}' \frac{\partial \sigma_{\eta}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \sum_{i=1}^{n} \psi_{i} \frac{\partial \sigma_{\eta}(\boldsymbol{\psi})}{\partial \psi_{i}} = \psi_{1} \frac{\partial \sigma_{\eta}(\boldsymbol{\psi})}{\partial \psi_{1}} + \dots + \psi_{n} \frac{\partial \sigma_{\eta}(\boldsymbol{\psi})}{\partial \psi_{n}}.$$
 (10)

Equation (10) shows that the volatility of the permanent shock,  $\sigma_{\eta}(\boldsymbol{\psi})$ , can be expressed as the weighted sum of marginal contributions from each asset (or market *i*). The *i*-th term on the right-hand side of (10),  $\psi_i \frac{\partial \sigma_{\eta}(\boldsymbol{\psi})}{\partial \psi_i}$ , is asset *i*'s (or market *i*'s) contribution to the volatility of the permanent shock. In the spirit of Hasbrouck's information share,  $\psi_i \frac{\partial \sigma_n(\psi)}{\partial \psi_i}$  is a natural measure of an asset's (or market's) contribution to price discovery.

Our new order invariant measure of an asset's (or market's) price discovery share, denoted  $PDS_i$ , is its contribution divided by  $\sigma_{\eta}(\boldsymbol{\psi})$ :

$$PDS_{i} = \frac{\psi_{i} \frac{\partial \sigma_{\eta}(\psi)}{\partial \psi_{i}}}{\sigma_{\eta}(\psi)}.$$
(11)

By construction  $\sum_{i=1}^{n} PDS_i = 1$ . By the chain rule:

$$\frac{\partial \sigma_{\eta}(\boldsymbol{\psi})}{\partial \boldsymbol{\psi}} = \frac{\boldsymbol{\Sigma} \boldsymbol{\psi}}{\sigma_{\eta}(\boldsymbol{\psi})} = \sigma_{\eta}(\boldsymbol{\psi})\boldsymbol{\beta},\tag{12}$$

where  $\boldsymbol{\beta} = (\beta_1, ..., \beta_n)' = \frac{\Sigma \boldsymbol{\psi}}{\sigma_{\eta}^2(\boldsymbol{\psi})}$  with  $\beta_i$  defined as follows:

$$\beta_i = \frac{\operatorname{cov}(e_{it}, \eta_t^P)}{\operatorname{var}(\eta_t^P)} = \frac{\psi_i \sigma_i^2 + \sum_{j=1}^n \psi_{j \neq i} \sigma_{ij \neq i}}{\psi' \Sigma \psi}.$$
(13)

As a result, (11) may be re-expressed as

$$PDS_{i} = \psi_{i}\beta_{i} = \frac{\psi_{i}^{2}\sigma_{i}^{2} + \sum_{j=1}^{n}\psi_{i}\psi_{j\neq i}\sigma_{ij\neq i}}{\psi'\Sigma\psi}.$$
(14)

We denote  $\beta_i$  in equation (14) as the *price discovery beta* of asset *i* (or market *i*). The price discovery beta is the slope coefficient from the regression of  $e_{it}$  on  $\eta_t^P$  and summarizes the (normalized) covariance contributions of an asset's (or market's) innovation to the variance of the efficient price innovation. Equation (14) shows that PDS<sub>i</sub> is defined as asset *i*'s (or market *i*'s) contribution to the volatility of  $\eta_t^P$  weighted by its price discovery beta.

### **3** Comparisons of PDS to Existing Measures

### 3.1 Comparison to IS

We consider the bivariate case of n = 2, so that  $\mathbf{P}_t = (p_{1,t}, p_{2,t})'$ . Under the assumption of uncorrelated innovations (diagonal  $\Sigma$ ), we find that IS<sub>i</sub> and PDS<sub>i</sub> are identical:

$$IS_{i,diag} = \frac{\psi_i^2 \sigma_i^2}{\psi' \Sigma \psi} = \frac{\psi_i^2 \sigma_i^2}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2} = \psi_i \beta_i = PDS_{i,diag}.$$
 (15)

