Measuring Price Discovery:
The Variance Ratio, the R^2, and the Weighted Price Contribution

This version: April 2011

We analyze the statistical properties of three price discovery measures: The variance ratio, the weighted price contribution (*WPC*), and the R^2 of unbiasedness regressions. We find that, if the price process is a driftless martingale, only the *WPC* is an unbiased estimator for the return variance explained during a time interval. For autocorrelated processes with a drift, only the R^2 of the unbiasedness regression is consistent, but it is biased for small samples.

JEL Classification: G14, C00, C13

Keywords: market microstructure, price discovery, weighted price contribution

1. Introduction

In financial economics we are often interested in the informativeness of time intervals or trade categories. Since equilibrium prices reflect information, the basic question is "What proportion of the return variance during a benchmark interval (typically a day) is explained during a subinterval or a subset of trades?" To measure price discovery, researchers have used several techniques. In this study we compare the variance ratio (*VR*), the R^2 of unbiasedness regressions and the weighted price contribution (*WPC*).

If prices follow martingales, the variance ratio (*VR*) is the natural measure for the variance explained during different trading periods.[1] However, if price processes exhibit autocorrelation, the variance ratio is not an appropriate statistic to gauge price discovery, as it captures both heteroskedasticity and autocorrelation. Recent research has shown that autocorrelation of transaction prices or quote midpoints can be significant. Negative autocorrelation during shorter intervals may persist in equilibrium due to the bid-ask bounce (Roll, 1984), positive autocorrelation of can obtain due to price discreteness and market maker inventory control (see, e.g., Hasbrouck (1988, 1991) and Stoll (1989)) Variance ratios are now mainly used to assess autocorrelation, not heteroskedasticity or explanatory power (see, e.g. Lo and MacKinley, 1989).

An alternative method to gauge price discovery was developed by Barclay and Warner (1993). Interested in the informativeness of trades of different sizes, they compute, for trades within different size categories, the total price change as a proportion of the day´s

[1] Fama (1965) concludes, based on subperiod variances, that during daytime trading more information is incorporated into stock prices than overnight, while French (1980) finds that weekends have relatively little information content. Other studies that analyze in variance ratios include Stoll and Whaley (1990), French and Roll (1986), Harris (1986), Barclay *et al.* (1990), and Ederington and Lee (1993), among others.

price change. They then take the weighted average of these price contributions over the sample period, with the weights being the absolute values of the day´s price changes, so as to dampen the influence of extreme values.[2] Cao, *et al.* (2000) formally defined the resulting estimator and dubbed it the *weighted price contribution* (*WPC*). They used it to gauge the extent of price discovery during the mock trading that used to precede the Nasdaq trading day. The *WPC* has since been used by many researchers.[3] Conspicuously, the statistical properties of the *WPC* estimator have not yet been analyzed.

A final estimator for the explanatory power of a subset of trades is the R^2 of a regression of the total price change on the price change due to particular trades or during certain time periods. This method was developed by Biais *et al.* (1999), who analyze both the slope and the root mean standard errors (*RMSE*s) of these *unbiasedness regressions*.[4] Subsequent studies that used unbiasedness regressions to analyze price discovery, such as those by Cushing and Madhavan (2000), Madhavan and Panchapagesan (2000), Corwin and Lipson (2000), and Chakrabarty *et al.* (2009), among others, all reported the R^2 instead of the *RMSE*.[5]

[2] They find that approximately 95% of the price change prior a tender offer is due to medium sized trades. Because only 42% of all trades and 60% of the dollar turnover came from the medium size category, they interpreted their finding as evidence for the stealth trading hypothesis, which says that informed traders break up their trades so that they become medium sized.

[3] Other papers who use the WPC include Huang (2002), Bacidore and Lipson (2001), Barclay and Hendershott (2004, 2008), Chakraborty *et al.* (2009) and Mayhew *et al.* (2009), among others.

[4] They regress close-to-close returns on close-to-t returns, and find a slope coefficient that gradually increases (to unity) in t, and an *RMSE* that gradually decreases (implying an increasing R^2).

[5] Notice that for univariate OLS regressions we have $R^2 = 1 - \frac{RMSE^2}{RMST^2}$, where the *RMST* denotes the root of the mean squared totals, so that the *RMSE* is an inverse measure of explanatory power.

