Price Discovery Share-An Order Invariant Measure of Price Discovery with Application to Exchange-Traded Funds*

Syed Galib Sultan†

Eric Zivot‡

February 5, 2015

Abstract

Price Discovery is the process by which new information is impounded into asset prices through trading activity. A market is considered to contribute more to price discovery if it is the first to capture new information regarding the fundamental value of an asset. Hasbrouck's (1995) information share (IS) is the most widely used measure for price discovery contribution even though there is a well-documented concern with identification: its dependence on the ordering of the variable in the price vector and its non-uniqueness. We propose a new measure, "Price Discovery Share" (PDS) that is closely related to IS and resolves the identification problem inherent in the IS method. PDS is motivated by a widely used method in risk management literature called the "risk-budgeting" or additive decomposition of portfolio volatility. Using simulated data based on different structural asset pricing models, we find that PDS measures the structural price discovery contribution more accurately than IS. We also apply PDS to investigate the "duplication of Exchange-Traded Funds (ETFs)" phenomenon, a recent institutional trend in financial markets. We show that although there are multiple ETFs tracking the S&P 500 index, one specific S&P 500 ETF ('SPY') always contributes more to price discovery than the rest. We also find that PDS, unlike IS, is robust to the use of intra-day market price data sampled at different frequencies.

Keywords: price discovery, information share, Exchange-traded Funds (ETFs)
JEL classification: C32; G10

---

*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 and seminar participants at the Annual Conference of Canadian Economic Association (2014)for helpful comments. Financial support from the Robert Richards Chair is greatly appreciated.

†Department of Economics, Uninversity of Washington, Seattle, WA 98195, USA. E-mail: sgsultan@uw.edu

‡Department of Economics, Uninversity of Washington, Seattle, WA 98195, USA. E-mail: ezivot@uw.edu

1 Introduction

New market trends have made price discovery, the process by which new information is impounded into asset prices through trading activity, an important research agenda in the financial economics literature. Recent features of financial markets include the trading of identical stocks in multiple venues ("market fragmentation") and the trading of closely related assets (eg. derivatives, future and spot, ETFs tracking the same market index etc.) in the same or different venues. Hasbroucks (1995) information share (IS) is the most widely used measure to identify and quantify the process of price discovery. [1] The measure, IS, is typically interpreted as identifying "who moves first in the process of price adjustment" when a new trade related information or permanent shock is received. IS of a given market is defined as the contribution of that market to the innovation variance of the "efficient" price or the random walk component of asset price. [2]

Since its inception, IS has been used by numerous studies in different financial market related contexts. [3] However, it is also well-documented that IS has a potentially serious identification problem when idiosyncratic innovations to different market prices in Hasbrouck's (1995) model are contemporaneously correlated. When the correlation is significantly high, the IS measure, which is typically reported as a range, becomes very wide and it proves impossible to identify the price/information leader or the follower and their individual contribution to price discovery. The limitation is also referred in previous literature as the 'order-dependence problem of IS'. This is because the upper and the lower bound of the range that IS reports depend on the order that the prices enter into the vector of prices. [4]

[1] Special Issue on Price Discovery in the Journal of Financial Markets (2003) is an excellent source for a review of price discovery measures.

[2] According to Hasbroucks (1995) model, intraday stock price at any time of the day is assumed to contain two components- a permanent component and a transient component. The permanent component is called the "efficient" or "fundamental" price of asset. It is assumed to have martingale property and to follow random walk. Any change in "Efficient" price reflects the change in price due to the arrival of new trade-related information into the stock market.

[3] IS has been applied to cross-listed stock to determine the information or price leadership among different stock exchanges (Hasbrouck, 1995; Huang, 2002; Harris et al.,2002) . IS has also been used to determine the information or price leadership between quotes and trade prices of stock (Hasbrouck, 2002), between stock options and underlying stocks (Chakravarty et al., 2004), between futures and their spots (Mizarch and Neely, 2008; Lien and Shrestha 2009), among Credit Default Swap (CDS), bonds and stocks (Grammig and Peter 2014) etc. Also, a brief summary of different studies which use IS can be found in pp- 78 of Putnins (2013).

[4] An extensive discussion regarding this can be found in the section II of this paper.

Numerous studies have proposed different solutions and measures to address this shortcoming of IS. However, no consensus has emerged so far because all of these approaches (Hasbrouck, 1995; Baillie et al, 2002; Lien and Shrestha, 2009 and Grammig and Peter, 2014) have been found to be either effective in particular context or to have their own identification issues[5].

In this paper we address the identification problem of IS and propose an alternative statistical methodology for estimating price discovery which produces a unique measure and is also order-invariant. Our new measure of price discovery is closely related to IS and our methodology is motivated by a widely used method in risk management literature called the additive decomposition of portfolio volatility. [6] A notable feature of Hasbrouck's (1995) model is that the 'volatility of the efficient price innovation' (VEPI) is linearly homogeneous in the common factor weights of each market's innovation. We use this property and apply Euler's theorem to additively decompose the VEPI into components attributed to each market. Each of these components is defined as the contribution of each market to VEPI. Moreover, a key component of this decomposition is what we call the *"price discovery beta"* of a market. *"Price discovery beta"* of a market is the regression coefficient of a market's innovation on the efficient price innovation. We convert the calculated market contributions to the market shares by dividing these contributions by the VEPI. Our new measure of price discovery for each market is this contribution share which we call *"price discovery share" (PDS)*. PDS is also applicable to the general n-assets or n-markets model. As a special case, we demonstrate the analytical comparison between IS (the upper and lower bound), IS-mean

[5]Hasbrouck (1995) suggests sampling the trade and quote prices at a high enough frequency such that contemporaneous correlation among the innovations becomes negligible. However, numerous studies including ours find that even with the use of bid-ask quotes 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 as a unique measure of IS. This approach, while intuitively appealing, is ad hoc. Lien and Shrestha (2009) correctly point out that the average of the two bounds of IS cannot be derived as a result of any particular factor structure. In addition, the calculated means or mid-points of the estimated individual price discovery contributions often do not add up to 100% in applications with more than two prices. Lien and Shrestha (2009) alternatively propose a modified information share (MIS) measure that is derived from the squared root of the eigenvalues of the innovation correlation matrix. A limitation of this approach is that it considers only the positive value of squared root of the eigenvalues in order to reach a unique result. If the negative values of the squared root of the eigenvalues are considered, then it would produce a different value for the MIS. Grammig and Peter (2014) propose another unique measure for IS which is derived by exploiting two properties of price changes -fat tails and tail-dependence.

[6]See Bruder and Roncalli (2012) for a nice description of "risk budgeting" or the additive decomposition of portfolio volatility and use of Euler theorem to determine individual risk contribution of assets in a portfolio.

and PDS in a simple bi-variate case. We also compare IS and PDS using simulated market data. We generate simulated asset price data following four different structural asset pricing models and compare the moments of IS and PDS. In every case, PDS is found to estimate the true structural price discovery contribution more accurately than IS.[7]

Another contribution of this paper to price discovery literature is embedded in our application of PDS. In the second half of the paper, we apply PDS for an empirical investigation of price discovery in the market of exchange-traded funds (ETFs) that track S&P 500 index. ETFs are defined as securities that track the performance of market index, commodity and bonds like an index fund. ETFs can be traded like close-end mutual funds, that is, they can be traded throughout the day like a stock. ETFs are usually highly liquid assets with high tax-efficiency and very low expense ratio compared to mutual funds. All these favorable features have made ETFs extremely popular among stock market investors.

After the Flash Crash of May 6, 2010 ETF trading has also caught the attention of regulators and academicians alike. Sharp price falls across a disproportionate number of ETFs during the Flash Crash have been deemed to be responsible for the abrupt market crash.[8] An investigative report by Borkovec, Domowitz, Serbin and Yegerman (2010) finds that price discovery failed "dramatically" for these ETF securities during the flash crash. In this backdrop, understanding the price discovery dynamics in the ETFs has become an important research agenda.

A recent trend in financial market, which is still relatively unexplored by the financial economists is the proliferation of multiple ETFs tracking the same index. In this study, we denote this as "duplication of ETFs". The first order research agenda that rises from this trend is to find the rationale behind the "duplication of ETFs" and its effect on price discovery. In this study, we mainly focus on the effect of ETF duplication on price discovery. We select a particular index ETF, the S&P 500 ETF and study the price discovery process in two nearly identical and competing S&P 500 ETFs. The two S&P 500 ETFs we consider

[7]Due to the popularity of Baillie et al (2002) measure in different literature, we have decided to report this measure along with upper and lower bound of IS in our analysis. We call this measure IS-mean.

[8]According to the joint SEC/CFTC flash Crash Report, "ETFs accounted for 70% of all US-listed securities that declined by 60% or more during the May 6, 2010 Flash Crash". See Borkovec, Domowitz, Serbin and Yegerman (2010)

are SPY (issued by SPDR) and IVV (issued by iShare).[9] We investigate the price discovery pattern across these two competing ETFs during a regular trading week in 2012 and also, during two highly volatile trading days in recent years- the Flash Crash on May 6th, 2010 and the stock market fall on August 8th, 2011 (two days after the US lost its AAA credit rating). SPY is found to be price leader in all instances and particularly, during the Flash Crash, SPY is found to be the absolute price leader with more than 95% contribution to price discovery.

We close our paper with a comparative assessment of IS and PDS in two different empirical settings. First, we use our "duplication of ETF" application to show that even with one-second tick-by-tick quotes, IS can report very misleading results with wide ranges of price discovery contribution. By contrast, PDS always provides a clean decomposition and a unique value of price discovery contribution. Second, we use quotes of a single day (Dec 3rd, 2012) of a cross-listed stock ("SPY") to calculate price discovery contribution across two stock exchanges (NASDAQ and BATS). We compare the estimate of IS and PDS using quotes sampled at different frequencies. IS reports misleading results due to its serious identification problem. In contrast, PDS results are consistent and also robust to the use of quotes data with higher time intervals.

The remainder of the paper is organized as follows. In Section II we describe the reduced-form cointegration framework used by Hasbrouck (1995) for modeling price discovery in arbitrage linked market and also define our new measure of price discovery, PDS. In Section III, we compare PDS to IS using simulated market data generated from different structural models of asset prices. Section IV presents our empirical application in S&P 500 ETFs and evaluates the effect of "duplication of ETFs" on price discovery. In Section V, we compare the performance of IS and PDS in two different empirical settings. Finally, we conclude with a brief summary of the paper's findings and provide some guidelines for future research.

[9]SPY is the largest, oldest and most popular ETF of its kind which was issued by SPDR State Street Global Advisors for public trading in 1993. IVV is the second most popular ETF that tracks S&P 500 index. It was first issued in 2000 by iShares. More information regarding these ETFs can be found in section 4.1.

2 Model description

We use the arbitrage linked cointegration model approach of Hasbrouck (1995). Let $P_t = (p_{1,t}, ., p_{n,t})'$ denote an $n \times 1$ vector of I(1) log prices. In the price discovery literature, P_t either represents the vector of prices of a single asset that is traded in n market locations and linked by arbitrage or it represents the vector of prices of n similar or closely related and arbitrage linked assets that are being traded in the same market location.