However, this is not the case when  $\Sigma$  is non-diagonal. Let,  $\Sigma = \mathbf{F}\mathbf{F}'$  where  $\mathbf{F}$  is the 2 × 2 lower triangular matrix (Cholesky factor):

$$\mathbf{F} = \begin{bmatrix} \sigma_1 & 0\\ \rho \sigma_2 & \sigma_2 (1-\rho)^{1/2} \end{bmatrix},\tag{16}$$

where  $\rho^2 = \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}$ . Then, using (7), IS<sub>i</sub> (i = 1, 2) is given by:

$$IS_{1,non-diag} = \frac{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 \rho^2 + 2\psi_1 \psi_2 \sigma_{12}}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}},$$
(17)

$$IS_{2,non-diag} = \frac{\psi_2^2 \sigma_2^2 - \psi_2^2 \sigma_2^2 \rho^2}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}}.$$
(18)

When the ordering of prices is reversed,  $\mathbf{P}_t = (p_{2,t}, p_{1,t})'$ , the subscripts 1 and 2 get reversed in the above expressions. Inspection of (17) and (18) shows that the highest (lowest)  $\mathrm{IS}_i$ value occurs when the price of asset *i* (or market *i*) is ordered first (last) in  $\mathbf{P}_t$ . This produces the upper and lower bounds for  $\mathrm{IS}_i$  based on the ordering of prices. To get a unique value for  $\mathrm{IS}_i$ , Bailie et al. (2002) proposed to use the mean of the upper and lower bounds of  $\mathrm{IS}$ :

$$IS_{1,mean} = \frac{\psi_1^2 \sigma_1^2 + (\psi_2^2 \sigma_2^2 - \psi_1^2 \sigma_1^2) \rho^2 / 2 + \psi_1 \psi_2 \sigma_{12}}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}},$$
(19)

$$IS_{2,mean} = \frac{\psi_2^2 \sigma_2^2 + (\psi_1^2 \sigma_1^2 - \psi_2^2 \sigma_2^2) \rho^2 / 2 + \psi_1 \psi_2 \sigma_{12}}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}}.$$
 (20)

From equation (14), the PDS<sub>i</sub> (i = 1, 2) values for non-diagonal  $\Sigma$  are:

$$PDS_{1,non-diag} = \frac{\psi_1^2 \sigma_1^2 + \psi_1 \psi_2 \sigma_{12}}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}},$$
(21)

$$PDS_{2,non-diag} = \frac{\psi_2^2 \sigma_2^2 + \psi_1 \psi_2 \sigma_{12}}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}}.$$
(22)

From equations (17) - (22), we make the following observations when  $\Sigma$  is non-diagonal. First,  $PDS_i$  distributes the covariance contributions to the permanent shock variance more evenly across assets than IS does. Second, the calculation of  $PDS_i$  is invariant to the ordering of prices. However, for n > 2, the calculation of the upper and lower bounds of IS<sub>i</sub> requires recalculation for all the possible orderings of prices. Also, for each ordering, we will get different values of  $IS_i$  from which we have to pick the highest and the lowest value in order to define the range of IS. Third, it is possible for  $PDS_i$  to be negative. This can happen if  $\psi_i$  is negative and  $\beta_i$  is positive in equation (14) and vice-versa. However, it is unusual for either  $\psi_i$  or  $\beta_i$  to be negative. It can be shown (cf. Zivot and Yan, 2010) that  $\psi \propto \alpha_{\perp}$ where  $\alpha_{\perp}$  is a 2 × 1 vector such that  $\alpha'_{\perp}\alpha = 0$  and  $\alpha = (\alpha_1, \alpha_2)'$  is the 2 × 1 vector of error correction coefficients from the VECM in (8) when n = 2. In typical applications,  $\alpha_1$ and  $\alpha_2$  have opposite signs so that  $\psi_1$  and  $\psi_2$  are both positive. However, it is possible to have a stable VECM with  $\alpha_1$  and  $\alpha_2$  having the same sign. In that case,  $\psi_1$  and  $\psi_2$  will have opposite signs and value of of  $PDS_i$  will be negative. On the other hand, if  $\psi_1$  and  $\psi_2$ have the same sign, then  $\beta_i = \operatorname{cov}(e_{it}, \eta_t^P) = \psi_1 \sigma_1^2 + \psi_2 \sigma_{12}$  can still be negative if  $\sigma_{12}$  is a sufficiently large negative number.<sup>3</sup>