In this paper we analyze the statistical properties of the three above mentioned estimators and several variants, under varying assumptions for the price process. We analytically show that if the underlying process is a driftless (heteroskedastic) martingale, all three estimators are consistent.

However, for small samples the *VR* and R^2 are (upward) biased, while the *WPC* is unbiased. We show that the *WPC* gives an unbiased estimate even if we have only two observations, in which case the *VR* and the R^2 are meaningless. An additional advantage of the *WPC* is that its distribution approaches the Normal much faster than the *VR* and the R^2, which have severely skewed distributions for small samples. The advantages of the *WPC* do not come at a high cost. Using Monte Carlo analyses we find that the precision of the *WPC* is only slightly smaller than that of the *VR* and higher than that of the R^2.

When we consider processes with a drift and/or autocorrelation, we find that the *WPC* is an inconsistent and downward biased estimator of the explanatory power of designated trades or intervals. The bias can be remedied by demeaning the independent variables in the *WPC* calculation. The resulting estimator, the *WPC*1, is a consistent and unbiased estimator for martingales with a drift.

If the underlying process is driftless but exhibits positive (negative) autocorrelation, both the *VR* and the *WPC* are inconsistent and biased, although the bias of the *WPC* is less severe than that of the *VR*.

Notice that our concept of price discovery is different from the information share measure developed in Hasbrouck (1995, 2002) employed by Theissen (2002), Hasbrouck (2003) and Yan and Zivot (2010), among others. The information share recognizes that transaction prices chase an unobservable value (often denoted the *efficient price*), but deviate from it due to microstructure noise and/or asymmetric information. The information share is defined as the proportion of the variance in *value* explained by a one of several simultaneous price processes, and estimated with a VAR regression, which assumes that the price processes are covariance stationary. When we are interested in systematic differences in (co)variances between intervals or trade categories, VAR regressions are unsuitable.

In the next section we define the estimation problem and show analyze the price discovery estimators properties in the case of a driftless (but heteroskedastic) martingale, In section three we consider all stochastic processes, and present the results of a Monte Carlo experiment, and section four concludes.

2. Problem definition and the martingale case

Consider a stochastic process $X(t)$ defined over a small period such as a day. We denote the beginning and end of this period $t = 0$ and $t = T$. Let τ be a set of M non-overlapping intervals $\{(t_{1b},t_{1e}], (t_{2b},t_{2e}],..,(t_{Mb},t_{Me}]\}$ with $0 \le t_{1b} < t_{1e} <\ldots< t_{Me} \le T$. Define the random variable $\tilde{x} \equiv X(t_{1e}) - X(t_{1b}) +\ldots+ X(t_{Me}) - X(t_{Mb})$, the cumulated X-differential attributed to τ, $\tilde{z} \equiv X(t_{1b}) - X(0) +\ldots+ X(T) - X(t_{Me})$, the X-differential achieved outside τ, and $\tilde{y} \equiv X(T) - X(0)$, the change of X during the interval $(0,T)$, so that $\tilde{y} = \tilde{x} + \tilde{z}$. In many application τ is a single subinterval of $(0,T)$.

In the following we will assume that $\tilde{x}$ and $\tilde{z}$ follow Normal distributions with expected values μ_x and μ_z and standard deviations σ_x and σ_z respectively. Apart from allowing X to have interval dependent drifts (μ_x and μ_z), the process may be autocorrelated. If we denote ρ the correlation between $\tilde{x}$ and $\tilde{z}$, we have $\tilde{y} \sim N\left(\mu_x + \mu_z, \sigma_x^2 + \sigma_z^2 + 2\rho\sigma_x\sigma_z\right)$.

Our assumption that price changes are Normally distributed is for convenience only. It can be shown that our results hold if $\tilde{x}$ and $\tilde{z}$ follow mixtures of Normal distributions, such as those that are generated by jump-diffusion processes that lead to kurtosis.

