How Well Does the Weighted Price Contribution Measure Price Discovery?

Jianxin Wang and Minxian Yang
January 2010

Abstract

Using the information share (IS) measure as a benchmark for price discovery, this paper examines the theoretical properties and empirical performance of the weighted price contribution (WPC). We show that the WPC does not generally measure the proportions of mean returns as it appears. When returns are normally distributed, the large-sample WPC converges to a function of the mean, variance, and correlation of the returns. It converges to the IS when the returns follow independent normal distributions with zero mean. The empirical comparison is based on estimating the overnight price discovery for the S&P 500 index. Confirming the theoretical analysis, the performance of the WPC is sensitive to return serial correlation: as the overnight and daytime return correlation becomes higher in recent years, the deviations between the IS and the WPC becomes larger. Since the IS can be estimated by a simple procedure with the same data requirement as the WPC, we recommend the use of the IS as a price discovery measure.

Keywords: price discovery, the weighted price contribution, the information share, information flow, the efficient price, overnight return, daytime return, the S&P 500 index.

Jianxin Wang (jx.wang@unsw.edu.au) and Minxian Yang (m.yang@unsw.edu.au) are both from the Australian School of Business, University of New South Wales, Sydney, Australia. Jianxin Wang is the corresponding author.
I. Introduction

A core function of financial markets is price discovery, the process that incorporates economic information into asset prices. Price discovery is by far the most important theme in the market microstructure literature. We have achieved considerable understanding of the impact on the price discovery process from different trading mechanisms, market participants, and changes in regulations. However, our ability to empirically measure and compare price discovery across markets, trading venues, and trading periods, remains limited. Hasbrouck (1995) and Harris, et al. (2002) are the two dominant approaches for comparing price discovery across different markets and trading venues. Adopting these two approaches to various applications has become “a mid-sized cottage industry” (Lehmann, 2002). A special issue of the Journal of Financial Markets in 2002 was devoted to the comparison of the two models. Recently Yan and Zivot (2010) use a structural cointegration model to bring new insights to the comparison. De Jong and Schotman (2010) propose a new measure based on market-specific structural innovations.

One popular non-parametric method for measuring price discovery is the weighted price contribution (WPC) proposed by Barclay and Warner (1993). Originally it was used to measure price movements associated with different transaction sizes. Cao, et al, (2000) is the first to use it as a price discovery measure and term it the “weighted price contribution”. It has been used to measure price discovery during the pre-opening period (Cao, et al. 2000), across trading venues (Huang, 2002), during overnight trading (Barclay and Hendershott, 2003, 2008), and during opening and closing call auctions (Ellul, et al. 2005). Note that the models of Hasbrouck (1995) and Harris, et al. (2002) are designed to compare price discovery across parallel markets\(^1\) where

---

trading takes place simultaneously. The WPC, however, can be used to compare price discovery across parallel markets as well as non-overlapping trading periods. Together with its simplicity, this flexibility of the WPC greatly enhances its popularity, particularly for supplementing and supporting the core methodology and findings, e.g. Owens and Steigerwald (2005) and Agarwal, et al. (2007).

Although the WPC is widely used, little has been said about what it exactly measures, its statistical properties and empirical performance. The validity of the WPC as a price discovery measure seems to come from its definition: the weighted average return ratio attributed to a market or trading period. This paper goes beyond the definition and explores the theoretical relationship between the WPC and the characteristics of the return series: its mean, variance, and serial correlation. We compare the WPC and the information share (IS), which is the variance ratio of efficient price changes, see Hasbrouck (1995). We show that the WPC does not generally measure the proportional mean return as it appears. Under normality assumption, the WPC converges to a function of the means, variances, and correlations of returns. It becomes a consistent estimator of the IS only when returns are uncorrelated and have zero means. The difference between the IS and the WPC crucially depends on return serial correlations.

We demonstrate the empirical difference between the IS and the WPC by estimating the overnight and daytime price discovery for the S&P 500 index. The overnight price discovery is reflected in the price change between today’s market close and next day’s market open. Wang and Yang (2009) provide a consistent estimator of the IS using a structural vector autoregression (VAR). Several studies have documented significant overnight or pre-opening price discovery

---

2 To our best knowledge, van Bommel (2009) is the first to systematically examine conditions under which the WPC is an unbiased and/or a consistent estimator of a benchmark price discovery measure. However, we argue in section II that the benchmark measure he uses is not based on changes of the efficient price therefore does not reflect the underlying information flow.
when the organized exchanges are closed, e.g. Cao, et al. (2000), Barclay and Hendershott (2003, 2004, 2008), and Moulton and Wei (2005). Tompkins and Wiener (2008) and Cliff, et al. (2008) document positive overnight returns and negative daytime returns across major international markets. We show that the IS estimates for overnight price discovery increased significantly in the past decade and have twice reached over 30% of the daily price discovery. The differences between the estimated values of the IS and the WPC are very sensitive to the correlation between overnight and daytime returns. While the annual estimates of the IS and the WPC are broadly similar from 1993 to 2006, the correlation between overnight and daytime returns increased significantly from 2007 to 2009, resulting large differences between the IS and the WPC.