#### 3.2 Comparing PDS to MIS

Lien and Shrestha (2009) define a modified IS (MIS) measure to eliminate the order dependence problem of IS:

$$\mathrm{MIS}_{i} = \frac{\left([\boldsymbol{\psi}'\mathbf{F}^{*}]_{i}\right)^{2}}{\boldsymbol{\psi}'\boldsymbol{\Omega}\boldsymbol{\psi}} = \frac{\left(\psi_{i}^{*}\right)^{2}}{\sum_{i=1}^{n}(\psi_{i}^{*})^{2}},$$
(23)

<sup>&</sup>lt;sup>3</sup>In the risk management context, an asset's contribution to portfolio volatility can be negative if it has a negative weight in the portfolio or if its beta with respect to the portfolio is negative (natural risk reducer). In the latter case the asset is negatively correlated with the portfolio.

where  $\psi^{*'} = \psi' \mathbf{F}^*$ ,  $\mathbf{F}^* = [\mathbf{G} \mathbf{\Lambda}^{-1/2} \mathbf{G}' \mathbf{V}^{-1}]^{-1}$  with **G** being a matrix with eigenvectors of the correlation matrix of the reduced-form residuals as columns, and  $\mathbf{\Lambda}$  representing the diagonal matrix with the corresponding eigenvalues as diagonal elements.

In the bivariate case it can be shown that

$$\mathbf{G} = \begin{pmatrix} \frac{1}{\sqrt{2}} & \frac{1}{\sqrt{2}} \\ \frac{1}{\sqrt{2}} & -\frac{1}{\sqrt{2}} \end{pmatrix}, \quad \mathbf{\Lambda} = \begin{pmatrix} 1+\rho & 0 \\ 0 & 1+\rho \end{pmatrix},$$

and

$$\mathbf{F}^* = \begin{pmatrix} 0.5(\sqrt{1+\rho} + \sqrt{1-\rho})\sigma_1 & 0.5(\sqrt{1+\rho} - \sqrt{1-\rho})\sigma_1 \\ 0.5(\sqrt{1+\rho} - \sqrt{1-\rho})\sigma_2 & 0.5(\sqrt{1+\rho} + \sqrt{1-\rho})\sigma_2 \end{pmatrix}$$

Substituting the relevant terms into (23) gives the following analytic formula for MIS:

$$MIS_{i} = \frac{\psi_{i}^{2}\sigma_{i}^{2}(1+\sqrt{1-\rho^{2}})/2+\psi_{j}^{2}\sigma_{j}^{2}(1-\sqrt{1-\rho^{2}})/2+\psi_{i}\psi_{j}\sigma_{i,j}}{\psi'\Omega\psi},$$
(24)

for i = 1, 2. As the above expression shows, the MIS measure decomposes the variance contribution to each market more equally than the IS measure does and coincides with IS and PDS when  $\rho = 0$ .

As shown in Lien et al. (2022), the difference between the MIS measures across two markets is no less than the corresponding difference between the average IS. Hence, the dominate-satellite relationship between markets is more prominent when using MIS to measure price discovery than when using IS-mean:

$$|\mathrm{MIS}_{1} - \mathrm{MIS}_{2}| = \sqrt{1 - \rho^{2}} |\frac{\psi_{1}^{2} \sigma_{1}^{2} - \psi_{2}^{2} \sigma_{2}^{2}}{\psi' \Omega \psi}| \ge |\mathrm{IS}_{1,mean} - \mathrm{IS}_{2,mean}| = (1 - \rho^{2}) |\frac{\psi_{1}^{2} \sigma_{1}^{2} - \psi_{2}^{2} \sigma_{2}^{2}}{\psi' \Omega \psi}|.$$