We are interested in the proportion of the variation of $\tilde{y}$ that can be explained by the variation during τ, $\tilde{x}$. Or, our estimant of interest is the following coefficient of determination:

$$\theta = 1 - \frac{E\left[(\tilde{y} - E[\tilde{y}|\tilde{x}])^2 | \tilde{x}\right]}{E\left[(\tilde{y} - E[\tilde{y}])^2\right]} = 1 - \frac{(1-\rho^2)\sigma_z^2}{\sigma_y^2} \tag{1}$$

The second equality follows from the projection theorem. The quotient´s numerator is the unexplained variance from the linear projection of $\tilde{y}$ on $\tilde{x}$, so that the former variable can be written as $\tilde{y} = \alpha + \beta\tilde{x} + \tilde{\varepsilon}$, where $\alpha = \mu_z - \rho\frac{\sigma_z}{\sigma_x}\mu_x$, $\beta = 1 + \rho\frac{\sigma_z}{\sigma_x}$, and the variance of the independent error term $\tilde{\varepsilon}$ is $(1-\rho^2)\sigma_z^2$.

The general problem is that we do not know the parameters $\{\mu_x, \mu_z, \sigma_x, \sigma_y, \rho\}$ and have to estimate them. In the following we assume that we have *N* independent observations of $\tilde{x}$ and $\tilde{y}$ to estimate θ. We shall denote these observations x_i and y_i respectively.

Clearly, if $X(t)$ is a martingale, $\rho = 0$ and $\theta = \frac{\sigma_x^2}{\sigma_y^2}$, so that a natural estimator is the variance ratio given by:

$$VR = \frac{\sum_{i=1..N}(x_i - \bar{x})^2}{\sum_{i=1..N}(y_i - \bar{y})^2} \tag{2}$$

The well known problem with this estimator is that its distribution is severely skewed, and moreover, biased. This is because *VR* is a quotient of two random variables. Naturally, if $X(t)$ exhibits autocorrelation ($\rho \neq 1$), *VR* is inconsistent.

If we are not sure whether $X(t)$ follows a martingale, a natural estimator for θ is the R^2 of a linear regression of $\{y_i\}$ on $\{x_i\}$. We know that if we do not constrain the intercept to be zero, this estimator is given by:[6]

$$R^2 = \frac{\left(\sum_{i=1..N}(x_i - \bar{x})(y_i - \bar{y})\right)^2}{\sum_{i=1..N}(x_i - \bar{x})^2 \sum_{i=1..N}(y_i - \bar{y})^2} \tag{3}$$

From the econometrics literature we know that also this estimator is biased. To alleviate the R^2 bias, several shrinkage estimators have been suggested. The most popular one is due to Ezekiel (1929), which defines the *adjusted* R^2.

$$R^2_{Adj.} = 1 - \frac{N-1}{N-1-K}\left(R^2 - 1\right), \tag{4}$$

where K is the number of independent variables. In our case $K = 1$.[7] In the current paper we only consider the R^2 and *adjusted* R^2 given by (6).

[6] If we constrain the intercept to be zero, we have $R^2 = \frac{\left(\sum_{i=1..N} x_i y_i\right)^2}{\sum_{i=1..N} x_i^2 \sum_{i=1..N} y_i^2}$

[7] Alternative shrinkage estimators are due Ezekiel (1930), Wherry (1931), Darlington (1968), and others. Olkin and Pratt (1958) identify the unique minimum variance unbiased estimator of R2, which is impractical as it involves a complex hypergeometric function. See Yin and Fan (2001), for an analysis and comparison of different shrinkage estimators.

The final estimator for θ that we consider in this paper is the *weighted price contribution* (*WPC*), which is defined as follows:

$$WPC = \sum_{i=1..N} \left(\frac{|y_i|}{\sum_{j=1..N} |y_j|} \right) \frac{x_i}{y_i} \tag{5}$$

The intuition behind this estimator is that the average $\frac{x_i}{y_i}$, is a natural candidate estimator for θ. However, it suffers from infinite variance, and hence undefined mean.[8] To remedy this, each $\frac{x_i}{y_i}$ is weighted by the standardized absolute value of y_i, so that observations with small $|y_i|$ receive relatively little weight.

In the appendix we prove proposition 1, which says that for martingale processes all three estimators are consistent but only the *WPC* is unbiased.