Given that the WPC does not generally measure the characteristics of efficient price changes, we recommend the consistent IS estimator based on the structural VAR model. The IS measure is on a firm economics ground: the variance of the efficient price is a natural measure of the information flow. The structural VAR model in this context is very easy to implement: an OLS regression with the Cholesky decomposition. It has the same data requirement as the WPC. As discussed in Wang and Yang (2009), when intraday prices are available, one should use the estimator based on realized variance after filtering out the impact of the noise term.

This paper is organized as the following: section II defines and motivates the IS as a benchmark measure for price discovery. Section III explores the relationship between the WPC and return characteristics, and draws theoretical comparison between the IS and the WPC. Section IV presents the structural VAR estimation of the IS and the empirical comparisons based on the overnight and daytime returns of the S&P 500 index. Section V concludes.
II. Information Flow and Price Discovery

Price discovery is commonly defined as the incorporation of economic information into asset prices. Economic information includes anything that affects the fundamental value of the asset, also termed the efficient price. While the variation of the observed price can be driven by liquidity trading, the variation in the efficient price in a market over a given period reflects the information flow in that market during that period. On the other hand, if the relevant information, public or private, is not fully incorporated into the asset price, it indicates a failure of price discovery. Therefore a natural measure for information flow or price discovery is the variance of the efficient price change. The efficient price, which is a random walk process, is unobservable but the variance of its change can be readily estimated.

In this section, we extend the information share measure of Hasbrouck (1995, 2002) to the case of trading in non-overlapping markets or periods. This also provides the setting for analysing the WPC. Consider a stock traded on an organized exchange. A trading day is divided into $n$ consecutive trading periods. Let $p_{i,t}$ and $r_{i,t} = p_{i,t} - p_{i-1,t}$ be the log price and the return for the $i^{th}$ period in day $t$. Note that $p_{0,t} = p_{n,t-1}$ and the daily return is $r_t = \sum_{i=1}^{n} r_{i,t} = p_{n,t} - p_{n,t-1}$. Our aim is to measure price discovery during the $i^{th}$ trading period relative to the rest of the trading day. The returns of the periods are subject to period-specific price shocks $\eta_{i,t}$. The shocks are serially uncorrelated and interpreted as period-specific news. They can be unexpected changes in economic fundamentals, short-term mispricing, or changes in liquidity and microstructure factors (e.g. bid-ask bounce or inventory control). In general, only part of a shock, the permanent component, enters the efficient price.

The end-of-period price $p_{i,t}$ can be written as $p_{i,t} = m_{i,t} + u_{i,t}$, where $m_{i,t}$ is the efficient price reflecting new information on economic fundamentals, and $u_{i,t}$ is a noise term resulting
from transitory factors. The changes in the efficient price are $\Delta m_{i,t} = m_{i,t} - m_{i-1,t}$, $i=1,\ldots,n$. They are serially uncorrelated and capture the permanent or information components in price innovations $\eta_{i,t}$. The information flow in the $i$-th period is measured by $\text{var}(\Delta m_{i,t})^4$. The change of the efficient price over day $t$ is $\Delta m_t = \sum_{i=1}^{n} \Delta m_{i,t}$. The information share of period $i$ on day $t$ is defined as

$$\text{IS}_i = \frac{\text{var}(\Delta m_{i,t})}{\sum_{i=1}^{n} \text{var}(\Delta m_{i,t})}, \quad i=1,\ldots,n.$$  

The above measure is in the same spirit of Hasbrouck (1995). He measures price discovery across parallel markets where trading takes place simultaneously. Section IV discusses the estimation of $\text{var}(\Delta m_{i,t})$ and $\text{var}(\Delta m_t)$ from the observed price changes over period $i$ and the entire trading day $t$.