Here, we can show that difference between PDS across the two markets is even larger than the corresponding difference between MIS:

$$|PDS_{1} - PDS_{2}| = |\frac{\psi_{1}^{2}\sigma_{1}^{2} - \psi_{2}^{2}\sigma_{2}^{2}}{\psi'\Omega\psi}| \ge |MIS_{1} - MIS_{2}| = \sqrt{1 - \rho^{2}}|\frac{\psi_{1}^{2}\sigma_{1}^{2} - \psi_{2}^{2}\sigma_{2}^{2}}{\psi'\Omega\psi}|.$$
 (25)

and we have equality when  $\rho = 0$ . Hence, PDS can provide an even more prominent dominant-satellite relationship in measuring price discovery than MIS and IS-mean.<sup>4</sup>

#### 3.3 Comparison to De Jong-Schotman IS

De Jong and Schotman (2010), using a structural unobserved components model, define an order-invariant IS measure of the form:

$$\mathrm{IS}_j = \gamma_j \beta_j,\tag{26}$$

where  $\beta_j$  is the regression coefficient of the price innovation on the efficient price, and  $\gamma_j$  is the coefficient in the reverse regression. Comparing (26) to (14), we can see that our PDS measure has a similar form as the De Jong-Schotman IS measure.

However, because of the structural unobserved component model adopted in De Jong and Schotman (2010), the correspondence of their IS measure with PDS computed from reduced-form VECM estimates is only clear under specific restrictive assumptions. When the correlation among competing markets' transitory shocks is solely due to the common efficient price innovation (with a diagonal  $\Sigma$  with positive diagonal elements), it can be shown that their IS measure coincides with the PDS under the Beveridge-Nelson normalization rule.

### 4 Applications to simulated market data

In this section, we examine how data frequency might affect the accuracy of various price discovery measures. We generate data from the following partial adjustment model used in Yan and Zivot (2010):

$$p_{it} = p_{i,t-1} + \delta_i (m_t - p_{i,t-1}) + b_{0,i}^T \eta_t^T, \quad m_t = m_{t-1} + \eta_t^P$$
(27)

<sup>&</sup>lt;sup>4</sup>All three price discovery measures (PDS, MIS, IS-mean) will give the same qualitative leadership results for the bivariate case. For example, when  $\psi_1^2 \sigma_1^2 > \psi_2^2 \sigma_2^2$ , all these measures will take a larger value for the first market, indicating the first market to be the dominant market. Even though these three measures tend to yield the same leadership results, PDS will give the most prominent lead-lag relationship estimates.

for i = 1, 2 with  $\eta_t = (\eta_t^P, \eta_t^T)'$  being Guassian white noise with covariance matrix  $diag(\sigma_P^2, \sigma_T^2)$ . For simplicity, we assume both markets' responses to transitory shocks are equal and set as  $(b_{0,1}^T, b_{0,2}^T) = (0.5, -0.5)$ . The variance of the permanent shock is set as unity  $(\sigma_P^2 = 1)$ . To make the reduced-form residuals uncorrelated for these two markets, we further set the variance of the transitory shock as  $\sigma_T^2 = \frac{\delta_1 \delta_2}{-b_{0,1}^T b_{0,2}^T}$ . We set  $\delta_2 = 1 - \delta_1$  and let  $\delta_1$  take values from 0.9 to 0.1 with a reduction of 0.1. When  $\delta_1 > \delta_2$  (i.e.,  $\delta_1 > 0.5$ ), Market 1 has a greater speed of price discovery than Market 2.

For each parameterization, we simulate 1000 samples with a sample size of 21600, mimicking the 1 second-level data for a trading day. For each simulated 1-second sample (1s), we resample the data at 5-second (5s), 10-second (10s), 30-second (30s), 1-minute (1min), and 5-minute (5min) intervals, respectively. With each re-sampled data, we re-estimate the empirical VECM and calculate price discovery estimates. Averages of each price discovery measure over the 1000 samples are summarized in Table 1.5