Proposition 1 (*X is a driftless martingale*): *If* $\mu_x = \mu_z = 0$ and $\rho = 0$,

- *(i)* *The VR is the maximum likelihood estimator for* θ. *It is consistent and biased.*
- *(ii)* *The* R^2 *is a consistent and biased estimator for* θ.
- *(iii)* *The WPC is a consistent, unbiased, and asymptotically Normal estimator for* θ.

The properties of the variance ratio and R^2 are well known. The consistency and unbiasedness of the *WPC* is somewhat surprising as it is not immediately obvious from expression (5). The unbiasedness finding, which is proved by finding the expected value of the *WPC* through integration (see appendix), is important: no other unbiased coefficient of determination for a linear regression model exists. The *WPC*′s distributional property, which follows from the Central Limit Theorem, is desirable for

[8] This is because x_i/y_i follows a Cauchy distribution.

hypothesis testing. Moreover, from simulations we find that the estimator´s distribution reaches Normality much faster than the R^2.

Unfortunately, the consistency and unbiasedness do not carry over to autocorrelated processes, which we discuss in the next section.

3. A coefficient of determination for non-martingales.

In this section we consider price process X that are not martingales, that is μ_x, μ_z, $\rho \neq 0$. The properties of the first two estimators are well known: the *VR* is inconsistent due to the non-zero ρ, while the R^2 is consistent but biased. By evaluating the expectation of the *WPC*, we find it to be biased and inconsistent.

Proposition 2 (*Non-martingales*). *For the three* θ *estimators, we have:*

(i) If $\rho = 0$ (*X is a martingale with a drift*), *the VR is consistent but biased.*

(ii) If $\rho \neq 0$ (*X is autocorrelated*), *the VR is inconsistent.*

(iii) For all cases, the R^2 *is consistent but biased.*

(iv) If $\rho = 0$ (*X is a martingale with a drift*), *the WPC is inconsistent.*

(v) If $\rho \neq 0$ (*X is autocorrelated*), *the WPC is inconsistent.*

Although the presence of a drift (μ_z, $\mu_x \neq 0$) does not change the properties of the *VR* estimator, autocorrelation ($\rho \neq 0$) is not picked up by the *VR*. To see this, notice that *VR* converges to $\frac{\sigma_x^2}{\sigma_y^2}$, while we know $\theta = 1 - \frac{(1-\rho^2)\sigma_z^2}{\sigma_y^2} = \frac{\sigma_x^2 + 2\rho\sigma_x\sigma_z + \rho^2\sigma_z^2}{\sigma_y^2}$. Hence, the asymptotic bias of the *VR* is negative (positive) for $\rho >$ (<) 0.

For the *WPC*, both the drift and autocorrelation enters the estimator. In the appendix we show that if $\rho = \mu_x = 0$, $\mu_z \neq 0$ the *WPC* is downward biased. If $\rho = \mu_z = 0$, $\mu_x \neq 0$ the *WPC* is upward biased. It can also be shown that $\lim_{N \to \infty} E[WPC|\mu_x = \mu_z = 0]$ = $\frac{\sigma_x^2 + \rho\sigma_z\sigma_x}{\sigma_y^2}$. Since $\theta = 1 - \frac{(1-\rho^2)\sigma_z^2}{\sigma_y^2} = \frac{\sigma_x^2 + 2\rho\sigma_x\sigma_z + \rho^2\sigma_z^2}{\sigma_y^2}$, we find the direction of the bias.[9]

The analysis also shows that we can correct the *WPC* bias that is due to drifts (of either *x* or *z*), can be remedied by defining demeaning the variables that make up the *WPC*, as follows:

$$WPC1 = \sum_{i=1..N} \left(\frac{|y_i - \bar{y}|}{\sum_{j=1..N} |y_j - \bar{y}|} \right) \frac{x_i - \bar{x}}{y_i - \bar{y}} \tag{6}$$

In the appendix we proof that:

Proposition 3: *For martingales with a drift* (μ_x, $\mu_z \in \mathbb{R}$, $\rho = 0$), *the WPC*1 *given by* (6) *is a consistent, unbiased, and asymptotically Normal distributed estimator for* θ.