The IS defined above is generally different from the price discovery measure of Bommel (2009). He measures price discovery in period $i$ ($i=1,\ldots,n$) as

$$\theta_i = 1 - \frac{\text{var}(r_{i,t}|r_{-i,t})}{\text{var}(r_t)},$$  

which is the population $R^2$ for the regression

$$r_t = \alpha + \beta r_{i,t} + \epsilon_t.$$  

He considers the conditions under which the WPC is an unbiased and/or a consistent estimator of $\theta_i$. We argue that $\theta_i$ is not a desirable measure for price discovery. Let $r_i = r_{i,t} + r_{-i,t}$, where $r_{-i,t}$ is the sum of the returns other than the $i$th period. Define $\sigma^2 = \text{var}(r_t)$, $\sigma_i^2 = \text{var}(r_{i,t})$, $\sigma_{-i}^2 = \text{var}(r_{-i,t})$, $\sigma^2 = \text{var}(r_{i,t}|r_{i,t})$, and $\rho = \text{cor}(r_{i,t}, r_{-i,t})$. Since $r_i = r_{i,t} + r_{-i,t}$, equation (3) becomes

$$r_{-i,t} = \alpha + (\beta-1)r_{i,t} + \epsilon_t,$$

1 Equation (11) below and Wang and Yang (2009) illustrate the relationship between $\Delta m_{i,t}$ and $\eta_{i,t}$.
2 If the efficient price $m_t$ follows a standard continuous diffusion process, the variance of $\Delta m_t$ over a trading day is the instantaneous variance integrated over the trading day. See Andersen and Benzoni (2008).
which leads to $\hat{\beta} = 1 + \rho(\sigma_i/\sigma_t)$ and $\sigma^2_i = (1-\rho^2)\sigma^2_{-1}$. Since $\sigma^2 = \sigma^2_t + \sigma^2_{-1} + 2\rho\sigma_t\sigma_{-1}$, we find

$$\theta_i = \frac{\sigma^2 - \sigma^2_i}{\sigma^2} = \frac{\sigma^2_t + \rho^2\sigma^2_{-1} + 2\rho\sigma_t\sigma_{-1}}{\sigma^2}.$$

It is the same as the IS only when the returns are (serially) uncorrelated. When $\rho \neq 0$, $\theta_i$ depends on price movements in other periods $\sigma^2_{-i}$. Clearly this contradicts the definition of price discovery as the process of incorporating new information into prices. The fundamental difference between IS$_i$ and $\theta_i$ is that the former measures the variation in the efficient price but the latter measures the variation in the observed price. When $\rho = 0$, the price follows a martingale: $r_{i,t} = \Delta m_{i,t}$ and $\theta_i = IS_i$.

### III. Understanding the WPC

In this section, we explore what the WPC actually measures under different conditions, and compare it with the information-based price discovery measure discussed in section II. The daily return is $r_t = \sum_{i=1}^n r_{i,t}$. Following Barclay and Warner (1993) and Cao, et al. (2000), the WPC of the $i^{th}$ trading period is defined as

$$WPC_i = \sum_{t=1}^T \frac{r_{i,t}}{r_t} \left( \frac{|r_t|}{\sum_{s=1}^T |r_s|} \right), \; i=1,\ldots,n.$$

The ratio $r_{i,t}/r_t$ measures the proportion of return on day $t$ attributed to period $i$. As discussed by Barclay and Warner (1993) and Cao, et al. (2000), the term in the bracket is the weight that removes the impact of small $|r_t|$. We rewrite the WPC in (6) as

$$WPC_i = \frac{\sum_{t=1}^T \text{sign}(r_t) r_{i,t}}{\sum_{t=1}^T \text{sign}(r_t) r_t} = \frac{\sum_{t=1}^T \text{sign}(r_t) r_{i,t}}{\sum_{t=1}^T \text{sign}(r_t) r_t}, \; i=1,\ldots,n.$$
where \( \text{sign}(x) \) is the sign of \( x \), being 1 for positive \( x \) and -1 for non-positive \( x \). The WPC then can be interpreted as the ratio of the weighted average returns, where the weight is \( \text{sign}(r_t) \). However the WPC does not generally converges to either \( \frac{E(r_{1i})}{E(r_t)} \) or \( \frac{E(r_{ii})}{E(r_i)} \).

By the law of large numbers,

\[
WPC_i \rightarrow \frac{E[\text{sign}(r_t)r_{1i}]}{E(|r_t|)} = \frac{E[\text{sign}(r_t)r_{1i}]}{E[\text{sign}(r_t)r_t]}, \quad i = 1,\ldots,n.
\]

in probability as \( T \rightarrow \infty \). To further analyse the express in (7), we assume that returns are normally distributed. Following the notations in section II, the returns \( (r_{1t}, r_{-i,t}) \) are jointly normally distributed with means \( (\mu_i, \mu_{-i}) \), variances \( (\sigma_i^2, \sigma_{-i}^2) \) respectively and correlation \( \rho \). Define \( \mu = E(r_t) = \mu_i + \mu_{-i} \). By the Appendix, we find