#### [Insert Table 1 about here.]

As the results show, IS-mean, MIS and PDS yield different estimates when the data is sampled at a frequency lower than 1-second, even though the original data generating process involves uncorrelated reduced-form errors. Data sampled at a more coarse frequency produces correlated reduced-forms errors and causes the price discovery measures to differ. As the data frequency decreases from 1-second to 5-minute (from Panel A to Panel F), we see that IS-mean estimates become less informative as a leader identifier. For the case with  $\delta_1 = 0.9$ , the IS-mean estimate of Market 1 decreases from 0.9 to 0.5 when data frequency decreases from 1-second to 5-minute. We find the same deterioration of the MIS's performance as data become coarse. The PDS measure performs much better than these two measures. For 5-minute data, PDS can correctly identify Market 1 as the leader for 90% of the time when  $\delta_1 = 0.9$ , compared with 50% of IS-mean and MIS measures. The simulations show that PDS is more resilient to data frequency than IS-mean and MIS.

<sup>&</sup>lt;sup>5</sup>To save space, we only keep two numbers after the decimal point. Each row of Table 1 corresponds to a specific value of  $\delta_1$ .

# 5 Conclusion

In this paper we propose a new measure, PDS, for quantifying a market's contribution to price discovery that is closely related to IS but is unique and order invariant. Our measure is equal to IS and MIS when reduced-form innovations are uncorrelated. In the bivariate case, we show that our PDS measure can provide more prominent lead-lag relationship estimates than MIS and IS do. Moreover, our PDS measure is more robust to samples with lower frequencies. Our expectation is that PDS will be adopted widely in the future discourse on price discovery.

### References

- Baillie, R. T., Booth, G. G., Tse, Y., Zabotina, T.,2002, Price discovery and common factor models, Journal of Financial Markets 5, 309-321.
- [2] Beveridge, S., Nelson, C. R., 1981, A new approach to the decomposition of economic time series into permanent and transitory components with particular attention to the measurement of the 'business cycle', Journal of Monetary Economics 7, 151-174.
- [3] Bruder, B., Roncalli, T., 2012, Managing Risk Exposure Using the Risk Budgeting Approach, Working Paper, Lyxor Asset Management.
- [4] De Jong, F., Schotman, P.C., 2010, Price discovery in fragmented markets, Journal of Financial Econometrics, 8(1), pp.1-28.
- [5] Grammig, J., Peter, F. J., 2013, Tell-Tale Tails: A data driven approach to estimate unique market information shares, Journal of Financial and Quantitative Analysis, Vol. 48, pp. 459-488.
- [6] Hasbrouck, J., 1995, One Security, Many Markets: Determining the Contribution to Price Discovery, Journal of Finance, Vol- 50, No. 4, 1175-1199.
- [7] Lien, D., Shrestha, K., 2009, A New Information Share Measure, The Journal of Futures Markets, Vol 29, No. 4, 377-395.
- [8] Lien, D., Shrestha, K., 2014, Price discovery in interrelated markets. Journal of Futures Markets, 34(3), pp.203-219.
- [9] Lien, D., Shrestha, K. and Lee, L.M.Q., 2022. Analytical properties of Hasbrouck and generalized information shares. Finance Research Letters, 49, p.103185.
- [10] Yan, B., Zivot, E., 2010, A structural analysis of price discovery measure, Journal of Financial Markets 13, 1-19.

Table 1: Data Frequency with Equal Transitory Responses and Uncorrelated Residuals This table reports price discovery measure estimates from the price data simulated from the following 2-market model:

$$p_{1t} = p_{1,t-1} + \delta_1(m_t - p_{1,t-1}) + b_{0,1}^T \eta_t^T,$$
  
$$p_{2t} = p_{2,t-1} + \delta_2(m_t - p_{2,t-1}) + b_{0,2}^T \eta_t^T,$$

where  $m_t = m_{t-1} + \eta_t^P$ ,  $\boldsymbol{\eta}_t = (\eta_t^P, \eta_t^T)'$  are Guassian white noise with diagonal covariance matrix  $diag(\sigma_P^2, \sigma_T^2)$ . The simulation parameterization is set as  $\delta_2 = 1 - \delta_1$ ,  $(b_{0,1}^T, b_{0,2}^T) = (0.5, -0.5)$ ,  $\sigma_P^2 = 1$ ,  $\sigma_T^2 = \frac{\delta_1 \delta_2}{-b_{0,1}^T b_{0,2}^T}$ .