It is assumed that there is a common stochastic component or fundamental value that drives all prices. As a result, there are $n-1$ cointegrating vectors θ_i such that, $\theta_i' p_t \sim I(0)$. Furthermore, since the difference between any two prices of P_t is I(0), it is convenient to use the following $(n-1) \times n$ matrix of rank n-1 as a basis for the cointegrating space:

$$\Theta' = \begin{bmatrix} \theta_1' \\ \vdots \\ \vdots \\ \theta_{n-1}' \end{bmatrix} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 \\ 1 & 0 & -1 & \cdots & 0 \\ \vdots & \vdots & \ddots & \cdots & \vdots \\ 1 & 0 & 0 & \cdots & -1 \end{bmatrix} = \begin{pmatrix} 1_{n-1} & \vdots & -I_{n-1} \end{pmatrix} \tag{1}$$

where 1_{n-1} is an $(n-1) \times 1$ vector of ones and I_{n-1} is the identity matrix of dimension n-1.

Since ΔP_t is I(0), it has a Wold representation:

$$\Delta P_t = \Psi(L) e_t = e_t + \Psi_1 e_{t-1} + \Psi_2 e_{t-2} + ... \tag{2}$$

where $\Psi(L) = \sum_{k=0}^{\infty} \Psi_k L^k, \Psi_0 = I_n$, $e_t = (e_{1t}, ., e_{nt})'$ and $e_t \sim iid(0, \Sigma)$. It is assumed that elements of $\Psi(L)$ are 1-summable and $\Psi(1) \neq 0$. Also,

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_n^2 \end{bmatrix} \tag{3}$$

Using the Beveridge-Nelson decomposition (Beveridge and Nelson, 1981), we can write:

$$P_t = P_0 + \Psi(1) \sum_{j=0}^{t} e_j + \Psi^*(L) e_t \tag{4}$$

where P_0 is $n \times 1$ vector of initial values, $\Psi(1) = \sum_{k=0}^{\infty} \Psi_k$, $\Psi^*(L) = \sum_{k=0}^{\infty} \Psi_k^*$ and $\Psi_k^* = -\sum_{j=k+1}^{\infty} \Psi_j$. Also, $\Psi^*(L)e_t \sim I(0)$ and $\theta'\Psi(1) = 0$. The restriction $\theta'\Psi(1) = 0$ implies that the $n \times n$ matrix, $\Psi(1)$ has rank one and can be expressed as

$$\Psi(1) = 1_n \psi' = \begin{bmatrix} \psi_1 & \cdots & \psi_n \\ \vdots & \ddots & \vdots \\ \psi_1 & \cdots & \psi_n \end{bmatrix} \tag{5}$$

where $\psi = (\psi_1, ..., \psi_n)'$ is an $n \times 1$ vector. The matrix $\Psi(1)$ contains the cumulative impacts of the innovation e_t on all future price movements, and thus measures the long-run impact of e_t on prices. Since the rows of $\Psi(1)$ are identical, the long-run impact of e_t on each price is identical. Substituting (5) into (4) gives us

$$P_t = P_0 + 1_n \sum_{j=0}^{t} \eta_j^P + \tilde{\varepsilon}_t \tag{6}$$

Here, $\tilde{\varepsilon}_t = \Psi^*(L)e_t$ is an I(0) pricing error vector, $\eta_t^P = \psi' e_t = \sum_{k=1}^{n} \psi_k e_{kt}$ and $\sum_{j=0}^{t} \eta_j^P$ is the random walk component that is common to all prices. The following representation of equation (6) makes the role of η_t^P easier to interpret:

$$P_t = P_0 + 1_n m_t + \tilde{\varepsilon}_t \tag{7}$$

where, $m_t = m_{t-1} + \eta_t^P$. And $\eta_t^P = \psi' e_t$ is therefore the innovation to the random walk component of the price. Hasbrouck (1995) describes η_t^P as the component of the price change that is permanently impounded into the price due to new information. Transient pricing errors such as bid-ask bounces and inventory adjustments are absorbed by the I(0) component $\tilde{\varepsilon}_t$.

2.1 Hasbroucks Information Share

Hasbrouck (1995) defines a market's price discovery contribution of i th market as its contribution to the permanent shock variance, $var(\eta_t^p) = \psi' \Sigma \psi$.

In practice, $IS_i (i = 1, ..., n)$ is computed from the estimated parameters of an empirical VECM (K-1) for asset prices:

$$\Delta P_t = A(\Theta' P_{t-1} - \mu) + \sum_{j=1}^{K-1} \Gamma_j \Delta P_{t-j} + e_t \tag{8}$$

where Γ_j is an $n \times 1$ matrix. The lag length, K, is typically chosen by some model selection criterion such as BIC or AIC. Because the cointegrating matrix Θ' 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 and the estimation coefficients (A, Θ and Γ_js) from the VECM:

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

where $\Theta_\perp$ and$A_\perp$ are vector satisfying $\Theta'\Theta_\perp = 0$ and $A'A_\perp = 0$, respectively. Also, $\Gamma(1) = I_n - \sum_{j=1}^{K-1} \Gamma_j$.

Information share (IS) measure for price discovery is defined as follows:

Case 1. Σ is diagonal:

$$IS_i = \frac{(\psi_i \sigma_i)^2}{\psi' \Sigma \psi}, i = 1, ..., n \tag{10}$$

Case 2. Σ is non-diagonal:

$$IS_i = \frac{((\psi'_i F)_i)^2}{\psi' \Sigma \psi}, i = 1, ..., n \tag{11}$$

where $(\psi'_i F)_i$ is the i-th element of $\psi'_i F$ and F is a lower triangular matrix (Cholesky factor) such that $FF' = \Sigma$. The value of F and hence, also the value of IS_i, depends on the ordering in which the individual prices enter into the vector of price, P_t. Therefore, when Σ is non-diagonal 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.

2.2 New Order Invariant Measure of Price Discovery: Price Discovery Share

Our new measure of price discovery is motivated by the additive decomposition of portfolio volatility that is widely used in risk management. Recall, the permanent shock is defined as a weighted average of individual market innovations $\eta_t^P = \psi' e_t = \sum_{k=1}^{n} \psi_k e_{kt}$. The volatility of the permanent shock is $\sigma_\eta(\psi) = (\psi'\Sigma\psi)^{1/2}$. An interesting property of $\sigma_\eta(\psi)$ is linearly homogenous in ψ since $\sigma_\eta(c\psi) = c\sigma_\eta(\psi)$ for any constant c. We apply Euler's theorem and derive the following additive decomposition of $\sigma_\eta(\psi)$:

$$\sigma_\eta(\psi) = \psi' \frac{\partial\sigma_\eta(\psi)}{\partial\sigma} = \sum_{i=1}^{n} \frac{\partial\sigma_\eta(\psi)}{\partial\sigma_i} = \psi_1 \frac{\partial\sigma_\eta(\psi)}{\partial\sigma_1} + ... + \psi_n \frac{\partial\sigma_\eta(\psi)}{\partial\sigma_n} \tag{12}$$

Hence, the volatility of the permanent shock, $\sigma_\eta(\psi)$ 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 (eq. 12), $\psi_i \frac{\partial\sigma_\eta(\psi)}{\partial\psi_i}$, is asset i's (or market i's) contribution to the volatility of the permanent shock. In the spirit of Hasbrouck's information share, $\psi_i \frac{\partial\sigma_\eta(\psi)}{\partial\psi_i}$ 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 $\sigma_\eta(\psi)$

$$PDS_i = \frac{\psi_i \frac{\partial\sigma_\eta(\psi)}{\partial\psi_i}}{\sigma_\eta(\psi)} \tag{13}$$

By construction $\sum_{i=1}^{n} PDS_i = 1$. A small amount of mathematical derivation gives us the following result :

$$\frac{\partial\sigma_\eta(\psi)}{\partial\psi_i} = \frac{\Sigma\psi}{\sigma_\eta(\psi)} = \sigma_\eta(\psi)\beta \tag{14}$$

where $\beta = (\beta_1, ..., \beta_n)' = \frac{\Sigma\psi}{\sigma_\eta^2(\psi)}$ and $\beta_i = \frac{cov(e_{it}, \eta_t^P)}{var(\eta_t^P)} = \frac{\psi_i\sigma_i^2 + \sum_{j=1}^{n-1} \psi_{j\neq i}\sigma_{ij\neq i}}{\psi'\Sigma\psi}$. We deduce the following analytic expression for PDS_i,

$$PDS_i = \psi_i\beta_i = \frac{\psi_i^2\sigma_i^2 + \sum_{j=1}^{n-1} \psi_i\psi_{j\neq i}\sigma_{ij\neq i}}{\psi'\Sigma\psi} \tag{15}$$

We denote β_i in equation (15) as the "*price discovery beta*" of asset i (or market i). The *price discovery beta* is the slope coefficient from the regression of η_t^P on e_{it} and summarizes the (normalized) covariance contributions of an asset's (or market's) innovation to the variance of the efficient price innovation. Hence, PDS_i can be defined as the asset i's (or market i's) contribution to the volatility of η_t^P weighted by its price discovery beta.

2.3 Comparing PDS to IS in the case of Two Assets/Markets

We consider the case of n = 2, so that $P_t = (p_{1,t}, p_{2,t})'$. This allows us to analytically compare PDS to IS. Under the assumption of uncorrelated innovations or diagonal Σ and using equation (10) and equation (14), we find that IS_i and PDS_i are identical:

$$IS_{i,diag} = \frac{\sigma_i^2 \sigma_i^2}{\psi' \Sigma \psi} = \frac{\sigma_i^2 \sigma_i^2}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2} = \psi_i \beta_i = PDS_{i,diag} \tag{16}$$

However, this is not the case when Σ is non-diagonal. Let, $\Sigma = FF'$ where F is the 2×2 lower triangular matrix (Cholesky factor)which is defined as follows:

$$F = \begin{bmatrix} \sigma_1 & 0 \\ \rho\sigma_2 & \sigma_2(1-\rho)^{1/2} \end{bmatrix} \tag{17}$$

where $\rho^2 = \frac{\sigma_{12}^2}{\sigma_1^2 \sigma_2^2}$.Then, using equation (11), IS_i is given by:

$$IS_{1,non-diag} = \frac{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 \rho^2 + 2\psi_1 \psi_2 \sigma_{12}}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}} \tag{18}$$

$$IS_{2,non-diag} = \frac{\psi_2^2 \sigma_2^2 - \psi_2^2 \sigma_2^2 \rho^2}{\psi_1^2 \sigma_1^2 + \psi_2^2 \sigma_2^2 + 2\psi_1 \psi_2 \sigma_{12}} \tag{19}$$