\[
\frac{E[\text{Sign}(r_t)r_{1i}]}{E(|r_t|)} = \frac{2\mu \left[ 0.5 - \Phi \left( -\frac{\mu}{\sigma} \right) \right] + \sqrt{\frac{2}{\pi}} \exp \left( -\frac{\mu^2}{2\sigma^2} \right) \left( \sigma_i^2 + \rho \sigma_i \sigma_{-i} \right) / \sigma}{2\mu \left[ 0.5 - \Phi \left( -\frac{\mu}{\sigma} \right) \right] + \sqrt{\frac{2}{\pi}} \exp \left( -\frac{\mu^2}{2\sigma^2} \right) \sigma}, \quad i = 1,\ldots,n.
\]

where \( \Phi \) is the standard normal cumulative distribution function. Equation (8) reveals some theoretical properties of the large-sample WPC under the assumption of normality. First, Figure 1 depicts the surface of (8) as functions of \( \mu \) and \( \rho \). Figures 1a and 1b assume that \( \mu_i = -0.2\mu \) and \( 0.2\mu \) respectively, with \( \sigma_i = 1 \) and \( \sigma_{-i} = 2.5 \). The figures indicate that the value of (8) is not very sensitive to changes in \( \mu \), but declines quickly as \( \rho \) moves from 0.5 to -0.5. Therefore the WPC is very sensitive to return serial correlation \( \rho \). Second, equation (8) indicates that \( WPC_i \rightarrow \mu_i / \mu \) when \( \frac{\mu^2}{\sigma^2} \rightarrow \infty \), i.e. when return variance is very small relative to its mean. This is never the case in real financial data. Therefore the WPC is not the proportional mean return as it appears. In large samples, it is a complex function of return characteristics, i.e. \( \mu_i, \mu, \sigma_i, \sigma, \) and \( \rho \). Third,

\footnote{Note that \( \mu_i = -0.2\mu \) is for a specific period, not all \( i = 1,\ldots,n \). In this case, \( \mu_{-i} = 1.2\mu \) so that \( \mu_i + \mu_{-i} = \mu \).}
when $\mu \to 0$, $WPC_i \to \frac{\sigma_i^2 + \rho \sigma_i \sigma_{i-1}}{\sigma^2}$.  If $\rho = 0$, $p_{i,t}$ follows a random walk and is the efficient price. In this case, $r_{i,t} = \Delta m_{i,t}$ and $\sigma_i^2 = \text{var}(\Delta m_{i,t})$ in equation (1). Therefore equation (8) indicates that

\[
WPC_i \to \frac{\sigma_i^2}{\sigma^2} = \frac{\sigma_i^2}{\sum_{i=1}^{n} \sigma_i^2} = IS_i \quad \text{as} \quad \mu \to 0 \quad \text{and} \quad \rho \to 0.
\]

In summary, we show that the WPC is generally not a measure for the weighted proportional return in a period. We make a theoretical connection between the WPC and the benchmark information share measure defined in (1). The WPC, and it’s difference with the IS, are very sensitive to the return serial correlation. These predictions are confirmed in the empirical analysis below.

**IV. Empirical Comparison between the IS and the WPC**

In this section, we explore the empirical difference between the IS and the WPC. Using daily opening and closing values of the S&P 500 index, we estimate the IS and the WPC during daytime trading versus the overnight period. The results confirm our discussion in section III: the IS and the WPC have large differences when the correlation between the daytime return and the overnight return is high.

**The Structural VAR Estimation of the IS**

Depending on data availability, Wang and Yang (2009) provide two ways to estimate $\text{var}(\Delta m_{i,t})$ and $IS_i$ in equation (1). If transaction prices are available within each period, $\text{var}(\Delta m_{i,t})$ can be estimated using the realized variance after filtering out the impact of the noise term. If only the end-of-period prices are available, a structural vector autoregression (VAR) can be used

\(^6\) Note the similarity between this expression and $\theta_i$ in equation (5). Compared to Bommel (2009), equation (8) provides a unified approach to examine the relationship between WPC and $\theta_i$. Clearly WPC is not a consistent estimator of $\theta_i$ when $\rho \neq 0$ ($\beta \neq 1$ in equations (3) and (4)) or $\mu$ is non-zero. This corresponds to Bommel’s proposition 5(i) and 5(ii). When $\mu=\rho=0$, WPC converges to $\theta_i$ (Bommel’s proposition 3(i)). As discussed in section II, $\theta_i$ is the same as $IS_i$ when $\rho=0$.  

8
to estimate \( \text{var}(\Delta m_{i,t}) \) and IS\(_i\). Since the index values are not available overnight, the overnight realized variance cannot be estimated. We use the structural VAR to estimate the IS. It has the same data requirement as the WPC.