3. Monte Carlo Analysis.

To better understand the distributional properties of the estimators, we conduct a Monte Carlo experiment. We generate *N* random vectors $\begin{pmatrix} x_i \\ y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \mu_x \\ \mu_z \end{pmatrix}, \begin{pmatrix} \sigma_x^2 & \rho\sigma_x\sigma_z \\ \rho\sigma_x\sigma_z & \sigma_z^2 \end{pmatrix} \right)$, set $y_i = x_i + z_i$, and compute the different coefficients of determination. We do this 1,000,000

[9] Notice that when $\sigma_z < 2\sigma_x$, we have that the bias of the WPC is positive (negative) *iff* ρ < (>) 0. If however $\sigma_z > 2\sigma_x$, the bias can be negative for $\rho << 0$.

times, for different combinations of μ_x, μ_z, and ρ, and various values of N. Throughout we set $\sigma_x = 1$ and $\sigma_z = 2$. Table 1 gives the result so this analysis.

--- Table 1 around here ---

From panel A we see that, if the price process follows a driftless martingale, only the WPC and the WPC1 are unbiased, that *VR* and R^2 are biased upward and that the adjusted R^2 overcompensates. Consistency of the estimators is evident from the fact that the bias decreases as N increases. From the standard deviation of the estimator population we also see that for most N, the standard error of the *VR* is the smallest while that of the R^2 is the largest. We attribute this to the fact that the R^2 assumes drifts and a non-zero correlation, introducing noise in the estimate. However, we find that for very small N the standard error of *VR* becomes the highest, and that of the R^2 the lowest. The reason for this is that now the additional noise in the R^2 is good, as it dampens the huge kurtosis that the *VR* needs to cope with. Clearly for $N = 2$ (not reported) the *VR* has infinite variance, while the R^2 would always be unity. The *WPC* however is meaningful (and unbiased) even for $N = 2$.

From panels B and C we see that if the price follows a martingale with a drift, the *WPC*1 is the only unbiased estimator. Notice that the direction of the sign of the drift does not affect the bias and precision of the estimator. In panels D and E we assess the bias due to autocorrelation. Consistent with our analysis, all statistics are biased and only the R^2 and the adjusted R^2 are consistent. For $\rho = 0.25$, The *VR* converges to $\frac{\sigma_x^2}{\sigma_y^2} = \frac{1}{6}$, and the *WPC* to $\frac{\sigma_x^2 + \rho\sigma_z\sigma_x}{\sigma_y^2} = \frac{1}{4}$.

Table 1 also gives, where relevant, the percentage of negative observations. Both the *WPC* and the adjusted R^2 can be negative. Negative values become however extremely unlikely for samples greater than $N = 100$, so that in general this should not be a problem. In any case, we find the negative value problem to be more severe for the R^2 than for the *WPC*. If we *do* have small sample sizes, we may want to further adjust the *WPC* measure to assure positive values.

--- Figure 1 around here ---

In Figure 1 we display the estimators asymptotic distributions for the driftless martingale case and $N = 10$ and $N = 100$. We observe that, for low *N*, the distributions of *VR* and R^2 are significantly skewed and non-Normal. The range of the *VR* is $\mathbb{R}^+$, that of the R^2 is [0,1]. The *WPC*′s range is $\mathbb{R}$, and its distribution is not visibly different from the Normal Distribution. The $N = 10$ panel also illustrates the problem of the adjusted R^2: Its distribution is highly non-Normal and it returns a high proportion of negative numbers. With increasing *N*, the *VR* and R^2 distributions are still significantly non-Normal. Their skewness coefficients are 0.56 and 0.26 respectively. For the WPCs it is not significantly different from zero.

4. Summary and Conclusions

In this paper we compare the statistical properties the variance ratio (*VR*), the R^2, and the weighted price contribution (*WPC*), as measures of price discovery. We find that if the price process follows a driftless martingale, the *WPC*, is the only unbiased estimator. If we the price process is autocorrelated, the *WPC* is, like all other measures, biased. More importantly, it is inconsistent. The only consistent price discovery measure is the R^2. The *WPC* is also biased if the price process follows a drift. To correct for the drift bias we propose an adjusted *WPC,* the *WPC*1, which is consistent and unbiased.