When the ordering of prices is reversed, that is $P_t = (p_{2,t}, p_{1,t})'$, the subscripts 1 and 2 get reversed in (18)−(19). Inspection of (18) and (19) show that the highest (lowest) IS_i value occurs when price of asset i (or market i) is ordered first (last) in the vector of prices. This gives rise to upper and lower bounds for IS_i based on ordering of prices. To get a unique value for IS_i, Bailie et al. (2002) proposed to use the mean of the upper and lower bounds of IS derived from (18) and (19).

$$IS_{1,Bailie} = \frac{\psi_1^2\sigma_1^2 + (\psi_2^2 - \psi_1^2)\sigma_2^2\rho^2 + \psi_1\psi_2\sigma_{12}}{\psi_1^2\sigma_1^2 + \psi_2^2\sigma_2^2 + 2\psi_1\psi_2\sigma_{12}} \tag{20}$$

$$IS_{2,Bailie} = \frac{\psi_2^2\sigma_2^2 + (\psi_1^2 - \psi_2^2)\sigma_2^2\rho^2 + \psi_1\psi_2\sigma_{12}}{\psi_1^2\sigma_1^2 + \psi_2^2\sigma_2^2 + 2\psi_1\psi_2\sigma_{12}} \tag{21}$$

From equation (13), PDS_i for non-diagonal Σ can be derived as:

$$PDS_{1,non-diag} = \frac{\psi_1^2\sigma_1^2 + \psi_1\psi_2\sigma_{12}}{\psi_1^2\sigma_1^2 + \psi_2^2\sigma_2^2 + 2\psi_1\psi_2\sigma_{12}} \tag{22}$$

$$PDS_{2,non-diag} = \frac{\psi_2^2\sigma_2^2 + \psi_1\psi_2\sigma_{12}}{\psi_1^2\sigma_1^2 + \psi_2^2\sigma_2^2 + 2\psi_1\psi_2\sigma_{12}} \tag{23}$$

From equations (18) - (23), we make the following observations regarding IS, PDS and IS-mean when Σ is non-diagonal. First, PDS_i distributes the covariance contributions of each asset (or market) to the permanent shock variance or $\psi_1\psi_2\sigma_{12}$, evenly across assets (or markets). PDS_1 differs from PDS_2 only due to the difference between $\psi_1^2\sigma_1^2$ and $\psi_2^2\sigma_2^2$. Second, for n $>$ 2, the calculation of the upper and lower bounds of IS_i requires recalculation of equation (11) for all the possible orderings of prices. For example, when n = 5 there are 120 possible orderings of prices which need to be considered. Also, for each ordering, we will get different value of IS from which we have to pick the highest and the lowest value in order to define the range of IS. The calculation of PDS_i is invariant to the ordering of prices. Third, if $\psi_1 = \psi_2$ the mid-point IS_i (i = 1,2) in equation (20) and (21) is equal to PDS_i (i = 1,2) in equation (22) and (23). Fourth, it is possible for PDS_i to be negative. This can happen if ψ_i is negative and β_i is positive in equation (15) and vice-versa. It is unusual for either ψ_i or β_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 $\alpha_\perp'\alpha = 0$. $\alpha = [\alpha_1, \alpha_2]'$ is the 2×1 vector of error correction coefficients from the VECM in equation (8) when n=2. In equation 8, we use the $n \times 1$ version of error correction matrix and express it as "A". In typical applications, α_1 and α_2 have opposite signs so that ψ_1 and ψ_2 are both positive. However, it is possible to have a stable VECM with α_1 and α_2 having the same sign. In that case, ψ_1 and ψ_2 will have opposite signs. On the other hand, if ψ_1 and ψ_2 have the same sign then $\beta_i = cov(e_{it}, \eta_t^P) = \psi_1\sigma_1^2 + \psi_2\sigma_{12}$ can still be negative if σ_{12} is a sufficiently large negative number.[10]

3 Applications to simulated market data

In this section we use simulated market data to provide comparison between IS and PDS. The simulated market data are generated from three different stylized structural models of asset prices described in Hasbrouck (2002). We also propose a modified version of one of these three models as an additional example. Using each of these simulated market data sets, we compute IS (upper and lower bound), IS-mean (average of these two bounds) and PDS. Using Monte Carlo simulations, we report mean, standard errors and 95% confidence intervals of these three price discovery measures.[11]

3.1 Application to simulated market data: two-market "Roll" model

The first example is a model of high-frequency trade prices suggested by Roll (1984). Hasbrouck (2002) uses a simplified two-market version of this model and compare IS with another measure of price discovery called "component share". This model suggests that the trade price at any time during the trading hours has two components. One is the "efficient" or "fundamental" price which has the martingale property and follows a random walk. The other is the transient component which mostly arises from bid-ask bounce, inventory effects, discreteness etc. This model assumes that there are two markets where a cross-listed identical stock is being traded at price p_{1t} in Market 1 and at p_{2t} in Market 2. The common efficient price of this stock is defined as follows,

$$m_t = m_{t-1} + u_t \tag{24}$$

[10]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.

[11]95% Confidence interval is calculated as, $CI_{95} = mean \pm 2 * s.e.$

where, $u_t \sim N(0, \sigma^2)$ There is a common and identical fixed cost per trade in each market denoted by *"c"*. The bid-price at time t is $m_t - c$ and the ask-price at time t is $m_t + c$. The trade direction indicator is denoted by q_t which takes a value equal to 1 if the trader is buying and -1 if the trader is selling. Buys and sells are assumed to be equally likely and serially independent. The traders are also assumed to buy or sell independently of the innovation to efficient price (denoted by u_t). Therefore, trade direction and transaction price are defined as follows,

Trade direction: $q_{it} = \pm 1$, each with probability $\frac{1}{2}$ for i=1,2,

Transaction price: $p_{it} = m_t + cq_{it}$ for market =1,2.

From the set-up of the model, it is clear that two markets are structurally identical and therefore the true contribution of each market to price discovery is 50%. Following Hasbrouck (2002) we set, $c = 1$ and $\sigma_u = 1$. We generate 100,000 observations of transaction prices, p_{1t} and p_{2t}, for each market. We use the prices to estimate the VECM as described in equation (8) and use the estimated parameters and covariance matrix to calculate IS, IS-mean and PDS. We repeat this process 1000 times and calculate the mean, standard error and 95% confidence interval of these 1000 estimated IS, IS-mean and PDS .Table 1 reports the simulation results for Market 1.

[Insert Table 1 here]

From Table 1 we find that both the mean of PDS (50.1%) and IS-mean (50%) of Market 1 are very close to the structural price discovery contribution or 50%. However, the upper and lower bound of IS report a large range of price discovery contribution. The upper bound IS identifies Market 1 as the price leader (with estimated 78.9% contribution to price discovery) and the lower bound IS identifies Market 1 as the price follower (21.2% of estimated contribution to price discovery).

3.2 Application to simulated market data: two markets with private information

We also take our second example from from Hasbrouck (2002). In this model, it is assumed that all the informed trading happen in Market 1. This implies that changes in efficient price are driven only by trading activities in Market 1.The efficient price is defined as following,

$$m_t = m_{t-1} + \lambda q_{1t} \tag{25}$$

where λ is defined as liquidity parameter and assumed to be strictly positive. Trade directions (q_{it}) are defined in similar way as before. The transaction price of Market 2 depends on the lagged value of m_t and this assumption essentially defines the Market 2 as the price follower.

Trade direction: $q_{it} = \pm 1$, each with probability $\frac{1}{2}$ for i=1,2,

Transaction price: $p_{1t} = m_t + cq_{1t}$ and $p_{2t} = m_{t-1} + cq_{2t}$

The structural model suggests that all the price discovery happens in Market 1 and therefore, the structural price discovery contribution of Market 1 is 100%. Following Hasbrouck (2002), we set $c = 1$ and $\lambda = 1$. We again generate 100,000 sample observations, calculate the IS, IS-mean and PDS and repeat this process 1000 times. Table 2 summarizes the moments and 95% confidence interval of each price discovery measure.

[**Insert Table 2 Here**]

From Table 2, we find that in the set-up of Example 2 , all the three price discovery measures estimate the contribution of Market 1 to price discovery as 99.9% which is almost equal to the structural contribution (100%). Therefore, in this setup, we do not find any difference among the three measures and all of them estimate the price discovery contribution with almost perfect accuracy.

3.3 Application to simulated market data: two markets with private and public information

In Example 3, the efficient price, m_t contains a non-trade public information component (u_t) and a private information component (λq_{1t}) which is driven by Market 1s trade. Efficient price is defined as follows,

$$m_t = m_{t-1} + \lambda q_{1t} + u_t \tag{26}$$

where $u_t \sim N(0, \sigma_u^2)$. Trade direction is defined same as before. Trade cost for each market is now different from each other. The transaction price of Market 2 is again assumed to depend on lagged information regarding efficient price.

Trade direction: $q_{it} = \pm 1$, each with pr. $\frac{1}{2}$ for i=1,2,

Transaction price: $p_{1t} = m_t + c_1 q_{1t}$ and $p_{2t} = m_{t-1} + c_2 q_{2t}$

Market 1 is again the price leader in the structural model with 100% contribution to price discovery. The trading cost in Market 1 (c_1) is higher than Market 2 because the cost of market making is higher in Market 1, where all the informed traders are trading. Trades are done cheaply at stale prices in Market 2. For simulation, we set $c_1 = 1$, $c_2 = 0$, $\lambda = 1$ and $\sigma_u = 1$. We conduct the simulation in the same way as before.

[**Insert Table 3 Here**]

Table 3 summarizes the result. The IS reports a upper bound of 98.4% and a lower bound of 90% for price discovery contribution of Market 1. The range in IS in this model is significantly larger than before and suffer from lack of identification. PDS estimate do not have any identification issue and reports a comparatively accurate estimate (96%) of actual price discovery contribution than that of IS-mean (94.2%). It should also be noted that the 95% confidence interval of lower bound of IS, IS-mean and PDS do not contain the true value of price discovery contribution.

3.4 Application to simulated market data: Modified two-market "Roll" Model

All the three models defined earlier are either 50-50 or all-or-nothing situation. In this example, we modify the first model ("Roll" model) in such a way so that it produces a structural price discovery contribution of 70% for Market 1 and 30% for Market 2. We define a binary variable D such that D=1 with probability 0.7 and D = 0 with probability 0.3. The assumption is that efficient price is driven by i.i.d. non-trade information which is revealed contemporaneously only to Market 1 70% of the time and the rest of the time to Market 2. We keep the model very simple by assuming no liquidity effect ($\lambda = 0$) and identical trade cost (c) for both of them. Efficient price is defined the same way as in first example (eq. 23) and the transaction prices are defined as follows:

$$p_{1t} = Dm_t + (1 - D)m_{t-1} + cq_{1t} \tag{27}$$

$$p_{2t} = (1 - D)m_t + Dm_{t-1} + cq_{2t} \tag{28}$$

The structural price discovery shares are 70% for Market 1 and 30% for Market 2. We proceed with the simulation exercise in the same way as before.