For the empirical analysis, a trading day \( t \) is defined from the market close on day \( t-1 \) to the market close on day \( t \). It is divided into overnight and daytime periods: \( n=2 \). Let \( p_{o,t} \) and \( p_{c,t} \) be the log opening and closing values of the S&P 500 index respectively. The overnight return is \( r_{N,t} = p_{o,t} - p_{c,t-1} \) and the daytime return is \( r_{D,t} = p_{c,t} - p_{o,t} \). Wang and Yang (2009) model the return vector, \( R_t = [r_{N,t}, r_{D,t}]' \), as a structural VAR process:

\[
B_0 R_t = a + \sum_{k=1}^{K} B_k R_{t-k} + \eta_t,
\]

where \( \eta_t = [\eta_{N,t}, \eta_{D,t}]' \) is the vector of structural shocks and \( a=[a_N, a_D]' \) is the vector of intercepts.

As discussed in section II, \( \eta_{N,t} \) and \( \eta_{D,t} \) are serially uncorrelated and reflect respectively the night-specific and day-specific changes in economic fundamentals, short-term mispricing, or microstructure factors. Their variances are normalized to one. Therefore \( E(\eta_t) = 0; E(\eta_t \eta_{t-k}') = 0 \) for \( k \neq 0; E(\eta_t \eta_{t}') = I, a 2 \times 2 \) identity matrix. \( B_0 \) is a lower triangular matrix because the periods are sequential: within the same trading day \( t \), \( r_{N,t} \) affects \( r_{D,t} \) but not vice versa. The impact of daytime trading on overnight returns is captured by the lagged returns on the right hand of (10).

The corresponding reduced form of the structural VAR is given by \( A(L)R_t = \alpha + \varepsilon_t \), where \( A(L) = I - A_1 L - ... - A_K L^K \), \( A_k = B_0^{-1} B_k \), and \( \alpha = B_0^{-1} a \). The vector of reduced-form shocks is given by \( \varepsilon_t = B_0^{-1} \eta_t \). As discussed in section II, the daily closing price \( p_{c,t} \) can be viewed as a combination between an efficient price \( m_t \) that follows a random walk, and a serially-correlated

---

7 Note that \( R_t \) differs from \( \rho_t \), which is the daily return \( p_{c,t} - p_{c,t-1} \) defined in section II.

8 An alternative and equivalent parameterization is to normalize the diagonal elements of \( B_0 \) as unity and leave the variance of \( \eta_t \) as a positive diagonal matrix.

9 Our definition of a trading day implies that the overnight period precedes the daytime trading period. As shown by Wang and Yang (2009), rotating the periods does not affect the structural VAR estimation.
noise component. Although the efficient price is not observable, Wang and Yang (2009) show that the daily change of the efficient price $m_t$ is given by

$$\Delta m_t = \mu + \tau A(1)^{-1} B^{-1}_0 \eta_t = \mu + h' \eta_t = \mu + h_N \eta_{N,t} + h_D \eta_{D,t},$$

where $\mu = A(1)^{-1} B^{-1}_0 a$, $\tau = [1,1]'$, and $h' = [h_N, h_D] = \tau A(1)^{-1} B^{-1}_0$. Since $\text{var}(\Delta m_t) = h_N^2 + h_D^2$, the IS defined in (1) becomes

$$\text{IS}_i = \frac{h_i^2}{h_N^2 + h_D^2}, \ i = N \text{ or } D.$$

Note that $A(1)$ in the reduced-form VAR is easily estimated by OLS and the $B^{-1}_0$ matrix is the lower triangle Cholesky factor of the estimated variance matrix of $\epsilon_t$. Hence the IS is almost as easy to compute as the WPC. Conceptually, the structural VAR provides a clean measure of the variance of the efficient price, whereas it is difficult to give an economic interpretation of the WPC in equations (6) and (8).

**Overnight versus Daytime Price Discovery for the S&P 500 Index**

We draw empirical comparisons between the IS and the WPC by estimating the overnight and daytime price discovery for the S&P 500 index. Our data set is the daily open and closing values of the S&P 500 index from January 1, 1993 to December 31, 2009 from DataStream. Table 1 reports the overnight and daytime returns and volatility over the 17-year sample period. It shows that overnight return and volatility were very low before 1999, and became larger since 2000. The “bad-day and good-night” return pattern generally does not hold for the S&P 500 index. Only five of the seventeen years have the average overnight return higher than the average daytime return. The result is consistent with Tompkins and Wiener (2008) but in contrast with Cliff, et al. (2008). The magnitudes of the daytime return and volatility are much larger than overnight return and volatility. Again the differences are smaller after 2000.
The annual estimates of the overnight IS and the overnight WPC are reported in Table 2, together with the correlation between overnight and daytime returns. The IS is estimated from the structural VAR in (10) and the WPC is based on equation (6). The lag length of the structural VAR is based on the Schwarz criterion. There are several features in Table 2. First, when the correlation between night return and day return is small, the overall similarity between the IS and the WPC is evident. Both were relatively low before 1999, rose up from 2001 to 2005, and dropped sharply in 2006. Second, confirming the discussion in section III, the difference between the IS and the WPC becomes large when the overnight and daytime correlation is large. The bold numbers in Table 2 indicate that high values of Cor(r_N,r_D) are always associated with large differences between the IS and the WPC; but the large differences in 2002 and 2004 are not associated with high Cor(r_N,r_D). Clearly high Cor(r_N,r_D) is not the only condition that leads to large deviations between the IS and the WPC. Other parameters (e.g. μ/σ) and the normality assumption may also play a role. The correlation between IS_N-WPC_N and Cor(r_N,r_D) is 0.66. Third, despite the overall similarity, the numerical difference between the IS and the WPC can be quite large, particularly relative to the estimated values of the IS and the WPC. Therefore using the IS or the WPC may lead to different empirical conclusions.