Appendix

Proof of Proposition 1(iii) - consistency of the WPC estimator

To prove unbiasedness we evaluate the expected value of the estimator for $N = 2$:

$$E[WPC_{N=2}] = E\left[\frac{|y_1|}{|y_1|+|y_2|}\cdot\frac{x_1}{y_1}+\frac{|y_2|}{|y_1|+|y_2|}\cdot\frac{x_2}{y_2}\right] = 2E\left[\frac{x_1}{|y_1|+|y_2|}\cdot\frac{|y_1|}{y_1}\right] \qquad \text{(A1)}$$

The latter inequality follows from both x_1 and x_2, and y_1 and y_2 being I.I.D. Due to symmetry and Normality we can write:

$$E[WPC] = 2\int_0^\infty\left(\int_0^\infty \frac{E\left[x_1|y_1\right]}{y_1+y_2}\frac{2}{\sigma_y\sqrt{2\pi}}e^{-\frac{y_1^2}{2\sigma_y^2}}dy_1\right)\frac{2}{\sigma_y\sqrt{2\pi}}e^{-\frac{y_2^2}{2\sigma_y^2}}dy_2 \qquad \text{(A2)}$$

Since $E\left[x_1|y_1\right] = y_1\frac{\sigma_x^2}{\sigma_x^2+\sigma_\varepsilon^2} = y_1\frac{\sigma_x^2}{\sigma_y^2}$, we have:

$$E[WPC] = \frac{4\sigma_x^2}{\sigma_y^4\pi}\int_0^\infty\int_0^\infty \frac{y_1}{y_1+y_2}e^{-\frac{y_1^2+y_2^2}{2\sigma_y^2}}dy_1dy_2 \qquad \text{(A3)}$$

This integral can be solved with polar integration: instead of integrating $dy_1{\cdot}dy_2$ over the positive quadrant of (y_1, y_2) space, we integrate $rd\theta{\cdot}dr$ over the same area, where $r \equiv \sqrt{y_1^2+y_2^2}$ and θ is the angle with the y_1 axis, so that (A3) can be written as:

$$E[WPC] = \frac{4\sigma_x^2}{\sigma_y^4\pi}\int_0^{\frac{1}{2}\pi}\int_0^\infty \frac{\cos(\theta)}{\cos(\theta)+\sin(\theta)}\cdot re^{-\frac{r^2}{2\sigma_y^2}}drd\theta \qquad \text{(A4)}$$

which becomes:

$$E[WPC] = \frac{4\sigma_x^2}{\sigma_y^4\pi}\cdot\int_0^{\frac{1}{2}\pi}\frac{\cos(\theta)}{\cos(\theta)+\sin(\theta)}d\theta\cdot\int_0^\infty re^{-\frac{r^2}{2\sigma_y^2}}dr = \frac{4\sigma_x^2}{\sigma_y^4\pi}\cdot\frac{\pi}{4}\cdot\sigma_y^2 = \frac{\sigma_x^2}{\sigma_y^2} \qquad \text{(A5)}$$

Thus proving that for N=2 the expected value of *WPC* is finite and equals θ. By repeating the argument, it can be established that also for $N > 2$ $E[WPC] = \theta = \frac{\sigma_x^2}{\sigma_y^2}$. *Q.E.D.*

The proofs of proposition 2, available upon request from the author, are based on similar analysis.

References

Bacidore, J., and Lipson, M.L. (2001), The effects of opening and closing procedures on the NYSE and Nasdaq. Unpublished Working paper.

Barclay, M.J, and Warner, J.B. (1993), Stealth trading and volatility: which trades move prices?, *Journal of Financial Economics* 34, 281-305.

Barclay, M.J, and Hendershott, T. (2004), Liquidity externalities and adverse selection: evidence from trading after hours, *Journal of Finance* 59, 681-710

Barclay, M.J, and Hendershott, T. (2008), A comparison of trading and non-trading mechanisms for price discovery, *Journal of Empirical Finance* 15, 839-849.

Barclay, M.J., Litzenberger, R.H., and Warner, J.B. (1990), Private information, trading volume, and stock return variances, *Review of Financial Studies* 3, 233-235.

Biais, B., Hillion, P., and Spatt, C. (1999), Price discovery and learning during the preopening period in the Paris Bourse, *Journal of Political Economy* 107, 1218-1248.