[Insert Table 4 Here]

Table 4 summarizes the simulation results. IS upper bound assigns a price discovery contribution of 80.5% to Market 1 (or identify Market 1 as the price leader) and lower bound assigns a contribution of 50.8% (no clear price leader or follower). PDS assigns a price discovery contribution of 67.5% to Market 1 which is very close to its true structural contribution (70%) and the 95% confidence interval of PDS contains the true contribution. IS-mean estimates a contribution of 65.6% contribution which is also close to actual contribution but its 95% confidence interval does not contain the true structural contribution.

4 Empirical Application: Price Discovery in S&P 500 ETF Market

In this section, we apply our new price discovery measure, PDS to quantify and analyze price discovery in the market for ETFs which track the S&P 500 index. In particular, we examine two competing S&P 500 ETFs (SPY and IVV) and discuss the effect of duplication of ETFs on price discovery. We first give an overview of the market for S&P 500 ETFs. We then review the existing literature on S&P 500 ETFs, describe our data and present our results and analysis.

4.1 Price Discovery in S&P 500 Exchange-Traded Funds (ETFs): An Overview

In the past decade, the U.S. stock market has been characterized by a new market phenomena which we call "duplication of ETFs". More precisely, this refers to the proliferation of ETFs that track the identical index. For example, SPY (issued by SPRD), IVV (issued by iShares) and VOO (issued by Vanguard) track S&P 500 index, IWM (issued by iShares), VTWO (issued by Vanguard) and TWOK (issued by SPDR) track Russel 2000 index and QQEW (issued by First Trust) and QQQE (issued by Direxion) track NASDAQ-100 equal weighted index. A natural question to ask is why this duplication is occurring and what effect does this duplication have on the market. In our application, we address the second question from a price discovery perspective. "Duplication of ETFs" shares similarity with "market fragmentation" as far as competition is concerned. Therefore, price discovery analysis of competing ETFs fits agreeably in this line of research.

Among the aforementioned index ETFs, we particularly focus on the S&P 500 ETFs because they are the most popular and highly traded. S&P 500 ETFs are also highly liquid assets with very low expense ratios and command larger portion of market share among other index ETFs. There are currently three ETFs that track S&P 500 index- SPY, IVV and VOO.

[Insert Table 5 here]

Table 5 and 6 provide a brief comparison among SPY, IVV and VOO. SPY was introduced to the market first in 1993 and in fact, was the very first ETF of its kinds. IVV was issued next in 2000 and VOO was introduced very recently in 2010. Table 5 also reveals that they are fairly similar to each other in terms of performance measures. Although VOO is the newest ETF, it is very popular among traders due to its impressively low expense ratio (0.05%). However, VOO still hasn't managed to capture significant market share since SPY and IVV commands almost 91.75% of the total market capitalization.

[Insert Table 6 here]

Analysis of Table 6 reveals the fact that in terms of top 10 holdings SPY and IVV are more similar to each other than VOO. This is also true when we see the sector-wise decomposition of these three ETFs. The market price data also reveals this fact in the sense that at any given time, prices of SPY and IVV are very close to each other whereas price of VOO is significantly different from the rest.

There are two additional features of S&P 500 ETFs that need to be discussed here in order to have a better understanding of the price dynamics of these ETFs. These are the tracking error of ETFs and the arbitrage opportunities in ETF trading.

The S&P 500 index utilizes market capitalization weighing structure to construct its portfolio. However, S&P 500 ETFs cannot exactly replicate these portfolio weights for several reasons. First, there is a copyright issue. And second, portfolio weights of S&P 500 index are constantly changing depending on the change in market capitalization of the underlying assets. But instant portfolio re-balancing for an ETF is costly and therefore, their portfolio weights are not identical to the S&P 500 index. This creates tracking error in the price of each S&P 500 ETF. Since a low tracking error of an ETF makes it more attractive for the ETF investors, all the ETFs issuers have strong incentives to reduce this tracking error as much as possible. Therefore, the tracking errors of each ETF are bounded.

Arbitrage opportunities can be created in ETF trading in two different ways. First, arbitrage between ETF price and its Net Asset Value or NAV (see Ben-David, Franzoni, Moussawi, 2014 ; Madhavan and Sobczyk, 2014).[12][13][14] Second, arbitrage between two similar ETFs which track the identical index (see Marshall et al, 2013). The "Authorized Participants" or the APs can only take advantage of the first kind of arbitrage.[15] If during the closing hour of trading, the ETF price exceeds its NAV, the APs can buy underlying securities of that ETF from the secondary market and submit them to the ETF issuers in exchange of new ETFs in the primary market. The APs can then sell the ETFs in the secondary market at a premium. The APs do exactly the opposite when the price of an ETF falls below its NAV. They buy ETF at discount from the secondary market, redeem the ETFs into its underlying stocks in the primary market and then sell the underlying securities at

[12] NAV per share is computed once a day based on the closing market prices of the underlying securities in the ETF's portfolio.

[13] Ben-David, Franzoni, Moussawi (2014) investigate the effect of ETFs on their underlying stocks and find that stocks owned by ETFs have significantly higher intra-day and daily volatility.

[14] Madhavan and Sobczyk (2014) utilize the arbitrage link between ETF closing price and its NAV and propose a state-space model of ETF price dynamics.

[15] APs are institutional investors of ETFs who usually has legal contract with the ETF issuers.

a profit in the secondary market. In contrast, the arbitrage between two similar or nearly identical ETFs can occur at any time during the trading hours. Any ETF investors (retail or institutional) can take advantage of this. Let us consider the case of SPY and IVV. If price of SPY exceeds the price of IVV at any time during the day, there is a strong incentive for investors who treat both of them as close substitutes, to sell SPY at the high price and buy IVV at the low price. This would allow the investors to sell the newly bought IVV at a premium when the price of IVV finally catches up with SPY.

4.2 Price Discovery in S&P 500 ETFs (SPY and IVV): Previous Literature and Research Agenda

Marshal et al (2013) argues that even though SPY and IVV are not perfect substitutes, investors consider them as close substitutes and when mispricing is allowed, an arbitrage opportunity is created between SPY and IVV. They also point out that one possible source of these ETFs being mispriced is the difference in price discovery between these two assets. More importantly, their findings suggest that the prices of SPY and IVV should not diverge from each other and whenever there is significant dispersion between the two prices, arbitrage opportunities make the two prices converge to each other. As a result, these two prices are co-integrated with co-integrating vector $[1, -1]'$. This also makes them an ideal candidate to analyze their price discovery in Hasbrouck's (1995) VECM setting.

Fang and Sanger (2012) examine price discovery across SPY, IVV and the reconstructed price series for the S&P 500 index. The constructed price series of the S&P 500 index captures the second-by-second price movement in the underlying securities of S&P 500 ETFs. They find that SPY and IVV contributes half of the price discovery share compared to their underlying stocks. They also look into the price discovery across SPY and IVV and find that both of them contribute equally in price discovery. A problem with their findings is the lack of interpretation in their reported results. For example, in Table 1 of their paper, they report that during 2006 Q4, SPY contributes 51.4%, IVV contributes 52.4% and underlying component assets contribute 29.5%. The contributions add up to more than 100%. Similarly, in Table 2, when they look into SPY and IVV, they report SPY contributes 85.5% and IVV contributes 82.4% of the price discovery in the whole sample period. The reason behind this uninformative result is that they follow IS to measure price discovery.

Our analysis is motivated by the results in Marshal et al (2012) and Fang and Sanger (2012). We use PDS to give a clean decomposition of price discovery between SPY and IVV on a selected of set of trading days across a number of exchanges. Our objective is to uncover the price discovery contributions of these two ETFs in normal and unusual trading environments. In other words, we seek to evaluate the relative importance of these two ETFs as a source of information for the traders under different market conditions.

4.3 Price Discovery in S&P 500 ETFs (SPY and IVV): Interpretation of PDS Estimates

When traders are buying or selling either SPY or IVV, they are essentially betting on their perceived forecast regarding future market performance. Any permanent change in S&P 500 ETF prices reflect the arrival and absorption of new information regarding the future movement of the market index. Price discovery across different S&P 500 ETFs describes this process.

To interpret our price discovery measure, it is important to define the set of information available to the traders of S&P 500 ETFs. According to Hasbrouck's (2003), the price vector P_t should be a "comprehensive set of prices" which serves as an information set that all the traders possess at time t and also "a poor proxy for common public information". The observable public information that a typical S&P 500 ETF trader has are the bid and ask price and trade price of the ETFs. We leave out the trade or transaction price from the information set of SPY and IVV in order to reduce the micro-structure noise (e.g. bid-ask bounce) in these prices. Following the previous literature, we only include the bid-ask mid-point of the two ETFs to construct the information set, so that $P_t = [P_{SPY,Bid-Askmid,t}, P_{IVV,Bid-askmid,t}]$. We use this price vector, P_t in the VECM in equation (9) and use the estimated parameters to derive PDS from equation (14).

Given our definition of P_t, it is straightforward to interpret the price discovery share of individual ETF. Suppose, SPY is estimated to have an X% price discovery share (according to IS or PDS), it can be interpreted in following ways:

1. SPY contributes X% of the volatility of the innovation to the common random walk efficient price.

2. If $X\% > 100\%$-$X\%$, then SPY will be dominant in price discovery. In other words, SPY will be the price leader. Therefore, SPY price will be the first to adjust to a new information about the fundamental value. It will be considered as the more important source of information regarding future movement in market index or the overall market performance.

Based on the analysis in Yan and Zivot (2010), Putnin (2013) argues that an asset price which is first to adjust (the price leader) also has the potential to be more noisy than the follower. When the noise level differs significantly in two prices, IS may end up assigning higher price discovery share to the less noisy asset price even though it may not be the price leader. As a result, IS may identify the wrong asset as the price leader. Our measure, PDS also shares this property as IS. However, in our application to S&P 500 ETFs the noise levels in the prices of SPY and IVV are found to be very similar and small in magnitude so that it is unlikely that IS or PDS will misidentify the price leader due to difference in noise level.

4.4 Price Discovery in S&P 500 ETFs (SPY and IVV): Data Description and Descriptive Statistics

We use the NYSE TAQ database as our source of high-frequency quotes for SPY and IVV. We choose two snapshots of data in two distinct trading environment. For a normal trading period, we choose intra-day quotes data from Dec 3rd to Dec 7th in 2012. For an extremely volatile trading period, we choose the quotes data of May 6th in 2010 (the day of Flash Crash) and Aug 8th, 2011 (the day of the worst stock market fall in US since 2008)

For Dec 3rd - Dec 5th, 2012 we collect dataset for the following stock exchanges- BATS, Nasdaq, Arca, CBOE, NSX, Boston, Philadelphia and EDGE A. For May 6th, 2010 and Aug 8th, 2011 we collect dataset for BATS, Nasdaq and Arca. The reason for excluding the rest of the stock exchanges is either SPY or IVV were not traded in those stock exchanges during 2010 or even if they were, then the frequency was too low for a meaningful discussion.

In order to be consistent with the first abnormal period, we also look into quotes only from these three stock exchanges for 8th August, 2011.