The above empirical analysis confirms the discussion in section III based on equations (8) and (9). Table 1 shows that the average overnight returns are mostly within the range of (-3, +3) basis points and the average daytime returns are mostly within (-10, +10) basis points. So the average daily return is approximately zero. The performance of the WPC, measured by its closeness to the IS, is very sensitive to the return correlation across intraday periods. While equation (8) is based on the strong assumption of normality, it works reasonably well with
returns of large indices and large portfolios. Most financial returns are not normally distributed, making it difficult to assess the empirical performance of the WPC.

V. Conclusion

Price discovery is a central function of financial markets and a central theme in market microstructure literature. One popular measure for price discovery is the WPC originally proposed by Barclay and Warner (1993). Bommel (2009) provides the only systematic examination of the statistical properties of the WPC. We argue that the benchmark used by Bommel (2009) should be revised to reflect the changes in the efficient price thus the underlying information flow. Our benchmark measure for price discovery is in the same spirit as Hasbrouck’s information share. We identify conditions required for the WPC to be close to the IS, i.e. normally distributed returns with small mean and small serial correlation. Since these conditions are not easily satisfied by financial asset returns, the IS should be the preferred measure for price discovery. It measures the variation in the efficient price and the structural VAR is easy to estimate. As pointed out in Wang and Yang (2009), realized variance should be used if intraday trade-by-trade data are available. Our analysis is based on sequential trading periods or markets. Future research should explore the performance of the WPC when trading takes place simultaneously in parallel markets.
Appendix: Proof of Equation (8)

Following the definitions and notations in sections II and III, we aim to find expressions for $E(|r_t|)$ and $E[\text{sign}(r_t)r_{i,t}]$, where $\text{sign}(\cdot)$ is the sign function. We can write $E(|r_t|)$ as

$$E(|r_t|) = E[r_t I(r_t > 0)] - E[r_t I(r_t \leq 0)]$$

$$= E(r_t) - 2E[r_t I(r_t \leq 0)]$$

$$= 2 \mu \left[ 1 - \Phi \left( -\frac{\mu}{\sigma} \right) \right] + \sqrt{\frac{2}{\pi}} \exp \left( -\frac{\mu^2}{2\sigma^2} \right) \sigma,$$

where $I(\cdot)$ is the indicator function. Similarly

$$E[\text{sign}(r_t)r_{i,t}] = E[\text{sign}(r_t)r_{i,t} I(r_t > 0)] + E[\text{sign}(r_t)r_{i,t} I(r_t \leq 0)]$$

$$= E[r_{i,t} I(r_t > 0)] - E[r_{i,t} I(r_t \leq 0)]$$

$$= E(r_{i,t}) - 2E[r_{i,t} I(r_t \leq 0)].$$

Define $\mu_{i|-i} = \mu_i + (\rho \sigma_i / \sigma_-)(r_{-i,t} - \mu_{-})$ as the conditional mean of $r_{i,t}$ given $r_{-i,t}$. Using the identity

$$r_{i,t} = \mu_i + \left( r_{i,t} - \mu_{i|-i} \right) + (\rho \sigma_i / \sigma_-)(r_t - \mu) \right] / (1 + \rho \sigma_i / \sigma_-),$$

we find

$$E[r_{i,t} I(r_t \leq 0)] = \mu_i E[I(r_t \leq 0)] + \frac{E[(r_{i,t} - \mu_{i|-i})I(r_t \leq 0)] + (\rho \sigma_i / \sigma_-)E[(r_t - \mu)I(r_t \leq 0)]}{[1 + (\rho \sigma_i / \sigma_-)]}.$$