Cao, C., Ghysels, E., and Hathaway, F. (2000), Price discovery without trading: Evidence from the Nasdaq pre-opening, *Journal of Finance* 55, 1339-1365.

Chakrabarty, B., Corwin, S.A., and Panayides, M.A. (2009) When a halt is not a halt: an analysis of off-NYSE trading during NYSE market closures. Working paper.

Corwin, S.A., and Lipson, M.L. (2000), Order flow and liquidity around NYSE trading halts, *Journal of Finance* 55 1771-1801.

Cushing, D., and Madhavan, A. (2000), Stock returns and trading at the close, *Journal of Financial Markets* 3, 45-67.

Darlington, R.B. (1968), Multiple regression in psychological research and practice, *Psychological Psychological Bulletin* 69, 161-182.

Ederington, L.H., and Lee, J.H. (1993), How markets process information: News releases and volatility, *Journal of Finance* 48, 1161-1191.

Ezekiel, M. (1929), The application of the theory of error to multiple and curvilinear correlation, *American Statistical Association Journal* 24, 99-104

Ezekiel, M. (1930), Methods of correlational analysis. New York: Wiley

Fama, E.F. (1965), The behavior of stock market prices, *Journal of Business* 28, 34-105.

French, K. (1980), Stock returns and the weekend effect, *Journal of Financial Economics* 8, 55-70.

French, K., and Roll, R. (1986), Stock return variances: the arrival of information and the reaction of traders, *Journal of Financial Economics* 17, 5-26.

Harris, L. (1986), A transaction data study of weekly and intradaily patterns in stock returns, *Journal of Financial Economics* 16, 99-117.

Hasbrouck, J. (1988), Trades, quotes, inventory and information, *Journal of Financial Economics*, 22, 229-252.

Hasbrouck, J.. (1991), Measuring the information content of stock trades, *Journal of Finance*, 46, 179-207.

Hasbrouck, J. (1995), One Security, Many Markets: Determining the Contributions to Price Discovery, *Journal of Finance*, 50, 4, 1175-1199.

Hasbrouck, J. (2002), Stalking the "efficient price" in market microstructure specifications: an overview, *Journal of Financial Markets*, 5, 1-19.

Hasbrouck, J. (2003) Intraday price formation in U.S. equity markets, *Journal of Finance*, 58, 2375-2400.

Lo, A.W., and MacKinlay, A.G. (1989), The size and power of the variance ratio test in finite samples: A Monte Carlo investigation, *Journal of Econometrics* 40, 203-244.

Madhavan, A., and Panchapagesan, V. (2000), Price discovery and auction markets: A look inside the black box, *Review of Financial Studies* 13, 627-658.

Mayhew, S., McCormick,T. , and Spatt, C. (2009) The information content of market-on-close imbalances, the specialist and NYSE equity prices, unpublished working paper.

Olkin, I., and Pratt, J.W. (1958), Unbiased estimation of certain correlation coefficients, *The annals of mathematical statistics* 29, 201-211.

Roll, R. (1984), A simple implicit measure of the effective bid-ask spread in an efficient market, *Journal of Finance*, 39, 1127-1139.

Stoll,H.R. (1989), Inferring the components of the bid-ask spread: Theory and empirical tests, *Journal of Finance*, 44, 115-134.

Stoll, H.R., and Whaley, R.E. (1990), Stock market structure and volatility, *Review of Financial Studies*, 3, 37-71.

Theissen, E. (2002), Price discovery in screen and floor trading systems, *Journal of Empirical Finance*, 9, 455-474.

Wherry, R.J. (1931), A new formula for predicting the shrinkage of the coefficient of multiple correlation, *Annals of mathematical statistics* 2, 440-451.

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

Yin P., and Fan, X. (2001), Estimating R^2 shrinkage in multiple regression: A comparison of different analytical methods, *Journal of Experimental Education* 69, 203-244.

Table 1: Monte Carlo Analysis

We generate 1,000,000 samples of N random numbers $x_i \sim N(\mu_x, 1)$ and $z_i \sim N(\mu_z, 4)$, with correlation coefficient ρ, and compute $y_i = x_i + z_i$. This table gives the average values of the estimators, and their standard deviations for various values of μ_x, μ_z, ρ, and N. Asterisks indicate difference from the true value θ at the 1% significance level.