[Insert Table 7 here]

Table 7 reports the S&P volatility index, VIX or the "fear index" of these different days. The average of closing price of VIX during the normal trading period is 16.53. In contrast, VIX on May 6th, 2010 was twice (closing VIX=32.8) as large and on Aug 8th, 2011 it was three times as large (closing VIX=48) than the normal period.

On May 6th, 2010 the US stock market experienced an abrupt crash. The abnormal plunge in the market index was first seen at 2:42 pm and the fall in prices continued for next 20 minutes. The Dow Jones Industrial Average experienced the biggest one-day point decline during that day.

August 11th, 2011 is considered to be the worst day in Wall Street since the crisis of 2008. All three major stock market indexes (S&P 500, Dow Jones Industrial Average, NASDAQ composite) fell sharply (between 5% to 7%) during that day. The day was also known for the wide-spread panic among the investors regarding the US losing its AAA credit rating on Aug 6th, 2011.

Tick-by-tick raw trade and quote data typically contain numerous types of data errors. It is required to be thoroughly cleaned prior to being analyzed. We use the data cleaning procedure recommended for the TAQ data described by Barndorff-Nielsen et al (2008) and implemented in the R package highfrequency. Data-cleaning steps for the bid-ask quotes involved the following- 1. Restrict data to exchange hours (9:30 am to 4:30 pm) 2. Delete entries with zero quotes. 3. Delete entries with negative spreads 4. Delete entries if spread maximum*median daily spread 5. Delete entries for which the mid-quote is outlying with respect to surrounding entries 6. Restrict data to a specific exchange for analysis. After this step we get intra-day time-series dataset of SPY quotes and IVV quotes for each stock exchange separately. 7. For each stock exchange dataset, we delete entries with same time stamp and use median quotes. 8. In each stock exchange dataset, when there is a time-stamp with no ask/bid price reported for it, we use the last observed ask/bid price to replace the missing values. Therefore, for each ETF we have 25201 observations for a given stock exchange and for a given day.

[Insert Table 8 here]

Table 8 contains descriptive statistics for the intraday quotes of SPY and IVV in eight different stock exchanges during a normal trading week in December, 2012. We first calculate the average 1 second continuously compounded returns for each day in that week and then report the average of these calculated returns. We do the same for 1 second return volatility, bid-ask spreads and number of shares traded. Table (8) reveals that in terms of return and volatility, in every stock exchange except NSX, both ETFs were performing almost similarly. In the National Stock Exchange (NSX), IVV is found to be abnormally volatile compared to SPY and this irregularity is also captured in the spread of their average returns. NASDAQ, BATS and Arca are the exchanges with highest number of trades per day and also have the lowest average bid-ask spread. CBOE, which has the lowest trades per day, also has the highest average bid-ask spread. In every stock exchanges SPY is much more heavily traded than IVV.

[Insert Table 9 here]

Table 9 reports descriptive statistics of the day of Flash Crash in 2010. Compared to our sample of a normal trading week, the average 1 second return volatility on May 6th,2010 was more than 1000 times higher. The average 1 second returns are also 10 times lower than our previous sample. NASDAQ, BATS and Arca are again the stock exchanges with the highest number of shares traded and lowest average bid-ask spread. In all of these three exchanges the return volatility for IVV is much higher than that of SPY.

[Insert Table 10 here]

Table 10 reports the descriptive statistics of the day of stock market fall in August 8th, 2011. The loss in 1-sec return is on average 100 times larger than that of regular trading period. The 1-sec return volatility is also high for both SPY and IVV in all the three stock exchanges. One interesting point here is that the bid-ask spread of both the ETFs were reasonably lower compared to May 6th, 2010. This indicates that the market turmoil on August 8th did not have any significant effect on the liquidity of SPY or IVV.

[Insert Diagram 1 through 6 here]

Diagram 1 through 6 provide visuals of these three different period. From these diagrams, it is very evident that both the ETF prices move in tandem throughout the day (i.e. highly cointegrated) . The only exception was the short period in the afternoon during the Flash Crash when SPY price plummeted by a significant amount compared to IVV.

4.5 Price Discovery in S&P 500 ETFs (SPY and IVV): Estimation and Results

We estimate and report PDS for each exchange in Table (11) during the two selected periods. For the normal trading period (Dec 3rd Dec 7th, 2012), we first estimate PDS of SPY and IVV in every day for a given stock exchanges and then report the daily average of PDS. We also do the same for abnormal periods- May 6th, 2010 and Aug 8th, 2011.

[Insert Table 11 here]

The third column of Table (11) reports PDS for the normal trading week in December, 2012 in eight stock exchanges. SPY leads IVV in price discovery in every stock exchange considered. On average, SPY contributes 61.25% of the price discovery compared to IVV across all stock exchanges. Interestingly, for the NASDAQ and Philadelphia, IVV price is found to be an almost equally important source of trade-related information as it contributes 47% of the price discovery. Although, the daily average PDS is indicating that SPY is leading IVV everywhere, there are sufficient fluctuations in PDS of SPY if we look into each day separately. Among the 40 cases (eight stock exchanges in five days), IVV contributes more to price discovery than SPY in 9 occasions. The fourth column reports PDS across SPY and IVV on the day of Flash Crash in NASDAQ, BATS and Arca. We find that SPY leads IVV in every stock exchange by a large margin. On an average, in every stock exchange SPY contributed 95% of the price discovery. The fifth column reports the result during August 11, 2011. Here, we again find that in every stock market SPY contributes a major portion to price discovery compared to IVV. On an average, the contribution to price discovery of SPY in each stock exchange was 75%.

We interpret these results in the following way. Our goal is to check whether duplication in S&P 500 ETFs provides the traders with a better source of trade related information. Looking at three snapshots in two different environments, we find the consistent result. That is, although there are more than one ETF tracking S&P 500 index, traders still consider SPY to be a better source of information. This reliance becomes extreme or significantly high in the event of abnormal jumps or falls in intra-day prices (Flash Crash, 2010 or August stock market fall in 2011). Particularly, on the day of Flash Crash, IVV price is found to be very less informative as far as capturing the new information in a timely manner is concerned.

A hypothesis that we think is feasible to explain this outcome is the attributes of the buyers of ETFs. The answer to the question that "Who buys S&P 500 ETFs and Why?" can give us a much better idea to explain the dominance of one particular S&P 500 ETF (namely SPY) in price discovery. If more informed traders choose to trade in one particular S&P 500 ETF most of the time, then that by definition that ETF should contribute more to price discovery. Therefore, we think that the distribution of institutional and retail buyers of S&P 50 ETF and their objectives (eg. hedging against market makers) behind investing in these ETFs can help us understand the dominance of SPY over IVV. In previous literature, liquidity of assets has also been found to be an important factor in explaining the price discovery contribution. However, in this case, we have observed that both ETFs are highly liquid and their liquidity measures like the bid-ask spreads are very close to each other in both normal and volatile conditions. Trading volume is also a key factor behind the dominance of SPY. Analyzing the descriptive statistics tables (Table 8, 9 and 10), we find that IVV is comparatively highly traded in three stock exchanges- NASDAQ, BATS and Arca. However, even in these stock exchanges, during the normal period, the number of shares traded of IVV is only about 5% of that of SPY. During the volatile period, this ratio becomes less than 1%. This complete dominance of SPY over IVV in terms of trading volume can also explain the higher price discovery contribution of SPY.

5 Empirical Application: Comparative Assessment between PDS and IS

In this section we provide a comparative assessment between our measure, PDS and IS. First, we show their difference in a setting which is similar to first empirical application. That is, measuring price discovery across SPY and IVV in different stock exchange. But here we only consider one day from the normal trading period to show our result. Next, we demonstrate a more conventional application of price discovery which is analyzing price discovery of a cross listed stock. We pick SPY as the cross listed asset and measure price discovery between NASDAQ and BATS where it is most heavily traded. We report our result using different time intervals and compare the estimates of PDS and upper bound, lower bound and mean of IS.

5.1 Comparative Assessment between PDS and IS: S&P 500 ETFs

Here we utilize a small portion of our previous application to demonstrate comparison between IS and PDS. One of the major drawbacks of IS, as we have discussed earlier is that it often reports the contribution of a given asset price to the price discovery in a range and this range can be so wide that it becomes almost meaningless. Hasbrouck (1995) recognizes this lack of identification issue with IS and proposes the use of high-frequency data (e.g. quotes at every second) to get a tighter bound for IS. Here we look into the price discovery contributions between SPY and IVV only on December 3rd, 2012 which was a normal trading day and we conduct our analysis in eight different stock exchanges. We find that even with high frequency 1-sec interval quotes data, we can end up with IS with a very large range.

We look into the bid-ask mid-quotes of these as before and estimating the VECM equation to derive the parameters necessary for the calculation of IS and PDS. We use equation (11) to calculate IS and equation (15) to calculate PDS. For the IS calculation, it is important to note which of the two prices was placed first in the price vector. We first put log of bid-ask mid-quotes of SPY as the first element in the price vector and calculate the upper (lower) bound IS for SPY (IVV). We then place log of bid-ask mid-quotes of IVV as the first element and calculate the IS again. This gives us the lower (upper) bound IS for SPY (IVV). We also take the average of these two bounds and report it. We call it "IS-mean" which is a method proposed by Baillie et al (2002) for a unique measure of IS.

[**Insert Table 12 here**]

The 3rd and 4th column of Table (12) reports the upper and lower bound of IS. If the difference between upper and lower bound gets larger, we face the identification problem in determining the price leader out of the two. A closer look at these two columns tells us that in the case NASDAQ, Arca, Boston, Philadelphia and EDGE A, the upper bounds of IS for each ETF identify themselves as price "leaders" and the lower bounds of the same ETF identify themselves as "followers". On the other hand, our measure, PDS gives a clean decomposition, reports a unique value for each ETF and more importantly, identifies the price "leader" correctly in every occasion. The IS-mean also identifies the same leaders as PDS but the contribution reported by IS-mean is found to be under-estimated in every case. Particularly, in the case of BATS and CBOE, IS-mean under-estimates the price discovery contributions by more than 10%. We also estimate the standard error for both measures (IS and PDS) in every case by bootstrap method .[16] For both measures, the reported boot-strap standard errors are very low and we also do not find a significant difference between IS and PDS in terms of standard errors reported.

5.2 Comparative Assessment between PDS and IS: Quotes at higher time interval

Another drawback of IS is that the gap between upper and lower bound gets extremely large if data at higher interval is used for analysis. Gramming et al. (2005), Theissen (2002) and Huang (2002) all of these studies while studying quotes data with more than 1-sec interval find a substantial contemporaneous residual correlation and wide range of upper and lower bounds for the IS.

[16] Bootstrap method is done following Grammig et al (2005). First, VECM parameters are estimated and we derive the estimated residuals from the difference of actual data and fitted data. Estimated residuals are used to simulate price series using original price series as starting values. The price discovery measure is calculated from the simulated data and the whole process is repeated 1000 times to get the mean and standard error of IS and PDS.