Therefore we have

$$E[I(r_t \leq 0)] = \Phi(-\mu/\sigma),$$

$$E \left[ (r_{i,t} - \mu_{i|t}) I(r_t \leq 0) \mid r_{-i,t} \right] = -\left( \frac{(1-\rho^2)\sigma^2}{2\pi} \right)^{1/2} \exp \left( -\frac{(r_{i,t} - \mu_{i|t})^2}{2(1-\rho^2)\sigma^2} \right),$$

$$E \left[ (r_{i,t} - \mu_{i|-i}) I(r_t \leq 0) \right] = -\frac{(1-\rho^2)\sigma^2}{\sqrt{2\pi} \sigma} \exp \left( -\frac{\mu^2}{2\sigma^2} \right),$$

$$E[(r_t - \mu) I(r_t \leq 0)] = -2\sigma \exp \left( -\frac{\mu^2}{2\sigma^2} \right),$$

where the last expression illustrates how these expectations are evaluated. Finally, putting the above together, we obtain

$$E[r_{i,t} I(r_t \leq 0)] = \mu_i \Phi \left( -\frac{\mu}{\sigma} \right) - \frac{\sigma^2 + \rho \sigma_i \sigma_-}{\sqrt{2\pi} \sigma} \exp \left( -\frac{\mu^2}{2\sigma^2} \right).$$

$$E[\text{sign}(r_t)r_{i,t}] = 2\mu_i \left[ 1 - \Phi \left( -\frac{\mu}{\sigma} \right) \right] + \sqrt{\frac{2}{\pi}} \exp \left( -\frac{\mu^2}{2\sigma^2} \right) \frac{\sigma^2 + \rho \sigma_i \sigma_-}{\sigma}.$$

Equation (8) is given by $\frac{E[\text{sign}(r_t)r_{i,t}]}{E(|r_t|)}$. 

13
Reference


Table 1: Overnight and Daytime Returns of the S&P500 Index

In this table, “$r_N$” and “$r_D$” are the average overnight and daytime return respectively; “$\sigma_N$” and “$\sigma_D$” are the overnight and daytime return volatility respectively; “$r_{\text{year}}$” is the annual index return.

<table>
<thead>
<tr>
<th>Year</th>
<th>$r_N$ (bp)</th>
<th>$\sigma_N$ (%)</th>
<th>$r_D$ (bp)</th>
<th>$\sigma_D$ (%)</th>
<th>$r_{\text{year}}$ (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>0.25</td>
<td>0.03</td>
<td>2.48</td>
<td>0.54</td>
<td>6.9</td>
</tr>
<tr>
<td>1994</td>
<td>0.11</td>
<td>0.04</td>
<td>-0.72</td>
<td>0.62</td>
<td>-1.6</td>
</tr>
<tr>
<td>1995</td>
<td>0.01</td>
<td>0.08</td>
<td>11.63</td>
<td>0.49</td>
<td>29.4</td>
</tr>
<tr>
<td>1996</td>
<td>0.48</td>
<td>0.06</td>
<td>6.79</td>
<td>0.74</td>
<td>18.5</td>
</tr>
<tr>
<td>1997</td>
<td>0.26</td>
<td>0.03</td>
<td>10.42</td>
<td>1.14</td>
<td>27.0</td>
</tr>
<tr>
<td>1998</td>
<td>0.50</td>
<td>0.04</td>
<td>8.88</td>
<td>1.28</td>
<td>23.6</td>
</tr>
<tr>
<td>1999</td>
<td>2.82</td>
<td>0.10</td>
<td>4.26</td>
<td>1.11</td>
<td>17.8</td>
</tr>
<tr>
<td>2000</td>
<td>1.74</td>
<td>0.28</td>
<td>-5.98</td>
<td>1.34</td>
<td>-10.7</td>
</tr>
<tr>
<td>2001</td>
<td>-2.72</td>
<td>0.42</td>
<td>-2.91</td>
<td>1.27</td>
<td>-14.0</td>
</tr>
<tr>
<td>2002</td>
<td>-6.80</td>
<td>0.80</td>
<td>-3.76</td>
<td>1.43</td>
<td>-26.6</td>
</tr>
<tr>
<td>2003</td>
<td>3.16</td>
<td>0.52</td>
<td>6.14</td>
<td>0.98</td>
<td>23.4</td>
</tr>
<tr>
<td>2004</td>
<td>1.16</td>
<td>0.34</td>
<td>2.26</td>
<td>0.63</td>
<td>8.6</td>
</tr>
<tr>
<td>2005</td>
<td>4.01</td>
<td>0.25</td>
<td>-2.84</td>
<td>0.59</td>
<td>3.0</td>
</tr>
<tr>
<td>2006</td>
<td>-0.47</td>
<td>0.05</td>
<td>5.56</td>
<td>0.63</td>
<td>12.8</td>
</tr>
<tr>
<td>2007</td>
<td>0.05</td>
<td>0.11</td>
<td>1.34</td>
<td>0.98</td>
<td>3.5</td>
</tr>
<tr>
<td>2008</td>
<td>-1.77</td>
<td>0.28</td>
<td>-17.44</td>
<td>2.46</td>
<td>-48.6</td>
</tr>
<tr>
<td>2009</td>
<td>-2.63</td>
<td>0.26</td>
<td>10.99</td>
<td>1.59</td>
<td>21.1</td>
</tr>
<tr>
<td>Full</td>
<td>0.01</td>
<td>0.30</td>
<td>2.19</td>
<td>1.16</td>
<td>94.1</td>
</tr>
</tbody>
</table>
Table 2: Overnight Price Discovery for the S&P500 Index