	Panel A: Frequency=1s							Panel B: Frequency=5s						Panel C: Frequency=10s					
	IS-mean		MIS		PI	PDS		IS-mean		MIS		PDS		IS-mean		MIS		PDS	
$\delta_1$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	
0.90	0.90	0.10	0.90	0.10	0.90	0.10	0.75	0.25	0.82	0.18	0.93	0.07	0.64	0.36	0.75	0.25	0.94	0.06	
0.80	0.80	0.20	0.80	0.20	0.80	0.20	0.64	0.36	0.72	0.28	0.83	0.17	0.58	0.42	0.67	0.33	0.86	0.14	
0.70	0.70	0.30	0.70	0.30	0.70	0.30	0.58	0.42	0.64	0.36	0.73	0.27	0.55	0.45	0.61	0.39	0.76	0.24	
0.60	0.60	0.40	0.60	0.40	0.60	0.40	0.54	0.46	0.57	0.43	0.62	0.38	0.52	0.48	0.55	0.45	0.64	0.36	
0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	0.50	
0.40	0.40	0.60	0.40	0.60	0.40	0.60	0.46	0.54	0.43	0.57	0.38	0.62	0.48	0.52	0.45	0.55	0.36	0.64	
0.30	0.30	0.70	0.30	0.70	0.30	0.70	0.42	0.58	0.36	0.64	0.27	0.73	0.45	0.55	0.39	0.61	0.24	0.76	
0.20	0.20	0.80	0.20	0.80	0.20	0.80	0.36	0.64	0.28	0.72	0.16	0.84	0.42	0.58	0.33	0.67	0.13	0.87	
0.10	0.10	0.90	0.10	0.90	0.10	0.90	0.25	0.75	0.17	0.83	0.07	0.93	0.36	0.64	0.25	0.75	0.06	0.94	
	Panel D: Frequency=30s					Pane	Panel E: Frequency=1min						Panel F: Frequency=5min						
	IS-mean		MIS		PI	PDS		IS-mean		MIS		PDS		IS-mean		MIS		PDS	
$\delta_1$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	$p_{1t}$	$p_{2t}$	
0.90	0.56	0.44	0.67	0.33	0.96	0.04	0.53	0.47	0.63	0.37	0.97	0.03	0.51	0.49	0.55	0.45	0.90	0.10	
0.80	0.53	0.47	0.61	0.39	0.89	0.11	0.52	0.48	0.58	0.42	0.91	0.09	0.50	0.50	0.53	0.47	0.79	0.21	
0.70	0.52	0.48	0.57	0.43	0.79	0.21	0.51	0.49	0.55	0.45	0.81	0.19	0.50	0.50	0.52	0.48	0.70	0.30	
0.60	0.51	0.49	0.53	0.47	0.65	0.35	0.50	0.50	0.53	0.47	0.67	0.33	0.50	0.50	0.50	0.50	0.53	0.47	
0.50	0.50	0.50	0.50	0.50	0.49	0.51	0.50	0.50	0.50	0.50	0.51	0.49	0.50	0.50	0.49	0.51	0.36	0.64	
0.40	0.49	0.51	0.47	0.53	0.34	0.66	0.50	0.50	0.48	0.52	0.36	0.64	0.50	0.50	0.49	0.51	0.28	0.72	
0.30	0.48	0.52	0.43	0.57	0.20	0.80	0.49	0.51	0.45	0.55	0.21	0.79	0.50	0.50	0.47	0.53	0.15	0.85	
0.20	0.47	0.53	0.39	0.61	0.10	0.90	0.48	0.52	0.42	0.58	0.10	0.90	0.50	0.50	0.46	0.54	0.05	0.95	
0.10	0.44	0.56	0.33	0.67	0.04	0.96	0.47	0.53	0.37	0.63	0.04	0.96	0.49	0.51	0.44	0.56	0.04	0.96	

Notes: Numbers shown are averages of price discovery measure estimates of 1000 samples. For the 1s sample, the sample size is N = 21600. We resample the 1s sample at the 5-second, 10-second, 30-second, 1-minute, 5-minute intervals to generate the 5s, 10s, 30s, 1min, and 5min samples, respectively.