[Insert Table 13 here]

Table 13 reports the results of an application where we focus on a single cross-listed stock (SPY) and the price discovery contribution between two competing stock exchanges (NASDAQ and BATS). We start our analysis with the tick-by-tick data (1 second interval) and then continue to calculate IS and PDS at lower frequency observations (data with 5 seconds-, 10 seconds-, 20 seconds-, 30 seconds, 40 seconds- , 50 seconds- , 1 minute- , 1 minute 30 seconds, 2 minutes- intervals). We find that as we move from high to lower frequency data, the gap between upper and lower bounds of IS becomes very large. We also report the IS-mean and find that for all the time intervals, IS-mean identifies both stock exchanges contributing almost equally to the price discovery. In contrast, according to PDS, NASDAQ is the price "leader" in every sample with different time intervals. This result proves that our measure PDS, unlike IS, is robust to the choice of time intervals and more importantly, it always gives a consistent and conclusive results when IS and IS-mean fail to do so.

6 Conclusion

This paper proposes a solution to the arbitrariness and lack of identification that come with Hasbrouck's (1995) IS. We propose a new method in quantifying the contribution to price discovery. We have shown analytically that our measure, PDS is identical to IS when there is zero contemporaneous correlation among the innovations to different market prices. When the correlation is non-zero, IS reports a range that can be excessively wide with strong correlation. On the other hand, PDS reports a unique measure under similar conditions. Our method also demonstrates a systematic way of distributing covariance contribution of each asset price to the permanent shock variance evenly across prices. We also demonstrate that PDS performs better compared to IS in identifying the price discovery contribution using simulated market data with known structural price discovery contribution.

Our empirical application investigates the new financial market phenomenon of "duplication of ETFs". Our paper contributes both into the literature of price discovery and the ETFs by showing that the presence of multiple near-identical ETFs do not result in creating equally important sources of information. We have analyzed the price discovery across SPY and IVV, two popular and highly traded S&P 500 ETFs in two different trading conditions-normal and extremely volatile. SPY is found to be information leader in both cases. This indicates that despite the existence of multiple S&P 500 ETFs, a stock market trader treats the quotes of SPY as the dominant source of information. This dominance is found to be absolutely leaning toward SPY during extreme volatile trading condition a.k.a. Flash Crash, 2010 and the stock market fall of August, 2011.

Our application also provides a useful information for ETF investors. As there are numerous ETFs that track the S&P 500 index, investors often face dilemma in choosing an ETF for their investment portfolio. Low tracking error, low expense ratio and low tax burdens are the common determinants behind this decision. ETF's price discovery contribution can also become an important consideration for choosing an ETF for an investor's portfolio. In addition, if a clear pattern of price leadership is identified between two competing ETFs in a particular stock exchange, it may be possible for arbitragers to make profits by adopting pair-trading. In both cases, our new measure, PDS can provide the investors with useful information regarding ETFs.

In our second application we provide comparative assessment between IS and PDS. We show that even using one second interval quotes, IS can report uninformative results. By contrast, PDS always gives clean decomposition of price discovery contribution and identifies the information leader correctly. In a separate application, we also demonstrate PDS's robustness to the use of quotes with high time intervals.

Our expectation is that, our new order invariant measure of price discovery, PDS will be adopted widely in the future discourse on price discovery. We also hope that our study on the S&P 500 ETFs will attract more attention in future to this class of assets and their price dynamics.

References

[1] 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] Barndorff-Nielsen, O. E., Hansen, P. R., Lunde, A., Shephard, N., 2008, *Realised Kernels in Practice: Trades and Quotes*, Econometrics Journal, Vol 4, 1-33

[3] 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

[4] Boehmer, B., Boehmer, E.,2003, *Trading your neighbors ETFs: Competition or fragmentation?*,Journal of Banking & Finance 27, 1667-1703

[5] Borkovec, M, Domowitz, I., Serbin, V., Yegerman, H., 2010, *Liquidity and Price Discovery in Exchange-Traded Funds*, Investment Technology Group (ITG), Inc.

[6] Boudt, K., Cornelissen, J., Payseur, S., *Highfrequency: Toolkit for the analysis of highfrequency financial data in R*

[7] Bruder, B., Roncalli, T., 2012, *Managing Risk Exposure Using the Risk Budgeting Approach*, Working Paper, Lyxor Asset Management

[8] Caporale, G. M., Girardi, A., 2013, *Price discovery and trade fragmentation in a multimarket environment: Evidence from the MTS system*, Journal of Banking & Finance 37, 227-240

[9] Chakravarty, S., Gulen, H., Mayhew, S., 2004, *Informed trading in stock and option markets*, Journal of Finance, Volume 59, No. 3, 1235-1258

[10] Fang, Y., Sanger, G. C., 2012, *Index Price Discovery in the Cash Market*, Working paper,Midwest Finance Association 2012 Annual Meetings Paper

[11] Gonzalo, J., Granger, C. W. J., 1995, *Estimation of common long-memory components in cointegrated systems*, Journal of Business and Economic Statistics 13, 27-35

[12] 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.

[13] Harris, F. H. deB., McInish, T. H., Wood, R. A., 2002, *Security price adjustment across exchanges: an investigation of common factor components in Dow Stocks*, Journal of Financial Markets 5, 277-308

[14] Hasbrouck, J., 1995, *One Security, Many Markets: Determining the Contribution to Price Discovery*, Journal of Finance, Vol- 50, No. 4, 1175-1199

[15] Hasbrouck, J., 2002, *Stalking the efficient price in market microstructure specifications: an overview*, Journal of Financial Markets 5, 329-339

[16] Hasbrouck, J., 2003, *Intraday Price Formation in U.S. Equity Index Markets*, Journal of Finance, Vol LVIII, No. 6, 2375-2399

[17] Hedge, Shantaram P., McDermott, John B., 2004, *The market Liquidity of DIAMONDS, Qs and their underlying stocks*, Journal of Banking & Finance 28, 1043-1067

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

[19] Lien, D., Shrestha, K., 2009, *A New Information Share Measure, The Journal of Futures Markets*, Vol 29, No. 4, 377-395

[20] Marshall, B., R., Nguyen, N. H., Visaltanachoti, N., 2013, *ETF Arbitrage: Intraday evidence*, Journal of Banking & Finance 37, 3486-3498

[21] Mizarch, B., Neely C. J., 2008, *Information Shares in the U.S. Treasury Market*, Journal of Banking & Finance 32, 1221-1233

[22] Putnins, T. J., 2013, *What do price discovery metrics really measure?* , Journal of Empirical Finance 23, 68-83

[23] Tse, Y., Martinez, V., 2007, *Price discovery and informational efficiency of international iShares funds*, Global Finance Journal 18, 1-15

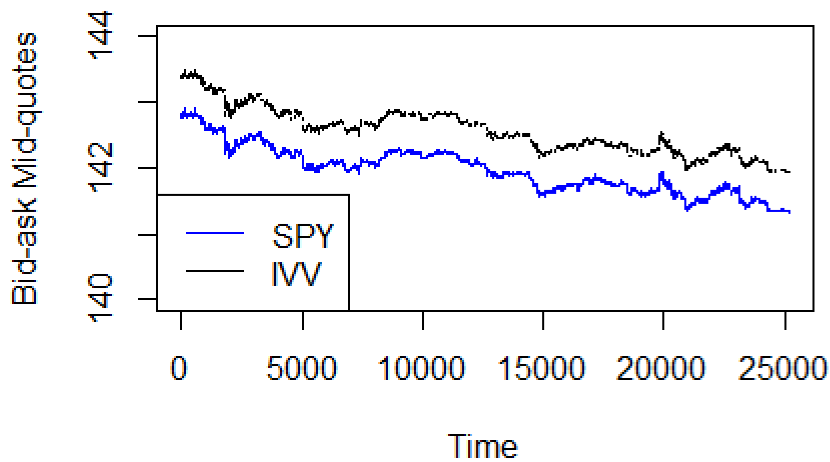
[24] Winne Rudy D., Gresse, C., Platten, I., 2014, *Liquidity and Risk Sharing Benefits from Opening an ETF market with Liquidity Providers: Evidence from CAC 40 Index*, International Review of Financial Analysis, Forthcoming

[25] Yan, B., Zivot, E., 2007, *The Dynamics of Price Discovery*, University of Washington Mimeo,

[26] Yan, B., Zivot, E., 2010, *A structural analysis of price discovery measure*, Journal of Financial Markets 13, 1-19

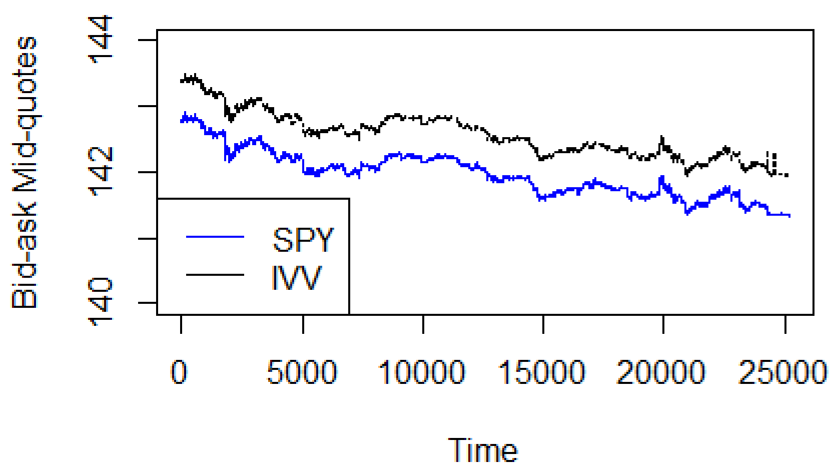
[27] Yu, L., *Basket securities, price formation, and informational efficiency*, 2005, Working paper, University of Notre Dame