In this table, “Cor(r_N,r_D)” is the correlation between overnight and daytime returns; “SVar Lags” is the number of lags in the structural VAR model based on the Schwarz criterion; “IS_N” and “WPC_N” are the overnight information share and the overnight weighted price contribution respectively. The bold numbers represent large values of Cor(r_N,r_D) and large differences between the IS and the WPC.

<table>
<thead>
<tr>
<th>Year</th>
<th>Cor(r_N,r_D)</th>
<th>SVar Lags</th>
<th>IS_N (%)</th>
<th>WPC_N (%)</th>
<th>IS_N-WPC_N (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1993</td>
<td>0.185</td>
<td>1</td>
<td>5.2</td>
<td>0.6</td>
<td>4.6</td>
</tr>
<tr>
<td>1994</td>
<td>0.075</td>
<td>1</td>
<td>0.8</td>
<td>1.3</td>
<td>-0.5</td>
</tr>
<tr>
<td>1995</td>
<td>-0.072</td>
<td>1</td>
<td>0.04</td>
<td>1.8</td>
<td>-1.7</td>
</tr>
<tr>
<td>1996</td>
<td>0.030</td>
<td>1</td>
<td>1.0</td>
<td>1.3</td>
<td>-0.3</td>
</tr>
<tr>
<td>1997</td>
<td>0.098</td>
<td>1</td>
<td>2.8</td>
<td>0.6</td>
<td>2.2</td>
</tr>
<tr>
<td>1998</td>
<td>0.030</td>
<td>1</td>
<td>0.2</td>
<td>0.5</td>
<td>-0.3</td>
</tr>
<tr>
<td>1999</td>
<td>0.241</td>
<td>3</td>
<td>12.1</td>
<td>3.1</td>
<td>9.0</td>
</tr>
<tr>
<td>2000</td>
<td>0.111</td>
<td>2</td>
<td>3.9</td>
<td>6.4</td>
<td>-2.5</td>
</tr>
<tr>
<td>2001</td>
<td>0.054</td>
<td>2</td>
<td>11.1</td>
<td>11.8</td>
<td>-0.7</td>
</tr>
<tr>
<td>2002</td>
<td>-0.006</td>
<td>1</td>
<td>17.7</td>
<td>23.4</td>
<td>-5.7</td>
</tr>
<tr>
<td>2003</td>
<td>-0.068</td>
<td>1</td>
<td>17.5</td>
<td>18.0</td>
<td>-0.4</td>
</tr>
<tr>
<td>2004</td>
<td>-0.048</td>
<td>1</td>
<td>35.6</td>
<td>22.8</td>
<td>12.7</td>
</tr>
<tr>
<td>2005</td>
<td>0.013</td>
<td>1</td>
<td>17.0</td>
<td>14.2</td>
<td>2.9</td>
</tr>
<tr>
<td>2006</td>
<td>-0.016</td>
<td>1</td>
<td>1.2</td>
<td>0.9</td>
<td>0.3</td>
</tr>
<tr>
<td>2007</td>
<td>0.197</td>
<td>7</td>
<td>8.9</td>
<td>3.9</td>
<td>5.0</td>
</tr>
<tr>
<td>2008</td>
<td>0.380</td>
<td>2</td>
<td>11.9</td>
<td>5.0</td>
<td>6.8</td>
</tr>
<tr>
<td>2009</td>
<td>0.419</td>
<td>1</td>
<td>32.4</td>
<td>7.5</td>
<td>24.9</td>
</tr>
<tr>
<td>Full</td>
<td>0.088</td>
<td>2</td>
<td>10.1</td>
<td>7.9</td>
<td>2.1</td>
</tr>
</tbody>
</table>
Figure 1: The WPC as a function of $\mu$ and $\rho$

The figures depict the WPC as a function of daily mean return $\mu$ and the cross-period return correlation $\rho$, i.e. equation (8). The parameters for return volatility are set as $\sigma_i = 1$ and $\sigma_{-i} = 2$. Figure 1a assumes $\mu_i = -0.2\mu$, hence $\mu_{-i} = 1.2\mu$. Figure 1b assumes $\mu_i = 0.2\mu$, hence $\mu_{-i} = 0.8\mu$. 