	N		*VR*	R^2	*Adj.* R^2	*WPC*	*WPC* 1
$\mu_x = 0.2$, $\mu_z = 0$, $\rho = 0$, ($\theta = 0.2$)	1,000	Mean	0.2003 **	0.2005 **	0.1997 **	0.2000	0.2000
		St.Dev.	0.011	0.023	0.023	0.016	0.016
	100	Mean	0.203 **	0.205 **	0.197 **	0.2000	0.2000
		St.Dev.	0.037	0.071	0.072	0.051	0.051
		% neg.	-	-	0.09	0.05	0.06
	10	Mean	0.246 **	0.261 **	0.169 **	0.2000	0.2000
		St. Dev.	0.184	0.210	0.236	0.174	0.182
		% neg.	-	-	31.43	11.33	12.18
$\mu_x = 0.2$, $\mu_z = 0$, $\rho = 0$, ($\theta = 0.2$)	1,000	Mean	0.2002 **	0.2005 **	0.1997 **	0.2062 **	0.2000
		St.Dev.	0.011	0.023	0.023	0.016	0.016
	100	Mean	0.203 **	0.205 **	0.197 **	0.2064 **	0.2000
		St.Dev.	0.037	0.071	0.072	0.051	0.051
		% neg.	-	-	0.10	0.04	0.06
	10	Mean	0.246 **	0.261 **	0.169 **	0.2063 **	0.2000
		St. Dev.	0.184	0.210	0.236	0.176	0.182
		% neg.	-	-	31.37	10.90	12.16
$\mu_x = 0$, $\mu_z = 0.2$, $\rho = 0$, ($\theta = 0.2$)	1,000	Mean	0.2003 **	0.2005 **	0.1997 **	0.1984 **	0.2000
		St.Dev.	0.011	0.023	0.023	0.016	0.016
	100	Mean	0.203 **	0.205 **	0.197 **	0.1985 **	0.2000
		St.Dev.	0.037	0.071	0.071	0.051	0.051
		% neg.	-	-	0.10	0.07	0.08
	10	Mean	0.246 **	0.261 **	0.169 **	0.1982 **	0.2000
		St. Dev.	0.184	0.210	0.236	0.175	0.182
		% neg.	-	-	31.36	11.38	12.13
$\mu_x = 0$, $\mu_z = 0$, $\rho = 0.25$, ($\theta = 0.375$)	1,000	Mean	0.1669 **	0.3754 **	0.3748 **	0.2500 **	0.2500 **
		St.Dev.	0.008	0.024	0.024	0.013	0.013
	100	Mean	0.169 **	0.377 **	0.370 **	0.250 **	0.250 **
		St.Dev.	0.027	0.076	0.077	0.041	0.041
	10	Mean	0.196 **	0.402 **	0.327 **	0.250 **	0.250 **
		St.Dev.	0.128	0.228	0.256	0.141	0.147
		% neg.	-	-	13.12	3.79	4.30
$\mu_x = 0$, $\mu_z = 0.2$, $\rho = -0.25$, $\theta = 0.0625$	1,000	Mean	0.2505 **	0.0633 **	0.0624 **	0.1250 **	0.1250 **
		St.Dev.	0.015	0.015	0.015	0.019	0.019
	100	Mean	0.255 **	0.071 **	0.061 **	0.125 **	0.125 **
		St.Dev.	0.050	0.047	0.048	0.061	0.061
		% neg.	-	-	6.27	2.11	2.09
	10	Mean	0.316 **	0.155 **	0.051 **	0.125 **	0.125 **
		St.Dev.	0.258	0.167	0.188	0.211	0.220
		% neg.	-	-	53.28	26.01	26.74

Figure 1: Empirical distribution of coefficient of determination estimators

We generate 1,000,000 samples of N independent random numbers $x_i \sim N(0, 1)$ and $z_i \sim N(0, 4)$, and compute $y_i = x_i + z_i$. We display the empirical distributions of the VR, R^2, the adjusted R^2, the WPC and the $WPC1$, for $N = 10$ and $N = 100$.