Figure 1: Quotes of SPY and IVV at NASDAQ on Dec 3,2012 from 9:30 am- 16:30 pm



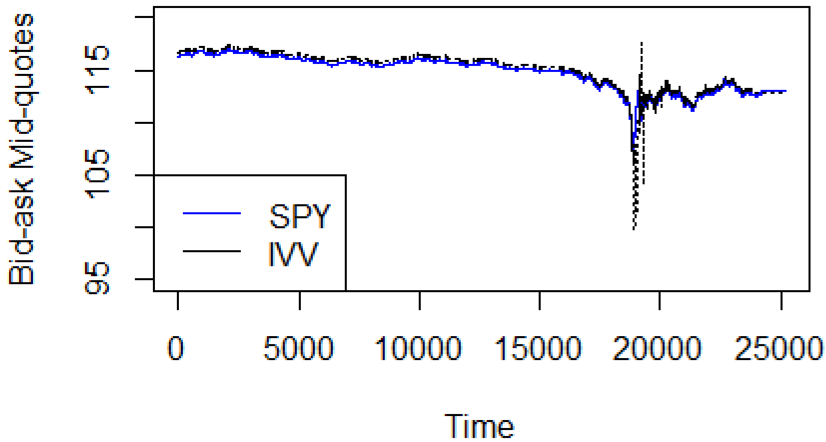
Sources: NYSE TAQ Database

Figure 2: Quotes of SPY and IVV at BATS on Dec 3,2012 from 9:30 am- 16:30 pm



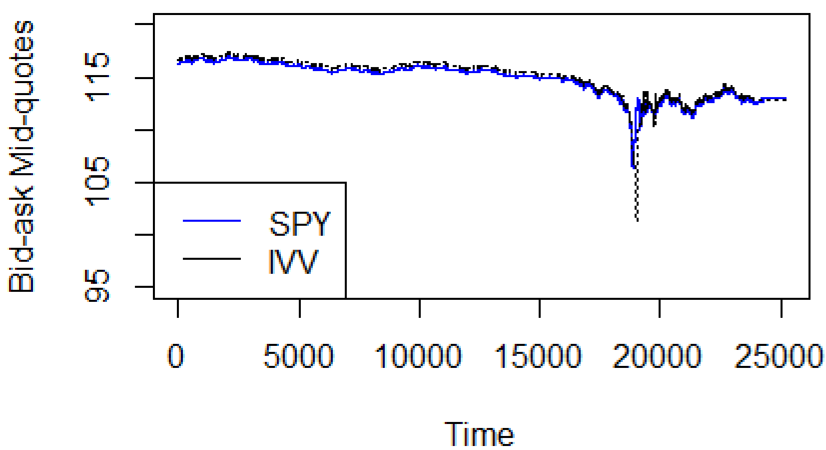
Sources: NYSE TAQ Database

Figure 3: Quotes of SPY and IVV at NASDAQ on May 6,2010 from 9:30 am- 16:30 pm



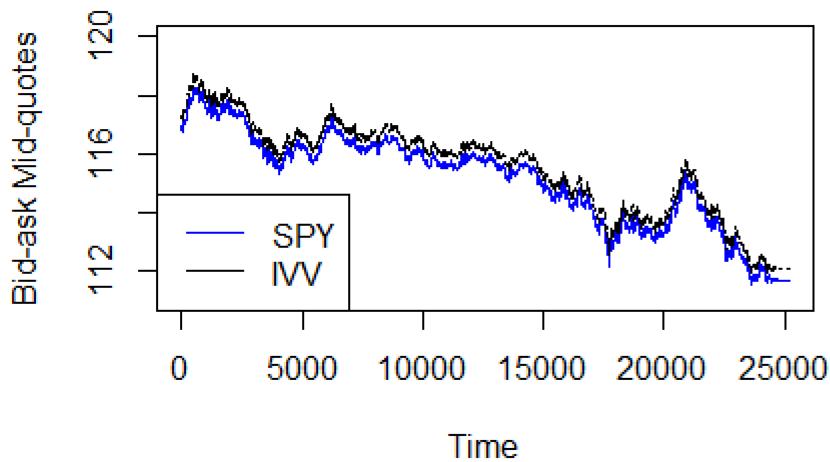
Sources: NYSE TAQ Database

Figure 4: Quotes of SPY and IVV at BATS on May 6,2010 from 9:30 am- 16:30 pm



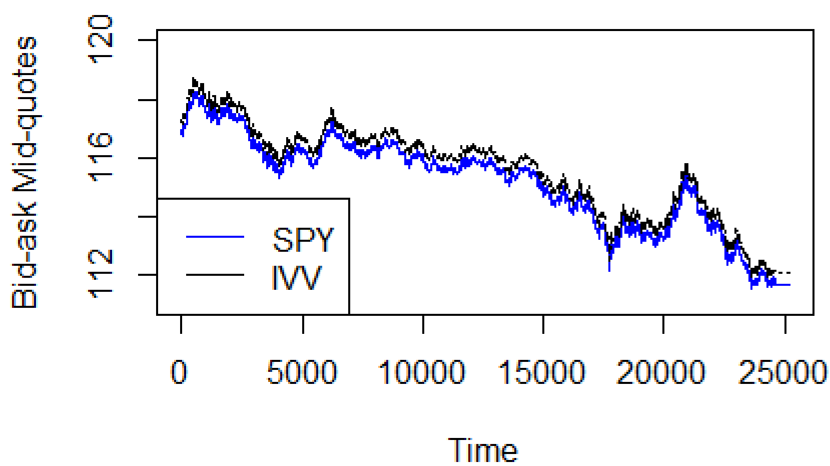
Sources: NYSE TAQ Database

Figure 5: Quotes of SPY and IVV at NASDAQ on Aug 8,2011 from 9:30 am- 16:30 pm



Sources: NYSE TAQ Database

Figure 6: Quotes of SPY and IVV at BATS on Aug 8,2011 from 9:30 am- 16:30 pm



Sources: NYSE TAQ Database

Table 1: Two Markets "Roll" Model

Efficient price: $m_t = m_{t-1} + u_t, \qquad u_t \sim N(0, \sigma^2)$
Trade direction: $q_{it} = \pm 1$, each with pr. $\frac{1}{2}$ for i=1,2,
Transaction price: $p_{it} = m_t + cq_{it}$ for =1,2
where $c = 1$ and $\sigma_u = 1$ so that both markets share of price discovery is 50%.
Model 1 is simulated for 1,000 samples of 100,000 observations.
IS and PDS are computed from the estimated VECM (eq. 8) with 20 lags.

Model 1 : structural price discovery share of market 1 = 0.5	Hasbrouck (1995) model: IS for Market 1 Upper Bound	Lower Bound	IS-Mean	PDS
mean	0.78	0.21	0.500	0.5011
Standard Deviation	0.011	0.011	0.011	0.017
95% confidence interval	[0.766, 0.812]	[0.188, 0.235]	[0.478,0.522]	[0.4662, 0.5359]

Table 2: Two Markets with Private Information

Efficient price: $m_t = m_{t-1} + \lambda q_{1t}$
Trade direction: $q_{it} = \pm 1$, each with pr. $\frac{1}{2}$ for i=1,2,
Transaction price: $p_{1t} = m_t + c_1 q_{1t}$ and $p_{2t} = m_{t-1} + c_2 q_{2t}$
where $c_1 = 1, c_2 = 1, \lambda = 1$ so that market 1s share of price discovery is 100%.
Model 2 is simulated using parameter values for 1,000 samples of 100,000 observations.
IS and PDS are computed from the estimated VECM (eq. 8) with 20 lags.

Model 2 : structural price	Hasbrouck (1995) model: IS for Market 1			
discovery share of market 1 = 1.0	Upper Bound	Lower Bound	IS-mean	PDS
mean	0.999	0.999	0.999	0.999
Standard Deviation	0.0002	0.0002	0.0002	0.0002
95% confidence interval	[0.999, 1.0]	[0.99, 1.0]	[0.99, 1.0]	[0.999, 1.0]

Table 3: Two Markets with Public and Private Information

Efficient price: $m_t = m_{t-1} + \lambda q_{1t} + u_t$ where $u_t \sim N(0, \sigma_u^2)$
Trade direction: $q_{it} = \pm 1$, each with pr. $\frac{1}{2}$ for i=1,2,
Transaction price: $p_{1t} = m_t + c_1 q_{1t}$ and $p_{2t} = m_{t-1} + c_2 q_{2t}$
where $c_1 = 1, c_2 = 0, \lambda = 1, \sigma_u = 1$ so that the share of price discovery in market 1 is 100%.
Model 3 is simulated for 1000 samples of 100,000 observations.
IS and PDS are computed from the estimated VECM (eq. 8) with 20 lags.

Model 3 : structural price	Hasbrouck (1995) model: IS for Market 1			
discovery share of market 1 = 1.0	Upper Bound	Lower Bound	IS-mean	PDS
mean	0.984	0.900	0.942	0.960
Standard Deviation	0.003	0.008	0.005	0.006
95% confidence interval	[0.978, 0.990]	[0.884, 0.916]	[0.932, 0.952]	[0.948, 0.972]

Table 4: Modified two-market "Roll" Model

Efficient price: $m_t = m_{t-1} + u_t, u_t \sim N(0, \sigma^2)$ Trade direction: $q_{it} = \pm 1$, each with pr. $\frac{1}{2}$ for i=1,2, $p_{1t} = Dm_t + (1-D)m_{t-1} + cq_{1t}$ $p_{2t} = (1-D)m_t + Dm_{t-1} + cq_{2t}$ Where, $D = 1$ with probability 0.7 and $D = 0$ with probability 0.3. $c = 1$ and $\sigma_u = 1$. Market 1's share of price discovery is 70%. Model 1 is simulated for 1,000 samples of 100,000 observations. IS and PDS are computed from the estimated VECM (equation 8) with 20 lags.				
Model 4 : structural price	Hasbrouck (1995) model: IS for Market 1			
discovery share of market 1 = 0.7	Upper Bound	Lower Bound	IS-mean	PDS
mean	0.805	0.508	0.656	0.675
Standard Deviation	0.011	0.014	0.012	0.014
95% confidence interval	[0.827, 0.783]	[0.536, 0.480]	[0.632, 0.680]	[0.703, 0.647]

Table 5: Comparison among SPY, IVV and VOO (Overview and Performance)

	SPY	IVV	VOO
Overview			
Issuer	State Street SPDR	iShares	Vanguard
Inception	22, Jan-1993	15, May-2000	7, Sept-2010
Asset Under Management	$165,308.6 M	$61,743.0 M	$21,794.6 M
Shares Outstanding	868.6 M	322.3 M	124.9 M
Expense Ratio	0.09%	0.07%	0.05%
Performance Comparison			
1 Week Return	3.39%	3.36%	3.38%
4 Weeks Return	-1.99%	-2.01 %	-1.99%
26 Weeks Return	4.27%	4.32%	4.33%
1 Year Return	13.44%	13.53%	13.48%
5 Year Return	96.65%	96.05 %	n/a
Beta	1.00	0.99	0.99
P/E ratio	15.24	15.24	15.37
Annual Dividend Rate	$3.68	$3.64	$3.38
Annual Dividend Yield	1.93%	1.90%	1.94%
5 day Volatility	18.23%	18.56%	20.63%
200 day Volatility	10.93%	11.01%	11.20%

Souurce: ETF Database. Web link: http://etfdb.com/tool/etf-comparison/IVV-SPY/ .
All the results are reported on October 21st , 2014

Table 6: Comparison among SPY, IVV and VOO (Holdings and Sector Breakdown)

	SPY	IVV	VOO
Top 10 Holdings			
Apple Inc.	3.43%	3.44%	3.46%
Exxon Mobil Corporation	2.28%	2.28%	2.39%
Microsoft Corporation	2.17%	2.18%	1.91%
Johnson & Johnson	1.71%	1.71%	1.64%
General Electric Co	1.46%	1.46%	1.46%
Berkshire Hathaway Inc Class B	1.43%	1.43%	1.27%
Wells Fargo & Co	1.40%	1.40%	1.38%
Procter & Gamble Co	1.29%	1.29%	1.26%
Chevron Corp	1.29 %	1.29%	1.38%
JPMorgan Chase & Co	1.29%	1.29%	1.26%
Sector Breakdown			
Technology	17.94%	17.98%	17.64%
Financial Service	14.89%	14.92%	14.70%
Health Care	14.24%	14.28%	14.02%
Industrial	10.99%	10.99%	10.89%
Consumer Cyclical	10.17%	10.19%	10.37%
Energy	9.65%	9.67%	10.34%
Consumer Defensive	9.47%	9.49%	9.33%
Communication Services	4.01%	4.02%	3.96%
Basic Materials	3.28%	3.29%	3.35%
Utilities	2.97%	2.99%	3.00%
Real Estate	1.91%	1.92%	1.96%
Others	0.48%	0.26%	0.44%

Souurce: ETF Database. Web link: http://etfdb.com/tool/etf-comparison/IVV-SPY/ .
All the results are reported on October 21st , 2014

Table 7: S&P 500 volatility index (VIX or the fear index) in normal trading period (Dec 3rd-5th , 2012) and in abnormal (volatile) trading period (May 6th, 2010, Aug 8th, 2011)

Date	Open	High	Low	Close
Normal Trading Period				
3-Dec,12	15.81	16.69	15.76	16.64
4-Dec,12	16.66	17.37	16.38	17.12
5-Dec,12	16.95	17.53	16.27	16.46
6-Dec,12	16.59	16.85	16.31	16.58
7-Dec,12	16.12	16.65	15.73	15.87
Average	16.42	17.01	16.09	16.53
Abnormal/volatile Trading Period				
6-May, 10	25.88	40.71	24.43	32.8
8-Aug,11	36.9	48	35.29	48
Average	31.39	44.35	29.86	40.4

Source: Yahoo! Finance

Table 8: Descriptive Statistics of SPY and IVV in different Stock Exchanges on Dec 3rd - Dec 7th, 2012. The prices are the log of bid-ask mid-quotes of SPY and IVV from 9:30 am-4:30 pm each day.

Stock Exchanges	Vectors of prices	Average 1-sec return($\times 10^{-8}$)	Average 1-sec return volatility ($\times 10^{-5}$)	Intra-day average numbers of shares traded	Intra-day average bid-ask spread
NASDAQ	SPY	-7.52	0.003	23,960,312	0.01
	IVV	-7.46	0.004	1,155,608	0.02
BATS	SPY	-7.48	0.004	18,935,201	0.01
	IVV	-7.55	0.004	964,979	0.02
Arca	SPY	-7.52	0.004	26,708,530	0.01
	IVV	-7.52	0.004	744,723	0.03
CBOE	SPY	-6.92	0.006	331,915	0.06
	IVV	-9.28	0.004	2980	0.06
NSX	SPY	5.97	0.007	235,790	0.02
	IVV	-9.49	0.036	4973	0.03
Boston	SPY	-7.58	0.003	3,018,644	0.02
	IVV	-7.57	0.004	107,219	0.04
Philadelphia	SPY	-7.78	0.004	1,963,004	0.02
	IVV	-7.49	0.004	123,952	0.03
EDGE A	SPY	-7.71	0.004	2,204,523	0.02
	IVV	-7.60	0.004	63,747	0.06

Table 9: Descriptive Statistics of SPY and IVV in different Stock Exchanges on May 6th, 2010 (Flash Crash) from 9:30 am-4:30 pm.

Stock Exchange	Vector of Prices	Average 1-sec return($\times 10^{-8}$) on May 6th, 2010	Average 1-sec return volatility($\times 10^{-5}$) May 6th, 2010	Intra-day average no of shares traded, May-6th, 2010	Intra-day average bid-ask spread, May 6th, 2010
NASDAQ	SPY	-58.0	10.0	201,085,629	0.02
	IVV	-67.0	100.0	3,754,730	0.09
BATS	SPY	-59.0	20.0	99,711,462	0.02
	IVV	-67.0	50.0	3,712,119	0.07
Arca	SPY	-58.0	10.0	145,771,969	0.02
	IVV	-59.0	90.0	2,771,224	0.07

Table 10: Descriptive Statistics of SPY and IVV in different Stock Exchanges on Aug 8th, 2011 from 9:30 am-4:30 pm.

Stock Exchange	Vector of Prices	Average 1-sec return($\times 10^{-8}$) on Aug 8th, 2011	Average 1-sec return volatility($\times 10^{-5}$) Aug 8th, 2011	Intra-day average no of shares traded, Aug 8th, 2011	Intra-day average bid-ask spread, Aug 8th, 2011
NASDAQ	SPY	-180	14	151,577,895	0.01
	IVV	-181	13	4,760,661	0.04
BATS	SPY	-181	14	128,578,031	0.01
	IVV	-181	13	3,245,591	0.04
Arca	SPY	-181	14	140,039,777	0.01
	IVV	-181	13	3,660,737	0.04

Table 11: PDS between SPY and IVV in Different Stock Exchange : Normal Trading Condition vs. Extremely Volatile Trading Condition

Stock Exchange	Vectors of Prices	Daily average of PDS on Dec 3rd-7th, 2012	PDS on May 6th, 2010 (Flash Crash)	PDS on Aug 8th, 2011
NASDAQ	SPY	0.53	0.92 (0.002)	0.83 (0.009)
	IVV	0.47	0.08 (0.002)	0.17 (0.009)
BATS	SPY	0.59	0.99 (0.005)	0.62 (0.012)
	IVV	0.41	0.01 (0.005)	0.38 (0.012)
Arca	SPY	0.62	0.93 (0.005)	0.79 (0.016)
	IVV	0.38	0.07 (0.005)	0.21 (0.016)
CBOE	SPY	0.56		
	IVV	0.44		
NSX	SPY	0.69		
	IVV	0.31		
Boston	SPY	0.58		
	IVV	0.42		
Philadelphia	SPY	0.53		
	IVV	0.47		
EDGE A	SPY	0.80		
	IVV	0.20		

Table 12: Price discovery across different ETFs (log of bid-ask mid-quotes of SPY and IVV) in the same stock exchange (9:30 am-4:30 pm, Dec 3rd, 2012), bootstrap standard errors are reported in parenthesis

Vector of Prices	Stock Exchange	IS- Upper bound	IS- lower bound	Mid-point IS	PDS
NASDAQ	SPY	0.9266	0.1325	0.5295	0.05813
		(0.023)	(0.023)	(0.023)	(0.023)
	IVV	0.8675	0.0734	0.4705	0.4187
		(0.023)	(0.023)	(0.023)	(0.023)
BATS	SPY	0.9867	0.5736	0.7802	0.9093
		(0.023)	(0.023)	(0.023)	(0.023)
	IVV	0.4264	0.0133	0.2198	0.0907
		(0.023)	(0.022)	(0.022)	(0.023)
Arca	SPY	0.9236	0.1427	0.5331	0.5865
		(0.01)	(0.01)	(0.01)	(0.01)
	IVV	0.4787	0.0325	0.2556	0.1494
		(0.02)	(0.02)	(0.02)	(0.02)
Chicago Board Option Exchange (CBOE)	SPY	0.9675	0.5213	0.7444	0.8506
		(0.021)	(0.02)	(0.02)	(0.02)
	IVV	0.4787	0.0325	0.2556	0.1494
		(0.02)	(0.02)	(0.02)	(0.02)
National Stock Exchange (NSX)	SPY	0.999	0.996	0.998	0.999
		(0.001)	(0.001)	(0.001)	(0.001)
	IVV	0.004	0.001	0.002	0.001
		(0.001)	(0.001)	(0.001)	(0.001)
Boston Stock Exchange	SPY	0.7810	0.1088	0.4449	0.3975
		(0.021)	(0.021)	(0.021)	(0.021)
	IVV	0.8912	0.2190	0.5551	0.6025
		(0.021)	(0.021)	(0.021)	(0.021)
Philadelphia Stock Exchange	SPY	0.8720	0.1982	0.5351	0.5647
		(0.02)	(0.02)	(0.02)	(0.02)
	IVV	0.8018	0.1280	0.4649	0.4353
		(0.021)	(0.021)	(0.021)	(0.021)
EDGE A Stock Exchange	SPY	0.8676	0.4180	0.6428	0.6845
		(0.02)	(0.02)	(0.02)	(0.02)
	IVV	0.5820	0.1324	0.3572	0.3155
		(0.021)	(0.021)	(0.021)	(0.021)

Table 13: Price discovery in log of bid-ask mid-quotes of SPY in BATS and NASDAQ (from high frequency to low frequency data), bootstrap standard errors are reported in parenthesis.

Time-lag between each observation	SPY traded at BATS and NASDAQ	IS-upper bound	IS-lower bound	IS mid-point	PDS
1 sec interval	BATS	0.9215	0.0227	0.4721	0.3428
		(0.02)	(0.02)	(0.02)	(0.02)
	NASDAQ	0.9733	0.0785	0.5279	0.6572
		(0.02)	(0.02)	(0.02)	(0.02)
5 sec interval	BATS	0.9656	0.01	0.4828	0.0162
		(0.05)	(0.04)	(0.05)	(0.05)
	NASDAQ	0.99	0.0344	0.5172	0.9838
		(0.05)	(0.05)	(0.05)	(0.05)
10 sec interval	BATS	0.98	0.01	0.4922	0.1028
		(0.05)	(0.05)	(0.05)	(0.05)
	NASDAQ	0.99	0.02	0.5078	0.8972
		(0.05)	(0.05)	(0.05)	(0.05)
20 sec interval	BATS	0.99	0.001	0.4963	0.1821
		(0.09)	(0.09)	(0.09)	(0.09)
	NASDAQ	0.999	0.01	0.5037	0.8179
		(0.09)	(0.09)	(0.09)	(0.09)
30 Sec interval	BATS	0.995	0.001	0.4978	0.2019
		(0.10)	(0.10)	(0.10)	(0.10)
	NASDAQ	0.999	0.005	0.5022	0.7981
		(0.10)	(0.10)	(0.10)	(0.10)
40 sec interval	BATS	0.995	0.001	0.4978	0.1353
		(0.11)	(0.11)	(0.11)	(0.11)
	NASDAQ	0.999	0.005	0.5022	0.8647
		(0.11)	(0.11)	(0.11)	(0.11)
50 sec interval	BATS	0.998	0.002	0.4999	0.4726
		(0.11)	(0.11)	(0.11)	(0.11)
	NASDAQ	0.998	0.002	0.5001	0.5274
		(0.11)	(0.11)	(0.11)	(0.11)
1 min interval	BATS	0.998	0.001	0.4993	0.3726
		(0.12)	(0.12)	(0.12)	(0.12)
	NASDAQ	0.999	0.002	0.5007	0.6724
		(0.12)	(0.12)	(0.12)	(0.12)
1.5 min interval	BATS	0.996	0.001	0.4983	0.0283
		(0.12)	(0.12)	(0.12)	(0.12)
	NASDAQ	0.999	0.004	0.5017	0.9717
		(0.12)	(0.12)	(0.12)	(0.12)
2 min interval	BATS	0.999	0.001	0.50	0.1168
		(0.15)	(0.15)	(0.15)	(0.15)
	NASDAQ	0.999	0.001	0.50	0.8823
		(0.15)	(0.15)	(0.15)	(0.15)